UNIVERSIDADE DE LISBOA
FACULDADE DE CIÊNCIAS
DEPARTAMENTO DE FÍSICA

A generative adversarial network approach to synthetic-CT creation for MRI-based radiation therapy

Mariana Ferreira Teixeira da Silva

Mestrado Integrado em Engenharia Biomédica e Biofísica
Perfil em Radiações em Diagnóstico e Terapia

Dissertação orientada por:
Dr. Nuno Matela
Dr. Raj Jena

2019

Acknowledgements

This project would not have been possible without the financial support of the European Union and the University of Lisbon through the Erasmus+ programme.

I want to start by expressing my gratitude to Dr. Raj Jena from the University of Cambridge for his supervision, for allowing me to work independently while also being promptly available to provide advice, motivation, and expertise to allow this project to be brought to success, and for providing me with the opportunity to present my work to multiple audiences. I would like to thank Karl Harrison for his constructive suggestions during the entire development of the project, and help with computational resources. Feedback given by Dr. David Noble and Megan Wilson during our small group meetings was greatly appreciated.

I would like to thank Professor Nuno Matela for his comments and guidance, not only during the writing of this dissertation, but also in various course modules that I had the pleasure to attend during my years at the Faculty of Sciences of the University of Lisbon.

I would like to extend my thanks to my open-plan office buddies at the Maxwell Centre for helping me on a daily basis, on work and beyond, and to all my friends at IBEB and FCUL for all their much appreciated support.

Finally, I wish to thank my family for providing me with continuous encouragement throughout my studies and through the process of researching and writing this dissertation.

Resumo

Em radioterapia, a atual norma clínica passa pela utilização de imagens médicas multimodais para o planeamento do tratamento. A tomografia computadorizada (CT, do inglês *computed tomography*) é geralmente utilizada como modalidade principal, sendo necessária para efeitos de cálculo de dose, uma vez que fornece informação acerca da atenuação dos tecidos, representada pelas unidades de Hounsfield (HU, do inglês *Hounsflied units*). A imagem por ressonância magnética (MRI, do inglês *magnetic ressonance imaging*) é utilizada como modalidade secundária, e para vários locais do corpo é a modalidade de preferência para deliniação do tumour e órgãos circundantes, devido ao alto contraste de tecidos moles que fornece. Com o avanço da tecnologia MRI em radioterapia, tem existido um crescente interesse no seu uso como modalidade única para planeamento e re-planeamento inter- e intra-fracionário ao longo do curso de tratamento. No entanto, ao contrário da CT, a ressonância magnética não fornece informação sobre a densidade eletrónica dos tecidos, pelo que esta informação tem de ser obtida por criação de imagens de CT sintéticas (sCT, do inglês *synthetic computed tomography*) a partir da MRI. Uma possível solução para este problema é o uso de inteligência artificial, através do treino de um algoritmo com pares de imagens de MRI e CT de vários pacientes, para aprender uma relação entre intensidades em MRI e CT. O algoritmo pode depois ser aplicado a novas imagens de MRI para criar CTs sintéticas correspondentes. Este projeto apresenta a aplicação de uma rede adversária generativa (GAN, do inglês *generative adversarial network*) na criação de CT sintéticas a partir de imagens de ressonância magnética volumétricas, com o objetivo de produzir imagens sintéticas suficientemente realistas para o cálculo da distribuição de dose em planeamento de radioterapia baseado em MRI, para terapia com fotões e protões. Um algoritmo GAN tridimensional para síntese de CT a partir de volumes de MRI foi desenvolvido em *PyTorch*, baseado na arquitetura *pix2pix* para transferência de conteúdo de imagens 2D. As imagens utilizadas neste projeto foram pares de CT e MRI (ponderada em T1) da região da cabeça de 54 pacientes, adquiridas para planeamento de radiocirurgia de Schwannoma Vestibular, tumores localizados no nervo vestíbulo-coclear. O algoritmo foi treinado com pares de CT e MRI previamente co-registados ($256\times256\times192$ vóxeis), divididos em subvolumes de $64\times64\times64$ vóxeis. Na fase de teste, os subvolumes gerados para cada paciente foram concatenados para criar a CT sintética final. Para comparação, o algoritmo 2D original foi também treinado e testado com imagens axiais dos volumes. Um método de validação cruzada foi usado para otimizar o algoritmo utilizando 42 pacientes (grupo A): o grupo foi dividido em 36 pacientes para treino e 6 para teste do algoritmo, sendo o processo repetido 7 vezes. Um segundo conjunto de dados de 12 pacientes (grupo B) foi usado para validação final do modelo otimizado. O desempenho da GAN foi avaliado pela análise qualitativa das imagens geradas e pelo cálculo de medidas de qualidade entre as sCT e as CT reais correspondentes: erro médio (ME, do inglês *mean error*) e erro médio absoluto (MAE, do inglês *mean absolute error*) das intensidades, índice de similaridade estrutural (SSIM, do inglês *structural similarity index measure*) e coeficiente de Dice (DSC, do inglês *Dice similarity coefficient*) do osso. Avaliação dosimétrica foi realizada para 33 pacientes do grupo A contendo tumores e órgãos contornados nas imagens. Planos de terapia modulada com fotões e com protões foram desenvolvidos utilizado modelos de feixes cónicos a energias terapêuticas, usando

um plano de 5 feixes coplanares para a terapia com fotões e 2 feixes para a terapia com protões. Os planos foram otimizados na CT sintética com base em objetivos e restrições clínicos. Os mesmos planos foram recalculados na CT real, e as distribuições de dose calculadas foram comparadas para cada paciente. Esta comparação foi realizada com base em diferenças percentuais de dose para todos os vóxeis, comparação de histogramas dose-volume (DVH, do inglês *dose-volume histogram*) e cálculo de índices gama entre distribuições calculadas na sCT e na CT real. O tempo de geração de uma CT sintética a partir da MRI foi inferior a 30 segundos, com recurso a uma unidade de processamento gráfico. Para as CT sintéticas geradas pelo algoritmo 3D otimizado, o MAE médio no grupo A para os vóxeis pertencentes ao corpo foi de 69 $\pm$ 10 HU, correspondendo a uma redução de 20% no erro quando comparado ao resultado obtido utilizando a arquitetura original da GAN 2D (MAE = 87 $\pm$ 11 HU). Os MAE médios para vóxeis pertencentes a classes de ar, osso e tecidos moles foram respetivamente 272 $\pm$ 32 HU, 146 $\pm$ 26 HU e 38 $\pm$ 5 HU. As CT sintéticas conseguiram reproduzir corretamente a estrutura geral da CT, como comprovado por valores de SSIM = 0,96 $\pm$ 0,03 e coeficiente de Dice DSC = 0,89 $\pm$ 0,03. O pior desempenho da GAN na geração das sCT foi visível nas vias aéreas e estruturas ósseas de pequena dimensão, sendo que erros de corregisto entre os pares CT e MRI utilizados para treino da GAN foram propagados na rede e diminuíram a qualidade da síntese nas regiões do nariz, orelhas e pescoço. O MAE médio quando considerando apenas vóxeis na região nasal foi de 107 HU. Os resultados das métricas de qualidade não diferiram estatisticamente para o conjunto de imagens de validação do grupo B (p = 0,09). Para a simulação de terapia com fotões, as diferenças médias de dose no tumour para planos calculados na sCT e na CT foram inferiores a 2% da dose planeada, para todos os pacientes. As diferenças nos órgãos em risco foram inferiores a 0,2%. Todos os pacientes tiveram taxas de passagem no teste gama superiores a 98% para critérios de 2%/2mm. Os baixos desvios entre as doses calculadas usando a CT e a sCT traduziram-se também em curvas de DVH semelhantes. Diferenças mais elevadas foram visíveis para o feixe anterior-posterior que atravessa cavidades aéreas. Para os planos do feixe de protões os desvios entre as distribuições foram, em média, superiores aos resultados com os feixes de fotões. 21 pacientes apresentaram desvios de dose médios inferiores a 2% no tumour, enquanto que os restantes 12 apresentaram desvios médios entre 2% e 8%. A taxa de passagem no teste gama mínima foi de 94%. Desvios mais elevados deveram-se a diferenças no alcance dos feixes de protões causadas por diferenças de HU entre sCT e CT no caminho que atravessam, em particular em cavidades aéreas de pequena dimensão e em diferenças de espessura do osso do crânio. Os resultados das métricas de qualidade de imagem comparam-se favoravelmente com algoritmos publicados anteriormente usando inteligência artificial e métodos baseados em bibliotecas de imagens. As modificações do algoritmo GAN 2D para 3D permitiu melhorias significativas na qualidade das imagens. O método desenvolvido superou os métodos que usam bibliotecas em tempo de geração das imagens sintéticas, que se reduz da ordem dos minutos para a ordem dos segundos, permitindo o uso desta técnica em fluxos de replaneamento de tratamento online. As diferenças dosimétricas obtidas usando fotões comparam-se às da literatura quando considerando tumores localizados em regiões da cabeça semelhantes. O método aqui desenvolvido e analisado mostra potencial para criação de CT sintéticas em fluxos clínicos de radioterapia guiada por MRI. Atenção especial deve ser dada aos feixes que atravessam pequenas estruturas ósseas e vias aéreas, principalmente em terapia com protões. O algoritmo GAN deve ser otimizado para melhorar o desempenho da síntese nessas regiões, assim como treinado e testado com um maior número de pacientes.

Palavras-chave: Planeamento de Radioterapia, Redes Generativas Adversárias, Tomografia Computorizada, Ressonância Magnética, CT sintética

Abstract

This project presents the application of a generative adversarial network (GAN) to the creation of synthetic computed tomography (sCT) scans from volumetric T1-weighted magnetic resonance imaging (MRI), for dose calculation in MRI-based radiotherapy workflows. A 3-dimensional GAN for MRI-to-CT synthesis was developed based on a 2-dimensional architecture for image-content transfer. Co-registered CT and T1-weighted MRI scans of the head region were used for training. Tuning of the network was performed with a 7-fold cross-validation method on 42 patients. A second dataset of 12 patients was used as the holdout dataset for final validation. The performance of the GAN was assessed with image quality metrics, and dosimetric evaluation was performed for 33 patients by comparing dose distributions calculated on true and synthetic CT, for photon and proton therapy plans. sCT generation time was <30 s per patient. The mean absolute error (MAE) between sCT and CT on the cross-validation dataset was 69 $\pm$ 10 HU, corresponding to a 20% decrease in error when compared to training on the original 2D GAN. Quality metric results did not differ statistically for the holdout dataset (p = 0.09). Higher errors were observed for air and bone voxels, and registration errors between CT and MRI decreased performance of the algorithm. Dose deviations at the target were within 2% for the photon beams; for the proton plans, 21 patients showed dose deviations under 2%, while 12 had deviations between 2% and 8%. Pass rates (2%/2mm) between dose distributions were higher than 98% and 94% for photon and proton plans respectively. The results compare favourably with published algorithms and the method shows potential for MRI-guided clinical workflows. Special attention should be given when beams cross small structures and airways, and further adjustments to the algorithm should be made to increase performance for these regions.

Key-words: Radiotherapy Planning, Generative Adversarial Network, Computed Tomography, Magnetic Resonance Imaging, Synthetic-CT

Contents

List of Figures

List of Tables

List of Abbreviations

ART Adaptive Radiation Therapy.

CBCT Cone Beam Computed Tomography.

CNN Convolutional Neural Network.

CPU Central Processing Unit.

CT Computed Tomography.

CTV Clinical Target Volume.

DCNN Deep Convolutional Neural Network.

DD Dose Difference.

DECNN Deep Embedding Convolutional Neural Network.

DICOM Digital Imaging and Communications in Medicine.

DSC Dice Similarity Coefficient.

DTA Distance to Agreement.

FCN Fully Convolutional Network.

FLAIR Fluid-Attenuated Inversion Recovery.

GAN Generative Adversarial Network.

GPU Graphics Processing Unit.

GRE Gradient Recalled Echo.

GTV Gross tumour Volume.

HU Hounsfield Unit.

IGRT Image Guided Radiation Therapy.

IMPT Intensity Modulated Proton Therapy.

IMRT Intensity Modulated Radiation Therapy.

MAE Mean Absolute Error.

ME Mean Error.

ML Machine Learning.

MR Magnetic Resonance.

MRI Magnetic Resonance Imaging.

OAR Organ at Risk.

PET Positron Emission Tomography.

PTV Planning Target Volume.

RT Radiation Therapy.

sCT Synthetic Computed Tomography.

SE Spin-Echo.

SSIM Structural Similarity Index Measure.

TE Echo Time.

TI Inversion Time.

TPS Treatment Planning System.

TR Repetition Time.

UTE Ultrashort Echo Time.

WE Water Equivalent.

1 Introduction

Cancer is a leading cause of death worldwide, with approximately 9 million cancer related deaths having occurred in 2016, according to the World Health Organization [1]. Radiation therapy (RT) is the most common treatment for cancer, being used in approximately 60% of all cancer treatments, either alone or combined with surgery or chemotherapy. An increase of 15% in the number of radiotherapy treatment courses is expected in Europe by 2025 [2]. Modern radiation therapy is a complex multi-step process, optimising the many features of a modern linear accelerator in order to deliver a high dose of radiation to the tumour yet minimising exposure to surrounding organs.

The current clinical norm uses a multimodality image based planning workflow: computer tomography (CT) is utilised as the primary set, and is used for dose planning, based on the information that it provides on tissue attenuation; magnetic resonance imaging (MRI) is the secondary modality, often used as a basis for monitoring and delineation of target and organs at risk, due to its increased soft tissue contrast [3, 4]. As the integration of MR simulators is becoming more common in radiation oncology departments, there has been a growing interest in the integration of MRI into the different stages of RT, and in the move from a CT-based to an MRI-based workflow [5, 6].

One advantage of using MRI as the primary set for treatment planning is the removal of the registration step between CT and MR. This registration has been shown to introduce geometrical uncertainties in the order of 2 mm [3], which are systematic and propagate through the treatment. Removing this uncertainties can allow to fully take advantage of the better tumour volume and target definitions that are offered by the soft tissue contrast in MRI, and decrease systematic planning margins that are added to these structures for irradiation.

Secondly, the exclusion of planning CT scans can lead to a more efficient and cost-effective workflow, and exposure to ionising radiation can be reduced. This rationale is of particular importance for the use of MRI not only at the initial planning steps of the RT treatment, but also for MRI-driven adaptive re-planning, i.e. to adapt the plan to changes in anatomy or biology observed at different time points of the course of treatment [7]. MRI can provide high-quality images on a daily basis, without exposure to additional ionising radiation, and functional MRI could serve as a basis for adaptation of treatment. Moreover, the introduction of combined MRI and treatment units, MRI-linacs, could bring further advantages in areas such as online MRI-driven adaptive treatment, where re-planning could be executed on the acquired MRI to account for deformations and motion not only between radiotherapy fractions but also prior to and during beam-delivery [8].

One of the main challenges of using MRI both for delineation and optimisation of the RT treatment is the lack of electron density information: CT provides a map of the attenuation properties of tissue, represented by Hounsfield units (HU), information that is necessary for accurate dose calculations. MR is inherently not a quantitative imaging modality, with its contrast being determined by a multitude of factors such as proton density, relaxivity of the tissue, diffusion and chemical composition [9], and therefore can not be used directly for dose calculation purposes.

Thus, it has become desirable to create synthetic-CT (sCT) images from the corresponding MRI scan for dose planning in MRI-based treatment planning workflows. The nonlinearity between intensity values in these two image modalities brings difficulties to this image-to-image translation task. Approaches to this problem can be divided in 5 main groups: bulk density override, voxel-based, patch-based, atlas-based and learning-based methods. Bulk-density override techniques segment the MRI in different tissue classes and attribute an equivalent density to each one of them [10–18]. Voxel-based techniques involve the prediction of HU values from MRI intensities (usually from multiple sequences) using statistical regression models [19–28]. Both these groups of techniques require either manual segmentation of structures or the use of multiple non-conventional MRI sequences that are not frequently employed in the oncology department [29]. Atlas-based methods make use of previously prepared pairs of MRI and CT atlases and deformable registration [30–37], but are highly dependent of registration accuracy. As in patch-based models [38–40], the generation time in the order of minutes means these methods are unsuitable for online re-planning tasks. To overcome these challenges, there has been an increasing search for machine-learning approaches to this image synthesis task [41–50], in order to rapidly create realistic synthetic CT images from conventional MRI sequences.

In this work I investigate the use of a generative adversarial network (GAN) to create synthetic CT images of the head from standard MRI sequences for MRI-only radiotherapy planning of intracranial tumours, in an end-to-end study that spans from the definition of the network architecture to the evaluation of the synthetic-CTs for dose planning. The main goals of this work included (1) examining the use of a 3-dimensional deep learning network for the MRI-to-CT synthesis, (2) creating sCT volumes suitable for online adaptive re-planning throughout the treatment and (3) evaluating the quality of the synthetic-CT for dose calculation with both photon and proton therapy plans.

This dissertation is divided into 5 chapters. Chapter 2 summarises the theoretical concepts behind MR and CT imaging in radiotherapy planning, and the state of the art on algorithms for synthetic-CT creation. Chapter 3 provides information on the data and methods used for the sCT generation and subsequent evaluations. The results of the developed project are presented and discussed in Chapter 4. Finally, Chapter 5 presents the main conclusions drawn from the study, and further work is considered.

2 Background Information

2.1 Imaging in Radiation Therapy

Imaging is a crucial component of the process which a patient undergoes in the radiotherapy department. The most important step in external-beam radiation therapy planning is the localisation of the tumour, as well as the assessment of the involvement of surrounding structures and proximity to critical organs. With the few exceptions of visible skin tumours, this can only be provided through appropriate radiological investigations.

The evolution of imaging modalities in recent years and widespread availability of high-speed CT, MRI, and functional imaging, has let to significant changes in radiotherapy practice. Nowadays, treatment is frequently planned and evaluated using information from CT-based datasets, and information from MRI imaging is often used in an integrated manner. Other imaging techniques, such as positron emission tomography (PET), single photon emission computed tomography (SPECT) or functional MRI can also be used in a parallel way with the diagnostic CT information during radiotherapy planning [4].

2.1.1 Computed Tomography

Fundamentals of CT Imaging

Computed tomography is an X-ray tomographic technique that provides cross-sectional images of the body by the use of X-ray measurements taken from different angles. Clinical CT is currently based on helical CT scanners, in which images are acquired using a continuously rotating X-ray tube and by moving the table on which the patient lies through the scan plane. Detectors around the patient measure the intensity of the attenuated radiation beam as it emerges from the body. These intensities are transformed into X-ray attenuation values based on the original intensity of each ray, producing the CT raw data [51, 52]. The projection measurement, p, which relates the initial and measured intensities is given by the following equation:

$$p = -\ln\left(\frac{I}{I_0}\right) = \mu x \tag{2.1}$$

where I_0 is the entrance X-ray intensity, I the intensity at the detector and μ is the total attenuation coefficient along the path x of the ray. For a nonuniform object, as the body, it can be broken down into its components in each small path length Δx:

$$\mu x = \mu_1 \Delta x + \mu_2 \Delta x + \mu_3 \Delta x + ... + \mu_n \Delta x \tag{2.2}$$

From the raw data, a CT reconstruction algorithm is used to determine these different attenuation

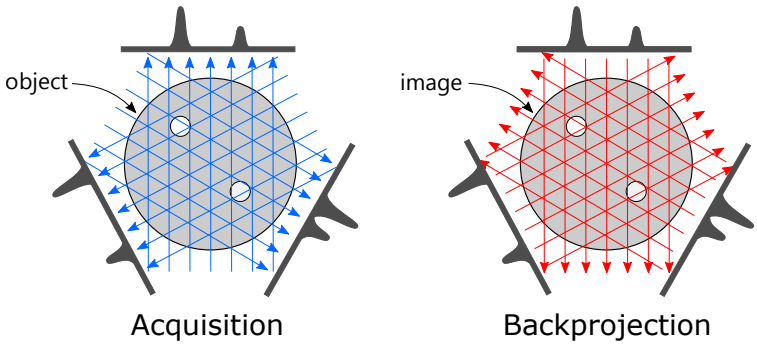

Figure 2.1: Data acquisition in computed tomography at numerous angles around the object (left). In the image reconstruction phase (right), each projection measurement is backprojected onto a digital matrix. Areas of high attenuation are positively reinforced through the backprojection process whereas other areas are not. Adapted from [51].

coefficients and produce the CT images. The most traditional algorithm is filtered backprojection, in which each projection measurement is backprojected onto a digital matrix, after being mathematically filtered using convolution kernels. The image is built up in the computer from the collection of information from multiple detectors. Figure 2.1 shows a simple scheme of the principles behind CT data acquisition and reconstruction. In helical acquisition, interpolation has to be performed during image reconstruction to generate a planar dataset for each table position and to produce artefact-free images [4].

The goal of the reconstruction algorithm is to compute the attenuation coefficient of each voxel on the image. Before storing and displaying, CT images are normalised and truncated to integer values. Each voxel is assigned a Hounsfield Unit (HU) representing its relative attenuation to water:

$$HU = \frac{\mu_i - \mu_{water}}{\mu_{water}} \times 1000 \tag{2.3}$$

where μ_i is the attenuation coefficient on voxel i and μ_{water} is the attenuation coefficient of water at standard temperature and pressure conditions.

By definition, water is assigned a value of 0 HU, while air has a value of –1000 HU. Soft tissues such as fat, muscle, and other body tissues have values ranging from –100 HU to 100 HU. Cortical bones are more attenuating and have CT numbers from 250 HU to over 1000 HU. There is no upper limit to the Hounsfield scale, and foreign bodies such as metal object can have values of over 10000 HU. Medical scanners however typically work in a range of -1024 HU to +3071 HU, reflecting the 12-bit integer range used for data storage in early CT scanners [52].

The result of CT image reconstruction are axial CT images: square matrices in which each pixel intensity represents the HU value of a voxel. These axial images can be manipulated further to produce images in any secondary plane and three-dimensional images. Typical display devices use eight-bit grayscales, representing 256 different shades of gray. Visualising the original dynamic range of over 2000 HU leads to a compressing of the values and little intensity variation can be visualized for soft tissues. Therefore, CT images are typically displayed with a modified grayscale. In the case of brain images, these are often displayed in a window [-40, 80] HU, so than all values below -40 HU are displayed as black, all values above 80 HU are displayed as white, and a linear transformation is applied to display all values inside the window in grayscale, allowing to differentiate gray and white matter within the brain [52].

CT as a Source of Electron Density Information

Precise calculation of dose distributions in radiotherapy can be performed on the basis of knowledge of radiological parameters of the tissues derived from CT imaging. The dose algorithms in many treatment planning systems are based on models calculated in water and adapted to all media other than water using a heterogeneity scaling method. For photon energies between 100 keV and 10 MeV, the Compton interaction is the most important interaction process for energy absorption, and the Compton attenuation coefficient is linearly related to the electron density (ED) of a tissue. This heterogeneity correction can therefore be performed based on the knowledge of electron density of the tissues in the body [53]. The electron density of a material is given by:

$$\rho_e = \rho \times \sum_i f_i \left(\frac{Z}{A}\right)_i \tag{2.4}$$

where ρ is the physical density of the material, f_i the fraction by weight of the chemical element i of the material and $(Z/A)_i$ the atomic number to atomic weight ratio for that element.

Treatment-planning systems (TPS) usually convert HU values from the CT to relative electron density normalized to water ($\rho_e/\rho_{e\,H_20}$) based on HU-to-ED conversion curves [54]. Since Hounsfield numbers for a given tissue depend on the quality of the X-ray beam and individual scanner parameters, this HU-to-ED conversion curve is usually determined empirically [55]. This is done with the use of phantoms with tissue-equivalent materials, i.e., materials that have a known atomic composition and that is similar to human tissues. Data can usually be fitted with a double straight line approach, using one slope for materials with values up to 0-50 HU and a second line for bone-like materials with higher CT numbers [56, 57].

For proton treatment planning, the construction of the calibration curve follows the same procedure as for X-rays, but in this case the required information is the relative proton stopping power, ρ_s. This is calculated using the Bethe–Bloch formula, which can be approximated by:

$$\rho_s = \rho_e \left\{log[2m_e c^2\beta^2/I_m(1-\beta^2)] - \beta^2\right\} = \rho_e\, K \tag{2.5}$$

where βc is the velocity of the proton, c is the speed of light in vacumm, m_e is the mass of the electron, I_m is the mean ionization energy of the target atoms and K is a constant for a given material and energy [56].

2.1.2 MR Imaging

The basis of magnetic resonance techniques is the measurement of radiofrequency radiation, resulting from transitions induced between nuclear spin states of tissue hydrogen atoms in the presence of a strong external magnetic field.

When positively-charged protons in the body are exposed to an external magnetic field, B_0, their magnetic moments align with the direction of the field. The individual protons do not actually line up, but rather rotate around the axis with a frequency directly proportional to the field strength. This frequency is termed the Larmor frequency and is given by the Larmor equation:

$$\omega_0 = \gamma\, B_0 \tag{2.6}$$

where ω_0 is the angular frequency of rotation, γ is the gyromagnetic ratio and B_0 is the magnetic field

strength in tesla (T). Protons have a gyromagnetic ratio of γ = 42.58 MHz/T [58].

The individual spins tend to align parallel or anti-parallel to the magnetic field depending on the energy. Under steady-state conditions, a larger fraction aligns parallel to the main magnetic field, because this is equivalent to spins residing in a more favourable energy state. This small difference produces the net magnetisation, M_0. For most applications, this is taken to lie along the main (Z) axis of the scanner, and therefore along the patient.

When a radiofrequency (RF) energy pulse is emitted with a bandwidth of frequencies centred around the Larmor frequency, this produces a circularly polarised magnetic field, termed B_1 field, which locally overcomes the main field B_0. The spins in the body tip to rotate around the new field direction, generating a transverse component to the magnetisation vector. The RF pulses can either be given enough energy to flip the spins exactly orthogonal to their starting position (90° pulse), antiparallel to it (180° pulse), or any intermediate angle. When the RF pulse ceases, the orientation of the spins returns to its base position via an exponential decay, characterised by relaxation times, termed T1 (spin-lattice relaxation) and T2 (spin-spin relaxation). The variation in the transverse magnetization is detected and processed using sensitive receivers [4, 9]. This process is represented in figure 2.2.

MRI scans are acquired using specific sequences, i.e. series of different radio-frequency (RF) pulses, applied at particular times in a specified way to obtain an image. There are two main types of MR sequences: spin echo (SE) and gradient recalled echo (GRE); other sequences are variations of these two. Variables which can be changed by the scanner operator include the repetition time (TR, the time between the application of an RF excitation pulse and the start of the next RF pulse), echo time (TE, the time between the application of the RF pulse and the peak of the echo detected) and flip angle. The contrast in MR images is determined by differences in T1, T2, and proton density (i.e., the number of ^{1}H nuclei) in various tissues. A particular type of contrast can be emphasised by adjusting these parameters, varying the amount of relaxation in the spins before signal readout [59].

T1-weighted scans have short TE and short TR. In these images, water tissues - which have a long T1 - appear relatively dark while fat - with a short T1 - appears bright. Contrast agents such as gadolinium shorten T1 where they are taken up and therefore highlight these tissues. T2-weighted scans have long TE and long TR, and are useful for showing fluid around tumours or tumours in prostate peripheral zones. In both T1-weighted and T2-weighted images, both air and cortical bones show low signal intensity, appearing as dark [4, 59].

An important feature of MRI is that there is no standard imaging sequence, and even when utilising a particular type of weighting, there is flexibility and variability in the imaging parameters used. This means that MRI contrast can, to some extent, be tailored to meet clinical requirements, while taking into

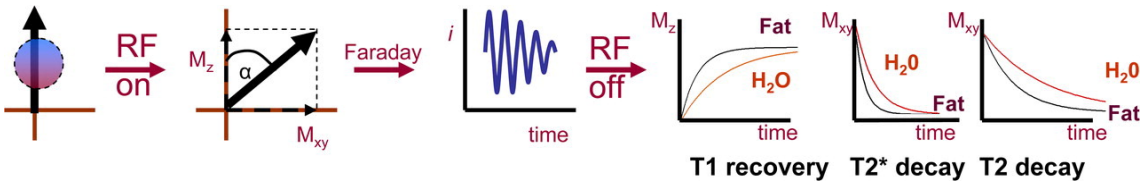

Figure 2.2: Basic physics of the MR signal. As ^{1}H nuclei spin, they induce their own magnetic field. When they are exposed to an external magnetic field (B_0), they align with it. When an RF pulse is applied, the net magnetization vector is flipped at an angle, which produces two magnetization components: longitudinal magnetization (M_z) and transverse magnetization (M_{xy}). As the transverse magnetization precesses around a receiver coil, it induces a current (i). When the RF generator is turned off, T1 recovery and T2 decay occur [59].

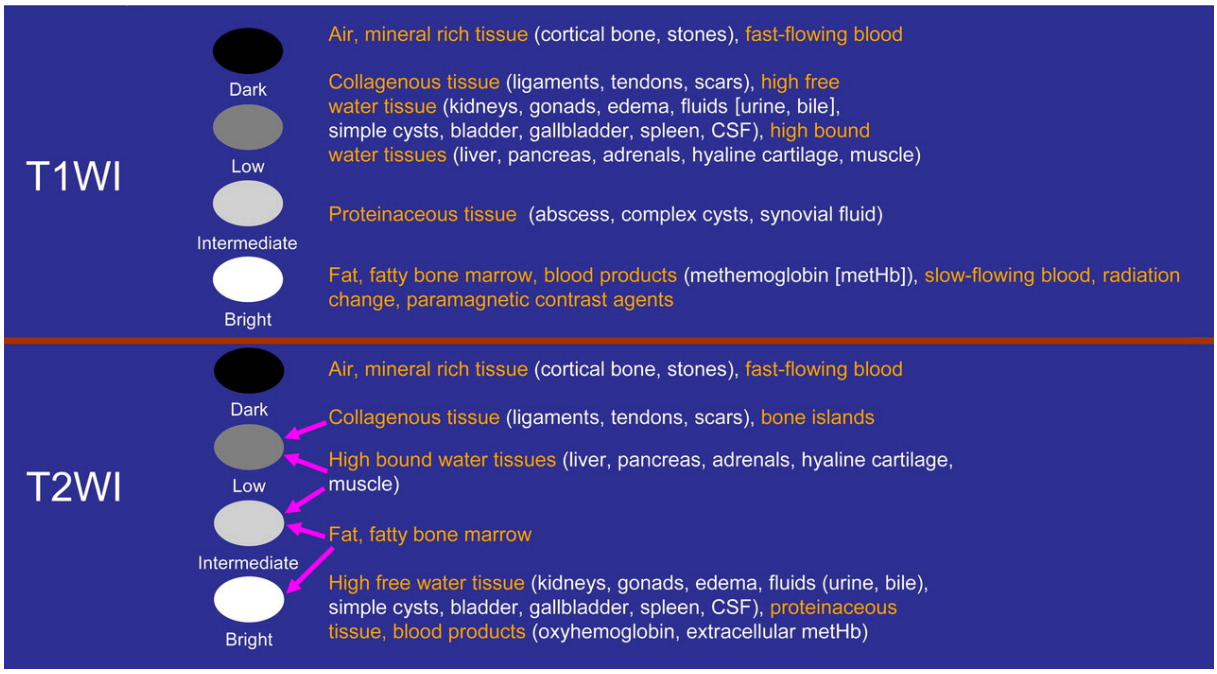

Figure 2.3: Diagram showing the signal intensity of various tissues at T1- and T2-weighted imaging [59].

account image resolution, the scanning volume and scanning time. Technological developments, such as those which have made parallel imaging a clinical tool, allowed a speeding-up of image acquisition, increasing resolution and reducing scan time [4].

MR pulse sequences can be either 2-dimensional, with one section acquired at a time, or 3-dimensional, with a volume of multiple sections obtained in a single acquisition. Most volume sequences provide heavy T1 weighting with excellent grey-white matter differentiation in the brain. Commonly used 3D sequences are BRAVO from GE healthcare and MPRAGE from SIEMENS. Thin sections allow detection of small lesions, such as small metastases or granulomas [60].

Comparisons between MR and CT Imaging

The main advantage of MR compared to CT for use in radiation therapy is its ability to better demonstrate and characterise tumours and soft tissues with high definition. MRI is the imaging modality of choice for pelvis, brain, spinal cord, and some head and neck tumours [3]. In the brain, MRI has been shown to resolve tumour boundaries not resolvable on CT and identify peritumoural edema [61]. For prostate, MRI is extremely beneficial for accurately identifying the prostate, areas of high tumour burden and the prostatic apex, which cannot be identified on CT [62]. ESTRO guidelines have concluded that MRI provides the most reliable delineation for gynecological cancer [63].

In CT, regions surrounded by thick bone will absorb X-rays, producing artefacts and decreasing the visualisation of nearby soft tissue tumours. This effect is absent with MR because cortical bone does not emit an MR signal in conventional MRI sequences and appears as a dark area. Tumours within the posterior fossa or brainstem and tumours centred at bone prominences are therefore better defined [3]. However, this also poses a challenge, making MRI not suitable for investigating bone deformities.

Detailed MR images can be acquired in any orthogonal or non-orthogonal plane and volumetric datasets can be obtained without loss of spatial resolution. Recent developments in MRI have produced

faster imaging and MRI that can be used to assess organ motion [64]. MR can also provide physiological and biochemical tumour information, with the use of MR angiography, MR spectroscopy and Diffusion-Weighted Imaging (DWI), which may allow individual modification of radiation dose or prompt alternative treatments for nonresponders. This way, MRI provides a wider range of applications while CT is limited in these aspects, showing only anatomical images with contrast based solely on differences in X-ray attenuation [3].

The main disadvantage of using MRI in radiation therapy is that whereas for CT electron density - required for dose planning - can be derived from Hounsfield units, MR signal intensity has no such correlation with beam attenuation. Another set-back on the use of MR images for treatment planning is the effect of MR distortions, which needs to be considered and corrected. Whilst MRI does not carry the same risks of ionising radiation exposure as CT, special precautions to remove all ferromagnetic materials from the patient and to carefully assess prosthetic implants are required [4]. General differences between MRI and CT are listed in table 2.1.

Table 2.1: General comparisons between MR imaging and conventional CT imaging [3].

Subject	**Parameters**	**MRI**	**CT**
Patient	Ionizing radiation dose	Nil	Present
	Magnetic safety concerns	Present	Nil
	Radiofrequency heat deposition	Present	Nil
	Claustrophobia in scan tube	Present	Minimal
	Scanning noise	Present	Minimal
	Contrast materials allergy:		
	- Iodinated contrast	Not applicable	Present
	- Gd DTPA (gadolinium)	Minimal	Not applicable
Characteristics	Soft tissue contrast	Excellent	Moderate
	Cortical bone contrast	Poor	Excellent
	Detection of calcifications	Poor	Excellent
	Metallic artefacts		
	- Non-ferromagnetic material	Minimal	Present
	- Ferromagnetic material	Present	Present
	Electron density information	Nil	Present
Machine	Size of tunnel aperture	Smaller	Larger
	Image resolution	Good	Excellent
	Scanning time	Moderate	Short
	Functionality and technical sequences	Large	Limited
	Geometrical image accuracy	Distortion present	Excellent
	Multiplanar imaging	Any plane	Limited
	Multiplanar reconstructions	Available	Available
	Cost	Higher	Lower
	Availability	Moderate	Widely available

Comparisons are given in relative scale, in which Nil < Minimal < Present, and Poor < Moderate < Good < Excellent.

2.2 Radiation Therapy Workflow

2.2.1 Initial Treatment Planning

Radiation therapy treatment planning involves choosing an appropriate patient positioning, identifying the location and contour of the tumour and of the organs at risk (OAR), selecting a suitable beam arrangement, calculating the treatment machine settings to deliver the required plan and finally evaluating the resulting dose distribution [3].

CT-based treatment planning has become the standard of practice in radiotherapy. When acquiring treatment-planning images, CT imaging is acquired in the radiotherapy treatment position. This requires the substitution of the convex couch top that is normally used for diagnostic imaging with a flat one so that the set-up mimics a radiotherapy treatment couch [65]. The treatment position needs to be both comfortable and reproducible. The set-up often includes the use of markers, the placement of radio-opaque fiducial reference points and/or the use of immobilisation masks when appropriate. For head immobilisation, it is frequent to use thermoplastic meshes indexed and mounted to a rigid frame, attached to the therapy couch [66]. For planning MRI, the standard practice is to acquire the scans in the same position as for diagnostic MRI, and regularly using curved cushion lined couches. However, MRI-simulation platforms have recently emerged in radiotherapy departments. These differ from diagnostic MRI by including larger bore size, flat tabletops to accommodate immobilisation devices, external laser systems and dedicated imaging protocols, allowing to acquire MRI scans in treatment position as it is done in standard practice for CT [5, 6].

The CT and MR planning images are downloaded into a treatment planning system. In CT-based radiotherapy planning, the CT scan is assigned as the primary set of patient data for treatment planning. The tabletop is removed from the CT images and the external patient contours are extracted on each transverse image using automatic segmentation algorithms, namely edge-detection techniques [3]. The external contour is essential for planning because it represents the interface between air and tissue and therefore allows the planning computer to take into account the variations in beam attenuation and obliquity of incidence to calculate dose distributions. In order to incorporate MR images in the RT planning, the MRI dataset (secondary dataset) has to be co-registered with the CT scan (reference dataset), which was acquired in treatment position. Upon the complexity of the sites to be coregistered and depending on the versatility of the virtual simulation software used within the clinic department, transformation models can be rigid, affine, and deformable. For sites such as the head, where not much deformation is expected between the planning MR and CT scans, affine registration is commonly employed: this registration allows for translation/rotation, uniform scaling and sheer operations. For sites where more movement can occur between the image acquisitions, such as the abdomen or pelvis, deformable registration can be used [67, 68].

Co-registration of the planning images allows for visualisation of structures seen on the MR scans superimposed onto the CT image set, for delineation of the internal organs and tumours. The images can be overlaid on top of each other, with the possibility to switch instantaneously from one modality to the other, or blended into a single image with the possibility of interactively changing the relative weight of each of them [3]. The International Commission on Radiation Units and Measurements (ICRU) reports 50 [69] and 62 [70] stipulate standard definitions to the volumes relevant to radiotherapy treatment planning. This includes the description of gross tumour volume (GTV), clinical target volume

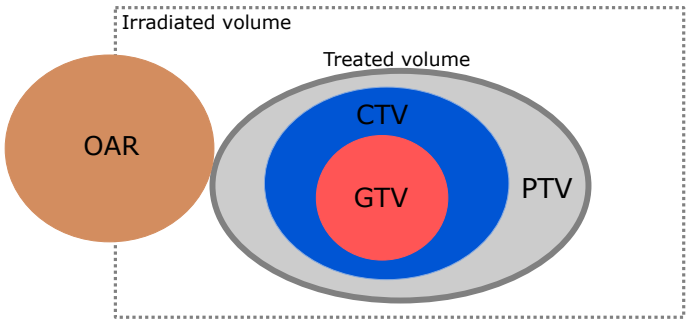

Figure 2.4: Schematic representation of the radiotherapy volumes defined by ICRU reports 50 and 62. Adapted from [72].

(CTV) and planning target volume (PTV). The GTV is the gross visible extent of the tumour, while the CTV is a volume encompassing the GTV and a margin accounting for its microscopic extension, which can be obtained by adding an empirical margin to the GTV or through individual clinical judgement based on results of clinico-pathological studies. The PTV is obtained from a safety margin added to the CTV to take into account the organ motion and the uncertainties inherent to radiotherapy setup delivery. In addition to the target, the clinicians also outline several OAR, anatomical structures with important functional properties located in the vicinity of the target volume. The OARs to be considered depend on the anatomical region being treated and the location of the PTV [71]. A schematic representation of these volumes is depicted in Figure 2.4.

Based on this information, a virtual patient can be constructed for 3-dimensional treatment planning. Intensity-modulated radiation therapy (IMRT) makes use of inverse planning. The physician and dosimetrist commonly select the beam angles, or arcs for arc-based therapy, and energies up front. The criteria that a plan must achieve is defined in the form of objective functions. IMRT optimisation systems use dose-volume-based criteria, such as percentage of the PTV that must receive a specified dose, and dose-volume limits to OARs. Physicians assign the priority of each of these input variables based on clinical guidelines [73]. The fluence profiles for all beam ports are adjusted by the optimisation algorithm, which runs through various iterations, until it converges to a plan that meets the input criteria. Dose calculations are performed using electron density information obtained from the planning CT, as described in Section 2.1.1.

After review of the computer-optimised dose distribution, some modification of the planned outcome and adjustment of the relative importance of each parameter might be needed if the physician is not satisfied with the dose at the target or OARs. Finally, the plan parameters and treatment machine settings are saved to be used in the treatment [3, 65].

2.2.2 Image-Guided and Adaptive Radiation Therapy

The ability to deliver conformal dose distributions through techniques such as intensity-modulated radiotherapy and charged particle radiotherapy has led linear accelerator manufacturers to develop integrated imaging systems in the treatment room to allow for image-guided radiotherapy (IGRT) in order to improve geometric irradiation accuracy [74].

The initial use of image guidance in radiotherapy has been for identifying internal organ position and assessing the subsequent isocenter shifts before initiating radiation therapy. In addition, IGRT enables imaging to permit dosimetric plan adjustments in response to changes in anatomy. Adaptive radiotherapy

(ART) makes use of anatomical variations in each patient to feedback into the plan and dose optimisation during the treatment course [75]. The goal is to ensure that the planned dose is accurately delivered taking into account the anatomy at the time, so PTV margins can be reduced and ultimately allow dose escalation in the primary tumour to improve treatment outcome. ART can be implemented offline between fractions, or online, either immediately prior to a fraction or during a fraction in real-time adaptive re-planning. In online ART, in-room imaging, deformable image registration (DIR) between planning and daily images, accelerated re-planning and quality assurance need to be performed in small time scales and with the patient lying in the treatment position in the treatment room [7, 8].

CT Guidance

CT-based IGRT has become the gold-standard for daily in-room treatment imaging. Several CT-guidance systems have been developed for this purpose. Kilovoltage cone-beam CT (CBCT) uses imaging components mounted orthogonally to the treatment beam, while megavoltage cone-beam CT (MV-CBCT) uses the accelerator's treatment beam and its imaging system to acquire CT images at megavoltage energies. Fan-beam MVCT has been introduced clinically by integrating it with helical tomotherapy machines, using the helical treatment beam to produce the images. To overcome the image quality problems existing in CBCT and MVCT, the "CT-on-rails" concept has been introduced. It is a fan-beam kVCT system mounted on rails with a specialised couch, providing diagnostic-quality daily images [76].

Although daily imaging doses from CT scans are generally small compared to therapeutic doses in radiotherapy treatments, they are distributed over the entire imaged volume. Dosimetric studies report dose ranging from 0.1 to 2 cGy/scan for kV-CBCT, 0.7 to 10.8 cGy/scan for MV-CBCT and 0.7 to 4 cGy/scan for fan-beam MVCT images, depending on the selected CT pitch and the imaged anatomy volume [76]. This means that, over a complete course of treatment, imaging doses can range from 3 to 370 cGy. Facilities should evaluate the doses associated with each IGRT implementation. The process of justification and optimisation is of particular importance for infants and young children since the risk of inducing a secondary malignancy is higher, due partly to the increased sensitivity to radiation of bone marrow in children and, for young girls, developing breast tissue [4].

MRI Guidance

Compared to CBCT or fan-beam MVCT, MRI offers superior soft-tissue definition with no associated ionising radiation risk, allowing for continued monitoring with repeated scans. MRI is therefore a suitable candidate for guidance during the course of the radiotherapy treatment, enabling direct tumour visualisation as well as OAR definition in soft tissue that is not well visualised in CBCT [77, 78].

MRI guidance has been implemented using offline protocols without an integrated MRI radiotherapy system by using either a diagnostic MRI scanner, MRI simulator, or a shuttle-based MRI-guided radiotherapy system (MRI-on-rails) [79]. These systems have been assessed for daily position verification using daily out-of-room MRI simulation [80] and weekly offline IMRT dosimetric re-planning [81].

Recently, a number of integrated MR-guided radiation therapy (MRgRT) systems have been implemented in the clinic. ViewRay commercialises two MRIndian systems, one combining a split magnet 0.35T MR imaging system with 3 multileaf collimated cobalt sources, in clinical use since 2014, and

an upgraded version with a linear accelerator, in use since 2017 [8, 82]. The Elekta MRI-linac unit, introduced in 2016, consists of a 7-MV linac that rotates around a closed 1.5 T MRI scanner [83]. Integrated MR-linac systems allow to acquire high resolution MR images just prior to radiation and during treatment using fast MRI sequences [78]. In addition to daily online ART, Both the ViewRay and Elekta MRI-linac solutions are capable of tumour tracking as well as gating using cine-MRI, which addresses intra-fractional motion, ultimately allowing for real-time plan adaptation [8]. Proof of principle of these intra-fraction ART workflows have been published [84, 85]. In clinical settings, the Elekta MRI-linac acquires continuous seven-second MRI during irradiation, which are presented for patient monitoring but are not used for intra-fraction plan adjustments [83].

Currently, clinical MRI-linac treatment workflows still make use of a pre-treatment planning CT, and therefore are not an MRI-only workflow. Prior to the course of treatment, both a CT and MR planning images are acquired for target definition and ED information, similarly to that described in section 2.2.1. On the MRI-Linac treatment day, an online MRI is acquired and the contours and the Hounsfield values are defined by registration of the pre-treatment CT data with the online MRI using a DIR software, thus generating a daily pseudo-CT by registration [83, 86].

2.3 State of the Art of synthetic-CT Generation

The use of MRI alone for radiation therapy planning requires that HU or ED information can be derived from the MRI scan. A number of different methods have been developed for MRI-based synthetic-CT generation, and can be grouped into 5 major categories: bulk density override, voxel-based approaches, patch-based approaches, atlas-based approaches, and learning-based approaches. Some studies on sCT generation are focused on their application for MRI-only radiation therapy, with results reported for dose calculation, while several others are more general in their applications and report results on image quality of the generated sCT volumes [29, 87].

2.3.1 Bulk Density

The first approaches to be introduced for synthetic-CT creation from MRI scans were based on applying bulk density overrides for dose calculation. The simplest method is to assign the entire patient volume a water equivalent (WE) electron density. This method was tested for brain [10–12] and prostate [13], but it results in dose calculations that are not considered clinically acceptable compared to calculations on the original heterogeneous densities, with average discrepancies for the whole volume greater than 2%, and in particular when the beam passes through air cavities in the body [14]. Additionally, it is not possible to use these images for patient positioning [29].

Dosimetric results can be improved to be within 2% of dose calculations performed on original CTs by assigning different classes of electron density for bone, soft tissue and air [15, 18]. This method requires segmentation of these structures, with some studies resorting to manual segmentation [14, 16, 17], which is highly time consuming and impractical for clinical workflows as it represents an extra segmentation step. The segmentation of bone is a challenge in conventional MR sequences because cortical bone does not produce signal, which results in a lack of contrast between air and bone. To overcome this, the use of ultrashort echo-time (UTE) sequences has been proposed for the separation of bone and air [88–90]. In addition, Dixon sequences can be used to segment water and fat tissues [91].

A commercially available solution for MRI-only treatments, Philips' MRCAT [92], generates synthetic CTs based on a dual echo 3D mDIXON MRI sequence. The MRI is automatically segmented using a classification algorithm and bulk HU values are assigned to five categories: air, fat, water, spongy bone and compact bone. This method was shown to result in dosimetric discrepancies between the original CT and the sCT below 0.5% for the prostate site. A version of the software that can provide a larger range of HU values is being developed. Although this method does not require any manual segmentation, it requires special MRI sequences and can be impractical for some MRI-based workflows.

2.3.2 Voxel-based

Voxel-based techniques for sCT generation involve the prediction of HU values from MRI intensities from a number of different sequences using statistical regression models. When using routine MRI sequences [19–21], manual segmentation of bone and airways is often required. As in the case of bulk density override methods, a large portion of voxel-based techniques use UTE sequences for automatic classification of bone tissue, usually integrated with other sequences [22–27]. In some voxel-based approaches, individual pixel intensities are used for the estimation [22, 24], and therefore no spacial information is used. This results in poor quality in air/soft-tissue and bone/soft-tissue interfaces [22]. Johansson *et al.* [28] proposed the inclusion of spacial information (x, y and z positions and distance to the body contour) to the statistical regression model, with improved quality results in complicated anatomical areas of the head.

Studies using voxel-based techniques on multiple MRI sequences have reported mean absolute errors (MAE) under 140 HU for the brain region [22, 23] and mean dose parameter differences lower than 1%, with Jonsson *et al*. [27] reporting average dose deviations at the PTV of 0.3% using a Gaussian mixture regression model, although only using 5 patients. The main drawback of voxel-based techniques is the need for multiple MRI sequences, which increases acquisition time, and particularly the use of UTE sequences for automatic bone and air separation, which are not conventionally used in clinical settings [93].

2.3.3 Patch-based

In patch-based approaches, 3D patches (i.e. small cuboidal image subregions) are extracted from the input MRI and a search for the most similar patches is performed in a preacquired database of MRI scans. The corresponding CT patches are then applied to the new sCT for the input patient. Using this method on 5×5×5 patches of T1-weighted MRI images, Andreasen *et al.* [38] reported an MAE of 85 HU and dose deviations lower than 0.5% at the PTV for the brain, but their method was tested using only 5 patients. Aouadi *et al.* [39] implemented a multi-scale patch method using T1 and T2 sequences, and reported dose deviations metrics under 1%. Speier *et al.* [40] used a multimodal approach to the patch-based model, including a recursive step in which the preliminary sCT is added as an additional modality to the library search and is compared with the real CT of the library group to generate the final sCT. Using this method in T1 and T2-weighted images simultaneously, they achieved an MAE of 73 HU for the head region. Due to the need for library search, these methods are highly time-consuming, with reported generation times of over 30 minutes [38, 40].

2.3.4 Atlas-based

Atlas-based approaches are based on previously prepared co-registered MRI and CT atlases. Each MRI atlas is registered to the new target MRI using deformable registration. Then, the resulting displacement fields are applied to the corresponding atlas' CTs and finally the deformed atlas' CT images are fused into a single synthetic-CT [30–35]. The simplest atlas-based methods use an average atlas, as developed by Dowling *et al.* [30]. Improvements in sCT quality have been reported when using multiple atlases in combination with patch-based fusion and sparse coding pattern recognition methods [32–35].

The main advantages of atlas-based methods are the use of a single conventional MRI sequence and the process being fully automated. These techniques have resulted in dose deviations below 1% [32, 35] and images with average MAEs below 50 HU for the prostate site [32] and 125 HU for the brain region [33]. However, the performances of the above methods are highly related to the registration accuracy, with large errors reported in patients with atypical anatomy [31]. Moreover, the use of multiple atlases and robust strategies to fuse the warped CT images is computationally intensive and highly time consuming [32]. Farjam *et al.* [36] reported a sCT generation time of at least 20 minutes for a multi-atlas algorithm.

2.3.5 Learning-based

Learning-based approaches for sCT generation have started gaining momentum in the last few years with the growing applications of machine learning (ML) in medical imaging [94]. One of the contributors to the rise in popularity of ML approaches has been the growing availability of Graphics Processing Units (GPU), allowing for faster computing in image synthesis tasks, in addition to the release of open-source software packages designed for deep learning, such as Caffe [95], Tensorflow [96] and Torch [97].

Regardless of the learning method used, ML approaches are all based on the concept that the MRI-CT relation can be learned from a training set and then applied to a target MRI scan for CT image prediction. Learning-based methods seek to include contextual spacial information by using either small 3-dimensional patches or entire 2-dimensional images as input to the networks. Unlike the previously described methods, learning-based approaches do not require any manual segmentation, and generally use conventional MRI sequences. Table 2.2 summarises the methods and main results from published studies on CT synthesis using ML.

Huynh *et al*. [41] presented a structured random forest approach to the generation of synthetic-CT, using a 3D patch method, which is, partitioning each MRI into a set of small patches and applying a random forest to predict each CT patch, using an auto-context model to iteratively improve the prediction. This type of method had been previously used in image synthesis to reconstruct high-resolution T2-weighted and FLAIR MR images from T1-weighted and T2-weighted sequences [98, 99]. Their method was shown to yield better results than voxel-based and atlas-based methods. Also using small patches of the volumes for the learning process, Nie *et al*. [42] applied a fully convolutional neural network (FCN) for the CT synthesis task of the prostate region.

Han [43] was the first to apply a deep convolutional neural network (DCNN) method for CT synthesis from T1-weighted MRI, using a network built on CNN architectures developed for object segmentation. This model was designed for 2D images, mapping each MRI slice to a corresponding sCT slice independently, which can cause discontinuities in the sCT volume. Later, Xiang *et al*. [44] applied a deep embedding convolutional neural network (DECNN) for the same task. Their network consisted of

Table 2.2: Summary of methods and main results from studies on MRI-to-CT synthesis using learning-based approaches.

Reference	**ML Method**	**Number of Patients**	**Main Results**	**Notes**
Huynh *et al*. (2016) [41]	Random forest	16 (brain) 22 (prostate)	MAE = 99.9 $\pm$ 14.2 HU MAE = 48.1 $\pm$ 4.6 HU	Synthesis performed using small 3D patches of the MRI and CT volumes.
Nie *et al*. (2016) [42]	FCN	22 (prostate)	MAE = 42.4 $\pm$ 5.1 HU	Synthesis performed using small 3D patches of the MRI and CT volumes.
Han (2017) [43]	DCNN	18 (brain)	MAE = 84.8 $\pm$ 17.3 HU	Synthesis performed using 2D slices.
Xiang *et al*. (2018) [44]	DECNN	16 (brain) 22 (prostate)	MAE = 85.4 $\pm$ 9.4 HU MAE = 42.5 $\pm$ 3.1 HU	Synthesis performed using 3 consecutive slices (quasi 3D approach).
Dinkla *et al*. (2018) [45]	Dilated CNN	52 (brain)	MAE = 67 $\pm$ 11 HU Mean dose differences $< 1.5\%$ at the PTV	Synthesis performed using orthogonal slices (2.5D approach); study includes dosimetric evaluation.
Nie *et al*. (2017) [46]	GAN	16 (brain) 22 (prostate)	MAE = 92.5 $\pm$ 13.9 HU MAE = 39.0 $\pm$ 4.6 HU	Synthesis performed using small 3D patches of the MRI and CT volumes.
Emami *et al*. (2018) [47]	GAN	15 (brain)	MAE = 89.3 $\pm$ 10.3 HU	Synthesis performed using 2D slices.
Maspero *et al*. (2018) [48]	GAN	91 (pelvic)	MAE = 61 $\pm$ 9 HU Dose differences < 1%	Synthesis performed using 2D slices; use of Dixon MR sequences; use of MR-CAT software for automatic air segmentation; study includes dosimetric evaluation.
Wolterink *et al.* (2018) [49]	CycleGAN	24 (brain)	MAE = 73.7 $\pm$ 2.3 HU	Synthesis performed using 2D slices; use of unpaired data for training.
Yang *et al.* (2018) [50]	CycleGAN	45 (brain)	MAE = 129 $\pm$ 16 HU	Synthesis performed using 2D slices; use of unpaired data for training.

convolutional and concatenation operations. They adopted a quasi-3D mapping, by synthesising every 3 consecutive axial slices and combining the outputs into the final result by simple averaging. These methods were shown to outperform atlas-based and random forest methods in quality of the synthesised CT images.

Dinkla *et al*. [45] proposed the use of a 2-dimensional dilated CNN for this synthesis problem, using a 2.5D approach in which axial, sagittal and coronal 2D images are used as input to the network. Their study focused on the use of sCT for MRI-based radiation therapy workflows, involving dosimetric evaluation for photon therapy, in addition to the image quality metrics. Clinical treatment plans were calculated in both the original and synthetic CT using a Monte Carlo dose algorithm. They reported mean voxel-wise differences on dose distributions calculated on the sCT lower than 1.5% for the PTV, as well as deviations in the maximum dose to the organs at risk well below 1%.

Generative adversarial networks (GAN) for MRI-based sCT generation were first applied by Nie *et al*. in 2017 [46]. Their proposed network used small 3D patches ($16\times16\times16$ voxels) of T1-weighted MRI volumes and showed improved results in sCT blurriness compared to FCN networks. Emami *et al*. [47] also proposed the use of a GAN for this task, here using 2D slices, and achieved better sCT image quality when compared to CNN networks tested on the same dataset. These two studies do not however evaluate the dosimetric accuracy of the generated sCT. Maspero *et al*. [48] tested the use of sCT generated with an existing 2D GAN architecture for accurate dose calculation in the pelvic area, and reported average dose differences lower than 0.3%. They acquired dual gradient-recalled echo MRI sequences for the sCT generation, using Dixon reconstruction to obtain in-phase, fat and water images and Philips' MRCAT to automatically identify air regions on the scan. This requires additional acquisition time and specific software, which may not be practical for MRI-based planning where conventional MRI sequences are used.

The aforementioned learning-based methods for MRI-to-CT synthesis require a training set of paired MRI and CT volumes. To benefit from additional MRI or CT training volumes from patients who were not scanned using both modalities, Wolterink *et al.* [49] applied a CycleGAN model to synthesise CT data from 2D MRI of the brain region using unpaired data for training. To further improve structure consistency on unpaired data training, Yang *et al.* [50] implemented a structure-constrained CycleGAN on the same type of data. These studies however lack evaluation of the synthesised images for RT treatment simulation.

2.4 Generative Adversarial Networks

Generative adversarial networks were first introduced in 2014 by Goodfellow *et al*. [100]. In a GAN two models are simultaneously trained: a generative model G that captures the data distribution, and a discriminative model D that estimates the probability that a sample came from the training data rather than G, i.e. if the data is real or fake. The training procedure for G is to maximise the probability of D making a mistake (fool the discriminator), while D is in a feedback loop with the ground truth label (real or fake) to improve its predictions. This framework corresponds to a zero-sum two-player game.

In a simple GAN, a neural network $G(z, \theta_1)$ is used to model the generator mentioned above. The role of this network is to map input noise variables z to the desired data space y (e.g. images). Conversely, a second neural network $D(y, \theta_2)$ models the discriminator, which is fed generated and real images, and

outputs the probability that the sample came from the real dataset, in the range (0,1). In both cases, θ represents the weights or parameters that define each neural network.

The goal for D is then to maximise the probability of assigning the correct label to both real images y and generated samples $y' = G(z)$. This means to maximise the probability that any real data input y is classified as belonging to the real dataset - maximising the function $D(y)$ - while minimising the probability that any fake image $y' = G(z)$ is classified as belonging to the real dataset - minimising $D(G(z))$ or maximising $1 - D(G(z))$. The generator is trained to fool the discriminator by generating data as realistic as possible, meaning that the generator's weights are optimised to maximise the probability that any fake image is classified as belonging to the real dataset. Formally, this means that the objective function used for this network maximises $D(G(z))$ or minimises $1 - D(G(z))$.

Mathematically, D and G play the following two-player min-max game with value function $V(G, D)$:

$$\min_G \max_D V(D, G) = \min_G \max_D \left[\mathbb{E}_y[logD(x)] + \mathbb{E}_z[log(1 - D(G(z)))]\right] \tag{2.7}$$

Conditional GANs [101] are a class of GANs trained in such a way that both the generator and the discriminator models are conditioned on input data x. In this case, G learns a map to y from both an input image x and a random noise vector z. The discriminator D also observes the input data. For a GAN in the conditional setting, the value function in Equation 2.7 becomes:

$$V(D, G) = \mathbb{E}_{x,y}[logD(x, y)] + \mathbb{E}_{x,z}[log(1 - D(x, G(x, z)))] \tag{2.8}$$

A schematic representation of a conditional GAN is given in Figure 2.5.

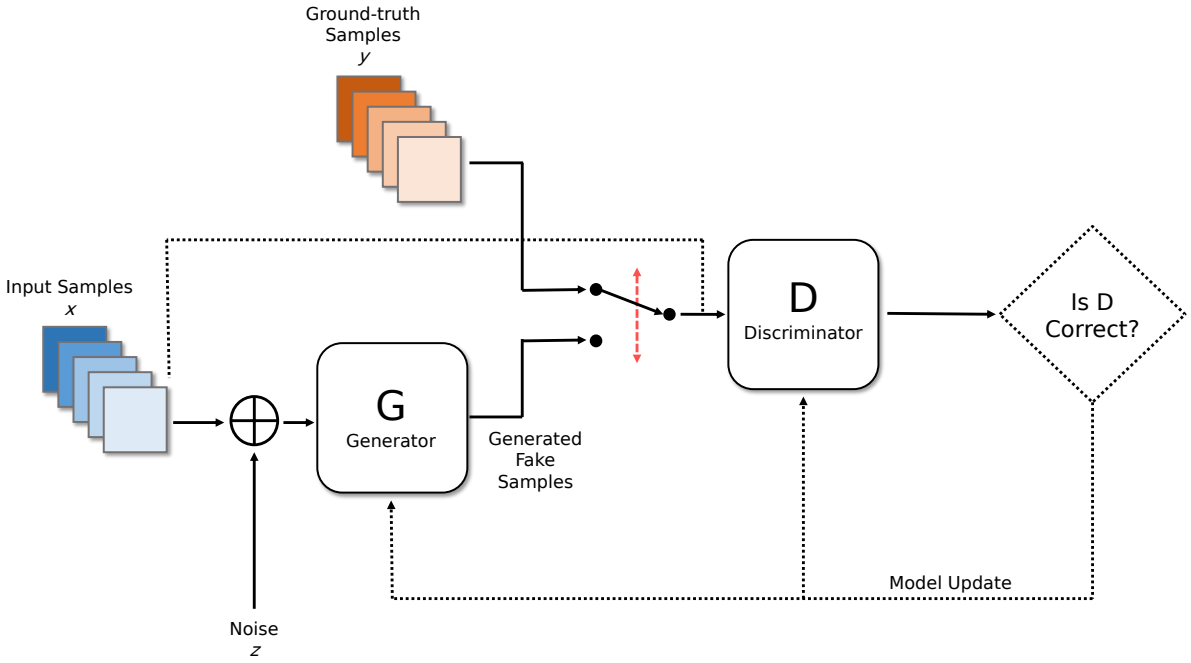

Figure 2.5: Representative scheme of a conditional generative adversarial network. The generator produces fake samples based on the input x and noise z. The discriminate model observes x in addition to either the generated samples or the real samples y, and its output is used to update both models based on the cost function.

Since both the generator and discriminator are modelled with neural networks, a gradient-based optimisation algorithm can be used to train a GAN. The parameters θ (weights) that define each network are randomly initialised and are then updated during training using the backpropagation of the error, according to the cost function. At each iteration, the weights are updated using the gradients of the errors and the defined learning-rate of the network. In the training stage, the weights of the generator are kept constant when training the discriminator, and vice-versa. The discriminator can be implemented directly by configuring the discriminator model to predict a probability of 1 for real images and 0 for fake images and minimising the binary cross-entropy loss. It is always easier for the discriminator to distinguish the generated images from real images in early training, causing the value function to saturate as $log(1 - D(x, G(x, z)))$ rapidly approaches 0. The generator is therefore usually trained to maximise $logD(x, G(x, z))$ rather than minimise $log(1 - D(x, G(x, z)))$, to provide stronger gradients and facilitate training [102]. At the testing stage, only the trained generator model is used to generate fake images $y' = G(x, z)$ from input images x.

GANs have been applied to a variety of tasks, but one of its biggest applications is the task of image-to-image synthesis [103–107]. Applications of GANs to medical image synthesis include cross-sequence MRI synthesis [108–112], generation of synthetic-PET from CT [113, 114] and MRI [115], retinal image synthesis [116, 117], and high-quality image estimation from low-dose scans [118–120], in addition to MRI-to-CT synthesis as described in section 2.3.

In the case of the MRI-to-CT synthesis task, the input samples x represented in Figure 2.5 correspond to the MRI scans and the ground-truth samples y correspond to the real CT images, which are both fed to the network during training, while the generated fake samples are the synthetic CT images.

3 Materials and Methods

3.1 Data Description

54 patients who underwent treatment planning for radiosurgery of vestibular schwannomas at Addenbrooke's Hospital (Cambridge University Hospitals NHS Foundation Trust), between August 2017 and January 2019, were retrospectively selected for this project. Vestibular shwannomas are small benign tumours located close to the inner ear. Dataset A consisted of 42 patients whose data were retrieved in a first phase and Dataset B consisted of additional data from 12 patients retrieved at a later stage. For all patients, a planning CT and a 3D T1-weighted MRI of the head were available. Additionally, delineations of the GTV and OAR were available for 33 patients of Dataset A.

This dataset was chosen based on its characteristics that facilitate CT synthesis using machine learning: (1) patients undergoing radiosurgery have small tumours, so there is no significant deformation of the brain caused by the tumour; (2) the MR images acquired for radiosurgery planning have high resolution, which allows for subsampling of the data; (3) the MR images for radiosurgery planning are always acquired using the same sequence (3D BRAVO T1-weighted), which is not the case for radiotherapy planning in this clinical department, where different sequences are used.

3.1.1 Data Acquisition

Patients were scanned following the clinical protocol of the radiotherapy department at Addenbrooke's Hospital. All CT scans were acquired on a Toshiba Aquilion LB scanner with tube current of 100 mA and tube potential of 120 kV. Images had an in-plane size of 512×512 pixels and voxel size = 0.781×0.718×1 mm^3. The CT scans were acquired in treatment position using a head immobilisation mask. MRI scans were performed on a GE Optima MR450w scanner with a field strength of 1.5 T. Images were acquired after the administration of a gadolinium contrast agent. No immobilization mask was used for the acquisition of the MRI. Acquisition parameters for the 3D T1w BRAVO sequence were as follow: repetition time (TR) = 8.46 ms, echo time (TE) = 3.25 ms, inversion time (TI) = 450 ms, flip angle = 12° and readout bandwidth = 244 Hz/pixel. The field of view (FOV) was 256×256×252 mm^3 with voxel size = 1×1×1 mm^3. For all patients, the planning CT and MR were acquired in the same day. The average time span between the acquisitions was 1.4 h (range: 0.1 to 4.5 h).

3.1.2 Image Pre-processing

The MRI and CT DICOM files were imported to MATLAB 2018a (MathWorks, Natick, MA), where the data was pre-processed. In order to simulate an MRI-only workflow, the MRI grid was used as the reference. For each patient, the CT was registered to the MRI space using MATLAB's affine registration algorithm based on mutual information. For each patient, a binary mask of the body region was created

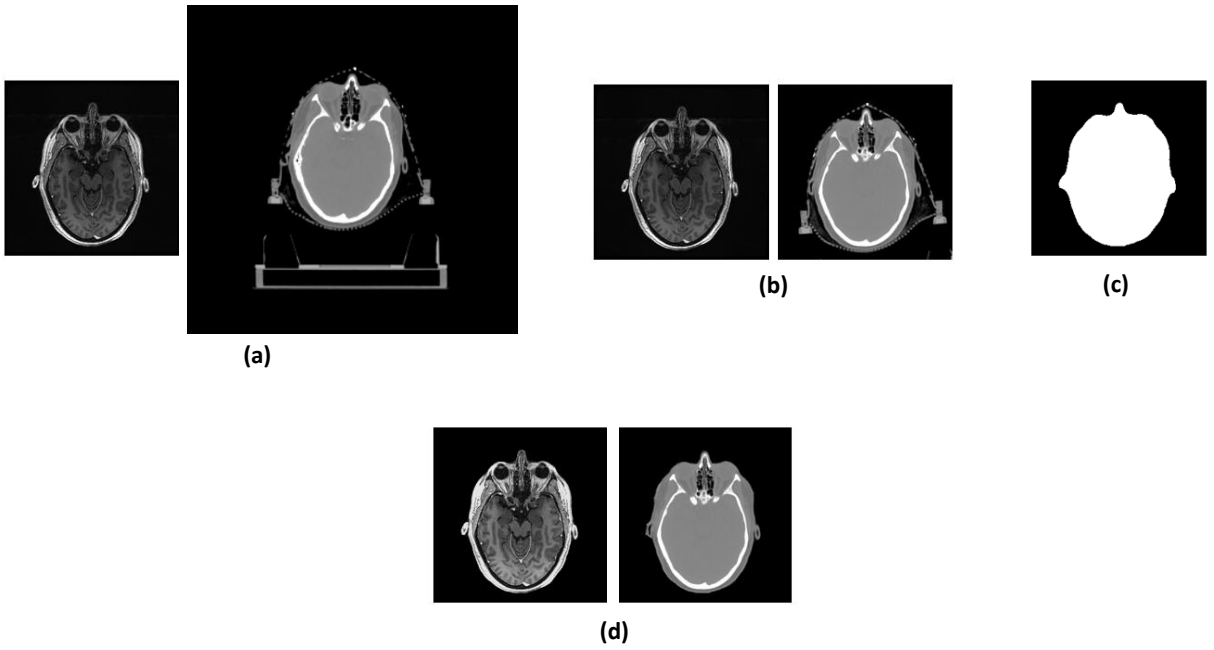

Figure 3.1: Pre-processing steps: **(a)** original volumes; **(b)** the CT volume was registered to the MRI space using affine registration; **(c)** a body mask was created from the MRI volume using binarizing and closing operations; **(d)** the body mask was applied to the volumes, MRI intensity values were normalised and CT values were cropped to the interval [-1000,1500] HU.

from the MRI by threshold and closing operations. This mask was applied to the registered CT, and all voxels outside the body region were set to air (-1000 HU), in order to remove the immobilization mask and table visible on the scan. Each MR scan was normalised based on its 95th percentile of intensities, to account for possible differences in MR intensity ranges and to decrease the effect of high signal intensity where the contrast agent was fixed. Histogram matching between subjects was not performed, as this would result in a overfit of the method to the particular dataset used and decrease generalisation when applying the sCT generation algorithm to new MRI scans. CT values were cropped to the interval [-1000, 1500] HU. Figure 3.1 shows these steps of the process on an representative pair of images.

In addition to the scans, the RT structure set DICOM files were also imported to MATLAB, and were saved as binary masks of the corresponding contoured OARs and tumours. For each patient, the same transformation previously obtained from the registration algorithm was applied to these binary masks, which were originally in the CT grid. The CT and MRI volumes were cropped at the bottom end of the brainstem using the corresponding binary mask. This was done to include the region of the tumour, while leaving out the lower region of the FOV where the MRI signal becomes slightly degraded and higher deformation can occur (neck region). Padding was performed at the top of the head in order to have the same amount of slices on the volumes for all patients. The final pre-processed volumes were of size $256 \times 256 \times 192$ voxels with voxel size = $1 \times 1 \times 1$ mm^3.

3.2 sCT Generation

3.2.1 Network Architecture

This project made use of the *pix2pix* conditional GAN architecture provided in an implementation for PyTorch [102] as the basis for the sCT generation method. The *pix2pix* network was designed to be applied to 8-bit 2-dimensional RGB image data, for image-to-image translation tasks.

This GAN architecture uses a structured cost function, in which in addition to the GAN objective function from Equation 2.8, the generator is also trained using an absolute difference (L1 loss). This way, the generator is tasked to not only fool the discriminator, but also for the output $y' = G(x, z)$ to be near the ground truth in absolute value:

$$\mathcal{L}_{L1} = \mathbb{E}_{x,y,z} \left(|y - G(x, z)|_1 \right) \tag{3.1}$$

The generator is modelled using a 256 $\times$ 256 U-net, consisting of an encoding path until a bottleneck layer, at which point the process is reversed with a decoding path, to return a 256 $\times$ 256 output. Both the encoder and decoder are CNN models. Skip-connections are added to concatenate all channels from symmetric layers of both paths, i.e. layer i and layer $n-i$, where n is the total number of layers. During the contraction path feature information is extracted from the input, by the use of multiple convolutional layers with different filters (kernels) followed by downsample operations. The expansive pathway combines the feature and spatial information through a sequence of up-convolutions and concatenations with the features from the contracting path. The learnable weights of the network correspond to the values of the various filters used in each convolutional layer, each filter responsible for a feature map.

The discriminator consists of a series of convolutional layers, followed by a sigmoid funtion to return values between 0 and 1. The discriminator is actually a patch model, with a 70$\times$70 receptive field, and therefore returns an N $\times$ N probability matrix as opposed to single scalar corresponding to the entire image. The output is then compared to a N $\times$ N label matrix of ones for the real images and zeros for the fake images.

In this project, the original *pix2pix* architecture was adapted to the particular task of 3D image-to-image translation for medical volumes. The main changes to the algorithm were as follow:

1. In order to use volumetric medical data, the input and output format of the network were changed to NIfTI files and to allow for 16-bit data as opposed to 8-bit, in order to use the full HU range of the images. Since the original code made use of libraries specific to 8-bit PNG/JPEG images, this involved having to write new functions, including when reading the images from the folder and to convert the input volumes to tensors normalised to a -1 to 1 float scale in the format used by Torch. Inverse transformations to scale the output sCT to the corresponding HU values were defined.

2. The original algorithm was developed to allow for images with 3 channels (RGB), which was changed to 1 channel files only, as is the case of the medical images here used.

3. All the 2-dimensional operations in the networks were modified to 3-dimensional operations, such as convolution and max-pooling (downsampling) operations in the encoding path of the generator, transpose convolutions in the decoding path, batch normalization and concatenation operations.

4. The generator architecture was changed to allow for smaller input size in each dimension (the reasoning for decreasing the size is further explained in section 3.2.2). The network was therefore modified from a 2-dimensional 256 U-net to a 3-dimensional 64 U-net. This involved decreasing the number of layers in the encoding and decoding path. The network architecture was slightly modified to add an additional convolution block in the first (outermost) layer before downsampling, and in the last (outermost) layer before producing the output.

5. The discriminator architecture was modified to have a 34×34×34 receptive field size.

Figure 3.2 shows the architecture of the modified 3D U-net generator and the convolutional discriminator. In the encoding path of the generator, the convolutions conserve the output size and, with the exception of the innermost and outermost blocks, duplicate the number of feature maps. 2×2×2 max pooling operations downsample the outputs of the convolutions by a factor of two. In the decoding path, the convolution operations upsample by a factor of 2 by using a stride of 2, with the exception of the outermost layer. Skip-connections that concatenate layers of equal resolution are represented by the grey arrows. Batch normalization (normalization of the weights) is applied to all layers except the innermost layer. The final convolution in the U-net maps the result to 1 channel, and is followed by a Tanh function to return values in the range -1 to 1. The discriminator, has 4 convolutional layers, where the first two downsample by a factor of 2, and the last layer is followed by a sigmoid function to return a 10×10×10 probability matrix, each value corresponding to a 34×34×34 receptive field on the input volume.

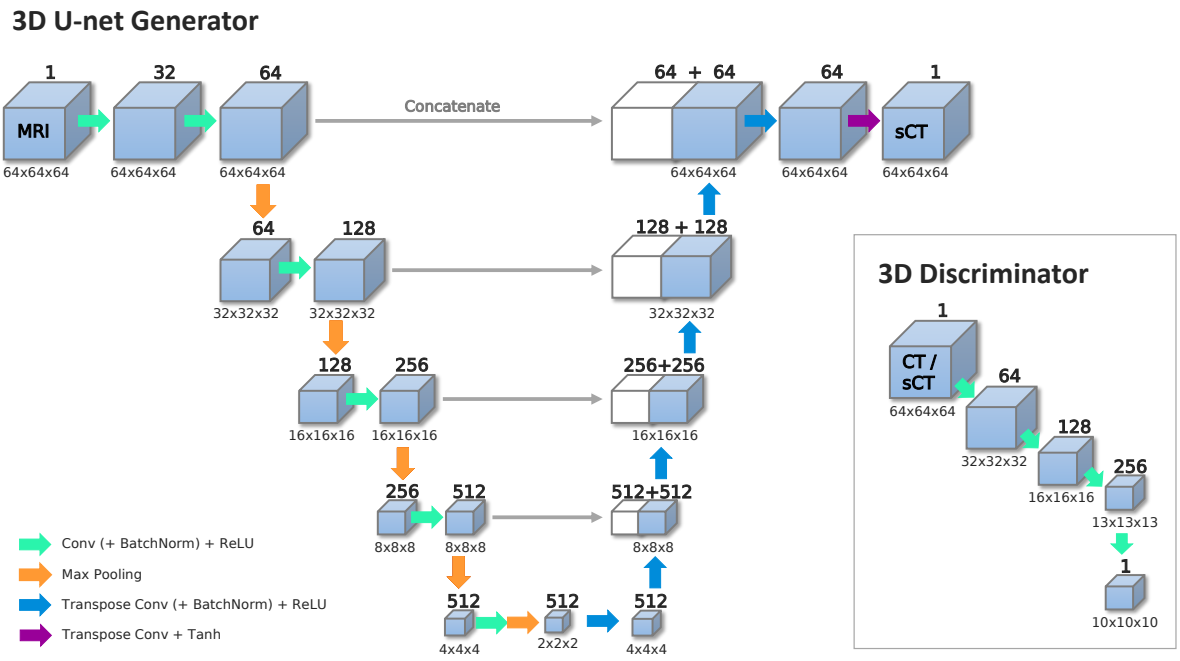

Figure 3.2: Architecture of the 3D U-net generator and discriminator networks used in the GAN. The cube sizes represent the resolution of the layer, which is detailed under each cube, while the number of channels (or feature maps) is detailed on top of each cube (corresponding to the fourth dimension)

3.2.2 Training

Training and validation of the model was performed on Dataset A using a 7-fold cross-validation technique. The dataset was randomly divided in 7 subsamples of 6 patients each. One subsample was

retained as the validation data for testing the model, and the remaining 6 subsamples (36 patients) were used as training data. The cross-validation process was repeated 7 times, with each of the 7 subsamples used once as the validation data, and with no leakage between training and validation sets in each fold.

For comparison of the networks, the original 2D *pix2pix* code was also trained and tested, using pairs of 2D axial slices of the MRI and CT volumes. The HU values were converted to a 0-255 scale by linear scaling. The images were then saved as 8-bit PNG files. Since a 3D network is more complex that a 2D network, its training requires high computational power. Taking this fact into account, and considering the small number of patients used in this study, the network was trained with subvolumes, or patches, of the pairs of volumes. The volumes were therefore divided into intersecting subvolumes of size 64×64×64 voxels, for a total of 245 subvolumes per scan, saved as 16-bit NIfTI files. These patches are small enough to allow for training on a 4 GB GPU, but large enough to be able to capture anatomic structures. Using subvolumes of the original image allows for more training data (245 cubes per patient), at the same time as the network has less parameters (learnable weights) than if using the original 256×256×192 volume, for the same amount of convolutional layers. This means that there is a smaller chance of overfitting the model to the training data, which is common when training a network with a very large number of parameters and a small training dataset. In addition, data augmentation was performed by left/right flipping the training images, both for the 2D and the 3D network. The synthesis of the final sCT volumes at the generation time was done by assembling the subvolumes together. For the voxels containing more than one prediction (maximum of 8) due to the intersection of the pacthes, simple averaging was performed.

Optimisation of hyperparameters (parameters whose values are set before the learning process begins) was performed for the networks, by analysing the performance of the GAN on the cross-validation data for a set of different hyperparameters. The performance was assessed using the mean absolute error (MAE) between the generated and real CT volumes (detailed in section 3.3). When training with the original algorithm, most default hyperparameters were kept the same, and only the number of iterations was optimised. When adapting the architecture to 3D, further hyperparameters were tuned: kernel size of the generator (values: 3×3×3, 4×4×4), number of feature maps in the initial convolutional block (values: 32, 64) and learning rate (values: 2×10^{-2}, 2×10^{-4}, 2×10^{-5}). Hyperparameter optimisation was not done extensively due to time limitations, and because it did not represent the main aim of this project, as the original 2D GAN had already been optimised to an extent.

The weights of the networks were randomly initialised from a Gaussian distribution with mean 0 and standard deviation 0.02, and a stochastic gradient descent algorithm was used for training. A mini-batch size (number of training samples passed through the network at each iteration) of 1 was used.

Dataset B (12 patients) was used to assess the generalisation of the algorithm, and therefore used as a holdout dataset. Testing on this data was carried out using the same GAN, for the best architecture and set of hyperparameters found through cross-validation, and trained with 36 patients (randomly selected) of dataset A. No further optimisation was implemented using this dataset.

Training and testing was executed on PyTorch (Python version 3.7), and using a NVIDIA GeForce GTX 1050 Ti GPU with 4GB random access memory. The final sCT volumes were exported as NIFTI files to MATLAB for analysis. Additionally, they were exported as DICOM files to simulate the process for a clinical workflow.

3.3 sCT Evaluation

The generated sCT images were compared with the pre-processed clinical CT images, visually and quantitatively, using MATLAB 2018a. To visually compare the generated CTs with the real CTs, difference images were produced by subtracting the HU values of the sCT to the CT on a voxel by voxel basis. To compare the CT images for different tissue classes, binary masks were created for bone, air and soft tissue regions by thresholding on the clinical CT. Voxels with an intensity higher than 250 HU were classified as bone, voxels with value lower than -200 HU were classified as air and the remaining voxels inside the body contour were classified as soft tissue, as done by Aouadi *et al.* [39].

3.3.1 Image Quality Metrics

To quantitatively evaluate the quality of the synthetic CTs, a set of image quality metrics were calculated:

- **Mean Error -** represents the bias in the estimation of the HU values, and is given by:

$$ME = \frac{1}{N}\sum_{i=1}^{N} sCT_i - CT_i \tag{3.2}$$

  where CT_i and sCT_i represent the HU value of the clinical and synthetic CT for each voxel i.

- **Mean Absolute Error -** is the average absolute error in the estimation of the synthetic CT, and therefore does not consider the direction of the error:

$$MAE = \frac{1}{N}\sum_{i=1}^{N} |sCT_i - CT_i| \tag{3.3}$$

  Both the ME and the MAE were calculated for all the voxels inside the body contour, as well as for the bone, air and soft tissue classes defined on the real CT.

- **Structural Similarity Index -** is a quality metric used to measure the similarity between two images. It is defined by the combination of three terms that correspond to three characteristics of an image - luminance, contrast and structure:

$$SSIM = \left[\frac{2\mu_{CT}\mu_{sCT} + C_1}{\mu_{CT}^2 + \mu_{sCT}^2 + C_1}\right] \cdot \left[\frac{2\sigma_{CT}\sigma_{sCT} + C_2}{\sigma_{CT}^2 + \sigma_{sCT}^2 + C_2}\right] \cdot \left[\frac{\sigma_{CT\,sCT} + C_3}{\sigma_{CT} + \sigma_{sCT} + C_3}\right] \tag{3.4}$$

  where $\mu_{CT}, \mu_{sCT}, \sigma_{CT}, \sigma_{sCT}$ and $\sigma_{CT\,sCT}$ are the local means, standard deviations and cross-covariance for clinical CT and sCT, and $C_1 = (0.01 \times L)^2$, $C_2 = (0.03 \times L)^2$ and $C_3 = C_2/2$, where L is the dynamic range of the images. In this case L = 2500 as the images have intensities ranging from -1000 to 1500 HU.

  The SSIM was calculated using a sliding Gaussian window of size $11 \times 11 \times 11$ and standard deviation of 1.5, and the mean SSIM (MSSIM) was calculated for the entire images by averaging. The MSSIM has a value between -1 and 1.

- **Dice Similarity Coefficient of Bone -** is a measure of the overlap between the bone class in the clinical CT and in the synthetic CT, being an indicator of structural quality of the image, and is defined as:

$$DSC_{bone} = \frac{2(V_{sCT}^{bone} \cap V_{CT}^{bone})}{V_{sCT}^{bone} + V_{CT}^{bone}} \tag{3.5}$$

where V_{sCT}^{bone} and V_{CT}^{bone} represent the volume, defined by the number of voxels, of the bone class region in the synthetic CT and real CT, respectively.

3.3.2 Statistical Comparisons

To compare the improvements in MAE values with the use of the 3D GAN architecture versus the 2D GAN architecture, a one-sided paired Wilcoxon signed rank test (H_0: the median of the distributions is equal), was used. To compare the MAE values - using the same GAN architecture and hyperparameters - on the cross-validation stage (dataset A) and test stage (dataset B), a Mann-Whitney test (H_0: the distributions of both samples are equal) was used. Non-parametric test were chosen due to the non-normality of the MAE distributions, assessed by histogram analysis. All tests were performed using the statistics tool R and analysed for a standard level of significance of 0.05.

3.4 Dosimetric Evaluation

3.4.1 Treatment Planning

The dataset used in this project consisted of images acquired for radiosurgery planning. The images were extracted before being transfered to the treatment planning stage, therefore no radiotherapy plans were available. To evaluate the quality of the generated sCT scans for photon and proton therapy, simple yet clinically relevant treatment plans were created using matRad, an open source software implemented on MATLAB for radiation treatment planning for research and education purposes.

The CT and sCT were imported to matRad. In order to convert the CT and sCT numbers to electron

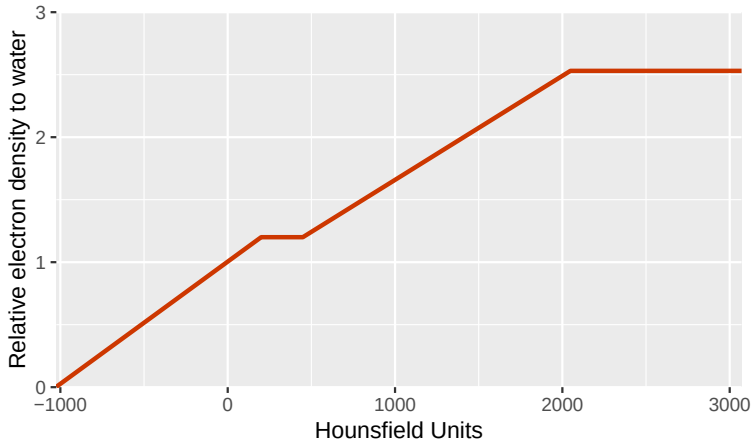

Figure 3.3: Default HU-to-density conversion curve used by matRad.

density information - necessary for dose calculations - matRad's default Hounsfield Units lookup table (HULT) was used. The conversion from HU values to relative electron density based on this HULT is ilustrated in Figure 3.3

The clinically defined RT structure sets were uploaded to matRad and used to define volumes of interests in the planning CT and sCT. As these structures sets contained only the GTV contour, and not the CTV or PTV, this was used as the target for the treatment planning. For all patients, an objective was defined to deliver 60 Gy of dose to the target in 30 fractions of 2 Gy each. The brainstem was defined as an OAR for its proximity to the tumour in all patients, with a constraint of $D_{max} < 54$ Gy, following published clinical guidelines [121]. Table 3.1 shows the parameters used for treatment optimisation as defined on matRad.

Table 3.1: Fluence optimisation parameters and constraints

VOI	VOI type	Priority	Objective Function	Penalty	Dose Parameter
GTV	Target	1	Square Deviation	1000	60 Gy
Brainstem	OAR	2	Max Dose Constraint	1000	54 Gy

matRad uses a decomposed pencil beam algorithm for dose calculation. The setup of the beams (number, gantry angle and couch angle) are defined by the user. The beams are divided into a set of smaller beamlets and the geometry of irradiation (source point of each beamlet and number of beamlets per beam) is calculated by matRad based on the beam setup and geometry of the target. Dosimetric information is generated by pre-computing dose influence matrices for the inverse planning. Then, fluence optimisation can be performed to find a set of beamlet weights to modulate the radiation intensity to yield the best possible dose distribution according to the clinical objectives and constraints.

In this case, fluence optimisation was computed only on the synthetic CT, to simulate an MRI-only planning workflow. Therefore, for each patient, the dose distribution in the clinical CT was directly calculated with the irradiation geometry and beamlet weights previously optimised in the corresponding sCT, without any further optimisation. This allowed to examine the quality of the sCT for plan optimisation and to analyse dose differences resulting from HU variations between sCT and CT only.

Photon Therapy

For photon therapy treatment planning, a 5-beam IMRT setup with fixed gantry angles of 0°, 72°, 144°, 216° and 288° was implemented. Coach angles were fixed at 0° and treatment plan isocenters were located at the centre of mass of the GTV for each patient. Dose calculations were performed using 6 MV photons and over a $1\times1\times1$ mm^3 grid. The treatment plan parameters and beam configuration used for photon IMRT planning on matRad are shown in Figure 3.4. Note that the figure shows the matRad GUI, which was used to define the parameters and assess the dose distribution for a small sample of patients. The script version of matRad was then modified and used when defining the parameters and calculating the dose for all patients, in both sCT and CT, in order to automate the process.

Proton Therapy

For proton therapy, a setup with 2 perpendicular (wedged) beams was defined. For patients with the tumour located on the left side of the head, gantry angles were set for 72° and 162°; for patients with tumours located on the right side of the head, beams were simulated at 198° and 288°. Once again, coach angles were fixed at 0° and treatment plan isocenters were located at the centre of mass of the GTV for each patient. Dose calculations were performed over a $1\times1\times1$ mm^3 grid. The treatment plan parameters and beam configuration used for intensity modulated proton therapy (IMPT) planning on matRad are shown in Figure 3.5.

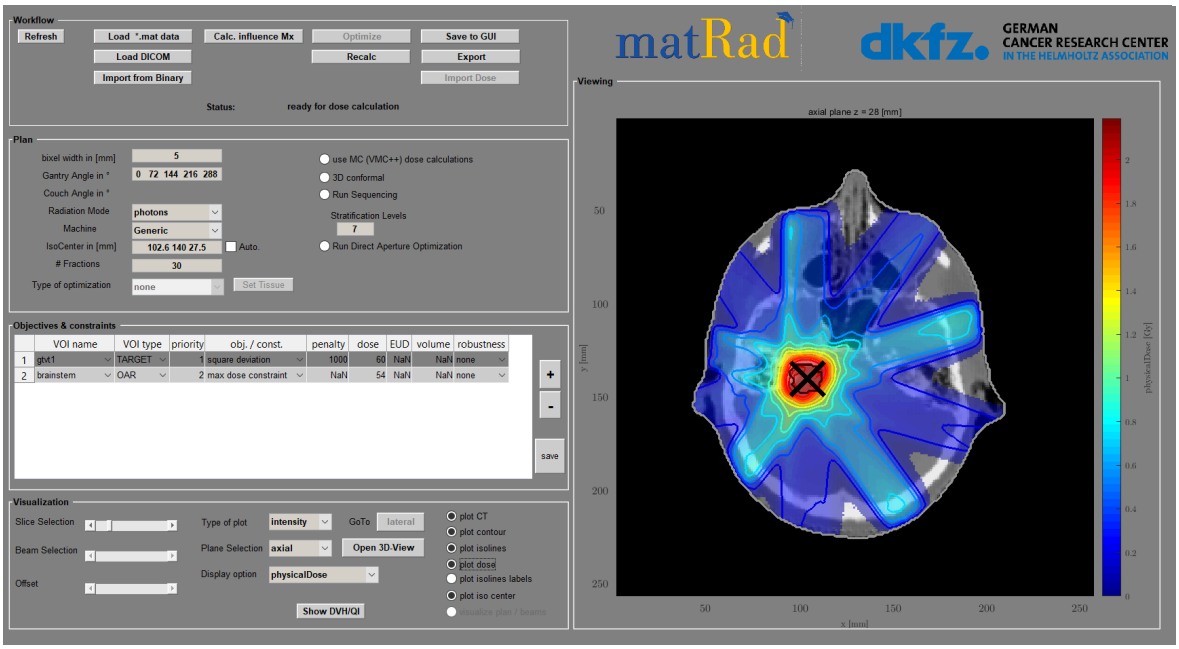

Figure 3.4: matRad configurations and beam arrangement defined for the IMRT treatment plan

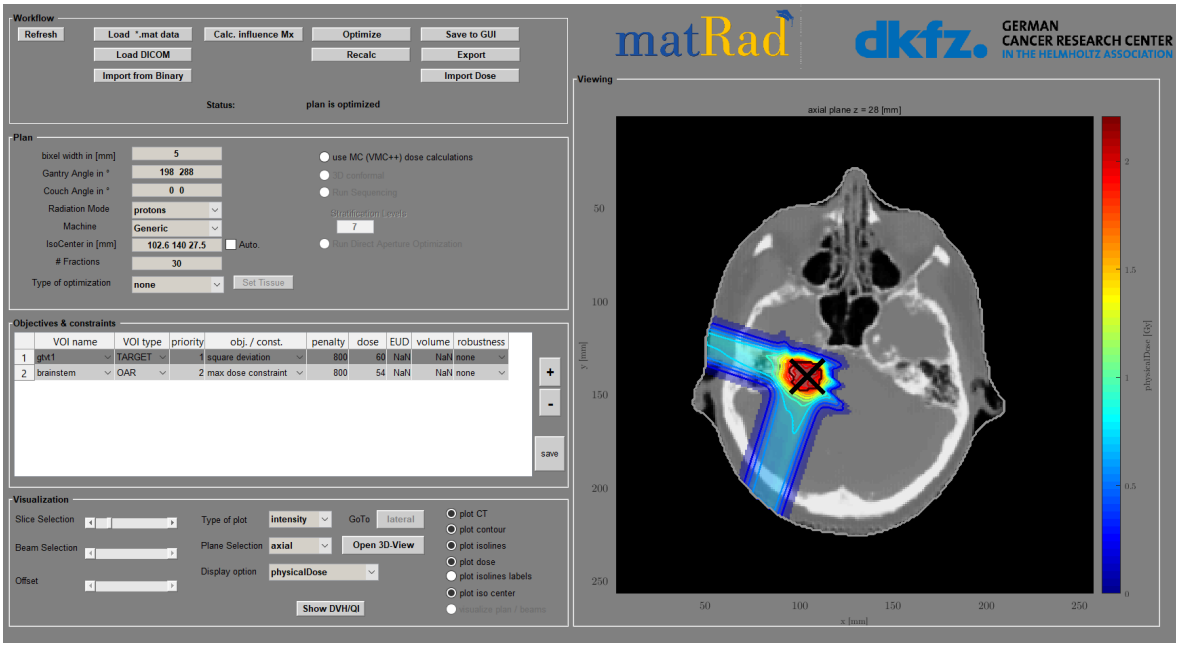

Figure 3.5: matRad configurations and beam arrangement defined for the IMPT treatment plan

3.4.2 Dose Comparison

Dose cubes were analysed in MATLAB. For each patient, the dose distribution that was optimised on the sCT was considered as the reference for comparisons. Difference maps were obtained by calculating the difference between the dose distributions in the real CT and synthetic CT on a voxel-by-voxel basis. The mean differences were computed over the entire body region, for the GTV and for OARs.

Gamma Analysis

The dose distributions were also analysed using gamma analysis based on the gamma index [122,123]. This analysis is commonly used for quality assurance of dose distributions in radiotherapy settings.

The gamma index compares an evaluated dose distribution with a reference dose distribution according to an acceptance criteria of dose difference (DD), ΔD, and distance to agreement (DTA), Δd. In this case the reference was defined as the dose distribution in the sCT and the evaluated distribution was the dose computed on the clinical CT. The gamma index for a point $\boldsymbol{r}_{sCT} = (x_{sCT}, y_{sCT}, z_{sCT})$ is given by:

$$\gamma(\boldsymbol{r}_{sCT}) = \min\{\Gamma(\boldsymbol{r}_{sCT}, \boldsymbol{r}_{CT})\}\forall\{\boldsymbol{r}_{CT}\} \tag{3.6}$$

where

$$\Gamma(\boldsymbol{r}_{sCT}, \boldsymbol{r}_{CT})\} = \sqrt{\frac{|\boldsymbol{r}_{CT} - \boldsymbol{r}_{sCT}|^2}{\Delta\, d^2} + \frac{(D_{CT}(\boldsymbol{r}_{CT}) - D_{sCT}(\boldsymbol{r}_{sCT}))^2}{\Delta\, D^2}} \tag{3.7}$$

The gamma index is therefore taken as the minimum gamma value calculated over all points from the evaluated distribution. The Δd and ΔD criteria form an ellipsoid around the reference point. The reference voxel passes the gamma index analysis if $\gamma < 1$, meaning an evaluated point is located within this ellipsoid. For each patient, the gamma pass rate is the percentage of voxels of the sCT dose distribution that pass the gamma index analysis for a certain pair of DTA and DD criteria.

The gamma analysis was performed excluding low dose points, by using a low-dose threshold of 10% of the prescribed dose, as frequently done in clinical setting [124]. Both 2mm/2% and 3mm/3% acceptance criteria were considered. The gamma indeces and pass rates were computed using a MATLAB implementation of the algorithm.

4 Results and Discussion

4.1 sCT Generation

For the initial 2D GAN implementation using the original *pix2pix* algorithm, the best average MAE results were obtained for 10 epochs (number of times an entire dataset is passed through the network), corresponding to approximately 54000 iterations. Training time was approximately 10 h per fold (training with 36 patients), totalling 70 h for the cross-validation process. Synthesis of the sCTs took approximately 15 s per sCT volume. For the adapted 3D GAN, the best MAE results were obtained for the following tested hyperparameters: kernel size $3 \times 3 \times 3$, number of filters in the first layer = 32, initial learning rate = 0.0002. Training was performed for 20 epochs, corresponding to approximately 175000 iterations. Training the final network on 36 patients took approximately 30 h. The total duration of the cross-validation process for each set of hyperparameters tested was approximately 210 h. Generation time was less than 30 s per patient.

4.2 Image Quality Evaluation

4.2.1 2D sCT Generation

Table 4.1 presents the results for the ME and MAE metrics for the 7-fold cross-validation, on a first stage, using the original pix2pix algorithm. The average MAE between the pre-processed CT and the sCT for the voxels within the body contour, over the entire dataset, was 87 $\pm$ 11 HU (range: 60-124 HU).

Table 4.1: Error metrics for the sCT volumes generated using the original 2D GAN architecture with a 7-fold cross-validation method on Dataset A.

	ME_{Body} (HU $\pm$ 1 SD)	MAE_{Body} (HU $\pm$ 1 SD)	MAE_{Air} (HU $\pm$ 1 SD)	MAE_{Bone} (HU $\pm$ 1 SD)	$MAE_{SoftTissue}$ (HU $\pm$ 1 SD)
Fold 1	-5 $\pm$ 10	90 $\pm$ 7	305 $\pm$ 45	190 $\pm$ 18	50 $\pm$ 5
Fold 2	-15 $\pm$ 11	79 $\pm$ 9	297 $\pm$ 16	176 $\pm$ 20	47 $\pm$ 7
Fold 3	3 $\pm$ 20	85 $\pm$ 10	319 $\pm$ 58	173 $\pm$ 28	50 $\pm$ 5
Fold 4	-9 $\pm$ 10	82 $\pm$ 12	280 $\pm$ 47	181 $\pm$ 37	49 $\pm$ 17
Fold 5	-6 $\pm$ 15	87 $\pm$ 10	289 $\pm$ 43	176 $\pm$ 25	51 $\pm$ 8
Fold 6	6 $\pm$ 29	83 $\pm$ 11	340 $\pm$ 43	160 $\pm$ 18	46 $\pm$ 7
Fold 7	-23 $\pm$ 19	99 $\pm$ 11	335 $\pm$ 53	228 $\pm$ 40	52 $\pm$ 4
Mean	**-6 $\pm$ 17**	**87 $\pm$ 11**	**309 $\pm$ 47**	**183 $\pm$ 32**	**49 $\pm$ 9**

The mean signed error in HU estimation was negative but close to 0 HU, showing there was no systematic trend for the network to underestimate or overestimate CT values. On average, the largest absolute differences were observed for the voxels belonging to the air class, followed by the bone and soft-tissue classes.

4.2.2 3D sCT Generation

The error metrics calculated for the synthetic CT volumes generated using the adapted 3D GAN architecture are shown in Table 4.2. The average MAE calculated over the 42 patients, when considering the voxels inside the body contour, was 69 $\pm$ 10 HU (range: 44-98 HU). The mean bias over the entire patient dataset was again close to 0 HU.

Table 4.2: Error measures for the sCT volumes generated using the 3D GAN architecture with a 7-fold cross-validation method on Dataset A.

	ME_{Body} (HU $\pm$ 1 SD)	MAE_{Body} (HU $\pm$ 1 SD)	MAE_{Air} (HU $\pm$ 1 SD)	MAE_{Bone} (HU $\pm$ 1 SD)	$MAE_{SoftTissue}$ (HU $\pm$ 1 SD)
Fold 1	1 $\pm$ 6	77 $\pm$ 8	298 $\pm$ 37	162 $\pm$ 21	42 $\pm$ 4
Fold 2	1 $\pm$ 8	62 $\pm$ 8	274 $\pm$ 10	126 $\pm$ 18	35 $\pm$ 4
Fold 3	3 $\pm$ 9	70 $\pm$ 10	270 $\pm$ 33	152 $\pm$ 25	39 $\pm$ 6
Fold 4	-4 $\pm$ 4	64 $\pm$ 9	241 $\pm$ 37	140 $\pm$ 32	36 $\pm$ 6
Fold 5	-11 $\pm$ 6	65 $\pm$ 8	246 $\pm$ 28	142 $\pm$ 19	34 $\pm$ 4
Fold 6	10 $\pm$ 11	67 $\pm$ 8	294 $\pm$ 17	126 $\pm$ 12	40 $\pm$ 3
Fold 7	-8 $\pm$ 9	76 $\pm$ 9	282 $\pm$ 23	173 $\pm$ 23	40 $\pm$ 4
Mean	**-1 $\pm$ 9**	**69 $\pm$ 10**	**272 $\pm$ 32**	**146 $\pm$ 26**	**38 $\pm$ 5**

When comparing the results in terms of quality metrics for the sCT volumes generated using the original 2D GAN with the images generated using the optimised 3D GAN, it can be concluded that the 3D approach is able to generate sCT volumes with lower mean error measures. The decrease occurs for

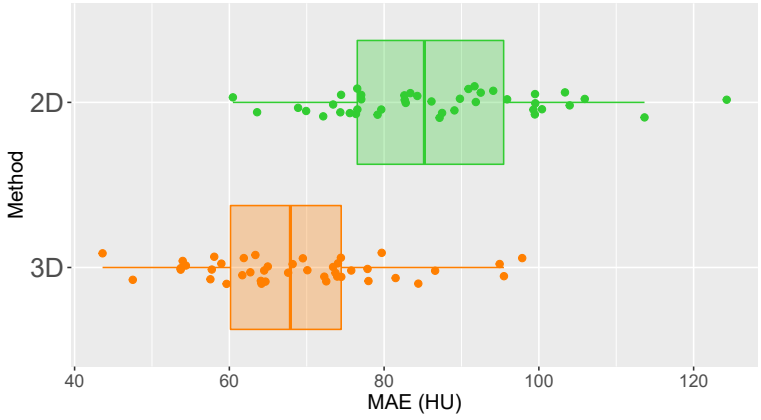

Figure 4.1: Boxplots representing the MAE calculated over the entire body region for synthetic CT images generated using the original 2D GAN architecture (green) and the modified 3D network (orange). The respective points represents the MAE for each patient in dataset A.

all classes of tissues. The average MAE_{Body} across the dataset saw a decrease of 20%. Figure 4.1 shows boxplots of the MAE values obtained across the patient dataset for sCT generated using both methods. It is visible that the use of the improved 3D GAN led to a shift in the values across the entire patient population. By analysing the boxplots it is also visible that the distribution of MAE values is slightly more symmetrically distributed around the median for the 3D method, and the interquartile range is smaller, indicating lower variability in values. The most extreme outlier in the first boxplot had a decrease in MAE from 124 to 98 HU. The effect of using an adapted 3D method over the original 2D model was tested with a one-sided paired Wilcoxon test. The null hypothesis can be rejected ($p = 5 \times 10^{-13}$), concluding that the medians of the distributions differ, and therefore the improvements in MAE were statistically significant.

Figure 4.2 shows sagittal and axial views of MRI, CT and sCT, as well as difference maps produced by subtracting the HU values in the CT (measured values) from the HU values in the sCT (estimated values), for 2 patients. The first patient is an example average patient, for whom the MAE in the generated sCT was 62 HU, while the second patient is a worst performing case in terms of the image quality metrics, with an MAE of 98 HU.

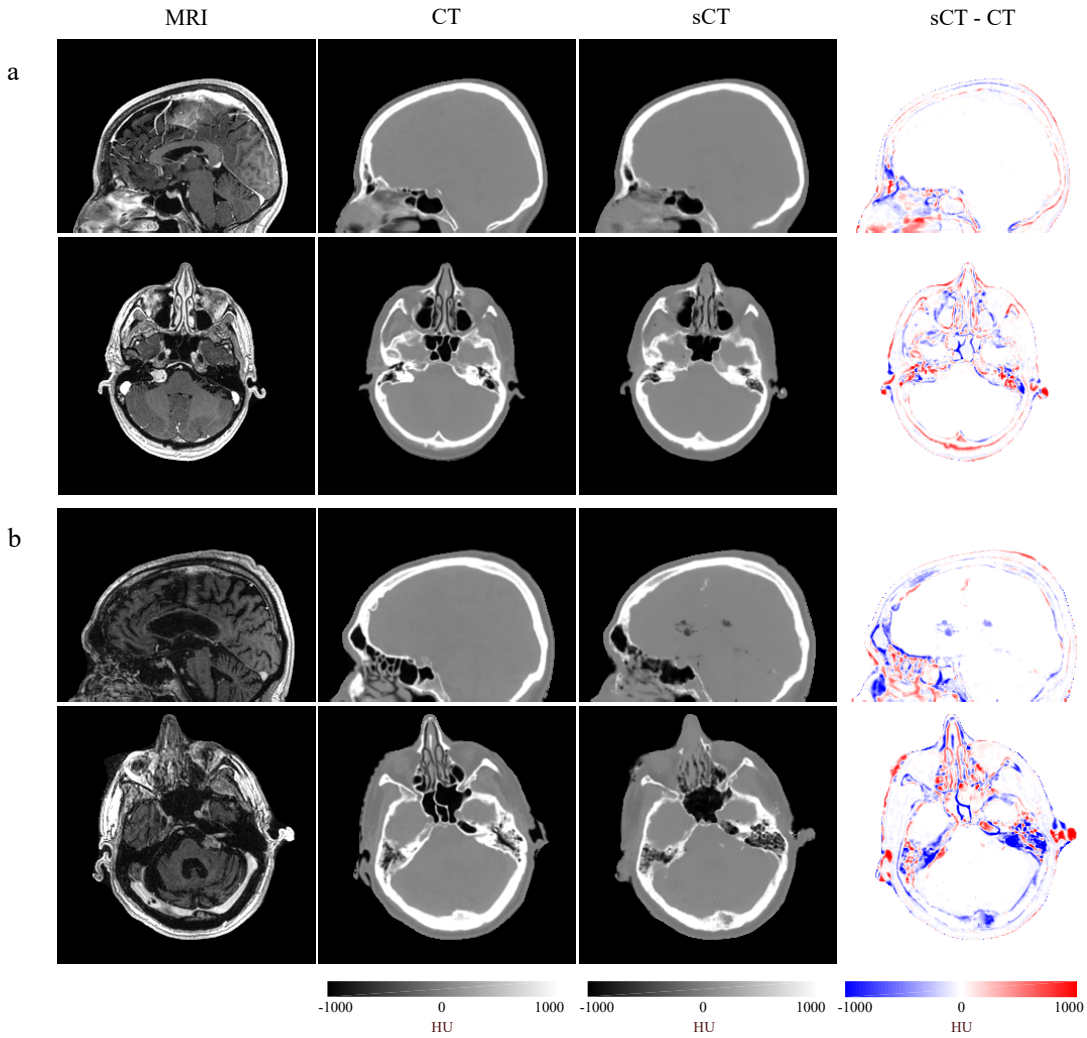

Figure 4.2: Sagittal (mid-plane) and axial (nasal level) views of MRI (input to 3D GAN), CT, synthetic CT and corresponding difference maps for a) a representative patient with MAE_{Body} = 62 HU and b) worst quality case with MAE_{Body} = 98 HU.

Analysing the pictures for both patients, its is visible the synthesized CT is capable of generally capturing the value distribution of the original CT. The network has the best performance in the soft tissue, particularly in the brain region, where the difference map is closer to 0 HU. The difference maps also show there is greater accuracy in the neurocranium region, while the network struggles to reproduce high accuracy results in the viscerocranium region. This is visible in the blue and red regions of the difference maps, which are more prominent in the tissue/air interfaces of the nasal region, as well as small bone structures existent in the base of the skull. This higher discrepancies also translate in the average MAE metric across all patients: when considering only voxels belonging to the nasal region, the value rises to 107 $\pm$ 15 HU when compared with 69 $\pm$ 10 HU for the entire body region.

The higher errors between the sCT and CT for the worst case patient in Figure 4.2 b) are visible in the regions of the airways, where the sCT does not correctly capture the distribution of the HU values, resulting in blurry regions, as well as in the brain and in the thickness of the occipital and frontal bones, with the GAN algorithm underestimating the HU values for these regions. This translates in high errors for the voxels belonging to the bone class (MAE_{Bone} = 217 HU) and a highly negative mean signed error (ME_{Body} = -28 HU). Poor results obtained for this and other two patients with MAE values higher than 90 HU can be due to differences between training and testing subjects, both in patient anatomy and in MRI intensity distributions in the body. Errors can also be attributed to differences in positioning - as the MRI was not acquired in treatment position and no immobilisation mask was used - and registration errors between CT and MRI pairs. The effect of image registration and patient positioning on the quality of the generated synthetic CT images is further evaluated and discussed in section 4.2.4.

In order to further visualize the differences in HU values between sCT and corresponding CT across the head, profile plots were created for lines crossing the head at 3 different axial planes, which are visible in Figure 4.3. In plot a), corresponding to the upper portion of the head, it is visible the synthetic-CT has a profile very close to the CT, as the line crosses the skull and brain. In plot b), the synthetic-CT correctly mimics the profile for the region of the back of the head, but it fails to correctly reproduce small air cavities and bone structures when the line crosses the nasal region. This is again visible in plot c), where the profile line for the sCT fails to reproduce an upward spike (around voxel 173) in HU values, corresponding to a fine bone structure.

Figure 4.4 shows sCT estimation absolute errors as function of the CT value for a representative patient. It can be seen the smallest errors occur for the voxels corresponding to soft tissue (-100 to 100 HU), which is also the range that has the highest frequency in the image, mostly corresponding to the brain region that occupies most of the volume. In accordance to what is observed in Figures 4.2 and 4.3, the highest errors occur for values between -900 and -200 HU. These ranges of values mostly correspond to transitions between structures and often are due to partial volume effects. This indicates that the higher MAE obtained for the air class is due to larger errors for this range, while the error decreases for HU values inferior to -900 HU, corresponding to well defined, larger air cavities. A similar trend is observed in the other end of the spectrum, with higher diferences for low-density bone structures and the error decreasing for HU values higher than 800 HU, corresponding to the skull. Difficulties shown by the method to accurately reproduce small structures and interfaces can be due to a tendency of the convolutional network to estimate average values for low-value/high-value interfaces. This can be further accentuated by the use of intersecting subvolumes of the MRI as input to the network, and producing the final sCT volume by averaging for the voxels for which there is more than one predicted value. Although this method leads to fewer discontinuities between patches in the final sCT, it can lead to further blurring of these regions.

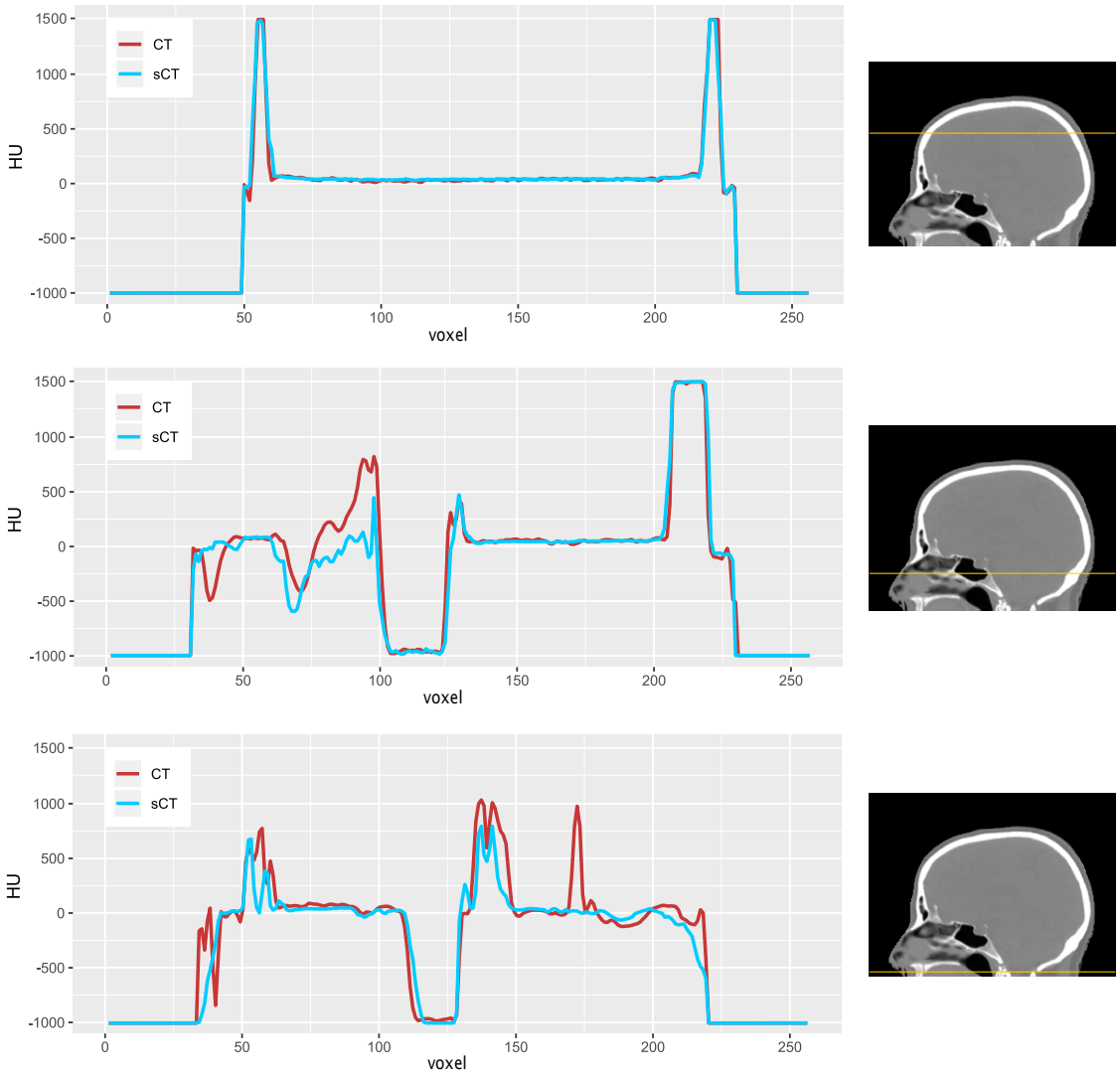

Figure 4.3: Profile plots of the HU values in real and synthetic CT at the mid-sagittal plane for three different axial planes, for a representative patient with MAE_{Body} = 70 HU. The corresponding path lines are represented in yellow on the CT scan.

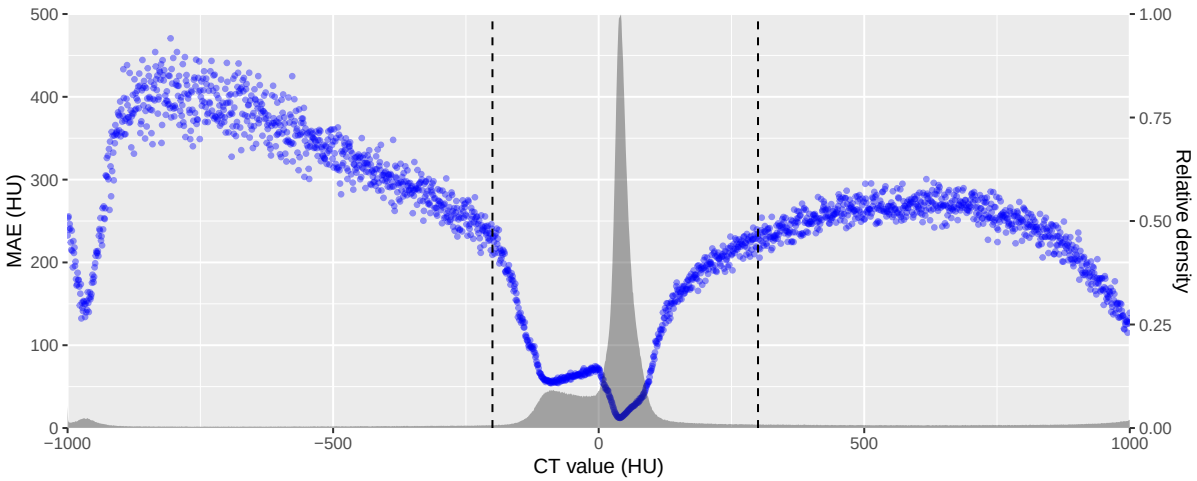

Figure 4.4: MAE as a function of the HU value in the original CT (blue) for the voxels inside the body contour for a representative sCT volume. Each point represents the bin-wise mean MAE for a bin size of 1 HU. The secondary plot, in gray, represents the relative density (normalised to the maximum value) of voxels for each HU value in the CT.

The measured Dice Similarity Coefficient for the bone voxels across the dataset was DSC_{bone} = 0.89 $\pm$ 0.03 (range: 0.77-0.94), indicating a strong overlap between the bone regions in the real CT and the synthetic-CT. The average MSSIM was 0.96 $\pm$ 0.03 (range: 0.92-0.98), expressing a very high correlation between synthetic and real CT regarding their luminance, contrast and structure.

4.2.3 Holdout Dataset

When applying the trained GAN model to the MRI images from dataset B, the obtained error metrics were as follow: MAE_{Body} = 75 $\pm$ 10 HU (range: 53-89 HU), MAE_{Air} = 267 $\pm$ 68 HU, MAE_{Bone} = 146 $\pm$ 31 HU, $MAE_{SoftTissue}$ = 44 $\pm$ 6 HU. The MSSIM and $Dice_{Bone}$ metrics did not suffer significant changes when calculated on this new dataset when compared to the results for the cross-validation set (Table 4.3). Comparing the datasets using the Mann-Whitney test based on the MAE_{Body} metric (shown in Figure 4.5), the null hypothesis that the distributions of both populations are equal cannot be rejected for a significance level of 0.05 ($p = 0.09$), and it can be concluded that the results obtained for the holdout dataset did not differ statistically from the results for the cross-validation. These results indicate that there was no overfit of the network to the training/validation dataset when selecting the hyperparameters, as using the same parameters on a new datset produced results of similar quality. Although the results using this holdout method are a better indicator of the generalization of the algorithm when compared to using cross-validation alone, it is important to refer that although these patients are independent, they are still part of the same population, i.e. they have the same type of tumours, and the MRI and CT scans were obtained using the same protocols. This project did not explore the performance of the trained GAN on patients with differences in pathology (different types of tumours), nor MRI scans obtained with different scanners.

Table 4.3: Quality metrics (mean $\pm$ 1 SD) for the sCT volumes generated for the 7-fold cross-validation using Dataset A (42 patients) and for Dataset B (12 patients).

	MAE Body (HU)	DSC_{Bone}	MSSIM
Cross-validation (Dataset A)	69 $\pm$ 10	0.89 $\pm$ 0.03	0.96 $\pm$ 0.03
Holdout dataset (Dataset B)	75 $\pm$ 10	0.88 $\pm$ 0.03	0.95 $\pm$ 0.04

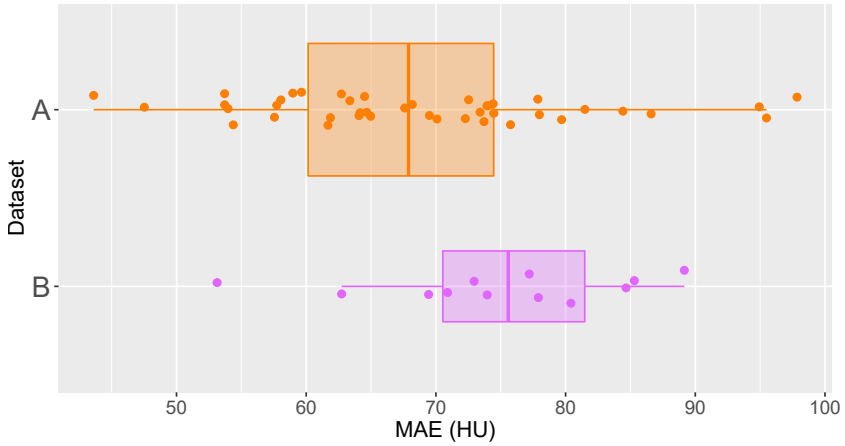

Figure 4.5: Boxplots representing the MAE calculated over the entire body region for 42 patients in the 7-fold cross-validation process (orange) and for the 12 patients in the holdout dataset (purple). The dots represent the MAE for each patient.

4.2.4 Effect of Suboptimal Image Registration

The results presented in the previous sections make use of the body contour that was created from the MRI scan. This was done to keep the MRI as the reference dataset, as it would happen in a MRI-only workflow where the CT scans would not be present. However, as it is visible in Figure 4.11, the registration of the pairs of MRI and clinical CT is suboptimal for some patients, in particular in the areas of the neck, nose and ears. This happens because even though the MRI and CT scan were acquired in the same day and no changes in anatomy occurred, there were differences in the acquisition of the scans, such as the use of an immobilization device in the CT. In addition, in some patients, portions of the immobilization device are still visible in the processed CT after the application of the MRI body binary mask.

Figure 4.6 shows sagittal and axial views of an example case where registration of the volume pair is suboptimal. It is visible from the figure that the contour of the generated sCT is more similar to the contour of the MRI, i.e. the GAN algorithm can, to some extent, overcome the registration errors and produce a volume that is structurally closer to the MRI. This however translates in inflated errors when comparing the sCT with the CT for the areas of misregistration. In the first difference map of the figure, this is visible in the red and blue zones in the peripheral areas of the neck and nose, as well as in the ear, which correspond to voxels that are within the patient contours of sCT but are background in the CT or vice-versa, which translates in voxel-wise absolute errors close to 1000 HU.

To overcome this, a body mask based on the CT was created using binarization and closing operations, and only the voxels in the intersection of the sCT and CT contours were considered for comparison, as done by Maspero *et al.* [48]. When calculatig a difference map using this mask, the differences in the contour of the body due to registration errors are not accounted for, as visualized in the second difference map in Figure 4.6, where the areas of high errors are eliminated from the edge of the image. When considering only the voxels belonging to the interception of these body contours, the MAE for the patient decreased from 94 HU to 85 HU, while the average MAE across the 42 patients of dataset A decreased from 69 $\pm$ 10 HU to 61 $\pm$ 11 HU.

These results show that the misregistration of the MRI and real CT translated in worse results for the quality metrics on the contour region of the patient. However, these registration errors are not limited to

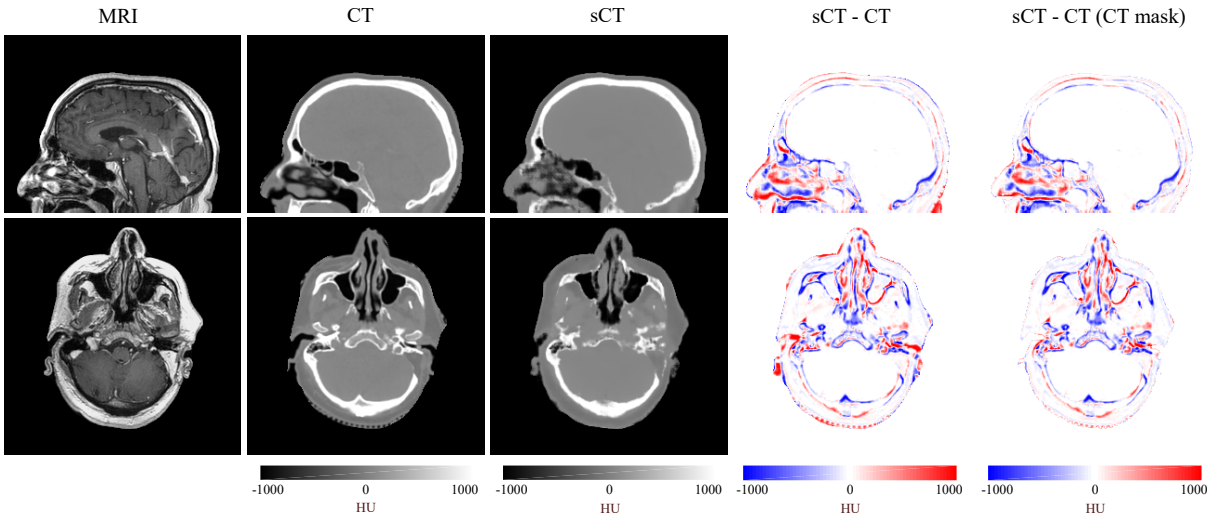

Figure 4.6: Sagittal and axial views of MRI, CT, synthetic-CT and corresponding difference maps based on the MRI body mask and on the CT body mask for a patient with suboptimal registration.

the body edges, and can also occur for the structures inside the patient's body. The suboptimal registration of MRI and CT has a dual effect, as it can influence the GAN algorithm during the training stage, leading to worse performance and conversely it can inflate the MAE metrics, because the sCT is not being compared to the exact density map of the MRI from which it was created - as it ideally would - but to the CT. Determining whether the differences between CT and corresponding sCT are due to poor performance by the GAN algorithm or due to registration errors is not an easy task, as there is currently no way to acquire a CT image that is in the exact same position and anatomy as the MRI. However, this effect can be greatly minimised if CT and MRI are both acquired in treatment position and using the same immobilization devices, which was not the case for the dataset here analysed. Another option would be the use of deformable registration in order to minimise the registration errors. This was not done in this project because the clinical norm for the head region is to use affine registration algorithms for registration of the planning image sets.

4.3 Dosimetric Evaluation

Evaluation of the quality of generated synthetic-CT volumes for dose calculation was performed using the synthetic-CT volumes generated using the 3D GAN architecture for 33 patients of Dataset A.

4.3.1 Photon Therapy

Mean voxel-wise differences between dose distributions calculated on synthetic-CT and real CT volumes for the 5-beam IMRT plans are shown in Table 4.4. Mean dose agreement between sCT and CT in the target volume was within 2% of the prescribed dose for all patients. The worst case mean differences at the tumour were -1.6%, -1.2%, -1.1% and +1.1%, while for the remaining 29 of 33 patients the mean dose difference was within 1% of the prescribed dose, corresponding to 0.6 Gy for the complete treatment course. Deviations in the considered OAR were small and within 0.5% of the prescribed dose for all patients. The average mean dose difference across the dataset was close to 0%, showing no tendency for the dose to be overestimated or underestimated when using the sCT.

Table 4.4: Mean voxel-wise differences between dose calculated on sCT and CT when considering all voxels inside the body contour, the GTV, and the OAR, for the 5-beam IMRT plan. Dose difference was calculated subtracting the dose calculated on the CT from the dose calculated on the sCT. Values are given in percentage of the prescribed dose.

	Body	GTV (Target)	OAR
Mean Dose Difference (%)			
Mean	-0.002	-0.12	0.02
Range	-0.03 to 0.03	-1.6 to 1.1	-0.12 to 0.19

Figure 4.7 shows results for the dosimetric evaluation on a representative patient for which the mean dose difference at the GTV was 0.7% of the prescribed dose. The difference map shows that the dose distributions are similar, with small discrepancies visible through the paths of the beams. The largest discrepancies occurred at the entry points of the beams and close to the tumour volume, where the dose builds up and is also higher. The differences visible at the entrance points of the beams can be due to subotpimal registration between MRI and CT as mentioned in section 4.2.4, as dose calculations were

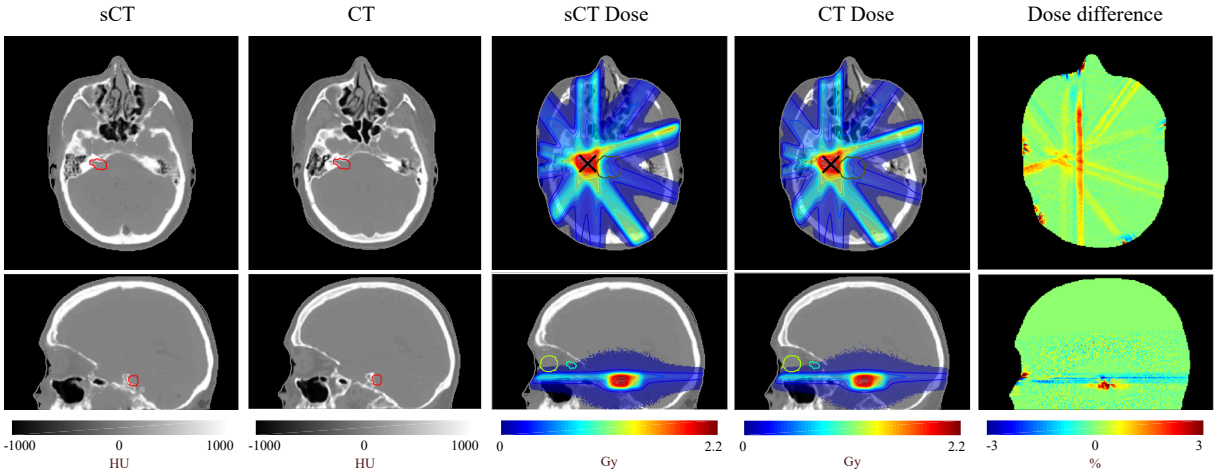

Figure 4.7: sCT, CT, dose calculated on the sCT and CT for the 5-beam IMRT plan and dose difference between the two for a representative patient with a vestibular schwannoma. Sagittal and axial views at the tumour isocentre plane.

performed using the original body contour, and not only for the intersection of the sCT and CT voxels. It is also visible in the difference map that the highest dose diferences occur in the anterior-posterior beam, shown in red. This beam travels through the air cavities of the nasal region, where it was previously shown that the errors in HU estimation are larger. Similar distributions of the dose differences were visible for the remaining patients of the dataset, as all tumours are located in similar regions of the brain.

For the same dose distributions represented in Figure 4.7, Figure 4.8 shows the cumulative dose-volume histogram (DVH) for all structures that were contoured on the original CT. It can be seen from the figure that the lines for the two calculated doses nearly overlap for the majority of the structures, indicating a very strong correlation between the plan calculated on the sCT and dose calculated using the real CT, for both the tumour volume and the organs at risk. It is important to note that although the figure shows the DVH curves for all structures in the CT, the plan was optimised only based on the target and the brainstem, with no dose constraints defined for the remaining OARs (and not all patients have the same number of organ contours). However, in this case, the goal of analysing the DVH was not to assess

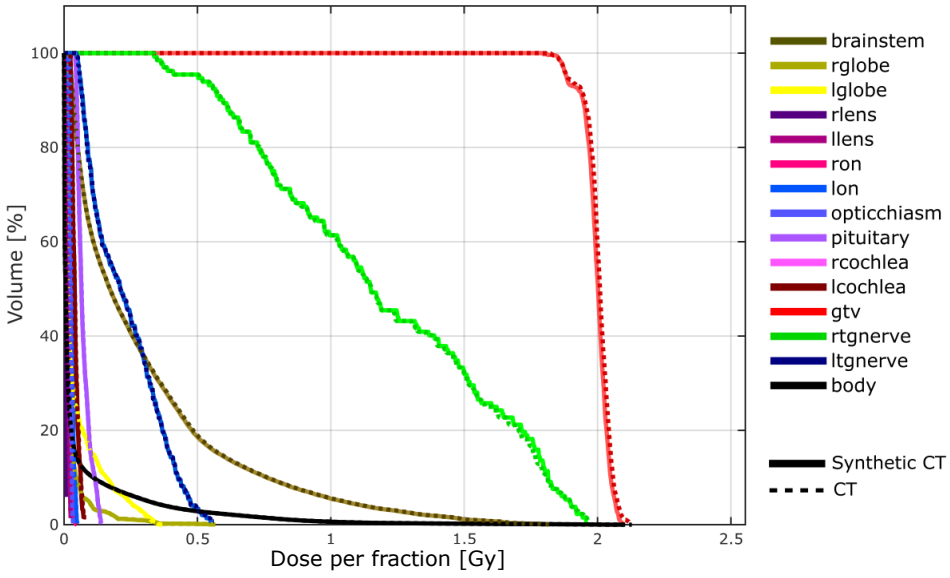

Figure 4.8: Cumulative DVH for the dose distribution calculated on the sCT (full line) and CT (dotted line) for the 5-beam IMRT plan for all structures delineated in the patient receiving dose.

the overall quality of the curves themselves, but rather the quality of the sCT when compared to the CT, i.e. the differences between both curves for each volume.

The DVH is a representation that assesses the overall dose planned for each volume, however it does not provide spatial information about the dose. To further analyse the quality of the dose calculated on the sCT, gamma analysis was performed between the two dose distributions. For a 3%/3mm criteria, all patients had a gamma pass rate superior to 99%, and the mean pass rate was 99.8 $\pm$ 0.1%. When using slightly stricter parameters of 2%/2mm, the mean pass rate across the dataset only slightly decreases to 99.4 $\pm$ 0.4%: 7 of the 33 patients had pass rates between 98% and 99%, while the remaining all had pass rates higher than 99%. These metrics indicate a very strong agreement between the dose distribution calculated on the synthetic-CT and using the CT, which is visible across the entire patient dataset.

4.3.2 Proton Therapy

Table 4.9 shows the results for the dose differences between CT and sCT calculated across the dataset for the proton therapy plan. In this case, the mean dose difference for the voxels belonging to the GTV was much larger than for the proton plans, and it ranged from -7.5% to 3.8%, corresponding to -5.3 Gy and 2.5 Gy, respectively for the total treatment plan of 60 Gy. For 21 of the 33 patients, the absolute mean dose differences at the target were within 2%. The mean across the entire dataset was negative and equal to -1.27%, with 23 of the 33 patients having negative dose differences, indicating a slight tendency for the planned dose to the tumour on the CT to be lower than the planned dose on the sCT. Larger differences when simulating proton therapy were also encountered for the OAR, where the voxel-wise mean dose difference for the worst case patient was 2.2%.

Table 4.5: Mean voxel-wise differences between dose calculated on sCT and CT when considering all voxels inside the body contour, the GTV, and the OAR, for the 2-beam proton therapy plan. Values are given in percentage of the prescribed dose.

	Body	GTV (Target)	OAR
Mean Dose Difference (%)			
Mean	-0.001	-1.27	-0.16
Range	-0.02 to 0.03	-7.5 to 3.8	-0.6 to 2.2

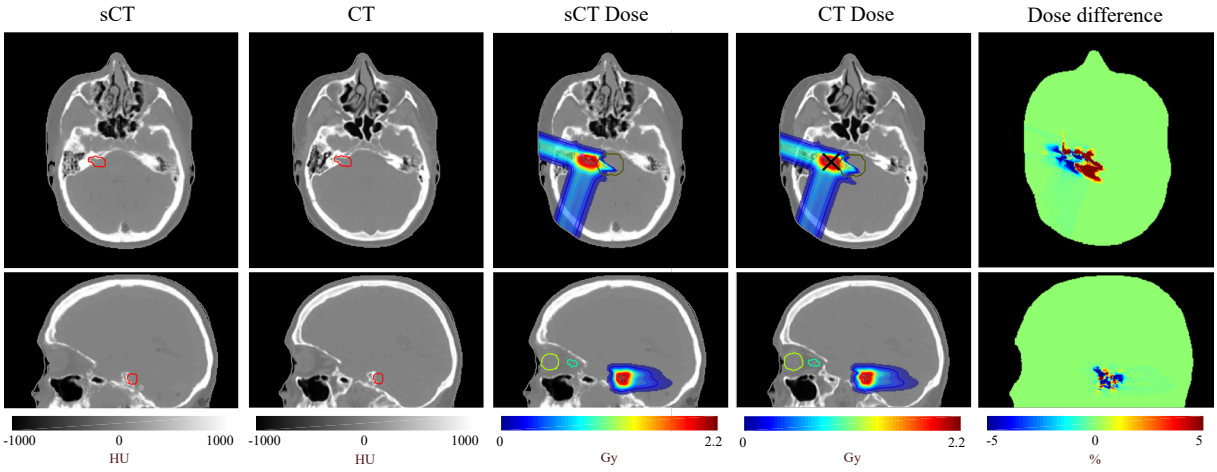

Figure 4.9: sCT, CT, dose calculated on the sCT and CT for the IMPT plan and dose difference between the two for a representative patient with a vestibular schwannoma. Sagittal and axial views at the tumour isocentre plane.

Figure 4.9 shows the dose distributions on the sCT and CT for the two-beam proton therapy plan for an example patient. The difference map shows that the highest dose differences occur in the region of the tumour, which is where the proton beams deposit most of their energy. It is also visible from the figure that the voxel-wise difference inside and around the tumour volume is highly heterogeneous. The high differences in dose when considering protons are due to the nature of charged particles, which make them more susceptible to changes in HU values across their path. These changes are detected in the small structures in the vicinity of the tumour, as previously mentioned, and are also visible in the thickness of the skull, which is crossed by one of the beams of the defined configuration. These differences result in changes in the range of the proton beams.

Higher differences between the dose calculated on the CT and the synthetic-CT when simulating a proton therapy plan are visible when analysing the DVHs for the plans. Figure 4.10 shows a DVH for the same patient represented in the previous figure. The plot shows there are discernible deviations between the curves for both the target and organs at risk, including the brainstem, as opposed to what was obtained for the photon therapy plan. For this patient, around 70% of the target volume had a planed dose equal to or greater than 2 Gy on the plan calculated on the sCT, while this value decreased to around 30% when the CT was used for calculation.

These higher discrepancies for proton plans are reflected in the measured gamma pass rates. For tolerance values of 3% and 3mm, the sCT volumes showed a mean pass rate of 99.1 $\pm$ 0.5% (range: 97.8-99.91%) for the dose distribution in comparison with the CT. For tighter error parameters of 2%/2mm, the mean pass rate decreases to 97.7 $\pm$ 1.1% (range: 94.0-99.5%), with 3 patients having pass rates around 95%, and 7 patients having rates higher than 99%. Although lower in comparison to the photon plans, most patients still showed high gamma pass rates for the IMPT, since this a measure of the overall accuracy of the plan for all voxels with D > 10% of prescribed dose, while the higher discrepancies in the proton plans are confined to the voxels inside and surrounding the target.

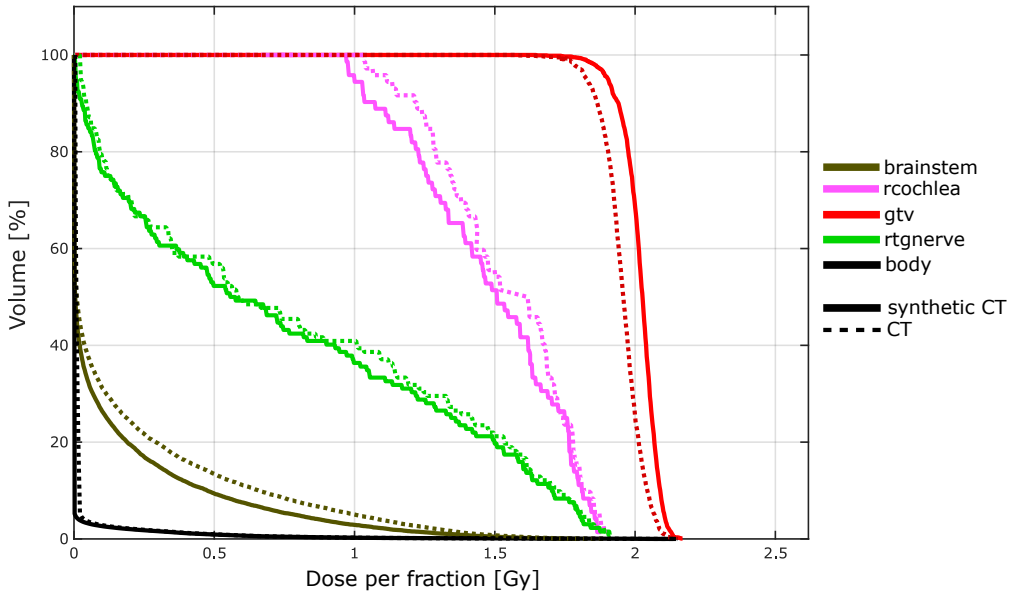

Figure 4.10: Cumulative DVH for the dose distribution calculated on the sCT (full line) and CT (dotted line) for the proton therapy plan for all structures delineated in the patient receiving dose.

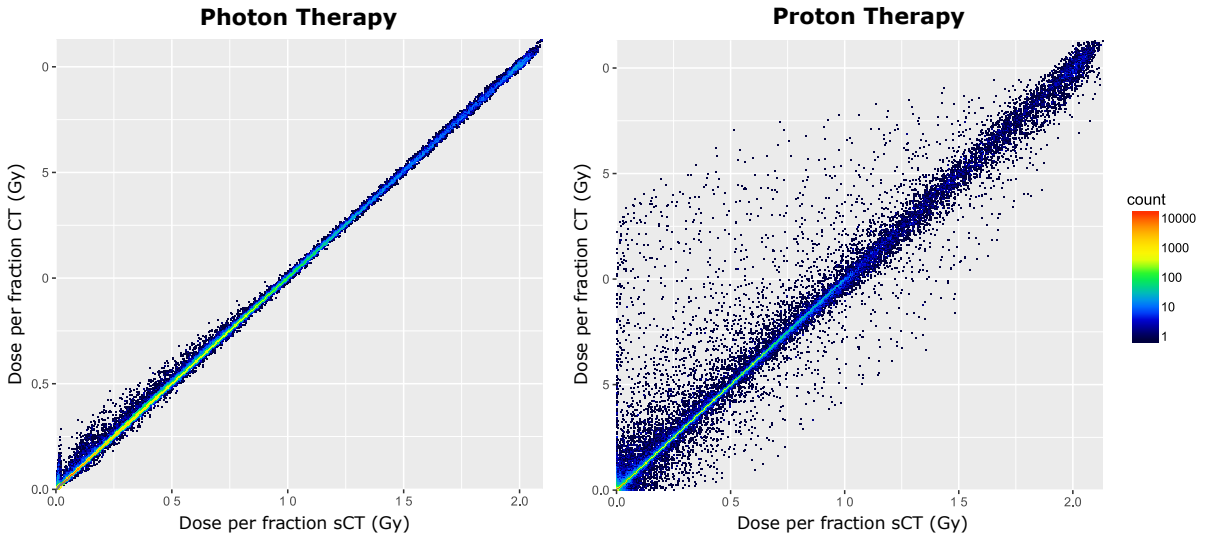

Figure 4.11: Density plots showing the relation between voxel-wise dose calculated on the sCT volume (horizontal axis) and on the CT volume (vertical axis) for a representative patient for the IMRT plan (left) and IMPT plan (right)

Figure 4.11 shows density plots that relate the dose calculated on the synthetic-CT and the dose calculated in the CT for each voxel of a patient for photon and proton plans. In the case of the 5-beam IMRT plan, it can be seen that there is a very high correlation between both dose distributions. The bigger discrepancies between values occur for the low-dose points. While maximum voxel-wise differences observed are around 0.25 Gy per fraction, these occur only in a very low number of voxels - less than 0.1% of voxels receiving dose show differences higher than 0.1 Gy per fraction. Therefore, these may not be clinically relevant, although an analysis of the location of the voxels is necessary due to the number of small critical structures in the head. In the case of the proton beams, however, there is a much larger dispersion of points away from the identity line, which occurs across the entire interval of dose values. Particular attention should be focused on the high differences that occur in points on the left side of the plot. Some voxels that have dose levels close to 0 Gy in the plan optimised using the sCT have values higher than 1 Gy per fraction when the same plan is applied to the CT. These differences can be attributed to changes in the proton beam range due to HU differences between CT and sCT, which resulted in changes in location of the distal dose falloff of the beams. Low-dose voxels in the sCT dose distribution, located in the distal zone of the beam, therefore receive high doses on the CT calculation if the range of the beam increases due to the difference in materials crossed. The opposite effect can also occur if ranges increase, and was seen across the patient dataset.

4.4 Comparison with Literature

To further analyse the results obtained in this study, these were compared with relevant synthetic-CT methods found in the literature. Methods were selected that focused on the brain region, and included dosimetric evaluation of the generated synthetic-CTs. Table 4.6 compares metric results for a variety of synthetic-CT generation methods.

It is important to note that although all these studies are relative to the head region, the selected FOV for the sCT generation differs, as well as the tumour location. Both the study by Dinkla using the CNN [45] and the study by Speier using the multimodal PATCH method [40] use a larger FOV in the

Table 4.6: Comparison of quality metrics for various methods of sCT generation for MRI-based RT planning and the current study. Standard deviations and range of values are reported when available.

Method	MAE_{Body} (HU)	Plan Type	Mean Dose Dif. Target (%)	ΔDose Isocenter (%)	Pass Rates, 2%/2mm
3D GAN	69 ± 10	IMRT	within 2 (-1.6 to 1.1)	0.5 ± 0.5	99.4 ± 0.4 (98.0 to 99.9)
		IMPT	within 7.5 (-7.5 to 3.8)	3.4 ± 4	97.7 ± 1.1 (94.0 to 99.5)
2D Dilated CNN [45]	67 ± 11	VMAT and IMRT	within 1.5 (-1.4 to 0.9)	-	99.1 ± 0.8 (96.2 to 99.9)
Multimodal PATCH [40]	73 ± 6	VMAT and IMRT	-	0.4 (VMAT), 0.3 (IMRT)	99.7
Patch-based [38]	85 ± 14	IMRT	within 1	-	-
Heterogeneous Bulk-Density [18]	34	IMPT	within 1.5 (-1.4 to 0.9)	-	99.5 (98.0 to 100)

axial direction than the one used in this study, with head scans that span until the chin region. Both these methods report MAE metrics similar to those obtained in this project, however these FOV differences need to be taken into account when comparing results. Cropping the scans at a slightly higher slice as was done in this study can lead to slightly lower HU errors as part of the airways are left out of the generation process. However, the value of 67 HU reported when using the CNN was calculated only for the intersecting voxels between the sCT and CT contours, and using this criteria in the present method results in lower MAE metrics of 61 HU. The opposite effect occurs for the study by Koivula using the heterogeneous bulk density method [18], which excludes any air cavities and considers only the top portion of the head consisting of brain and skull, resulting in a calculated MAE of only 34 HU across the dataset.

The results obtained for dose differences at the target when comparing sCT and CT for dose calculation in this study are comparable to the the ones reported by Dinkla using a CNN. They analysed 52 patients with different types of tumour, and reported worse results for patients with tumours located in the base of the brain and near small bone structures, including vestibular schwannomas. The worst case patients had a mean dose difference at the target of 1.4%, which is a similar result to the worst case dose difference obtained using the 3D GAN, of 1.6%, for targets located in similar regions. The multimodal PATCH study does not report mean dose differences at the target, but obtained slightly smaller absolute dose differences at the isocenter of the target, as well as superior results in gamma pass rate metrics; however the tumours analysed in this study were located in higher, more homogeneous parts of the brain. The patched-based method used by Andreasen [38] to generate sCT volumes was analysed dosimetrically using two simulated PTV contours, one in the cerebrum and one positioned behind the nasal cavity, and reported dose deviations at the target below 1% for both cases, slightly inferior to those reported in this current study. The main stepback of their study, however, was the use of only 5 patients for the analysis.

sCT generation for proton therapy MRI-planning has been assessed only for tumours located in the top portion of the cranium. Koivula reported values of mean dose differences at the target within 1.5% and pass rates superior to 98% when using sCTs when compared to CT for calculation, revealing greater accuracy than in the current study, justifiable from the difference in location of the tumours,

as well as lower HU uncertainty in sCT by excluding the portion of the head containing air-cavities from the generation process. The beam locations in that study were selected to cross only high-density homogeneous bone and homogeneous brain matter, for which bulk density methods perform reasonably well. In the current study, all patients had tumours positioned behind air cavities and near small structures of the ear. It is important to note that coplanar beams were selected that crossed regions where the sCT generation was known to perform worse, in order to estimate the upper limits of dose difference. When using a wedged pair of beams in clinical setting, this plan would normally be optimised to include a tilt to avoid the bulk of the petrous temporal bone.

The plans selected for this work were designed to be simple enough to allow for easy implementation for all patients using the matRad software. These were plans with a small number of co-planar fixed beam positions, not adjusted to each particular patient. VMAT plans were not defined because they are not currently supported in matRad and require high computational resources. With the exception of the patch-based method by Andersen, the remaining studies in the table resort to previously defined plans, and therefore are personalized to each individual patient and more clinically relevant. While they often further analyse other dose measures for each particular organ - such as maximum and minimum dose and DVH metrics - this would not be sensible in this project. For this reason, it was chosen to focus on mean differences and gamma analysis to assess the dose calculation accuracy in the synthetic-CTs in more general comparisons. In addition, no robustness optimisation was performed on the plans. Creating plans on the sCT that are optimised to be more robust to small deviations could result in smaller errors when applied to the CT containing differences in HU. As it is done in similar studies, the plans used here were optimised on the sCT before being applied to the CT for re-calculation, without further optimisation. Speier *et al.* has shown that performing the optimisation on the CT and then calculating the dose on the sCT has no significant impact in resulting mean dosimetric differences when compared to performing the optimisation on the sCT first [40].

As previously mentioned, one main setback of the methods in the present work was the difference in acquisition of CT and MRI, leading to registration errors between the two modalities that reflect on the generation of the synthetic-CTs to an extent that cannot be easily measured. Koivula *et al.* reported the use of a matching body-countour for sCT and CT based on the intersection of the voxels to remove uncertainties from body outline changes between MR and CT from the dose calculation [18]. In the present study, the generated sCT based on the MRI contour was chosen for dose analysis, to eliminate the need for any post-processing of the sCT before dose calculation. It is possible that the use of only intersecting body contours between sCT and CT for the dosimetric calculations could lead to slightly smaller deviations between dose distributions, mimicking the change in results obtained for the MAE metric. Another setback of the acquisition of the MRI not having been performed in an RT simulator is that the generated sCT volumes are not in treatment position. The plans defined for all the patients included beams in the same plane, however the position of the head is slightly tilted for some patients, meaning the direction of the beams in relation to the head is not always the same. A possibility to overcome this would be to apply to the sCT the inverse of the transform obtained from co-registering CT and MRI, to obtain a synthetic-CT in the original CT space. This, however, would induce an extra registration-related uncertainty. The use of MRI and CT images acquired both in treatment position would decrease registration errors and improve the accuracy of dose calculation comparisons using sCT volumes.

Previously reported criteria for the use of MRI-only RT planning were deviations of less than 2% on the PTV coverage when compared to CT-based dose calculations. Although using a simple plan,

the results obtained in this study for photon therapy fulfil this criteria for dose differences in the target (here, the GTV), when considering tumours located in theoretically difficult regions. Previously reported worst case deviations of 1.5% at the target, similar to the results here obtained, have not been deemed clinically relevant when compared to the total uncertainty of the RT treatment chain, which is within 5% for photon therapy [45]. The high metrics obtained when performing gamma analysis further support the accuracy of the dose calculations using synthetic-CTs generated by the 3D GAN approach. In the case of proton therapy, however, the synthetic-CTs do not show clinically acceptable results for all patients. Proton therapy plans already account for range uncertainties in the order of 3.5%, including uncertainty from CT imaging itself and scanner calibration, which can be around $\pm$0.5% for today's technology [125]. Although deviations under 2% were found for two thirds of the patients, which can be considered acceptable when considering these margins, mean dose differences at the target can be as high as 7.5% for the remaining. Further improvements to the generation method are therefore necessary for the synthetic-CT volumes to be accurate sources of dose calculation for proton therapy in some head and neck targets located near heterogeneous structures.

In addition to image quality and dosimetric accuracy, another relevant aspect when analysing CT generation methods is the time necessary for the generation of the images. This is of particular importance for the use of synthetic-CT images for adaptive re-planning, and in particular for online ART. Although using the 3D network slightly increased the generation time when compared to the 2D network, generation time for the method here presented was within 30 seconds. The training stage of the GAN is highly time consuming (hours to days), however training the network is only necessary once for each anatomical site and scanners. In addition, the time-consuming pre-processing steps such as image registration are only required at the training/validation stage. When generating new sCT scans from the MRI, only intensity normalization of the MRI is performed. The study using a multimodal PATCH algorithm reports generation times superior to 30 minutes, and therefore is only suitable for static applications. Although times are highly dependable of the computational resources used in each study, and therefore not directly comparable, the difference in time scales between ML methods and methods based on libraries or atlases is highly significant. The method developed by Speier also makes use of both T1 and T2-weighted sequences simultaneously. The method here presented uses only a single sequence, T1-weighted, further facilitating its use in adaptive re-planning without the need for acquiring additional sequences, which is time consuming. Although not tested, the nature of the GAN algorithm suggests is would be feasible to use other MRI sequences as input for training the network, in order to utilise the same sequence that is clinically relevant in MR-guidance for each site. The use of a pre-contrast T1-weighted MRI as opposed to a post-contrast MRI sequence should also be explored, at it could lead to fewer differences in intensity distributions between MRI scans. No pre-contrast scans were available for the patients used in this project.

Generative adversarial networks have previously been used for the MRI-to-CT synthesis task for brain sites, but most studies resort to 2D networks [47,49,50], and the original 2D *pix2pix* GAN implementation was previously used to generate sCT slices, but for the prostate site [48]. The only other study in literature that using GANs in a 3D setting for this task was by Nie *et al.* [46], using a FCN as the generative model. This method used patches of size 16$\times$16$\times$16 voxels, 4 times smaller than the patch size used in this project. A larger subvolume size allows for improved contextual information, translated in the difference of MAE metrics (93 vs 69 HU), although differences in data need to be considered. The study by Nie did not perform dosimetric evaluation. This project was the first study to evaluate the use of synthetic-CTs created by GANs for dose calculation using both photon and proton plans. In the current project, the

3D method using subvolumes improved the results when compared to the 2D network. With access to a larger training dataset, and more powerful GPUs, the method could be implemented using the entire image volumes.

5 Conclusions

This project demonstrated the use of a modified 3-dimensional generative adversarial network to create synthetic-CT volumes from a single conventional MRI sequence, and showed the potential of this method for MRI-based radiotherapy planning.

The results presented in the previous section showed the 3D method outperformed a 2D architecture. This type of method could be transferable to clinical practice with the use of full volumes in the training process, as opposed to using patches of the image only, as more training data would be available. Even when using volumetric data, this type of deep-learning method is capable of generating sCT images in less than 30 seconds, making it possible to use in clinical applications were fast generation of images is necessary, in particular for MRI-guided online ART workflows. The fast generation of sCT images from a single conventional T1w MRI sequence and easy implementation within a clinical workflow are the main advantages of the method here presented, in comparison with previously published methods of sCT generation such as atlas-based or bulk-density approaches.

This was the first study to evaluate the use of synthetic-CTs created by GANs for dose calculation using both photon and proton plans. In addition, it was chosen to analyse tumours located at the base of the skull, as well as beam configurations crossing small structures and airways, where the sCT was shown to perform worse. For photon treatment plans, dose differences at the target within 2% demonstrate MRI-based dose calculations can be deemed clinically acceptable for these type of tumours. Although sCT images have previously been shown to provide accurate dose calculations for proton beams crossing the top of the skull, improvements in sCT quality are necessary for accurate dose calculations when the beams cross air cavities and small bone structures.

Special attention should be given to the effect of registration between MRI and real CT on the generated synthetic-CT images. When one of the main reasonings for an MRI-only planning workflow is the decrease of the registration errors arriving from the multimodal workflow, it is important to eliminate the propagation of this type of errors when generating synthetic-CTs. With the growing use of MRI simulators in radiotherapy, which allow for the acquisition of MR scans in treatment position and with reduced geometric distortions, the future of sCT generation studies will surely make use of this type of data for network training. Larger, more diverse datasets are required to support the validity of the method for clinical settings, and should also provide improvements when training the network. The method should also be tested for other anatomical sites, such as the abdominal and pelvic areas.

In order to evaluate the use of synthetic-CT volumes for adaptive re-planning, further analysis should be performed using image data acquired at multiple steps of the treatment, in particular for patients where strong deformations occur with the course of treatment, such as head and neck cancer patients. Synthetic-CTs generated from the MRI should be evaluated and compared to the use of CT guidance and the use of MR guidance with deformable registration using a pre-treatment CT. In addition to dose calculation accuracy using sCT images, the advantages of removing the CT from the RT workflow should be weighted. Such workflows would be particular advantageous for paediatric patients, for whom the

acquisition of multiple CT scans throughout the course of treatment presents higher associated risks. Paediatric data was unfortunately not available for this project, however this would be an area of great interest.

Another aspect that requires further exploration in the use of machine learning for the creation of synthetic CTs is that present solutions are largely black box methods. Patient specific QA is therefore necessary to guarantee the quality of the generated images when no ground-truth CT images are available. With the growth of the search for interpretable machine learning methods for clinical applications, this is an area that could be explored for the development of such QA methods.

The focus of this project was the application of MRI-based synthetic-CT generation for MRI-based radiation therapy planning. However, this task is also very relevant for the implementation of hybrid MRI/PET imaging systems. Although deeper analysis would be necessary to study the quality of the generated images for attenuation correction, the results in MAE obtained in this project can lead to the conclusion that this method would be suitable for this application. Additionally, the proposed methodology using a 3-dimensional approach can also be applied to other medical imaging synthesis tasks, such as the synthesis of one MRI sequence from another, or the synthesis of full-quality kVCT images from MVCT or CBCT scans. This last example could also be of value for conventional radiotherapy workflows, in particular for adaptive re-planning using CT guidance.

Bibliography

[1] World Health Organization, *World health statistics 2018: monitoring health for the SDGs, sustainable development goals*. 2018.

[2] J. M. Borras *et al.*, "How many new cancer patients in Europe will require radiotherapy by 2025? an ESTRO-HERO analysis," *Radiotherapy and Oncology*, vol. 119, no. 1, pp. 5 – 11, 2016.

[3] P. Mayles, A. Nahum, and J. C. Rosenwald, *Handbook of Radiotherapy Physics: Theory and Practice*. Boca Raton, Florida: Taylor & Francis, 1 ed., 2007.

[4] P. Hoskin and V. Goh, *Radiotherapy in Practice: Imaging*. New York: Oxford University Press, 1 ed., 2010.

[5] S. Devic, "MRI simulation for radiotherapy treatment planning," *Medical Physics*, vol. 39, no. 11, pp. 6701–6711, 2012.

[6] A. M. Owrangi, P. B. Greer, and C. K. Glide-Hurst, "MRI-only treatment planning: benefits and challenges," *Physics in Medicine & Biology*, vol. 63, no. 5, p. 05TR01, 2018.

[7] S. Lim-Reinders, B. M. Keller, and A. Kim, "Online adaptive radiation therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 99, no. 4, pp. 994 – 1003, 2017.

[8] A. Hunt *et al.*, "Adaptive radiotherapy enabled by MRI guidance," *Clinical Oncology*, vol. 30, 2018.

[9] C. Westbrook, C. K. Roth, and J. Talbot, *MRI in Practice*. Chichester, West Sussex: Wiley-Blackwell, 4th ed., 2011.

[10] L. R. Schad *et al.*, "Radiosurgical treatment planning of brain metastases based on a fast, three-dimensional MR imaging technique," *Magnetic Resonance Imaging*, vol. 12, no. 5, pp. 811 – 819, 1994.

[11] C. R. Ramsey and A. L. Oliver, "Magnetic resonance imaging based digitally reconstructed radiographs, virtual simulation, and three-dimensional treatment planning for brain neoplasms," *Medical Physics*, vol. 25, no. 10, pp. 1928–1934, 1998.

[12] R. Prabhakar *et al.*, "Feasibility of using MRI alone for 3D radiation treatment planning in brain tumors," *Japanese Journal of Clinical Oncology*, vol. 37, no. 6, pp. 405–411, 2007.

[13] D. Pasquier *et al.*, "MRI alone simulation for conformal radiation therapy of prostate cancer: technical aspects," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, pp. 160–163, 2006.

[14] D. Weber *et al.*, "Open low-field magnetic resonance imaging for target definition, dose calculations and set-up verification during three-dimensional crt for glioblastoma multiforme," *Clinical Oncology*, vol. 20, no. 2, pp. 157 – 167, 2008.

[15] J. H. Jonsson, M. G. Karlsson, and T. Nyholm, "Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions," *Radiation Oncology*, vol. 5, no. 1, p. 62, 2010.

[16] B. H. Kristensen *et al.*, "Dosimetric and geometric evaluation of an open low-field magnetic resonance simulator for radiotherapy treatment planning of brain tumours," *Radiotherapy and Oncology*, vol. 87, no. 1, pp. 100 – 109, 2008.

[17] L. Chen *et al.*, "Magnetic resonance–based treatment planning for prostate intensity-modulated radiotherapy: Creation of digitally reconstructed radiographs," *International Journal of Radiation Oncology*Biology*Physics*, vol. 68, no. 3, pp. 903 – 911, 2007.

[18] L. Koivula, L. Wee, and J. Korhonen, "Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: Dose calculation accuracy in substitute CT images," *Medical Physics*, vol. 43, no. 8Part1, pp. 4634–4642, 2016.

[19] M. Kapanen and M. Tenhunen, "T1/T2*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning," *Acta Oncologica*, vol. 52, no. 3, pp. 612–618, 2013.

[20] J. Korhonen *et al.*, "A dual model HU conversion from MRI intensity values within and outside of bone segment for MRI-based radiotherapy treatment planning of prostate cancer," *Medical physics*, vol. 41, no. 1, 2014.

[21] J. Kim *et al.*, "Implementation of a novel algorithm for generating synthetic CT images from magnetic resonance imaging data sets for prostate cancer radiation therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 91, no. 1, pp. 39 – 47, 2015.

[22] A. Johansson, M. Karlsson, and T. Nyholm, "CT substitute derived from MRI sequences with ultrashort echo time," *Medical Physics*, vol. 38, no. 5, pp. 2708–2714, 2011.

[23] A. Johansson *et al.*, "Voxel-wise uncertainty in CT substitute derived from MRI," *Medical Physics*, vol. 39, no. 6Part1, pp. 3283–3290, 2012.

[24] S. Roy *et al.*, "PET attenuation correction using synthetic CT from ultrashort echo-time MR imaging.," *Journal of nuclear medicine : official publication, Society of Nuclear Medicine*, vol. 55, no. 12, pp. 2071–7, 2014.

[25] A. Johansson, A. Garpebring, and T. Nyholm, "CT substitutes derived from MR images reconstructed with parallel imaging," *Medical Physics*, vol. 41, no. 8, p. 302, 2014.

[26] J. M. Edmund *et al.*, "A voxel-based investigation for MRI-only radiotherapy of the brain using ultra short echo times," *Physics in Medicine & Biology*, vol. 59, no. 23, p. 7501, 2014.

[27] J. H. Jonsson *et al.*, "Accuracy of inverse treatment planning on substitute CT images derived from MR data for brain lesions," *Radiation Oncology*, vol. 10, no. 1, p. 13, 2015.

[28] A. Johansson *et al.*, "Improved quality of computed tomography substitute derived from magnetic resonance (MR) data by incorporation of spatial information – potential application for MR-only radiotherapy and attenuation correction in positron emission tomography," *Acta Oncologica*, vol. 52, no. 7, pp. 1369–1373, 2013.

[29] J. M. Edmund and T. Nyholm, "A review of substitute CT generation for MRI-only radiation therapy," *Radiation Oncology*, vol. 12, no. 1, p. 28, 2017.

[30] J. A. Dowling *et al.*, "An atlas-based electron density mapping method for magnetic resonance imaging (MRI)-alone treatment planning and adaptive MRI-based prostate radiation therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 83, no. 1, pp. e5 – e11, 2012.

[31] J. Uh *et al.*, "MRI-based treatment planning with pseudo CT generated through atlas registration," *Medical Physics*, vol. 41, no. 5, p. 051711, 2014.

[32] J. A. Dowling *et al.*, "Automatic substitute computed tomography generation and contouring for magnetic resonance imaging (MRI)-alone external beam radiation therapy from standard MRI sequences," *International Journal of Radiation Oncology*Biology*Physics*, vol. 93, no. 5, pp. 1144 – 1153, 2015.

[33] J. Sjölund, D. Forsberg, and H. Knutsson, "Generating patient specific pseudo-CT of the head from MR using atlas-based regression," *Physics in Medicine & Biology*, vol. 60, no. 2, p. 825, 2015.

[34] N. Burgos *et al.*, "Robust CT synthesis for radiotherapy planning: Application to the head and neck region," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pp. 476–484, Springer International Publishing, 2015.

[35] D. Andreasen *et al.*, "Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain," *Medical Physics*, vol. 42, no. 4, pp. 1596–1605, 2015.

[36] R. Farjam *et al.*, "Multiatlas approach with local registration goodness weighting for MRI-based electron density mapping of head and neck anatomy," *Medical Physics*, vol. 44, no. 7, pp. 3706–3717, 2017.

[37] C. Siversson *et al.*, "Technical note: MRI only prostate radiotherapy planning using the statistical decomposition algorithm," *Medical Physics*, vol. 42, no. 10, pp. 6090–6097, 2015.

[38] D. Andreasen *et al.*, "Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain," *Medical Physics*, vol. 42, no. 4, pp. 1596–1605, 2015.

[39] S. Aouadi *et al.*, "Generation of synthetic CT using multi-scale and dual-contrast patches for brain MRI-only external beam radiotherapy," *Physica Medica*, vol. 42, pp. 174 – 184, 2017.

[40] C. Speier *et al.*, "Advanced multimodal methods for cranial pseudo-CT generation validated by imrt and vmat radiation therapy plans," *International Journal of Radiation Oncology*Biology*Physics*, vol. 102, no. 4, pp. 792 – 800, 2018. Imaging in Radiation Oncology.

[41] T. Huynh *et al.*, "Estimating CT Image From MRI Data Using Structured Random Forest and Auto-Context Model.," *IEEE transactions on medical imaging*, vol. 35, no. 1, pp. 174–83, 2016.

[42] D. Nie *et al.*, "Estimating CT image from MRI data using 3D fully convolutional networks," in *Deep Learning and Data Labeling for Medical Applications* (G. Carneiro, D. Mateus, L. Peter, A. Bradley, J. M. R. S. Tavares, V. Belagiannis, J. P. Papa, J. C. Nascimento, M. Loog, Z. Lu, J. S. Cardoso, and J. Cornebise, eds.), pp. 170–178, Springer International Publishing, 2016.

[43] X. Han, "MR-based synthetic CT generation using a deep convolutional neural network method," *Medical Physics*, vol. 44, no. 4, pp. 1408–1419, 2017.

[44] L. Xiang *et al.*, "Deep embedding convolutional neural network for synthesizing CT image from T1-Weighted MR image," *Medical Image Analysis*, vol. 47, pp. 31–44, 2018.

[45] A. M. Dinkla *et al.*, "MR-Only Brain Radiation Therapy: Dosimetric Evaluation of Synthetic CTs Generated by a Dilated Convolutional Neural Network," *International Journal of Radiation Oncology*Biology*Physics*, vol. 102, no. 4, pp. 801–812, 2018.

[46] D. Nie *et al.*, "Medical image synthesis with context-aware generative adversarial networks," in *Medical Image Computing and Computer-Assisted Intervention MICCAI 2017*, pp. 417–425, Springer International Publishing, 2017.

[47] H. Emami *et al.*, "Generating synthetic CTs from magnetic resonance images using generative adversarial networks," *Medical Physics*, vol. 45, no. 8, pp. 3627–3636, 2018.

[48] M. Maspero *et al.*, "Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy," *ArXiv e-prints*, 2018.

[49] J. M. o. Wolterink, "Deep MR to CT Synthesis using Unpaired Data," *ArXiv e-prints*, 2018.

[50] H. Yang *et al.*, "Unpaired brain MR-to-CT synthesis using a structure-constrained cyclegan," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, pp. 174–182, Springer International Publishing, 2018.

[51] J. R. Bushberg, J. A. Seibert, and J. M. Boone, *The Essential Physics of Medical Imaging*. Philadelphia: Lippincott Williams & Wilkins, 2 ed., 2002.

[52] J. Hsieh, *Computed Tomography: Principles, Design, Artifacts and Recent Advances*. Bellingham, Washington: SPIE, 2 ed., 2002.

[53] J. Seco and P. M. Evans, "Assessing the effect of electron density in photon dose calculations," *Medical Physics*, vol. 33, no. 2, pp. 540–552, 2006.

[54] W. Skrzyński *et al.*, "Computed tomography as a source of electron density information for radiation treatment planning," *Strahlentherapie und Onkologie: Organ der Deutschen Röntgengesellschaft*, vol. 186, pp. 327–33, 2010.

[55] C. B. Saw *et al.*, "Determination of CT-to-density conversion relationship for image-based treatment planning systems," *Medical Dosimetry*, vol. 30, no. 3, pp. 145 – 148, 2005.

[56] S. Thomas, "Relative electron density calibration of CT scanners for radiotherapy treatment planning," *The British journal of radiology*, vol. 72, pp. 781–6, 1999.

[57] U. Schneider, E. Pedroni, and A. Lomax, "The calibration of CT hounsfield units for radiotherapy treatment planning," *Physics in Medicine and Biology*, vol. 41, no. 1, pp. 111–124, 1996.

[58] D. Weishaupt and V. D. K. B. Marincek, *How does MRI Work? An Introduction to the Physics and Function of Magnetic Resonance Imaging*. Berlin: Springer, 2 ed., 2008.

[59] R. Bitar *et al.*, "MR pulse sequences: What every radiologist wants to know but is afraid to ask," *RadioGraphics*, vol. 26, no. 2, pp. 513–537, 2006.

[60] S. G. Wetzel *et al.*, "Three-dimensional, T1-weighted gradient-echo imaging of the brain with a volumetric interpolated examination," *American Journal of Neuroradiology*, vol. 23, no. 6, pp. 995–1002, 2002.

[61] E. L. Chang *et al.*, "Evaluation of peritumoral edema in the delineation of radiotherapy clinical target volumes for glioblastoma," *International Journal of Radiation Oncology*Biology*Physics*, vol. 68, no. 1, pp. 144 – 150, 2007.

[62] N. G. Zaorsky *et al.*, "Acr appropriateness criteria® external beam radiation therapy treatment planning for clinically localized prostate cancer, part i of ii," *Advances in Radiation Oncology*, vol. 2, no. 1, pp. 62 – 84, 2017.

[63] R. Pötter *et al.*, "Recommendations from gynaecological (gyn) gec estro working group (ii): Concepts and terms in 3D image-based treatment planning in cervix cancer brachytherapy—3D dose volume parameters and aspects of 3D image-based anatomy, radiation physics, radiobiology," *Radiotherapy and Oncology*, vol. 78, no. 1, pp. 67 – 77, 2006.

[64] M. J. Ghilezan *et al.*, "Prostate gland motion assessed with cine-magnetic resonance imaging (cine-MRI)," *International Journal of Radiation Oncology*Biology*Physics*, vol. 62, no. 2, pp. 406 – 417, 2005.

[65] Intensity Modulated Radiation Therapy Collaborative Working Group, "Intensity-modulated radiotherapy: current status and issues of interest," *International Journal of Radiation Oncology*Biology*Physics*, vol. 51, no. 4, pp. 880 – 914, 2001.

[66] A. Thornton *et al.*, "A head immobilization system for radiation simulation, CT, MRI, and pet imaging," *Medical Dosimetry*, vol. 16, no. 2, pp. 51 – 56, 1991.

[67] K. K. Brock *et al.*, "Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the aapm radiation therapy committee task group no. 132," *Medical Physics*, vol. 44, no. 7, pp. e43–e76, 2017.

[68] K. S. Oh Seungjong, "Deformable image registration in radiation therapy," *Radiat Oncol J*, vol. 35, no. 2, pp. 101–111, 2017.

[69] D. Jones, "ICRU report 50 — prescribing, recording and reporting photon beam therapy," *Medical Physics*, vol. 21, no. 6, pp. 833–834, 1994.

[70] T. Landberg *et al.*, "ICRU Report 62," *Journal of the International Commission on Radiation Units and Measurements*, vol. os32, no. 1, 2016.

[71] A.-L. Grosu, L. D. Sprague, and M. Molls, *Definition of Target Volume and Organs at Risk. Biological Target Volume*, pp. 167–177. Berlin, Heidelberg: Springer Berlin Heidelberg, 2006.

[72] T. Pavel, *Feasibility of magnetic resonance imaging-based radiation therapy for brain tumour treatment*. PhD thesis, 2017.

[73] G. A. Cefaro, D. Genovesi, and C. A. Perez, *Radiation Dose Constraints for Organs at Risk: Modeling and Importance of Organ Delineation in Radiation Therapy*, pp. 49–73. Milano: Springer Milan, 2013.

[74] G. T. Y. Chen, G. C. Sharp, and S. Mori, "A review of image-guided radiotherapy," *Radiological Physics and Technology*, vol. 2, no. 1, pp. 1–12, 2009.

[75] T. Gupta and C. Narayan, "Image-guided radiation therapy: Physician's perspectives," *Journal of Medical Physics*, vol. 37, no. 4, pp. 174–182, 2012.

[76] J.-P. Bissonnette *et al.*, "Quality assurance for image-guided radiation therapy utilizing CT-based technologies: A report of the aapm tg-179," *Medical physics*, vol. 39, pp. 1946–63, 2012.

[77] J. M. Pollard *et al.*, "The future of image-guided radiotherapy will be MR guided," *The British Journal of Radiology*, vol. 90, no. 1073, p. 667, 2017. PMID: 28256898.

[78] M. B. van Herk *et al.*, "Magnetic resonance imaging-guided radiation therapy: A short strengths, weaknesses, opportunities, and threats analysis.," *International journal of radiation oncology, biology, physics*, vol. 101 5, pp. 1057–1060, 2018.

[79] D. A. Jaffray *et al.*, "A facility for magnetic resonance–guided radiation therapy," *Seminars in Radiation Oncology*, vol. 24, no. 3, pp. 193 – 195, 2014. Magnetic Resonance Imaging in Radiation Oncology.

[80] T. Bostel *et al.*, "Prospective feasibility analysis of a novel off-line approach for MR-guided radiotherapy," *Strahlentherapie und Onkologie*, vol. 194, no. 5, pp. 425–434, 2018.

[81] J. Stewart *et al.*, "Automated weekly online replanning for imrt of cervix cancer," *International Journal of Radiation Oncology*Biology*Physics*, vol. 72, no. 1, Supplement, p. S18, 2008. Proceedings of the American Society for Therapeutic Radiology and Oncology 50th Annual Meeting.

[82] S. Mutic *et al.*, "Tu-h-bra-08: The design and characteristics of a novel compact linac-based MRI guided radiation therapy (MR-igrt) system," *Medical Physics*, vol. 43, no. 66, pp. 3770–3770, 2016.

[83] B. W. Raaymakers *et al.*, "First patients treated with a 1.5 t MRI-linac: clinical proof of concept of a high-precision, high-field MRI guided radiotherapy treatment," *Physics in Medicine & Biology*, vol. 62, no. 23, pp. L41–L50, 2017.

[84] C. Kontaxis, G. H. Bol, and B. W. Raaymakers, "Towards adaptive IMRT sequencing for the MR-linac," *Physics in Medicine and Biology*, vol. 60, no. 6, pp. 2493–2509, 2015.

[85] C. Kontaxis *et al.*, "Towards fast online intrafraction replanning for free-breathing stereotactic body radiation therapy with the MR-linac," *Physics in Medicine & Biology*, vol. 62, no. 18, pp. 7233–7248, 2017.

[86] A. M. Werensteijn-Honingh *et al.*, "Feasibility of stereotactic radiotherapy using a 1.5t MR-linac: Multi-fraction treatment of pelvic lymph node oligometastases," *Radiotherapy and Oncology*, vol. 134, pp. 50 – 54, 2019.

[87] E. Johnstone *et al.*, "Systematic Review of Synthetic Computed Tomography Generation Methodologies for Use in Magnetic Resonance Imaging Only Radiation Therapy," *International Journal of Radiation Oncology*Biology*Physics*, vol. 100, no. 1, pp. 199–217, 2018.

[88] M. D. Robson, M. Bydder, and G. M. Bydder, "Magnetic resonance: an introduction to ultrashort TE (UTE) imaging.," *Journal of computer assisted tomography*, vol. 27, no. 6, pp. 825–46, 2013.

[89] E. M. Johnson *et al.*, "Improved cortical bone specificity in UTE MR imaging," *Magnetic Resonance in Medicine*, vol. 77, no. 2, pp. 684–695, 2017.

[90] A. Ribeiro, *Bone recognition in UTE MR images by artificial neural networks for attenuation correction of brain imaging in MR/PET scanners*. Dissertation, University of Lisbon, 2012.

[91] J. Ma, "Dixon techniques for water and fat imaging," *Journal of Magnetic Resonance Imaging*, vol. 28, no. 3, pp. 543–558, 2008.

[92] N. Tyagi *et al.*, "Dosimetric and workflow evaluation of first commercial synthetic CT software for clinical use in pelvis," *Physics in Medicine & Biology*, vol. 62, no. 8, p. 2961, 2017.

[93] P. Margosian, T. Takahashi, and M. Takizawa, "Practical implementation of UTE imaging," *Encycl. Magn. Reson.*, vol. 1, 2012.

[94] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, 2017.

[95] Y. Jia *et al.*, "Caffe: Convolutional architecture for fast feature embedding," in *Proceedings of the 22Nd ACM International Conference on Multimedia*, MM '14, (New York, NY, USA), pp. 675–678, ACM, 2014.

[96] M. Abadi *et al.*, "TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems," *ArXiv e-prints*, 2016.

[97] R. Collobert, K. Kavukcuoglu, and C. Farabet, "Torch7: A matlab-like environment for machine learning," in *BigLearn, NIPS Workshop*, 2011.

[98] A. Jog, A. Carass, and J. L. Prince, "Improving magnetic resonance resolution with supervised learning," in *2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, vol. 2014, pp. 987–990, IEEE, 2014.

[99] A. Jog *et al.*, "Random Forest Flair Reconstruction from T1,T2 and PD-weighted MRI," *Proceedings. IEEE International Symposium on Biomedical Imaging*, vol. 2014, pp. 1079–1082, 2014.

[100] I. J. Goodfellow *et al.*, "Generative Adversarial Networks," *ArXiv e-prints*, 2014.

[101] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *CoRR*, vol. abs/1411.1784, 2014.

[102] P. Isola *et al.*, "Image-to-image translation with conditional adversarial networks," in *Computer Vision and Pattern Recognition (CVPR), 2017 IEEE Conference on*, 2017.

[103] D. Pathak *et al.*, "Context encoders: Feature learning by inpainting," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[104] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," *CoRR*, 2015.

[105] D. Yoo *et al.*, "Pixel-level domain transfer," *Lecture Notes in Computer Science*, pp. 517—532, 2016.

[106] J. Wu *et al.*, "Learning a probabilistic latent space of object shapes via 3D generative-adversarial modeling," *CoRR*, 2016.

[107] R. Hermoza and I. Sipiran, "3D reconstruction of incomplete archaeological objects using a generative adversarial network," in *Proceedings of Computer Graphics International 2018*, CGI 2018, (New York, NY, USA), pp. 5–11, ACM, 2018.

[108] B. Yu *et al.*, "3D cgan based cross-modality MR image synthesis for brain tumor segmentation," in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 626–630, 2018.

[109] Q. Yang *et al.*, "MRI image-to-image translation for cross-modality image registration and segmentation," *CoRR*, 2018.

[110] C. Han *et al.*, "GAN-based synthetic brain MR image generation," in *2018 IEEE 15th International Symposium on Biomedical Imaging*, pp. 734–738, 2018.

[111] P. Welander, S. Karlsson, and A. Eklund, "Generative Adversarial Networks for Image-to-Image Translation on Multi- Contrast MR Images - A Comparison of CycleGAN and UNIT," *ArXiv e-prints*, 2018.

[112] X. Gu, H. Knutsson, and A. Eklund, "Generating Diffusion MRI scalar maps from T1 weighted images using generative adversarial networks," *ArXiv e-prints*, 2018.

[113] Ben-Cohen *et al.*, "Virtual PET images from CT data using deep convolutional networks: Initial results," in *Simulation and Synthesis in Medical Imaging* (S. A. Tsaftaris, A. Gooya, A. F. Frangi, and J. L. Prince, eds.), pp. 49–57, Springer International Publishing, 2017.

[114] A. Ben-Cohen *et al.*, "Cross-Modality Synthesis from CT to PET using FCN and GAN Networks for Improved Automated Lesion Detection," *ArXiv e-prints*, 2018.

[115] Y. Pan *et al.*, "Synthesizing missing PET from MRI with cycle-consistent generative adversarial networks for alzheimer's disease diagnosis," in *Medical Image Computing and Computer Assisted Intervention MICCAI 2018* (A. F. Frangi, J. A. Schnabel, C. Davatzikos, C. Alberola-López, and G. Fichtinger, eds.), pp. 455–463, Springer International Publishing, 2018.

[116] P. Costa *et al.*, "End-to-end adversarial retinal image synthesis," *IEEE Transactions on Medical Imaging*, vol. 37, no. 3, pp. 781–791, 2018.

[117] H. Zhao, H. Li, and L. Cheng, "Synthesizing retinal and neuronal images with generative adversarial nets," *Medical Image Analysis*, vol. 49, pp. 14 – 26, 2018.

[118] J. M. Wolterink and other, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Transactions on Medical Imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[119] Y. Wang *et al.*, "3D conditional generative adversarial networks for high-quality PET image estimation at low dose," *NeuroImage*, vol. 174, pp. 550 – 562, 2018.

[120] D. Korkinof *et al.*, "High-Resolution Mammogram Synthesis using Progressive Generative Adversarial Networks," *ArXiv e-prints*, 2018.

[121] S. Scoccianti *et al.*, "Organs at risk in the brain and their dose-constraints in adults and in children: A radiation oncologist's guide for delineation in everyday practice," *Radiotherapy and Oncology*, vol. 114, no. 2, pp. 230 – 238, 2015.

[122] D. A. Low, W. B. Harms, and J. A. Purdy, "A technique for the quantitative evaluation of dose distributions," *Medical Physics*, vol. 25, no. 5, pp. 656–661, 1998.

[123] M. Hussein, C. Clark, and A. Nisbet, "Challenges in calculation of the gamma index in radiotherapy – towards good practice," *Physica Medica*, vol. 36, pp. 1–11, 2017.

[124] G. A. Ezzell *et al.*, "IMRT commissioning: Multiple institution planning and dosimetry comparisons, a report from AAPM task group 119," *Medical Physics*, vol. 36, no. 11, pp. 5359–5373, 2009.

[125] H. Paganetti, "Range uncertainties in proton therapy and the role of monte carlo simulations," *Physics in Medicine and Biology*, vol. 57, no. 11, pp. 99–117, 2012.

Appendix

A. Suplementary Data

Table A1: Image quality metrics obtained for the sCT volumes generated with the modified 3D GAN for all patients of Dataset A using the 7-fold cross-validation method. ME and MAE values are given in HU.

Patient	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**
ME$_{Body}$	1	15	1	-10	5	-7	-12	1	-4	17	8	-1	-2	0
MAE$_{Body}$	70	84	76	65	74	94	57	44	74	62	73	59	54	74
MAE$_{Air}$	314	317	263	233	287	376	264	259	272	277	291	285	237	294
MAE$_{Bone}$	147	178	147	144	146	209	104	96	150	140	141	125	99	155
MAE$_{ST}$	38	45	43	35	42	52	38	24	39	34	42	35	31	42
DSC$_{Bone}$	0.89	0.86	0.88	0.90	0.88	0.83	0.92	0.94	0.89	0.91	0.88	0.90	0.93	0.88
MSSIM	0.95	0.95	0.95	0.96	0.95	0.93	0.96	0.98	0.95	0.96	0.95	0.96	0.97	0.95

Patient	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**
ME$_{Body}$	28	2	4	-14	-7	-5	-10	-5	4	0	-12	-20	-17	-4
MAE$_{Body}$	72	62	95	65	68	54	65	58	87	54	48	58	78	70
MAE$_{Air}$	278	236	338	239	237	197	245	214	350	204	223	205	245	267
MAE$_{Bone}$	176	130	199	151	144	112	130	108	203	114	120	112	176	155
MAE$_{ST}$	35	35	52	36	37	31	37	32	52	30	24	34	35	38
DSC$_{Bone}$	0.90	0.90	0.84	0.90	0.90	0.93	0.91	0.92	0.77	0.92	0.93	0.93	0.89	0.88
MSSIM	0.95	0.97	0.93	0.95	0.96	0.98	0.96	0.97	0.91	0.97	0.98	0.97	0.96	0.96

Patient	**29**	**30**	**31**	**32**	**33**	**34**	**35**	**36**	**37**	**38**	**39**	**40**	**41**	**42**
ME$_{Body}$	0	-14	6	19	13	30	-9	4	6	-28	-16	-2	-5	-6
MAE$_{Body}$	63	73	78	64	63	80	54	64	68	98	74	73	81	58
MAE$_{Air}$	241	297	300	304	274	311	315	263	262	323	309	283	259	257
MAE$_{Bone}$	137	153	125	116	109	162	117	127	168	217	167	180	194	114
MAE$_{ST}$	34	41	42	39	41	46	31	40	39	47	40	37	44	33
DSC$_{Bone}$	0.89	0.87	0.91	0.90	0.91	0.86	0.93	0.90	0.87	0.87	0.90	0.88	0.86	0.92
MSSIM	0.97	0.95	0.96	0.97	0.95	0.95	0.97	0.96	0.95	0.93	0.94	0.94	0.95	0.97

Table A2: Image quality metrics obtained for the sCT volumes generated with the modified 3D GAN for all patients of Dataset B (holdout dataset). ME and MAE values are given in HU.

Patient	**43**	**44**	**45**	**46**	**47**	**48**	**49**	**50**	**51**	**52**	**53**	**54**
ME$_{Body}$	0	-5	-30	-20	22	-10	-17	-16	-18	20	3	-1
MAE$_{Body}$	85	71	77	78	89	53	73	69	80	74	85	63
MAE$_{Air}$	131	239	211	251	255	270	321	290	244	325	403	252
MAE$_{Bone}$	225	140	131	161	162	104	127	145	156	149	143	107
MAE$_{ST}$	45	41	49	46	50	36	46	41	52	39	42	36
DSC$_{Bone}$	0.83	0.88	0.91	0.88	0.87	0.92	0.90	0.90	0.88	0.85	0.90	0.91
MSSIM	0.94	0.97	0.97	0.96	0.96	0.97	0.97	0.98	0.95	0.96	0.94	0.97

Table A3: Comparison of MAE and dose evaluation metrics (dose differences and pass rates) for the 33 patients of Dataset A containing contoured tumour and OAR data.

Patient	**1**	**2**	**3**	**4**	**5**	**7**	**9**	**10**	**11**	**12**	**13**
MAE$_{Body}$	70	84	76	65	74	57	74	62	73	59	54
Photon Therapy:											
Mean ΔDose GTV (%)	0.2	0.1	-0.3	0.2	0.5	-0.8	-0.7	0.6	-0.1	-0.3	0.3
Pass Rate 3%/3mm (%)	99.9	99.9	99.9	99.9	99.9	99.8	99.9	99.9	100	99.9	99.9
Pass Rate 2%/2mm (%)	99.9	99.9	99.7	99.8	99.7	99.2	99.9	99.7	99.4	99.9	99.8
Proton Therapy:											
Mean ΔDose GTV (%)	1.5	3.8	-1.3	1.4	-2.8	-6.2	-4.3	-0.1	-0.3	-0.9	1.7
Pass Rate 3%/3mm (%)	99.5	98.8	99.9	99.8	99.3	98.6	98.9	99.4	99.5	99.7	98.8
Pass Rate 2%/2mm (%)	99.1	97.9	99.6	99.2	98.1	97.5	97.6	98.7	98.9	99.3	98.9

Patient	**14**	**15**	**16**	**17**	**18**	**19**	**20**	**21**	**22**	**23**	**24**
MAE$_{Body}$	74	72	62	95	65	68	54	66	58	87	54
Photon Therapy:											
Mean ΔDose GTV (%)	-0.2	0.8	0.4	0.9	-0.6	-0.3	-0.4	-1.6	-0.5	0.6	0.2
Pass Rate 3%/3mm (%)	99.9	99.6	99.9	99.5	99.7	99.9	100	99.7	99.9	99.7	99.9
Pass Rate 2%/2mm (%)	99.8	98.6	99.3	98.0	99.0	99.3	99.8	98.5	99.6	99.1	99.9
Proton Therapy:											
Mean ΔDose GTV (%)	1.2	-0.9	-1.7	0.6	-1.5	-0.4	1.6	-5.5	-2.9	-0.1	0.9
Pass Rate 3%/3mm (%)	98.8	99.4	98.4	97.8	98.8	98.4	99.6	99.0	99.1	99.4	99.3
Pass Rate 2%/2mm (%)	97.5	98.6	98.6	97.3	95.4	97.6	98.3	97.9	97.7	98.0	98.8

Patient	**26**	**27**	**28**	**29**	**30**	**34**	**36**	**37**	**39**	**41**	**42**
MAE$_{Body}$	58	78	70	63	73	80	64	68	74	81	58
Photon Therapy:											
Mean ΔDose GTV (%)	-0.7	-1.1	0.6	0.4	-0.4	1.1	-0.7	-0.4	-0.3	-0.5	-1.2
Pass Rate 3%/3mm (%)	99.8	99.6	99.8	99.9	99.8	99.5	99.9	99.7	99.9	99.9	99.8
Pass Rate 2%/2mm (%)	99.5	98.2	99.6	99.7	98.9	98.6	99.9	99.4	99.5	99.8	99.1
Proton Therapy:											
Mean ΔDose GTV (%)	-2.6	-5.4	-0.9	-0.5	2.3	1.5	0.6	-7.5	-3.6	-0.1	-4.5
Pass Rate 3%/3mm (%)	99.6	98.3	99.4	99.6	99.2	99.1	99.4	98.3	97.1	99.2	98.3
Pass Rate 2%/2mm (%)	97.9	95.0	98.7	99.1	97.5	97.9	98.7	95.0	97.2	98.2	97.3