



# Deep learning in MR-guided radiotherapy

Kévin Brou Boni

## ► To cite this version:

Kévin Brou Boni. Deep learning in MR-guided radiotherapy. Medical Imaging. Université de Lille, 2021. English. NNT : 2021LILUB008 . tel-03523701

**HAL Id: tel-03523701**

**<https://theses.hal.science/tel-03523701>**

Submitted on 12 Jan 2022

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

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

Université de Lille

THÈSE

Présentée en vue d'obtenir le grade de

**Docteur**

En

**Automatique, Génie informatique, Traitement du signal et image**

Par

**Kévin N. D. Brou Boni**

Titre de la thèse

Deep learning in MR-guided radiotherapy

*Radiothérapie guidée par IRM via réseaux de neurones profonds*

Soutenue le 6 octobre 2021 devant le jury d'examen :

<i>Présidente</i>	Jocelyne Troccaz	DR CNRS, Université de Grenoble-Alpes
<i>Rapporteurs</i>	Su Ruan	PR, Université de Rouen
	David Sarrut	DR CNRS, INSA Lyon
<i>Examineur</i>	Philippe Lambin	Professor, Maastricht University
<i>Directeurs de thèse</i>	Nick Reynaert	Professeur, Université libre de Bruxelles
	John Klein	MCF (HDR), Université de Lille
	David Pasquier	MCU-PH, Université de Lille

Thèse préparée au

*Centre de Recherche en Informatique Signal et Automatique de Lille*

Université de Lille, Centrale Lille, CNRS, UMR 9189 - CRISAL

Ecole Doctorale SPI 072





# Contents

0	INTRODUCTION	11
	General introduction to radiotherapy	11
	Techniques of irradiation	11
	Radiotherapy in clinical workflow	12
	Magnetic Resonance in Radiotherapy	14
	Description of the project	15
	Problem statement and contributions	15
	Outline of the manuscript	16
1	TOOLS AND THEORETICAL ASPECTS	19
1.1	Energy transport and deposit	19
1.1.1	<i>Photons, the carriers</i>	19
1.1.2	<i>Electrons, the ionizers</i>	20
1.1.3	<i>Dose distribution and Monte Carlo</i>	20
1.2	Magnetic resonance imaging	22
1.2.1	<i>Nuclear magnetic resonance</i>	22
1.2.2	<i>Image reconstruction</i>	24
1.3	Computed tomography	25
1.3.1	<i>X-ray tomography</i>	25
1.3.2	<i>In radiotherapy planification</i>	25
1.4	Performance comparison indicators	26
1.4.1	<i>Voxelwise comparison</i>	26
1.4.2	<i>Dose volume histogram</i>	27
1.4.3	<i>Voxelwise dose difference</i>	27
1.4.4	<i>Gamma dose distribution evaluation tool</i>	27
1.5	Deep learning in computer vision	28
1.5.1	<i>The mechanics of machine learning</i>	29
1.5.2	<i>Perceptron</i>	29
1.5.3	<i>Multiple layer perceptron</i>	31
1.5.4	<i>Convolutional neural net</i>	33
1.5.5	<i>Generative adversarial net</i>	33
1.5.6	<i>Feature scaling or normalization</i>	35



## 2 SYNTHESIZING CT FROM MR IMAGES 37

- 2.1 Bulk density method 38
- 2.2 Atlas-based method 38
- 2.3 Voxelwise conversion 39
- 2.4 Machine Learning 40
  - 2.4.1 Patch-based 41
  - 2.4.2 Convolutional neural network 41
  - 2.4.3 Generative adversarial network 42
- 2.5 Conclusion 42

## 3 MR TO CT SYNTHESIS WITH PAIRED DATA 45

- 3.1 Introduction 45
- 3.2 A conditional GAN for MR-to-CT synthesis 46
  - 3.2.1 cGAN baseline 46
  - 3.2.2 The *pix2pixHD* network 47
  - 3.2.3 Tailored architecture for *sCT* generation 48
- 3.3 Experimental material and implementation details 48
  - 3.3.1 Patient data collection 48
  - 3.3.2 Image pre-processing 49
  - 3.3.3 Training of the network 49
  - 3.3.4 *sCT* evaluation 50
- 3.4 Results 50
- 3.5 Discussion and Conclusion 51

### APPENDICES

- 3.A Network details 55

## 4 MR TO CT SYNTHESIS WITH UNPAIRED DATA 57

- 4.1 Introduction 57
- 4.2 From cycle GANs to an augmented cycle GAN for MR-to-CT synthesis 59
  - 4.2.1 Unsupervised learning of one-to-many 59
- 4.3 Experimental material and implementation details 62
  - 4.3.1 Patient data collection 62
  - 4.3.2 Image pre-processing 62
  - 4.3.3 Training of the networks 64
  - 4.3.4 *sCT* evaluation 64
- 4.4 Results 65
  - 4.4.1 Image comparison 65
  - 4.4.2 DVH analysis 66
  - 4.4.3 Dose difference 66

4.4.4 <i>CycleGAN comparison</i>	67
4.5 Discussion	67
4.6 Conclusion	70
APPENDICES	
4.A Fiducial markers	71
4.B Network details	72
DISCUSSION AND CONCLUSION	
	75
BIBLIOGRAPHY	
	83
ACRONYMS	
	91



# List of Figures

1	Cyberknife linear particle accelerator.	11
2	Different type of Cyberknife collimators.	12
3	Illustration of a conformal treatment compared to a beam intensity modulation.	12
4	Treatment course in radiotherapy.	12
5	Different target volumes in radiotherapy.	13
6	Typical interface of a TPS	13
7	Comparison of axial CT and T2 weighted image.	14
1.1	Photoelectric effect.	19
1.2	Compton scattering.	19
1.3	Pair production.	19
1.4	Electron collision.	20
1.5	Bremsstrahlung effect.	20
1.6	Dose calculated on the TPS for a Cyberknife plan compared to a MC dose.	21
1.7	The trajectory of the tip of $\vec{M}$	23
1.8	FID oscillating at Larmor frequency but damped by the $T_2$ decay.	24
1.9	Tissue classification on the HU scale.	25
1.10	Acquisition of an IVDT.	26
1.11	Gamma distribution between two dose distributions.	28
1.12	The Mechanical Turk.	28
1.13	Example of task performed by ML in radiotherapy.	29
1.14	An artificial neuron with 3 inputs.	29
1.15	multi-layer perceptron.	32
1.16	CNN architecture.	34
2.1	Atlas-based method using deformable image registration.	39
3.1	Generative Adversarial Net.	46
3.2	MR image, CT, sCT and difference (CT – sCT).	50
3.3	DVH parameters comparison.	52
3.4	Dose calculated on CT, sCT and dose difference (CT – sCT).	52
4.1	Paired and unpaired MR-CT images.	58
4.2	Distributions on MR intensity according different scanners.	60
4.3	Augmented cycle generative adversarial net.	62
4.4	MR, CT, sCT images, and the difference (CT–sCT).	65
4.5	DVH parameters comparison.	66
4.6	Dose calculated on CT, sCT and dose difference (CT – sCT).	67
4.7	MRI and CT cycle loss in training phase	68
4.8	sCT generated from two site 5 patients.	69
4.9	sCT generated from two site 4 patients.	69
4.A.1	Bias effect on AugCGAN generation.	71

4.B.1 Cycle starting from an MRI.	73
4.2 Illustration of the workflow used to evaluate our methods.	76
4.3 Comparison between fake and real segmentation.	79
4.4 Segmentation exported in Raystation.	79
4.5 Dose refinement net.	81

# List of Tables

3.1	Acquisition settings for the three sites.	49
3.2	Average MAE and ME in HU when training with pix2pixHD based model and pix2pix.	51
3.3	Mean dose difference between CT and sCT.	53
4.1	Acquisition settings for the five sites.	63
4.2	Average MAE and ME in between sCT and real CT.	65
4.3	Mean dose difference between CT and sCT.	66



# Introduction



## 0.1 GENERAL INTRODUCTION TO RADIOTHERAPY

Radiotherapy is a cancer treatment that has always been associated for a long time with medical imaging. By definition, radiotherapy involves the use of high doses radiation to kill cancer cell or slows their growth and shrink tumors by damaging their DNA.

Even if the benefits of ionizing radiation have been proven (Morgan, 1969; Kogel, Joiner, and Van der Kogel, 2009), their harmfulness remains a concern; the dosimetry was born. Radiation dosimetry is the measurement, calculation and assessment of the ionizing radiation dose absorbed by the human body. It allows an optimization of dose delivery in radiotherapy. The physical unit used to predict the effects of radiation on cells is called the ionizing energy deposited in the tissues per unit mass, also called dose and measured in Gray<sup>1</sup>. To ensure accurate treatment planning, work was done to reduce uncertainties in the dose delivered in tumor while decreasing the dose delivered in healthy tissues. To maintain an acceptable level of complication, the total dose delivered to the tumor, and by extension to nearby healthy tissues, must be limited and spread over time, according to radiobiological criteria dependent on tumor characterization, anatomical location, and treatment technique. External beam radiotherapy includes all treatments in which the source of ionizing radiation is external to the human body. On the other hand, brachytherapy uses internal sources of ionizing radiation.

## 0.2 TECHNIQUES OF IRRADIATION

Various types of techniques can be used in radiotherapy. In most of the cases, a [linear particle accelerator \(Linac\)](#) (see Figure 1) is used to deliver a photon beams.

Inside the [Linac](#), an electron beam is generated by a source and then accelerated by several electronic oscillators. The high speed beam, by hitting a tungsten target will produce a divergent photon beam<sup>2</sup>.

At the beginning of radiation therapy, the photon beam was shaped at the end of the [Linac](#) with a basic collimator resulting in the irradiation of the target together with a significant volume of healthy tissues. Nowadays, radiotherapy devices include a [multi-leaf collimator \(MLC\)](#) (see Figure 2 for comparison) enabling conformational radiotherapy with intensity modulation ([IMRT](#)) in which the photon beam is shaped in conformance with the target by moving the leafs during the emission of the beam. An illustration in term of treated volume is shown in Figure 3.

0.1	General introduction to radiotherapy	11
0.2	Techniques of irradiation	11
0.3	Radiotherapy in clinical workflow	12
0.4	Magnetic Resonance in Radiotherapy	14
0.5	Description of the project	15
0.6	Problem statement and contributions	15
0.7	Outline of the manuscript	16

<sup>1</sup>  $1\text{ Gy} = 1\text{ J.kg}^{-1}$

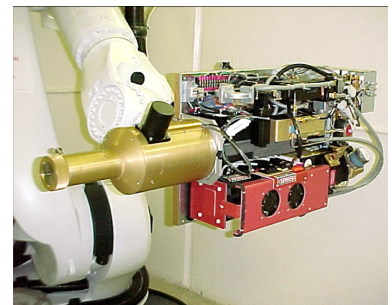


Figure 1: Cyberknife [linear particle accelerator](#) (Accuray Inc., Sunnyvale, CA).

<sup>2</sup>Bremsstrahlung X-rays: electrons by decelerating lose kinetic energy, which is converted in radiation i.e., photons. X-rays penetrate more deeply, but the dose absorbed decreases exponentially along thickness



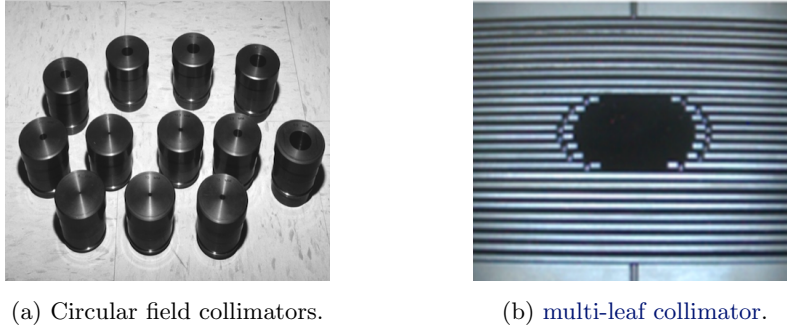


Figure 2: Different type of Cyberknife collimators: (a) Circular fixed collimators. (b) A MLC, each leaf can be positioned independently.

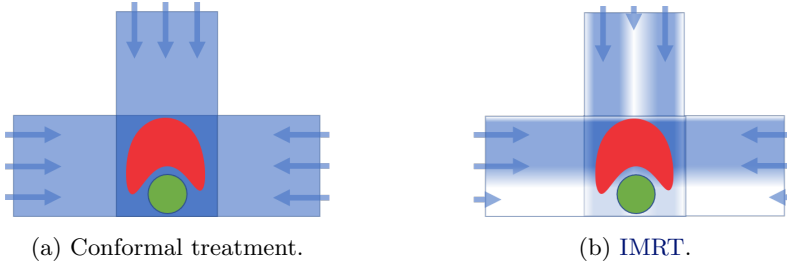


Figure 3: Illustration of a conformal treatment compared to a beam intensity modulation. Target in red and in green an organ to protect.

Other types of radiation can be used and are based on particle beams such as protons or heavier positive ions. In hadron therapy the dose deposit is maximum near the end of the range of the particle. This kind of treatment requires to a much heavier equipment.

### 0.3 RADIOTHERAPY IN CLINICAL WORKFLOW

Between the radiotherapy treatment prescription and the first irradiation, the patient undergoes several steps. These steps are illustrated in Figure 4 and explained below

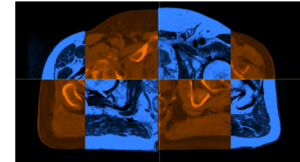
*Pretreatment imaging* The patient is placed in treatment position with specific immobilization devices in order to optimize the reproducibility of positioning during the treatment sessions. A CT (computed tomography) is then acquired. This imaging modality provides a map of the physical densities of the internal tissues of the patient in Hounsfield unit (HU).

*Image registration* Complementary informations can be added to this simulation CT by merging the benefits of other imaging modalities. For instance higher contrast image or a functional imaging of the tumor, this can be done with the use of the MRI (Magnetic Resonance Imaging) or PET (positron emission tomography). The overlaying mitigates the weakness of each technique by using image registration algorithms. Rigid registration is most commonly used in clinical practice, only translations and rotations are allowed. In contrast, non-rigid (deformable) registration uses geometric transformation that does not preserve the Euclidian distance between every pair of points.

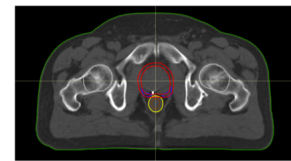
#### 1. Pre-treatment imaging CT



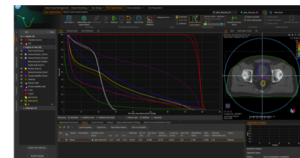
#### 2. Registration MR/PET to CT



#### 3. Target and OAR delineation



#### 4. Treatment planning



#### 5. Treatment delivery



Figure 4: Treatment course in radiotherapy.

*Delineation* This step consists of localizing the tumor volume and the **organs-at-risk (OARs)** that must be protected during treatment. They are delineated by a radiation oncologist on the **CT** or on other modalities. Conventions were defined by **International Commission on Radiation Units and Measurements (ICRU)** in order to standardize the definition of the target. The report clearly defines three volumes of interest. The **gross tumor volume (GTV)** corresponds to the tumor visible on the imagery. Then the **clinical target volume (CTV)** that encompasses the **GTV** with a margin taking into account regions known to carry malignant cells. Finally, the **planning target volume (PTV)** taking into account positioning uncertainties of the patient and the movement of the target by adding a margin to the **CTV**. Figure 5 represents the **ROIs**.

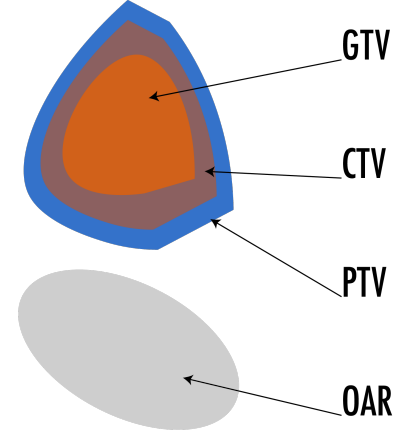


Figure 5: Different target volumes in radiotherapy.

*Treatment planning* In order to plan a radiotherapy treatment and to calculate the dose distribution inside the patient, a virtual simulation including the modeling of the patient and the treatment beams is performed on a **treatment planning system (TPS)** by a dosimetrist or a medical physicist. During the virtual treatment planning, the beam arrangement, the collimator arrangement, and the active beam time are to be determined, either manually (direct planning) or automatically (inverse planning), in order to design the treatment plan. The treatment plan is established in accordance to the prescribed dose to the **PTV** while limiting the irradiation to the **OARs**. The patient is modeled from the previously acquired **CT** scan. This image, in addition to providing information on the patient geometry, allows the calculation of the dose by the **TPS** algorithm thanks to the information in **HU** contained in the voxels. **CT** is thus the reference modality for any radiotherapy treatment. **Dose-volume histograms (DVHs)** representing the dose received to a percentage of each of the **ROIs** and the dose distribution in 3D are used in combination during optimization and for the evaluation of the treatment (see Figure 6).

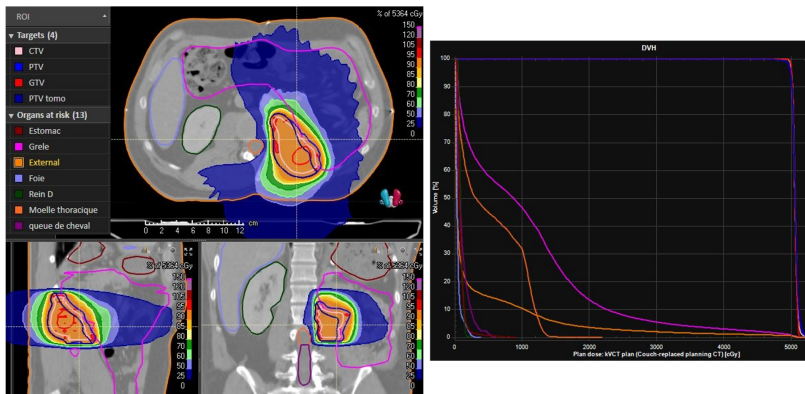


Figure 6: Typical interface of a **TPS**. Left is a visualization of the **CT** with the isodose lines superimposed in colors, the legend represents the **ROIs**. Right is the **DVHs** curves for each of the **ROIs**: the targets appear as steep lines at the right of the graph (homogeneous doses), whereas the **OARs** show smoother curves with various dose levels depending on the sparing reached for this specific plan (RayStation, RaySearch Laboratories).

*Treatment* During each session, the patient must be correctly positioned, using the lasers and/or skin markers. Motion restriction is carried out by the same immobilization devices as in the pre-treatment

imaging. Then an image is acquired using a **Linac**-based on-board or an in-room imaging system enabling a comparison with the images from the simulation. If no major changes, the treatment is then delivered.

The accuracy of dose delivery is prone to several errors and uncertainties related to the workflow introduced above.

#### 0.4 MAGNETIC RESONANCE IN RADIOTHERAPY

Since its introduction in the late 70s until now, **MRI** is more and more used in radiotherapy. It is particularly appreciated for its superior soft tissue contrast compared with **CT** (Figure 7).

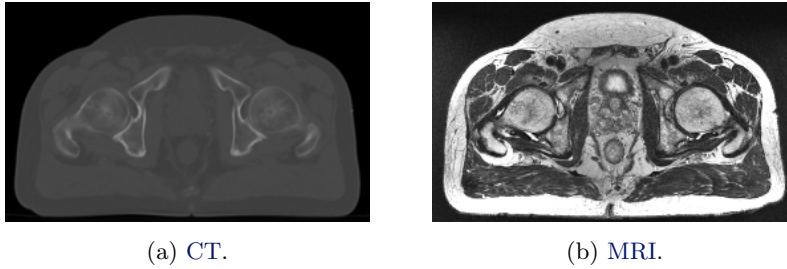


Figure 7: Comparison of axial **CT** (a) and T2 weighted **MRI** (b).

As a result, **MRI** helps to better discern tumors than a **CT**, and has therefore been integrated into radiotherapy practices as a diagnostic image. Tumor volume is delineated more precisely on **MRI** compared with **CT**, whether in the case of prostate tumors (Debois et al., 1999; Rasch et al., 1999), brain lesions (Prabhakar et al., 2007), or head and neck tumors (Lemort, Canizares, and Kampouridis, 2006). These studies have shown that **MRI** was able to reduce inter-observer variability during the delineation. In practice, after the **MRI** acquisition, **ROIs** are contoured directly on this image by a radiation oncologist. Since the simulation **CT** serves as the basis for dose calculation, the contours drawn on the **MRI** are propagated to the **CT** by using a registration algorithm. This step is considered in the community as the weakest link in radiotherapy workflow (Rasch et al., 1999; Njeh, 2008). The image registration introduces spatial uncertainties more or less significant depending on the localisation. They are due to inter-scan differences in positioning between the two examinations or anatomical changes of deformable organs such as the filling of the rectum and/or bladder which influences the position of the prostate. The total uncertainty was estimated at 3.7 mm (head-foot direction) for prostate treatment by Nyholm et al. (2009), meaning that "virtual" target can therefore be shifted a few millimeters from the "real" target. In addition, they have also estimated the total uncertainty at 2.9 mm by using a **MRI**-only workflow. The idea behind is to replace the **CT** simulator by an **MRI** simulator, reducing systematic spatial uncertainties introduced by the inter-scan differences. The removal of a **CT** acquisition in the workflow allows to eliminate the registration error but also the tissue density informations needed for dose calculation.

**MRI** does not map tissue density, this issue will be the focus of

the current thesis, different solutions, based on the conversion of [MRI](#) images into so-called [synthetic CT \(sCT\)](#). In a [sCT](#) image, [HU](#) or tissue densities are mapped from [MRI](#) intensities.

## 0.5 DESCRIPTION OF THE PROJECT

The genesis of this thesis stems from the European INTERREG [Cooperative Brachytherapy \(CoBra\)](#) project. This project aims to improve the quality of both diagnosis and treatment of localized cancers, by developing a new medical robot prototype for brachytherapy and biopsy under [MRI](#) guidance.

The goal is to achieve an adaptive tumor tracking and dose control in real-time and to reach the overall objective, which is to enhance quality of life for patients and reduce cancer mortality.

Co-financed by the European Regional Development Fund in the 2 Seas area, [CoBra](#) is a cross-border cooperation with partners from coastal areas of England, France and the Netherlands.

The lead partner is the university of Lille which is in charge of building the robot. Core partners are TU Delft in charge of developing specific steerable needles and Demcon which is a company working on their biopsy guidance devices. All these tasks are meant to be solved under [MRI](#) control. NHS Portsmouth and the university of Portsmouth work on the time optimization of the treatment plan.

The present study falls within the range of dose calculation under [MRI](#). This task has to be fast since the treatment plan needs to be optimized on the fly. The workflow can be split in three steps: ideally, the [sCT](#) should be generated along with the delineation of the [ROIs](#). With these two at hand, one can compute a dose distribution according to a treatment plan given by an optimization algorithm. All of these steps must be performed in real time.

## 0.6 PROBLEM STATEMENT AND CONTRIBUTIONS

The original ambition of this thesis was to develop a dose calculation method based solely on the [MRI](#) as part of the [CoBra](#) project. This task implies [CT](#) image synthesis and segmentation as well as dose calculation. All of these sub-tasks are non trivial scientific challenges. Among them, segmentation is probably the one that received the greatest attention from the medical image processing community. Consequently, it was decided to focus on the other two with a focus on [sCT](#) generation as it is the starting point of the desired pipeline.



In this thesis, we tackle generalization problem related to deep learning methods in [sCT](#) generation task.

Recent advances in deep learning have made it possible to simplify the [sCT](#) generation workflow. Less and less pre and post-processing while being fast and accurate. These methods rely on a dedicated [MRI](#) sequence (or a set) as an input; conflicting with the [magnetic resonance \(MR\)](#) intensities variability in clinical reality. In some extent, a model is fitted for one [MRI](#) sequence.

In an attempt to generalize the idea of a model not only dedicated to a single sequence, we investigated a way to induce the MR intensities variability in the training of a deep learning model. This will allow in the future to generate a sCT without the need of a dedicated dataset.

BELOW IS A LIST OF OUR CONTRIBUTIONS.



*Journal paper (cf. Chapters 3 and 4)*

-  K. N. B. Boni, J. Klein, L. Vanquin, A. Wagner, T. Lacornerie, D. Pasquier, and N. Reynaert. 2020. *MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network*. Physics in Medicine & Biology.
-  K. N. Brou Boni, J. Klein, A. Gulyban, N. Reynaert, and D. Pasquier. 2021. *Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data*. Medical Physics.


*Collaboration in journal paper*

-  A. Wagner, K. B. Boni, E. Rault, F. Crop, T. Lacornerie, D. Van Gestel, and N. Reynaert. 2020. *Integration of the M6 Cyberknife in the Moderato Monte Carlo platform and prediction of beam parameters using machine learning*. Physica Medica.

*Conference papers*

-  K. N. Brou Boni, A. Wagner, L. Vanquin, J. Klein, D. Pasquier, and N. Reynaert. 2019b. *High-Resolution Synthetic-CT Generation with Conditional Generative Adversarial Networks*, in *Magnetic Resonance in Radio-Therapy*. MRinRT 2019, Toronto, Canada, July 2019.
-  K. Brou Boni, A. Wagner, L. Vanquin, J. Klein, N. Reynaert, and D. Pasquier. 2019a. *Génération de tomodensitométrie synthétique par apprentissage profond pour la radiothérapie du cancer de la prostate basée sur l'IRM seule*. 30e Congrès national de la Société française de radiothérapie oncologique, Cancer/Radiothérapie.

*Workshop papers*

-  K. B. Boni, J. Klein, L. Vanquin, D. Pasquier, and N. Reynaert. 2019. *Génération d'un CT synthétique à partir d'une séquence IRM avec un réseau antagoniste génératif en radiothérapie*. In GRETSI 2019.

## 0.7 OUTLINE OF THE MANUSCRIPT

THIS THESIS is an attempt to provide some key elements to the dissemination of dose calculation on MRI in clinical routine.

CHAPTER 1 introduces the ingredients for a proper comprehension of this subject. In particular, imaging and irradiation techniques related to our work. The indicators for comparing the respective performances of sCT generation methods are presented in the last section.

CHAPTER 2 discusses several methods available to generate an accurate sCT. The trade-offs of each type of method are summarized at the end of this chapter.

CHAPTER 3 explores the first attempts to use a multi-centric cohort in order to induce MR intensities variability in a deep learning model. Importantly the approach does not require any type of MRI histograms normalization. Hence it demonstrates promising results but suffer from the need of non-rigid registration for the training, which is time

consuming. This chapter includes material accepted to an international journal (Boni et al., 2020).

CHAPTER 4 goes further by presenting a method that does require only rigid registration, thus alleviating once again the workflow. This chapter includes material accepted to an international journal (Brou Boni et al., 2021).

THE FINAL CHAPTER in page 75 presents conclusions and a discussion on the state of the work in the continuation of this thesis.





# Tools and theoretical aspects

# 1

In this chapter, the medical imaging techniques used in this thesis are presented with the key concept of dose. MRI and CT, which were briefly described in the introduction, will be explained in more details from their basic physics concepts to image property. Secondly, the key indicators for comparing the respective performances of sCT generation methods will also be introduced. Since this thesis focuses mainly on convolutional neural networks (CNN) solutions, we will finish this chapter with the key concepts concerning this type of neural net.

## 1.1 ENERGY TRANSPORT AND DEPOSIT

This section presents the mechanism of transport and interaction between a photon beam and the medium. This explanation will help us to understand how the dose is deposited in the body.

### 1.1.1 Photons, the carriers

The photon has zero mass and zero electric charge. Contrary to charged particles, its displacement does not necessarily imply an interaction with the environment and can therefore just cross a body without any interaction. The probability  $p$  that a photon interacts through a linear section  $dx$  is given by

$$p = \nu \cdot dx \quad (1.1.1)$$

where  $\nu$  is the linear attenuation coefficient ( $cm^{-1}$ ) and depends on the energy of the photon and the electron density of the material.

When a photon interacts with an atom, a transfer of energy towards an electron ( $e^-$ ) of this atom is made. This transfer of energy has the effect of setting the electron in motion outside the electronic cloud (ionization). This interaction is called the photoelectric effect (Figure 1.1).

The second possible interaction is the Compton scatter. The incident photon will set in motion an electron of the peripheral layers outside the electronic cloud (ionization). The incident photon is deviated from its initial trajectory, it is scattered (Figure 1.2). It is the predominant effect in radiotherapy.

Last inelastic phenomenon is pair production occurring only at higher energy level. Near a nucleus, the incident photon is absorbed and a positron-electron pair is produced (Figure 1.3). Ultimately, the positron ( $e^+$ ) will meet an electron. The two particles will annihilate each other by emitting two annihilation photons of equal energy in opposite directions (concept used in PET scan).

All these interactions imply a transfer of energy and not a deposit

1.1 Energy transport and deposit	19
Photons, the carriers	
Electrons, the ionizers	
Dose distribution and Monte Carlo	
1.2 Magnetic resonance imaging	22
Nuclear magnetic resonance	
Image reconstruction	
1.3 Computed tomography	25
X-ray tomography	
In radiotherapy planification	
1.4 Performance comparison indicators	26
Voxelwise comparison	
Dose volume histogram	
Voxelwise dose difference	
Gamma dose distribution evaluation tool	
1.5 Deep learning in computer vision	28
The mechanics of machine learning	
Perceptron	
Multiple layer perceptron	
Convolutional neural net	
Generative adversarial net	
Feature scaling or normalization	

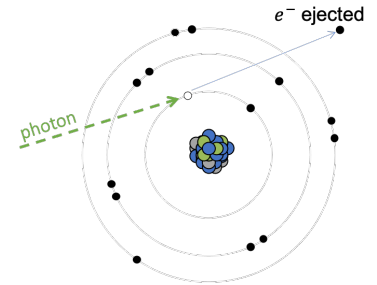


Figure 1.1: Photoelectric effect.

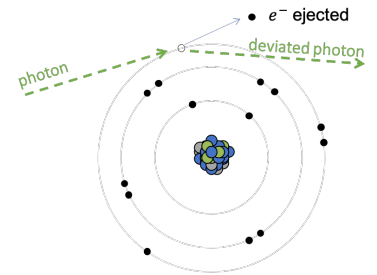


Figure 1.2: Compton scattering.

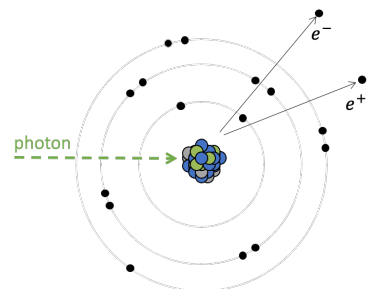


Figure 1.3: Pair production.



of energy. Photon beam is an indirectly ionizing radiation. It is the released electrons that will deposit their kinetic energy on their path by Coulombic interactions in the medium.

### 1.1.2 Electrons, the ionizers

Unlike photons, electrons are charged particles. They are therefore subject to many more interactions with the medium, which has the effect of quickly slowing them down, until they stop completely.

Electrons passing through the medium can interact with an electronic cloud. The collision between the incident electron and one at equilibrium will cause sometimes just a excitation or the ejection of this one (ionization, Figure 1.4), at the cost of a loss of energy of the incident electron.

This is called inelastic collision. A hole appears on the electronic cloud which is quickly filled by an electron from a neighboring layer generating a X-ray. The emitted energy is characteristic of the difference in energy between two electronic level. An electron of sufficient energy has to interact inelastically several hundred times before losing all its kinetic energy. The succession of ionization is the main source of biological damage due to localized radiation (ionization).

Less often, the electron is slowed down and it is deviated while passing in the vicinity of a nucleus. The energy lost by the electron is carried away by a photon, called secondary (Bremsstrahlung X-ray, Figure 1.5).

Electrons tend to lose kinetic energy continuously as they travel, the rate of energy lost is called the stopping power<sup>1</sup> and is related to the two previous phenomenons.

The energy lost  $E$  by the collision of electrons in an infinitesimal sample  $\partial V$  of density  $\rho$  is related to the dose received in the sample

$$D = \frac{E}{\rho \cdot \partial V} = \frac{E}{m} \quad (1.1.2)$$

where  $m$  is the mass of the infinitesimal sample. Other phenomenons of increasing complexity can alter the dose deposit, we just give a brief explanation.

*Inverse-square law* The fluence which is the number of crossing photons over a defined surface area is inversely proportional to the square of the distance from the source.

*Heterogeneity and interfaces* When several materials compose the medium, the dose distribution takes a less regular form. The greater the electronic density of the medium, the more the electron interacts, and the shorter its range.

### 1.1.3 Dose distribution and Monte Carlo

Several algorithms calculate a dose distribution in the patient according to the established treatment plan, i.e. a unique set of beam param-

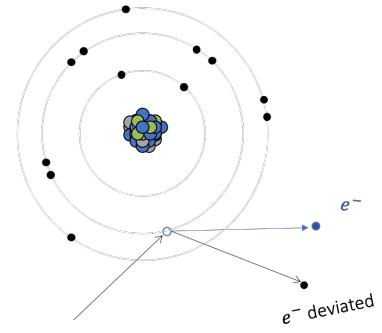


Figure 1.4: Electron collision.

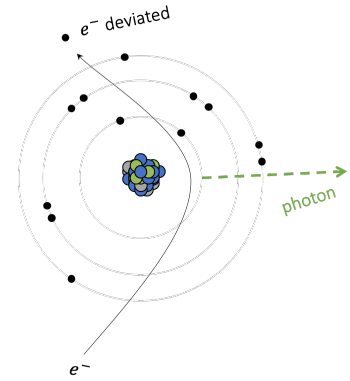


Figure 1.5: Bremsstrahlung effect.

<sup>1</sup>Similar to the attenuation coefficient but for charged particles.

eters (number of beams, orientation, MLC). Among the current dose calculation algorithms, Monte Carlo (MC) algorithms are considered to be the most accurate, because tissue heterogeneity and other complex interactions are fully taken into account for the dose calculation.

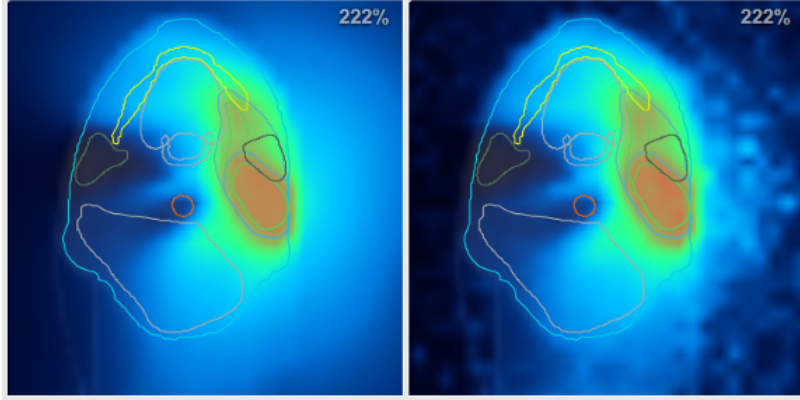


Figure 1.6: Dose calculated on the TPS compared to a MC dose re-calculated on Moderato (Reynaert et al., 2016).

MC algorithms are stochastic methods for solving numerical problems for which no analytical formulation is defined<sup>2</sup>. From the probability distributions governing the interactions of electrons and photons, the transport of these particles in medium are simulated. The resulting physical quantities, such as the deposited dose, can be calculated by generating a very large number of simulated particles, called histories. The system uses the physical properties of the radiation in combination with random number generators to determine when the particles will interact in the medium, and the type of interaction. Each history can be summarize in two successive steps

*Straight line motion* the particle (photon or electron) moves in a straight line and without any interaction. The length of its path is obtained by a random draw on the probability distribution of the length of the path of a particle. We can therefore know the position where the particle interacts (or not).

*Interaction* The position of the interaction having been simulated, its type and characteristics must be determined. The type of interaction is first determined by random draw, from the probabilities of each interaction. Then the interaction itself is simulated, new random draws give the energy loss and the new direction of the incident particle, but also the characteristics of any secondary particle created. These secondary particles then enter in the *Straight line motion* step.

However, a large number of histories have to be simulated to obtain an acceptable level of uncertainty. The calculations can therefore take several hours. The distribution probability of the different interactions depends mainly on the energy of the particle, as well as on the electronic density of the medium. Their description<sup>3</sup> is out of scope of this thesis.

<sup>2</sup>Monte Carlo typically allows to obtain an estimate of an intractable integral  $I = \int f d\mu$  where  $f$  is some (integrable) function of interest and  $\mu$  a probability measure. The approximation is  $I \approx \frac{1}{n_k} \sum_{k=1}^{n_k} f(x^{(k)})$  where the  $x^{(k)}$  are drawn from  $\mu$ . The estimator is consistent, i.e. asymptotically correct when the number of samples  $n_k$  goes to infinity, a property arising from the law of large numbers.

<sup>3</sup>Practical aspects of Monte Carlo simulation by Salvat et al. (1999) is a good start.

## 1.2 MAGNETIC RESONANCE IMAGING

### 1.2.1 Nuclear magnetic resonance

Magnetic Resonance Imaging is based on the fundamental concept of nuclear magnetic resonance (NMR). As its name indicates, this process relates to atom nucleus, here the hydrogen nucleus. The hydrogen nucleus contains only a proton<sup>4</sup> and is widely represented in the human body because it is a component of fat or water molecules. A proton is a positive charged particle and it is animated by a movement of rotation on itself (often referred to as spin) and thus induces a magnetic field of very small amplitude. This magnetic field is characterized by the nuclear magnetic moment<sup>5</sup>  $\vec{\mu}$  which is proportional to its quantum spin number  $S = \{-1/2, 1/2\}$

$$\vec{\mu} = \gamma \hbar \vec{S} \quad (1.2.1)$$

where  $\gamma$  is the gyromagnetic ratio, a constant equal to 42.6 MHz/T for an hydrogen nucleus and  $\hbar$  the reduced Planck constant.

The average nuclear magnetic moment from all the protons in a volume  $V$  is called net magnetization  $\vec{M}$

$$\vec{M} = \frac{1}{\#V} \sum_{\text{proton} \in V} \vec{\mu}(\text{proton}) \quad (1.2.2)$$

where  $\#$  denotes the set cardinality operator. Outside any external magnetic field, the magnetic moments of the protons are completely arbitrary and tend to compensate each other on average, i.e.  $\vec{M} = \vec{0}$ . Inside an external magnetic field  $\vec{B}_0$ , protons precess<sup>6</sup> around this field and the magnetic moments will orient themselves according to the rules of quantum mechanics. They tend to sort themselves into either an aligned state (lower energy and same direction as  $\vec{B}_0$ ) or an unaligned state (higher energy and opposite direction of  $\vec{B}_0$ ). The energy of a particular energy level is given by

$$E = -\vec{\mu} \cdot \vec{B}_0. \quad (1.2.3)$$

The exact distribution of magnetic moments into aligned or unaligned states can be predicted by the Boltzmann distribution

$$\frac{N^+}{N^-} = \exp\left(-\frac{\Delta E}{kT}\right) \quad (1.2.4)$$

where  $N^-$  and  $N^+$  are respectively the number of proton in aligned and unaligned state,  $\Delta E$  the energy gap between two nuclear spin state,  $k$  is the Boltzmann's constant equal to  $1.380649 \times 10^{-23} \text{ J/K}$  and  $T$  is the absolute temperature. This result states that only a few spins per million participate in the effective magnetization  $\vec{M}$  aligned with  $\vec{B}_0$ .<sup>7</sup> At equilibrium, for total number of spins  $N$ , the amplitude

<sup>4</sup>Because a proton is the sole sub-atomic particle inside hydrogen nucleus, the words proton and hydrogen atom are often used interchangeably in the MRI literature.

<sup>5</sup>The nuclear magnetic moment can be pictured as a tiny compass needle that will align with stronger magnetic fields when placed in one such field.

<sup>6</sup>The external magnetic field generates a force and the proton will start moving accordingly. But remember that a proton is already (somewhat) spinning, so when moving an object which already has a rotational inertia, one gets an oscillating movement. This oscillation is called precessing. It is a similar kind of movement that a spinning top will have when gravity tries to have it falling.

<sup>7</sup>At body temperature,  $\frac{\Delta E}{kT} \ll 0$ , therefore  $\frac{N^+}{N^-} \approx 1 - \frac{\Delta E}{kT} < 1$ . One can understand  $\frac{\Delta E}{kT}$  as the proportion of spin in excess.

of the net magnetization can be denoted  $M_0$

$$\begin{aligned} M_0 &= \frac{1}{V}(N^- - N^+)\mu = \frac{1}{V}(N^- - N^+)\frac{\gamma\hbar}{2} \\ &\approx \frac{\gamma\hbar N^-}{2V} \frac{\Delta E}{kT} \approx \frac{\gamma\hbar N^-}{2V} \frac{\gamma\hbar B_0}{kT} \\ &\stackrel{(a)}{\approx} \frac{\gamma^2\hbar^2}{4kT} \frac{N}{V} B_0 \approx \frac{\gamma^2\hbar^2}{4kT} \rho B_0 \end{aligned} \quad (1.2.5)$$

where  $\rho$  is the number of protons per unit volume or spin density and  $\hbar$  the reduced Plank constant equal to  $1.054 \times 10^{-34} Js$ . It is this magnetization or signal (1.2.5) which is measured in NMR because it is proportional to the number of protons present in the sample and also to the amplitude of the magnetic fields.

Besides, a magnetic moment is animated by a precession movement around the  $\vec{B}_0$  axis. The precession frequency  $f_L$  is directly related to the amplitude of the magnetic fields and the gyromagnetic ratio by the Larmor relationship

$$f_L \stackrel{(b)}{=} \frac{\omega_L}{2\pi} = \gamma B_0. \quad (1.2.6)$$

Even if all the individual nuclei are precessing,  $\vec{M}$  is not. It possesses only a longitudinal component  $M_z$  which is not measurable in that direction. The goal here is to tilt  $\vec{M}$  in the transverse plane (xy) in order to have a measurable value. This is done by using a radio-frequency coil  $\vec{B}_1$  rotating around the body at the Larmor frequency in the transverse plane. When transmitting, the induced signal will cause  $\vec{M}$  to precess in the transverse plane near the Larmor frequency. When the radio-frequency coil stop transmitting,  $\vec{M}$  will still precess while returning<sup>8</sup> to its equilibrium state. The net magnetization has now three time varying components, a longitudinal  $M_z(t)$  and two transverses  $M_x(t)$ ,  $M_y(t)$ . Bloch (1946) has described the behavior of these components when returning to equilibrium by introducing two relaxation time constants  $T_1$  and  $T_2$

$$M_x(t) = M_0 e^{-\frac{t}{T_2}} \sin(\omega_L t) \quad (1.2.7a)$$

$$M_y(t) = M_0 e^{-\frac{t}{T_2}} \cos(\omega_L t) \quad (1.2.7b)$$

$$M_z(t) = M_0 (1 - e^{-\frac{t}{T_1}}). \quad (1.2.7c)$$

$T_1$  is the longitudinal relaxation time and is associated with the exponential regrowth of the longitudinal component  $M_z(t)$ .  $T_2$  is the transverse relaxation time and is associated with the exponential decays of the transverse components  $M_x(t)$  and  $M_y(t)$ . An illustration of the trajectory of the tip of  $\vec{M}$  back to equilibrium is proposed in Figure 1.7. These times are characteristic of the immediate surrounding of the nuclei, and vary from a few microseconds ( $T_2$  in cortical bone) to several seconds ( $T_1$  in cerebrospinal fluid) in the human body. These time lapses are the pieces of information allowing to provide ultimately tissue contrast in the image that will be reconstructed from these recordings.

(a)  $N \approx 2N^-$  the total number of protons.

(b) The cyclic frequency is denoted  $f_L$  while the angular frequency is denoted  $\omega_L$ .

<sup>8</sup>In general, protons are represented only in a state “up” or “down” but people often tend to forget that there is a gradual transition between these two states. This result in the fact that the net magnetization amplitude  $M$  does not fully develop instantly.

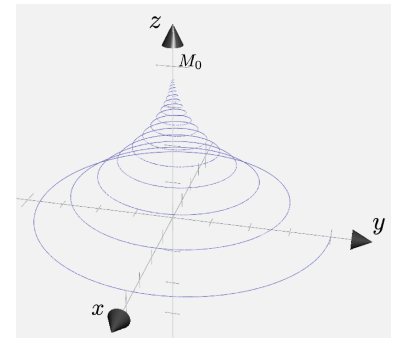


Figure 1.7: The trajectory of the tip of  $\vec{M}$  is a spiral pattern. Parameters are  $M_0 = 1$ ,  $T_1 = 4$ ,  $T_2 = 2$  and  $\omega_L = 10$ .

The transverse magnetization produces a signal in the form of an electromagnetic wave at the Larmor frequency which can be collected by a receiver coil placed (or almost) in contact with the patient. The origin of this signal comes from the rotational movement of the magnetization through a receiver coil according to the Faraday's laws of induction. The shape of the signal collected after the magnetization has been tilted in the transverse plane is called the **free induction decay (FID)** signal. The resulting signal follows a sine wave with an exponentially decaying amplitude according to the Bloch equations, as in Figure 1.8.

The real **MR** signal is obtained by using the **FID** of the two orthogonal directions ( $fid_x, fid_y$ ). The representation of the signal as a complex number

$$signal = fid_x + i \cdot fid_y \quad (1.2.8)$$

of magnitude  $\sqrt{fid_x^2 + fid_y^2}$  and phase  $\arctan \frac{fid_x}{fid_y}$ . It is the magnitude which is used in clinical **MRI**.

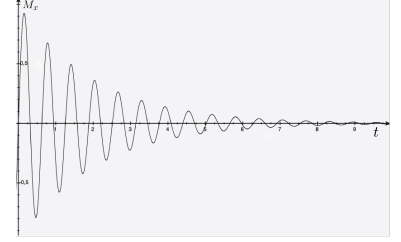


Figure 1.8: **FID** oscillating at Larmor frequency but damped by the  $T_2$  decay.

### 1.2.2 Image reconstruction

Up until now, the **MR** signal have arisen from a single voxel in an homogeneous sample. Indeed, the signal received is the global magnetization of the human body, which is not spatially localized. To produce an image from the **NMR** mechanism, it is necessary to encode the information spatially. This is the role of magnetic field gradients which produce a linear variation of the magnetic field in the three possible dimensions which is added to the main field  $B_0$ . With the right timing, it is possible to differentiate each voxel on a slice.

The simultaneous transmission of the radio-frequency pulse and the z-gradient allows to select the slice. Only the protons in the slice generate a signal. The phase gradient is then activated for a given time, which assigns the protons varying phases along the y-axis. Finally, the frequency gradient comes into play during the measurement of the signal, so that when the information is recorded, the rotation speeds vary along the x-axis. An illustration is given in Figure.

The **FID** for a slice will be the superposition of several signals with different amplitudes, frequencies and phases. The 2D image is reconstructed by using two successive Fourier transforms. Each voxel represents the magnitude of the signal received at this position. The term sequence in **MRI** refers to a set of periodically repeated radio frequency pulses, magnetic field gradients, and signal acquisition windows that put the magnetization to a desired state, and fill the Fourier space integrally.

An important remark on **MR** image reconstruction is that the values of the voxels will have different meanings depending on parameter choices made by the **MR** technician. These parameters are the echo time and the repetition time. Depending on their values, the voxels values will be closely related to  $T_1$ ,  $T_2$  or  $M_0$ . Each of these choices produce tissue contrasted images of much different aspects. These

categories of MR images are often referred to as “sequences” and are a great source of variability among MR images<sup>9</sup>.

### 1.3 COMPUTED TOMOGRAPHY

#### 1.3.1 X-ray tomography

CT is based on the same principle as radiology, i.e. the use of an X-ray source and a detector on each side of the body. It allows a 3D image reconstruction by simultaneously rotating and translating the X-ray source and the detector around the body.

X-ray transmission depends on the linear attenuation coefficients  $\nu$  of the tissue along its path. The attenuation coefficient reflects how easily a tissue can be penetrated by the X-ray. The source intensity  $I$  is attenuated exponentially in the x-axis by:

$$I = I_0 e^{-\int \nu(x) dx} \quad (1.3.1)$$

where  $I_0$  is the initial beam intensity and  $dx$  is the cross section.

A set of projections at different angles allows to compute the attenuation coefficient at each position. The reconstructed 2D image is a mapping of these attenuation coefficients normalized to the water attenuation coefficient  $\mu_{water}$  in the Hounsfield unit according to the formula

$$HU = 1000 \left( \frac{\nu}{\nu_{water}} - 1 \right). \quad (1.3.2)$$

It is an inverse problem solved by the Radon transform. The translation of the source will lead to the reconstruction of the 3D image. Because each tissue has a different attenuation coefficient, a tissue contrasted image volume is thus obtained.

#### 1.3.2 In radiotherapy planification

In radiotherapy, the HU has an important property. Indeed, the HU and the tissue mass density are bilinearly correlated. This relation (when known) implies that the density of a tissue can be determined from a CT image. This physical quantity is precisely the one necessary for the dose calculation algorithms. The relationship between the HU and the mass density is dependent on the energy spectrum of the beam used inherently by each scanner. For this reason, a relationship between the HU and the mass density must be established for each machine and also over the time. This relationship is represented by a calibration curve, also called *image value to density table (IVDT)*. Basically, to determine the calibration curve, the following protocol is run: several test objects are acquired and these objects have known homogeneous densities covering the range of density encountered in human tissue. The CT number of an object measured on the image is then assigned to the density of this object, thus composing a data set of about a dozen points depending on the test objects used. A

<sup>9</sup>There is obviously an inter-class variability but also an intra-class one. Indeed, for a same patient, two technicians might go for slightly different parameters even if they target the same MR sequence. In addition, the surrounding magnetic field is also subject to variations over time. All this accounts for the limited reproducibility of MR images which is one of its drawback compared to other modalities.

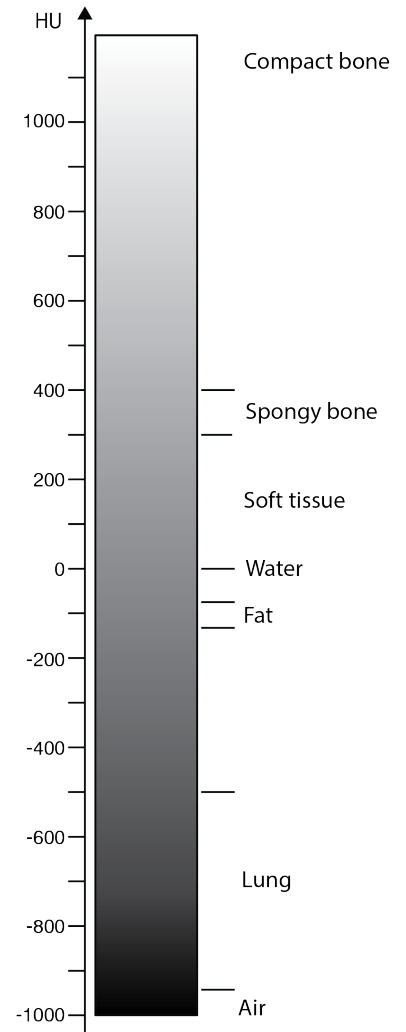


Figure 1.9: Tissue classification on the HU scale.



linear regression in two segments is then performed to extrapolate the calibration curve. A visual representation in Figure 1.10.

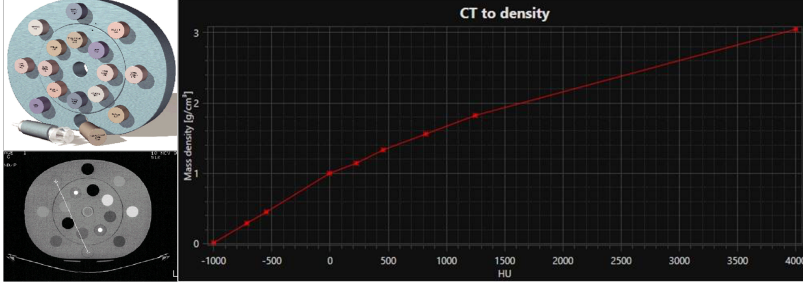


Figure 1.10: Acquisition of an IVDT: a phantom containing inserts of known variable densities (upper left) is scanned to obtain a CT image (lower left) in which the HU of each insert is measured, providing a bilinear correlation between the HU and the mass density (curve on the right).

## 1.4 PERFORMANCE COMPARISON INDICATORS

In order to evaluate the accuracy of a sCT generation method, we first have to see the different comparison metrics used to compare a CT with a sCT. There are several of them because none of them is uniformly better than the others in every aspects. Basically, independently they give us just few hints but their combination brings us deeper insights into their relative qualities.

### 1.4.1 Voxelwise comparison

The mean absolute error (MAE) and the mean error (ME) are two global indicators of the correspondence between the HU of each voxel in two CT images, typically an sCT and the associated CT image of the patient. The latter is considered as the reference since the image is actually produced during a CT examination. The respective formulas<sup>10</sup> for the mean absolute error and the mean error are

$$\begin{aligned} MAE &= \frac{1}{N} \sum_{i=0}^N |ct(i) - sct(i)| \\ ME &= \frac{1}{N} \sum_{i=0}^N ct(i) - sct(i) \end{aligned} \quad (1.4.1)$$

where  $N$  is the total number of voxels and  $i$  the voxel index. They are both expressed in HU.

These indicators do not take into account the spatial distribution of the error, but rather the amplitude and the occurrence of these errors. They are often introduced in the field of sCT generation to summarize the overall performance of a method. It is a good trade-off when it is difficult to perform a dosimetric comparison.

On the one hand, ME is more permissive as it allows the errors of one area to compensate the errors of another one even if they are not adjacent. On the other hand, unlike ME, MAE<sup>11</sup> is more compelling as does not allow such compensations between adjacent areas. A better exploitation of their results will be to use them on the different ROIs instead of the whole body.

<sup>10</sup>The formulas are given as if image volume tensors had been flattened to vectors and can be indexed using only one integer  $i$  instead of the usual  $(i, j, k)$  triplet used to index a voxel.

<sup>11</sup>MAE is also known as the Manhattan distance or the  $L_1$  norm distance. ME cannot qualify as a distance because  $ME = 0$  does not imply that  $sct = ct$ .

#### 1.4.2 Dose volume histogram

The final interest of **sCT** generation methods is to produce an accurate dose distribution in regard to the one computed on the real **CT**. It is therefore essential to study the dosimetric impact of **sCT** generation methods.

One tool commonly used in radiotherapy for plan evaluation or comparing different plans is the **DVHs**. It is a decreasing cumulative histogram counting the number of voxels that received at least a given dose level. **DVHs** are usually given for different **ROIs** (tumour or organ at risk). Because it is more convenient to convert the number of voxels into a volume, the volume of each **ROIs** receiving each dose level is known. The axes are displayed either in absolute (cubic centimeters and Gray) or in relative (% volume and % max or prescribed dose).

To compare the doses computed on the **CT** and the **sCT**, one may uses some **DVHs** points. They are expressed as  $Dx$ , where  $x$  represents a percentage volume. The value of  $Dx$  is therefore the minimum dose that  $x$  receives or equivalently the pseudo-inverse of the DVH for a given volume. The **DVHs** parameters studied through the different studies are the following: D98, D50 and D2. Once again, this indicators discards all spatial information. The comparison is done by computing the relative difference on these **DVHs** parameters

$$\Delta Dx_{\%} = 100 \left( \frac{Dx_{ct} - Dx_{sct}}{D_p} \right) \quad (1.4.2)$$

where  $D_p$  is the dose prescribed.

#### 1.4.3 Voxelwise dose difference

The difference between two dose distributions is a qualitative analysis of the spatial difference. It is a visual tool which is the most direct and informative. However, there is a huge amount of information to process as we are dealing with 3D images. The relative dose difference is simply given by

$$\Delta D_{\%} = \frac{100}{N} \sum_{i=0}^N \frac{D_{ct}(i) - D_{sct}(i)}{D_p} \approx \frac{1}{Gr} \sum_x \Delta Dx_{\%} \quad (1.4.3)$$

where  $D_{ct}$  and  $D_{sct}$  are respectively the dose distribution computed on the **CT** and **sCT**,  $N$  is the total number of voxels,  $i$  the voxel index and  $Gr$  the number of dose levels considered in the body. Similarly as the ME, this measure allows error compensations therefore for a qualitative assessment,  $D_{ct} - D_{sct}$  may be displayed as an image thereby revealing which regions received an overdose / underdose depending the sign of the difference.

#### 1.4.4 Gamma dose distribution evaluation tool

The gamma index (Low, 2010) is used to quantitatively compare two dose distributions, in our case  $D_{ct}$  and  $D_{sct}$ . For each pair of voxel



positions  $(\vec{v}_{ct}, \vec{v}_{sct})$  inside the **sCT** and **CT** respectively, the gamma index is given by

$$\Gamma(\vec{v}_{ct}, \vec{v}_{sct}) = \sqrt{\frac{\|\vec{v}_{ct} - \vec{v}_{sct}\|_2^2}{\Delta d^2} + \frac{[D_{ct}(\vec{v}_{ct}) - D_{sct}(\vec{v}_{sct})]^2}{\Delta D^2}} \quad (1.4.4)$$

where  $D_{ct}(\vec{v}_{ct})$  and  $D_{sct}(\vec{v}_{sct})$  are respectively the doses at this position on the **CT** and the **sCT**,  $\Delta d$  and  $\Delta D$  are two operator-imposed criterion representing the distance to agreement (3, 2 or 1 mm) and the accepted dose difference (3, 2 or 1%). The minimum value of  $\Gamma$  is the value of the  $\gamma$  index at this position

$$\gamma(\vec{v}_{ct}) = \min_{\vec{v}_{sct} \in \mathcal{S}} \Gamma(\vec{v}_{ct}, \vec{v}_{sct}) \quad (1.4.5)$$

where  $\mathcal{S} = \{\vec{v} : \|\vec{v} - \vec{v}_{ct}\|_2 \leq \Delta d\}$  is a sphere of center  $\vec{v}_{ct}$  and radius  $\Delta d$ . The proportion of voxels such that  $\gamma \leq 1$  is called the passing rate. An illustration is given in Figure 1.11.

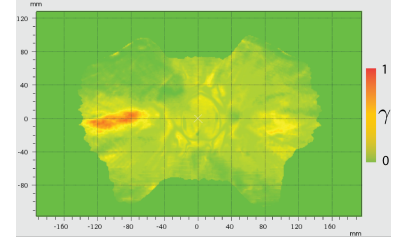


Figure 1.11: Gamma distribution between two dose distributions. Each voxel represents the  $\gamma$  index.

## 1.5 DEEP LEARNING IN COMPUTER VISION

Computer vision is about software that can interpret images. One important aspect to consider in computer vision are features. This term refers to interesting areas/patterns on an image that characterize it. Features can correspond to contours, points, colors that can constitute “a lower dimensional representation” of image content. A relevant feature must be unique enough to be able to differentiate two different classes of images, and generic enough to be able to easily recognize images of the same class despite the variability of instances of this class.

Not so long ago, depending on the given problem, computer vision scientists used to craft pre-processing steps of learning algorithms which must rely on carefully chose features in order to solve a task such as automatic classification or segmentation. Feature engineering is a tedious job that in some cases jeopardizes the sought level of autonomy of machine intelligence. Although these contributions required hard work and could achieve remarkable results on dedicated applications, they were time consuming and required constant adaptation whenever the task evolves (even moderately) which, to some extent, is reminiscent of the Mechanical Turk, see Figure 1.12.

Progress has been made since, but the biggest breakthrough came in 2012 at the annual ILSVRC<sup>12</sup> computer vision competition when a **deep learning** (DL) algorithm based on a **convolutional neural networks** (CNN) architecture called AlexNet killed the competition, by outperforming state-of-the-art approaches with a never-seen-before margin. The key aspect of this breakthrough comes from the ability of deep learning to find itself the best suited visual features to solve the learning task.

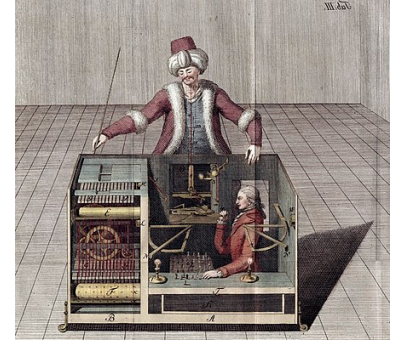


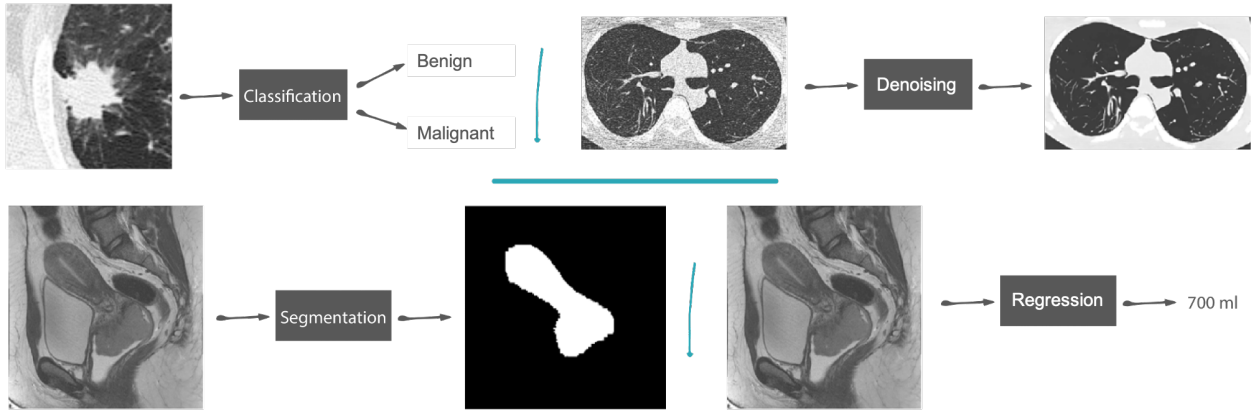
Figure 1.12: The Mechanical Turk was a famous hoax built at the end of the 18th century: it was an alleged automaton with the ability to play chess, it was actually a man hidden inside.

<sup>12</sup>Annual challenge to both promote improvement in computer vision and benchmark it.

### 1.5.1 The mechanics of machine learning

As humans, we certainly do not have to count the number of legs or measure the shape of an ear to learn how to recognize a cat. We are able to do so by seeing several examples and being corrected when we make the wrong guess. In other words, we do not improve our vision with a large list of formulas but with examples. As we grow up, our brain experiences several kinds of stimulations in which its connections are reinforced or changed in order to assimilate new information.

This learning process has been mimicked by computer vision scientists with **machine learning (ML)** in which the classical paradigm of AI has changed. **ML** allows to efficiently perform complex tasks such as classification, segmentation or regression (see Figure 1.13) by learning (instead of following) a set of rules.



**DL** is a sub-category of compositional **ML** models, i.e. such models are organized in layers which are stacked together to create a more general mapping from inputs to predicted outputs. Most **DL** models are based on neural nets and we will focus on these latter in this manuscript. Basically, in **DL**, we shift the burden of feature extraction from the programmer to the machine. First of all, let us start with a short review of shallow neural networks, starting with the perceptron.

Figure 1.13: Example of task performed by **ML** in radiotherapy.

### 1.5.2 Perceptron

The perceptron was designed to mathematically mimic the processing of information by biological neural networks. Perceptrons are to a computer what biological neural networks are to the cortex of mammals.

The building brick of the perceptron is a neural unit (Figure 1.14) which simple model allowing to perform a linear combination followed by a threshold operation usually embodied by the sign function. If a neural unit belongs to the first layer, it computes a linear combination of the input plus some intercept/bias constant. For some  $d$ -dimensional

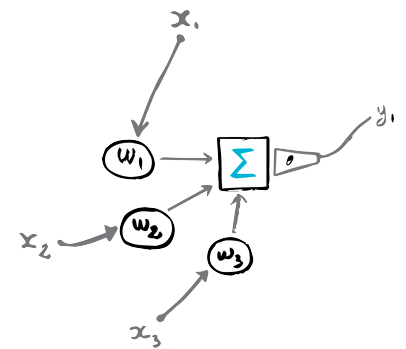


Figure 1.14: An artificial neuron with 3 inputs. Its approximation capacity is therefore necessarily limited.

input  $x = (x_i)_{0 \leq i \leq d}$ , the neural unit output is

$$\hat{y} = \text{sign}(w_0 + \sum_{i=1}^d w_i x_i), \quad (1.5.1)$$

where the weights  $w = (w_i)_{0 \leq i \leq d}$  are learnable parameters. If the network contains only one layer which itself contains one neuron, then  $\hat{y}$  is the prediction of the perceptron. Because the sign function binarizes the linear combination, such models were initially used to solve binary classification tasks. Also, for this model to work, it is necessary to adjust the weights to appropriate values. Under linear separability assumption, a simple (and provably converging) learning rule can be used<sup>13</sup>.

Obviously, this core neural unit model is not very flexible and more general versions of it are now preferred. In particular, the sign function is usually replaced with a differentiable activation function<sup>14</sup>. Not only does this allow to tackle more general problems than binary classification, but it also make it possible to use more subtle learning rules whose success is not tied to the linear separability condition. The most powerful such learning algorithm relies on the definition of a loss function and the computation of its gradients w.r.t. learning parameters, hence the need for differentiability.

More formally, for a training set  $\mathcal{D} = (x^{(j)}, y^{(j)})_{0 \leq j < N}$  containing  $N$  observations, the training loss is defined as

$$L(w) = \sum_{j=1}^N \ell(y^{(j)}, f_w(x^{(j)})), \quad (1.5.2)$$

where  $f_w$  is the neural network and  $\ell$  is a loss function that provides feedback on the ability of the network to produce a relevant prediction  $\hat{y}^{(j)} = f_w(X^{(j)})$  compared to the true value  $y^{(j)}$ . In a classification task, the network output  $\hat{y}^{(j)}$  is a vector of class membership probabilities. If there are  $K$  possible class labels, the most usual loss is cross-entropy which writes

$$\ell(y^{(j)}, \hat{y}^{(j)}) = - \sum_{k=1}^K 1_{y=k} \log \hat{y}_k^{(j)}. \quad (1.5.3)$$

The predicted probability for label  $k$  is  $\hat{y}_k^{(j)}$  so if the true label is indeed  $k$ , the cross-entropy will penalize values of  $\hat{y}_k^{(j)}$  that are small.

In a regression task, the usual loss is the quadratic loss, in which case the training loss is often referred to as mean square error (MSE). In this case, we have

$$\ell(y^{(j)}, \hat{y}^{(j)}) = (y - \hat{y}^{(j)})^2. \quad (1.5.4)$$

This loss obviously penalizes large discrepancies between prediction and true value.

Now the learning problem of the weights has been cast into an optimization problem and the training loss is minimized using [stochastic gradient descent](#)<sup>15</sup> (SGD). The opposite direction of the gradient is

<sup>13</sup>As for biological neurons, the weights of a perceptron are not found in one try. They are determined using an iterative rule. This rule consists in adding to the weights  $x \times y$  where  $y \in \{-1; +1\}$  is the true class label of input  $x$ . This rule must be repeated for each  $(x, y)$  pair in the training set. This will guarantee that the sign of the dot product between any input and the weights agree with the true class label.

<sup>14</sup>identity for regression, sigmoid/softmax for classification.

<sup>15</sup>The gradient descent algorithm uses the gradient of  $L$  w.r.t. weights. However, when  $N$  is large and because differentiation and summation can be exchanged, it becomes advantageous to compute the gradient for only one point, i.e. the gradient of  $\ell(y^{(j)}, \hat{y}^{(j)})$ . This is what the stochastic version of gradient descent does. Sometimes, gradients are average on a small subset of training data called a mini-batch.

where the loss will be smaller, therefore we update the weights for each observations  $(X^{(i)}, y^{(i)})$  as

$$w_i := w_i - \eta \frac{\partial \ell(y^{(j)}, \hat{y}^{(j)})}{\partial w_i}, \forall i \in [0, d], \quad (1.5.5)$$

where  $\eta$  is an hyperparameter called the learning rate. More recently, advanced update rules are used which provide a form of gradient memory through update steps and will adapt the speed of convergence to the shape of the loss function. The most popular such is ADAM which was introduced by Kingma and Ba, 2014. Nonetheless, the learning rate remains the most important hyperparameter of the rule and deserves great care to be tuned.

When all examples in the training set have been used one time (iteratively) as part of the update rule, the algorithm has completed a so-called epoch. The general convergence of SGD is usually controlled either in number of epochs or through a more subtle procedure called early stopping. This latter procedure uses another set of data, called validation set, and check if the validation loss keeps decreasing. If not, the training is stopped. Most often, the algorithm is preferably stopped when the validation increased for several epochs in a row. The number of epochs in question is often called the patience hyperparameter. Another possible stopping criterion consists in checking if the gradient norm is below a given predefined value which usually indicates that SGD has reached a local minimum<sup>16</sup> of the loss function.

So far, we have mostly mentioned a basic situation in which a neural net is organized in a few neural units in parallel that maps the input to the output which might be a scalar or a vector. Sequential connections of neural units are usually preferred so that the network can pick a predictive function from a larger functional space and thus learn more complicated patterns. These multiple-layer alternatives are presented in the next subsection.

### 1.5.3 Multiple layer perceptron

A set of  $s$  neural units which map in parallel the same input quantity to an  $s$ -dimensional vector is called a layer. When multiple such layers are plugged sequentially, the corresponding network is called a **multi-layer perceptron** (MLP). The most widely used MLP architecture is the one-hidden-layer MLP. It comprises

*an input layer* which just an abstraction where each unit is simply meant to contain one of the entries of the input  $x$  to be processed,

*a hidden layer* which usually contains several neural units that process the input in different ways to trigger signals that reveal the presence of a pattern,

*an output layer* that maps the intermediate representation issued by the hidden layer to the desired type of output.

See Figure 1.15 for an illustration of one-hidden-layer MLP architecture.

The partial derivative is carried out by using the chain rule of differentiation for computational efficiency. It is thus important that the components of a neural net are (easily) differentiable functions. The chain rule requires to start the computation of the derivatives from the predicted output back to the first layer and the whole procedure is thus often referred as the backpropagation algorithm.

<sup>16</sup>Neural network training does not allow in general to reach the optimal values of the parameters. The obtained parameters after training are dependent on the initialization of these latter which is randomized. This is one notorious drawback of neural networks which may not systematically converge to a function with the same predictive power.

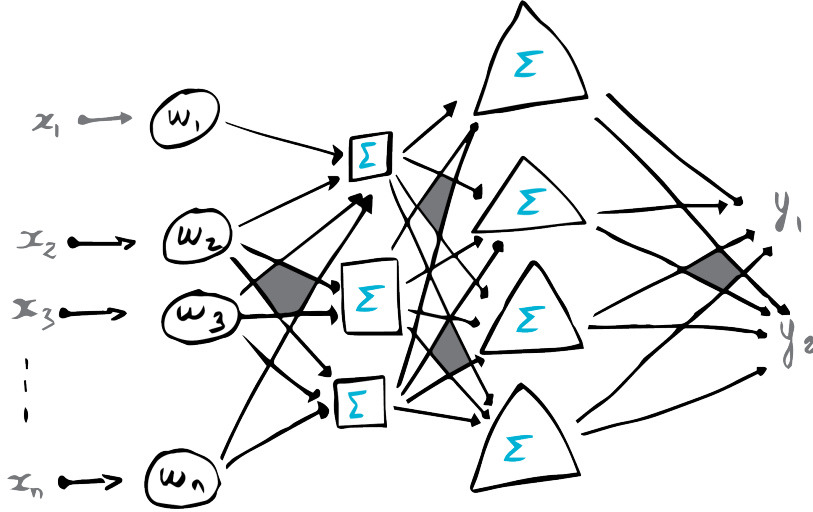


Figure 1.15: How to artificially learn ? For a given input sample  $x$  the network should predict an output  $y$ . To begin with, the weights are randomly initialized. Secondly, we compute the output of the network  $y$  and the error associated  $\epsilon$ . By the chain rule we compute the partial derivative of  $\epsilon$  with respect to each parameters  $w_k$ . We finish by updating the weights using the gradient descent algorithm. Repeat for each instance.

When one uses several units and layers, there multiple sets of weights to learn. For a notation simplification purpose, all these weights are concatenated in a single vector of parameter usually denoted by  $\theta$ . The training loss minimizing must thus be solved w.r.t.  $\theta$ .

Basically, the perceptron (as depicted in the previous subsection) is a linear model with an approximation capability necessarily limited. Minsky and Papert (1969) in their book have nipped in the bud all development by stating that practical applications was futile, dramatically decreasing the research by the late 60s even by connecting multiple perceptrons together<sup>17</sup>.

Surprisingly, Cybenko (1989) and Hornik (1991) have proven that an appropriate combination of perceptrons in the form of a one-hidden-layer MLP can achieve universal approximation and renewed the interest of the community for this class of models. The architecture of MLPs is more closely related to the functioning of the human brain to perform the task for which they were developed.

Stacking hidden layers (deeper net) allows to model arbitrarily complex functions. This is what gives deep neural net the predictive power that makes them so successful today. Although there is no rigorous definition specifying how many layers is “deep”, networks with only four or five hidden layers used to be considered very deep and almost impossible to train in practice as they were involving far too many trainable parameters.

The concept of convolutional layer and parameter sharing is a game changer that allowed the actual training of very deep networks and was the starting point of the success story of deep neural networks. The following paragraphs give a brief introduction to convolutional architectures. Note that there are many other neural net architectures than the one presented here, which allow to model particular types of data (images, sounds, temporal dependencies...), once again by mimicking the brain. See Zarándy et al. (2015) for more details.

<sup>17</sup>In a re-edition, they contrasted their words with the technical means of the time and the lack of basic theory.

#### 1.5.4 Convolutional neural net

**CNN** (Krizhevsky, Sutskever, and Hinton, 2012) are a class of neural nets that use convolutional layers. Unlike usual fully connected layers which are made of neural units with their own private weights, convolutional layers are made of units that share parameters, thereby drastically reducing the amount of parameters to be learned. **CNN** leverage parameter sharing in order to learn deep models, i.e. networks with a lot of layers. In contrast, deep fully connected networks cannot be trained because the numbers of parameters in them lead to overfitting<sup>18</sup>.

The key idea is that, unlike **MLP** architectures, a neural unit is not connected to each single entry of the layer input. Suppose this input is a multi-dimensional array (or tensor), then the unit will only process a small sub-array of consecutive cells. Consequently, while an **MLP** layer (also called dense layer in this context) has to learn as many parameters as cells in the array (+1 for the intercept), a convolutional layer will use only a number of parameters equal to the number of cells of the sub-array (+1 for the intercept). The convolutional layer then slides to the next sub-array. To some extent, this is like several neural units were processing each sub-array but were forced to use the same weights, hence the term parameter sharing.

Parameter sharing units in layer can be reshaped as a mathematical operation called convolution. Intuitively, convolutions are well suited to inputs that are signals or images since it has been known for long that convolutions can filter signals and, for instance disambiguate noise from informational content. Consequently, neural units are often called filters or kernels in this context. The output of the convolution operation is not a scalar in general. It is usually a tensor of the same size as the input one, also called feature map. The feature map is also mapped element-wise by a chosen activation function.

In order to reduce the dimensionality of intermediate representations issued by convolutional layers, one can use a pooling layer. This layer has no trainable parameters. The feature maps is divided into small (usually non-overlapping) sub-arrays and each sub-array is replaced by a scalar. If the scalar in question is the maximal value in the neighborhood defined by the small sub-array, then this is max-pooling. Another possibility is to use the mean of the sub-array (mean-pooling).

The typical architecture of a **CNN** consists in alternating between a few convolutional layers and pooling ones. When the dimensionality of a feature map is small enough, we stop using convolutions and we plug a small **MLP** to complete the network (see Figure 1.16).

#### 1.5.5 Generative adversarial net

Generative adversarial networks (**GAN**, Goodfellow, 2016) are models that are at the interplay between game theory of neural networks. In the game in question, there are two players (respectively called the generator and the discriminator) trying iteratively to adapt their

<sup>18</sup>When a learning model is very flexible, it is pretty easy to minimize the train loss. For example, the trained model  $f$  could be such that 
$$f(x) = \begin{cases} y & \text{if } (x, y) \in \mathcal{D} \\ 0 & \text{otherwise} \end{cases}.$$
 Although this function achieves a zero train loss, it has no chance to generalize to unseen data, i.e. a pair  $(x, y)$  that does not belong to the training set  $\mathcal{D}$ . This function also exhibits sharp variations which are usually indicating a form of overfitting. A workaround is thus to forbid the algorithm to learn such functions with strong variations.



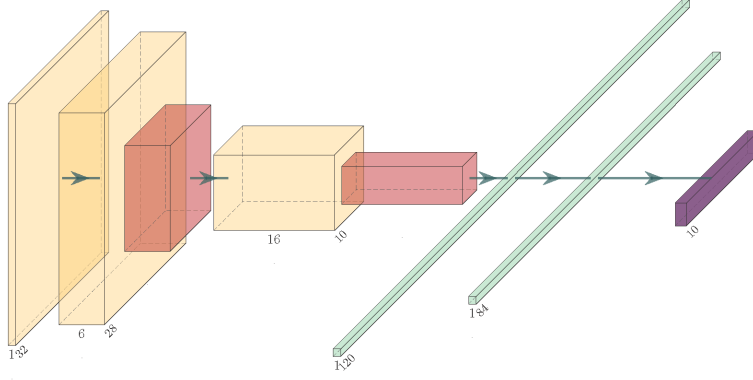


Figure 1.16: A CNN inspired by LeCun et al. (1989). Convolution in yellow, pooling in red, MLP in green and softmax output in purple.

strategies with respect to the opponent. Goodfellow (2016) uses the following metaphor

$G$  is a counterfeiter trying to produce undetectable fake currency.

$D$  is the police trying to discriminate true currencies from counterfeited ones.

In the context of this manuscript, both  $G$  and  $D$  are CNN and we are not interested in money but in CT images. The game consists in training  $D$  in an almost usual way, i.e. to maximize the probability of assigning the correct label to either true training examples or to counterfeited ones produced by  $G$ . Alleging that  $G = G^0$  is fixed, this amounts to solving

$$\max_{\theta_d} \sum_{i=1}^m \log D(ct^{(i)}) + \log \left( 1 - D \left( G^0(\epsilon^{(i)}) \right) \right) \quad (1.5.6)$$

where  $\theta_d$  is the vector of trainable parameters of the discriminator,  $m$  is the size of a mini-batch of images,  $ct^{(i)}$  is a true CT image and  $noise^{(i)}$  a noise vector that  $G$  maps to an sCT image. This maximization problem can be addressed by the usual backpropagation algorithm. Indeed, multiplying (1.5.6) by  $-1$  gives the usual cross-entropy loss. The obtained solution is not guaranteed to be the optimal solution of the problem but at least a “good” one if the training went alright. Before the first iteration, the parameters of  $G^0$  are initialized at random. At the first iteration ( $k = 1$ ), we thus obtain a discriminator  $D^1$  using the above maximization.

We can now move to the optimization of  $G$  assuming that  $D = D^1$  is fixed.  $G$  is the opponent of  $D$ , so it will try to minimize the function that  $D$  strove to maximize. We will now try to solve

$$\min_{\theta_g} \sum_{i=1}^m \log D^1(ct^{(i)}) + \log \left( 1 - D^1 \left( G(\epsilon^{(i)}) \right) \right) \quad (1.5.7)$$

where  $\theta_g$  is the vector of trainable parameters of the generator.

Again, backpropagation will optimize this function and we obtain an updated generator  $G^1$  at the end of the first iteration. We can go

on and iterate several times. From Goodfellow (2016, Proposition 2), this MinMax game has guaranteed convergence, although this proof relies on rather ideal assumptions. Note that, in practice, the noise generation is replaced with dropout<sup>19</sup>. Also, the discriminator should be optimized for several epochs before switching to one epoch of generator optimization.

<sup>19</sup>Dropout consists in turning off some neural units of the network at random during training. This forces the network to be able to provide good answers even if some information is “missing”.

### 1.5.6 Feature scaling or normalization

A good rule of thumb with CNN is that inputs and outputs should be scaled. It is only because the optimization is more stable if the inputs and outputs are centered and white. Mixing images that lie in  $[-1000, 300]$  with images that lie in  $[-1000, 2200]$  will usually result in failure.

Min-max scaling was chosen as normalization in upcoming studies. It restricts the range of values in the dataset between any arbitrary points  $a$  and  $b$  such as:

$$scale(X) = a + \frac{(X - X_{min})(b - a)}{X_{max} - X_{min}}. \quad (1.5.8)$$

This simple pre-processing will be used in both contributions that will be presented in Chapter 3 and Chapter 4 respectively. Concerning CT, we scale the features (images) from  $[-1024, 1500]$  to  $[-1, 1]$  following this function:

$$scale(ct) = \frac{2 \cdot (ct + 1024)}{2524} - 1. \quad (1.5.9)$$

Likewise, the MR images are scaled per site meaning that if the min and max intensity over all the patients of site  $k$  are  $min_k$  and  $max_k$ , then the scaling function from  $[min_k, max_k]$  to  $[-1, 1]$  for site  $k$  is:

$$scale_k(mr) = \frac{2 \cdot mr}{max_k - min_k} - 1. \quad (1.5.10)$$





# Synthesizing CT from MR images

## 2

Current TPS include both use of MRI and CT for dose calculation. CT is still needed for an accurate dose calculation, it gives necessary information on attenuation properties of the tissues which can be converted into mass densities. Contrary to MRI, CT offers a limited contrast between soft tissues. By synthesizing a CT from a MRI, it appears to be possible to introduce an MRI-only workflow.

Currently, there are a number of problems impeding the usage of MRI for treatment planning

- It is not always possible to scan the patient on MRI in treatment position. The radio frequency coils do not always allow the usage of fixation/immobilization equipment, such as masks. Recently, most MRI vendors have introduced systems dedicated to radiotherapy, allowing to position the patient in treatment position. Additional efforts are needed though. In brachytherapy, it is difficult to visualize the catheter.
- Some geometrical distortions can be caused by non-linearities in the magnetic gradients and the static field, by a chemical shift, or by magnetic susceptibility artifacts. These effects can be minimized by using actively shielded gradients or by measuring the distortions (Pasquier et al., 2006). A study on MR images of prostate patients (Kapanen and Tenhunen, 2013) highlighted that even for large scanning volumes (pelvis), geometrical errors concerning rectum wall and gold seed markers were below 2mm. In open MRI systems, the magnetic field is more homogeneous and distortions are less important (Devic, 2012). It can generally be concluded that the geometrical distortion is no longer a restrictive issue for a MRI-only workflow (Korhonen et al., 2014).
- Last but not least, there is no direct relation between MRI values and electron density, which is needed in current TPS for dose calculation.

The first two problems have been partially solved by the efforts provided by MRI manufacturers who all now provide solutions specific for radiotherapy. For the third problem, which is the focus of this manuscript, different solutions, based on conversion of MRI into sCT, have been proposed in literature. They are reviewed hereafter.

2.1 Bulk density method	38
2.2 Atlas-based method	38
2.3 Voxelwise conversion	39
2.4 Machine Learning	40
Patch-based	
Convolutional neural network	
Generative adversarial network	
2.5 Conclusion	42

## 2.1 BULK DENSITY METHOD

This method has been used for many years (Pasquier et al., 2007). The method is based on assigning homogeneous densities in delineated regions. This approximation can be compared to switching off inhomogeneity corrections in CT based dose calculations (Jonsson et al., 2010; Jonsson et al., 2013). In the last 20 years, dose calculation accuracy has continuously improved, evolving from correction based algorithms to Pencil Beam, convolution/superposition algorithms. Recently Monte Carlo calculations are being introduced within commercial TPS, increasing the accuracy even further.

Using bulk densities based on MR images seems to be a major step back concerning dose calculation accuracy. Obviously, the accuracy of the bulk density method will be improved by increasing the number of contoured regions. For the case of prostate treatment, using the conventional four beam set up (Pasquier et al., 2006) demonstrated that dose calculations without heterogeneity corrections lead to errors up to 4 % in the PTV for clinical cases. Automatic contouring of femoral heads and other pelvic structures, assigning bulk densities, leads to a much better agreement (within 1 %) with the original CT based plans. But this is practically not feasible for all localizations, even when using an automatic contouring tool, and the cases studied in literature were relatively conventional beams traversing simple geometries. Korsholm, Waring, and Edmund (2014) determined a criterion to check if for a certain treatment the MRI-only method using bulk densities would provide a high enough accuracy. This is not a practical solution though and is a direct consequence of the weakness of the usage of bulk densities. Stanescu et al. (2008) used an adaptive thresholding technique to correct image distortion, atlas-based<sup>1</sup> software to auto-segment relevant structures and bulk density assignment for treatment planning of intracranial lesions. They showed that MRI-based treatment planning for intracranial lesions was as precise as CT-based planning. Even recently this method is still being considered for sCT generation (Johnstone et al., 2018).

<sup>1</sup>explained below.

## 2.2 ATLAS-BASED METHOD

Another approach is to build an atlas including multiple pairs of CT and MR images of the same patient and to compute a deformable registration mapping an MRI in the atlas to the new acquired MRI. The chosen MRI in the atlas is usually the closest one in the atlas to the new one. The mapping is then applied to the CT associated to the selected MRI in the atlas. Various operations on the voxels can then be applied to the distorted CT atlas. The sCT is therefore a distorted CT included in the atlas. However, this method has a recognized risk of failure in case of unconventional patient anatomy. This method is explained in Figure 2.1.

According to Keereman et al. (2010), the main problem associated with the atlas-based method, is the need of anatomic reference data.

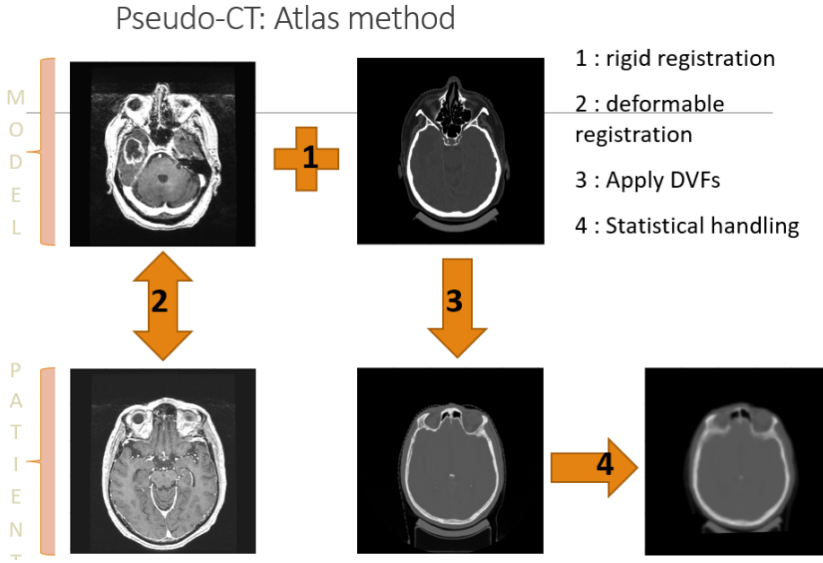


Figure 2.1: Atlas-based method using deformable image registration. The atlas consists of one or more pairs of aligned **CT** and **MRI** datasets (step 1). For a new patient, with only the **MRI** available, the **MRI** dataset is registered using deformable registration (step 2). The obtained deformable vector field (DVF) is applied to the corresponding atlas **CT** (step 3) providing a first estimate of the **sCT**. In step 4, different manipulations can be performed, ranging from simple denoising to more advanced correction methods. The atlas can consist of one single pair of **MRI** and **CT**, or of a larger number of patients: multi-atlas approach (Guerreiro et al., 2017).

Inter-patient differences are often too large to rely on these methods (Nyholm and Jonsson, 2014). This is why a (post-processing) correction is often applied. In a study performed by Dowling et al. (2012) two out of 39 patients needed to be excluded because of this. Consequently, increasing the number of **CT** cases that have been used to generate the atlas improves the performance of atlas based **sCT** generation (Uh et al., 2014) at the expense of computation time. This is also the method that was used in the Centre Oscar Lambret in the framework of several scientific projects (Boydev et al., 2017) as described in Figure 2.1.

### 2.3 VOXELWISE CONVERSION

Preferably, one would like to determine the tissue densities directly from the **MR** image contrast (Siversson et al., 2015; Su et al., 2015). The main problem is the low signal intensity of cortical bone because of the low hydrogen content<sup>2</sup>. Because of that, the transverse relaxation rate ( $R_2 = 1/T_2$ ) is very fast and the signal is lost before it is collected. A number of groups have introduced dual **ultrashort time echo (UTE)** **MRI** sequences to distinguish bone from air regions (Keereman et al., 2010; Johansson, Karlsson, and Nyholm, 2011; Grodzki, Jakob, and Heismann, 2012). These sequences are used to image tissues with short  $T_2$  relaxation time like the bones (Reichert et al., 2005). The bone signal in the first echo is relatively high while it has disappeared in the second. For other tissues both signal intensities are almost equal. So the difference between the two signals obtained in the two echoes will be maximal for bone and minimal for tissue and air (Catana et al., 2010). The signal intensity of the first echo is used to distinguish air from soft tissue (Keereman et al., 2010). The noise in the signal associated with air might complicate this technique, so one needs to filter the images and even use anatomical information to distinguish

<sup>2</sup>spin density is directly related to the signal intensity, see Equation 1.2.5

air from soft tissue.

These methods provide excellent attenuation maps for the **MRI** system, but the usage of three bulk densities (air, tissue, bone) is not adequate for treatment planning leading to deviations up to 9 % in organs at risk (Eilertsen et al., 2008). The combination of these **UTE** sequences with a spin density scan<sup>3</sup> should allow to finetune the tissue (composition and density) definition. For the moment the **UTE** sequences are only applied in the head but this is evolving (Nyholm and Jonsson, 2014). One of the problems might be that on **UTE** images, the blood vessels are dark (Hsu et al., 2013), so one still needs to combine with another sequence. Also, the **UTE** images used to characterize tissues need to be registered to the more conventional images ( $T_1/T_2$  e.g.) that are used for delineation, but this is an intra-modality registration (Nyholm and Jonsson, 2014).

Korhonen et al. (2014) used a conventional  $T_1$  Dixon sequence for direct conversion of intensity levels for prostate patients. This demands a delineation of femoral heads in a semi-automatic way though. The **MRI** is then calibrated as a function of **HU** using two different calibration curves (bone, soft tissue). This method ignores air cavities (or they would have to be contoured as well). The method was recently applied to proton dosimetry (Koivula, Wee, and Korhonen, 2016).

Johansson, Karlsson, and Nyholm (2011) used a  $T_2$  weighted 3D spin echo based sequence and two **DUTE** sequences and a trained algorithm to convert **MRI** into a **sCT**. They are able to convert on a voxel by voxel basis, not needing bulk densities. Their method was successfully applied to the head region, but severe artifacts were still obtained for pelvis and thorax regions (larger field of view). They also estimated the uncertainty on the calculated **HU** in the **sCT** images (Johansson et al., 2012), which allows estimating the error on the dose calculation on the **sCT** images. Gudur et al. (2014) used a combination of both **MRI** intensity and geometry information (atlas based method) to estimate electron density as both methods are not perfect. They claim that the conversion method using pixel intensity is not good enough, but this is because of the usage of a  $T_1$  image instead of using spin density weighted images.

<sup>3</sup>Sequences set that allow the measurement  $\rho$  in a voxel, see Equation 1.2.5

## 2.4 MACHINE LEARNING

As a general rule, **MRI**-only workflow would require an expert for the generation of an **sCT**. An **ML** approach can tackle this requirement by using an **ML** algorithm able to learn a direct end-to-end mapping from **MRI** to **CT**. It is important to note that it is difficult to compare the different results since a voxel-wise comparison between the **CT** and the **sCT** is not fully representative (because of misalignment) and not everyone can perform a dosimetric comparison.

### 2.4.1 Patch-based

One option is to divide the images into so-called patches that are matched. A patch is a (usually rather small) sub-region of an image (MRI). An intensity-based nearest neighbor search in the patch database (MRI-CT atlas) is done in order to find the corresponding CT patch. These methods are becoming more and more popular. Andreassen et al. (2015) and Andreassen, Van Leemput, and Edmund (2016) for example, used a  $K_D$  tree algorithm. The general principle consists of a comparison of local MRI intensity patterns with that of multiple MRI of the learning database. The sCT is generated using atlas CT values using coordinates that are determined using the best match between MRI images. This method provides dosimetric precision of 0.5%. This method demands important computing resources while being slow as the pattern comparison needs to be performed on multiple patches from several image datasets.

### 2.4.2 Convolutional neural network

From a training dataset consisting of a large set of CT-MR aligned image pairs, a direct mapping from MR to CT can be learned by training a CNN. At test time, this mapping is then applied to new MRI to generate an sCT. This class of model is the first to achieve an end-to-end sCT generation. The next paragraphs present recent CNN based approaches from the literature.

In Roy, Butman, and Pham (2017), a CNN (using patches) was developed based on dual-echo UTE images using Inception blocks (Szegedy et al., 2015). Bones and air are now taken in account. The use of a CNN and dual echo UTE sequence produces very good results, but there is no dosimetric comparison reported in this article concerning the brain.

Han (2017) has developed a deep CNN based on a U-Net (Ronneberger, Fischer, and Brox, 2015). U-net architectures are encoder-decoder type of architectures with additional shortcut-connections between symmetrically opposed layers. The proposed deep CNN method produced an MAE = 84 HU for all subjects tested, outperforming atlas based methods MAE = 94 HU.

The first use of 3D convolution (MAE = 41 HU), improving slightly the result compared to 2D (MAE = 38 HU), came with the study of Fu et al. (2018) in the pelvic area. However 3D convolutions requires to learn a lot more parameters than 2D convolutions which may increase the risk of overfitting or alternatively impose to use more massive datasets. It may also rise technical issues as this a memory demanding architecture which is a weakness of most GPUs.

A trade-off between dimensionality and informational content is proposed by Dinkla et al. (2018) who use a 2.5D CNN. They have built a model based on dilated convolution (Wolterink et al., 2016) allowing larger receptive fields while keeping the same number of parameters. This method has a MAE = 67 HU and a mean error dose of 0.05 %.

Instead of just an axial image for instance, the net is also fed with the two additional planes: coronal and sagittal.

The authors demonstrated the accuracy of radiation therapy dose calculations based on brain **sCT** generated by a dilated CNN. Recently, another study followed in the head and neck area (Dinkla et al., 2019) with a gamma pass rate  $\gamma_{2\%,2mm} = 95\%$ .

Without being fully exhaustive, the above list shows that it is possible (with a dataset of paired **MR** and **CT** images that underwent non-rigid registration) to generate an accurate **sCT** in approximately less than 2 minutes. These methods represent so far the best trade-off in the conversion of **MRI** into **CT**.

### 2.4.3 Generative adversarial network

By using a **GAN**, the **sCT** generation becomes an image-to-image translation problem. **GANs** have gained popularity in the past few years because they succeeded to generate accurate (new) samples in wide range of applications. Put most simply, they allow a network to learn to generate data with the same internal structure as other data, a procedure often referred to as style transfer (Gatys, Ecker, and Bethge, 2016).

Nie et al. (2017) were the first to use a **GAN** for **sCT** generation. A context model was developed in order to use the spatial information included in the **MRI** volume. It offered promising results:  $MAE = 39$  HU for the pelvic area and  $MAE = 92$  HU for the brain area. Wolterink et al. (2017) proposed an **sCT** synthesis using unpaired **CT-MR** images, it means that they do not use aligned pair of **CT-MR** images. Based on a Cycle**GAN** (Zhu, Park, et al., 2017), it seems to be robust to image misalignment. Their model outperforms paired model when facing region with air pockets. One major advantage is the only use of rigid registration in data preparation which is a new big step since the first use of **DL** in **sCT** generation.

As mentioned by Maspero et al. (2018), these two **GAN** approaches do not provide dosimetric evaluation. In their work, these other team of authors have trained a conditional **GAN** (Isola et al., 2016). They pre-processed **MR** image as input by labeling fat and water thanks to particular **MR** sequences called  $T_1$  Dixon. This sequence produces 4 different images contrasts (Fat, Water, In-phase and Out-of-Phase) in one-shot acquisition allowing a better segmentation. An image comparison gives an  $MAE = 65$  HU for the prostate,  $MAE = 56$  HU for the rectum and  $MAE = 59$  HU for the cervix. A dose comparison shows good results too, higher dose ( $D > 90\%$ ) about 0.1 to 0.3 % (dose difference between **CT** and **sCT**) for the prostate and in the worst case up to -1.6% for the cervix.

## 2.5 CONCLUSION

This chapter illustrates that **sCT** synthesis methods are numerous and of varying complexities. The frequent changes in procedure from one study to the other, such as the treatment technique, scanning parameters, the method of evaluation and the statistics of the number of cases

does not allow a global comparison of these different approaches.

It is difficult to draw conclusions about the dosimetric accuracy associated with each method, which is often a point discussed as additional information but does not constitute a study in its own right, especially in the atlas and intensity-based methods. In DL, dataset size matters too, the bigger the better. While some researchers claim that their method is better than others, the observed discrepancies may be mostly about dataset size and quality. The size and encoding of an image is also important, HU accuracy is (virtually) more easy to obtain on a smaller image with a narrow HU spectrum. Quantifying the differences between sCT and CT is an important analysis, but in the context of radiotherapy, the ultimate goal remains the quality of dosimetric accuracy.

The first step to a proper comparison is the use of a reference set of training and/or testing patients depending of the location. One must admit that this is not as simple as it seems to create an open-cohort when dealing with medical data. Efforts have been made in order to create datasets dedicated to research purposes, for instance the ADNI data<sup>4</sup> in the head and the Gold Atlas project in the pelvic area-part (Nyholm et al., 2018).

When facing clinical reality, some of the quoted references are not always feasible for many reasons. Some dedicated sequences depends on the scanner manufacturer or the version and if some license has been purchased. When possible, the acquisition time can also be a barrier. The guideline of our research is generalization of sCT synthesis regardless the MRI sequence chosen.

<sup>4</sup>Alzheimer's Disease Neuroimaging Initiative





# MR to CT synthesis with paired data

3


This chapter introduces one of the first sCT studies evaluating the performance of a GAN in a multicentric context with a paired MR-CT dataset<sup>1</sup>.

We focus on the issue of generalization in DL and data unavailability. Recall from the previous chapter that dedicated sequences for the convenience of sCT generation is not always possible in clinical routine. Moreover, the creation of a database is also a tedious job if one wishes to train its own DL model.

The presented method overcomes the limitation of a dedicated sequence by using standard sequences already in use in radiotherapy workflow. In addition, we demonstrated that it is also possible to generate an accurate sCT using MR sequence not included in the training phase. Our goal and contribution is a turnkey model usable by the greatest number thanks to its simplicity.

It is also beneficial for the CoBra project to the extent that an another partner is working on a brachytherapy optimized MRI sequence. As a robust model, we will no longer need to build a dataset for the chosen sequence.

The material in the next section has been published in Physics in Medicine & Biology journal

 K. N. B. Boni, J. Klein, L. Vanquin, A. Wagner, T. Lacornerie, D. Pasquier, and N. Reynaert. 2020. *MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network*. Physics in Medicine & Biology.

While some considerations and definitions are redundant with the introduction, a few notations are specific to this chapter.

## 3.1 INTRODUCTION

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

This problem is solved by converting an MRI to a so-called synthetic CT (sCT) or pseudo CT. Many different sCT generation methods have been proposed in the literature. These techniques recently underwent significant changes with the emergence of deep learning. Accuracy and

3.1	Introduction	45
3.2	A conditional GAN for MR-to-CT synthesis	46
	cGAN baseline	
	The pix2pixHD network	
	Tailored architecture for sCT generation	
3.3	Experimental material and implementation details	48
	Patient data collection	
	Image pre-processing	
	Training of the network	
	sCT evaluation	
3.4	Results	50
3.5	Discussion and Conclusion	51
	APPENDICES	
3.A	Network details	55

<sup>1</sup>Paired dataset refers to the need of deformable registration to obtain pairwise aligned MR and CT images of the same patient.

velocity have dramatically increased (Han, 2017; Dinkla et al., 2018). Generative Adversarial Networks (GAN) have boosted this trend with their ability to learn generating any data distribution in a paired (Nie et al., 2017; Maspero et al., 2018) or unpaired fashion (Wolterink et al., 2017).

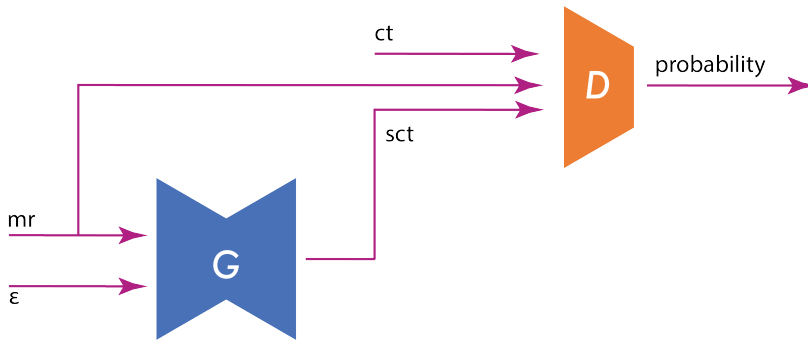
In this chapter, we discuss a new multi-scale approach by using an existing conditional GAN (cGAN, Wang et al., 2018) with paired data coming from different sites. This approach is among the first in the literature to use a deep learning-based method on data from different medical imaging centers using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

## 3.2 A CONDITIONAL GAN FOR MR-TO-CT SYNTHESIS

### 3.2.1 cGAN baseline

As explained in 1.5.5, GANs are characterized by two networks: the generator  $G(\epsilon)$  with  $\epsilon$  a noise vector and a discriminator  $D(y)$ . For the current application,  $y$  represents a CT image. All CT images are distributed according to an unknown probability distribution  $p_y$ .  $G$  attempts to transform the vector  $\epsilon$  into images so that a sample of size  $n$ ,  $G(\epsilon^{(1)}), \dots, G(\epsilon^{(n)})$  follows the probability distribution  $p_y$ .  $D$  attempts to separate the images actually distributed according to  $p_y$  from those produced by his opponent  $G$ . Actually,  $D(y)$  is understood as the probability that image  $y$  is a true CT<sup>2</sup>.

To convert an MRI into a CT, the networks have to be conditioned with an MR image  $x$ . A simple way to achieve this objective is to feed these two networks with  $x$  (as additional input). The generator and the discriminator therefore become  $G(x, \epsilon)$  and  $D(x, y)$  respectively (see Figure 3.1).



<sup>2</sup>This is simply a convention (without loss of generality) for the binary classification tasks pertaining to the discriminator. If the class label of a content  $y$  is denoted by  $c \in \{0, 1\}$  and  $c = 1$  means  $y$  is genuinely sampled from  $p_y$ , then  $D(y) = P(c = 1|y)$ . This must be kept in mind when it comes to write the GAN loss.  $D(y) = p(c = 1|y)$

Figure 3.1: Conditional generative adversarial net: the discriminator  $D$  learns to classify a real CT from a synthetic CT while the generator  $G$  learns to fool  $D$  following a min-max game.

As the training progresses,  $G$  must be able to generate samples that are more and more faithful to the distribution  $p_y$ , making it more and

more difficult for  $D$  to detect fakes CT images.  $G$  and  $D$  are trained alternately and share the same objective function. The discriminator tries to maximize it while the generator tries to minimize it. When both networks must play this game given an additional input  $x$ , the corresponding architecture is called a conditional GAN (cGAN)<sup>3</sup>.

The theoretical objective function of the cGAN is the following expected cross-entropy

$$\mathbb{E}_{x,y}[\log D(x, y)] + \mathbb{E}_{x,\epsilon}[\log(1 - D(x, G(x, \epsilon)))]. \quad (3.2.1)$$

However, as usual, we will work with an empirical version of the latter based on our paired training set  $\mathcal{D} = (ct^{(j)}, mr^{(j)})_{0 \leq j < N}$ :

$$\sum_j \log D(mr^{(j)}, ct^{(j)}) + \sum_j \log(1 - D(mr^{(j)}, G(mr^{(j)}, \epsilon^{(j)}))). \quad (3.2.2)$$

This network is optimized following the standard approach of Goodfellow (2016) by alternating the gradient ascent/descent steps between the generator and the discriminator. Noise instances  $\epsilon^{(j)}$  are induced by dropout (Hinton et al., 2012) in both the training and test phases.

### 3.2.2 The *pix2pixHD* network

In the architecture investigated in this chapter, we will use the cGAN introduced by Wang et al. (2018) whose motivation is to improve photorealism of the outputs when inputs are high resolution images. The traditional cGAN architecture is modified regarding four important aspects:

- Coarse-to-fine generator: the generator which has an encoder-decoder<sup>4</sup> architecture is separated in two sub-networks  $G = \{G_{global}, G_{local}\}$ . The first one is the center of an encoder-decoder architecture and is thus itself a (smaller) encoder-decoder. It is pre-trained on low resolution images. The local generator (the entire encoder-decoder structure) is then fine-tuned on high resolution images.
- Multi-scale discriminators:  $G$  has to fight against several discriminators  $D = \{D_1, D_2, D_3\}$ . Each of these discriminators works at a different image scale.
- A feature matching loss  $L_{FM}$  (Wang et al., 2018) is added in order to stabilize the training of the generator by promoting a match between intermediate representations (feature maps) in the different layers of the discriminators from real and synthesized images. The idea behind this additional loss term is that the generator will be forced to produce images with more natural statistics at different scales. If we denote  $D_k^{(i)}$  the  $i$ -th layers of  $D_k$ ,  $L_{FM}$  is then calculated as

$$L_{FM}(G, D_1, \dots, D_K) = \sum_{i,j,k} MAE(D_k^{(i)}(mr^{(j)}, ct^{(j)}), D_k^{(i)}(mr^{(j)}, G(mr^{(j)}, \epsilon^{(j)}))). \quad (3.2.3)$$

<sup>3</sup>The idea of cGANs was already mentioned in Goodfellow, 2016.

<sup>4</sup>A neural architecture, known as auto-encoder allows to learn a low dimensional representation of inputs. It comprises two sub-networks called encoder and decoder. The encoder  $Enc$  maps inputs to their low-dimensional versions and the decoder  $Dec$  then maps these representations back to the input space. In this case, the optimized loss is  $\|Dec(Enc(x)) - x\|$  (unsupervised paradigm). This architecture has proved to be beneficial in a wide range of applications.

- Instead of the usual cross-entropy cGAN loss (3.2.2), the authors recommend the Least Square GAN (LSGAN) loss (Mao et al., 2017) which is a quadratic version. This loss address the problem of vanishing gradient when updating the generator (Arjovsky, Chintala, and Bottou, 2017) using samples lying on the “True” decision boundary but still far from the real data distribution. LSGAN loss penalises these samples enabling faster convergence and more realistic image generation.

### 3.2.3 Tailored architecture for sCT generation

In our sCT generation implementation, it proved necessary to make several adjustments compared to pix2pixHD. First, pre-training the smaller resolution generator ( $G_{local}$ ) proved to be counterproductive and led to poorer results. The generator  $G$  used in our case follows instead the architecture proposed by Johnson, Alahi, and Fei-Fei (2016) and learns to synthesize a CT. We chose to work with  $K = 2$  discriminators working at different scales, both of them being trained to differentiate real and synthesized CT images. The first discriminator  $D_1$  operates at standard scale while the second  $D_2$  operates with downsampled images by a factor 2. These discriminators have identical architectures with different receptive fields. They follow the PatchGAN architecture (Isola et al., 2016) forcing the generator to produce consistent images while encouraging finer details. Training this model tends to produce realistic CT images but regarding HU, performances do not seem as good as they visually do. To overcome this difficulty without adding a post-processing step, we propose to add an additional  $L_1$  reconstruction loss (MAE) term between the generated sCT and the true CT. Finally, the full objective function  $L_{cGAN}$  writes:

$$L_{cGAN}(G, D_1, D_2) = \frac{\mu}{K} L_{FM} + \sum_{j,k} \left[ D_k \left( mr^{(j)}, ct^{(j)} \right) - 1 \right]^2 + \left[ D_k \left( mr^{(j)}, G \left( mr^{(j)}, \epsilon^{(j)} \right) \right) \right]^2 + \lambda \left\| ct^{(j)} - G \left( mr^{(j)}, \epsilon^{(j)} \right) \right\|_1 \quad (3.2.4)$$

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

## 3.3 EXPERIMENTAL MATERIAL AND IMPLEMENTATION DETAILS

### 3.3.1 Patient data collection

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

cancers that did not deform the pelvic anatomy and allowed realistic planning of prostate cancer radiotherapy.

Nineteen patients coming from three sites were selected and scanned in radiotherapy treatment position, T2-weighted MR and CT images were acquired following clinical protocol. Table 3.1 provides the acquisition settings.

	Site 1	Site 2	Site 3
Number of patients	8	7	4
CT			
Manufacturer	Siemens	Toshiba	Siemens
Model	Somatom Definition AS+	Aquilion	Emotion 6
Slice thickness (mm)	3	2	2.5
Kernel	B30f	FC17	B41s
T2-w			
Manufacturer	GE	Siemens	GE
Model	Discovery 750w 3T	- 1.5T	Signa PET/MR 3T
Sequence type	FRFSE	TSE	FRFSE
Slice thickness (mm)	2.5	2.5	2.5
Bandwidth (Hz/pixel)	390	200	390
Encoding direction	COL	ROW	COL
TR (ms)	6000–6600	12000–16000	6000–10000
TE (ms)	97	91–102	65

9 organs were segmented by five experts based on MRI, and consensus contours among the experts are also available. The open source library ITK (Johnson et al., 2015) was used to perform a deformable registration on the CT to fit the anatomy of the MRI, enabling the use of the delineations on the registered CT.

Table 3.1: Acquisition settings for the three sites. TSE stands for Turbo spin echo and FRFSE for Fast recovery fast spin-echo, COL for columns.

### 3.3.2 Image pre-processing

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

### 3.3.3 Training of the network

The 19 patients were separated into a training set containing 7 patients from site 2 and 4 patients from the third one. The 8 patients coming from the site 1 were used as testing set.  $256 \times 256$  sub-images were randomly cropped during training. The network was trained using

Adam optimizer with an initial learning rate of 0.0002 for 100 epochs, then for another 100 epochs with a linearly decay learning rate to zero.

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

### 3.3.4 *sCT evaluation*

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

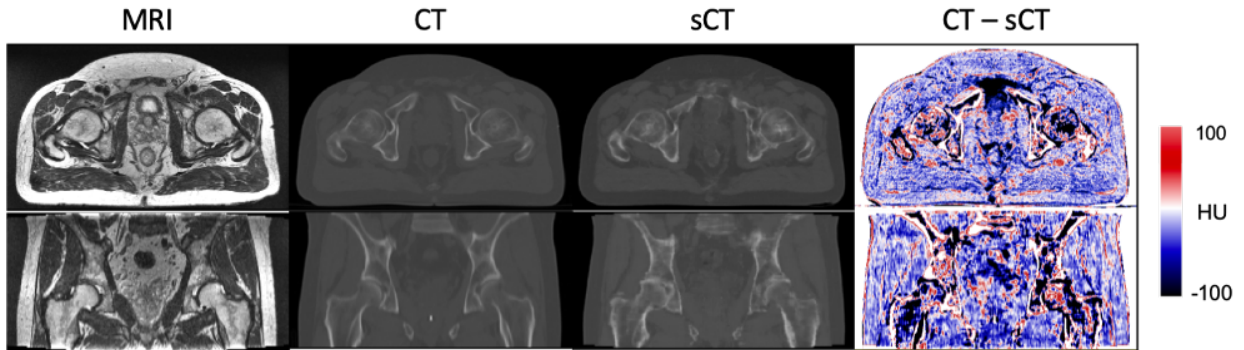
*Image comparison* Synthetic CT and registered CT were compared on a voxel-wise basis using the MAE and the Mean Error (ME). Considering the voxels within the body contours, MAE and in HU were calculated for each patient.

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

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

A dose volume histogram (DVH) analysis was performed after copying the structures (PTV, femoral heads, bladder wall and rectum wall) to CT. The chosen DVH points were D98, D50 and D2. Voxel-wise absolute dose differences in percentage were computed within a dose threshold of 90%, 50% and 10% of the prescribed dose  $D_p$ .

## 3.4 RESULTS



*Image comparison* CT synthesis took on average 7.5 s on GPU. Figure 3.2 shows an example of one of our test patients. As expected,

Figure 3.2: From left to right, MR image, CT, sCT and difference (CT – sCT). The images on top represent the axial plane, on the bottom, the frontal plane.

differences are most pronounced in the bone structures. Staircase patterns are visible on the bone in the frontal view. This may be due to the 2D generation technique used that does not take into account adjacent slices.

The proposed method produced an average MAE of  $48.5 \pm 6$  HU and an average ME of  $-18.3 \pm 9$  HU for our 8 patients. Vanilla pix2pix produced an average MAE of  $62.0 \pm 1$  HU and an average ME of  $-11.4 \pm 2$  HU. Table 3.2 provides the average MAE and ME for target volumes and organs at risk (OAR) for pix2pixHD and pix2pix.

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

Table 3.2: Average MAE and ME in HU ( $\pm\sigma$ ) between sCT and real CT for different locations when training with pix2pixHD based model and pix2pix.

*DVH analysis* The absolute difference between the DVH points on sCT and CT were always below 1.4%. Figure 3.3 shows a boxplot of the DVH point difference for the PTV and the OARs.

*Dose difference* Mean absolute dose differences were computed with several dose thresholds. Differences only appear in high dose regions and the body contour as shown on Figure 3.4.

The sCTs tend to have higher Hounsfield units (HU) resulting a global decreased dose inside the body. Inner negative dose differences are often due to lower HU on the sCT in bone area or air pocket not generated in sCT.

Table 3.3 reports the statistics in terms of mean dose difference related to the prescribed dose calculated on a threshold of 10%, 50% and 90% of the prescribed dose.

### 3.5 DISCUSSION AND CONCLUSION

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

The approach presented in this chapter is one of the very first allowing to derive a robust neural network trained and tested with data coming from different medical imaging centers. Without ever having seen an image from the test site, our model learns to synthesize a clinically acceptable sCT, which may be generalized to different MRI manufacturers. This process has the capability to tackle the images variability



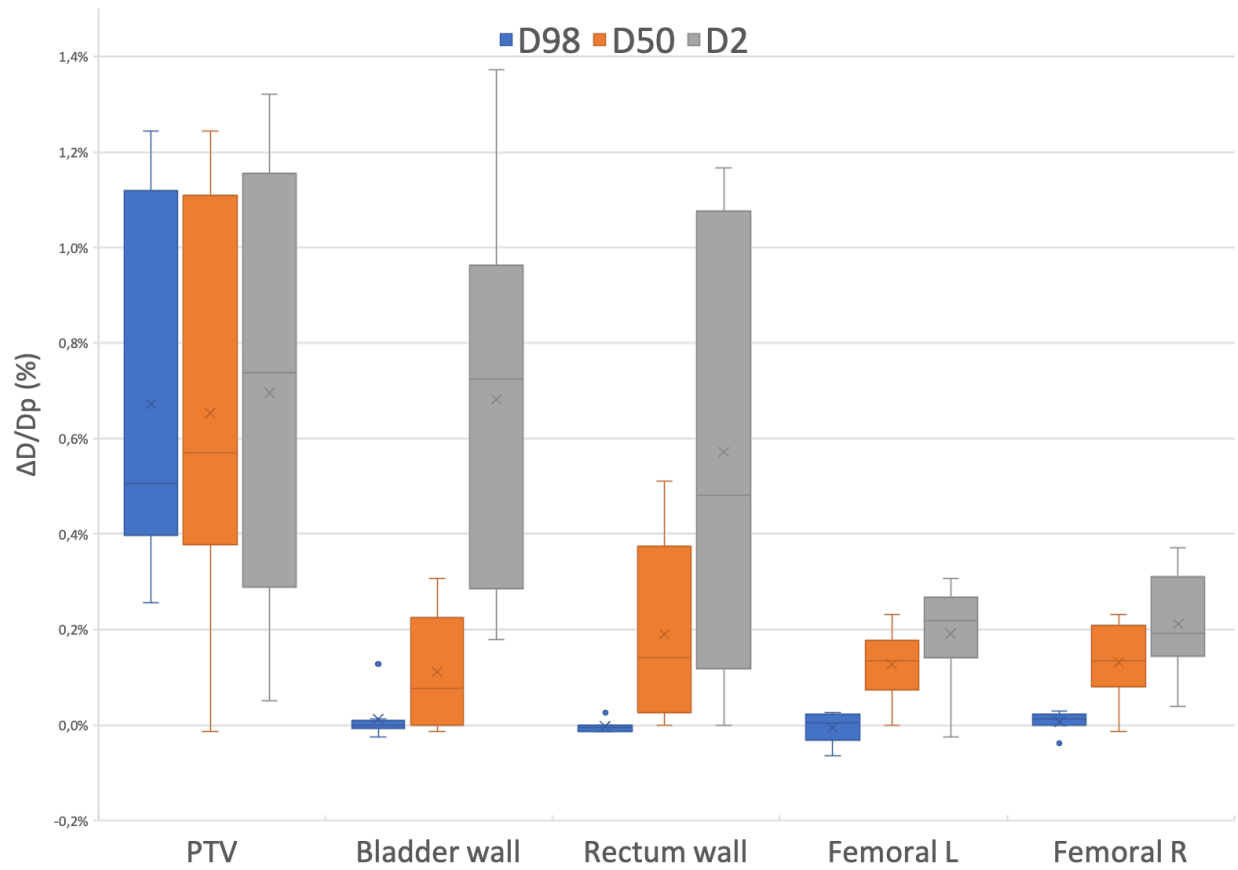


Figure 3.3: VH parameters comparison between dose on CT and sCT for the PTV and OARs.

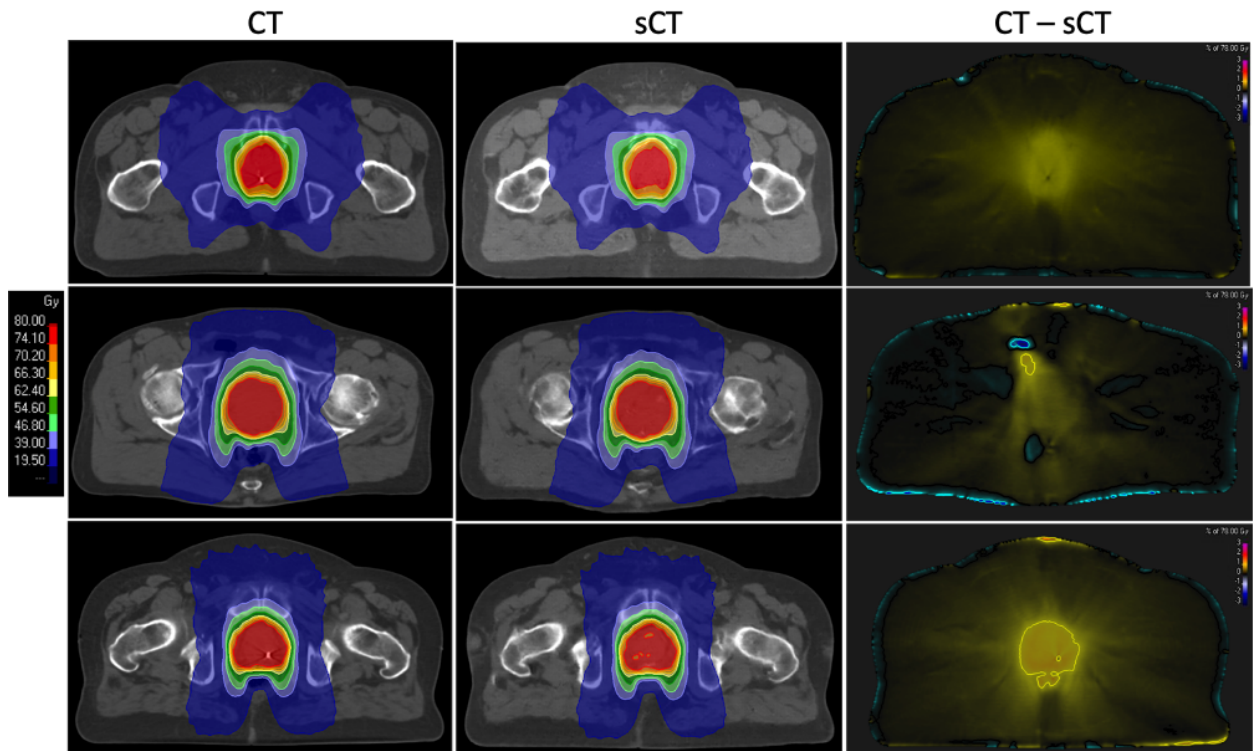


Figure 3.4: From left to right, dose calculated on CT, sCT and dose difference (CT - sCT).

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

Table 3.3: Mean dose difference ( $\pm\sigma$ ) between CT and sCT and range of values.

problem in clinical practices, since changes can happen in image acquisition parameters or with machine replacement for instance. This study was done using standard morphological sequence (T2-w Spin Echo) without the need of any dedicated sequences.

Results look promising although a presence of artifact patterns can be noted. This may be partially due to the low amount of data and to the transposed convolutions used in the decoder part in the generator. The use of a third discriminator seems to get rid of this problem without improving quantitative results. The average MAE (48.5 HU) and the dosimetric evaluation (dose differences within 1.4%) obtained in this study compare similarly with other state-of-the-art single center results (Maspero et al., 2018; Nie et al., 2017) in the literature for the pelvic area. These small differences would be suitable for clinical implementation. It is a well-known fact that deep learning models can benefit from more training data, which leads to the expectation that better results will be obtained when feeding our algorithm with more datasets. A direct comparison with other studies is not trivial since distinct datasets are used. The size of the dataset, the sequence(s) used, the diversity (artifact, specific case, etc.) and the misalignment between the sCT and the CT are some of the numerous factors that make a direct comparison difficult.

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

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

As a final word, an obvious limitation of the model studied in this chapter is the necessity of a paired dataset to train. This means that

one must possess a dataset of (CT,MR) pairs each of which was acquired from the same patient. Obviously, this is a drawback in terms of dataset size as the number of patients undergoing both exams is several order magnitude smaller than those performing only one. Also, many organs are deformable and might exhibit different shapes in the CT and MR exams. This issue coupled with registration requirement are somewhat inducing a noise in the training data which hinders the job of the cGAN. The next chapter will investigate another GAN-type architecture that can learn from unpaired data thereby circumventing the above mentioned issues.

## APPENDICES

### 3.A NETWORK DETAILS

The generator in pix2pixHD is basically a residual net (He et al., 2016a) with 9 residual block. The discriminators follow the same architecture as the PatchGAN (Isola et al., 2016). We follow the naming convention used in CycleGAN (Zhu, Park, et al., 2017) for their descriptions.

*Details of the generator architecture* Let c1-k denote a  $1 \times 1$  Convolution with k filters, c7ir-k denotes a  $7 \times 7$  Convolution-IN-ReLU layer with k filters, Ri-k denotes a IN residual block two  $3 \times 3$  convolutional layers with the same number of filters on both layers and ct3ir-128 denotes a  $3 \times 3$  Transposed Convolution-IN-ReLU layer with 128 filters.

Generator: c7ir-64, c7ir-128, c7ir-256, Ri-256, Ri-256, Ri-256, Ri-256, Ri-256, Ri-256, Ri-256, Ri-256, ct3ir-128, c3ir-64, c7-64, tanh.

*Details of the on discriminator architecture* Let c4-1 denote a  $4 \times 4$  Convolution with one filter and c4il-k denotes a  $4 \times 4$  Convolution-IN-LeakyReLU layer with k filters.

Discriminator: c4il-64, c4il-128, c4il-256, c4il-512, c4-1.




# MR to CT synthesis with unpaired data

## 4

We investigate in this chapter the sCT generation with an unpaired dataset<sup>1</sup>. We focus once again on generalization but the use of an unpaired dataset introduces a new challenge. State-of-the-art machine learning models for this situation are CycleGANs (Zhu, Park, et al., 2017)<sup>2</sup>. One major limitation of these models is that they only learn one-to-one mapping. Such a model assumes that a CT can only be associated (through a bijective mapping) with a single MRI sequence which is not correct.

Unpaired datasets have the advantage of being simple to create. They can be used with an unpaired-data-friendly model directly in theory without any form of registration between the MR and CT, even if in practise we chose to perform rigid registration in order to make the training easier. Moreover, models trained from unpaired data do not suffer from misalignment like paired model in which a slight voxel-wise difference can negatively impact the training of the model. However, unpaired models need more data compared to models trained from paired data since they basically use twice as many parameters. This can be balanced with the ease of fueling this unpaired database with patient not necessarily scanned with both modalities.

The material in the next section has been published in Medical Physics journal

 K. N. Brou Boni, J. Klein, A. Gulyban, N. Reynaert, and D. Pasquier. 2021. *Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data*. Medical Physics.

While some considerations and definitions are redundant with the introduction or other sections of the manuscript, a few notations are specific to this chapter.

## 4.1 INTRODUCTION

Since its introduction in the late 80s, magnetic resonance imaging (MRI) is being increasingly used in radiotherapy. MRI has been particularly appreciated for its superior soft-tissue contrast over computed tomography (CT). As a result, MRI can better discern tumors than CT images, and has therefore been integrated into the radiotherapy treatment workflow. In practice, after MR acquisition, regions of interest (ROIs) are contoured directly on the image by a radiation oncologist. However, CT is still required since it serves as the basis for dose calculation because of the need for electron-density information. The contours drawn on an MR image are propagated to a CT image

4.1	Introduction	57
4.2	From cycle GANs to an augmented cycle GAN for MR-to-CT synthesis	59
	Unsupervised learning of one-to-many	
4.3	Experimental material and implementation details	62
	Patient data collection	
	Image pre-processing	
	Training of the networks	
	sCT evaluation	
4.4	Results	65
	Image comparison	
	DVH analysis	
	Dose difference	
	CycleGAN comparison	
4.5	Discussion	67
4.6	Conclusion	70
	APPENDICES	
4.A	Fiducial markers	71
4.B	Network details	72

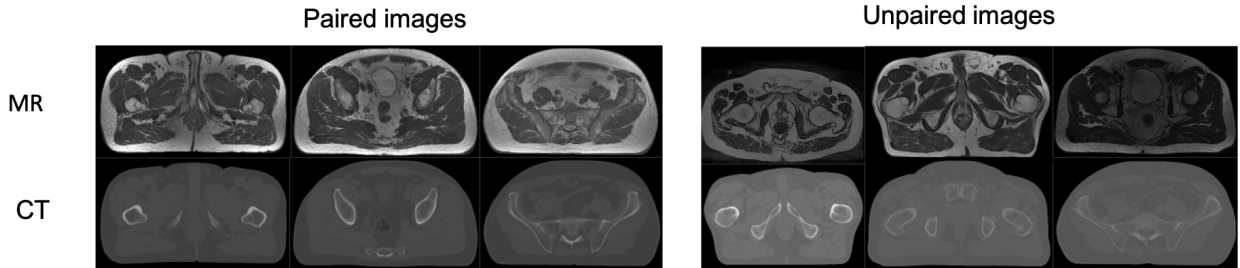
<sup>1</sup>Unpaired dataset refers to the need of only rigid registration.

<sup>2</sup>This other variant of GANs are actually relying on two GANs. One GAN would typically learn to map MRIs to CTs while the other one will learn the inverse mapping from CTs to MRIs. For any image  $x$  of an unpaired dataset, one can leverage a pipeline of two generators to produce another image  $\hat{x}$  that should be very close to  $x$ . A so called cycle-consistency loss  $\sum_x \|x - \hat{x}\|$  can be added to the objection function. Further explanations on CycleGANs are given in the next section.

after co-registration. In the professional community, this step is considered as the weakest link in the radiotherapy workflow (Njeh, 2008). The image-registration process introduces spatial uncertainties whose significance depends on the localization.

To reduce these systematic spatial uncertainties introduced by inter-scan differences, interest has been rapidly growing in MRI-only radiotherapy (Nyholm and Jonsson, 2014). However, the main challenge associated with this modality is that MRI intensities cannot be directly used to obtain information regarding electron density, which is required for dose calculation in radiation therapy treatment planning systems (TPS).

This problem can be partially solved by converting MRI into a so-called synthetic CT (sCT) which is the focus of this manuscript. As explained in the previous chapters, several methods for this purpose have recently emerged with the advent of deep learning, especially methods using Generative Adversarial Networks (GANs, Mao et al., 2017; Isola et al., 2016). These networks rely on learning the relationship between MRI and CT images in a paired (Maspero et al., 2018; Boni et al., 2020) or unpaired fashion (Wolterink et al., 2017; Hiasa et al., 2018) (see Figure 4.1).



As discussed in Chapter 3, learning in a paired fashion requires perfect registration of a set of MR and CT images, a time-consuming task that often involves deformable registration and consequently reintroduces geometrical uncertainty (Thor et al., 2013; Nyholm et al., 2009). Due to their adaptability/flexibility, unpaired models using at most rigid registration and based on CycleGAN (Zhu, Park, et al., 2017) have attracted more interest and offer the potential to improve sCT generation accuracy by using appropriate models such as a 3D model (Lei, Harms, et al., 2019) or a deeper model (Lei, Wang, et al., 2019). Nevertheless, these models are not robust to the MR variability arising in multicentric workflows, highlighting the common generalization issues in machine learning when dealing with out-of-distribution samples. Since a CycleGAN only learns one-to-one mappings (Almahairi et al., 2018) i.e., the model associates each CT with a single MR sequence, mapping several MRI instances to the same CT image represents a bottleneck that will become a major limitation in the clinical implementation of this approach (Fu et al., 2018).

In this chapter, we aim to improve the generalizability of MR-to-CT

Figure 4.1: In the paired fashion, the input MR and ground truth CT slices correspond to the same patient at the same anatomical localization. In contrast, models based on the unpaired fashion use CT and MR slices of different patients at different anatomical localizations during training.

synthesis with unpaired data. We devised an Augmented CycleGAN (AugCycGAN, Almahairi et al., 2018) that allowed the use of multi-centric data in both training and testing phases. We evaluated the accuracy of sCT images obtained with this approach by performing both image and dosimetric comparisons in order to show that our model can learn from several MRI distributions.

## 4.2 FROM CYCLE GANS TO AN AUGMENTED CYCLE GAN FOR MR-TO-CT SYNTHESIS

### 4.2.1 Unsupervised learning of one-to-many

*Unpaired learning* Given two distributions  $p_{mr}(mr)$  from MRI and  $p_{ct}(ct)$  from CT, we assume there exists a mapping between these two domains only if they are highly dependent, with the same anatomical localization for instance. To recover this mapping, Zhu, Park, et al. (2017) used two GANs with cycle consistency. In our MR-to-CT context, we can define the first GAN as  $\{G_{mr2ct} : MRI \mapsto CT, D_{ct} : CT \mapsto \{0, 1\}\}$  and the second as  $\{G_{ct2mr} : CT \mapsto MRI, D_{mr} : MRI \mapsto \{0, 1\}\}$ . These networks follow the standard optimization described by Goodfellow (2016) with an adversarial objective (marginal matching) formulated for the first CT-GAN as follows:

$$L_{GAN}^{mr2ct}(G_{mr2ct}, D_{ct}) = \sum_j \log D_{ct}(ct^{(j)}) + \sum_{j'} \log \left(1 - D_{ct}\left(G_{mr2ct}(mr^{(j')})\right)\right), \quad (4.2.1)$$

Similarly for the second MR-GAN  $\{G_{ct2mr}, D_{mr}\}$ , we have:

$$L_{GAN}^{ct2mr}(G_{ct2mr}, D_{mr}) = \sum_{j'} \log D_{mr}(mr^{(j')}) + \sum_j \log \left(1 - D_{mr}\left(G_{ct2mr}(ct^{(j)})\right)\right). \quad (4.2.2)$$

We use two different indexes  $j$  and  $j'$  in these cross-entropy loss terms to emphasize the fact that the sets of CT and MR images are not paired and would typically have different sizes. Observe that, compared to losses from Chapter 3, dropout noise instances  $\epsilon^{(j)}$  are no longer used. This is because one directly attempts to map  $p_{mr}$  to  $p_{ct}$  (or conversely) instead of mapping a noise distribution to  $p_{ct}$  having observed a sample from  $p_{mr}$ .

The goal here is that the output of each generator should match the distribution of the target domain. As usual, each discriminator is a binary classifier that discriminates genuine samples from generated ones. The main trick of cycle GANs allowing to learn from unpaired data is cycle consistency. Simply put, transferring a modality to the other and then back into it produces a reconstructed image close to the original. The cycle-consistency loss starting from an MRI slice is thus given by:

$$L_{cyc}^{mr}(G_{mr2ct}, G_{ct2mr}) = \sum_{j'} \left\| G_{ct2mr}\left(G_{mr2ct}(mr^{(j')})\right) - mr^{(j')} \right\|_1. \quad (4.2.3)$$



Similarly, for a CT slice, one has:

$$L_{cyc}^{ct}(G_{ct2mr}, G_{mr2ct}) = \sum_j \left\| G_{mr2ct} \left( G_{ct2mr} \left( ct^{(j)} \right) \right) - ct^{(j)} \right\|_1. \quad (4.2.4)$$

The global objective function of a CycleGAN is obtained as a weighted sum of these four loss terms (with appropriate positive weights). As explained in the next paragraph, this model cannot be applied in our context therefore we will leverage additional latent variables.

*Limitations of CycleGAN* In our medical image generation problem, the learning process of CycleGANs cannot be performed correctly because we need to learn a one-to-many mapping. Since the CycleGAN model generates deterministic mappings (Zhu, Zhang, et al., 2017), the generator  $G_{ct2mr}$  cannot be optimized properly when the MR outputs exhibit substantially differences due to (manufacturer proprietary solutions or sequence types); see Figure 4.2 for an illustration of these discrepancies in terms of voxel intensity distributions.

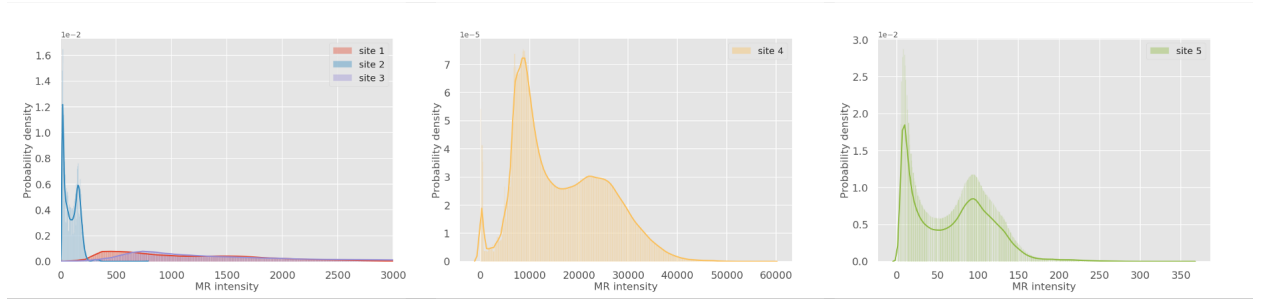


Figure 4.2: Distributions on MR intensity according different scanners. Zero values were not taken into account.

The main difficulty is that the cycle consistencies when plugged together amount to require that  $G_{mr2ct}(G_{ct2mr}(ct)) \approx ct'$ ; therefore, the generators must be inverse functions of one another and thus bijective:  $G_{mr2ct} = G_{ct2mr}^{-1}$ , which contradicts the fact that many MR images must be mapped to the same CT image. To overcome this modeling problem, we propose to resort to an “augmentation” as proposed by Almahairi et al. (2018) to capture any missing information that is necessary to generate / reconstruct an MR image from  $G_{ct2mr}$  of a given sub-category (e.g. T1 or T2 category).

*Augmented CycleGAN (AugCycGAN) in MR-to-CT* Instead of learning a mapping  $G_{ct2mr} : CT \mapsto MRI$  with a single input, this generator is augmented by including a latent space  $Z_{MRI}$  as an additional input space that captures missing information (i.e., “style” information characterizing the MRI distribution in terms of sequence type and manufacturer). The inverse mapping  $G_{mr2ct}$  causes the loss of this crucial information needed to synthesize or reconstruct the correct MRI style; thus, the latent space allows recovery of missing information when transforming an MRI into a CT. The CT-to-MR mapping becomes a stochastic mapping  $G_{ct2mr} : CT \times Z_{MRI} \mapsto MRI$  capable of correctly

optimizing the cycle consistency loss. The proposed model captures  $Z_{MRI}$  by using an encoder  $E_{mr} : CT \times MRI \mapsto Z_{MRI}$ . To summarize, the AugCycGAN designed for our purpose<sup>3</sup> is composed of five networks: two in the first GAN  $\{G_{mr2ct}, D_{ct}\}$ , two in the second GAN  $\{G_{ct2mr}, D_{mr}\}$ , and one encoder  $\{E_{mr}\}$ . A sixth (and final) denoted by  $D_{z_{mr}}$  network will be necessary to complete the architecture. as explained in the following paragraphs.

When resorting to latent variables, it often proves beneficial to chose a prior for these variables. Usually, this prior is chosen to make mathematical derivations simpler and/or to induce independence among the variables. To this end, Almahairi et al. (2018) use a standard Gaussian prior denoted by  $p_z$  over  $Z_{MRI}$ . They consequently adopt a marginal matching loss  $\mathcal{L}_{GAN}^{Z_{MRI}}$  for the latent space in order to encourage the encoder  $E_{mr}$  to produce realistic samples matching  $p_z = \mathcal{N}(0, \mathbf{I})$ :

$$L_{GAN}^{Z_{MRI}}(G_{mr2ct}, E_{mr}, D_{z_{mr}}) = \sum_{j''} \log D_{z_{mr}}(z^{(j'')}) + \sum_{j'} \log \left( 1 - D_{z_{mr}} \left( E_{mr} \left( G_{mr2ct} \left( mr^{(j')} \right), mr^{(j')} \right) \right) \right), \quad (4.2.5)$$

where the samples  $z^{(j'')}$  are drawn from  $p_z$ . This is where the sixth network is needed in order to discriminate samples genuinely sample from  $p_z$  from those produced by the encoder  $E_{mr}$ . Cycle consistency is also extended to the latent space:

$$L_{cyc}^{Z_{MRI}}(G_{ct2mr}, E_{mr}) = \sum_j \left\| z^{(j)} - E_{mr} \left( G_{ct2mr} \left( ct^{(j)}, z^{(j)} \right), ct^{(j)} \right) \right\|_1. \quad (4.2.6)$$

Again  $z^{(j)} \sim p_z$  and index  $j$  is used because we need as many latent vectors as CT images. This last cycle consistency loss force the encoder to extract a latent vector that is the same as the one needed by generator  $G_{ct2mr}$  to produce an MRI with appropriate style properties.

When training this vast architecture, it is simpler to summarize the optimization process with respect to input flows which respectively minimize different loss terms. Training in the forward direction CT-to-MR is done by minimizing the following terms:

$$L_{GAN}^{ct2mr}(G_{ct2mr}, D_{mr}) + L_{cyc}^{ct}(G_{ct2mr}, G_{mr2ct}) + \gamma L_{cyc}^{Z_{MRI}}(G_{ct2mr}, E_{mr}), \quad (4.2.7)$$

with  $\gamma = 0.025$ , training in the backward direction (MR-to-CT) is done by minimizing the following remaining terms:

$$L_{GAN}^{mr2ct}(G_{mr2ct}, D_{ct}) + L_{GAN}^{Z_{MRI}}(G_{mr2ct}, E_{mr}, D_{z_{mr}}) + L_{cyc}^{mr}(G_{mr2ct}, E_{mr}, G_{ct2mr}), \quad (4.2.8)$$

The AugCGAN is trained on both objectives simultaneously; Figure 4.3 summarizes both mappings.

For easier reproducibility, it should be emphasized that loss terms  $L_{GAN}^{ct2mr}$ ,  $L_{cyc}^{mr}$  and  $L_{cyc}^{ct}$  have evolved compared to equations (4.2.2), (4.2.3) and (4.2.4) because latent variables and the encoder are now at play. Their updated version are given below:

<sup>3</sup>Because  $G_{mr2ct}$  does not need to be one-to-many, there is no need for an extra latent variable for CT style therefore our model is simpler than the vanilla version introduced in Almahairi et al. (2018).

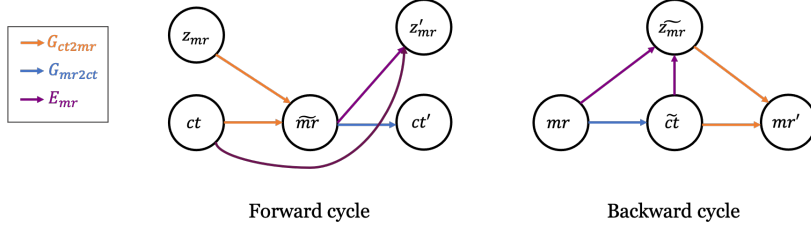


Figure 4.3: Augmented Cycle GAN consists of a forward and backward cycle in which a generator is augmented with a latent space to allow a better reconstructed MR image.

$$L_{GAN}^{ct2mr}(G_{ct2mr}, D_{mr}) = \sum_{j'} \log D_{mr}(mr^{(j')}) + \sum_{j,j'} \log \left( 1 - D_{mr} \left( G_{ct2mr} \left( ct^{(j)}, E_{mr} \left( mr^{(j')}, ct^{(j)} \right) \right) \right) \right), \quad (4.2.9)$$

$$L_{cyc}^{mr}(G_{mr2ct}, G_{ct2mr}) = \sum_{j,j'} \left\| G_{ct2mr} \left( G_{mr2ct} \left( mr^{(j')} \right), E_{mr} \left( mr^{(j')}, ct^{(j)} \right) \right) - mr^{(j')} \right\|_1, \quad (4.2.10)$$

$$L_{cyc}^{ct}(G_{ct2mr}, G_{mr2ct}) = \sum_{j,j'} \left\| G_{mr2ct} \left( G_{ct2mr} \left( ct^{(j)}, E_{mr} \left( mr^{(j')}, ct^{(j)} \right) \right) \right) - ct^{(j)} \right\|_1. \quad (4.2.11)$$

### 4.3 EXPERIMENTAL MATERIAL AND IMPLEMENTATION DETAILS

#### 4.3.1 Patient data collection

The experimental study carried out in this chapter included pelvic MR and CT images of 39 patients with prostate or rectal cancer obtained from 5 different sites. All scans were acquired in the radiotherapy treatment position. Half of the images were obtained from the public dataset named Gold Atlas project (GaP), 17 which aims to provide a source of training and validation images for segmentation as well as sCT generation. The remaining half were obtained from Institut Jules Bordet (IJB) and Centre Oscar Lambret (COL).

Table 4.1 provides the acquisition settings for both CT and MRI and site-wise splitting of the data for network training and evaluation.

Patients with hip implants were included in training and testing. Observe that different manufacturers and different parameters are used, thereby making this multicentric experimental setting inappropriate for cycle GAN without augmentation.

#### 4.3.2 Image pre-processing

To ensure a simple and smooth workflow, patient images underwent three pre-processing steps:

- Registration: CT images were rigidly registered and resampled to MR images. CT images were then cropped to match the field of view (FOV) of MRI. Registration for patients in site 5 was performed in Raystation (v 9A, RaySearch Laboratories, Sweden) by radiation oncologists, while the other patients were registered using Elastix (Klein et al., 2010).

	Training				Test	
	Site 1 (GaP)	Site 2 (GaP)	Site 3 (GaP)	Site 4 (LJB)	Site 5 (COL)	
Patients	8	7	4	8	8	
Slices	79–100	92–116	66–91	45–86	72–88	
CT						
Manufacturer	Siemens	Toshiba	Siemens	Toshiba	Siemens	
Model	Somatom Definition AS+	Aquilon	Emotion 6	Aquilon	Somatom Confidence	
Slice thickness (mm)	3	2	2.5	3	3	
Kernel	B30f	FC17	B41s	FC17	Br38f	
T2-w						
Manufacturer	GE	Siemens	GE	Siemens	Siemens	
Model	Discovery 750w 3T	- 1.5T	Signa PET/MR 3T	Magnetom Skyra 3T	Magnetom Sola 1.5T	
Sequence type	FRFSE	TSE	FRFSE	TSE	AX T2 SPACE	
Slice thickness (mm)	2.5	2.5	2.5	3	2	
Bandwidth (Hz/pixel)	390	200	390	255	651	
Encoding direction	Col	Row	Col	Row	Row	
TR (ms)	6000–6600	12000–16000	6000–10000	3500	1770	
TE (ms)	97	91–102	65	132	172	

Table 4.1: Acquisition settings for the five sites. TSE stands for turbo spin echo, FRFSE for fast recovery fast spin echo, and Col for columns.

- Exclusion of surrounding air/noise: A mask excluding surrounding air or outside structures was obtained on the CT and MR images by using the external ROI option (threshold level based) on Raystation. Voxels outside the body were automatically assigned to -1024 HU for CT and 0 for MR.
- Normalization: Limit values were arbitrarily set at 1500 HU and depending on the MR constructor 10000 or 15000. HU values were normalized, MR intensities as well site-wise following the min-max feature scaling.

#### 4.3.3 *Training of the networks*

We adapted the AugCycGAN implementation to make it suitable for radiotherapy image specifications. Modifications over the vanilla AugCycGAN have been made in order to use 16-bit grayscale images, modify input size, and use data augmentation. The 39 patients were separated into a training set containing 19 patients from site 1 to 3. The 19 patients from sites 4 and 5 were used for testing. The network was trained using RMSProp optimizer with an initial learning rate of 0.0002 for 150 epochs, then for another 150 epochs with a linearly decay learning rate to zero. Training took an average of 28h on an Nvidia Tesla V100 SXM2 (GPU) with a batchsize of 4. Original images were resampled to  $256 \times 256$ . Then,  $128 \times 128$  sub-images were randomly cropped during training. Data augmentation was performed by horizontal flip, increasing the size of the training set to 3328 image pairs.

#### 4.3.4 *sCT evaluation*

In the testing phase, the working image size was  $256 \times 256$  pixels. sCTs were generated using only the generator on the GPU (Nvidia Quadro P6000). The created image files were then converted into DICOM format by using MICE Toolkit (v1.1.3, NONPI Medical AB, Sweden).

*Image comparison* sCT and CT were compared on a voxel-wise basis by using the mean absolute error (MAE) and the mean error (ME). Considering the voxels within the body contours, MAE and ME in HU were calculated for each patient on MICE Toolkit. Detailed statistics by organ are only available for site 5, where delineated structures were copied and resampled from CT to the sCT.

*Dose comparison* Halcyon (VMAT) treatment plans were optimized on each CT image in Raystation by using the Collapsed Cone (v5.1) algorithm on a  $3 \times 3 \times 3 \text{ mm}^3$  grid. Optimized clinical plans were then recalculated on the sCT image for dose comparison. Only the 8 patients of site 5 underwent this procedure. A dose-volume histogram (DVH) analysis was performed after copying the structures (PTV, femoral heads, bladder, and rectum) to sCT. The chosen DVH points were

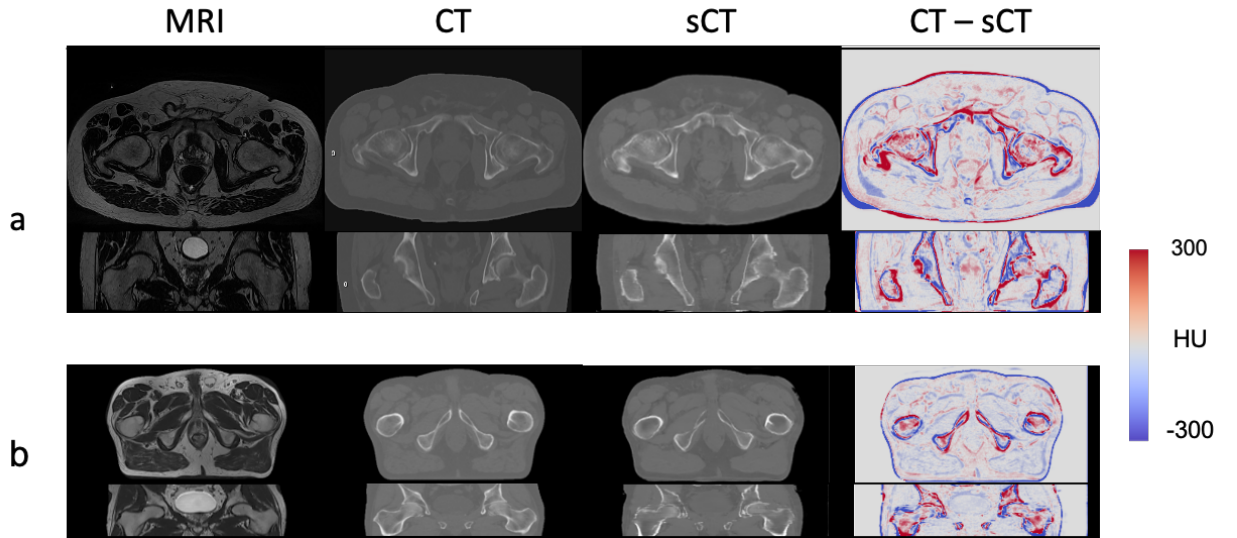
D98, D50 and D2. Voxel-wise absolute dose differences in percentage were computed within a dose threshold of 90%, 50%, and 10% of the prescribed dose  $D_p$ . Finally, two different gamma pass rates ( $\gamma_{3\%,3mm}$ ,  $\gamma_{2\%,2mm}$ ) of the dose distributions were calculated using four different lower dose thresholds (0%, < 10%, < 50%, and < 90%).

## 4.4 RESULTS

CT and MR scans were not necessarily acquired on the same day; the median interval between the scans was 2 days (range: 1h to 19 days). Interscan differences were therefore not taken into account in this study. CT synthesis took on average 8.5s on the GPU.

### 4.4.1 Image comparison

Figure 4.4 shows an example of two test patients from sites 4 and 5. Differences are most acute in the external geometry and bone structures.



The proposed method produced an average MAE of  $65.8 \pm 9$  HU and an average ME of  $-5.8 \pm 6$  HU for site 4. Patients from site 5 showed an average MAE of  $59.8 \pm 11$  HU and an average ME of  $-0.7 \pm 3$  HU, indicating low variability across sites. Table 4.2 provides the average MAE and ME by ROIs only for site 5.

	MAE	ME
Bladder	$65.3 \pm 29$	$-13.4 \pm 20$
Rectum	$60.3 \pm 25$	$-10.7 \pm 29$
Femoral Head L	$178.9 \pm 43$	$0.1 \pm 49$
Femoral Head R	$197.1 \pm 43$	$-1.0 \pm 64$

Figure 4.4: From left to right, MR, CT, and sCT images, and the difference (CT-sCT). The images on top represent the axial plane, while those on the bottom represent the frontal plane. (a) Site 5 and (b) site 4.

Table 4.2: Average MAE and ME in HU ( $\pm\sigma$ ) between sCT and real CT for different locations in site 5.

#### 4.4.2 DVH analysis

The differences between several DVH points on sCT and CT were computed and presented as a boxplot in Figure 4.5. The values are rescaled to the prescribed dose.

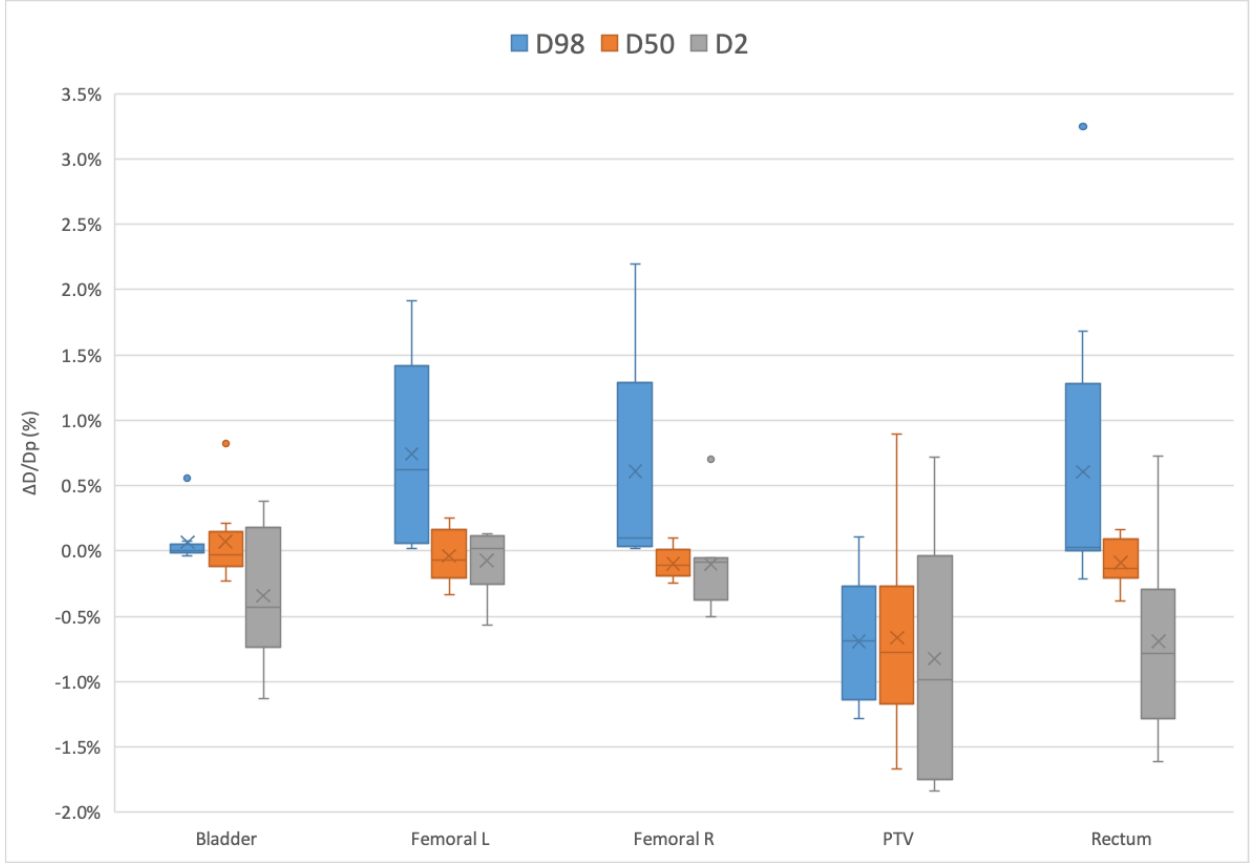


Figure 4.5: DVH parameters comparison between dose on CT and sCT for the PTV and OARs for site 5.

#### 4.4.3 Dose difference

The doses calculated on CT ( $D_{CT}$ ) and sCT ( $D_{sCT}$ ) along with their relative differences are presented in Figure 4.6 for three patients.

Mean absolute dose differences ( $\frac{|D_{CT}-D_{sCT}|}{D_p}$ ) were computed with several dose thresholds. Table 4.3 reports the statistics in terms of the mean dose difference related to the prescribed dose calculated on thresholds of 10%, 50%, and 90% of the prescribed dose and the passing rates of local gamma.

Volume	Dose difference (%)	$\gamma_{3\%,3mm}$ (%)	$\gamma_{2\%,2mm}$ (%)
Body	$0.13 \pm 0.07[0.05; 0.28]$	$99.5 \pm 0.4[9.88; 9.99]$	$95.5 \pm 2.2[92.5; 98.3]$
Dose > 10%	$0.20 \pm 0.07[0.08; 0.30]$	$99.5 \pm 0.5[98.7; 100]$	$94.8 \pm 2.3[90.7; 97.6]$
Dose > 50%	$0.48 \pm 0.24[0.24; 0.94]$	$99.9 \pm 0.2[99.3; 100]$	$97.6 \pm 1.3[95.6; 99.5]$
Dose > 90%	$0.70 \pm 0.38[0.28; 1.22]$	$99.8 \pm 0.4[99.8; 100]$	$98.8 \pm 0.4[98.8; 100]$

Table 4.3: Mean dose difference relative to the prescribed dose and gamma pass rate ( $\pm\sigma$ ) between CT and sCT and the range of values



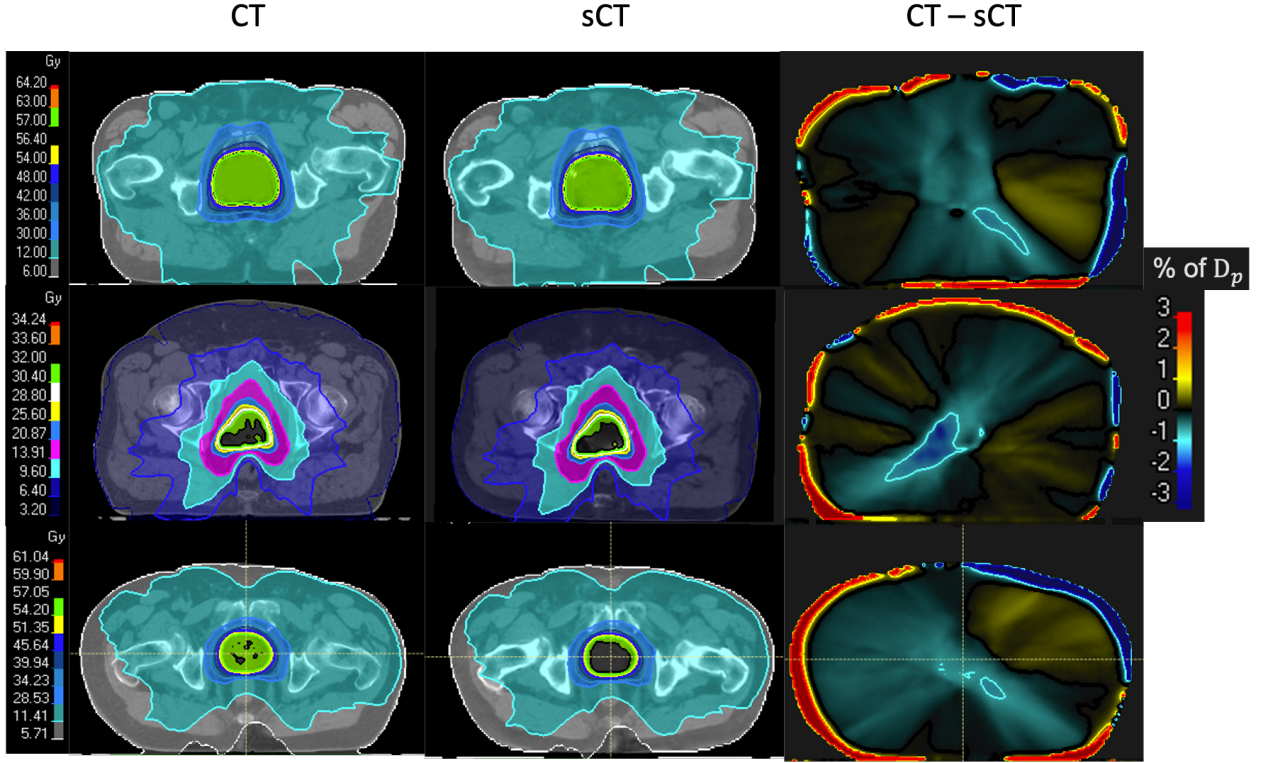


Figure 4.6: From left to right, doses calculated on CT and sCT and the relative dose difference (CT-sCT).

#### 4.4.4 CycleGAN comparison

A CycleGAN (Zhu, Park, et al., 2017) has been trained with the same data as our AugCycleGAN for qualitative comparison purposes.

*Training Losses* The cycle consistency loss which measures the similarity between two images of the same domain is evaluated in training phase. As shown in Figure 4.7, the AugCycleGAN improves dramatically the minimization and stability.

*Images synthesis* For site 5 testing patient, the MRI range of values is close to the ones seen in training. The trained CycleGAN manages to keep a good HU consistency, however a vertical and horizontal shift appears on the synthesized images. In addition, the CycleGAN does not recognize the bladder Figure 4.8 (a).

Contrary to site 5, site 4 MRI distribution is quite different. The CycleGAN struggles to recognize the shape resulting in a poorer image quality and appearance of artifacts. The shift is less serious in that case, Figure 4.9.

## 4.5 DISCUSSION

In this study, an MR-to-CT synthesis method was proposed to provide increased generalization performance in comparison with state-of-the-art approaches. We augmented the CycleGAN approach by



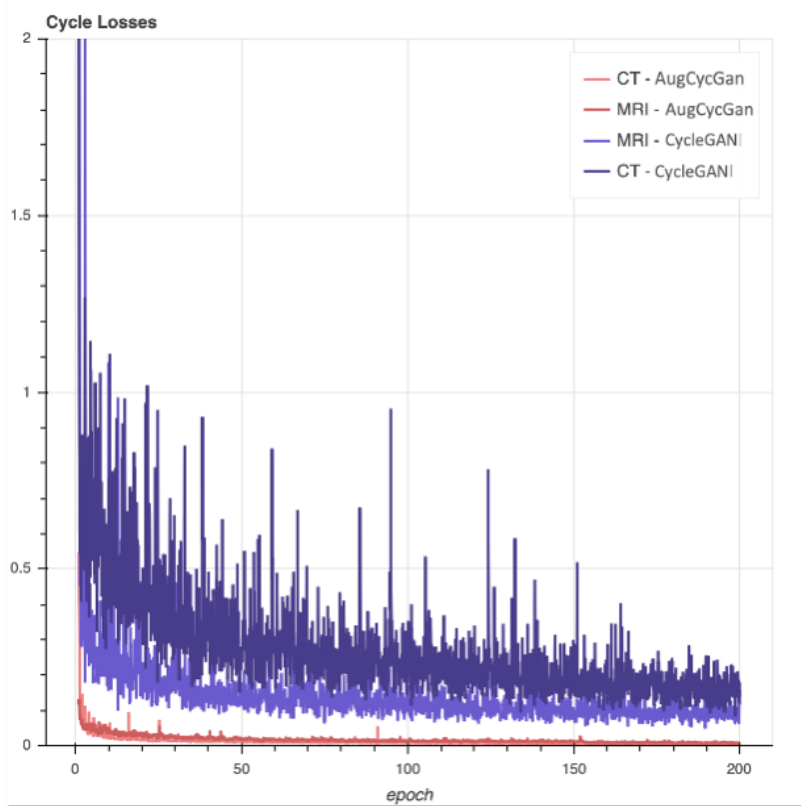


Figure 4.7: MRI and CT cycle loss in training phase (3-WMA).

adding a latent space to enable the reconstruction of several MR images, therefore making the training of this unsupervised model possible with multicentric data. The results suggest that our model can learn to synthesize a clinically acceptable sCT while using different scanners/parameters. With pre-training, this model can address the limitations of small datasets by finetuning it with a reduced sample.

The use of a vanilla CycleGAN in the same configuration produces unreliable results (see 4.4.4). A vanilla CycleGAN will not necessarily fail to converge but will fail to learn an appropriate model because it simply does not have the capacity to learn one-to-many mappings while stochastic CycleGANs do.

The average MAE of 65.8 HU and 59.8 HU for both sites were comparable to those obtained in a previous paired study (60 HU, Maspero et al. (2018)) or unpaired study (50.8 HU, Lei, Harms, et al. (2019)). However, several biases<sup>4</sup> in the training data may be observed on sCT. First, this method may show a tendency to generate higher HU values due to the different image value-to-density tables. Our model occasionally tries to generate a fiducial marker in the prostate or shows a higher HU value in the bottom of the bladder due to the presence of contrast agent. In clinical practice, it will be beneficial to standardize the training data with a clinical protocol by performing additional modifications, such as directly using the electron density and not HU values as the output of the model.

This study is the first to show sCT images generated with a Cycle-

<sup>4</sup>Here we mean that the training data has imperfections that are also unfortunately learned by our model whereas they are artifacts without genuine anatomical meaning. This is another source of uncertainty in medical data that is cumulated with more usual sources of uncertainty in ML such training sample finiteness (lack of knowledge / epistemic uncertainty) and noise (randomness / ontic uncertainty).

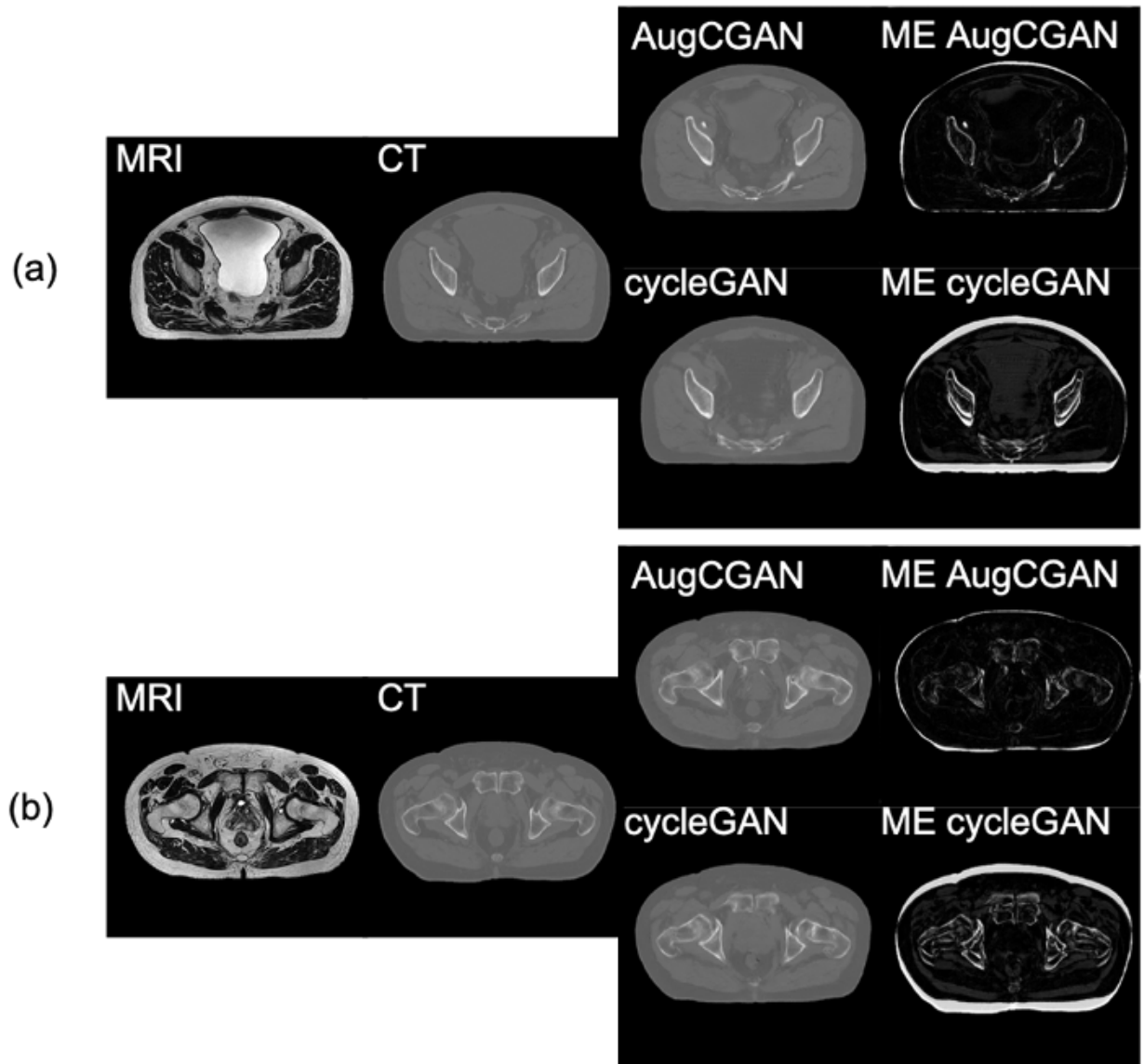


Figure 4.8: sCT generated from two site 5 patients.

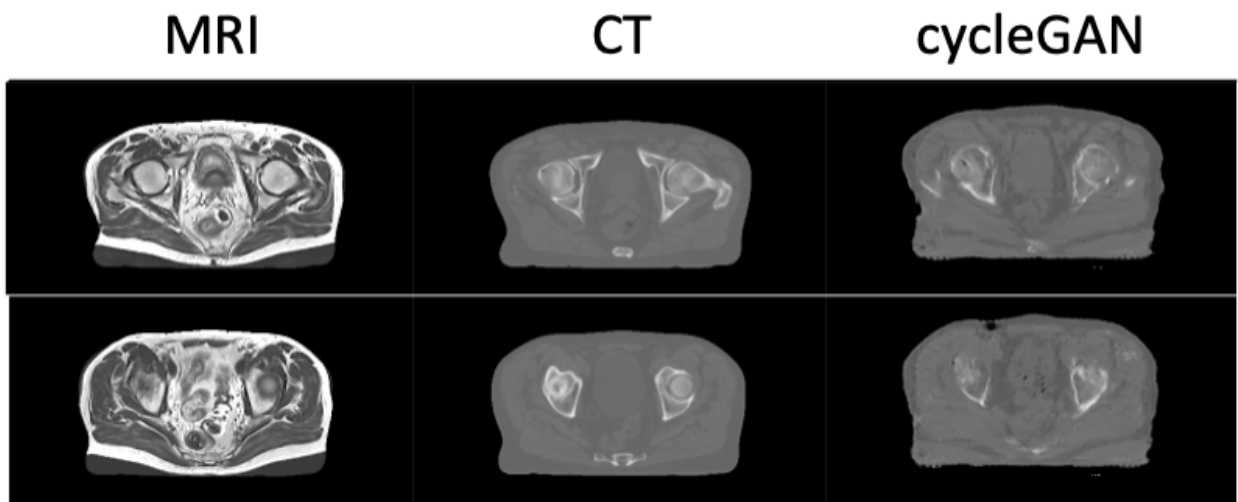


Figure 4.9: sCT generated from two site 4 patients.

GAN approach with unpaired multicentric data that underwent dosimetric evaluation in the pelvic area. Although interscan differences certainly influence the recalculation plan for the sCT images, all the DVH points recalculated on sCT were within  $\pm 3\%$  with respect to the CT images. Maximum relative dose difference in the high-dose region is 1.2%, which is consistent with previous studies (Maspero et al., 2018; Kemppainen et al., 2017). An advantage of increased generalization is that it will avoid the need to retrain a model for each new MR sequence or even a small change in sequence setting. Current CycleGAN-like models do not take into account MR variability. Collection of new data and retraining a model will substantially impede the use of deep learning in clinical practice.

## 4.6 CONCLUSION

In this chapter, we proposed an augmentation of CycleGAN to generate synthetic CT images using multicentric data in an unpaired fashion without the need for a dedicated MR sequence. The possibility of using several sequences with this approach will allow the development of a single-body model instead of an anatomy-specific model. Our method was designed to improve the generalization of MR-to-CT synthesis and not directly improve the accuracy of HU. Future studies should aim to further develop AugCycGAN (which shares much with CycleGAN) with deep learning methods such as 3D or dense blocks to improve accuracy as well as generalization.

## APPENDICES

## 4.A FIDUCIAL MARKERS

The presence of fiducial markers in the prostate or contrast agent in the training dataset has an effect in testing phase. The trained model tries sometimes to reconstruct fiducial markers in the prostate, see Figure 4.A.1 (a). Occasionally, it assigns higher HU in the bottom of the bladder.

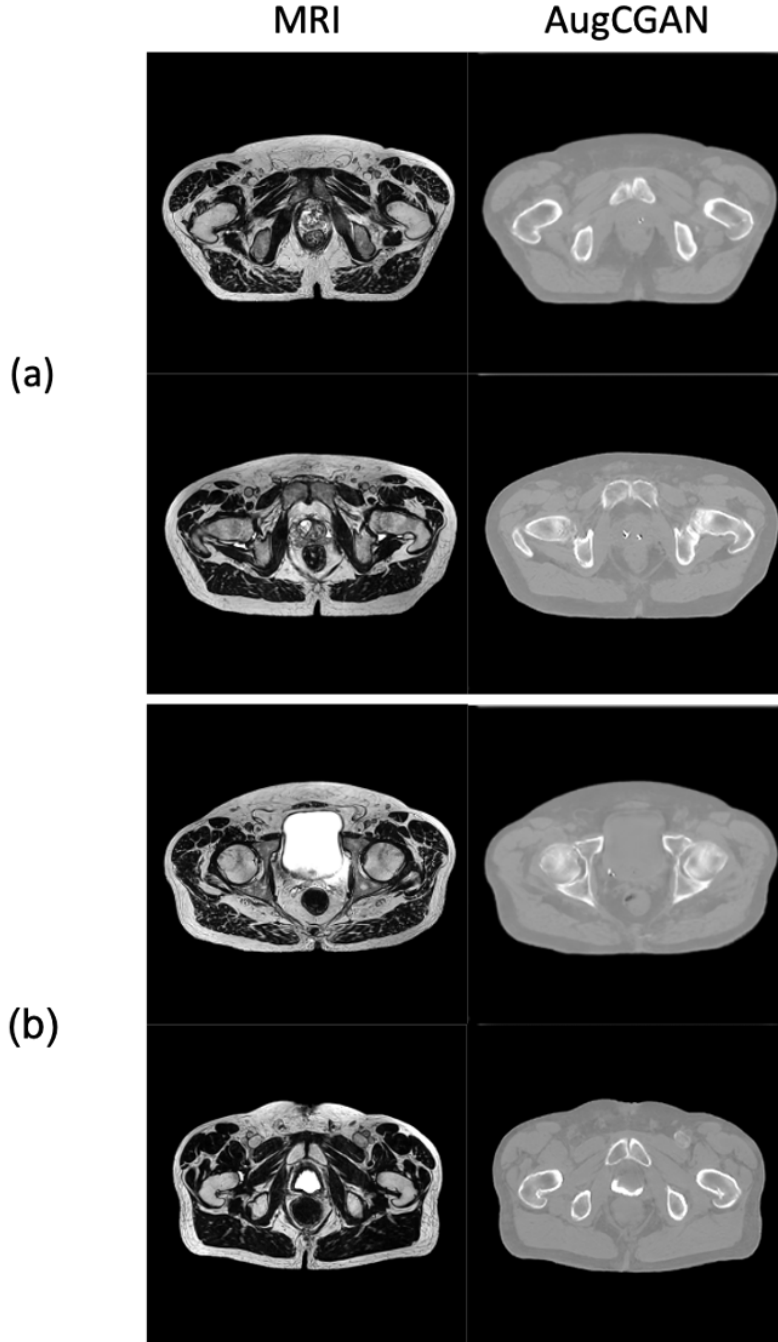


Figure 4.A.1: – Bias effect on AugCGAN generation.

## 4.B NETWORK DETAILS

An augmented CycleGAN (AugCycGAN, Almahairi et al., 2018) shares the same architecture as a CycleGAN (Zhu, Park, et al., 2017) concerning the generators and discriminators. Both networks use the same losses too, the differences between these two networks appears in the use on Conditional Instance Normalization (CIN, Dumoulin, Shlens, and Kudlur, 2016) instead of Instance Normalization (IN, Ulyanov, Vedaldi, and Lempitsky, 2016) in one generator (He et al., 2016b).

*Instance Normalization* A classical IN follows this formula for a 4D input tensor (a mini-batch of 2D inputs with additional channel dimension)  $x$ :

$$y = \frac{x - \mathbb{E}[x]}{\sqrt{\text{Var}[x] + \epsilon}} \times \gamma + \beta, \quad (4.B.1)$$

expectation  $\mathbb{E}[x]$  and variance  $\text{Var}[x]$  are computed per-dimension separately for each object in a mini-batches,  $\epsilon$  allows to avoid divisions by zero and  $\gamma$  and  $\beta$  are learnable parameters.

*Conditional Instance Normalization in style transfer* The learnable parameters are replaced by the outputs of two linear functions  $\mu$  and  $\nu$  which take a latent variable  $z_{mr}$  as input i.e.,  $\gamma = \mu(z_{mr})$  for the scaling factor and  $\beta = \nu(z_{mr})$  for the shifting factor. The latent variable  $z_{mr}$  is the output of the encoder  $E_{mr} : CT \times MRI \mapsto Z_{MRI}$  in charge of capturing the missing information when converting an MRI into a CT.

The intuition behind this operation is to apply an affine transformation specific to each sequence instead of learning an affine transformation corresponding to several sequences at once.

*Details of the encoder architecture* We follow the naming convention used in CycleGAN (Zhu, Park, et al., 2017). Let c1-k denote a  $1 \times 1$  Convolution with k filters, c3r-k denotes a  $3 \times 3$  Convolution-ReLU layer with k filters and c3br-k denotes a  $3 \times 3$  Convolution-BatchNorm-ReLU layer with k filters.

Encoder: c3r-32, c3br-64, c3br-128, c3br-256, c3br-512, c3br-1024, c3br-1024, c1-32.

The scaling factor  $\mu$  and the shifting factor  $\beta$  are two  $1 \times 1$  Convolution-ReLU with the same number of filters in the layers where they are applied.

*Details of the augmented generator architecture* Let c1-k denote a  $1 \times 1$  Convolution with k filters, c7cr-k denotes a  $7 \times 7$  Convolution-CIN-ReLU layer with k filters, Rc-k denotes a CIN residual block two  $3 \times 3$  convolutional layers with the same number of filters on both layers and ct3cr-128 denotes a  $3 \times 3$  Transposed Convolution-CIN-ReLU layer with 128 filters.

Augmented Generator: c7cr-64, c7cr-128, c7cr-256, Rc-256, Rc-256, Rc-256, Rc-256, Rc-256, Rc-256, Rc-256, ct3cr-128, c3cr-64, c7-64, tanh.

*Details of the latent discriminator architecture* Let l-k denote a linear operation of size k, lbr-k denotes a linear-batchsize-LReLU of size k.

Latent discriminator: lbr-64, lbr-64, lbr-64, l-1.

*Details of the (non-augmented) generator and discriminator architectures* Both have the same architecture like those described in the previous chapter, see Section 3.A.

A visualization of their interaction in the backward cycle is given in Figure 4.B.1.

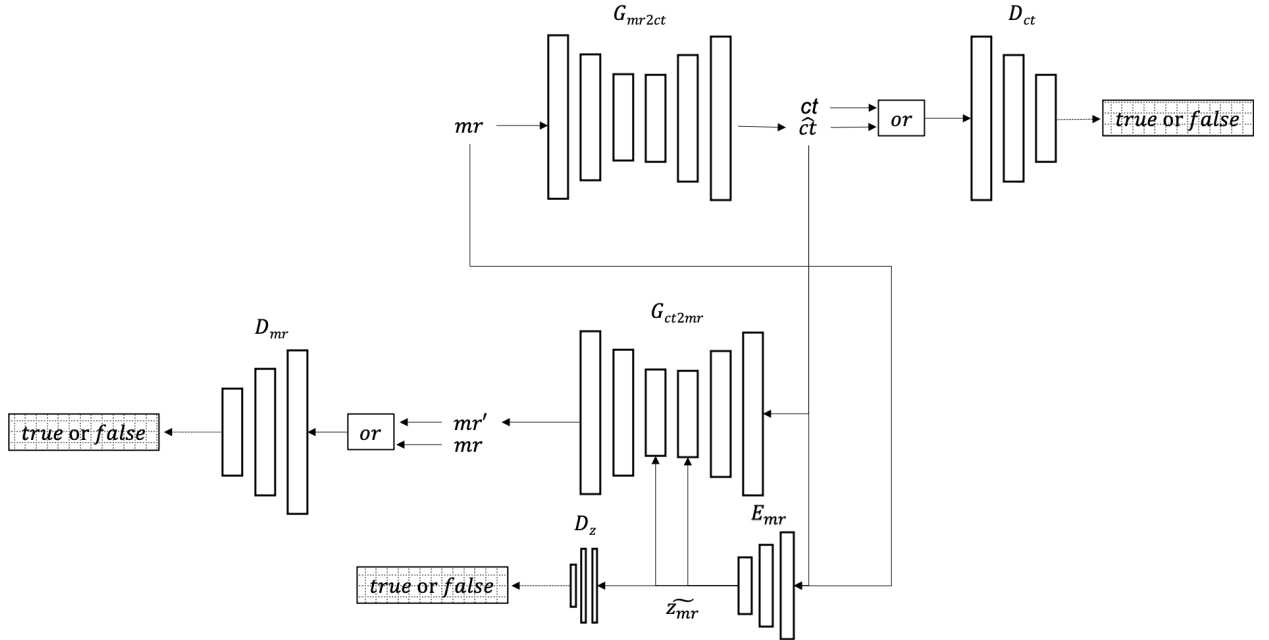


Figure 4.B.1: - Cycle starting from an MRI.



# Discussion and conclusion

This thesis has developed two new generic approaches for training a GAN for MR-to-CT synthesis in radiotherapy. The idea is simple: the sought generalization properties should be achieved through MR image variability which is induced in the training phase<sup>5</sup>. To do so, we essentially leverage two sources of variability during the training: different types of MR sequences and/or different medical imaging centers/acquisition devices. By doing this, we force the neural net to compose with a global representation of MRI thus alleviating preprocessing steps to standardize MR images.

The first proposed method was applied into a paired fashion with a conditional GAN. Unpaired fashion on the other hand was more challenging. In a second method, an augmentation of a cycle GAN has been devised in order to use different MR sequences in the training. A focus has been made in the pelvic area in accordance with the CoBra project, but this technique is independent from localization.

In the next paragraph, we summarize these two methods and discuss their respective advantages and limitations. We also provide insightful perspectives regarding the MR-to-CT synthesis task but also with respect to the more general workflow pertaining to the

<sup>5</sup>This starts from a simple observation: the factors of variation that explain the data should be easier to identify when the data exhibits sufficient diversity to cover a significant portion of the range of these factors. For example, suppose you want to learn a visual model of dogs. If your data contains only images of male chihuahua, it will be virtually impossible to identify factors of variations such as size, gender, hair style and color and to generalize to other kinds of dogs.

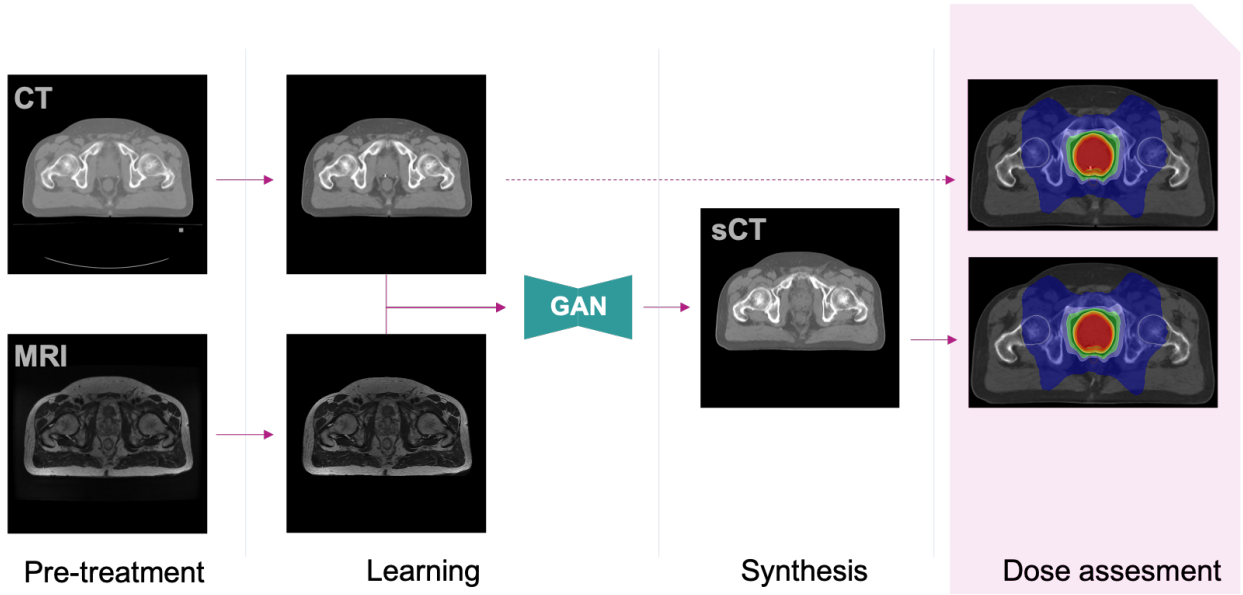
## A GENERAL DEEP LEARNING APPROACH IN MR-TO-CT SYNTHESIS IN RADIOTHERAPY

We decided to plainly use the deep learning capability in finding automatically appropriate features for our task. The preprocessing steps were limited to registration and normalization to keep the number (and level) of non-data-driven steps to a minimum. An illustration summarizing the workflow used for both studies is shown in Figure 4.2.

CHAPTER 3 described one of the first synthetic CT generated in a multicentric context. This proof-of-concept has been proposed for paired datasets (containing MR/CT pairs of the same patient) with a conditional GAN initially used for high definition image synthesis, thus allowing a gain in accuracy. A vanilla conditional GAN has been used for comparison. Both are able to reach convergence so we can deduce as a first result that any conditional GAN architecture should be able to deal with different MR sequences. Special attention was given to keep the workflow simple while not altering the size and the quality of the images.

This method is the most accurate and simple of our two approaches. However, the need of a paired dataset is an obvious limitation as many medical imaging centers might not have a number of paired examples. The high-resolution synthesis was first presented in Brou Boni





et al. (2019b) and the generalization in Boni et al. (2020).

CHAPTER 4 addressed one issue of vanilla cycle GAN in an unpaired dataset context. We provide in this chapter a method to counter this issue by using an augmentation, i.e. providing latent variables as additional inputs for the MRI generator. Classical cycle GANs fail to properly converge since it is not explicit for the generator to know which sequence it should generate. This results in a poorer cycle loss optimization. The augmentation is simply done by implicitly giving the generator this information. This conditioning was done using a conditional instance normalization<sup>6</sup>. It was the only effective way to condition the sequence information in our case.

Compared to our first contribution, this method is more complex and therefore more data demanding but this is justified by the use of an unpaired dataset. Hardware requirement forces us to downsample the images. Results are less accurate compared to the first approach, we can nevertheless underline some larger differences in sequence parameters between training and testing sequences.

In conclusion, the two proposed methods have (unsurprisingly) complementary advantages. The first method is more accurate but requires a higher level of supervision in the construction of the dataset. The second method can learn from more general datasets but cannot achieve the same level of accuracy.

In the following, some open questions related to this thesis are tackled. The provided answers are personal views of the author regarding these remaining challenges.

Figure 4.2: Illustration of the workflow used to evaluate our methods.

<sup>6</sup>It is a Feature wise instance modulation FiLM. It refers to a model which is conditioned or modulated by information extracted from an additional input. Concatenation is the most widely use type of FiLM in DL (see beginning of 4.B for more details).

ONE TO RULE THEM ALL ? Assuming that different sequences in training is no longer a problem, it becomes possible to mix several anatomical localizations<sup>7</sup> to build a multi-localization model. A dedicated model specific to each localization is likely to be more accurate compared to a global one. On the other side, one can reckon that a global model provides an alternative to collecting data from different localizations and then training repeatedly one model for each. We see this option as a trade-off between time and performance knowing that even a gain in HU consistency will not affect that much the dosimetric fidelity. Another compromise between genericity versus accuracy.

<sup>7</sup>In general, each localization is associated with a different set of sequence.

ACCURACY ABOVE ALL ? One may wonder if MR-to-CT synthesis is not just about accuracy. Indeed, current studies focus mainly on the performance of their model and it becomes increasingly difficult to find clinical relevant difference between a real CT and a synthetic CT. In my opinion, accuracy is no longer an obstacle to the establishment of such technique in clinical routine, reliability is. We cannot (at least not yet) count on our neural nets to explain why sometimes they fail which is why safe/explainable AI has become a hot trend in the ML community. Studies must be conducted by stressing out a bit our model to trigger their weaknesses if they are.

IS MY SCT IS RIGHT OR WRONG? Unlike semi-automatic task like segmentation in which an expert is in charge of assessing the prediction, sCT cannot be assessed by an expert. Another practical challenge to the establishment of deep learning solutions will be the acceptance. Quality guarantees must be given for a better acceptance and also for monitoring.

Within GAN based solutions, the discriminator could be a start. This network is systematically discarded at test time (because the original purpose is sCT generation) while it is so far the one which has the greatest knowledge of what can be right or wrong. We could associate the sCT with the prediction made by the discriminator as a first step. Calibration techniques could be employed to the output probabilities of the discriminator in order to provide more useful information. However, this process remains tied to a particular choice of the discriminator architecture and hyperparameters.

We conclude this manuscript with additional comments on two other tasks involved in the general workflow of the CoBra project, namely segmentation and dose calculation.

#### AUTOMATIC SEGMENTATION

Our recent experiments have taught us how time consuming it is to deal with the DICOM standard<sup>8</sup> for a newbie. We alleviate the burden of researchers or students by creating a segmentation toolbox in radiation therapy. `segmentation_rt`<sup>9</sup> is a python package designed to easily create a database and train a deep learning model. Inferred delineations are converted into readable RT Structures dicom.

<sup>8</sup>DICOM is the international standard to transmit, store, retrieve, print, process and display medical imaging information.

 [github.com/BrouBoni/segmentation\\_RT](https://github.com/BrouBoni/segmentation_RT)  
 [segmentation-rt.readthedocs.io](https://segmentation-rt.readthedocs.io)

*Public audience and objectives* Several commercial solutions exists for the auto-segmentation of **OARs** and **PTV** in radiotherapy, we do not plan to compete. We have just leveraged a **GAN** generator (3D) and our experience in "dicomization" to begin with this free framework.

The goal is to automate pre- and post-processing tasks in order to spare time for more challenging **DL** tasks. Indeed, raw data in radiotherapy are in DICOM while current deep learning frameworks use more conventional formats. When it is time to evaluate our results, we have to convert into DICOM, a tedious job when it comes to RT Structures. We therefore propose an end-to-end solution to perform all these task automatically.

In the world of open science, everyone is free to improve our package or just use a piece for convenience. It is a high level package divided in three modules

*rs2mask* create a deep learning friendly dataset from DICOM data.

*dl* deep learning module for training and testing.

*mask2rs* create RT Structure Set from a previously generated masks.

Other such modules already exist but they are organized independently to each other, meaning that there is not a smooth workflow to perform all these tasks without hand-crafted communication code.

*Examples in use* So far, our package has been used in the case of head/neck and breast irradiation by a physicist and a medical student. No quantitative study has been done yet, promising visuals are given in Figure 4.3 and 4.4.

## DOSE PREDICTION IN BRACHYTHERAPY

The Centre Oscar Lambret has a long experience on dose calculation for radiotherapy treatment planning. An in-house **MC** platform was developed (Reynaert et al., 2016; Wagner et al., 2017) for validation of Cyberknife and Tomotherapy plans. The system is currently extended for conventional **IMRT** and **volumetric modulated arc therapy (VMAT)** treatments (Alhamada et al., 2018). An option would be to introduce brachytherapy into the system. But again this would only be to provide a reference method as **MC** is too slow (double-checking). An analytical dose calculation based on the report of AAPM<sup>9</sup> Task Group-43 (TG-43, Rivard et al. (2004)) seems to be a more realistic option and are used by most brachytherapy **TPS** but self-attenuation is not taken into account.

Treatment plan optimization is the third brick of the current **CoBra** workpackage. In the **TPS**, the dose distribution is optimized using an objective function and gradient-descent optimization methods.

For prostate treatment using iodine seeds (**low dose rate, LDR**) one can even consider using geometrical optimization as an optimal

<sup>9</sup>The American Association of Physicists in Medicine

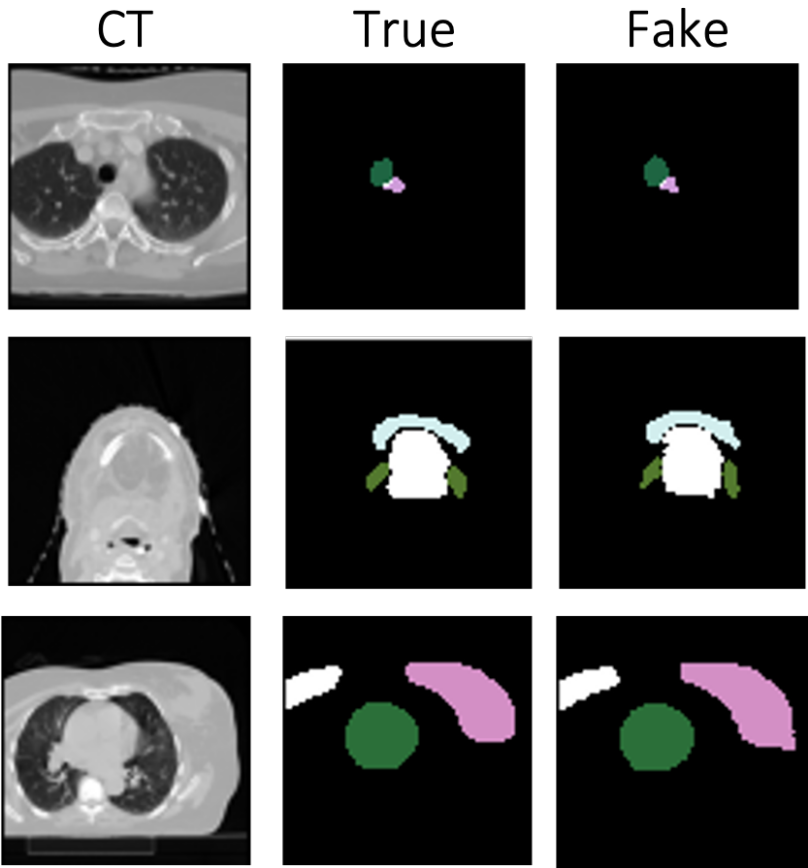


Figure 4.3: Comparison between generated and real segmentation for a thoracic, head and breast case.

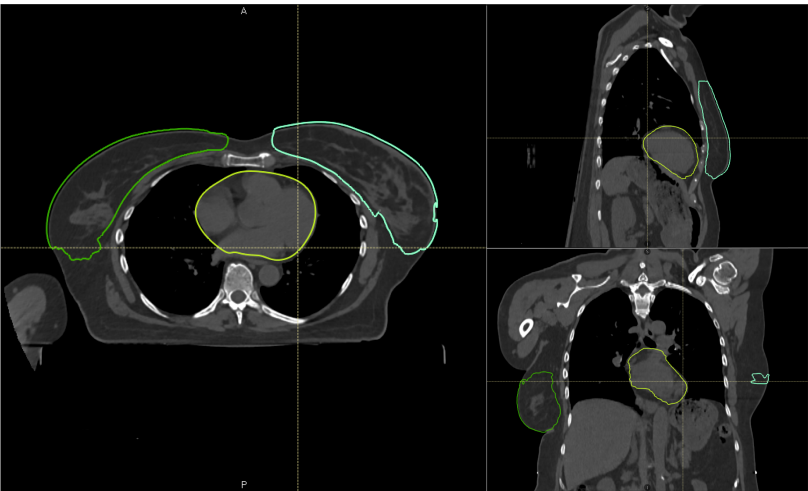


Figure 4.4: Breast case imported in a TPS (Raystation).

source configuration that automatically provides an optimal dose distribution<sup>10</sup>. In between insertion of different seeds, on-line MR images can be used to determine the actual source configuration and the actual deformed prostate geometry before inserting the next seed at the optimal position. In this workflow, a single dose calculation is only needed when all seeds are inserted, no need of an accurate dose calculation.

On the other hand, for the higher dose rate (HDR) the situation is more complicated, demanding an optimization of dose and thus intermediate dose calculations (on intermediate sCT images). Providing a fast enough and on-line adaptive solution can be a challenge for HDR brachytherapy but not for all localization.

*Deep learning speed with MC accuracy* MC algorithms are the gold standard in terms of accuracy but are time consuming. Interest has been shown in DL to mimic MC precision while being faster. CNN dose estimation has demonstrated good results both in LDR (Villa et al., 2021) and HDR (RapidBrachyDL, Mao et al., 2020) in the prostate. However, in terms of dose difference, MC algorithm compared to TG-43 does not provide a significant clinical gain (Ma et al., 2015) in prostate HDR. In comparison, there is a relative dose difference for the prostate D90 of 0.9% between MC/TG-43 and 0.7% between MC/RapidBrachyDL. There are important differences but mostly in areas where the doses are very low like the bones.

There is not a clear advantage mimicking MC with DL at least for HDR in the prostate. Locations with heterogeneities such as the neck or the presence of material with high atomic number (gynecological applicator) on the other hand would be more suitable.

*Experiments in Centre Oscar Lambret* Since the installation last year of a new MRI (Siemens Magnetom Sola 1.5 T) partially dedicated to the radiotherapy, an effort has been made to find sequences allowing an MRI-only brachytherapy (HDR) when possible. A set of two sequences is used,  $T_2$  weighted images are used for contouring and a  $T_1$  dixon to reconstruct the needle or catheter<sup>11</sup>.

Here, we mainly use the MRI for better contrast and we do not seek accuracy in dose calculation. The dose calculation is performed using the TG-43 by assigning the water density to all tissues, it is sufficient in terms of accuracy.

## DOSE PREDICTION IN EXTERNAL RADIOTHERAPY

Seeking MC precision in brachytherapy is not a subject of great interest in medical physics. The clinical benefits are not that significant because it is a very local treatment in which the treated area are quite homogeneous. Even in external beam radiotherapy, MC is not always used because the clinical difference compared to conventional dose calculation algorithms like the collapsed cone (Ahnesjö, 1989) are not significant in most cases.

<sup>10</sup>This is thanks the low energy of the emitted gammas, leading to a local dose deposition, no need of taking into account electron transport.

<sup>11</sup>This is needed by the TPS to optimize the source placement

The other cases, on the other hand, have more interest in being double-checked with **MC**. The most challenging treatment machine at our disposal is the Cyberknife “VSI”, a stereotactic body radiation therapy system. Higher doses per fraction is delivered to much smaller target volumes. Strong gradient around the target is made possible by the use of a very large number of angular incidence.

*Dose prediction* A preliminary study was conducted using 18 patients with multiple localizations. The number of incident beams ranges from 20 to 205 (including non-coplanar angles) with no **MLC**. The objective was to refine a dose calculated on the **TPS** in **MC**.

The chosen deep learning architecture is a **CNN** with **FiLM**. Basically, the design is the same as our stochastic generator presented in appendix 4.2.1 or the one from Mao et al. (2020). This architecture uses 3D convolutions. One of the main reason is the use of non-coplanar angles. The motivation behind this choice is to maintain a consistency in term of dose calculation more than increasing the learning capacity of the network. An illustration of the net is given in Figure 4.5.

Currently, a visual inspection gives encouraging results in terms of convergence. This work will be pursued in the coming months.

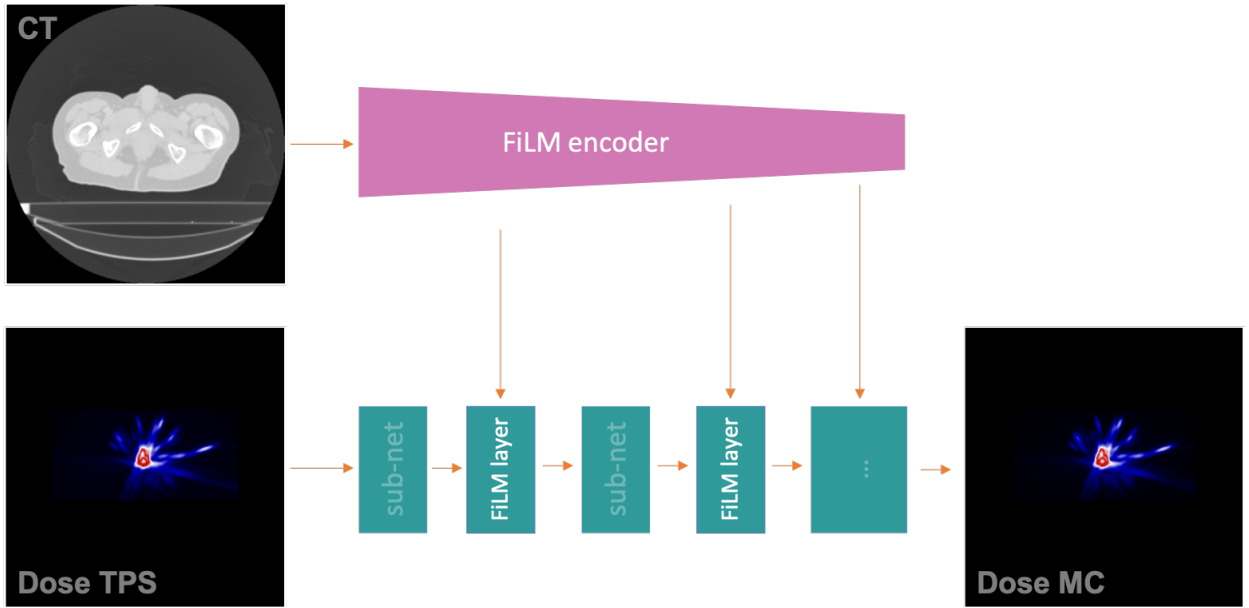


Figure 4.5: Dose refinement net: the **FiLM** encoder predicts parameters describing the patient geometry and composition. The dose refinement generator is conditioned by the **FiLM** parameters.



# Bibliography

- Ahnesjö, A. 1989. *Collapsed cone convolution of radiant energy for photon dose calculation in heterogeneous media*. Medical physics. (see p. 80).
- Alhamada, H., S. Simon, Y. Jourani, C. Vandekerckhove, N. Pauly, A. Dubus, and N. Reynaert. 2018. *[OA239] Monte Carlo as quality control tool of machine performance and treatment planning system, is it a luxury or a necessity?* Physica Medica: European Journal of Medical Physics. (see p. 78).
- Almahairi, A., S. Rajeswar, A. Sordoni, P. Bachman, and A. Courville. 2018. *Augmented CycleGAN: Learning Many-to-Many Mappings from Unpaired Data*. arXiv:1802.10151. (see pp. 58, 59, 60, 61, 72).
- Andreasen, D., K. Van Leemput, and J. M. Edmund. 2016. *A patch-based pseudo-CT approach for MRI-only radiotherapy in the pelvis*. Medical physics. (see p. 41).
- Andreasen, D., K. Van Leemput, R. H. Hansen, J. A. Andersen, and J. M. Edmund. 2015. *Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain*. Medical physics. (see p. 41).
- Arjovsky, M., S. Chintala, and L. Bottou. 2017. *Wasserstein generative adversarial networks*. In International conference on machine learning. PMLR. (see p. 48).
- Bloch, F. 1946. *Nuclear induction*. Physical review. (see p. 23).
- Boni, K. B., J. Klein, L. Vanquin, D. Pasquier, and N. Reynaert. 2019. *Génération d'un CT synthétique à partir d'une séquence IRM avec un réseau antagoniste génératif en radiothérapie*. In GRETSI 2019. (see p. 16).
- Boni, K. N. B., J. Klein, L. Vanquin, A. Wagner, T. Lacornerie, D. Pasquier, and N. Reynaert. 2020. *MR to CT synthesis with multicenter data in the pelvic area using a conditional generative adversarial network*. Physics in Medicine & Biology. (see pp. 16, 17, 45, 58, 76).
- Boydev, C., B. Demol, D. Pasquier, H. Saint-Jalmes, G. Delpon, and N. Reynaert. 2017. *Zero echo time MRI-only treatment planning for radiation therapy of brain tumors after resection*. Physica Medica. (see p. 39).
- Brou Boni, K., A. Wagner, L. Vanquin, J. Klein, N. Reynaert, and D. Pasquier. 2019a. *Génération de tomodensitométrie synthétique par apprentissage profond pour la radiothérapie du cancer de la prostate basée sur l'IRM seule*. 30e Congrès national de la Société française de radiothérapie oncologique, Cancer/Radiothérapie. (see p. 16).
- Brou Boni, K. N., J. Klein, A. Gulyban, N. Reynaert, and D. Pasquier. 2021. *Improving generalization in MR-to-CT synthesis in radiotherapy by using an augmented cycle generative adversarial network with unpaired data*. Medical Physics. (see pp. 16, 17, 57).
- Brou Boni, K. N., A. Wagner, L. Vanquin, J. Klein, D. Pasquier, and N. Reynaert. 2019b. *High-Resolution Synthetic-CT Generation with Conditional Generative Adversarial Networks*, in *Magnetic Resonance in Radio-Therapy*. MRinRT 2019, Toronto, Canada, July 2019. (see pp. 16, 75).



- Catana, C., A. van der Kouwe, T. Benner, C. J. Michel, M. Hamm, M. Fenchel, B. Fischl, B. Rosen, M. Schmand, and A. G. Sorensen. 2010. *Towards implementing an MR-based PET attenuation correction method for neurological studies on the MR-PET brain prototype*. Journal of nuclear medicine: official publication, Society of Nuclear Medicine. (see p. 39).
- Cybenko, G. 1989. *Approximation by superpositions of a sigmoidal function*. Mathematics of Control, Signals and Systems. (see p. 32).
- Debois, M., R. Oyen, F. Maes, G. Verswijvel, G. Gatti, H. Bosmans, M. Feron, E. Bellon, G. Kutcher, H. Van Poppel, et al. 1999. *The contribution of magnetic resonance imaging to the three-dimensional treatment planning of localized prostate cancer*. International Journal of Radiation Oncology\* Biology\* Physics. (see p. 14).
- Devic, S. 2012. *MRI simulation for radiotherapy treatment planning*. Medical physics. (see p. 37).
- Dinkla, A. M., M. C. Florkow, M. Maspero, M. H. Savenije, F. Zijlstra, P. A. Doornaert, M. van Stralen, M. E. Philippens, C. A. van den Berg, and P. R. Seevinck. 2019. *Dosimetric evaluation of synthetic CT for head and neck radiotherapy generated by a patch-based three-dimensional convolutional neural network*. Medical physics. (see p. 42).
- Dinkla, A. M., J. M. Wolterink, M. Maspero, M. H. Savenije, J. J. Verhoeff, E. Seravalli, I. Išgum, P. R. Seevinck, and C. A. van den Berg. 2018. *MR-only brain radiotherapy: Dosimetric evaluation of synthetic CTs generated by a dilated convolutional neural network*. International Journal of Radiation Oncology Biology Physics. (see pp. 41, 46).
- Dowling, J. A., J. Lambert, J. Parker, O. Salvado, J. Fripp, A. Capp, C. Wratten, J. W. Denham, and P. B. Greer. 2012. *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. (see p. 39).
- Dumoulin, V., J. Shlens, and M. Kudlur. 2016. *A Learned Representation For Artistic Style*. arXiv:1610.07629. (see p. 72).
- Eilertsen, K., L. Nilsen Tor Arne Vestad, O. Geier, and A. Skretting. 2008. *A simulation of MRI based dose calculations on the basis of radiotherapy planning CT images*. Acta Oncologica. (see p. 40).
- Fu, J., Y. Yang, K. Singhrao, D. Ruan, D. A. Low, and J. H. Lewis. 2018. *Male pelvic synthetic CT generation from T1-weighted MRI using 2D and 3D convolutional neural networks*. arXiv preprint arXiv:1803.00131. (see pp. 41, 58).
- Gatys, L. A., A. S. Ecker, and M. Bethge. 2016. *Image style transfer using convolutional neural networks*. In Proceedings of the IEEE conference on computer vision and pattern recognition. (see p. 42).
- Goodfellow, I. 2016. *Nips 2016 tutorial: Generative adversarial networks*. arXiv preprint arXiv:1701.00160. (see pp. 33, 34, 35, 47, 59).
- Grodzki, D. M., P. M. Jakob, and B. Heismann. 2012. *Ultrashort echo time imaging using pointwise encoding time reduction with radial acquisition (PETRA)*. Magnetic resonance in medicine. (see p. 39).
- Gudur, M. S. R., W. Hara, Q.-T. Le, L. Wang, L. Xing, and R. Li. 2014. *A unifying probabilistic Bayesian approach to derive electron density from MRI for radiation therapy treatment planning*. Physics in Medicine & Biology. (see p. 40).
- Guerreiro, F., N. Burgos, A. Dunlop, K. Wong, I. Petkar, C. Nutting, K. Harrington, S. Bhide, K. Newbold, D. Dearnaley, et al. 2017. *Evaluation of a multi-atlas CT synthesis approach for MRI-only radiotherapy treatment planning*. Physica Medica. (see p. 39).

- Han, X. 2017. *MR-based synthetic CT generation using a deep convolutional neural network method*. Medical physics. (see pp. 41, 46).
- He, K., X. Zhang, S. Ren, and J. Sun. 2016a. *Deep residual learning for image recognition*. In Proceedings of the IEEE conference on computer vision and pattern recognition. (see p. 55).
- . 2016b. *Deep Residual Learning for Image Recognition*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR). (see p. 72).
- Hiasa, Y., Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano, and Y. Sato. 2018. *Cross-Modality Image Synthesis from Unpaired Data Using CycleGAN*. (see p. 58).
- Hinton, G. E., N. Srivastava, A. Krizhevsky, I. Sutskever, and R. R. Salakhutdinov. 2012. *Improving neural networks by preventing co-adaptation of feature detectors*. arXiv preprint arXiv:1207.0580. (see p. 47).
- Hornik, K. 1991. *Approximation capabilities of multilayer feedforward networks*. Neural networks. (see p. 32).
- Hsu, S.-H., Y. Cao, K. Huang, M. Feng, and J. M. Balter. 2013. *Investigation of a method for generating synthetic CT models from MRI scans of the head and neck for radiation therapy*. Physics in Medicine & Biology. (see p. 40).
- Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros. 2016. *Image-to-Image Translation with Conditional Adversarial Networks*. arxiv. (see pp. 42, 48, 50, 55, 58).
- Johansson, A., M. Karlsson, and T. Nyholm. 2011. *CT substitute derived from MRI sequences with ultrashort echo time*. Medical physics. (see pp. 39, 40).
- Johansson, A., M. Karlsson, J. Yu, T. Asklund, and T. Nyholm. 2012. *Voxel-wise uncertainty in CT substitute derived from MRI*. Medical physics. (see p. 40).
- Johnson, H. J., M. M. McCormick, L. Ibanez, I. S. Consortium, et al. 2015. *The ITK software guide: design and functionality*. Kitware. (see p. 49).
- Johnson, J., A. Alahi, and L. Fei-Fei. 2016. *Perceptual losses for real-time style transfer and super-resolution*. In European conference on computer vision. Springer. (see p. 48).
- Johnstone, E., J. J. Wyatt, A. M. Henry, S. C. Short, D. Sebag-Montefiore, L. Murray, C. G. Kelly, H. M. McCallum, and R. Speight. 2018. *Systematic review of synthetic computed tomography generation methodologies for use in magnetic resonance imaging-only radiation therapy*. International Journal of Radiation Oncology\* Biology\* Physics. (see p. 38).
- Jonsson, J. H., A. Johansson, K. Söderström, T. Asklund, and T. Nyholm. 2013. *Treatment planning of intracranial targets on MRI derived substitute CT data*. Radiotherapy and Oncology. (see p. 38).
- Jonsson, J. H., M. G. Karlsson, M. Karlsson, and T. Nyholm. 2010. *Treatment planning using MRI data: an analysis of the dose calculation accuracy for different treatment regions*. Radiation oncology. (see p. 38).
- Kapanen, M., and M. Tenhunen. 2013. *T1/T2\*-weighted MRI provides clinically relevant pseudo-CT density data for the pelvic bones in MRI-only based radiotherapy treatment planning*. Acta Oncologica. (see p. 37).
- Keereman, V., Y. Fierens, T. Broux, Y. De Deene, M. Lonnew, and S. Vandenberghe. 2010. *MRI-based attenuation correction for PET/MRI using ultrashort echo time sequences*. Journal of nuclear medicine. (see pp. 38, 39).

- Kemppainen, R., S. Suilamo, T. Tuokkola, P. Lindholm, M. H. Deppe, and J. Keyriläinen. 2017. *Magnetic resonance-only simulation and dose calculation in external beam radiation therapy: a feasibility study for pelvic cancers*. Acta Oncologica. (see p. 70).
- Kingma, D. P., and J. Ba. 2014. *Adam: A method for stochastic optimization*. arXiv preprint arXiv:1412.6980. (see p. 31).
- Klein, S., M. Staring, K. Murphy, M. Viergever, and J. Pluim. 2010. *elastix: A Toolbox for Intensity-Based Medical Image Registration*. IEEE Transactions on Medical Imaging. (see p. 62).
- Kogel, A. J. van der, M. C. Joiner, and A. Van der Kogel. 2009. *The dose rate effect*. Basic clinical radiobiology. (see p. 11).
- Koivula, L., L. Wee, and J. Korhonen. 2016. *Feasibility of MRI-only treatment planning for proton therapy in brain and prostate cancers: Dose calculation accuracy in substitute CT images*. Medical physics. (see p. 40).
- Korhonen, J., M. Kapanen, J. Keyriläinen, T. Seppälä, and M. Tenhunen. 2014. *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. (see pp. 37, 40).
- Korsholm, M. E., L. W. Waring, and J. M. Edmund. 2014. *A criterion for the reliable use of MRI-only radiotherapy*. Radiation Oncology. (see p. 38).
- Krizhevsky, A., I. Sutskever, and G. E. Hinton. 2012. *Imagenet classification with deep convolutional neural networks*. Advances in neural information processing systems. (see p. 33).
- LeCun, Y., B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel. 1989. *Backpropagation applied to handwritten zip code recognition*. Neural computation. (see p. 34).
- Lei, Y., J. Harms, T. Wang, Y. Liu, H.-K. Shu, A. B. Jani, W. J. Curran, H. Mao, T. Liu, and X. Yang. 2019. *MRI-only based synthetic CT generation using dense cycle consistent generative adversarial networks*. Medical Physics. (see pp. 58, 68).
- Lei, Y., T. Wang, Y. Liu, K. Higgins, S. Tian, T. Liu, H. Mao, et al. 2019. *MRI-based synthetic CT generation using deep convolutional neural network*. In Medical Imaging 2019: Image Processing, edited by E. D. Angelini and B. A. Landman. SPIE. (see p. 58).
- Lemort, M., A. C. Canizares, and S. Kampouridis. 2006. *Advances in imaging head and neck tumors*. Current opinion in oncology. (see p. 14).
- Low, D. A. 2010. *Gamma dose distribution evaluation tool*. In Journal of Physics-Conference Series. (see p. 27).
- Ma, Y., F. Lacroix, M.-C. Lavallée, and L. Beaulieu. 2015. *Validation of the Oncentra Brachy Advanced Collapsed cone Engine for a commercial  $^{192}\text{Ir}$  source using heterogeneous geometries*. Brachytherapy. (see p. 80).
- Mao, X., J. Pineau, R. Keyes, and S. A. Enger. 2020. *RapidBrachyDL: rapid radiation dose calculations in brachytherapy via deep learning*. International Journal of Radiation Oncology\* Biology\* Physics. (see pp. 80, 81).
- Mao, X., Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley. 2017. *Least squares generative adversarial networks*. In Proceedings of the IEEE international conference on computer vision. (see pp. 48, 58).

- Maspero, M., M. H. Savenije, A. M. Dinkla, P. R. Seevinck, M. P. Intven, I. M. Jurgensliemk-Schulz, L. G. Kerkmeijer, and C. A. van den Berg. 2018. *Dose evaluation of fast synthetic-CT generation using a generative adversarial network for general pelvis MR-only radiotherapy*. *Physics in Medicine & Biology*. (see pp. 42, 46, 50, 51, 53, 58, 68, 70).
- Minsky, M., and S. Papert. 1969. *An introduction to computational geometry*. Cambridge tiass., HIT. (see p. 32).
- Morgan, K. Z. 1969. *Ionizing radiation: Benefits versus risks*. *Health physics*. (see p. 11).
- Nie, D., R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen. 2017. *Medical image synthesis with context-aware generative adversarial networks*. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. (see pp. 42, 46, 53).
- Njeh, C. 2008. *Tumor delineation: The weakest link in the search for accuracy in radiotherapy*. *Journal of medical physics/Association of Medical Physicists of India*. (see pp. 14, 58).
- Nyholm, T., and J. Jonsson. 2014. *Counterpoint: opportunities and challenges of a magnetic resonance imaging-only radiotherapy work flow*. In *Seminars in radiation oncology*. Elsevier. (see pp. 39, 40, 58).
- Nyholm, T., M. Nyberg, M. G. Karlsson, and M. Karlsson. 2009. *Systematisation of spatial uncertainties for comparison between a MR and a CT-based radiotherapy workflow for prostate treatments*. *Radiation oncology*. (see pp. 14, 58).
- Nyholm, T., S. Svensson, S. Andersson, J. Jonsson, M. Sohlin, C. Gustafsson, E. Kjellén, K. Söderström, P. Albertsson, L. Blomqvist, et al. 2018. *MR and CT data with multiobserver delineations of organs in the pelvic area—Part of the Gold Atlas project*. *Medical physics*. (see pp. 43, 48).
- Pasquier, D., N. Betrouni, M. Vermandel, T. Lacornerie, E. Lartigau, and J. Rousseau. 2006. *MRI alone simulation for conformal radiation therapy of prostate cancer: technical aspects*. In *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE. IEEE*. (see pp. 37, 38).
- Pasquier, D., T. Lacornerie, M. Vermandel, J. Rousseau, E. Lartigau, and N. Betrouni. 2007. *Automatic segmentation of pelvic structures from magnetic resonance images for prostate cancer radiotherapy*. *International Journal of Radiation Oncology\* Biology\* Physics*. (see p. 38).
- Prabhakar, R., K. Haresh, T. Ganesh, R. Joshi, P. Julka, G. Rath, et al. 2007. *Comparison of computed tomography and magnetic resonance based target volume in brain tumors*. *Journal of cancer research and therapeutics*. (see p. 14).
- Rasch, C., I. Barillot, P. Remeijer, A. Touw, M. van Herk, and J. V. Lebesque. 1999. *Definition of the prostate in CT and MRI: a multi-observer study*. *International Journal of Radiation Oncology\* Biology\* Physics*. (see p. 14).
- Reichert, I. L., M. D. Robson, P. D. Gatehouse, T. He, K. E. Chappell, J. Holmes, S. Girgis, and G. M. Bydder. 2005. *Magnetic resonance imaging of cortical bone with ultrashort TE pulse sequences*. *Magnetic resonance imaging*. (see p. 39).
- Reynaert, N., B. Demol, M. Charoy, S. Bouchoucha, F. Crop, A. Wagner, T. Lacornerie, F. Dubus, E. Rault, P. Comte, et al. 2016. *Clinical implementation of a Monte Carlo based treatment plan QA platform for validation of Cyberknife and Tomotherapy treatments*. *Physica Medica*. (see pp. 21, 78).
- Rivard, M. J., B. M. Coursey, L. A. DeWerd, W. F. Hanson, M. Saiful Huq, G. S. Ibbott, M. G. Mitch, R. Nath, and J. F. Williamson. 2004. *Update of AAPM Task Group No. 43 Report: A revised AAPM protocol for brachytherapy dose calculations*. *Medical physics*. (see p. 78).

- Ronneberger, O., P. Fischer, and T. Brox. 2015. *U-net: Convolutional networks for biomedical image segmentation*. In International Conference on Medical image computing and computer-assisted intervention. Springer. (see p. 41).
- Roy, S., J. A. Butman, and D. L. Pham. 2017. *Synthesizing CT from Ultrashort Echo-Time MR Images via Convolutional Neural Networks*. In International Workshop on Simulation and Synthesis in Medical Imaging. Springer. (see p. 41).
- Salvat, F., J. M. Fernández-Varea, J. Sempau, and J. Mazurier. 1999. *Practical aspects of Monte Carlo simulation of charged particle transport: mixed algorithms and variance reduction techniques*. Radiation and environmental biophysics. (see p. 21).
- Siversson, C., F. Nordström, T. Nilsson, T. Nyholm, J. Jonsson, A. Gunnlaugsson, and L. E. Olsson. 2015. *MRI only prostate radiotherapy planning using the statistical decomposition algorithm*. Medical physics. (see p. 39).
- Stanescu, T., H. Jans, N. Pervez, P. Stavrev, and B. Fallone. 2008. *A study on the magnetic resonance imaging (MRI)-based radiation treatment planning of intracranial lesions*. Physics in Medicine & Biology. (see p. 38).
- Su, K.-H., L. Hu, C. Stehning, M. Helle, P. Qian, C. L. Thompson, G. C. Pereira, D. W. Jordan, K. A. Herrmann, M. Traugher, et al. 2015. *Generation of brain pseudo-CTs using an undersampled, single-acquisition UTE-mDixon pulse sequence and unsupervised clustering*. Medical physics. (see p. 39).
- Szegedy, C., W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. 2015. *Going deeper with convolutions*. In Proceedings of the IEEE conference on computer vision and pattern recognition. (see p. 41).
- Thor, M., L. Bentzen, U. V. Elstrøm, J. B. B. Petersen, and L. P. Muren. 2013. *Dose/volume-based evaluation of the accuracy of deformable image registration for the rectum and bladder*. Acta Oncologica. (see p. 58).
- Uh, J., T. E. Merchant, Y. Li, X. Li, and C. Hua. 2014. *MRI-based treatment planning with pseudo CT generated through atlas registration*. Medical physics. (see p. 39).
- Ulyanov, D., A. Vedaldi, and V. Lempitsky. 2016. *Instance Normalization: The Missing Ingredient for Fast Stylization*. arXiv:1607.08022. (see p. 72).
- Villa, M., J. Bert, A. Valeri, U. Schick, and D. Visvikis. 2021. *Fast Monte Carlo-based Inverse Planning for Prostate Brachytherapy by Using Deep Learning*. IEEE Transactions on Radiation and Plasma Medical Sciences. (see p. 80).
- Wagner, A., K. B. Boni, E. Rault, F. Crop, T. Lacornerie, D. Van Gestel, and N. Reynaert. 2020. *Integration of the M6 Cyberknife in the Moderato Monte Carlo platform and prediction of beam parameters using machine learning*. Physica Medica. (see p. 16).
- Wagner, A., F. Crop, X. Mirabel, C. Tailly, and N. Reynaert. 2017. *Use of an in-house Monte Carlo platform to assess the clinical impact of algorithm-related dose differences on DVH constraints*. Physica Medica. (see p. 78).
- Wang, T.-C., M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro. 2018. *High-resolution image synthesis and semantic manipulation with conditional gans*. In Proceedings of the IEEE conference on computer vision and pattern recognition. (see pp. 46, 47).
- Wolterink, J. M., A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum. 2017. *Deep MR to CT synthesis using unpaired data*. In International Workshop on Simulation and Synthesis in Medical Imaging. Springer. (see pp. 42, 46, 58).

- Wolterink, J. M., T. Leiner, M. A. Viergever, and I. Išgum. 2016. *Dilated convolutional neural networks for cardiovascular MR segmentation in congenital heart disease*. In *Reconstruction, Segmentation, and Analysis of Medical Images*. Springer. (see p. 41).
- Zarándy, Á., C. Rekeczky, P. Szolgay, and L. O. Chua. 2015. *Overview of CNN research: 25 years history and the current trends*. In 2015 IEEE International Symposium on Circuits and Systems (ISCAS). IEEE. (see p. 32).
- Zhu, J.-Y., T. Park, P. Isola, and A. A. Efros. 2017. *Unpaired image-to-image translation using cycle-consistent adversarial networks*. arXiv preprint. (see pp. 42, 55, 57, 58, 59, 67, 72).
- Zhu, J.-Y., R. Zhang, D. Pathak, T. Darrell, A. A. Efros, O. Wang, and E. Shechtman. 2017. *Toward Multimodal Image-to-Image Translation*. arXiv:1711.11586. (see p. 60).



# Acronyms

*CNN* convolutional neural networks. 7, 19, 29, 33–35, 41, 80, 81

*CoBra* Cooperative Brachytherapy. 15, 45, 75, 77, 78

*CT* computed tomography. 7, 12–15, 19, 25–28, 34, 37–43, 45, 57, 75, 77

*CTV* clinical target volume. 13

*DL* deep learning. 29, 42, 43, 45, 76, 78, 80

*DVHs* Dose-volume histograms. 13, 27

*FID* free induction decay. 7, 24

*FiLM* feature instance linear modulation. 76, 81

*GAN* generative adversarial net. 34, 42, 45, 57, 75–78

*GTV* gross tumor volume. 13

*HDR* higher dose rate. 80

*HU* Hounsfield unit. 7, 12, 13, 15, 25, 26, 40, 43, 77

*ICRU* International Commission on Radiation Units and Measurements. 13

*IMRT* intensity modulated radiotherapy. 11, 12, 78

*IVDT* image value to density table. 7, 26

*LDR* low dose rate. 78, 80

*Linac* linear particle accelerator. 7, 11, 14

*MAE* mean absolute error. 26, 27

*MC* Monte Carlo. 7, 21, 78, 80, 81

*ME* mean error. 26, 27

*ML* machine learning. 7, 29, 40, 77

*MLC* multi-leaf collimator. 11, 12, 21, 81

*MLP* multi-layer perceptron. 7, 31–34

*MR* magnetic resonance. 15, 16, 22, 24, 25, 37–39, 41, 42, 45, 57, 75, 77, 80

*MRI* Magnetic Resonance Imaging. 12, 14–16, 19, 22, 24, 37–43, 45, 57, 75, 76, 80

*NMR* nuclear magnetic resonance. 22–24

*OARs* organs-at-risk. 13, 78



*PET* positron emission tomography. 12, 19

*PTV* planning target volume. 13, 38

*ROIs* regions of interest. 13–15, 27

*sCT* synthetic CT. 15, 16, 19, 26–28, 34, 38–43, 45, 57, 75, 77, 80

*SGD* stochastic gradient descent. 31

*TPS* treatment planing system. 7, 13, 21, 37, 38, 78–81

*UTE* ultrashort time echo. 39–41

*VMAT* volumetric modulated arc therapy. 78

## DEEP LEARNING IN MR-GUIDED RADIOTHERAPY

Modern radiation therapy practices greatly improve treatment success while sparing healthy tissues from damages as much as possible. This is done by constantly reducing the uncertainties in the treatment planning workflow. The starting point of the treatment planning is a 3D acquisition of the patient geometry. In current workflow, we merge the benefits of two imaging modalities. Magnetic resonance imaging (MRI) which has excellent capabilities of imaging soft tissues and the Computed Tomography (CT) that gives the attenuation properties of the tissues, needed for dose calculation. The combined use of MRI and CT requires to perform a registration between these two modalities to merge their respective benefits. However, the uncertainty introduced during the registration operation is not negligible and its magnitude amounts to several millimeters. It is generally accepted that it would be very interesting to entirely devise the treatment based on MRI only. The establishment of an MRI-only workflow in radiotherapy depends on the ability to convert the MRI intensities in order to get attenuation properties. Many different methods have been proposed in the literature to solve this problem often referred to as synthetic-CT (sCT) generation. These methods recently underwent significant changes with the emergence of deep learning. Accuracy and generation speed have dramatically increased. Generative Adversarial Networks (GAN) have brought new impulses, with their ability to learn to generate any data distribution. In this thesis, we aimed to improve the generalizability of MRI-to-CT synthesis using GAN. MRI variability has been induced in the training phase by mixing images coming from different centers. This training procedure is a way to force the network to be less attentive to MRI intensities. This process has the capability to tackle the image variability problem in clinical practices, since changes can happen in image acquisition parameters or with machine replacement for instance. Two GAN architectures are proposed to either learn from paired or unpaired data.

**Keywords:** MR-guided radiotherapy, synthetic-CT, generative adversarial networks, dose calculation

## RADIOTHÉRAPIE GUIDÉE PAR IRM VIA RÉSEAUX DE NEURONES PROFONDS

Les pratiques modernes en radiothérapie améliorent considérablement le succès des traitements tout en épargnant autant que possible les tissus sains. Cela se fait en réduisant constamment les incertitudes en amont du traitement. La planification du traitement débute avec une acquisition 3D de la géométrie du patient. Actuellement dans le parcours patient, nous fusionnons les avantages de deux modalités d'imagerie. L'imagerie par résonance magnétique (IRM), qui offre d'excellentes capacités de visualisation des tissus mous, et la tomodensitométrie (TDM, en anglais computed tomography ou CT), qui donne les propriétés d'atténuation des tissus, information nécessaire au calcul de la dose. L'utilisation combinée de l'IRM et du CT nécessite un recalage entre ces deux modalités afin de fusionner leurs avantages respectifs. Cependant, l'incertitude introduite lors de cette opération est loin d'être négligeable et se chiffre à plusieurs millimètres. Il est généralement admis qu'il serait très intéressant de planifier le traitement uniquement avec l'IRM. Le développement d'un parcours patient uniquement basé sur l'IRM en radiothérapie dépend de la capacité à pouvoir corréler les intensités de l'IRM avec les propriétés d'atténuation des tissus. De nombreuses méthodes ont été proposées dans la littérature pour résoudre ce problème, souvent sous l'appellation de génération de pseudo-CT (pCT). Ces méthodes ont récemment connu des changements importants avec l'émergence de l'apprentissage profond. La précision et la vitesse de génération ont considérablement augmenté. Les réseaux adverses génératifs (GAN) ont poursuivi cet élan avec leur capacité à apprendre à générer n'importe quelle distribution de données. Dans cette thèse, nous avons cherché à améliorer la généralisation de la synthèse IRM-CT en utilisant des GAN. Le caractère variable de l'IRM a été induit en phase d'entraînement en prenant des images de différents centres. Cette procédure d'entraînement est un moyen de forcer le réseau à être moins attentif aux intensités de l'IRM. Cette méthode permet de s'affranchir du problème de la variabilité des images en routine clinique puisque des changements peuvent survenir dans les paramètres d'acquisition ou lors du remplacement d'une machine par exemple. Deux architectures de GAN sont proposées pour apprendre à partir de données appariées ou non.

**Mots clés :** radiothérapie guidée par IRM, pseudo-CT, réseaux adverses génératifs