UCLouvain
Institute for Information
and Communication Technologies,
Electronics and Applied Mathematics

Synthetic 3DCT reconstruction using fluoroscopy and convolutional neural networks for patient-specific real-time image-guided proton therapy

Estelle Loÿen

Thesis submitted in partial fulfillment
of the requirements for the degree of
Docteur en Sciences de l'ingénieur et technologie

Dissertation committee:

Prof. Benoît Macq (advisor, UCLouvain, Belgium)
Prof. Christophe De Vleeschouwer (secretary, UCLouvain, Belgium)
Prof. John Lee (UCLouvain, Belgium)
Prof. Laurent Francis (chair, UCLouvain, Belgium)
Prof. Vincent Grégoire (Centre Léon Bérard, France)
Dr. Damien Dasnoy-Sumell (UCLouvain, Belgium)
Dr. Guillaume Janssens (Ion Beam Applications, Belgium)

November 2024.

During this thesis, Estelle Loÿen was supported by a research grant from the F.R.S.-FNRS - Télévie.

Contents

Abstract

External beam radiation therapy is a standard cancer treatment that uses a source of radiation to destroy the tumor. Proton therapy uses a beam of protons to irradiate cancerous tissue. It offers a physical advantage over conventional radiotherapy thanks to the very localised dose deposition of protons within the body. This decreases the risk of side effects because the dose delivered in the surrounding healthy tissue is lower. However, it also means that it is highly vulnerable to uncertainties. A variety of geometrical uncertainties may affect the accuracy of photon and proton therapy, such as respiratory motion, tumor delineation or inter-fraction setup errors. Those inaccuracies are generally overcome by applying safety margins around the target, but larger margins result in increased irradiated healthy tissue. Modern radiation therapy is generally performed using daily image guidance to reduce the uncertainty of overall tumor targeting. However, these technologies are expensive and require the installation of new dedicated devices, not all of which is suitable for proton therapy. The majority of radiation therapy treatment rooms are currently equipped with a projection radiography system. Adaptive radiation therapy is another modern radiation therapy technique that uses imaging information acquired during treatment to re-plan the treatment plan in order to improve target coverage and reduce treatment toxicity. However, the decision to re-plan is made by the radiotherapist and is subject to inter-physician variability.

In the context of real-time tumor tracking during treatment delivery, this thesis explores the use of artificial intelligence to reconstruct a 3DCT image from a fluoroscopy image. This research is motivated by the ease of acquiring a x-rays projection in the treatment room, and the need to have a 3DCT image to compute the radiation dose deposition. The recon-

structed 3DCT image can be used for several purposes: give a feedback to the machine on the 3D positions of the tumor and internal organs, and/or to compute the radiation dose delivered to the patient. The radiation dose can either be given as feedback to the machine or be used by the radiotherapist to decide whether re-planning is necessary.

The research approach taken in this thesis can be divided into three main contributions. The first contribution implements a data augmentation tool to overcome the lack of medical data available to train and validate neural networks. The second contribution focuses on the design of a methodology for reconstructing a 3DCT image from a projection radiography using a patient-specific training of a convolutional neural network. This contribution assesses the quality of the reconstructed images using similarity metrics. The third contribution deals with the use of these images in a proton therapy treatment. To this end, the delivery of a treatment plan on reconstructed 3DCT images is simulated. In each of these last two contributions, a base case and two variants are studied. The aim of the variants is to evaluate and compare the robustness of different training methods to events that may occur in the clinic, such as a change in layout and a change in image acquisition time.

Author's list of publications

Related papers in peer-review journals

Patient-specific three-dimensional image reconstruction from a single X-ray projection using a convolutional neural network for on-line radiotherapy applications
Loÿen E., Dasnoy-Sumell D., Macq B., *Physics and Imaging in Radiation Oncology*, vol. 26, April 2023.
DOI:10.1016/j.phro.2023.100444

Deep learning-based 3DCT image reconstruction from a fluoroscopy projection, robustness to on-line and off-line anatomical variations
Loÿen E., Dasnoy-Sumell D., Macq B., *submitted to Medical Physics.*

Related papers in conference proceedings

CT reconstruction from single X-ray projection
Loÿen E., Dasnoy-Sumell D., Macq B., *4D Treatment Planning Workshop for Particle Therapy*, November 2021, Delft, The Netherlands.

3DCT reconstruction from single X-ray projection using Deep Learning
Loÿen E., Dasnoy-Sumell D., Macq B., *The European Society for Radiation and Oncology*, May 2022, Copenhagen, Denmark. Radiotherapy and Oncology 170:S171-S172.
DOI:10.1016/S0167-8140(22)02317-9

3DCT reconstruction from a single X-ray projection using convolutional

neural network
Loÿen E., Dasnoy-Sumell D., Macq B., *IEEE International Conference on Image Processing*, October 2022, Bordeaux, France. IEEE Xplore.
DOI:10.1109/ICIP46576.2022.9897902

Dosimetric evaluation of synthetic 3DCTs reconstructed from a single x-ray radiograph using CNN
Loÿen E., Dasnoy-Sumell D., Macq B., *The European Society for Radiation and Oncology*, May 2024, Glasgow, United Kingdom. Radiotherapy and Oncology 194:S3758-S3759.
DOI:10.1016/S0167-8140(24)01028-4

OpenTPS: an open-source treatment planning system to foster research and innovation
Barragán-Montero A., Dasnoy-Sumell D.,Wuyckens S., Zhao W., Peeters E., Schyns R., Huet-Dastarac M., Loÿen E., Rotsart de Hertaing G., Janssens G., Hamaide V., Deffet S., Souris K., Sterpin E., Macq B., Lee J.A., *International Conference on the use of Computers in Radiation therapy*, July 2024, Lyon, France.

Related papers on arXiv

OpenTPS – Open source treatment planning system for research in proton therapy
Wuyckens S., Dasnoy-Sumell D., Janssens G., Hamaide V., Huet M., Loÿen E., Rotsart de Hertaing G., Macq B., Sterpin E., Lee J., Souris K., and Deffet S., *arXiv preprint*, March 2023.
arXiv:2303.00365v1

List of acronyms

3DCT three-dimensional computed tomography.

4DCT four-dimensional computed tomography.

Adam adaptive moment estimation.

AI artificial intelligence.

ANN artificial neural network.

AP anterior-posterior.

ART adaptive radiation therapy.

CBCT cone-beam computed tomography.

CC cranio-caudal.

CT computed tomography.

CTV clinical target volume.

DRR digitally reconstructed radiograph.

EBRT external beam radiotherapy.

ED euclidean distance.

GAN generative adversarial network.

GTV gross tumor volume.

HU Hounsfield units.

IGRT image-guided radiation therapy.

IMAT intensity-modulated arc therapy.

IMRT intensity-modulated radiation therapy.

ITV internal target volume.

linac linear accelerator.

MAE mean absolute error.

MRI magnetic resonance imaging.

MU monitor unit.

NRMSE normalised root mean squared error.

OAR organ at risk.

PDD percentage depth dose.

PSNR peak signal-to-noise ratio.

PTV planning target volume.

ReLU rectified linear unit.

RT radiation therapy.

sCT synthetic computed tomography.

SGD stochastic gradient descent.

SOBP spread-out Bragg peak.

SPR stopping power ratio.

SSIM structural similarity index measurement.

TPS treatment planning system.

WEPL water equivalent path length.

WHO world health organization.

Acknowledgments

Aucun de nous, en agissant seul,
ne peut atteindre le succès.

Nelson Mandela

At the end of my studies in biomedical civil engineering, I was hired as a pharmaceutical consultant by Quality by Design. I thought I would start my professional career working on projects for various clients, but I quickly realised that I was missing out on research. I took the gamble of engaging into a PhD and it turned out that this suited me much better. It was a wonderful, stimulating and motivating journey, during which I had the opportunity to put my knowledge to good use on a subject close to my heart: improving cancer treatment. While it is clear that this journey required a fair amount of effort and personal commitment, it was also made possible thanks to the help of a dedicated team. I feel extremely thankful for the support I received over these four years.

First of all, I am deeply grateful to my supervisor, Benoît Macq, who believed in me 4 years ago when I was thinking of changing career direction and engaging into a PhD. I would also like to thank you, Benoît, for the freedom you gave me over the past years and for always trusting my work. I would also like to thank all the members of the jury, whose constructive suggestions and comments greatly helped to improve this manuscript. In addition, I would like to express my deepest thanks to the F.R.S.-FNRS, Fonds de la Recherche Scientifique, for funding and supporting this research. In particular, as a Télévie grant holder, I would like to thank all the generous donors who actively contribute to cancer research through their donations.

Of course, these four years would not have been so fun and instructive without all the UCLouvain colleagues. Thank you to all the PiLAB members for the good times we have had together, whether during the more formal discussions at PiLunch, or the less formal discussions during social events. In particular, I would like to thank the members of the ARIES project for the good atmosphere in our office and all (more or less) interesting conversations we have had together. Damien, thank you for your constant help over the past 4 (even 5) years, your availability to answer all my questions and your enthusiasm for my project. Gauthier, thank you for dealing with the disk and Waldo issues, which enabled me to complete this thesis without all the worries. Mélanie, thank you for your constant good humor, and for listening whenever I needed to talk. I would also like to thank Antoine and Valentin who helped me a lot a the beginning of my thesis before leaving the PiLAB to pursue their careers.

I am also particularly grateful to my colleagues Manon, Mélanie and Tiffanie, the (dream) team with whom we launched the project EP'elles, which promotes civil engineering studies among young female pupils in order to achieve a better representation of the female gender in the auditoriums of the École Polytechnique de Louvain.

Working at ELEN is pleasant thanks to administrative and technical staff. Many thanks to Patricia, Isabelle, François and Brigitte for always kindly helping in various ways ranging from booking trips to setting up the equipment needed for this thesis. Also, I would like to thank Nicolas, and the openHub, for saving me with the loan of three computers when Waldo decided to let me down six months before the end of this thesis. I would also like to thank the MIRO team for their warm welcome in their offices, and in particular Rodin who helped me a lot with the optimisation of the treatment plans using RayStation.

Last but not least. I would like to thank my friends and family for always being by my side. Maman and Papa, thank you for encouraging me to do this PhD and for always believing in my abilities. Thank you, Maman, for all the delicious dishes you have lovingly concocted over the last few weeks to give me strength at this busy time. Chloë and Martin, thank you for always supporting and pampering your little sister. Thank you also to Alex, Adèle and Aria (my amazing goddaughter) for being part of

our beautiful family. And of course, many thanks to Louis, without whom I certainly would not be putting the finishing touches to this thesis (because it would never have started). Thank you, Louis, for supporting me in moments of doubt, and for celebrating moments of joy.

One thing is certain: without all of you, I would not have produced much. Thank you all for your support and your presence.

Introduction

Cancer is a pathology characterised by a mutation in a cell leading to the uncontrollable growth of abnormal cells. This mutation is due to a genetic instability that results in a loss of cell cycle control, insensitivity to apoptosis or abnormalities in the DNA repair process. This transformation can start in almost any organ or tissue of the body, leading to a specific type of cancer. Carcinoma is the most frequently diagnosed type of cancer and results from the development of a tumor from the cells of the epithelium. Leukemia is a blood cancer caused by the growth of white blood cells. Lymphoma is the cancer affecting the lymphocytes and targets the lymph nodes, increasing their sizes. Sarcoma is the cancer arising in connective tissues such as bones, muscles, blood vessels or cartilages [SMNJ23].

Most common cancer sites differ for men and women. Among cancers diagnosed in women, breast cancer accounts for 29.4%, colorectum cancer for 12.4% and lung cancer for 9.1%. For men, prostate cancer represents 22.6%, lung cancer 13.8% and colorectum cancer 13.5% [EC]. According to World Health Organisation (WHO) in [WHOa], there are around 20 million new cases of cancer and 10 million deaths every year due to this disease, making it the second leading cause of death after heart disease [FEL$^+$]. The number of cancer cases is predicted to increase by around 60% over the next two decades, further straining health systems, people and communities. It is estimated that the global burden will reach around 30 million new cancer cases annually by 2040, with the largest increases occurring in low- and middle-income countries. However, unlike other leading causes of death, the cancer death rate continues to decline, with an overall reduction of 33% since 1991 [WHOb].

The significant drop in the cancer mortality rate is the result of many years of research, which have led to the development of several types of treatment to cure the disease. Figure I.1 highlights the different types of treatment and their main characteristics. Hormone therapy stops or slows the development of cancer cells that use hormone to grow [AS16]. Surgery removes a part of the tissue affected by cancer [Ben14]. Bone marrow transplantation does not act directly against cancer. Rather, it restores the ability of the body to produce new blood cells after treatment with high doses of chemotherapy or radiation [WPZ06]. Chemotherapy administers a drug by an intravenous way to destroy cancerous cells by preventing their division and eliminating them from the whole body [ADC+23]. Targeted therapy targets proteins that control the division, growth and spread of the cancer cells [Shu22]. Radiation therapy irradiates the tumor to destroy the DNA of cancer cells and stop their proliferation [BLYY12]. Immunotherapy acts on the immune system to fight cancer [LYZ+22]. Surgery and radiation therapy are local treatments used to target a specific part of the body, whereas other treatments are systemic and affect the whole body. The most commonly used treatment methods are surgery, chemotherapy and radiation therapy, with usually a combination of them [atNIoH].

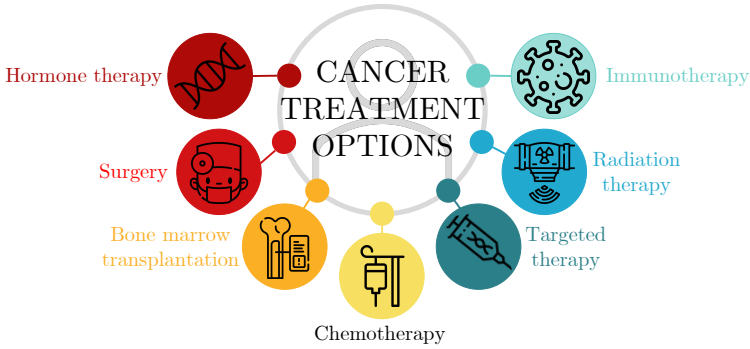


Fig. I.1 Different methods commonly used to treat cancer.

This thesis focuses on the use of radiation therapy in the treatment of lung and liver cancer. Today, around 40% of lung cancers and more than 50% of liver cancers are treated with radiation therapy, making it an indispensable component of comprehensive treatment. Besides, it offers a 3-year survival rate that is higher than surgery [CCD20]. This is made possible by medical physicists, who optimise the treatment plan in order

to deliver the prescribed dose to the target, while minimising the radiation dose delivered to organs at risk and surrounding healthy tissues. Specific strategies had to be deployed in the traditional radiotherapy workflow to consider the internal anatomical deformations generated by the breathing of the patient and ensure adequate target coverage through successive treatment sessions. These strategies are generally classified in two categories: off-line techniques and on-line techniques. Off-line techniques are all methods used before the treatment plan is delivered, whereas on-line techniques are all methods used during the delivery of the treatment plan. Among the on-line techniques, motion tracking consists in tracking the tumor movement in real-time and adapting the beam delivery accordingly in order to minimise the zone irradiated around the target. Two types of tumor monitoring are possible and available in the treatment room: direct monitoring based on imaging or indirect monitoring based on an external surrogate. Different image modalities exist for daily imaging. However, some require the installation of new, expensive dedicated devices or are not compatible with proton therapy, a particular type of radiation therapy that offers a very localised dose deposition within the body. This thesis focuses on projection radiography, which has the advantage of being already available in the majority of photon and proton therapy treatment rooms. However, this irradiating imaging modality only provides a two-dimensional image, on which it is difficult to localise the tumor without the use of a marker. Therefore, this thesis explores the use of artificial intelligence to reconstruct a 3DCT image from a fluoroscopy image.

I.1 Challenges

Although radiation therapy has been used for decades to improve the lot of patients with thoracic cancer, there are still many areas where the treatment can be improved.

Acquiring enough 3DCT images of the same patient is impossible. Computed tomography is a medical imaging modality that allows to observe the inside of the body. However, this modality uses x-rays, which have the disadvantage of being irradiating for the patient. The ALARA precautionary principle has therefore been defined to ensure that human exposure to radiation remains "as low as reasonably achievable" [HB05]. It is therefore not permitted to acquire a large number of 3DCT images of the same patient. This is a major problem for this thesis and other research works, as

all artificial intelligence algorithms require a large number of data to train and validate the neural network.

Tracking the movement of the target with real-time 3-dimensional image is limited. A number of specialised equipment have been developed in recent years to enable real-time 3D image-guided radiation therapy. However, common to all of these approaches is the need for dedicated devices to add to or replace standard-equipped radiation therapy systems. Real-time 2-dimensional image-guided radiation therapy with implanted markers has been implemented on standard-equipped linear accelerators. This is mainly based on room-mounted x-rays imagers, which enable fluoroscopy imaging [KNO+18].

Guaranteeing an unchanged anatomy of the patient during treatment is complex. Immobilising and positioning the patient on the treatment couch is essential to ensure that the dose is delivered as planned. In-room imaging technologies enable evaluation and correction of setup errors, anatomic changes related to weight loss, or internal organ motion. However, position changes are typically restricted to simple translational adjustments as most linear accelerators are not equipped with rotational adjustment systems [RBW11].

Deciding whether re-planning is necessary is difficult. The aim of adaptive radiation therapy is to adapt treatment plans to accommodate during-treatment anatomical changes due to weight loss, tumor regression and/or diminution of the volume of surrounding healthy tissues and organs at risk. The difficulties of decisions on re-planning arise from a range of factors such as uncertainty in treatment response and inter-patient heterogeneity. Clinical decisions are also primarily influenced by the professional experiences of the physician, which may result in inter-physician variability [NSJ+23].

Measuring the dose actually delivered during treatment is tricky. It is necessary to accumulate the dose delivered to the target and organs at risk as treatment progresses in order to guide decisions on treatment plan adaptation. This is achieved by non-rigidly aligning the daily cone-beam computed tomography image with the planning CT. However, non-rigid registration methods are deterministic and ignore uncertainties, which could lead to errors in the dosimetric evaluation [RBW11].

I.2 Contributions

In the previous section, five limitations of the current radiation therapy workflow were described. In this thesis, we attempted to remedy these major issues in radiation therapy through a number of contributions.

Implementation of a data augmentation tool. We propose a data augmentation tool to create a database of sufficient size for training and validating neural networks. This data augmentation tool requires the acquisition of a 4DCT and generates new 3DCT images of a patient representing intra- and inter-fractional anatomical deformations. This data augmentation tool is implemented in an open-source treatment planning system for research in proton therapy, OpenTPS[1].

Patient-specific 3DCT reconstruction from a single x-rays projection using a CNN for on-line radiotherapy applications. We propose a method able to reconstruct a 3DCT image from a single digitally reconstructed radiograph by means of patient-specific training of a convolutional neural network. Through neural network inference, this method provides a real-time volumetric image in the treatment room. In addition, this method does not require implanted markers to visualise the tumor as the computed tomography modality provides relatively good soft tissue contrast.

Robustness to changes in layout and in image acquisition time. We propose strategies to train the neural network able to counter errors in machine positioning, or to deal with during-treatment anatomical deformations.

Dosimetric evaluation of the synthetic 3DCTs. We propose a dosimetric evaluation of the synthetic 3DCTs reconstructed using the patient-specific trained neural network. This evaluation includes a simulation of the delivery of a treatment plan on the synthetic 3DCTs and provides quantitative metrics on the accumulated dose to help physicians make a decision on re-planning. The use of the synthetic 3DCT images has the major advantage of eliminating the errors caused by non-rigid registration algorithms.

[1] http://www.opentps.org/

I.3 Outline of the thesis

This thesis is divided into six chapters. The first two chapters are background chapters that lay out the theory and state-of-the-art methods. All the tools required to understand this thesis are presented in the third chapter. The next three chapters are based on peer-review articles that are either published or submitted. A list of publications is given at the beginning of the manuscript. Figure I.2 gives the organisation chart of the thesis. It highlights the different chapters with their main topic and how they are related to each other in the general context of this work.

Chapter 1 gives some background on radiation therapy. The first part of this chapter describes the radiation therapy treatment workflow. The second part presents modern radiation therapy techniques. The third part explains the characteristics of proton therapy. The last part of this chapter discusses the particularities of mobile tumors as well as the motion monitoring and mitigation techniques currently in use.

Chapter 2 gives some background on deep learning. The first part of this chapter presents the components needed to build a neural network. The second part explains the main steps involved in the training of a neural network. The last part of this chapter focuses on the use of artificial intelligence in radiation therapy and discusses the motivations, challenges and areas of application.

Chapter 3 presents the resources required to produce this thesis. The first part of this chapter defines the general context behind this work. The second part presents the medical imaging modalities involved and details the data augmentation tool developed for this thesis. This part is based on a paper published on arXiv [WDSJ+23]. The third part explains the features of the neural network. The last part of this chapter focuses on the treatment plan optimisation.

Chapter 4 details the patient-specific method able to reconstruct a 3DCT image from a single x-rays projection. Image quality metrics are used to assess the accuracy of the method. This chapter is based on a published article [LDSM23], and additional analyses are provided.

Chapter 5 discusses variants in the patient-specific 3DCT reconstruc-

tion method. This chapter is divided into two parts. The first part assesses the robustness of the method to changes in layout. This analysis is based on a conference paper [LDSM22a]. The second part assesses the robustness of the method to changes in image acquisition time. This analysis is adapted from a submitted paper.

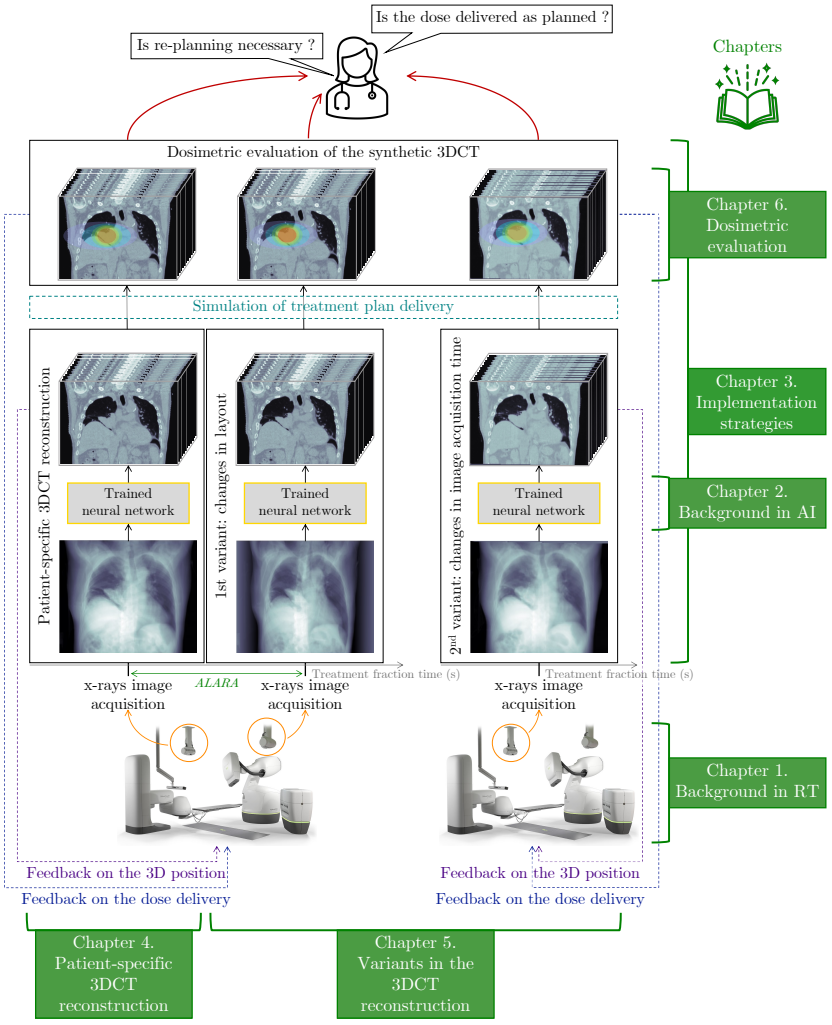


Fig. I.2 Organisation chart of the thesis. It highlights the different chapters with their main topic and how they are related to each other in the general context of this work.

Chapter 6 evaluates the dosimetric accuracy of the synthetic 3DCTs. The impact on using a synthetic 3DCT image on the estimate of the required proton energy and on the treatment plan delivery is studied. This chapter is adapted from a submitted paper.

Finally, the thesis is concluded in the last chapter with a review of the contributions to the field of radiation therapy and a discussion on the possible application scenarios.

1

An introduction to adaptive radiation therapy for mobile tumors

This chapter is inspired by the course of John Lee, Edmond Sterpin and Guillaume Janssens on proton therapy given at UCLouvain (LGBIO2070 - Engineering challenges in proton therapy) [LSJ20]. This chapter contains mainly theoretical sections, experienced authors can skip to the next chapter.

Radiation therapy (RT) is involved in about half of the cancer cases. It uses ionising radiations such as x-rays, gamma rays, electrons or protons to destroy or damage cancer cells. The ionising radiations carry enough energy to damage the genetic material of cancer cells, leading to the inability of cells to divide and proliferate. The energy deposited in tissues during those interactions is called the absorbed dose and is expressed in the unit of Gray (Gy). This unit represents the energy (Joule) absorbed per unit of mass (kg). Radiation therapy can be delivered to patients in three ways. Brachytherapy uses a radioactive source inserted in the body at the site of the tumor to destroy cancer cells. Systemic radiation uses a radioactive drug administered by infusion or orally, which circulates in the body, locates the tumor and kills cancer cells. External beam radiation

therapy (EBRT) uses a machine that directs high-energy rays from outside the body towards the tumor. The radiation therapy technique studied in this thesis is EBRT. This is the most common form of radiation therapy and involves targeting the tumor with ionising radiation while the patient is immobilised on a couch.

1.1 Treatment workflow

The external beam radiation therapy workflow is presented in figure 1.1. It aims at achieving the prescribed dose to the tumor while respecting the constraints on the organs at risk (OARs). It consists of five steps that are explained in the following sub-sections.

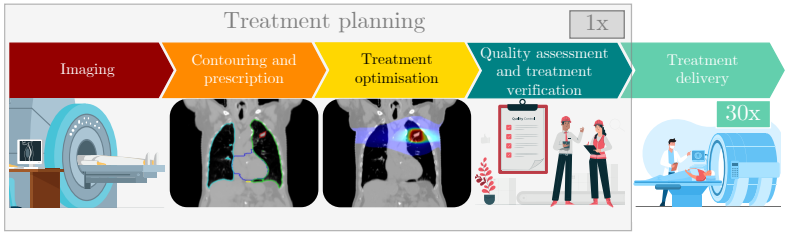


Fig. 1.1 The radiation therapy workflow is composed of 5 main steps: 1) imaging, 2) contouring of the organs and dose prescription, 3) treatment plan optimisation, 4) quality assessment and treatment plan verification, and 5) treatment plan delivery.

1.1.1 Imaging

The first step of the radiation therapy workflow consists of the acquisition of a medical image, a computed tomography (CT) scan. This image is acquired for treatment planning, and is therefore called planning CT. To produce this image, the patient is placed on the couch in the same position he will remain throughout the treatment. Some equipment may be used to support the patient in the right position, such as chest board, neck rest or arm pole.

1.1.2 Contouring and prescription

The second step of the radiation therapy workflow consists of the delineation of the target contour and organs at risk (OARs) contours on the planning CT scan by radiation oncologists. This phase is time-consuming

as it is performed by hand by physicians on a contouring software. Different target volumes encompassing the tumor are defined to account for different types of uncertainty. These are visible in figure 1.2. The gross tumor volume (GTV) is the tumor visible on the planning CT and is manually contoured by the physician. The clinical target volume (CTV) extends the GTV to include the possible infiltration of the tumor cells into surrounding tissues. If the tumor is located on a mobile organ, motion margins extend the CTV to cover the movement. The internal target volume (ITV) is the union of the CTVs defined in all breathing phases of the 4DCT scan (3D + time) acquired during the imaging step. Finally, the planning target volume (PTV) considers setup margins to include uncertainties in patient position. It extends the ITV in case of a mobile tumor, or the CTV in case of a static tumor.

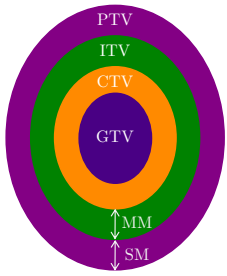


Fig. 1.2 Representation of the different target volumes. The GTV is visible on the CT scan and is manually delineated by a physician. The CTV includes tumor spread, the ITV covers motion of the tumor if located on a mobile organ and the PTV encompasses setup errors. Image adapted form [ASH+17].

Once the target volume and organs at risk have been contoured, the oncologist determines the dose to be prescribed in the target volume and the maximum permissible dose in the surrounding OARs. In general, these doses are described as inequality constraints based on the mean, the maximum or the minimum dose in the volumes of interest. The dose prescribed by the radiotherapist depends on the type and stage of the cancer to be treated.

Besides, it has been shown that healthy tissue has a greater capacity for regeneration than tumor cells and that cell survival is lower when a dose is delivered in a single irradiation session than when the same dose is delivered in several irradiation sessions. In order to preserve healthy tissue,

the dose is delivered to the patient in several sessions, called treatment fractions. The number of fractions required to deliver the treatment is also decided in this step of the workflow [HMZ14]. A typical prescription for lung cancer is given in table 1.1. Thanks to advances in techniques for controlling the position of the tumor, it is increasingly common to prescribe hypo-fractionated treatments in which the number of scheduled sessions is significantly reduced (down to around 5 sessions) and the dose delivered per fraction is much higher (7 to 20 Gy) [IWH+20, PDR22].

Table 1.1 Example of a radiation therapy prescription for lung cancer.

Prescribed dose	60 Gy
Number of fractions	30 fractions
Staggering treatment	6 weeks
Number of sessions per week	5
Dose per fraction	2 Gy

1.1.3 Treatment optimisation

The third step of the radiation therapy workflow consists in optimising the treatment plan to ensure optimum dose distribution in accordance with the prescription given in the previous step. Optimising radiation therapy treatment is a multi-criteria problem, since it involves balancing the dose between the tumor and the neighbouring organs in order to obtain the best possible quality of life for patients. This problem requires an individual solution for each patient because the anatomy of the patient is unique [BCvH19]. Figure 1.3 shows the numerical decomposition of the radiation therapy problem.

In this figure, the beam reaches the multi-leaf collimator and is then discretised into beamlets. These beams are the decision variables in the radiation therapy optimisation problem. The numerical value of these variables is the intensity of the beam after passing through the grid. The intensity of the beam is then converted into a dose when a numerical value is measured in the patient. The dose received by the patient is related to the beam intensity by a linear relationship:

$$d = d(x) = Ax \tag{1.1}$$

In equation 1.1, d is the dose per voxel, x is the beam intensity and A is

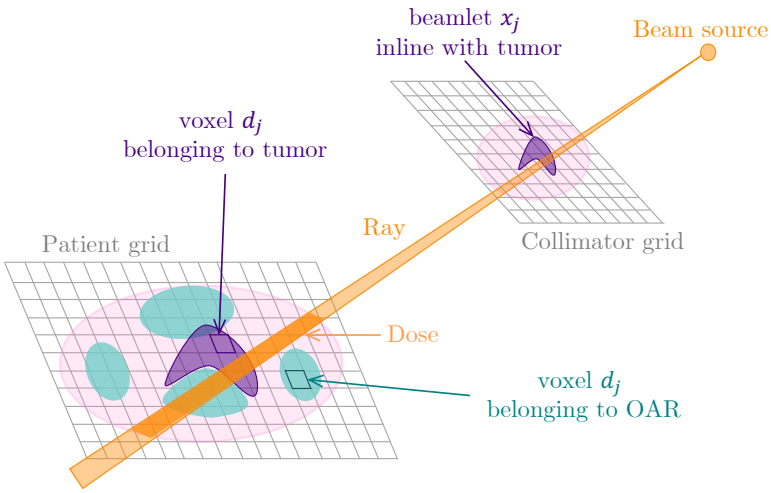


Fig. 1.3 Decomposition of the radiation therapy optimisation problem. Top part: ionising radiations are emitted by a beam source, passing through a modulation device. The beam is divided into beamlets. Bottom part: the patient is divided into voxels. Intensity modulation is possible by delivering multiple shape or different beamlets exposure times. Image adapted from [BCvH19].

the dose-fluence matrix. The matrix can be calculated by algorithms using scanner images and the beam position. Instead of optimising the dose over the entire scanned volume of the patient, the dose is divided into volumes of interest and the dose for each volume is optimised separately.

Consequently, the optimisation problem can be written as follows:

$$\begin{aligned} &\text{minimise}_x\ f(d_1) \\ &\text{subject to}\ g_1(d_2) \leq b_1 \\ &\qquad\qquad g_2(d_2) \geq b_2 \\ &\qquad\qquad g_3(d_3) \leq b_3 \\ &\qquad\qquad g_4(x) \leq b_4 \\ &\qquad\qquad x \geq 0 \\ &\text{where}\ d_1 = A_1 x \\ &\qquad\qquad d_2 = A_2 x \\ &\qquad\qquad d_3 = A_3 x \end{aligned} \tag{1.2}$$

In equations 1.2, d_1, d_2 and d_3 are the doses delivered in three different organs, while $f()$ and $g()$ are two cost functions. This problem involves minimising the dose delivered to organ_1 as a function of $f()$. This minimisation must consider that the dose to organ_2 has to lie between two constraints. A third structure is also considered in this optimisation problem. The constraint $x \geq 0$ ensures that the beam intensity remains positive.

In practice, the contours of the organs and the constraints imposed on them are given as input to a treatment planning system (TPS) which starts the optimisation process using physical and mathematical tools. The result is a treatment plan, i.e. machine parameters settings (energy levels, number of beams, beams angles, etc.) required to deliver a dose to the target sufficiently close to the prescribed dose and a dose to OAR below the maximum value authorised by the radiotherapist. The first plan produced by the TPS is rarely immediately accepted by the medical doctor. The constraints then have to be relaxed and the optimisation process repeated. It is therefore important to be able to compare different treatment plans.

1.1.4 Quality assessment and treatment verification

The fourth step of the radiation therapy workflow consists of two main stages. The physicist ensures that the machine delivers the right dose using a phantom specifically designed to mimic human anatomy, while the oncologist checks that the treatment plan meets all the necessary requirements before validating it. Physicians use several tools to assess the quality of a treatment plan. The dose-volume histogram (DVH) is a technique for quickly assessing the quality of a treatment plan and an example is represented in figure 1.4.

A DVH is a cumulative histogram of the radiation dose received in a volume of interest. It is represented by a curve, with the x-axis being the value of the dose and the y-axis the percentage of the volume receiving this dose value. In the dose-volume histogram, each curve represents an organ and reflects the proportion of its volume that receives that amount of dose. As figure 1.4 shows, the curves should remain as close as possible to the left for organs at risk, with the maximum volume receiving a zero dose. For the target volume, the curve should be oriented to the right, with a steep downward slope at the prescribed dose. A more quantitative assessment of the quality of the treatment plan is carried out using DVH metrics. The

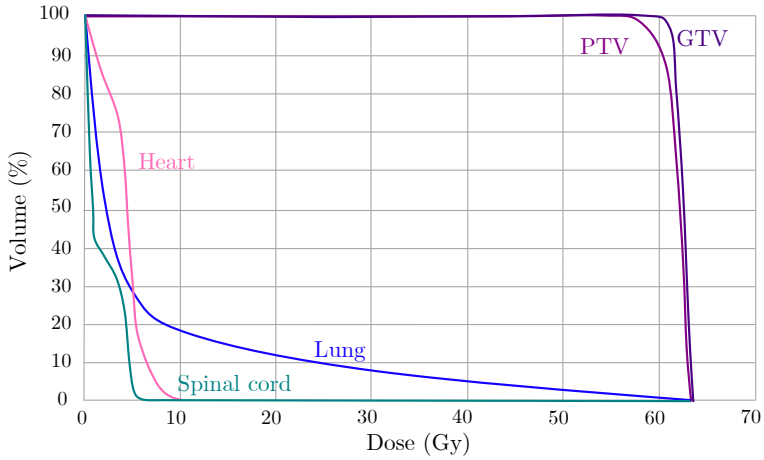


Fig. 1.4 Example of a dose-volume histogram. This tool is commonly used by radiotherapists to ensure that all constraints on the volumes of interest are satisfied.

metrics generally used are the mean, the maximum and the minimum dose in a given volume, but also the dose received in a percentage of the volume. For example, the metric $D_{95\%}$ is defined as the minimum dose received in at least 95% of the volume. The $D_{95\%}$ and $D_{5\%}$ metrics are commonly used to compare and validate treatment plans.

1.1.5 Treatment delivery

The last step of the radiation therapy workflow consists in administering the treatment plan to the patient. The treatment plan generally only contains instructions on the spot location (position and energy) and intensity (weight). The machine takes charge of the order in which the spots are shot. Usually, the spots are delivered layer by layer, from highest to lowest energy, and the lateral movement follows a serpentine pattern, as shown in figure 1.5. This method is called pencil beam scanning. The treatment plan is administered to the patient in several fractions to take advantage from the fact that healthy tissues recover faster from radiation than cancerous cells. In general, the patient has around 5 sessions per week during 5 to 8 weeks. The treatment plan can be adapted if the anatomy of the patient changes too much, which would induce a significant dosimetric change.

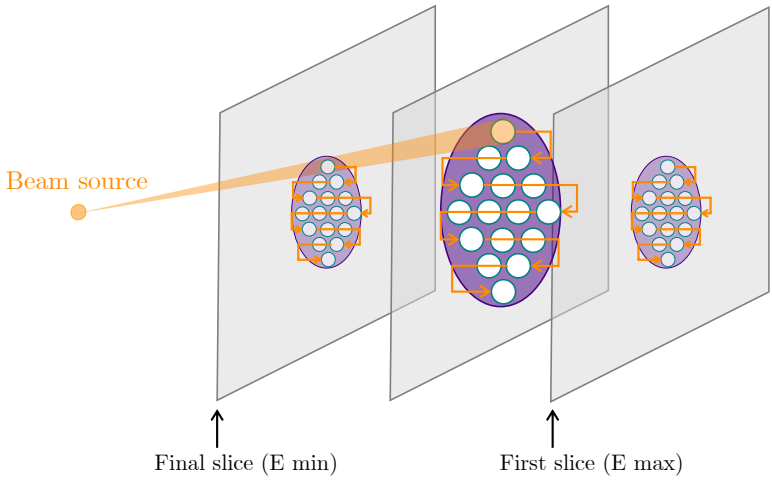


Fig. 1.5 Example of pencil beam scanning, a discrete spots delivery system.

1.2 Modern radiation therapy techniques

Over the last few decades, technological advances have enabled the development of new radiation therapy techniques, making treatment more effective by reducing safety margins while increasing the dose to the target volume. The techniques developed and commonly used in clinics are described in the following sub-sections.

1.2.1 Intensity-modulated radiation therapy

Intensity-modulated radiation therapy (IMRT) is a type of radiation therapy that uses a linear accelerator (linac) to deliver high-precision radiation therapy by shaping the radiation beam to closely fit the shape of the tumor. To this end, the linear accelerator is equipped with a device called a multi-leaf collimator, made up of thin leaves that move independently and are able to form shapes that precisely match the treatment area. This means that the tumor receives a high radiation dose, while nearby healthy tissues receive a much lower dose. This type of radiation therapy allows the dose to be shaped to the tumor by modulating the intensity of the radiation beam. The main advantage of this technique is that it creates a concave treatment zone. This avoids administering high radiation dose to structures that would otherwise be damaged by radiation therapy, thereby

reducing the risk of long term side effects [NDW00].

1.2.2 Volumetric-modulated arc therapy

Volumetric-modulated arc therapy (VMAT) is an evolution of IMRT. The linear accelerator rotates around the patient during irradiation at the same time as the leaves of the collimator move. The machine continuously reshapes and changes the intensity of the radiation beam as it moves around the body. The advantages of this radiation therapy technique are that it is very accurate, it shortens the treatment time and it uses a lower overall dose of radiation [Ott08, WYK+09, TEL+10].

1.2.3 Image-guided radiation therapy

Image-guided radiation therapy (IGRT) was defined by van Herk in 2007 as *"Increasing the precision by frequently imaging the target and/or healthy tissues just before treatment and acting on these images to adapt the treatment"* [vH07]. The key component of any image-guided radiation therapy device is an image acquisition system that provides good soft tissue contrast and/or adequate imaging of a fiducial marker. To be useful, the imaging system must be in a calibrated position and have a high acquisition and processing speed. There are a number of different IGRT techniques. A non-integrated option uses a CT scanner outside the treatment room. Integrated options in the treatment room use x-rays imagining, ultrasound, etc.

1.3 Characteristics of proton therapy

In most cases of external beam radiation therapy, x-rays are the ionising radiation type used. The photon beam is produced using a linear accelerator. Radiation therapy involving x-rays is referred as radiotherapy. Another radiation modality is particle therapy, which employs a beam of protons or heavier ions to treat cancer cells. Proton therapy is by far the most widely used particle therapy, accounting for 86% of such treatment, while carbon ions account for the remaining 14% [Dur24]. However, recent estimates suggest that only 1.6% of all external beam radiation therapy treatments in the United States are delivered with proton therapy [HSG+24].

Proton therapy involves using high-energy protons to irradiate cancer cells. The key difference between photons and protons lies in the way they

interact with tissues. Figure 1.6 displays the percentage depth dose curve of both ionisation types, which relates the absorbed dose deposited by a radiation beam in a medium as a function of depth along the beam direction. A photon beam delivers the highest dose at a depth of $1-3$ cm and then slowly decreases with depth. Rather, a proton beam shows a small dose deposition increasing with depth, a sharp increase at a certain depth followed by an extremely sharp fall off to zero. Protons lose energy and slow down as they move through tissues due to atomic and nuclear interactions. As they slow down, there are more and more interactions with orbiting electrons, resulting in a maximum energy release at the end of their range, called the Bragg peak.

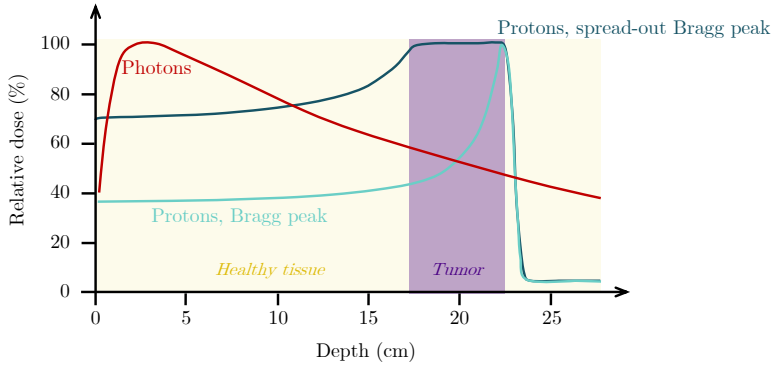


Fig. 1.6 Percentage depth dose curve of photons and protons.

The depth of the Bragg peak depends on the energy of the proton beam, meaning that the beam can be focused accurately on the area to be irradiated. The ideal scenario is a low dose deposition in front of the tumor, a high dose deposition in the tumor, and a minimal dose after the tumor. To cover the entire tumor, proton beams of different energies must be superposed by either passive scattering or spot scanning techniques, resulting in the so-called spread-out Bragg peak (SOBP). Both techniques result in a much greater dose deposition in front of the tumor and a similar fall-off after. Nevertheless, protons have the potential to be more precise than photons and are therefore better suited for targeting small tumors or tumors close to organs at risk.

Despite the physical advantage of protons, their precision is accompa-

nied by a high degree of vulnerability to uncertainties. In particular, the uncertainty of the proton range (determined by the position of the Bragg peak in the body) can lead to a deterioration in the quality of the treatment. The position of the Bragg peak depends on the energy of the proton beam and on the density and composition of the tissues across its path. If the density of tissues changes during treatment, e.g. due to breathing motion or weight gain/loss, the Bragg peak is moved to a different position from that originally expected. This can result in a missed target and an unwanted high dose in healthy tissues, leading to an overall deterioration in the optimised dose distribution. Uncertainties related to tumor position and tissue density are also present in photon therapy, but increasing the target volume with safety margins around tumors can mitigate dose deterioration. However, in proton therapy, the simple margin approach is sometimes not sufficient and more complex techniques are required.

1.4 Characteristics of mobile tumors

This work focuses on the treatment of mobile tumors, particularly lung and liver cancers. Lung cancer is the leading cause of cancer deaths, accounting for around 1.8 million deaths (18%) in 2020 [WHOb]. The movement of organs due to breathing is a source of uncertainties in the treatment of thoracic cancers [Goi04]. Tumors located in the lung or in the liver move with respiration, resulting in ballistic uncertainties. The potential risks are under-dosage of the tumor and over-dosage of surrounding healthy tissues. In order to better understand the problem of movement in radiation therapy, this section defines the directions of movement, the types of movement and the effects of respiratory movement in radiation therapy, as well as the techniques used to control the motion.

1.4.1 Definition of movement directions

Before going on to describe the different types of movement, it is important to define the terms used to describe the directions of movement. The anterior-posterior (AP) direction corresponds to an axis running from the front (chest) to the rear (back). The cranio-caudal (CC) direction corresponds to an axis running from the head to the foot. The left-right direction corresponds to an axis running from the left to the right. Figure 1.7 illustrates these designations.

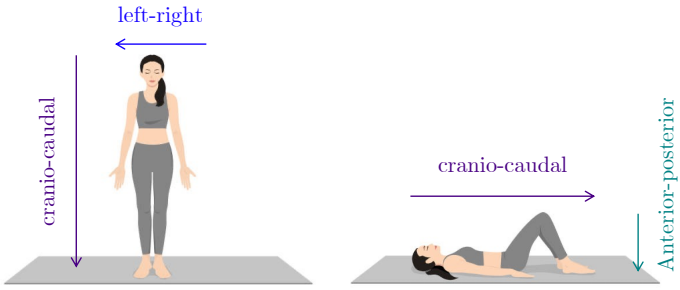


Fig. 1.7 Illustration of the three main directions of movement and definition of the terms used to refer to them.

1.4.2 Types of movement

The breathing motion is not the only movement making the treatment of mobile tumors more difficult. Two types of movement with different motion timings need to be considered and addressed adequately to avoid deteriorating the quality of the treatment.

Intra-fraction motion is the movement occurring on a small time-scale, in the scale of seconds during one treatment fraction. Intra-fraction motion includes periodic movements that are repeated many times during a single treatment session. The cardiac cycle contributes to these periodic movements, on a small spatial-scale and at a high frequency. The respiratory cycle is primarily responsible for intra-fraction motion. The respiratory cycle comprises two phases: inhalation and exhalation. During the inhalation phase, the diaphragm contracts, causing it to move downwards in the cranio-caudal direction. This enlarges the rib cage, causing the thorax to expand in the anterior-posterior direction. Through the pleura, this causes the lungs to expand and fill with air. During the exhalation phase, the diaphragm relaxes, causing it to be pushed upwards by intra-abdominal pressure. The rib cage and the lungs return to their initial positions. These movements cause the tumor to move in all three directions. Figure 1.8 shows the inhalation and exhalation mechanisms.

At rest, a healthy adult needs 12 to 15 respiratory cycles per minute to supply oxygen and eliminate carbon dioxide. However, in the presence of pulmonary pathologies, this frequency may increase. Moreover, it may

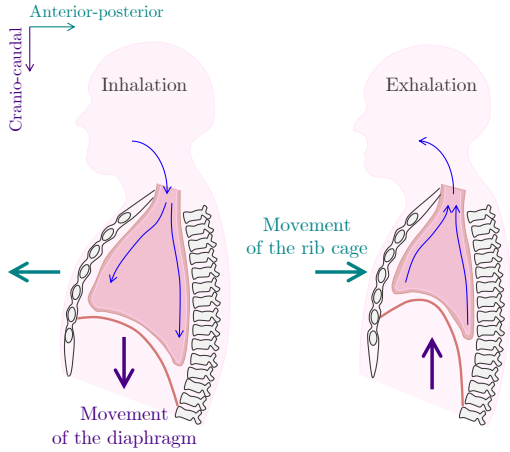


Fig. 1.8 Inhalation and exhalation mechanisms.

happen that the patient has an irregular breathing pattern (apnea or cough) during a treatment fraction, altering the periodicity of the movement. All this means that the respiratory movement is complex and variable from one patient to another and from one cycle to another, complicating the automation to predict the motion.

Inter-fraction motion is the movement happening on a large time-scale, over a period of several days or weeks. For instance, this can be due to a displacement of some organs, a change in anatomy due to weight loss or gain, or an expansion or shrinkage of the tumor size as a result of the dose already delivered within the tissue. A baseline shift is a systematic shift in the position of an organ, which means that its average position over the breathing phases is shifted. This may be caused by gastrointestinal activities such as bladder filling or peristalsis. Setup errors in patient positioning (small translation or rotation) may also be considered as inter-fraction motion. Inter-fraction motion does not strictly concern thoracic tumors.

The consequences of these movements affect both radiotherapy and proton therapy treatment quality, although it affects proton therapy to a greater extent. Two additional problems arise in proton therapy using pencil beam scanning.

Variation in proton range is the result of a variation in density along the beam due to tumor motion, shifting the expected proton range and worsening the overall dose distribution, as represented in figure 1.9.

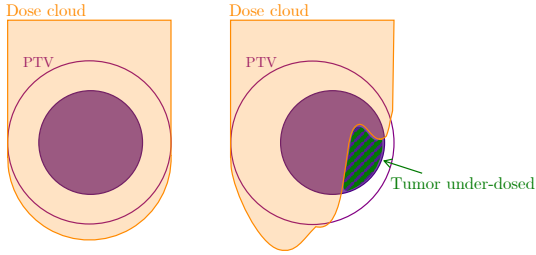


Fig. 1.9 Consequence of motion uncertainties in proton therapy. Adding PTV margins around the tumor does not guarantee coverage of the target due to density variations along the beam path.

Interplay effect is the result of the interaction between the tumor and the beam scanning motions as they occur on the same time-scale. This effect can cause hot and cold spots in the target, worsening the overall dose distribution [SRTP09]. This process is illustrated in figure 1.10.

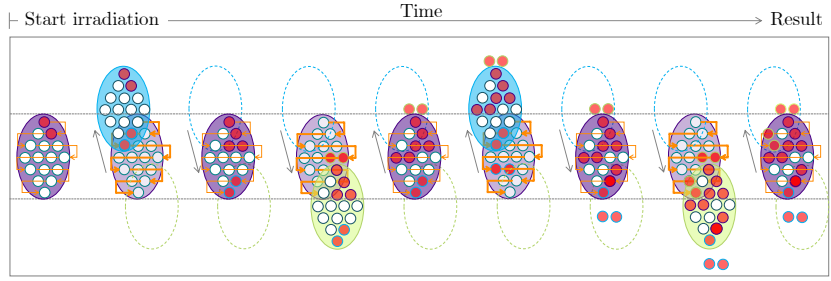


Fig. 1.10 Illustration of the interplay effect. In the presence of tumor movement, delivery of the scanning beam results in dose deterioration. During beam delivery, the target moves represented by the blue, purple and green circles, while the pencil beam is delivered at fixed coordinates. Target movement therefore alters the relative position of the pencil beam within the target. The result is a deterioration in the dose delivered.

1.4.3 Motion management techniques

This section is inspired by the review of Mori et al. on motion management in particle therapy [MKU18], as well as by the works of Keall [KMB+06] and Dhont

[DHC+20] on the management of respiratory motion in radiation oncology and image-guided radiotherapy.

Several solutions have been developed over the last few decades to manage mobile tumors and mitigate their negative dosimetric effect. Motion management techniques can be classified as off-line or on-line. Off-line techniques refer to methods that are used prior to treatment delivery, while on-line techniques refer to methods that are used during treatment delivery. These techniques are detailed in the following sub-sections and summarised in figure 1.11.

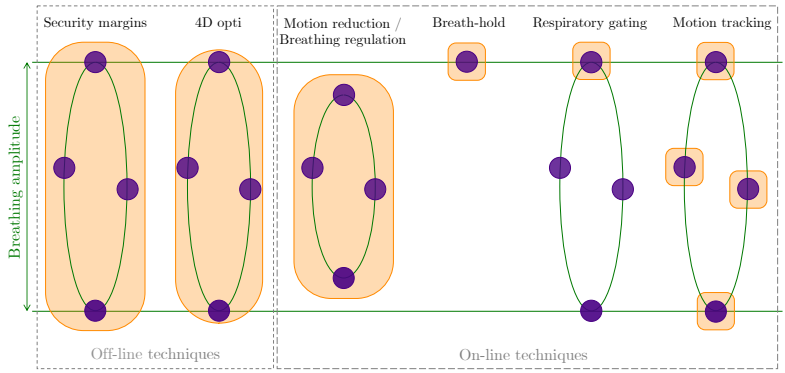


Fig. 1.11 Illustration of the different off-line and on-line respiratory motion management strategies and associated margins (in orange). The respiratory motion is shown in green, as hysteresis. This figure is adapted from [DHC+20].

Off-line motion management techniques

4D robust optimisation is the current treatment planning approach to deal with intra- and inter-fraction motions. Robust optimisation is designed to withstand small changes in anatomy between two treatment fractions by optimising the worst-case scenario. In addition to the usual 3D robustness scenarios, such as range and setup errors, 4D robust optimisation is designed to be robust to the multiple anatomical variations present in the 4DCT [LSC+16]. Although this approach is effective and current best practice for thoracic cancer treatment, two drawbacks can be formulated: the treatment is only designed to withstand the movements observed in the planning 4DCT, and the robustness to respiratory phases increases the dose received by the surrounding organs at risk [CZK+17].

Adaptive radiation therapy involves restarting the treatment planning process using a new image acquired at the start of the treatment fraction if it shows a too different anatomy from the one used for planning. The treatment plan optimised on the planning image is then deemed unsuitable by the physician and a replanning is necessary. This replanning is carried out taking into account the dose already delivered to the target during the previous treatment fractions [AMN+20, Bro19]. While adaptation has traditionally been done off-line with re-planning based on a new 3DCT image, on-line treatment adaptation based on on-board imaging has gained momentum in recent years thanks to advanced imaging techniques combined with treatment delivery systems [PBSW21].

Optimising the beam delivery parameters, such as spot size, spot spacing or beam velocity, has an impact on the interplay effect. Large spot sizes ($\sigma \approx 9 - 16$ mm) improve dose homogeneity in the target compared with small spot sizes ($\sigma \approx 2 - 4$ mm) [DGSP13]. A smaller spot spacing and a longer treatment duration improve dose homogeneity in the target and reduce the interplay effect. The sequence of spot delivery also has an impact on the dose uncertainty induced by respiratory movements. The optimal sequence of spot delivery consists in increasing the area covered by the proton beam over a certain period to reduce the fluence delivered to a given point over this period [LZZ15]. In a more recent study, Engwall et al. investigated the inclusion of uncertainties in the temporal structures of delivery within the robust optimisation [EFG18]. They considered multiple scenarios in which beam spots are distributed in the different breathing phases of the planning 4DCT. This allows to reduce the interplay effect, particularly for large tumor movements, when combined with rescanning.

Rescanning, also known as repainting, consists in delivering the dose in several iterations within a treatment fraction in order to attenuate the inhomogeneities resulting from the interplay effect by statistically averaging the motion effects. Rescanning is generally used when fractionation does not provide enough repetitions to mitigate the interplay effect [SRTP09]. There are two types of rescanning: volumetric rescanning and layered rescanning. Volumetric rescanning involves irradiating the entire volume once before the start of the next scan. In this case, the dose delivered to a point during one scan is equal to the total dose divided by the number of scans. Layered rescanning involves scanning the same iso-energy

layer several times before changing the beam energy to scan the next energy layer. In this case, the irradiation duration is limited by a maximum value. The choice between the two rescanning techniques depends on the treatment machine and the motion of the target.

On-line motion management techniques

Breathing regulation consists in regularising the breathing amplitude of the patient. One technique is mask ventilation, the patient is forced to breathe at a certain pattern [VODSL+19]. Another technique is audio-visual coaching, the patient is trained to breath in synchronisation with a sound heard through headphones [GSL+14].

Breath-hold consists in repeating breath-hold during the treatment delivery. Breath-hold can be achieved voluntary or with a computer-controlled device that can assist breath-hold via airway blocking and feedback approaches. This technique can be used for both moving targets in conjunction with gating in an attempt to decrease tumor motion, and for breast or lung cancer to maximise the distance between the tumor and the surrounding organs at risk. In this case, the patient holds his breath at the end of deep inhalation to benefit from the increase in lung volume at that precise moment. This leads to a reduction in the density of lung tissue and therefore a reduction in the percentage of healthy tissue irradiated. For an in-depth review of clinical applications, the reader can refer to [BHKSC+16].

Motion reduction consists in limiting the motion amplitude. In this case, mask ventilation can be used to force a breathing pattern of small amplitude [VODSG19]. Another method is abdominal compression, which employs a compression belt placed on the abdomen of the patient to apply a constant force and reduce the movement of the diaphragm. This technique has been shown to reduce organs movement and was proven effective in decreasing the interplay effect in liver tumors [SGK+16]. It has also been shown that this method reduces tumor movement in some cases of lung cancers [EPS+10, LSK+17]. However, the work of Bouilhol et al. in [BAR+13] has shown that abdominal compression does not provide clinically relevant improvements for patients with tumors located in the upper or middle lobes. For the patients included in this study, the use of abdominal compression reduced the amplitude of movement by an average of 3.5 mm for tumors located in the lower lobe of the lung, but only by 0.8 mm for those located in the upper and middle lobes.

Motion tracking consists in tracking the tumor movement in real-time and adapting the beam delivery accordingly. This is the optimum technique for mitigating movement, as it does not require the use of safety margins around the target and does not increase treatment time. Motion tracking techniques require continuous monitoring of the motion of the anatomy of the patient. Two types of monitoring are possible and available in the treatment room: direct monitoring based on imaging or indirect monitoring based on an external surrogate. Direct methods use an imaging device to monitor the movement of the tumor and internal organs. Imaging-based motion monitoring can be divided into two types of tracking method: fiducial marker tracking and marker-less tracking. Fiducial marker tracking involves inserting a fiducial marker on or near the tumor and extracting the 3D coordinates of the tumor using a triangulation method. Marker-less tracking has the advantage of being non-invasive and involves tracking the tumor using precise computer vision algorithms to find its position. Indirect methods use an external device, such as an optical surface imaging system, infrared reflectors or a pressure belt, to measure the displacement of the body surface. A correlation model is then required to map the movement of the skin surface and the internal movement of the tumor. While this approach has the advantage of being non-invasive, it lacks the accuracy of direct methods because tumor movement does not always correlate well with surface movement. The main direct and indirect monitoring methods are explained below and are schematically represented in figure 1.12.

X-rays imaging is the most commonly used imaging modality to monitor the motion of the tumor and internal organs. Unfortunately, soft tissue contrast on x-rays images is poor and most tumors are not easily visible. To overcome this problem, metallic implants are implanted in or around the target. The implants are tracked on the images in place of the tumor, and the position of the tumor is deduced from the position of the implants. This technique has several disadvantages: implant placement is an invasive procedure and x-rays imaging irradiates the patient, preventing its continuous use. Improving motion tracking and anatomy monitoring using x-rays imaging is one of the main motivations behind this thesis.

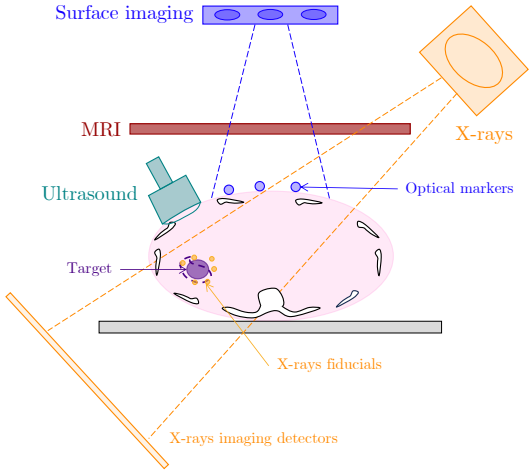


Fig. 1.12 Schematic representation of the four main imaging modalities used to continuously monitor the motion of the anatomy of the patient. This figure is inspired by [OBF+16].

The Cyberknife[1], developed by Accuray, is a device currently on the market and in clinical use with radiotherapy. The first version of this system appeared around thirty years ago [GA92]. This robot, described in numerous technical reports during the 1990s [ACM+97, AMCH99], combines the properties of both stereotaxy and tracking. This technique is capable of irradiating small thoracic lesions with a narrow beam in respiratory synchronisation. The linear accelerator is mounted on a robotised head with six degrees of freedom, enabling numerous independent targets to be reached and high-precision non co-planar treatments to be carried out. Two orthogonal x-rays images are taken to measure translational and rotational errors. These x-rays projections are compared with the digitally reconstructed radiographs generated from the planning 3DCT images. This process is very fast, enabling the treatment to be monitored and corrected according to the movements of the tumor during treatment. In the case of thoracic cancer treatment, tumor tracking is carried out using a number of optical markers detected by a synchrony camera. A motion model correlating the external movement of the patient with the movement of the tumor is integrated into the machine, providing

[1] `http://cyberknife.com`

real-time tracking of moving tumors. X-rays projections taken regularly during treatment ensure that the correlation is still valid. A full technical description of the Cyberknife system can be found in [KDK+10]. Figure 1.13 illustrates the main components of the Cyberknife system.

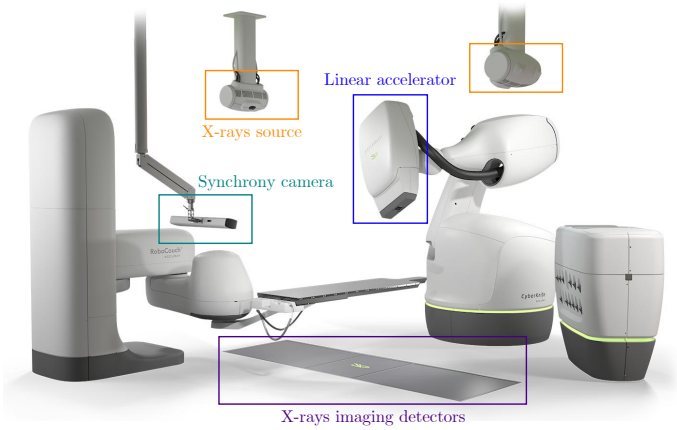


Fig. 1.13 Main elements of the Cyberknife system. It shows the camera, the x-rays sources, the linear accelerator as well as the x-rays imaging detectors.

The Heavy Ion Therapy CI-1000S[2], developed by Toshiba, is a tracking system that uses x-rays projection. This device is currently on the market and is used clinically with particle therapy.

Magnetic Resonance Imaging (MRI) is a non-irradiating imaging modality that provides excellent soft tissue contrast. MR-linac, a device combining a MRI scanner and a radiotherapy linac, is already commercially available for photon therapy. For example, the MRIdian™ system combines a radiation delivery device with a localisation software that allows tumors to be targeted and visualised on soft organs while tracking their positions and shapes during radiation delivery. The real-time analysis offered by the MRI technology allows the differentiation of tumors from other organs during treatment to prevent radiation from affecting healthy tissue [MLC+16]. However, MR-linac is not yet available for proton therapy, as this type of machine poses

[2] https://www.global.toshiba/ww/products-solutions/heavy-ion/about.html

additional technical problems due to the effect of the magnetic field on the particle beam.

Surface imaging is a non-irradiating and non-invasive technology used for the continuous localisation of the patient during the treatment fraction. This technique employs a combination of light projectors and optical cameras to generate 3D map of the topography of the patient. Surface imaging is mainly used to reduce the variability of the initial setup, to verify continuous immobilisation of the patient during treatment and to provide dynamic information about the body surface of the patient.

Ultrasound imaging is a non-irradiating modality that offers good soft tissue contrast, resolution and acquisition frequency. The main disadvantage of this modality is the difficulty to place the ultrasound probe in a reproducible manner. This one must be placed in contact with the skin with almost constant pressure so as not to cause displacement [GMO+18].

Respiratory gating consists in delivering the treatment beam at a precise moment in the periodic movement of the target during which its position is known with a high degree of confidence [LBS+07]. This method is based on the acquisition of an external surrogate signal acquired in real-time and correlated with the target motion, which controls the beam activation. The gating window can be defined on a single or multiple phases, allowing a trade-off between delivery efficiency and dose conformity [MKU18]. There are two types of gate: phase-based synchronisation and amplitude-based synchronisation. Phase-based synchronisation defines the activation window as part of the respiratory period, while amplitude-based synchronisation defines the activation window as part of the respiratory amplitude or in a specific position. Two types of external surrogate can be used to obtain information about the breathing signal: infrared reflectors and camera are used to follow the position of the reflectors placed on the skin, or a pressure belt is placed around the patient abdomen and measures the belt tension using a pressure sensor. The two main drawbacks of respiratory gating are the increase in the treatment duration and the use of an external surrogate, which requires a good correlation between its external motion and the internal movement of the tumor.

1.5 Summary

In this chapter, we set out the medical context of our research by describing the treatment of thoracic cancer by radiation therapy. We also discussed the special case of proton therapy, which enables tumors to be targeted more precisely thanks to the interactions of protons in the tissue. We addressed the issue of respiratory movement in chest cancer. This movement must be taken into account to avoid over-dosage of healthy tissue and under-dosage of the tumor. We presented the current techniques for motion management and monitoring in the last part of this chapter.

2

An introduction to deep learning for image-guided radiation therapy

This chapter is inspired by the course of John Lee and Michel Verleysen on machine learning given at UCLouvain (LELEC2870 - Machine learning: regression, deep networks and dimensionality reduction) [LV20] and by the technology review of Ana Barragán-Montero [BMJV+21]. This chapter contains mainly theoretical sections, experienced authors can skip to the next chapter.

For more than a decade, the development and adoption of artificial intelligence (AI) technologies has been accelerating. In medicine, it is used in fundamental and clinical research, hospital practice, medical examinations, care and logistics. It is contributing to the refinement of diagnoses and prognoses, even more personalised and targeted medicine, advances observation and analysis technologies, as well as surgical tools and other assistance robots. Numerous challenges specific to AI and medicine, such as the digitisation of data, respect for data privacy, algorithm explanation, design of inclusive AI systems and their repeatability, need to be overcome if hospital staff wants to have confidence in these tools. This requires a mastery of the fundamental concepts that are presented in this chapter.

2.1 Building a neural network

This work requires the use of a neural network to reconstruct a 3DCT image from a digitally reconstructed radiograph. In order to understand the neural network used in the proposed method, it is essential to master the main components of an artificial neural network.

2.1.1 Artificial neuron

An artificial neuron takes as input a vector $x \in \mathbb{R}^n$ and produces as output a scalar $y = g(w^\intercal x + b)$, where $w \in \mathbb{R}^n$ and $b \in \mathbb{R}$ are the model parameters, and g is a non-linear activation function [VDM86]. An example with $n = 2$ is illustrated in figure 2.1.

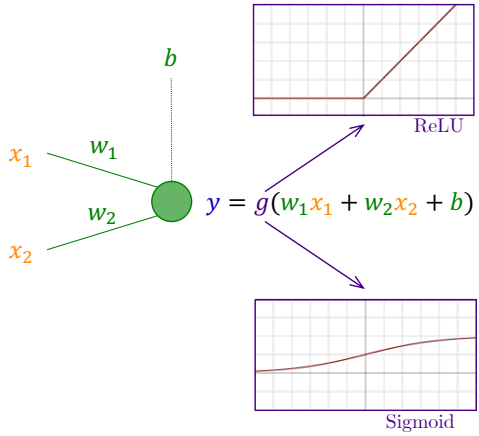


Fig. 2.1 Schematic representation of an artificial neuron. The sigmoid function and the rectified linear unit (ReLU) function are common activation functions of artificial neural networks.

2.1.2 Artificial neural network

An artificial neural network (ANN) is formed by stacking several artificial neurons together. This architecture is organised into three types of layer: input layer, hidden layer and output layer. The input layer is the first layer of the network, the output is the last, and the hidden layers correspond to the layers in-between. The non-linear activation function is used to activate or inhibit the neuron output as a function of its input value. Figure 2.2 shows the calculation of a single hidden layer with three hidden

neurons, and of an output layer with two output neurons. This ANN is fully-connected, meaning that the output of each neuron in one layer is transmitted to the input of each neuron in the next layer. The input of each neuron is then the weighted sum of the outputs of all the neurons in the previous layer [PBPPM09]. The weights w are called the hidden weights and are specific to each neuron. These hidden weights are computed during the training phase of the neural network, explained in detail in section 2.2.

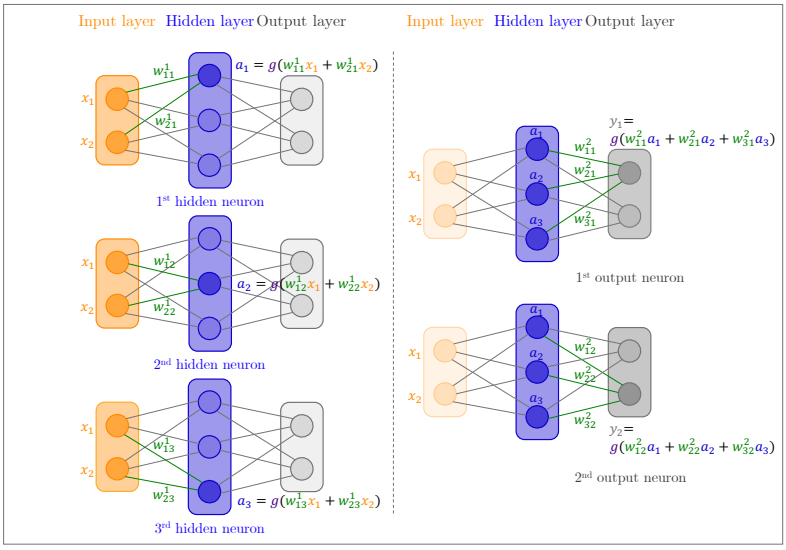


Fig. 2.2 Calculation of an artificial neural network containing an input layer composed of two neurons, one hidden layer composed of three neurons and an output layer composed of two neurons.

It has been shown that an ANN with a single hidden layer is a universal approximator, any function can be approximated by this model assuming that the hidden layer contains enough neurons [Cyb89, HSW89]. However, the number of neurons required to model a given function using a single hidden layer is often too high and the model may not generalise well. For a given number of neurons, arranging them in several hidden layers instead of a single one is often more practical and allows better generalisation. The study of these artificial neural networks with several hidden layers is known as deep learning [GBCB16]. More generally, assume a deep neural network with L hidden layers and an activation function $g^{[l]}$

at layer l whose output is governed by the following equations:

$$\begin{aligned} z^{[l]} &= W^{[l-1]} a^{[l-1]} \\ a^{[l]} &= g^{[l]}(z^{[l]}) \end{aligned} \tag{2.1}$$

for $1 \leq l \leq L$. In these equations, $a^{[0]} = x$ is the input of the network and $a^{[L]} = \hat{y}$ is the output of the network.

Activation functions

There are a large number of activation functions. Three functions are frequently used. The first function is the sigmoid defined by $g(z) = 1/(1 + \exp(1 - z))$ whose output is between 0 and 1. The second one is the hyperbolic tangent defined by $g(z) = \tanh(z)$ whose output is between -1 and 1. However, activation functions with such a small range are subject to the vanishing gradient phenomenon, which prevents the kernel weights from being updated during the training of deep neural networks. The third popular activation function is the rectified linear unit (ReLU) defined by $g(z) = \max(0, z)$. The response curves of the sigmoid and ReLU functions are depicted in figure 2.1.

Hyper-parameters

Hyper-parameters are all the parameters that are not chosen by the training process but that must be defined before it. There are two types of hyper-parameter: parameters determining the structure of the network and parameters determining the learning process. The hyper-parameters relative to the structure of the network are the number of hidden layers and neurons per layer, the initialisation of weights, the type of activation function and the dropout percentage. The hyper-parameters relative to the training process are the learning rate, the number of epochs and the batch size [Bro]. The learning rate controls the extent to which the model is modified in response to the estimated error each time the model weights are updated, the number of epochs defines the number of times the learning algorithm runs through the entire training dataset and the batch size defines the number of samples to be processed before updating the model parameters.

2.1.3 Convolutional neural network

The network presented in figure 2.2 is fully-connected, each neuron in a given layer is connected to all neurons of the next layer. Although this is

useful in many applications, it is not the most appropriate configuration when the inputs are images. Indeed, the number of connections between a 2D input image and the first hidden layer grows quadratically with the size of the input patch. If the input is a 3D image, the growth is cubic. Large input images are often required to provide sufficient contextual information, and large networks are more difficult to train. Besides, fully-connected networks lack a desirable property: spatial invariance. Small translations of the input image result in different outputs from the artificial neural network.

Convolutional layer

The convolutional layer has been proposed to exploit spatial component of an image [LBD+89]. This layer prevents the network from using pixels too far apart in the image to extract spatial features, while reducing the number of trainable weights of the network. To calculate the activation of a hidden neuron, an element-wise multiplication and a sum are calculated with a given set of weights. This set of weights is called the kernel. Figure 2.3 illustrates this principle. The kernel enables the weighted sum calculation to be limited to a particular region of the input image. Each element of the kernel can therefore be considered as the weight of the corresponding input pixel in the sum. For the next hidden neuron, the kernel is shift by a fixed number of elements, called the stride, and the process is repeated until the whole image is covered. A typical value of the stride is 1, but a larger stride can also be used, resulting in under-sampling of the output [YND18].

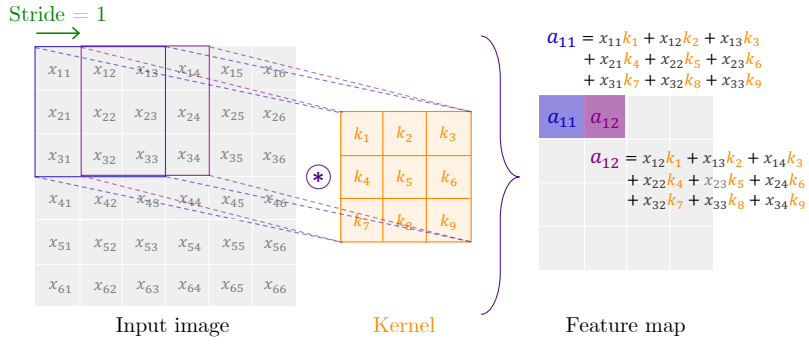


Fig. 2.3 Calculation of the two first neurons of the feature map in a convolutional layer.

In practice, different kernels are used to produce a multitude of output arrays. Hidden weights corresponding to a same kernel are considered as belonging to the same feature map. Each kernel is therefore an extractor of specific features. In a convolutional neural network, the first convolutions detect small patterns, such as edges. Intermediate layers build on these simple patterns to search for more complex ones, such as complicated curves. Deeper convolutions assemble these curves to find even more abstract objects.

Sometimes, convolution with padding is used. This means that zero entries are added around feature maps before convolution as represented in figure 2.4. The purpose of padding is to control the resolution of the feature map and to avoid a reduction in resolution.

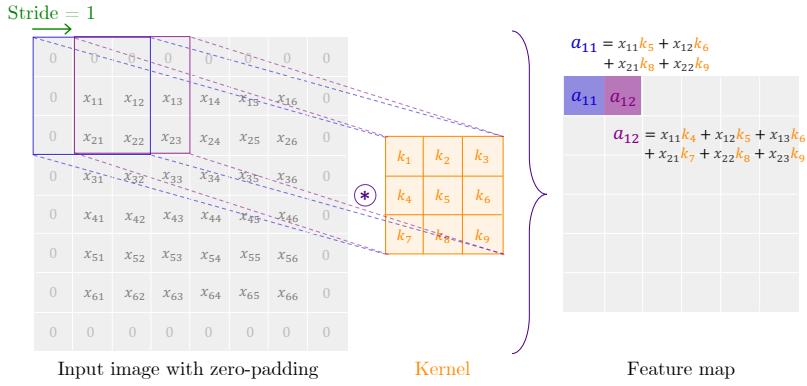


Fig. 2.4 Illustration of a convolutional layer with zero-padding. The input layer is extended with a border composed of zeros to produce a feature map of the same size than the input image.

The stride, the kernel size, the padding and the number of kernels are hyper-parameters of the neural network and are fixed before the training process.

Pooling layer

Convolutional layers use efficient weight sharing to search for patterns in images. However, they do not exhibit the property of spatial invariance. An artificial neural network composed only of convolutional layers is very sensitive to small translations in the input image as the extracted features depend on the exact positions in the image. Minor translations in the in-

put image result in a different feature map. One solution to mitigate this effect is to use a pooling layer after the convolution. This operation consists in replacing the neurons of a given region of the feature map by their maximum value in case of max-pooling or by their mean value in case of average-pooling. This introduces spatial invariance and reduces the number of weights which limits over-fitting and improves generalisation, but it also results in a loss of accuracy in feature localisation. This is graphically represented in figure 2.5.

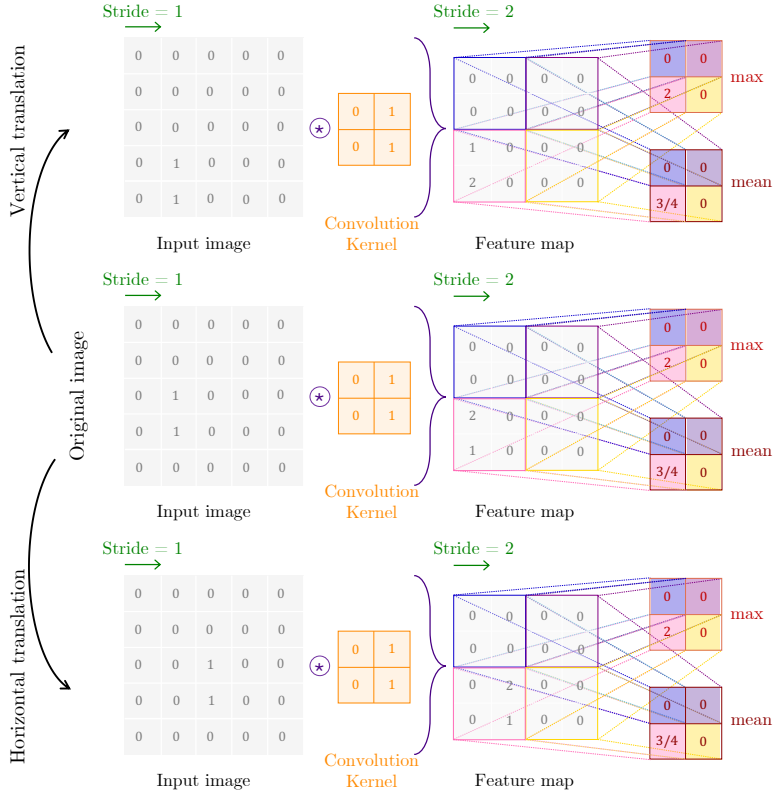


Fig. 2.5 Illustration of a pooling layer, either max-pooling or average-pooling. The output of the pooling layer is the same in all three cases, demonstrating spatial invariance with respect to small translations in the input image.

2.1.4 High resolution neural network

The convolutional layer and the pooling layer presented in the previous sub-section are the main building blocks of the neural networks that take an image as input. These elements must then be combined to perform the desired task. For image segmentation or image reconstruction tasks, high-resolution output is required [CGGS12]. For this reason, high resolution neural networks have been developed to maintain a high resolution representation throughout the network architecture.

Encoder-decoder

An encoder-decoder network, also called autoencoder, is a specific neural network used to produce high resolution output. An autoencoder consists of two parts: an encoder and a decoder. The first part of the network, the encoder, extracts the characteristics of the input to represent it in a compact state vector. This part looks like a classical CNN in the way that successive blocks use an increasing number of layers, with decreasing resolution. The second part of the network, the decoder, has a similar shape to a traditional CNN but the pooling layers are replaced by up-sampling layers to progressively extend the output of the encoder to full resolution [GBCB16].

Residual connections

Another characteristic of high resolution neural networks is the presence of residual connections. A skip connection is defined as a direct connection between the output of a contraction block and the input of an expansion block, it bypasses certain parts of the network to connect the output of certain blocks directly to the input of blocks further down the network. A skip connection allows the network to make its prediction based on a combination of low-level features and more global features. Low-level features are specific to a small group of pixels in the image and accurately located at earlier levels of the network, while global features characterise the full image and require a wider spatial context. These are not accurately located on the image due to the lower resolution of features maps at deeper levels.

2.2 Training a neural network

Training a neural network refers to determining the best set of weights for maximising its accuracy. There are several learning frameworks and strategies, but they are all based on the same technique, the gradient descent method.

2.2.1 Gradient descent

A neural network can be seen as a function $f(x, \theta)$, where x is the input and θ is a vector containing the weights and the biases of all layers. The loss function is a cost function designed to measure the dissimilarity between the network prediction and the ground-truth. Training a neural network aims to minimise the loss function over the samples of the training set. Stochastic gradient is an algorithm based on the derivatives of f with respect to θ, which aims to choose biases and weights that maximise the proximity between $f(x, \theta)$ and the true label of x. The gradient provides the neural network with a direction and an amplitude for adjusting the weights. The amplitude is scaled by the learning rate. A learning rate too small requires many iterations to converge to a minimum, while a large learning rate frequently results in a sub-optimal final set of weights. A desirable learning rate is low enough for the network to converge on something useful, yet high enough to train in a reasonable length of time. Adaptive learning rate algorithms dynamically adjust the learning rate during the training process in order to combine a rapid convergence to the minimum at the beginning of the gradient descent with a more stable behavior as the minimum is approached. This is illustrated in figure 2.6.

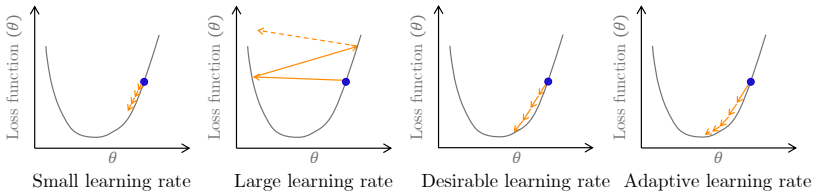


Fig. 2.6 Effect of different learning rates on the gradient descent process.

In order to reduce the cost of calculating the gradient, gradient descent is generally replaced by stochastic gradient descent (SGD). In stochastic gradient descent, the gradient is replaced by an estimate of the gradient calculated from a randomly selected subset of training data. A common extension of stochastic gradient descent is adaptive moment estimation (Adam), which involves combining gradient estimation with lower moments of the gradients [KB14]. Other notable and widely used variants of SGD include SGD with momentum [RHW86], RMSprop [ZS19] and AdaGrad [DHS11]. Readers can refer to the sources for a detailed explanation of these variations.

It is not enough to have good accuracy on the training set. The ultimate goal is to perform well on unseen data. Over-fitting occurs when the network performs well on the training set, but its performance drops dramatically on unknown data. In order to determine the optimal weights and to choose the hyper-parameters, the performance of trained models with different hyper-parameters is evaluated on another set, called the validation set [Yin19]. To obtain an accurate and unbiased approximation of the network quality on unseen data, another set is needed, the test set. This subset of data is used neither by the network for training, nor by the programmer to determine the optimal hyper-parameters of the network, but only after the training phase to evaluate the performance of the model.

2.2.2 Performance evaluation

The performance of a neural network can be evaluated in various ways. To this end, several metrics are used and depend on the task performed by the neural network.

Classification neural network

Accuracy measures the frequency with which the classifier correctly predicts. It is defined as:

$$\mathrm{A} = \frac{T_p + T_n}{T_p + T_n + F_p + F_n}$$

where T_p is the number of true positives, T_n is the number of true negatives, F_p is the number of false positives and F_n is the number of false negatives. High accuracy indicates that the network has a high number of correct predictions. Accuracy is useful when the target class is well balanced but is not a good choice for unbalanced classes.

Precision measures the relevance of positive results. It is defined as:

$$\mathrm{P} = \frac{T_p}{T_p + F_p}$$

High precision indicates that the samples assigned to the class in question have a high probability of actually belonging to that class.

Recall measures the ratio of positives detected. It is defined as:

$$\mathrm{R} = \frac{T_p}{T_p + F_n}$$

High recall indicates that the samples belonging to the class in question have a high probability of actually being assigned to that class by the model.

Precision and recall give a partial idea of the network performance. For example, consider a neural network designed to achieve binary classification. If the network predicts a true value for all pixels of the image, the recall is maximised but the precision is very low. On the other hand, if the network predicts a true value only for pixels with a very high confidence level, the precision is maximised but the recall is very low. However, even combined, these two metrics are not always sufficient.

F1 score gives a combined idea about precision and recall metrics and is maximum when precision is equal to recall. It is defined as:

$$\mathrm{F1} = \frac{2(P \times R)}{P + R}$$

Regression neural network

Mean absolute error computes the absolute difference between actual and predicted values. It is defined as:

$$\mathrm{MAE} = \frac{1}{N}\sum|y - \hat{y}|$$

where N is the number of data, y is the actual output and $\hat{y}$ is the predicted output. It has the advantage of returning a value in the same unit as the output variable.

Mean squared error computes the squared difference between actual and predicted values. It is defined as:

$$\mathrm{MSE} = \frac{1}{N}\sum(y - \hat{y})^2$$

It returns a value in the unit squared as the output variable. This metric aims to heavily penalise outliers.

Root mean squared error computes the square root of the squared difference between actual and predicted values. It is defined as:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum(y - \hat{y})^2}$$

It allows to return a value in the same unit as the output variable, but it is not that robust in case of outliers.

R2 score, also known as coefficient of determination, computes the performance of a model independently of the context. It is defined as:

$$\text{R2} = 1 - \frac{\sum(y - \hat{y})^2}{\sum(y - \bar{y})^2}$$

where $\bar{y}$ is the mean of the data. It makes it easy to compare different models. However, it is a metric difficult to interpret and it gives no information about the average error of the model.

2.2.3 Learning frameworks and strategies

Machine learning algorithms are classified into several categories depending on how the models are trained and the type of training data used [MN18].

Basic learning frameworks

Unsupervised learning algorithms work with a dataset containing many features and learn useful properties from the inherent structure of that dataset. The dataset is only composed of input data, without any associated output. This type of learning is useful to find unknown pattern in the dataset and is used for clustering, which consists in dividing a dataset into several groups that share similarities.

Supervised learning algorithms work with a dataset containing many features, but each example is also associated with a label. The model constructs a function that attempts to predict the output for each training input. It then compares these predictions with the actual desired outputs in an iterative process, and corrects its objective function accordingly to obtain predictions as close as possible to the labels.

Reinforcement learning algorithms teach an agent to evolve and interact in a specific environment on the basis of its previous experience with the aim to optimise a cumulative reward. Unlike supervised learning that uses targets to improve its predictions, reinforcement learning only needs a feedback loop between the learning system and its experiences.

Hybrid learning frameworks

Semi-supervised learning is a hybrid framework halfway between supervised and unsupervised learning. It involves data for which the desired results are only partially known. The groups identified as clusters by unsupervised learning can be used as possible class labels.

Self-supervised learning is a recent hybrid framework and can be considered halfway between supervised and unsupervised learning. The trick is to exploit the labels that can be extracted from the structure of the data itself. Self-supervised algorithms work in two steps. The first step is to pre-train the model to solve a preliminary task whose objective is to obtain supervisory signals from the data. The second step is to transfer the acquired knowledge and to fine tune the model to solve the main task.

Learning strategies

In addition to these basic and hybrid learning frameworks, there are also strategies for reusing previously trained models or for combining models.

Transfer learning reuses and refines the blocks and layers of a model that has already been trained with certain data and for a certain task, in order to apply it to other data and/or task. Transfer learning allows knowledge from different but related domains to be exploited, reducing the need for a large dataset for the target task and improving the model performance.

Ensemble learning combines the results of several models or algorithms to perform a task, improving overall performance and model stability.

2.3 Artificial intelligence in radiation therapy

This section is inspired by two reviews on the role of artificial intelligence in radiation therapy [HHG+20, LKT23].

The development of artificial intelligence in a wide range of fields has

led to growing interest in deep learning applications in the field of radiation therapy. However, the techniques developed in recent years are very rarely deployed in hospitals and traditional methods based on manual inspection of images by professionals still predominate. The section explores the motivations behind artificial intelligence in radiation therapy, its challenges and current areas of application.

2.3.1 Motivations

The interest in artificial intelligence for radiation therapy is part of the wider context of artificial intelligence for medicine. Today, doctors are under pressure to carry out numerous administrative tasks, forcing them to be more productive and to drastically reduce the average consultation time per patient. The increased pressure leads to symptoms of burnout in many doctors, which can contribute to misdiagnosis. The role of artificial intelligence would then be to reduce the number of diagnostic errors through technology, but also to reduce the number of burnout cases by reducing the workload on doctors [Top19].

One of the advantages of deep learning over manual inspection is the speed of diagnosis. Given sufficient computing power, an automated system can analyse a very large number of images simultaneously, at any one time. A computer processes data at an inhuman speed. This acceleration in diagnostic speed has been reinforced by advances in GPU-based machine learning [SKP10]. This technical advance means that real-time algorithms can support, and even automate, certain medical procedures [ZZC+19]. It also has the considerable advantage of speeding up certain tasks carried out by doctors, allowing them to devote more time to patients.

Another advantage of deep learning is the elimination of human perception bias in diagnosis. Physicians are human beings and are subject to biases, such as cognitive bias, anchoring bias or confirmation bias, all of which can lead to medical errors [CKC13]. Furthermore, an experienced physician is only exposed to a small sample of possible pathologies over the course of his or her career. An autonomous system based on deep learning can process data from many sources and have multiple examples of rare diseases that may be frequently overlooked. The artificial intelligence system could alert the practitioner to the need for further inspection or additional tests to rule out a specific pathology.

An additional advantage of image processing based on deep learning is the ability to detect objects in images that are invisible to the eye. This is made possible by the use of many filters in convolutional neural networks, which are based on correlations between surrounding structures.

2.3.2 Challenges

The use of artificial intelligence in clinics faces a number of challenges. These can be divided into two categories: technical challenges as well as ethical and legal challenges.

Technical challenges

One challenge of using artificial intelligence in medicine is the access to sufficient data. In most applications, the theoretical limit on the number of medical images available is less than what is needed to match human performance. To match or exceed human performance, the training dataset must contain at least millions of examples [GBCB16]. However, it is impossible to acquire such a database for most medical image analysis applications. Fortunately, although a ten-million-image database is often necessary in natural images to match human performance, a smaller amount of data has been shown to be sufficient for certain applications in medical image analysis. Rajpukar et al. in [RIB+18] achieve radiologist-level performance for pneumonia detection on chest x-rays using 112.120 images, and Esteva et al. in [EKN+17] achieve dermatologist-level performance for classification of skin cancer with 129.450 images. The creation of large databases is also complicated by the fact that many medical images are not stored. After acquisition, images are stored in the hospital database for a limited time and then deleted. This limits the number of images available. Moreover, image retrieval within a hospital is very time-consuming. Images can only be retrieved using a specific software, which is installed on only a few computers as it requires an expensive license. These computers are also used for clinical work, and their availability for image retrieval is therefore limited. In addition, most of these programs are not designed to extract large numbers of images, and often require many clicks of the user to extract images one by one, making it impossible to automate this step.

Another technical problem with medical data is the variation in image quality acquired by one scanner or another. A model trained on data from a single scanner is generally not generic enough to perform well on images

acquired by another machine.

An additional difficulty with medical images is the expertise required to label the data. This expertise is scarce and expensive, although essential for supervised learning. For natural images, anyone can determine the position of a person, a bike or a cat on an image. In medical imaging, expertise is rare and the availability of physicians for the annotation task is limited. Annotations take time to produce and are therefore costly. Another problem with these annotations is the lack of inter-hospital and inter-observer consensus on the labeling protocol. Different clinics or physicians give different names to the same organ or use different languages for annotation. A time-consuming pre-processing step is then essential to standardise all names.

Ethical and legal challenges

One of the main ethical challenges facing artificial intelligence in medicine relates to the data used. Medical data is personal, private and sensitive. This creates a tension between the need for more data to train and validate powerful algorithms and the obligation to preserve the privacy of personal health data. There is no good way to solve this conflict, because different cultures with different systems of beliefs lead to different choices.

Another problem is access to healthcare. Underprivileged populations currently have less access to healthcare and in consequence, the existing medical databases are biased towards the privileged populations. Moreover, the incidences of different pathologies vary between regions. As a result, underprivileged populations may have less access to algorithms adapted to their medical needs, creating an undesirable reinforcement of inequalities [NHA+21].

A major legal challenge for the application of artificial intelligence in medicine is legal liability in the event of misdiagnosis. It is difficult to assign responsibility for a medical error resulting from a faulty prediction of the algorithm. Is the error attributable to the medical doctor, the software engineer or the manufacturer of the imaging device? It is also important to determine the minimum performance required of the algorithm before deploying it on a large scale.

2.3.3 Areas of application

Artificial intelligence has the potential to improve radiation therapy for cancer patients by increasing the efficiency of the staff involved, improving the quality of treatment and providing additional clinical information and treatment response predictions to aid and improve clinical decision making. In this section, we discuss the promise of artificial intelligence to transform the treatment of cancers by radiation therapy. Examples of how AI could increase the efficiency of radiation therapy treatments and how it would integrate into the usual workflow are presented.

Initial treatment decision

The clinical workflow of radiation therapy starts with patient intake and evaluation. This step typically involves a consultation by the radiotherapist that includes a review of the symptoms of the patient, medical history, physical examination and an evaluation of the risk of toxicities from radiation therapy. On the basis of a synthesis of these data, the radiotherapist recommends a treatment plan. AI-based methods can automatically extract key clinical features from all this data to serve as a decision-support tool.

For example, Draguet et al. in [DBMCV+22] develop a fully automated clinical decision-support tool to refer a patient to either radiotherapy or proton therapy. This tool exploits recent advances in artificial intelligence applied to radiotherapy and combines them with state-of-the-art normal tissue complications probabilities models to recommend a specific treatment for each patient.

Synthetic CT generation

During the imaging step of the radiation therapy workflow, medical images are acquired for the preparation of the treatment plan. Computed tomography is the main imaging modality in radiation therapy as it provides a precise and high resolution geometry of the patient, and it enables direct conversion of the electron density needed to calculate the radiation dose. However, each acquisition of a CT image irradiates the patient. For these reasons, many research groups investigate the possibility of generating a CT image from an image of another modality.

Cone-beam computed tomography (CBCT) is used on a daily or weekly ba-

sis during the treatment for accurate patient positioning in image-guided radiation therapy. However, the inaccuracy of CT numbers prevents CBCT from performing advanced tasks such as dose calculation. Some studies focus on improving CBCT image quality for better image-guided radiation therapy, and other studies assess the validity of the synthetic image for dose calculation.

Liang et al. in [LCN+19] develop a cycle-consistent adversarial generative network to synthesise CT images from CBCT images. This model is capable of translating one image modality into the other using unpaired CT and CBCT, and unsupervised learning. The synthetic CT images generated by their model are visually and quantitatively similar to real CT images, with a mean absolute error of about 30 HU for head and neck cancer patients. They also demonstrate that dose distributions calculated on the synthetic CT images are more accurate than those calculated on the CBCT images.

Chen et al. in [CLS+20] use a U-Net architecture to take advantage of the anatomical structure of on-treatment CBCT and intensity information of the planning CT image. The synthetic CT generation U-Net model is trained using on-treatment CBCT and initial planning CT as input. The same-day CT is taken as reference. The mean absolute error is lower than 19 HU between the synthetic CT and the reference CT, while it is around 45 HU between the CBCT and the reference CT for head and neck cancer patients.

Liu et al. in [LLW+20] study the generation of a synthetic CT based on a CBCT for patients treated with radiotherapy for pancreatic cancer using a self-attention cycle generative adversarial network. The CBCT acquired before the first treatment fraction is registered on the planning CT for training and synthetic CT generation. CT-based contours and treatment plans are transferred to the CBCT and sCT of the first treatment fraction for dosimetric comparison. In the abdomen area, the mean absolute error is around 55 HU between the synthetic CT and the reference CT, while it is around 80 HU between the CBCT and the reference CT. They do not observe significant differences in dose-volume histogram metrics between CT- and sCT-based treatment plans, while significant differences are observed between CT-

and CBCT-based treatment plans.

Magnetic resonance imaging has also proved its added value in delineating tumors and organs at risk, thanks to its excellent soft-tissue contrast. To benefit from the complementary advantages offered by the different imaging modalities, MRI is generally registered with CT. However, residual registration errors and differences in patient positioning can introduce systematic errors that would affect the accuracy of the overall treatment. MRI-only radiation therapy has been proposed to eliminate residual registration error, to simplify and accelerate the workflow, and also to reduce the exposure of the patient to ionising radiation [FMD+87, LBCE+03]. The main obstacle to the introduction of MRI-only radiation therapy is the lack of tissue attenuation information required for accurate dose calculation. Various methods have therefore been proposed for converting MRI into CT equivalent representations using deep learning [BNC+21].

Han et al. in [Han17] develop a model with an architecture similar to a U-Net to generate a synthetic CT from magnetic resonance imaging. The model uses 2D slices of the 3D-MRI image to reconstruct the 2D slices of the synthetic CT. This 2D U-Net model directly learns the mapping between the 2D grey-scale MRI image and the corresponding CT slice. It is therefore necessary to predict all slices independently in order to reconstruct the whole 3DCT image.

Emami et al. in [EDNDGH18] use a generative adversarial network to generate synthetic CTs from magnetic resonance images. The generator part of the model learns the mapping between a T1-weighted MRI image as input and a real CT image, while the discriminator part of the network classifies the synthetic CT image as real or synthetic.

X-rays projections are acquired during a radiation therapy treatment fraction using the equipment available in the treatment room. However, these images only give 2D information and it is sometimes difficult to localise the tumor on them. Several studies have therefore focused on the reconstruction of a 3DCT image from these x-rays radiographs.

Montaya et al. in [MZL+21] generate 3D tomographic patient models from two-view scout images using a deep learning strategy. They also show that the 3D reconstructed patient models enable accurate organ-specific dose delivery in the subsequent CT scans.

Ying et al. in [YGM+19] propose a generative adversarial network to reconstruct a CT scan from two orthogonal x-rays. The proposed method is able to reconstruct the general structure accurately, but small anatomies still suffer from some artifacts.

Shen et al. in [SZX19] demonstrate that a deep learning model trained to map projection radiographs of a patient to the corresponding 3D anatomy can generate volumetric tomographic x-rays images of the patient from a single projection view.

Segmentation

The contouring step of the treatment workflow has a significant influence on the success of radiation therapy. However, manual segmentation is a tedious and time-consuming task for clinicians, and inter-observer variability can affect the results of the radiation therapy. These two disadvantages may be mitigated with automatic segmentation. For automatic segmentation, convolutional neural networks, and in particular U-Net, have become workhorses and produce clinically acceptable results. It is currently accepted that automatic segmentation is not to be used in an unsupervised manner, but that the correction of automatic contours saves time compared with a solely manual workflow.

For example, Baek et al. in [BHAea19] train a convolutional neural network to perform tumor segmentation, with no other information than physician contours. They show that the neural network is able to identify a rich set of survival-related image features with remarkable prognostic value. The CNN algorithm trained for tumor segmentation contains features having strong correlation with 2- and 5-year survivals.

Dong et al. in [DLW+19] propose an adversarial training strategy to train deep neural networks for the segmentation of multiple organs on thoracic CT images. The generator produces a segmentation map of multiple organs. The discriminator distinguishes between the ground-truth and the

segmented organs at risk produced by the generator. The generator and the discriminator compete against each other in an adversarial learning process to produce the optimal segmentation map of the multiple organs.

Dose prediction

During the treatment optimisation step, the medical dosimetrist creates the optimal treatment plan for the patient. The aim of this plan is to maximise the dose delivered to the tumor while sparing the surrounding organs. Treatment planing is a time-consuming process during which the medical dosimetrist optimises the dose distribution to achieve the objectives defined in the dose prescription. The quality of a treatment plan depends on a number of different human factors, such as the choice of radiation beams angles and plan optimisation parameters, leading to significant intra- and inter-institutional variations. The results of several studies demonstrate the ability of deep learning algorithms to predict the optimal entire dose distribution, which allows to steer the optimiser to directly achieve that dose distribution.

For example, Fan et al. in [FWC+19] use a residual neural network to predict a dose distribution based on the specific geometry of the patient and the prescribed dose for head-and-neck patients. The input data of the model is a CT image as well as the contours of the planning target volume and organs at risk on this image. The algorithm is trained to predict the dose distribution over the slices of the CT image. The results demonstrate that a deep learning method is able to predict clinically acceptable dose distributions as there is no statistically significant difference between the predicted and actual treatment plans for all relevant dose-volume histogram metrics.

Nguyen et al. in [NJS+19] propose a new deep learning-based dose prediction model, called Hierarchically densely connected U-Net. This new architecture is based on two popular network architectures: U-Net and DenseNet. This new architecture is able to accurately and efficiently predict the dose distribution. The model predicts, in the organs at risk, the maximal dose within 6.3% and mean dose within 5.1% of the prescription dose on the test data.

Image guidance and motion management

During the treatment delivery step, the patient is placed in the same position as the one used to create the treatment plan. Image-guided radiation therapy currently uses imaging during treatment to position the patient correctly. Artificial intelligence can be used to generate the corresponding synthetic 3DCT image, as explained in the previous sub-section on synthetic CT generation. Another challenge to consider during treatment delivery is the movement of the patient and of its internal organs. These movements can lead to over-dosage of the surrounding healthy tissues and under-dosage of the target. Current methods for motion management and mitigation aim to control and/or reduce the motion. However, there is considerable variability in motion between and within individuals in terms of amplitude and frequency, which complicates predictive modelling of the tumor motion. Artificial intelligence can be used to generate patient-specific dynamic motion management models that adapt to changes in patterns of motion in order to improve tumor tracking.

For example, Rotsart de Hertaing et al. in [RdHDSJM23] train a vision transformer network to forecast the motion of the tumor. The training of the neural network is patient-specific and the prediction error of the model is 1.30 mm in the three directions at an horizon of 1 s.

Adaptive radiation therapy

Throughout the treatment delivery, major changes can occur in the anatomy of the patient. These changes often reflect tumor shrinkage or internal anatomical variations that could potentially lead to a change in the doses delivered to the tumor and organs at risk. This may warrant re-planning. Adaptive radiation therapy involves the creation of a new treatment plan based on updated images of the anatomy of the patient. The potential use of artificial intelligence in all steps of treatment planning suggests that re-planning could be simpler and, above all, quicker. Another challenge of current adaptive radiation therapy is that the radiotherapist has to decide when the anatomical changes are significant enough to be clinically relevant, based on his or her own qualitative evaluation of the clinical parameters and images of the patient. Artificial intelligence could provide tools for predicting which patients need treatment adaptation and the ideal time to do so.

For example, Guidi et al. in [GMM+16] develop a machine-learning

classifier to analyse volume and dose variations of parotid glands, and to predict patients who would benefit from adaptive radiotherapy and re-planning intervention. A double-blind evaluation by two radiotherapists is carried out to validate the day or the week selected by the classifier for re-planning.

2.4 Summary

In this chapter, we set out the context of artificial intelligence. We began by explaining the different elements that constitute a neural network, from the simplest to the most complex. We also looked at the different strategies currently employed for training a neural network. We then discussed the motivations behind the use of artificial intelligence in radiation therapy and the challenges involved. In the last part of this chapter, we explained which stages of the clinical workflow could benefit from artificial intelligence and where research currently stands.

3

Implementation strategies

This chapter presents the implementation strategies pursued in this thesis. The first section of this chapter begins by explaining the general context of the thesis and where the implementation strategies are involved. The second section describes the data management strategy, the third section discusses the neural network architecture and the fourth section explains the treatment plan optimisation strategy.

3.1 Context

Figure 3.1 illustrates the general context of this thesis. The strategy developed in this thesis is based on a patient-specific training of a convolutional neural network that learns the mapping between a 2D x-rays projection and a 3DCT image. The purpose of the patient-specific feature is to refine the neural network to the deformations of that patient, while requiring a smaller sample of data than a generalised method. The neural network takes as input a digitally reconstructed radiograph (DRR) that simulates a daily projection radiography acquired using room-mounted x-rays imagers. The trained neural network outputs the associated 3DCT image. The 3DCT image can then be used for several purposes: give a feedback to the machine on the 3D positions of the tumor and internal organs, and/or to compute the dose delivered to the patient. This dose can either be given as feedback to the machine or be used by the radiotherapist to decide whether

re-planning is necessary.

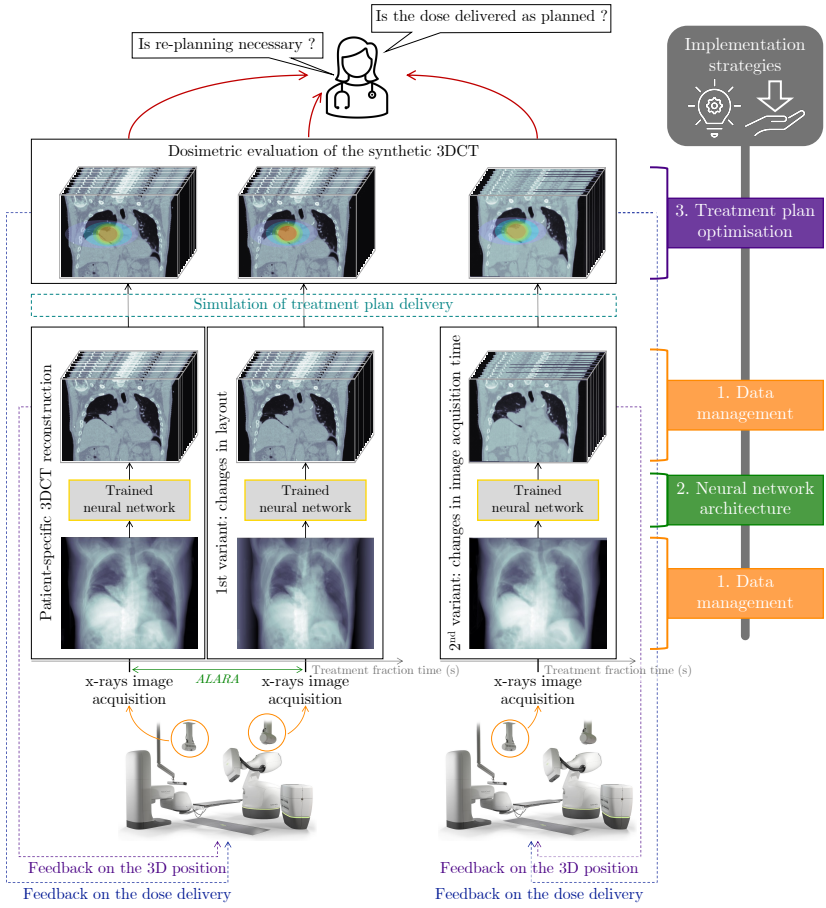


Fig. 3.1 Organisation chart of the thesis. It shows the different contributions and how they are related to each other in the general context of this work. It also highlights the implementation strategies pursued in this thesis and where they are involved.

The research approach taken in this thesis can be divided into several contributions, which are the topics of the next three chapters. The main contribution focuses on the design of a methodology for reconstructing a 3DCT image from a projection radiography using a patient-specific training of a convolutional neural network. This contribution assesses the qual-

ity of the reconstructed images using similarity metrics. The second contribution deals with the use of these images in a proton therapy treatment. To this end, the delivery of a treatment plan on reconstructed 3DCT images is simulated. In each of these two contributions, a base case and two variants are studied. The aim of the variants is to evaluate and compare the robustness of different training methods to events that may occur in the clinic, such as a change in layout and a change in image acquisition time.

To fully understand each contribution of this work, it is necessary to master the implementation strategies pursued. These are detailed independently in the following sections.

3.2 Data management

The theoretical parts of this section are based on the course of Benoît Macq, John Lee, Greet Kerckhofs and Frank Peters on medical imaging given at UCLouvain (LGBIO2050 - Medical Imaging) [MLKP20]. The last part of this section about the data augmentation tool was published on arXiv [WDSJ+23].

The medical images required for this thesis are acquired using radiography. Radiography is an imaging technique that uses x-rays, gamma rays or similar ionising and non-ionising radiation to visualise the internal structure of an object. As the body is composed of different tissues with different densities, ionising and non-ionising radiation allow to reveal the internal structure of the body on an image receptor by highlighting these differences through the attenuation or the absorption of x-rays photons. Medical radiography is generally acquired by radiographers, while image analysis is usually carried out by radiologists. Medical radiography comprises a range of modalities producing many types of image, each with a different clinical application.

3.2.1 Projection radiography

Projection radiography is the simplest form of radiography and produces two-dimensional images using x-rays. A radiograph is a two-dimensional view of the total absorption of x-rays through the body along a given axis, such that two objects placed in front of each other are superimposed on the image.

Basic principles

The system used to acquire x-rays projections is divided into several components, all of which are essential for correct image acquisition. The basic principles of this medical imaging modality are displayed in figure 3.2.

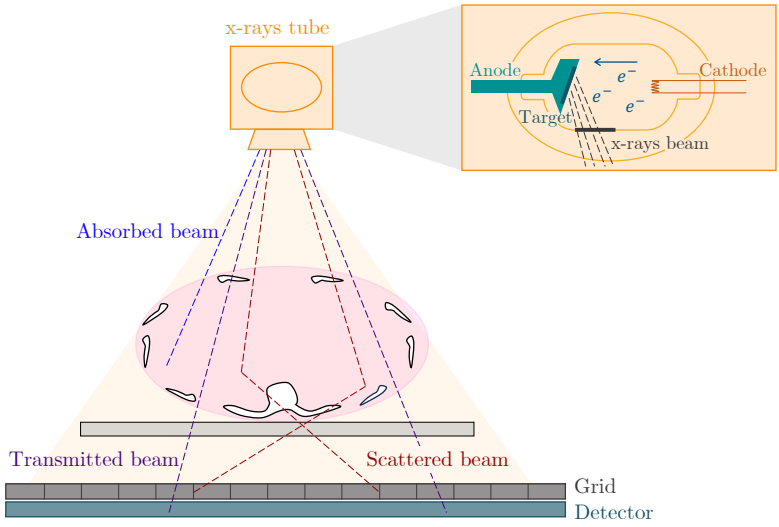


Fig. 3.2 Representation of the device used to acquire a projection radiography.

The x-rays tube consists of two electrodes placed in a vacuum envelope that can withstand high temperatures. The negative electrode is a filament that emits electrons by thermionic emission. These electrons are accelerated towards an anode by a peak voltage. X-rays are produced when highly energetic electrons interact with matter and convert their kinetic energy into electromagnetic radiation. The conversion of electrons kinetic energy into electromagnetic radiation takes place in the anode material. The electrons are deflected by a positively charged nucleus in the target and lose energy emitted in the form of Bremsstrahlung radiation. The subatomic distance between the electron and the nucleus determines the energy lost by the electron during this process: the closer the electron is to the nucleus, the more energy it loses. The probability of a direct impact of the electron with the nucleus is extremely low and lower x-rays energies are generally generated in greater quantities. The energy of the x-rays photons is equal to the energy of the incident electron and therefore depends on the peak voltage.

The x-rays emitted by the x-rays tube can be absorbed, scattered by the patient or transmitted without interaction. The transmitted photons, known as primary photons, produce the x-rays image. They provide information about the properties (thickness, density or atomic number) of the tissue passed through. Scattered photons, known as secondary photons, are not useful for imaging because they have lost their original direction and degrade the contrast of the image. In general, most secondary photons are eliminated by an anti-scatter grid placed between the patient and the detector. In computed radiography, storage phosphors are used as screens of photostimulable phosphor detectors. When a ray is absorbed by a photostimulable phosphor detector, some light is emitted rapidly as in an intensifying screen, but much of the absorbed energy is trapped in the screen and can be read out later. Photostimulable screens are made from a mixture of barium fluorohalides and are enclosed in a cassette. After x-rays exposure, the cassette is placed in a readout unit, where a red laser light scans the imaging plate and stimulates emission of the trapped energy in the form of visible light. The released light is then collected by a fiber-optic light guide and sent to a photo-multiplier tube, where it produces an electronic signal that is digitised and stored. The cassette is then exposed to bright white light to erase any residual trapped energy, and is ready for re-use.

Projection radiography in radiotherapy

Standard-equipped linear accelerators of radiation therapy contain room-mounted x-rays imagers. This enables fluoroscopy which allows real-time 2-dimensional image-guided radiation therapy with implanted markers. In-room imaging technologies enable evaluation and correction of setup errors, anatomic changes related to weight loss, or internal organ motion [Kor15, Kru18, VSV+18].

Digitally reconstructed radiograph

A digitally reconstructed radiograph (DRR) is a simulation of a conventional two-dimensional projection radiography created from a computed tomography image. Figure 3.3 illustrates the principle of simulating a digitally reconstructed radiograph and shows an example of the result. The x-rays beams are emitted from a source with a fixed initial energy. The energy of the beam decreases as it passes through the body of the patient. The energy of the attenuated beam is measured when the x-rays reach a

point on the detector, producing the DRR.

The most commonly used algorithm to compute a DRR is the Siddon algorithm. In this model, the image value at position (x, y) is the weighted average of the intensities of the voxels crossed by the beam, where the weight is the length of its intersection with the voxel:

$$DRR(x,y) = \sum_i \sum_j \sum_k \rho(i,j,k) l(i,j,k) \tag{3.1}$$

where $\rho(i,j,k)$ is the voxel intensity value and $l(i,j,k)$ is the length of the intersection between the beam and that voxel.

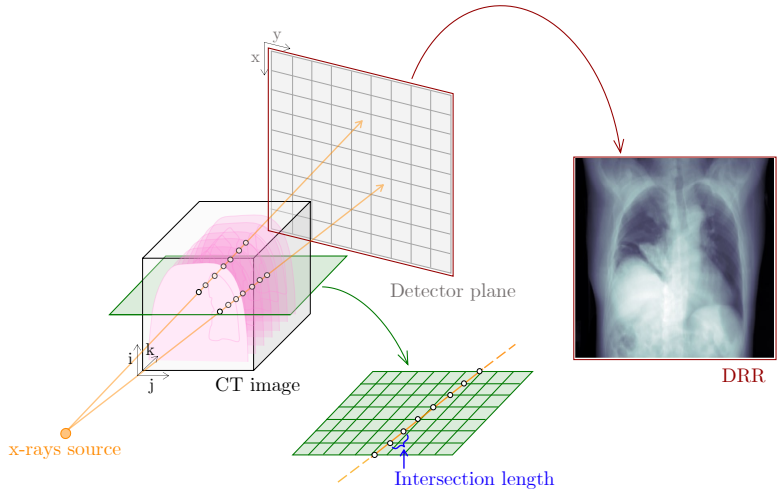


Fig. 3.3 Illustration of the DRR generation geometry.

In this work, digitally reconstructed radiographs are generated using the TomoPy Python library and a projection angle of 0° along the anterior-posterior axis [GDCXJ14]. The projection geometry is a 1440 x 1440 image with a pixel size of 0.296 x 0.296 mm^2.

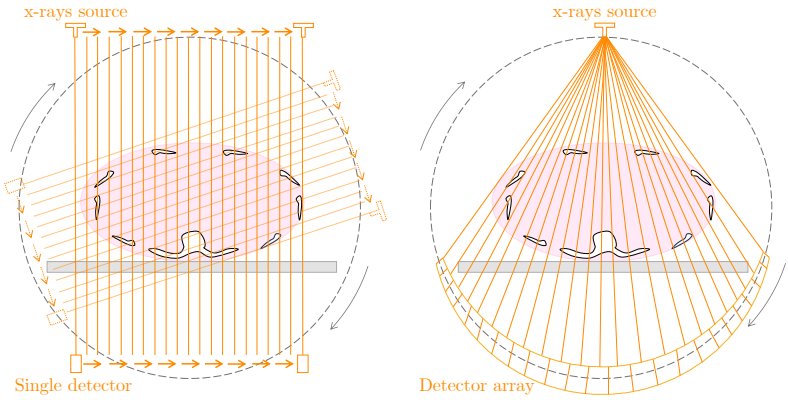
3.2.2 Computed Tomography

Computed tomography (CT) was the first imaging modality to probe the inner depths of the body, slice by slice. Since the first head CT scan in 1972, CT has gained in technical sophistication, with concomitant changes in the

quality of CT images. Imaging time has also been considerably improved, and modern computers allow images to be reconstructed in almost real-time.

Basic principles

CT scanners work by shooting multiple beams of x-rays with the machine rotating around the patient. The signals generated are then picked up by the detectors on the other side of the x-rays source and processed by a computer to generate a two-dimensional cross-sectional image, known as a slice. Figure 3.4 illustrates the two possible types of geometry. Figure 3.4(a) shows the parallel beam geometry implemented in the first CT scanners. In this case, all x-rays are parallel to each other. Figure 3.4(b) presents the fan beam geometry used in modern scanners. In this case, all x-rays diverge at a given projection angle.



(a) Parallel beam geometry (b) Fan beam geometry

Fig. 3.4 Two different projection geometries have been used in CT imaging: parallel beam geometry and fan beam geometry.

Each ray is a measure of the transmission through the patient along a line, where the detector measures the intensity of the transmitted x-rays (I_t). The intensity of the unattenuated x-rays (I_o) is also measured during the scan by a reference detector. These two intensities are machine-dependent values. For an ideal monoenergetic photon beam, I_t and I_o are

related by the following formulas:

$$\begin{aligned} I_t &= I_o . \exp(-\mu x) \\ p(E) &= -\ln(I_t / I_o) \end{aligned} \tag{3.2}$$

where x is the thickness of the patient along the ray, μ is the average linear attenuation coefficient along the same ray and p is the line integral of the attenuation coefficient μ along the ray. This calculation, which is a preprocessing step prior to reconstruction, reduces the dependency of the final image on machine-dependent parameters and explains the great clinical utility of CT images. After collecting a complete set of line integrals, a reconstructed distribution of attenuation coefficients can be obtained using filtered back-projection. The reconstructed value in each image pixel represents the linear attenuation coefficient for the corresponding tissue voxel.

Hounsfield units (HU) is the unit in which the pixels of a CT image are expressed. For a useful display of the image, the values obtained for each pixel are normalised and truncated into integer values using:

$$\mathrm{CT}(x,y) = 1000 \cdot [(\mu_t(x,y) - \mu_w)/\mu_w] \tag{3.3}$$

where $\mu_t(x,y)$ is the value obtained in pixel (x,y) before conversion and μ_w is the attenuation coefficient of water. The value of μ_w is approximately 0.197 cm^{-1} for the x-rays beam energies usually used in CT scanning. This standardisation makes it possible to obtain CT numbers ranging from -1000 HU to $+3000$ HU. Air corresponds to -1000 HU, soft tissue ranges from -300 HU to -100 HU, water is 0 HU, and dense bone or areas filled with contrast agent range up to $+3000$ HU. CT numbers are quantitative, allowing a more accurate diagnosis in certain clinical contexts. CT is also quantitative in terms of linear dimensions and can therefore be used to measure the volumes of organs of interest.

CT images generally have 12 bits of grey scale for a total of 4096 shades of grey, with CT numbers ranging from -1000 to $+3095$. However, the human eye has a limited ability to resolve relative differences in grey scale and 8 bits, 256 shades of grey, are considered sufficient for image visualisation. 12-bit CT images need to be reduced to 8-bit to suit most image display hardware. Contrast can be improved when displaying the image by windowing and levelling the CT image. The width of the window de-

termines the contrast of the image and the level corresponds to the CT number at the centre of the window. To facilitate image interpretation, the window is chosen to cover the tissue of interest.

Compared to x-rays radiography, CT scanning has significantly lower spatial resolution and better contrast resolution. The limiting spatial frequency for a CT scan is around 1 to 1.5 lp/mm, depending primarily on the size of the detector elements. The contrast resolution for a CT scan is approximately 0.5%. There are two main methods to reduce noise and improve contrast resolution: increasing the number of incident x-rays or increasing the number of photons absorbed by a pixel of the image. A compromise must clearly be found between spatial resolution and contrast resolution as the radiation dose administered to the patient must remain within acceptable limits.

Three-dimensional CT

The basic principles presented in the previous sub-section allows to generate a two-dimensional cross-sectional image. It is then necessary to translate the x-rays source in cranio-caudal direction to obtain a volumetric image of the patient. Modern scanners use multiple detector arrays to increase the axial coverage. The slices are stacked one on top of the other to form the three-dimensional CT (3DCT) image. A 3DCT image is usually composed of 150 to 180 slices. The slice thickness is determined by the detector size and is generally equal to 2 mm. Each slice is then composed of 512×512 pixels with a pixel spacing of 1.074 mm in both directions.

Digital Imaging and Communications in Medicine (DICOM) is the international standard for medical images and related information. It defines the format to be used to store medical images to enable easy data exchange and to guarantee the quality expected for clinical use. In practice, each two-dimensional cross-sectional image is saved in an independent DICOM file. 3DCT images are essential for radiation therapy as the treatment planning is based on a three-dimensional computed tomography scan.

Four-dimensional CT

Four-dimensional computed tomography (4DCT) is a type of CT scan that records multiple three-dimensional computed tomography volumes over a period of time, creating a dynamic volume dataset. A 4DCT is created by acquiring a large number of two-dimensional cross-sectional images over

a couple of minutes, while the patient breathes in a trained pattern. The slices are then sorted into several groups, called breathing phases, using an external surrogate signal. All slices corresponding to a same breathing phase are re-arranged in space to form the relevant 3DCT image.

Four-dimensional computed tomography aims to consider the motion of the tumor during treatment planning. A schematic representation of the tumor motion is displayed in figure 3.5(a). The treatment planning of a mobile tumor is based on the midpCT image. To create this image, deformation fields are computed from every breathing phase of the 4DCT to a reference phase using non-rigid registration, as shown in figure 3.5(b). This gives an average position called the mid-position. Every breathing phase is then deformed to the mid-position, where the median is computed to create the mid-position 3DCT (midpCT) image, as displayed in figure 3.5(c). The midpCT image represents the average position over the breathing cycle. This image is not blurred as the mean is directly performed on the deformation fields. On top of this, the median over all deformed breathing phases allows to remove noise and reconstruction artefacts. The mid-position image has the same size and resolution than the breathing phases in the 4DCT. The midpCT and the deformation fields are saved in the DICOM format, and are given to the treatment planning system to consider the breathing motion in case of a mobile tumor.

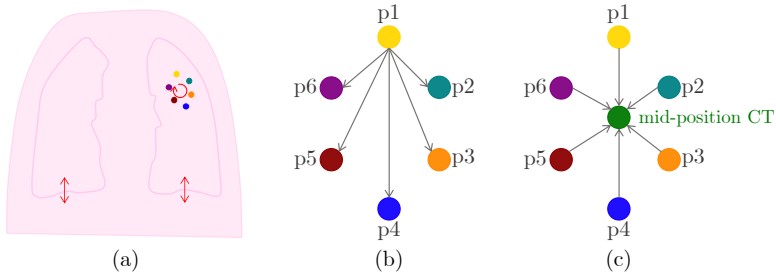


Fig. 3.5 (a) Schematic representation of the tumor motion represented by a hysteresis shape created by the positions of the tumor at the different breathing phases of the 4DCT. (b) Deformation fields (grey vectors) are computed using non-rigid registration between the first phase (p1) and all other phases. (c) All phases are deformed to the average motion position, called the mid-position (midp) 3DCT image (green circle).

ROI contours

The radiation therapy treatment workflow requires the target and organs at risk to be contoured on the planning 3DCT by the radiotherapist. The different contours defined by the physician are saved in the standard file format, DICOM RTstruct. This file contains, for each contour, the name of the organ and the list of the 3D points forming the mesh of the 3D contour in the coordinate system of the 3DCT image on which they are drawn.

In this thesis, only the contours of the GTV, heart, both lungs and body are considered. Figure 3.6 shows the contours of interest on different slices and views of a CT image for one patient studied in this thesis.

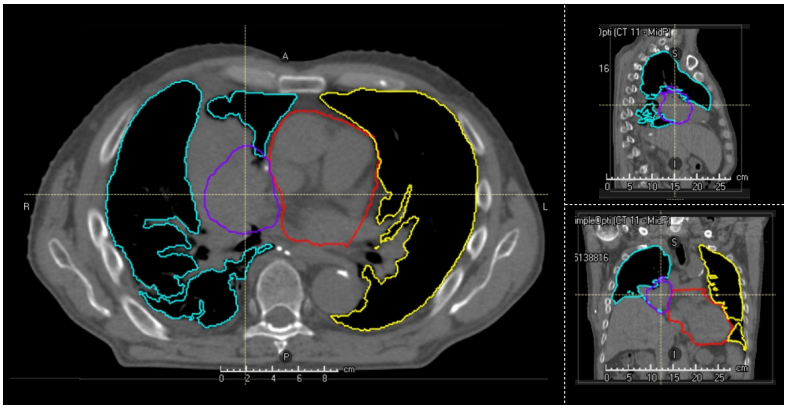


Fig. 3.6 Example of contours of different regions of interest: the right lung in blue, the left lung in yellow, the heart in red and the CTV in purple. These contours are displayed on transversal (left), sagittal (top right) and coronal (bottom right) CT slices.

3.2.3 Data augmentation tool

The ALARA principle has been defined to ensure that human exposure to radiation remains "as low as reasonably achievable" [HB05]. It is therefore not permitted to acquire a large number of 3DCT images of the same patient. This is a major problem for this thesis as all artificial intelligence algorithms require a large number of datasets to train and validate the neural network. One goal of this thesis was therefore to develop a data augmentation tool able to create a database of sufficient size for training and validating neural networks. This data augmentation tool requires the

acquisition of a 4DCT and generates new 3DCT images of a patient. This data augmentation tool is implemented in an open-source treatment planning system for research in proton therapy, OpenTPS[1].

OpenTPS is an open-source treatment planning system designed for research. The software is organised into two packages: the core package and the gui package. The core package is a library that defines data classes, data processing methods and in/out methods. The gui package offers a graphical user interface for viewing and interacting with the data. The core package of OpenTPS is the main software library and includes a range of features that are essential for proton therapy treatment planning. Some of the key features available in the core package of OpenTPS include: data management and processing, dose computation, treatment planning, treatment evaluation and data augmentation. In this section, the data augmentation tool developed and implemented in OpenTPS is described. For further information of the other features available in the software, the reader may refer to [WDSJ+23].

The data augmentation tool is based on the acquisition of a 4DCT. Deformations fields are computed from every breathing phase of the 4DCT to a reference phase using non-rigid registration in order to obtain the mid-position. Every breathing phase is then deformed to the mid-position, where the median is computed to create the midpCT image. The midpCT image coupled with the deformation fields from the mid-position to each breathing phase forms a motion model. Depending on the non-rigid registration algorithm used, the deformation fields are either displacement fields or velocity fields. The motion model is then given as input to the data augmentation algorithm. This algorithm aims to create new synthetic and realistic 3DCT images of the patient by deforming the midpCT image. The deformations are divided into two categories: inter-fraction deformations and intra-fraction deformations. Inter-fraction deformations represent setup errors and anatomical changes related to weight loss, while intra-fraction deformations represent internal organ motion.

Inter-fraction deformations

Figure 3.7 displays a schematic representation of the four inter-fraction deformations implemented in OpenTPS. The inter-fraction deformations are implemented as operations on the displacement fields. The displacement

[1] `http://www.opentps.org/`

is automatically computed from the velocity field when the deformation is applied using a field exponentiation operation. Figure 3.8 shows the results of applying the four types of inter-fraction deformation to a midpCT image, either independently or together. It is important to note that large deformation values have been expressly chosen here so that the deformations are perceptible in the figure. The deformation values chosen in the rest of this work are smaller and are inspired by the values actually observed in the treatment room. This will be explained in more detail in each contribution chapter.

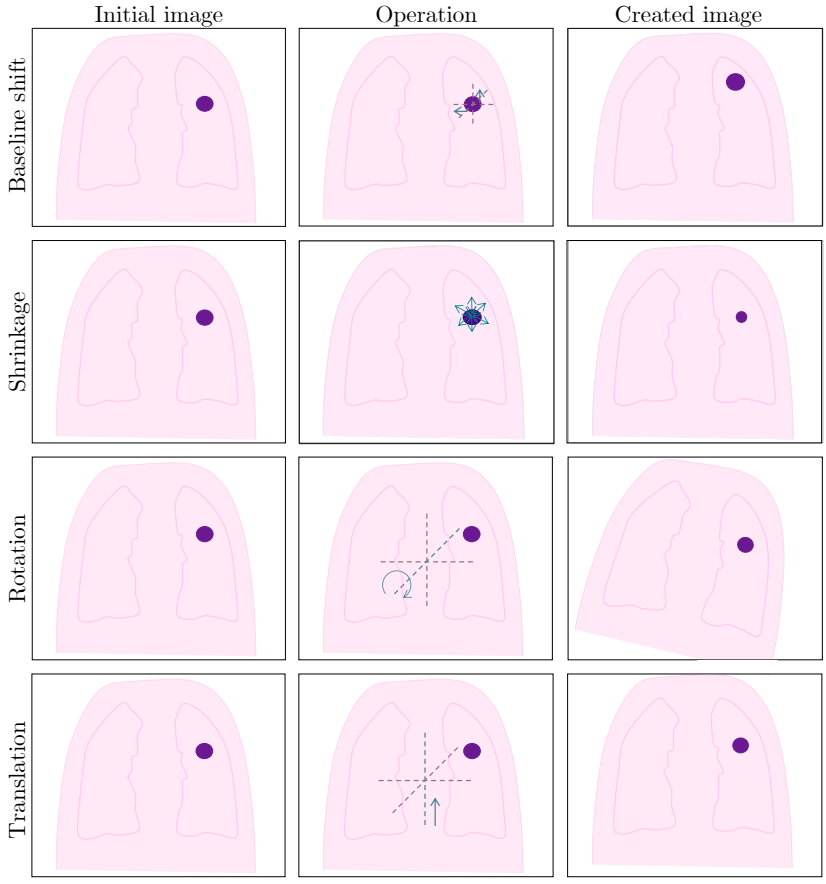


Fig. 3.7 Schematic representation of the four inter-fraction deformations implemented in OpenTPS: organ baseline shift, organ shrinkage, image rotation and image translation.

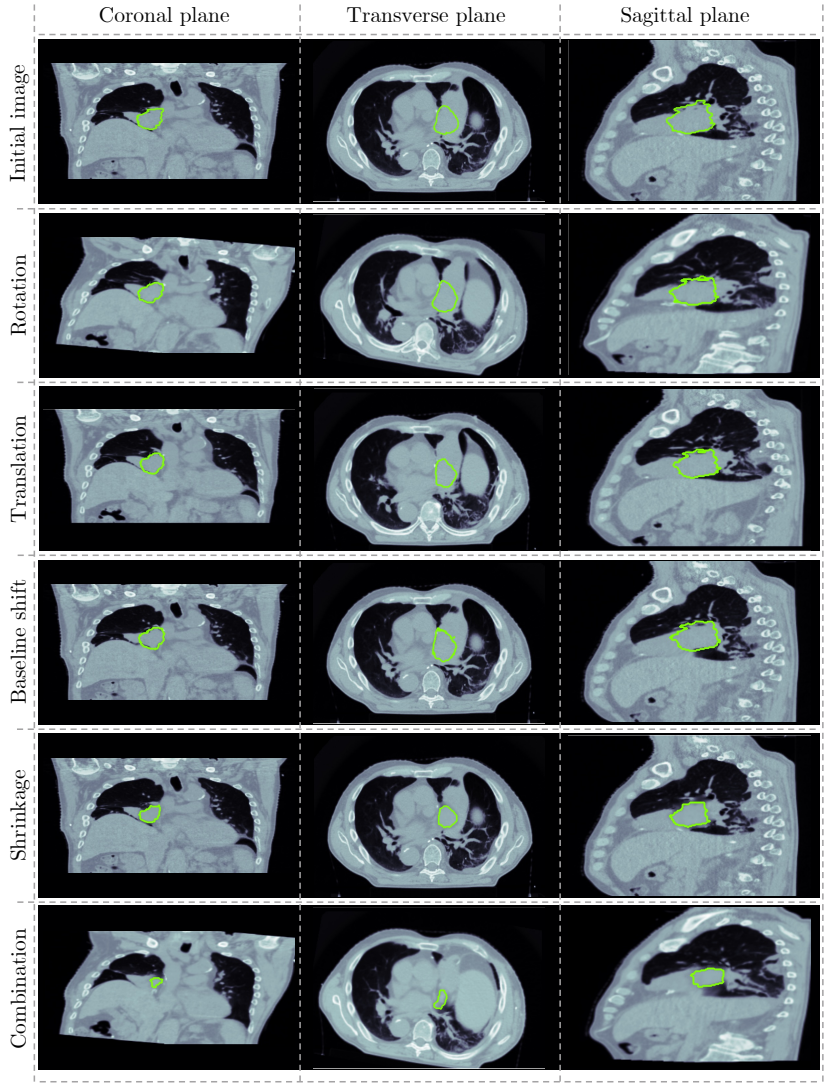


Fig. 3.8 Application of the four inter-fraction deformations available in the data augmentation tool on a midpCT image, either separately or together. For each plane, the slice shown is that of the centre of mass of the GTV in the initial image. In this example, the deformation values are 5 mm (left-right), 8 mm (cranio-caudal) and 10 mm (anterior-posterior) for the translation and the GTV baseline shift, 4°, 7° and 9° for the rotation and 2 mm, 1 mm and 3 mm for the GTV shrinkage.

Baseline shifts can be applied on a motion model using the ROI mask of

the target or of the organ to be translated. The baseline shift is constructed as a diffeomorphic displacement vector field modeling a local shift of the ROI mask while preserving the surrounding structures. The user selects the baseline shift value to be applied and the organ affected by this deformation. The value chosen is expressed in millimeters and it is possible not to give the same value in the three motion directions.

Shrinkages can be applied on a motion model using the ROI of the target or of the organ to be shrink. The shrinkage is constructed as an expansion of one voxel and an erosion of N voxels depending on the user input in millimeters. The dilated and eroded bands corresponding to the difference with the original mask are computed using these two masks. Every voxel of the eroded band is then replaced by a random sample picked from a normal distribution $\mathcal{N}(\mu, \sigma^2)$, where μ is the average value of the ten voxels from the dilated band closest to the voxel. The user selects the shrink value to be applied and the organ affected by this deformation. The value chosen is expressed in millimeters and it is possible not to give the same value in the three motion directions.

Rotations around the three main axes can be applied on a motion model. The rotation is applied around an axis at the center of the image and not around the image origin. In case of multiple rotations, the order is determinant as rotations are not commutative. The user chooses the three angles of rotation in each direction, which are expressed in degrees.

Translations in the three main axes can be applied on a motion model by translating the whole image by N voxels depending on the user input in millimeters. The user decides on the three translation values in each direction, which are expressed in mm.

Intra-fraction deformations

A breathing motion signal is generated to create a sequence of 3DCT images following this specific respiratory pattern. An ideal breathing signal is a sinusoidal wave oscillating at a specific frequency. However, over the image acquisition period, the patient does not always remain still, which may result in a shift of the sinusoidal wave. In addition, random events such as coughing, stress or apnea have a real impact on the simulation of breathing. Noise is also present in the data acquisition. Because of all these non-idealities, respiration must be modelled by a complex sine wave

whose amplitude and frequency vary with time. Therefore, a breathing signal is simulated using:

$$y(t) = A(t)\sin(2\pi f(t)) + s(t)$$

where $A(t)$ is the amplitude of the signal, $f(t)$ is the frequency of the signal and $s(t)$ is the shift. $A(t)$, $f(t)$ and $s(t)$ are random step functions where the intervals are random variables. A Gaussian noise is also added to each signal. The user can choose the breathing period and mean amplitude of the signal, as well as the standard deviation of the noise present in it. He can also choose to add or not irregularities. Figure 3.9 displays two synthetic breathing signals: a regular one in blue and an irregular signal in orange.

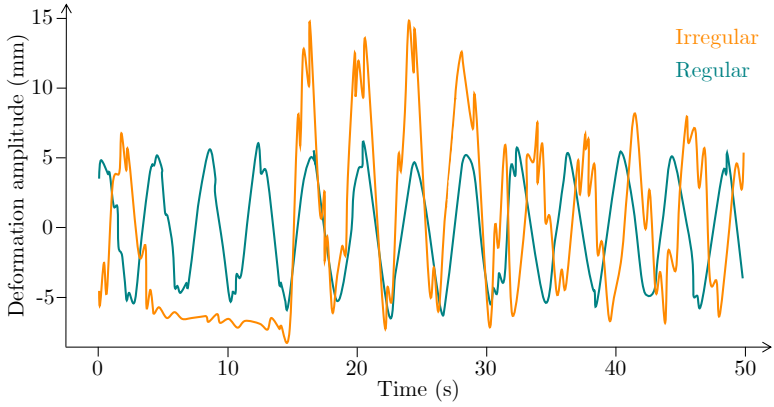


Fig. 3.9 Example of two synthetic breathing signals, a regular signal in blue and an irregular in orange.

The user can chose in this signal one or multiple points at which new images are created. Each point determines the phase and the amplitude of the new image. A polar coordinate system (r, n) related to the motion model is defined, where the origin is the midpCT image and n are the periodic breathing phases. One example is represented in figure 3.10. In this system, the deformation fields associated to the 10 breathing phases of the 4DCT are known and are $F(1, N)$, with $N \in 0, 0.1, ..., 0.9$. Then, to create the new image at breathing phase n and at a normalised distance r of the midpCT image, the deformation field $F(r, n)$ is computed using a linear interpolation between the two closest discrete breathing phases plus

a scaling :

$$F(r,n) = \left[F(1,N) + \left(F\left(1,N{+}0.1\right) - F\left(1,N\right)\right) \cdot 10 \cdot (n-N)\right] \cdot r \qquad (3.4)$$

where $N \leq n \leq N{+}0.1$. Using this method, it is possible to generate slightly different 3DCT images, spread around the ten original breathing phases of the 4DCT. Moreover, the user can chose to apply the breathing signal to one or multiple points of the image. If multiple points and signals are used, for example to partially decorrelate the tumor motion from the skin motion, they are combined using weight maps and linear interpolations.

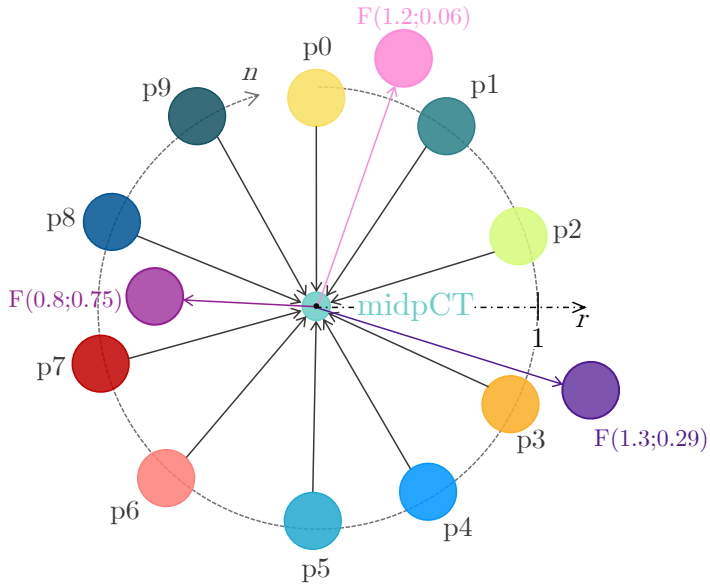


Fig. 3.10 Representation of the polar coordinate system used to reconstruct the new 3DCT images. The breathing phase and amplitude at which the new image is created are selected by the user on the synthetic breathing signal.

Figure 3.11 shows five synthetic 3DCT images created using random intra-fraction deformations implemented in the data augmentation tool.

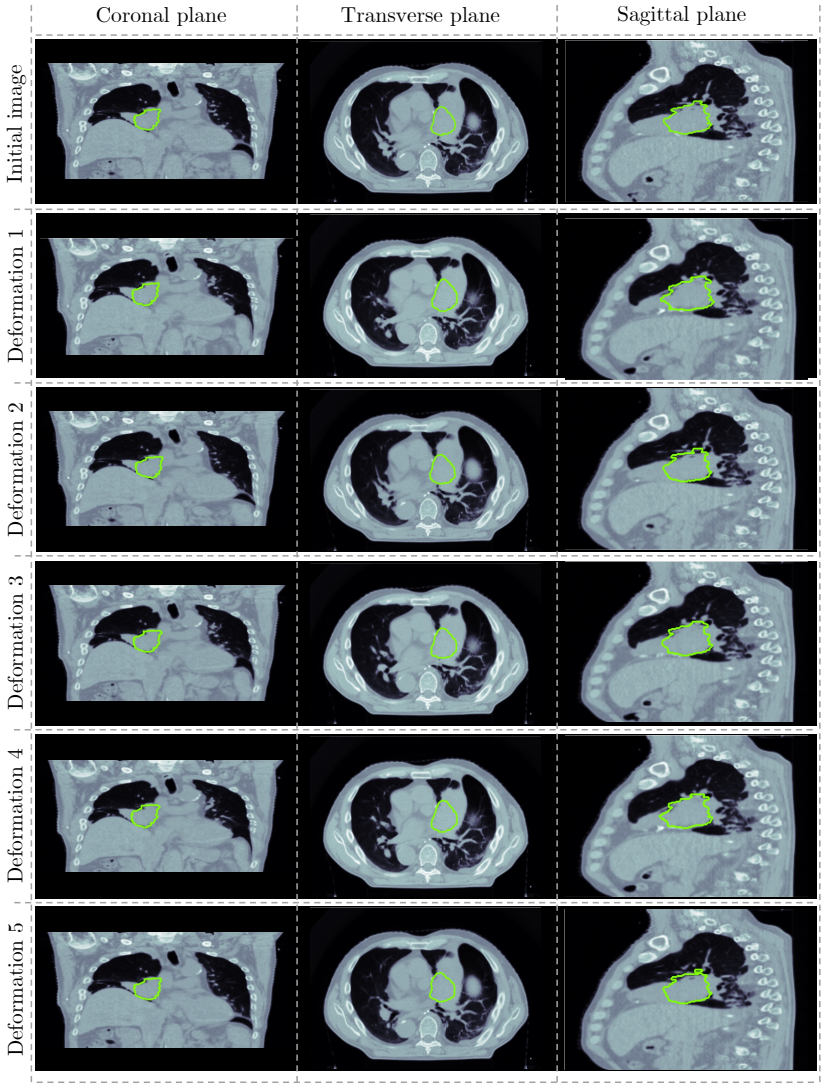


Fig. 3.11 Example of five 3DCT images created using random intra-fraction deformations of the midpCT image.

3.3 Neural network

The neural network used in this thesis was developed and published by Henzler et al. in [HRRR17]. This research group applied deep learning

in the form of a convolutional neural network to the challenge of inverting x-rays imaging. Although CNNs have had success in generating depth from opaque observations [EPF14] and inferring 3D volumes [WSK+15, REM+16], they were the first team to attempt to invert single x-rays images using this type of neural networks. The special feature of the neural network developed by Henzler et al. is that it takes as input a 2D image of size 256×256 and outputs a volume of size $128 \times 128 \times 128$, whereas typical neural networks generally have input and output of the same spatial resolution.

3.3.1 Architecture

The overall structure of the neural network is an encoder-decoder with skip connections and residual learning. An overview of the network is shown in figure 3.12. The rest of this section describes the architecture of the neural network.

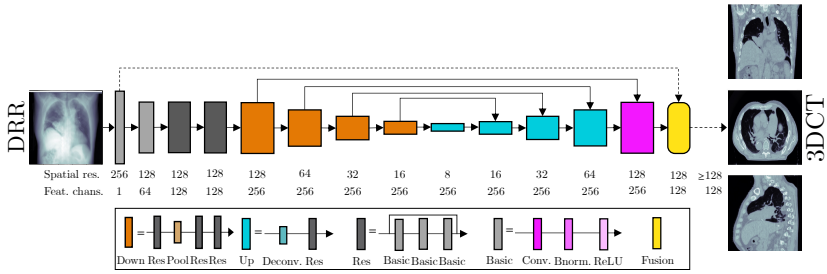


Fig. 3.12 Neural network used in this thesis to reconstruct a 3DCT image from a single x-rays projection. The network takes as input a digitally reconstructed radiograph of size 256×256 and outputs the corresponding 3DCT scan, with a size of $128 \times 128 \times 128$. This figure is adapted from [HRRR17].

Two basic blocks compose the first layers of the network. Each one consists of a convolution operation accompanied by a batch normalisation and a ReLU activation function. The convolution of the first basic block is defined by 64 kernels of size 7 and a stride of 2, allowing to pass to a tensor of size $128 \times 128 \times 64$. The convolution of the second basic block uses 128 kernels of size 1 and a stride of 1, giving a tensor of size $128 \times 128 \times 128$.

Two residual blocks are used after the basic blocks to increase the learnability. Instead of learning the convolutions directly, the network learns the

additive residual. This does not change the network expressiveness, but it significantly helps the training, making it easier, and improves generalisation. The residual blocks are made up of three successive basic blocks, where the sizes of the three convolution kernels are 1, 3 and 1 with unit strides. The number of filters are 64, 64 and 128 for the first block, giving a tensor of size $128 \times 128 \times 128$. For the second block, the number of filters are 128, 128 and 256, creating a tensor of size $128 \times 128 \times 256$.

Encoder starts at this point. The aim of an encoder is to convert an image into an internal representation that represents the information contained in the training data. To this end, the encoder reduces the spatial resolution of its input. In this network, a down block is composed of three residual blocks and a pooling operation. The pooling operation takes place after the first residual block with a filter size of 2 and a stride of 2, reducing spatial resolution by half. The filters of the three convolution operations are identical in the three residual blocks, with sizes of 1, 3 and 1 and a number of filters of 128, 128 and 256.

Decoder starts when the spatial resolution is reduced to 8. The aim of a decoder is to apply the internal representation of the encoder to a specific instance. The decoder increases the spatial resolution again, without reducing the number of feature channels. This is performed using an up block composed of a deconvolution operation and a residual block. The transposed convolution is composed of 256 kernels of size 4, using a stride of 2 and a zero-padding, meaning that the output spatial resolution is the input resolution multiplied by 2. The residual block has the same parameters than those used in the encoder step.

Skip connections are used to share the spatial details of a certain resolution at some level on the encoder part with the same resolution on the decoder part. These convert fine details of the input 2D image into details of the output 3D volume. Skip connections allow the high-resolution spatial layout to be used to locate features, for example on edges.

A last convolution with 128 kernels is used to reduce the number of feature channels to 128. The design of the network increases the number of feature channels from 1 to 256 using the first basic and residual blocks. The network retains this number of feature channels until this last step where it is reduced to the output resolution of 128. This ensures that the neural

network does not produce the same result for every slice in that direction.

Fusion is used to increase the resolution of the output to a desired spatial resolution, while maintaining the overall transparency (α) of the original 2D radiograph. This step distributes the density error $\mu_i - \bar{\mu}_i$ of a slice i to obtain a new density value $(\hat{\mu}_i)$ for each voxel, so that the voxels still form the correct value of the transparency, defined as the fraction of x-rays arriving at the sensor after traveling a volume with extinction coefficient. The extinction coefficient is defined as the sum of absorption and scattering. During the training step, the loss function encourages the inferred densities $(\bar{\mu}_i)$ to be as close as possible to the true densities (μ_i), but it has no influence in ensuring that the fraction of x-rays arriving at the detector is correct. This is achieved, as a post-processing step, by setting:

$$\hat{\mu}_i = \bar{\mu}_i - \Delta \frac{\bar{\mu}_i^2}{\sum_1^n \bar{\mu}_i^2} \tag{3.5}$$

where $\Delta = \log(1 - (\bar{\alpha} - \alpha))$ and $\bar{\alpha} - \alpha$ is the transparency error.

3.3.2 Learning

In this work, the mean squared error is the loss function used to train the network in a supervised way. The mean squared error computes the squared difference between actual and predicted values. It is defined by:

$$\text{MSE} = \frac{1}{N} \sum (y - \hat{y})^2 \tag{3.6}$$

This loss function is generally used for a regression task as is the case in this problem. Stochastic gradient descent is employed to reduce the cost of computing the gradient, as well as an adaptive learning rate algorithm to dynamically adjust the learning rate during the training process. This is performed using the common extension Adam which involves combining gradient estimation with lower moments of the gradient. The values of the other hyper-parameters of the neural network used in the rest of this work are task-specific. This will be explained in more detail in each contribution chapter. For each contribution, the model was trained on NVIDIA RTX 600.

It is essential to note that, for each contribution, the training of the neural network is patient-specific. This means that the neural network is

trained independently for each patient. The training and test sets only contain images of the same patient. However, it is also important to point out that the hyper-parameters are optimised for a task and not for a patient. This implies that the training performed for the different contributions do not have the same hyper-parameter values (for example, the number of images and the number of epochs required for the training are not the same), but within a same contribution, all patient-specific networks are trained with the same values of hyper-parameters.

3.4 Treatment plan optimisation

RayStation v.12B is used to create and optimise all treatment plans required in this thesis. RayStation is a treatment planning system (TPS) for external beam radiation therapy developed, sold and maintained by the Swedish software company RaySearch Laboratories A.B. [RSL]. This software plays a very important role in clinical routine, but the version used is dedicated solely to research. This software has several tabs. In this work, five of them are used.

Patient data management tab is used firstly to import the data of the patient to be planned. This tab enables to import in the software all images and contours of the patient. This tab is also important because it enables the calibration of the CT scanner to ensure that the Hounsfield Units and the tissue densities match.

Patient modeling tab is used to visualise the different cross-sections of the patient and display the contours of the target and organs at risk. This tab also includes various tools for creating new volumes or modifying and combining existing volumes. These tools are not needed for this thesis as the contours defined by the physicians are used.

Plan design tab is used to create the treatment plan. This step is important because it allows to define a certain number of parameters characterising the plan. These are listed in table 3.1 and differ according to the type of cancer of the patient, either lung cancer or liver cancer.

This tab also enables to add beams to the treatment plan. In this work, three beams are always defined. The parameters used to characterise the treatment beams are given in table 3.2. These are identical for lung and

liver cancers.

Table 3.1 Parameters set in RayStation when creating a treatment plan.

Treatment plan parameter	Lung cancer	Liver cancer
Planning image set	planning midpCT	planning midpCT
Patient treatment position	Head First Supine	Head First Supine
Modality	Protons	Protons
Treatment technique	Pencil Beam Scanning	Pencil Beam Scanning
Treatment machine	IBA ProteusONE	IBA ProteusONE
Treatment site	GTV	GTV
Prescription type	Median dose ($D_{50\%}$)	Median dose ($D_{50\%}$)
Prescribed dose	60 Gy	52 Gy
Number of fractions	30	6

Table 3.2 Parameters set in RayStation when adding a beam to the plan.

Beam parameter	Value
Isocenter	Center of the GTV
Gantry angle	see table 3.3
Couch angle	0°
Dose specification point	Isocenter
Spot tune ID	3.0
Snout	None
Range shifter	None

The gantry angle of each beam is specifically chosen according to the position of the target within the body. The gantry angle values used for the different patients studied in this thesis are given in table 3.3.

Table 3.3 Gantry angles used for the different patients studied in this thesis.

Patient ID	Cancer site	Beam 1	Beam 2	Beam 3
Patient 0	liver	0°	180°	270°
Patient 1	lung	10°	250°	310°
Patient 2	lung	30°	100°	320°
Patient 3	lung	210°	260°	320°
Patient 4	lung	0°	270°	315°
Patient 5	lung	20°	200°	290°
Patient 6	lung	30°	75°	120°
Patient 7	lung	27°	200°	290°
Patient 8	lung	25°	205°	295°
Patient 9	lung	15°	105°	195°
Patient 10	lung	15°	105°	195°
Patient 11	lung	190°	255°	320°
Patient 12	lung	215°	270°	325°
Patient 16	liver	0°	180°	270°
Patient 19	liver	0°	180°	270°

Plan optimisation tab is used to define all parameters required for the treatment plan optimisation, as well as to set clinical objectives and optimisation constraints.

The dose-based optimisation of a treatment plan, also known as inverse planning, is a technique that achieves the clinical objectives of a plan through iterative adjustments of the plan parameters using a numerical optimisation method. The optimisation process continues until one of the following three criteria is met: the variation in the value of the optimisation objective function is less than the optimisation tolerance (set to 10^{-6}) for 3 consecutive iterations, the maximum number of iterations has been reach (set to 150), or no direction of improvement can be found.

In proton therapy, the Monte Carlo algorithm is regularly used to compute the dose. This algorithm is recognised for its high accuracy and provides a good estimate of the dose distribution inside the patient. This algorithm models and simulates the particles electromagnetic and nuclear interactions in matter to predict their range and local dose deposition. Two parameters are used in the Monte Carlo algorithm: the number of ions per spot and the dose uncertainty. These are set to 5000 and 1%, respectively.

The optimisation constraints and clinical objectives depend on the type of cancer of the patient. The optimisation constraints and clinical objectives used for lung cancer patients are presented in table 3.4, while those used for liver cancer patients are presented in table 3.5.

Table 3.4 Optimisation constraints and clinical objectives used for the optimisation of proton therapy treatment planning in case of lung cancer patients.

ROI	Optimisation constraints	Clinical objectives
GTV	Min dose 58 Gy Max dose 63 Gy	At least 98% volume at 57 Gy
Heart	Max EUD 20 Gy Max dose 22.5 Gy	At most 20 Gy average dose At most 63 Gy dose at 0.04 cm^3 volume
Lungs−GTV	Max dose 63 Gy Max EUD 20 Gy	At most 20 % volume at 30 Gy At most 20 Gy average dose

Plan evaluation tab is used to check that the clinical objectives set by the physician are met with the treatment plan proposed using the optimisation algorithm. This tab includes two tools for analysing the resulting plan. The first one displays the DVH, which indicates the proportion of each organ volume receiving a specific dose. The second one displays isodose

curves, which indicate the areas of over- and under-dosage. If, after this evaluation, the treatment plan is not validated, the optimisation must be repeated with relaxed optimisation constraints. On the other hand, if the treatment plan is validated, it is saved in the appropriate DICOM format.

Table 3.5 Optimisation constraints and clinical objectives used for the optimisation of proton therapy treatment planning in case of liver cancer patients.

ROI	Optimisation constraints	Clinical objectives
GTV	Min dose 52 Gy Max dose 52 Gy	At least 99% volume at 52 Gy
Heart	Max EUD 25 Gy	At most 40 Gy dose at 0.01 cm^3 volume At most 1 cm^3 volume at 30 Gy
Liver	Max EUD 25 Gy	At most 18 Gy average dose At most 700 cm^3 volume at 21 Gy

3.5 Summary

In this chapter, we first set out the general context of this thesis. In the second section, we discussed data management. We presented the two medical imaging modalities used in this work and we also explained the data augmentation tool developed to deal with the problem of lack of data. We then looked at the neural network used in this work, and we described its architecture and the learning strategy pursued. In the last section of this chapter, we presented the treatment plan optimisation strategy and we also discussed the different parameters set in RayStation.

4

Patient-specific 3DCT reconstruction from a single x-rays projection using a CNN for on-line radiotherapy applications

This chapter covers the main method used to reconstruct a 3DCT image from a single x-rays projection using a convolutional neural network. The beginnings of this work were presented orally at the 4D Treatment Planning Workshop for Particle Therapy in Delft in 2021. A more detailed version was then presented orally at the European Society for Radiation and Oncology in Copenhagen in 2022 [LDSM22b]. Thanks to this oral presentation, the complete method was published in the special edition "Physics highlights from ESTRO 2022" of the journal Physics and Imaging in Radiation Oncology [LDSM23]. This chapter is based on this journal paper, and also provides additional results.

4.1 Context

Radiotherapy is one of the most widely used treatments in oncology and is prescribed for more than half of all cancer patients, either alone or in combination with surgery and chemotherapy [BLYY12]. In radiotherapy, ionising radiation is used to kill cancer cells. A trade-off must be made between delivering the prescribed dose to the target and not delivering large doses to healthy tissues, which could lead to undesirable effects and induce secondary cancer [WSS+04]. Applying radiotherapy to lung and liver cancers is even more challenging as the treatment must consider the respiratory motion. This requires specific strategies in the radiotherapy workflow to ensure adequate target coverage through successive treatment fractions. These strategies are generally classified in two categories.

The first category consists in acquiring a four-dimensional computed tomography (4DCT) scan prior to treatment and defining security margins. Safety margins ensure target coverage regardless of the breathing phase, but this method irradiates more the surrounding healthy organs [RB10]. The breathing motion in the treatment room may also differ significantly from the motion captured in the 4DCT from time to time [DVB+18].

The second category includes breathing-synchronised methods that aim to minimise the contribution of the motion of the tumor in the computation of the safety margins by monitoring position of the tumor or reducing/regularising its motion amplitude during breathing. These methods gather mechanically assisted ventilation [VODSLG19], audio coaching [NNM+09], abdominal compression [PHNACG18] and respiratory gating [MFD+10]. In these techniques, tumor monitoring is based on external surrogates for internal motion to avoid the use of invasive procedures, given that the placement of markers involves surgery before the treatment, but pinpoints the tumor position with greater accuracy [HWST19]. This approach requires a stable correlation between the internal tumor motion and its external surrogate, which is usually not the case when changes occur in the breathing movement of the patient.

Image-guided radiation therapy (IGRT) employs imaging techniques during each treatment session. By adding detailed images, it ensures that the radiation is narrowly focused on the target. A broad range of IGRT is now available [RLLL19a]. X-rays projections are commonly acquired to

estimate the position of the tumor, but their use often requires implanted markers to identify the tumor volume correctly and make it visible on the x-rays projection [SVM+02]. Another disadvantage of this method is that it does not provide 3D information.

All these methods result in a small reduction in the safety margins, while adapting the treatment in 3D and in real-time leads to a big reduction in the motion margins thanks to precise tracking of the 3D anatomical structures. To achieve this, the real-time positions of the target and surrounding organs must be known throughout treatment delivery. Most of the radiotherapy treatment rooms are equipped with 2D fluoroscopy to validate the position of the patient before the treatment. This work proposes to rely on this equipment to estimate the related 3D information.

Many studies that reconstruct a 3D volume from a 2D x-rays projection have already been performed. Different fields of application in the biomedical sector have been explored: Henzler et al. investigated how to reconstruct 3D volumes from 2D cranial x-rays by applying deep learning [HRRR17], while Liang et al. developed a new model architecture to reconstruct a tooth in 3D from a single panoramic radiograph [LSY+21]. Montaya et al. in [MZL+21], as well as Ying et al. in [YGM+19], demonstrated that it was possible to reconstruct a 3DCT image from biplanar x-rays projections using a neural network, and Shen et al. used a neural network to reconstruct a 3D image from a single projection view [SZX19].

In this context, the aim of the work described in this chapter was to use the 2D information available in the treatment rooms to obtain 3D information. To that end, this work is based on a convolutional neural network that reconstructs a high-quality 3DCT image based on a single x-rays projection. This image, predicted in real-time, can then be used by a real-time segmentation method [ZXBB22] in order to know the positions of the tumor and surrounding organs at the moment of acquisition. This process would make it possible to locate the tumor and neighboring structures accurately in 3D during the treatment without requiring implanted markers.

4.2 Methodology

Figure 4.1 summarises the workflow of the proposed method. The different steps of the process are detailed in the following sub-sections.

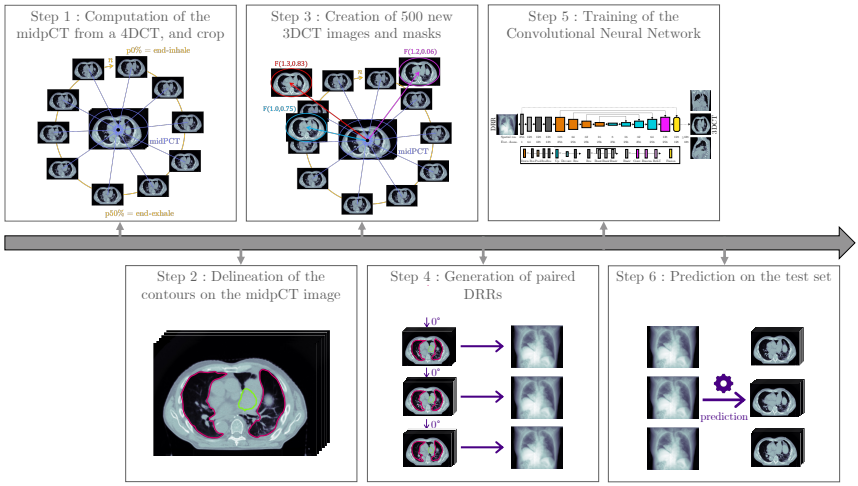


Fig. 4.1 Overview of the workflow of the proposed method.

4.2.1 Dataset generation

The data used in this work come from nine patients who were treated for lung or liver cancer at Cliniques universitaires Saint-Luc in Brussels between 2010 and 2015. This retrospective study was approved by the Hospital Research Ethics Committee (B403201628906). Table 4.1 shows patients information (tumor size and location, and its motion in the different sets). A planning 4DCT composed of 10 breathing phases evenly spread over the respiratory cycle was acquired for each patient prior to treatment delivery. The dimensions of each 3DCT image were $512 \times 512 \times 173$, and the voxel size was 1 mm^2 in plane with a slice thickness of 2 mm. The mid-position (midp) CT image, defined as the local mean position in the respiratory cycle, was computed using the average of all velocity fields obtained by non-rigid registration between the 4DCT phases [WSvHD08]. On the midpCT image, the gross tumor volume (GTV) and surrounding organs at risk were delineated manually by an experienced radiation oncologist.

As training a neural network requires a lot of data, it was necessary to generate new 3DCT images. To do so, a polar coordinate system (r, n) related to a breathing cycle is considered. Its origin is the midpCT image and n are the periodic phases. In this system, the deformation fields associated to the 10 breathing phases of the 4DCT are known and are $F(1, N)$,

Table 4.1 Patient characteristics. MR_{4DCT}, $MR_{TrainSet}$ and $MR_{TestSet}$ stand for the motion range in 3D of the centroid of the GTV in the 4DCT, training set and test set, respectively. The motion range is defined as the Euclidean distance between the two most distant positions.

Patient ID	Tumor location	GTV size [cm^3]	MR_{4DCT} [mm]	$MR_{TrainSet}$ [mm]	$MR_{TestSet}$ [mm]
Patient 0	Right lobe of liver	28.62	15.1	18.7	26.4
Patient 1	Right upper lobe of lung	137.1	11.1	17.2	17.9
Patient 2	Left upper lobe of lung	17.24	9.92	9.69	12.6
Patient 3	Right middle lobe of lung	153.7	24.4	32.4	34.7
Patient 4	Right upper lobe of lung	13.78	14.5	15.2	18.5
Patient 5	Left upper lobe of lung	315.1	9.68	10.1	11.4
Patient 6	Left upper lobe of lung	67.21	11.6	15.2	16.1
Patient 16	Right lobe of liver	80.36	27.1	29.9	30.8
Patient 19	Left lobe of liver	22.53	24.1	32.3	34.8

with $N \in \{0, 0.1 \ldots, 0.9\}$. Then, to generate the breathing phase n at a normalised distance r of the midpCT, the deformation field $F(r, n)$ is computed using a linear interpolation between the two closest discrete breathing phases plus a scaling:

$$F(r,n) = \big[F(1,N) + \big(F\,(1, N{+}0.1) - F\,(1,N)\big) \cdot 10 \cdot (n - N)\big] \cdot r \qquad (4.1)$$

where $N \leq n \leq N{+}0.1$. Using this method, based on a previous work of our team [DSASM22] and developed in [WDSJ+23], slightly different 3DCT images, spread around the ten original phases of the 4DCT, can be generated for every patient. The training set was composed of 500 images where n was a uniform random draw between 0 and 1, and r a random sample from a normal distribution $\mathcal{N}(1, 0.25)$ truncated between 0.4 and 1.1. A digitally reconstructed radiograph was generated from each of these images using the Beer-Lambert absorption-only model (implemented in the TomoPy Python library [GDCXJ14]) and a projection angle of 0° along the anterior-posterior axis. The projection geometry was a 1440×1440 image with a pixel size of 0.296×0.296 mm^2. The source-to-origin and source-to-detector distances were 1000 mm and 1550 mm. The training dataset of each patient was made up of 500 pairs containing the created 3DCT image and the associated DRR. An independent test set composed of 100 3DCT/DRR pairs was also created for each patient. For each image of the test set, the masks of the GTV, lungs and heart were also generated by deforming the 3D binary masks of the midpCT image.

The difference between the test and training sets comes from the normalised distance r used to generate the 3DCT image. In the case of the training set, r was a random sample from a normal distribution $\mathcal{N}(1, 0.25)$ truncated between 0.4 and 1.1, while r was a random sample from a normal distribution $\mathcal{N}(1, 0.5)$ truncated between 0.8 and 1.5 for the test set. This means that deeper breathing situations were present in the test set than in the training set, as observed in table 4.1. All breathing phases were used in both cases.

4.2.2 Network

The network used for the 3DCT reconstruction process is a convolutional neural network (CNN) that learns the mapping between a 2D image and a 3D volume. This network was proposed by Henzler et al. in [HRRR17] and the different hyper-parameters was tuned for this challenge. The overall structure of this network is an encoder-decoder with skip connections. The goal of the encoder is to condense the information contained in the training data into a low-dimensional representation, which the decoder then takes as input to predict the output [MBP+22]. The input of the network is a DRR of size 256×256, while the output consists of a $128 \times 128 \times 128$ 3DCT image. The details of the training dataset, namely 3DCT/DRR pairs, are explained in section 4.2.1.

4.2.3 Training specifications

The neural network training was patient-specific, meaning that a new network was trained independently for each patient. The same training strategy and hyper-parameters were used for all patients. The Adam optimiser was used to train the network with an initial learning rate of 10^{-3} and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The model was trained for a total of 300 epochs using a mini-batch size of 16 on a NVIDIA RTX 6000, which brought the training time down to roughly 8 hours. Then, it took around 50 μs to predict the output from a new input. This way of training the model is called *basic-training*.

4.2.4 Performance evaluation

In order to evaluate the performance of the proposed method, 100 3DCT images independent of the training set were created for each patient. These 3DCT images are called the ground-truth (GT) 3DCT images. 100 DRRs

were generated from these images to form the test set. The trained neural network was used on these radiographs to predict the corresponding 3DCT images, called the predicted (P) 3DCT images. The predicted 3DCT images were compared with the ground-truth 3DCT images to evaluate the performance of the model using several metrics.

Visual analysis was carried out to qualitatively assess the performance of both training methods. To that end, the difference was computed between a ground-truth 3DCT image and the corresponding predicted 3DCT. The image used in this analysis was the first image of the test set of Patient 19.

Difference was computed between all ground-truth 3DCT images and the corresponding predicted 3DCT images of the test set. The mean and the median of the difference were studied, as well as the percentage of the 3DCT volume having an absolute value of the difference below a certain threshold in order to quantify the proportion of the image that was correctly reconstructed.

Dice similarity coefficient (DSC) is a common overlap-based metric used to measure the performance of a segmentation algorithm, and is defined by :

$$\mathrm{DSC} = \frac{2|A \cap B|}{|A| + |B|} \cdot 100 \quad [\%] \tag{4.2}$$

where A and B are the sets containing the matrix indices of both binary masks A and B. In this work, the DSC was computed between a 3D binary mask in the ground-truth 3DCT image and the corresponding mask in the predicted 3DCT image to evaluate the quality of the predicted 3DCT image in terms of anatomical structure positions. The 3D binary masks of a predicted 3DCT image were obtained by computing the Morphons non-rigid registration [JJOdX+11], then applying the resulting deformation fields to deform the masks on the predicted image. This was done between this predicted image and either the ground-truth 3DCT image (*GT-based*), or the midpCT image (*midp-based*). Using the ground-truth 3DCT image serves as a post-training quality evaluation, to evaluate if a state-of-the-art registration algorithm sees a difference between the ground-truth and the prediction. Using the midpCT image simulates how the quality of the predicted images could be evaluated after each treatment fraction as the ground-truth 3DCT images are not available during the treatment. For

both versions, the DSC was computed for the same 50 images of the 100 items constituting the test set, for each organ and each patient. In either case, this metric was an evaluation tool and not part of the real-time process as the computation time of the Morphons is about 150 s.

Euclidean distance (ED) was computed between the centroids of two 3D binary masks. This metric was used to complement the DSC analysis by computing the Euclidean distance between the centroids of the masks in the ground-truth 3DCT image and in the predicted 3DCT image for the four organs. As for the DSC analysis, the masks on the predicted 3DCT are obtained using either the masks from the GT image (*GT-based*), or the masks from the midpCT image (*midp-based*).

Normalised root mean squared error (NRMSE) was computed between two images A and B, and is defined by :

$$\text{NRMSE} = \frac{\sqrt{\frac{\sum_{i=1}^{n}(A_i - B_i)^2}{n}}}{A_{max} - A_{min}} \tag{4.3}$$

where X_i is the voxel i in the image X. A_{max} and A_{min} stand for the maximum and minimum in image A, the ground-truth 3DCT image. Two analyses were carried out with this metric. The first analysis consisted of computing the NRMSE between the ground-truth 3DCT image and the corresponding predicted 3DCT image. The second analysis compared the NRMSE obtained using our method with a baseline. The baseline was computed between the ground-truth 3DCT image and the midpCT image. This was repeated for all images in the test set.

Three quality metrics, commonly used in the literature, were calculated to assess the quality of the reconstructed images. Two analyses were carried out with these three metrics. The first analysis consisted in computing the results of each metric between all ground-truth 3DCT images and the corresponding predicted 3DCT images in order to observe the distribution of these metrics in relation to the entire test set. The second analysis focused on the computation of these metrics on organs of interest. The analysis was based on the masks in the ground-truth 3DCT, and the comparison was voxel-wise. These three metrics are explained below.

1. **Mean absolute error** (MAE) was computed between two images A

and B, and is defined by:

$$\mathrm{MAE} = \frac{\sum_{i=1}^{n} |A_i - B_i|}{n} \tag{4.4}$$

2. **Peak signal-to-noise ratio** (PSNR) was computed between two images A and B, and is defined by:

$$\mathrm{PSNR} = 10 \cdot \log_{10} \left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}} \right) = 10 \cdot \log_{10} \left(\frac{\mathrm{MAX}_I^2}{\frac{\sum_{\mathrm{i}=1}^{\mathrm{n}} (\mathrm{A_i} - \mathrm{B_i})^2}{\mathrm{n}}} \right) \tag{4.5}$$

where MAX_I is the largest possible pixel value in the image. If the two images are significantly different, the PSNR is generally less than 15 dB, whereas it is greater than 30 dB when the images are almost identical.

3. **Structural similarity index measurement** (SSIM) was computed between two images A and B, and is defined by:

$$\mathrm{SSIM} = \frac{(2\mu_A\mu_B + c_1)(2\sigma_{AB} + c_2)}{(\mu_A^2 + \mu_B^2 + c_1)(\sigma_A^2 + \sigma_B^2 + c_2)} \tag{4.6}$$

where μ_X and σ_X are respectively the pixel mean and variance of image X, σ_{AB} is the covariance of A and B, and c_1 and c_2 are two variables to stabilise the division with weak denominators. A SSIM of 1 corresponds to an ideal match between both images, while a SSIM of less than 1 indicates a mismatch between both images.

Reverse process of the proposed method consists in obtaining a DRR image from one 3DCT image. The aim of this analysis was to evaluate the quality of the digitally reconstructed radiograph obtained from the 3DCT image predicted by the neural network in order to quantify the accuracy of the predicted 3DCT image. This was done by computing the mean and the standard deviation of the three image quality metrics (MAE, PSNR and SSIM) between the ground-truth DRR and the one obtained from the corresponding predicted image, for all images of the test set. Besides, for Patient 3, a visual analysis of the difference between the DRR generated using one ground-truth 3DCT of the test set and the DRR obtained using the predicted 3DCT was performed.

4.3 Results

4.3.1 Visual analysis

A representative example of Patient 19 (whose results are: DSC_{GT} (GTV) = 98.5%, DSC_{midp} (GTV) = 88.6%, NRMSE = 0.053, mean of the difference = -1.73 HU and $V_{<25\,HU}$ = 80.3%) obtained using the proposed method can be seen in figure 4.2. For a human eye, the predicted 3DCT image looks pretty close in terms of anatomical structures. The zoom shows that a red pixel (difference $\approx$ 200 HU) is commonly adjacent to a blue pixel (difference ≈ -200 HU) or surrounded by two turquoise pixels (difference ≈ -100 HU). This phenomenon is usually observed at tissue borders. Looking at the histogram, one sees that there are only few voxels with a significant difference and over 30% of the voxels have a difference between -5 HU and 5 HU.

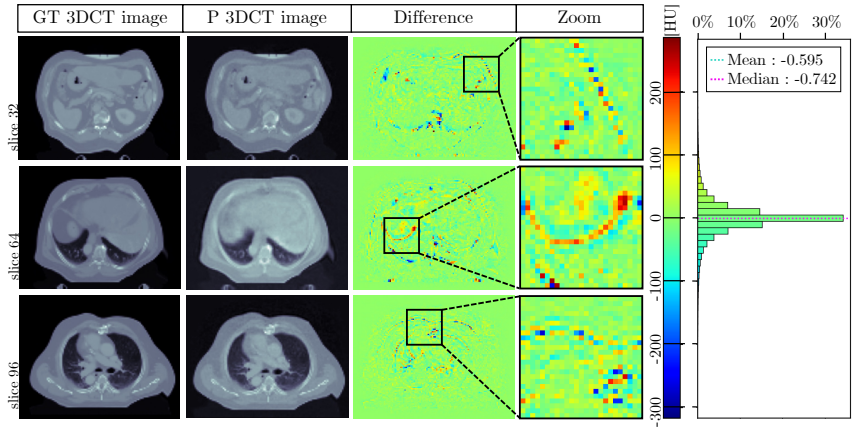


Fig. 4.2 Visualisation of three slices of the ground-truth 3DCT image of Patient 19 compared with the corresponding slices of the predicted 3DCT image, as well as the results of the difference analysis and a zoom of the boxed area. On the right of the color bar is the histogram of the difference concatenated for all patients and the 100 images of the nine test sets.

4.3.2 Difference

The results of the difference analysis are summarised in figure 4.3. The mean of the difference (orange box plots) is around 0 HU for most patients, while the median (red box plots) of this metric is rather around 0.5

HU. Patient 16 has a difference of approximately −1 HU, which means that the network tends to overestimate the value of the voxels intensity, whereas for Patient 19, the difference is approximately 2 HU, which shows an underestimation of the real values of the voxels. Besides, this figure highlights that 25.1% to 39.8% of the image volume has an absolute value of the difference lower than 5 HU (turquoise box plots), 69.9% to 81.9% below 25 HU (purple box plots), and 88.6% to 94.6% less than 50 HU (blue box plots). In summary, the difference between the ground-truth image and the predicted image is very small, with about 91% of the image volume having an absolute value of the difference smaller than 50 HU, which represents 1.25% of the range of possible values, since the scale of a 3DCT image typically runs from −1000 HU for air to 3000 HU for dense bone [BEP15].

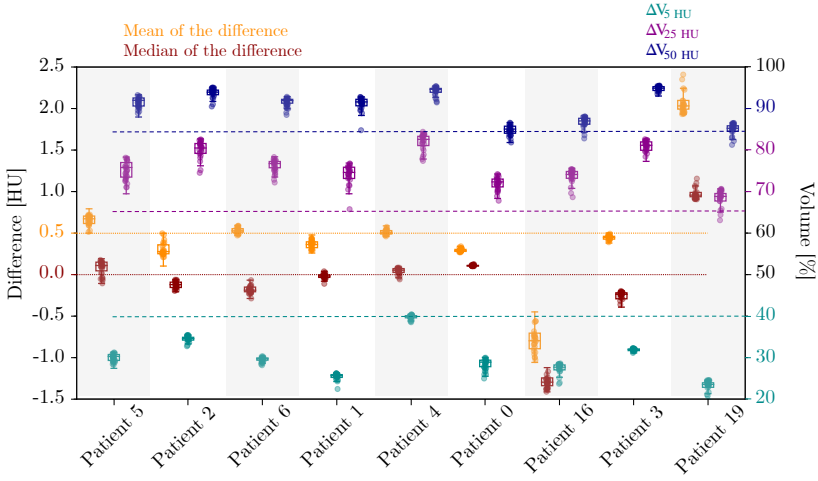


Fig. 4.3 Results of the difference analysis. The left axis is expressed in HU for the mean and median difference analyses, while the right axis is expressed in % for the analysis of the volume percentage with a difference below three thresholds: 5 HU, 25 HU and 50 HU. Patients are sorted by increasing motion range in the test set.

4.3.3 Dice similarity coefficient

The results of the DSC analysis for both GT-based and midp-based versions are summarised in figure 4.4. This figure shows that all DSC values

are above 75%, demonstrating the high quality of the reconstructed images. A more detailed analysis of this figure reveals that the two least well reconstructed organs are the GTV and the heart. The results for these two organs are generally between 75% and 90%, while the DSC values calculated for both lungs are mostly above 90%. For the GT-based version, the mean of the DSC varies respectively from 93.2% to 99.8% for the GTV; from 96.3% to 99.8% for both lungs; from 93.5% to 99.8% for the heart. While, for the midp-based version, the mean of the DSC varies from 76.7% to 90.6% for the GTV; from 90.9% to 97.3% for both lungs; from 78.1% to 90.1% for the heart.

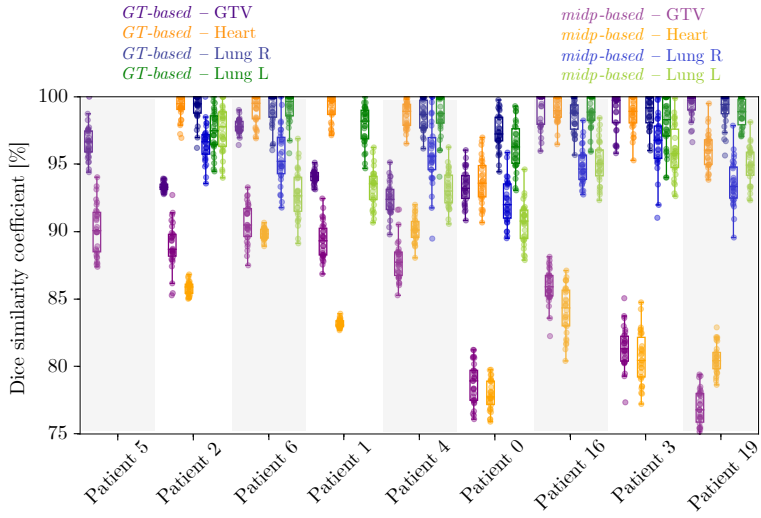


Fig. 4.4 Results of the DSC analysis on the four organs (GTV, heart, right and left lungs) for both GT-based (on the left) and midp-based (on the right) versions. The lungs and heart of Patient 5 were not delineated, as well as the right lung of Patient 1. Patients are sorted by increasing motion range in the test set.

The DSC results of the midp-based version are lower than those of the GT-based version, but still over 75%. As the same 50 images were used for both, the difference might be due to the approximations in the deformation and re-binarisation of the masks, which probably have a higher impact with deformations over multiple voxels, but this was not quantified.

4.3.4 Euclidean distance

The results of the ED analysis for both GT-based and midp-based versions are summarised in table 4.2. For the GT-based version, the mean, the median and the 95th percentile computed over the test set vary respectively from 0.11 mm to 2.15 mm, from 0.03 mm to 1.64 mm, and from 0.14 mm to 2.63 mm for the GTV; from 0.15 mm to 1.29 mm, from 0.27 mm to 1.42 mm, and from 0.39 mm to 1.92 mm for both lungs; from 0.14 mm to 1.68 mm, from 0.13 mm to 1.44 mm, and from 0.37 mm to 2.18 mm for the heart. For the midp-based version, the mean, the median and the 95th percentile computed over the test set vary respectively from 1.08 mm to 2.45 mm, from 0.61 mm to 2.80 mm, and from 0.87 mm to 3.08 mm for the GTV; from 0.18 mm to 1.58 mm, from 0.33 mm to 1.37 mm, and from 0.58 mm to 1.77 mm for both lungs; from 1.18 mm to 2.34 mm, from 0.87 mm to 1.73 mm, and from 1.11 mm to 2.69 mm for the heart.

Table 4.2 Results of the ED analysis for both GT-based and midp-based versions. ED_{GT} and ED_{midp} stand for the mean of the ED over the 50 images of the test set for the GT-based version and midp-based version, respectively. The lungs and heart of Patient 5 were not delineated, as well as the right lung of Patient 1.

Patient ID	GTV		Lung R		Lung L		Heart	
	ED_{GT} [mm]	ED_{midp} [mm]	ED_{GT} [mm]	ED_{midp} [mm]	ED_{GT} [mm]	ED_{midp} [mm]	ED_{GT} [mm]	ED_{midp} [mm]
Patient 0	2.08	2.32	0.91	1.56	1.29	1.12	1.68	2.34
Patient 1	1.68	1.08	NA	NA	0.86	0.51	0.25	1.72
Patient 2	2.09	1.21	0.63	1.12	0.89	0.23	0.14	1.58
Patient 3	0.11	1.83	0.52	1.21	0.18	0.18	0.48	2.07
Patient 4	2.15	1.47	0.68	0.92	0.69	0.52	0.68	1.18
Patient 5	0.95	0.74	NA	NA	NA	NA	NA	NA
Patient 6	0.59	1.38	0.15	1.58	0.20	0.41	0.16	1.23
Patient 16	0.21	1.56	0.75	0.29	0.77	0.24	0.32	1.69
Patient 19	0.24	2.45	0.46	0.37	0.55	0.28	0.96	1.87

Most of the time, the distance computed between the centroids of both 3D binary masks is under 2 mm. On average, the distance is smaller for the GT-based version, but this is not verified for every patient.

4.3.5 Normalised root mean squared error

The results of the NRMSE analysis are displayed in figure 4.5. The mean of this metric is lower for Patients 5, 2, 6 and 1 who have smaller motions in the test set (from 0.032 to 0.039) than the mean obtained for Patients 0,

16, 3 and 19 (from 0.047 to 0.051) who have larger motions. This is also observed for the median and the 95th percentile, which range respectively from 0.032 to 0.038, and from 0.039 to 0.045 for the first batch of patients, while they are respectively between 0.045 and 0.052, and between 0.051 and 0.059 for the second group of patients. This analysis also shows that the breathing phases have no impact on the reconstruction process as these are uniformly distributed along the NRMSE values range.

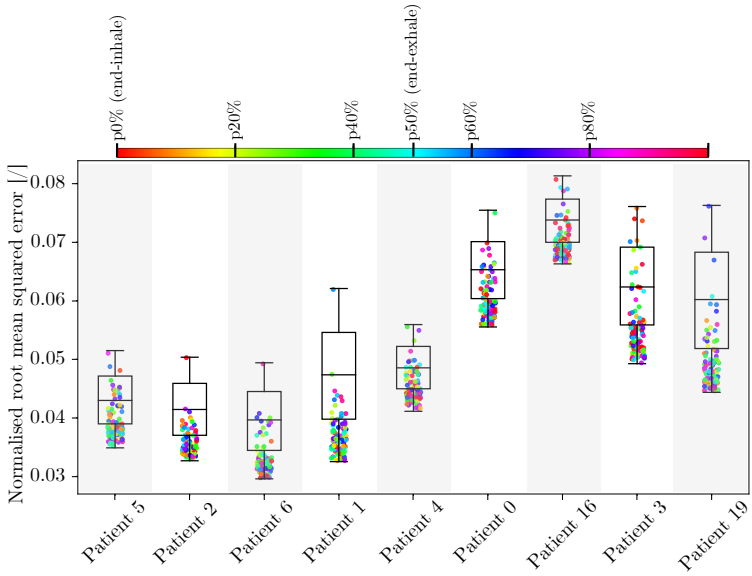


Fig. 4.5 Results of the NRMSE analysis. The NRMSE was computed between the ground-truth 3DCT image and the corresponding predicted 3DCT image for each test set data. The color of a dot represents the breathing phase at which the ground-truth 3DCT image was created. Patients are sorted by increasing motion range in the test set.

The results of the second part of this analysis, in which the results are compared with a baseline, are shown in figure 4.6. This figure shows that the method proposed in this work significantly reduces the error between the ground-truth image and the predicted image. In contrast, the baseline, which consists of predicting the midpCT image of each image in the test set, has a larger error, which is never less than 0.34 and can even be as high as 0.69.

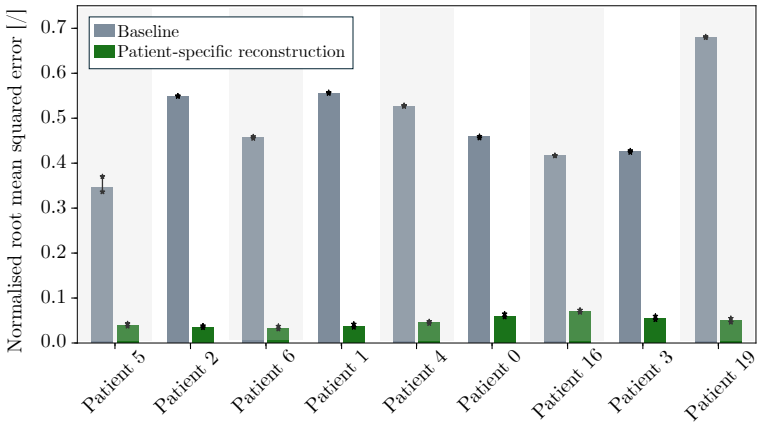


Fig. 4.6 Results of the NRMSE analysis. For the baseline, the NRMSE was computed between the ground-truth 3DCT image and the midpCT image. For our method, the NRMSE was computed between the ground-truth 3DCT image and the corresponding predicted 3DCT image. This was done for each data of the test set. Patients are sorted by increasing motion range in the test set.

4.3.6 Image quality metrics

The results of the three metrics assessing the quality of the images predicted by the neural network are presented in figure 4.7. This figure shows box plots of the (a) MAE, (b) PSNR and (c) SSIM for the 9 patients studied in this work. As was already the case in the NRMSE analysis, the MAE increases for patients with greater motion ranges in the test set. This distribution is reversed for the other two metrics because the PSNR and SSIM metrics must be maximised, whereas the goal is to minimise the NRMSE and MAE metrics. The MAE is lower for patients with smaller movements in the test set. The average value of the MAE is around 17.5 HU for the six patients with the smallest movements, while it is around 27.5 HU for the three patients with the largest movements. The mean of the PSNR for the six patients with the smallest movements is around 42.5 dB, whereas it is around 37.5 dB for the three patients with the largest movements. The mean of the SSIM for the six patients with the smallest movements is around 0.98, whereas it is around 0.95 for the three patients with the largest movements.

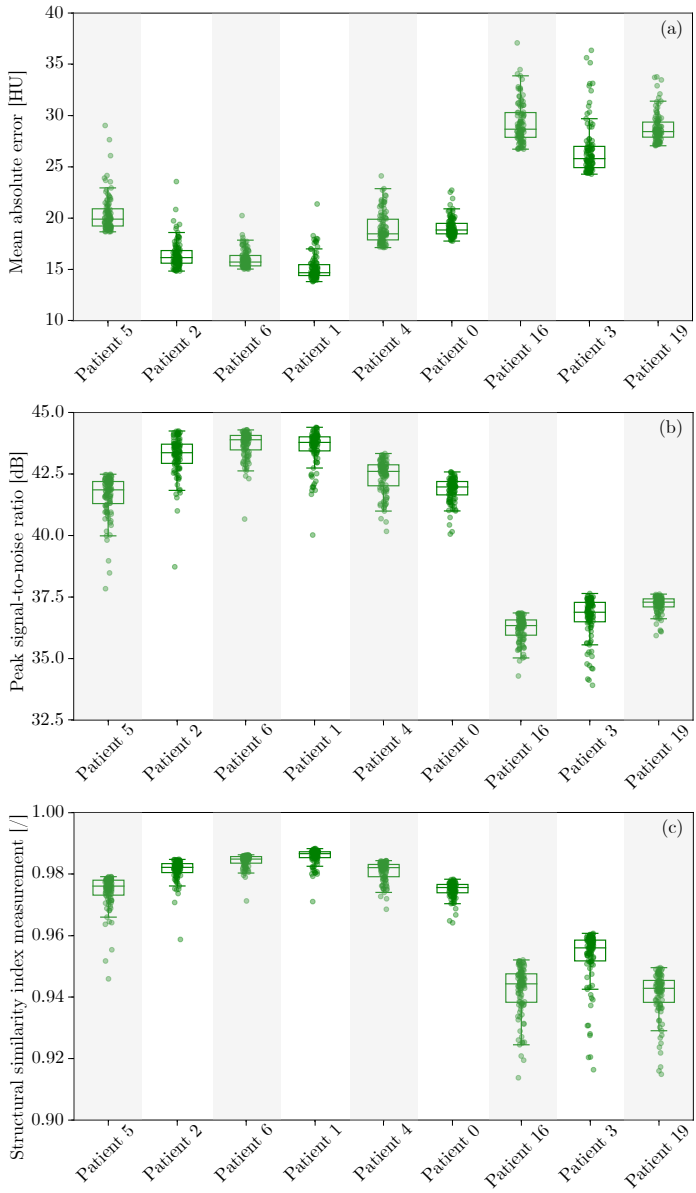


Fig. 4.7 Box plots of (a) MAE, (b) PSNR and (c) SSIM between all GT 3DCTs and corresponding predicted 3DCTs of the test set, for the nine patients. Patients are sorted by increasing motion range in the test set.

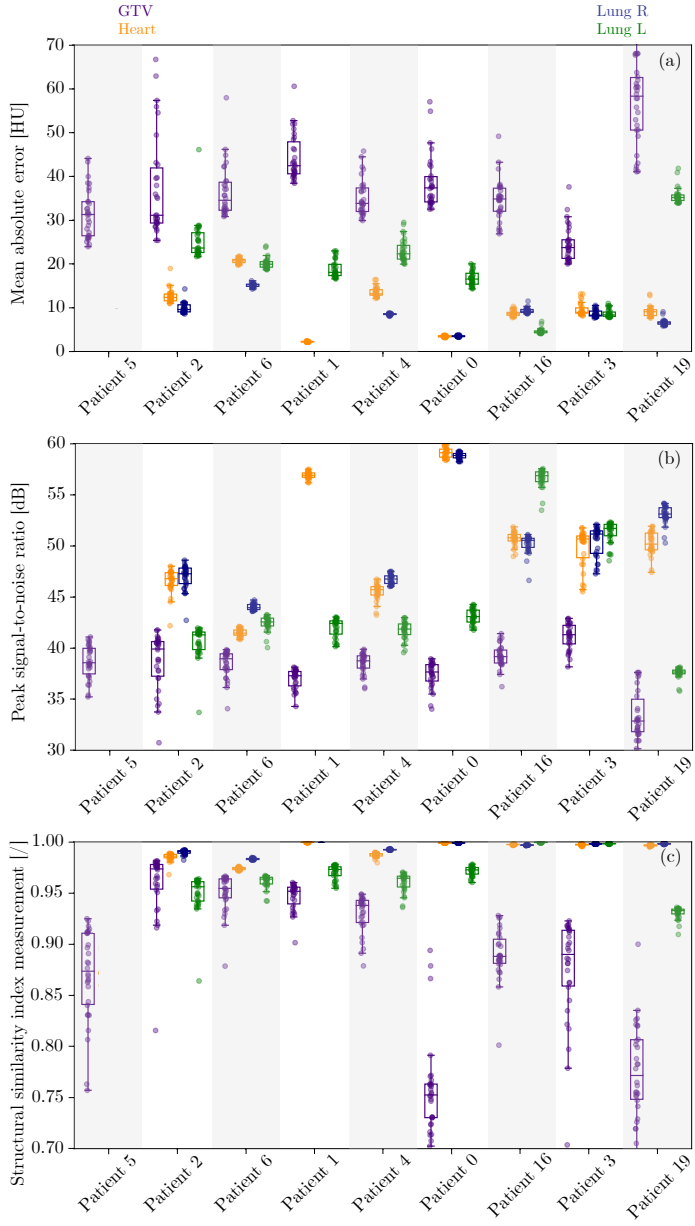


Fig. 4.8 Box plots of (a) MAE, (b) PSNR and (c) SSIM between all GT 3DCTs and corresponding predicted 3DCTs of the test set, for the nine patients focused on organs of interest. Patients are sorted by increasing motion range in the test set. The lungs and heart of Patient 5 were not delineated, as the right lung of Patient 1.

The results of the second part of this analysis, in which the metrics are computed on the organs of interest, are shown in figure 4.8. The results show that the GTV is generally the worst reconstructed organ. The means of the MAE, PSNR and SSIM computed over this organ are between 20 HU and 70 HU, 30 dB and 40 dB, and 0.70 and 0.95. In contrast, the reconstruction quality and accuracy on the other three organs appear to be similar, as the values of the metrics are generally within the same range of magnitudes. In most cases, the MAE calculated on these three organs is below 40 HU, the PSNR above 40 dB and the SSIM over 0.95.

4.3.7 Reverse process

The results of the comparison between the digitally reconstructed radiograph generated using one ground-truth 3DCT of the test set and the digitally reconstructed radiograph generated using the associated predicted 3DCT for Patient 3 are displayed in figure 4.9. Visually, the DRR generated using the predicted 3DCT image is identical to the one obtained using the corresponding ground-truth 3DCT image. However, the analysis of the difference computed between these two images shows that they are not identical, with an average difference equal to 0.055 kV. The biggest differences are located in the diaphragm area, with absolute values of around 2 kV. Generally, a red pixel is adjacent to a blue pixel, meaning that the edges of this organ are the zones most poorly reconstructed by the method.

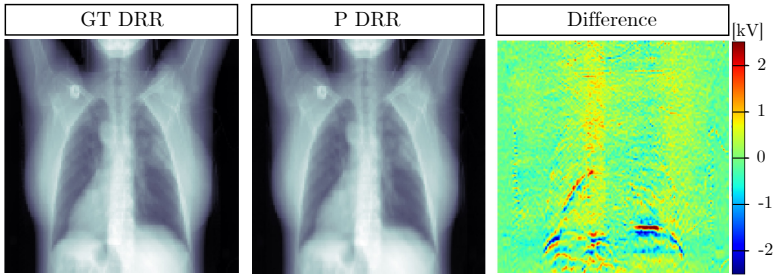


Fig. 4.9 Patient 3 results for the reverse process. The digitally reconstructed radiograph generated from one image of the test set (left column) is compared with the DRR generated using the corresponding predicted 3DCT image (centre column). The difference between both images is displayed in the right column.

Figure 4.10 shows the values of the (a) MAE, (b) PSNR and (c) SSIM metrics calculated between the ground-truth DRR and the predicted DRR, for the different patients. This figure shows that six patients obtain excel-

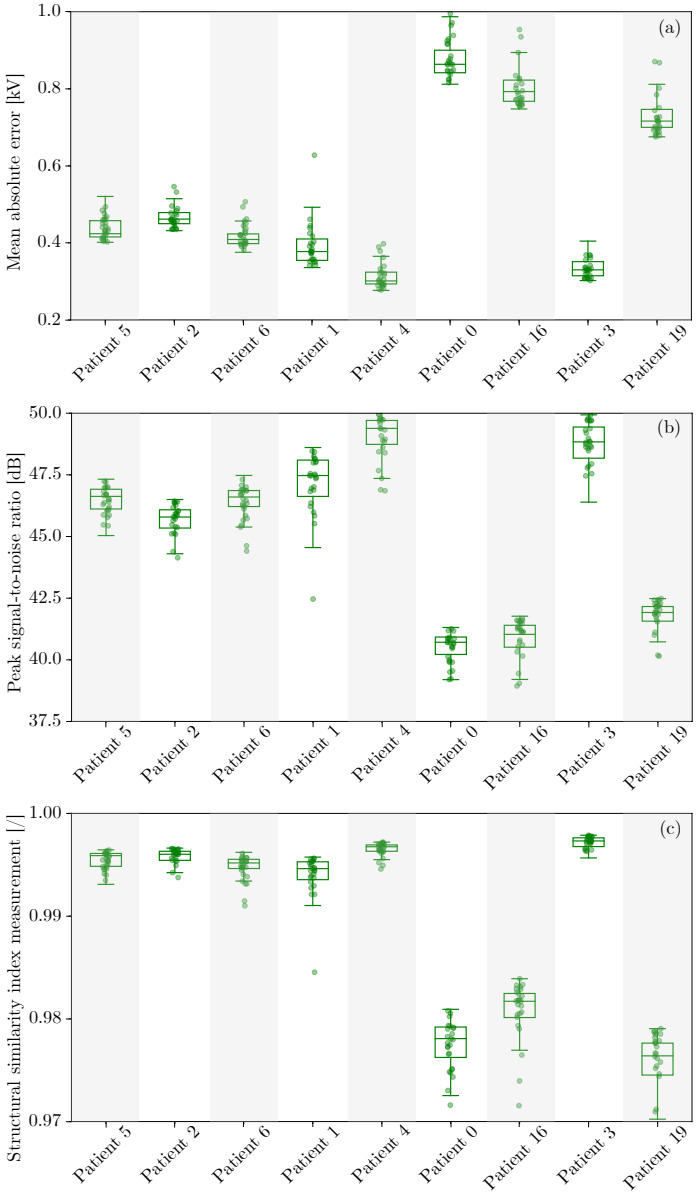


Fig. 4.10 Box plots of (a) MAE, (b) PSNR and (c) SSIM computed between the ground-truth DRR and the DRR generated from the predicted 3DCT, for all images of the test set. Patients are sorted by increasing motion range in the test set.

lent results: the MAE is less than 0.5 kV, the PSNR is greater than 45 dB and the SSIM is above 0.99. On the other hand, the other three patients have slightly poorer results, with MAE around 0.8 kV, PSNR around 40 dB and SSIM of 0.98. These three patients are in fact the three patients diagnosed with a liver cancer. These differences can probably be explained by a difference in the image acquisition protocol.

4.4 Discussion

In this chapter, it has been shown that the proposed CNN-based methodology allows to reconstruct a high-quality 3DCT image from a single digitally reconstructed radiograph. The proposed method requires a patient-specific training and is based on a patient database created using a data augmentation tool.

The Dice values computed between the masks of the predicted 3DCT image and the corresponding ground-truth 3DCT are all greater than 75%, which is reliable. The comparison of the results of the midp-based version obtained for lungs and heart (94.6% and 83.9%, displayed in figure 4.4) with the results of previous works [ZZQ+19, DLW+19, FQT+19], whose goal was to segment organs at risk in lung cancer utilising deep learning algorithms (best in [FQT+19] and equal to 97.5% and 92.5%), shows that lungs have similar results to the literature and the heart has a higher difference. However, our results should be taken in hindsight, given that the masks in the predicted image are defined as the manually segmented masks on the midpCT image deformed using the deformation fields obtained by the Morphons registration between both images.

The mean of the difference between the ground-truth image and the predicted image is small for each patient, with an average value of 0.45 HU over all patients. Comparing these results (figure 4.2) with those obtained by [SZX19] when they use only one view, the quality of our reconstructed image is similar to their own. Their method also performs less at tissue borders. However, there is no scale or numerical value in their difference analysis, so it is not clear that the difference values are similar.

The results of the reverse process show that the digitally reconstructed radiograph generated using the predicted 3DCT image is similar to the reference DRR. This highlights that the 3DCT reconstruction is properly

performed and that the depth of the image is consistent, which is critical for the dose calculation step.

One limitation of this study is that the CNN was trained using training sets composed of 3DCT images created from deformations of a planning 4DCT acquired prior to treatment and paired DRRs generated using the Beer-Lambert absorption-only model. This method supposes that inter-fractional anatomical variations such as tumor shrinkage, tumor baseline shift, and stomach and bladder fillings are not included in the training set, but are checked prior to treatment is administered. A next step of this work is to evaluate whether the network must be retrained for each fraction or whether these variations are negligible in the reconstruction process. Another possibility to counteract this limitation is to improve the data augmentation tool and incorporate inter-fraction motion in the training set. If this works, it would mean that it is possible to generate high-quality 3DCT images at any fraction of the treatment based solely on the planning 4DCT. This means that it is possible to have a good knowledge of the patient anatomy at all times, without increasing the imaging dose given to the patient too much. Using this method, it would be possible to generate 3DCT images based only on x-rays projections acquired to position the patient correctly, or others acquired during the fraction. This would not increase the time and cost of the treatment, but would drastically reduce the motion margins and therefore the dose delivered to the surrounding healthy tissues.

An additional potential purpose of the predicted 3DCT image would be to use it to compute the dose delivered during the treatment (either on-line or inter-fraction). To this end, the voxel value representing tissue density is a crucial piece of information to have the dose delivered at the right place. This chapter shows that, for the human eye, the predicted 3DCT image is really close to the ground-truth 3DCT image but the results of the difference should be discussed further and it will be necessary to assess whether the maximum of the difference is located on the trajectory of the beam or whether the difference, no matter how small, has too great an impact on the computed dose. Furthermore, in order to get a clinically usable dose, the standard resolution of a 3DCT scan would be needed. Therefore, the predicted 3DCT image should be oversampled to get the desired resolution.

4.5 Summary

This chapter presents a method that allows the reconstruction of a 3DCT image from a single digitally reconstructed radiograph. This method relies on a data augmentation algorithm and on a patient-specific training of a CNN. The performance of the method was evaluated according to several metrics. For example, the Dice similarity coefficient was computed between the masks of the ground-truth image and those of the predicted image, as well as the PSNR between these two images or on regions of interest. However, it is still necessary to integrate inter-fraction motion in the training set and to estimate the accuracy of the images in terms of dose delivery in order to confirm the potential clinical use of the method.

5

Variants in the patient-specific 3DCT reconstruction

Numerous variants of the main method proposed in this thesis could have been studied: the contribution of temporal knowledge of the anatomy by giving a sequence of DRRs as input to the network, the added value of spatial knowledge by combining the position of an external surrogate and a DRR at the input of the network, the benefit of inter-patient anatomy knowledge by including a multi-patients pre-training step, etc. Unfortunately, each change involves new data generation and new training, which require a huge amount of computational and storage resources, and, obviously, of time. This work focuses on two variations that have a considerable impact on the desired clinical implementation of the method. The first variant, investigated in section 5.1, aims to determine the robustness of the method to changes in layout. In particular, we study two training methods for adapting to a change in the orientation and number of the projection angles. The second variant, described in section 5.2, seeks to study the robustness of the method to changes in image acquisition time. More specifically, we study two training methods for adapting to inter-fractional anatomical changes.

5.1 Robustness to changes in layout

The basis of this work was presented at the IEEE International Conference on Image Processing in Bordeaux in 2022 [LDSM22a]. Additional analyses are provided in this thesis.

5.1.1 Context

Radiation therapy kills cancer cells by delivering radiation into the tumor, while sparing surrounding healthy tissue and organs at risk. Thoracic and abdominal tumors are challenging to treat with radiation therapy as the breathing motion induces the movement of organs. In case of mobile tumors, it is important to take into account the movement of the tumor during treatment planning. The acquisition of a planning 4DCT before the treatment allows to estimate the tumor motion, to derive safety margins, and to apply 4D-robust optimisation to ensure target coverage. Safety margins reduce the risk of mistreating the tumor but increase the dose delivered inside the surrounding healthy organs [RB10]. Unfortunately, the tumor motion in the treatment room can differ significantly from the movement captured in the 4DCT, leading to errors in the treatment delivery [DVB+18].

Different methods have been developed to address the motion-related issues and increase confidence in tumor localisation. Respiratory gating is a common method and employs an external surrogate to follow the breathing movement. However, this technique relies on a strong relationship between the tumor motion and the surrogate position [MFD+10]. The placement of fiducial markers allows to pinpoint the tumor position with greater accuracy, but then requires a heavy and risky surgery before the treatment [HWST19]. Other methods have also been studied to reduce the motion amplitude or to obtain a regular movement. These methods use abdominal compression [PHNACG18], audio coaching [NNM+09] or mechanically assisted ventilation [VODSLG19].

The total reduction of safety margins would be to adapt the treatment in real-time. To achieve this, the positions of the target and surrounding organs must be known throughout the treatment delivery. Image-guided radiation therapy (IGRT) uses frequent imaging during the treatment to improve the precision and accuracy of the treatment delivery. A broad

range of IGRT modalities is now available and adopted. X-rays projections are commonly acquired to evaluate and correct setup errors in order to correctly target the tumor with the radiation beam. However, this technique has two major drawbacks. One is that radiographs do not provide 3D information about the tumor volume or its spatial localisation [VDRT+08]. Another is that position changes are typically restricted to simple translational adjustments as most linear accelerators are not equipped with rotational adjustment systems.

Many studies that reconstruct a 3D volume from a 2D x-rays projection have already been performed. Different fields of application have been explored. Henzler et al. in [HRRR17] investigated how to reconstruct 3D volumes from 2D cranial x-rays by applying deep learning, while Liang et al. in [LSY+21] developed a new model architecture to reconstruct a 3D tooth from a single panoramic radiograph. Montaya et al. in [MZL+21] demonstrated that it is possible to reconstruct a 3D image from two scout views using deep learning and Shen et al. in [SZX19] used an encoder-decoder framework to reconstruct one 3D image from a single projection.

In this context, the aim of this part of the thesis was to study the robustness of the 3DCT reconstruction methodology to layout changes. In particular, the impact of the orientation and number of the projection angles is evaluated in order to quantify whether the method performs better when the x-rays imager is ceiling-mounted or integrated into the gantry, and to assess if it is possible to change the projection angle during the treatment. To this end, two training strategies were explored and compared.

5.1.2 Methodology

Dataset generation

The data used in this part of the work come from three patients treated for liver tumors by radiotherapy. For each patient, a planning 4DCT composed of 10 breathing phases was acquired before the start of treatment. However, the training of a deep learning model requires a large amount of data. To tackle this problem, new 3DCT images, representing intra-fractional anatomical deformations, were created using the method previously developed by our research group and described in section 3.2.3. This method is based on the computation of a motion model using the planning 4DCT and a random respiratory signal. The signal aims to reproduce a

maximum of different and possible anatomies, and can also generate extreme respiratory phases that are not observed in the original 4DCT. In this work, 20 random respiratory signals with various amplitudes and periods were simulated, from which 25 random samples were selected to create matching images with the motion model. This results in creating 500 3DCT images for each patient. From each created 3DCT image, 5 digitally reconstructed radiographs were generated with different projection angles: 0°, 30°, 45°, 60°, 90°. This can be observed in figure 5.1. To simulate real-world x-rays imaging, the Beer-Lambert absorption-only model implemented in the TomoPy Python package was used [GDCXJ14]. The 500 3DCT images and the associated DRRs were then divided into two distinct sets, the training set and the test set. 90% of the 500 images were used for training, and the remaining 10% for testing.

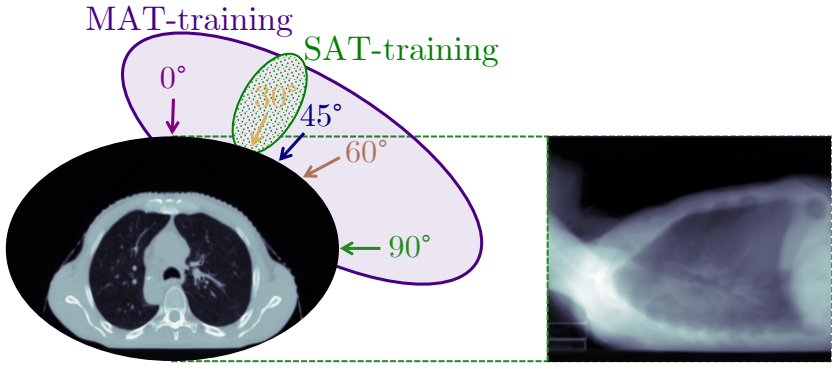


Fig. 5.1 Projection angles used to generate the DRRs from the 3DCT image, the DRR obtained with an angle of 90° is displayed. SAT-training and MAT-training are the two training methods studied in this work. SAT-training trains the network with DRRs generated with a single projection angle, while MAT-training uses DRRs generated with the 5 projection angles.

Network

To address the problem of reconstructing a 3DCT from a single 2D x-rays image, the neural network designed by Henzler et. al. in [HRRR17] and described in section 3.3 was used. This network is a convolutional neural network (CNN) that learns the mapping between a 2D x-rays image and a 3D volume. The input of this network is an image of size 256×256, while the output of the network consists of a $128 \times 128 \times 128$ volume. An overview of the network can be seen in figure 3.12.

Training specifications

The neural network training was patient-specific, it used patient-specific CT data to refer to individual features. The network was trained with 500 epochs. This work studied two main methods to train the model: *single-angle training* (*SAT-training*) and *multi-angles training* (*MAT-training*).

SAT-training method used as input of the network DRRs that were all obtained with the same projection angle. Therefore, the training was performed independently for each angle value.

MAT-training method used as input of the network DRRs that were obtained with the 5 different projection angles. In this case, the network was trained only once using all projection angles.

Performance evaluation

In order to evaluate the performance of the proposed method, the 50 pairs of DRR/3DCT of the test set were used. The outputs of the trained CNN, which are called the predicted 3DCT images, were compared with the ground-truth 3DCT images of the test set through various analyses.

Visual analysis was carried out to qualitatively assess the performance of both training methods. To that end, the absolute value of the difference was computed between a ground-truth 3DCT image and the corresponding predicted 3DCT. The image used in this analysis was the first image of the test set of Patient 16.

Various quantitative analyses have been carried out to evaluate the results of the proposed training strategies. The different quantitative analyses were based on four metrics commonly used in image processing: normalised root mean squared error, mean absolute error, peak signal-to-noise ratio and structural similarity index measurement. Each of these metrics was computed for the 50 images composing the test set, between the ground-truth image and the corresponding predicted image. Each quantitative analysis aims to answer a specific research question.

What is the impact on the predictions quality if the projection angle used to train the network is different? This analysis looks at the variation in NRMSE results when the projection angle used to generate the DRRs given as input to the network trained with SAT-training is different. In this anal-

ysis, SAT-training was performed with the five projection angles.

What is the impact on the predictions quality if several projection angles are used to train the network? This analysis looks at the variation in image quality metrics results between the two training methods. In this analysis, SAT-training was performed with a projection angle of 0°.

What is the impact on the predictions quality if the projection angle used to test the network is different from that used to train it? This analysis looks at the variation in NRMSE results of both training methods when the projection angle used to generate the DRR for testing the network is different from that used for training it. This analysis was divided into two parts: the first part consisted in studying large angle changes, while the second part focused on small angle changes. In both cases, 13 DRRs were generated with different projection angles for the 50 ground-truth 3DCT images of the test set. For the first part, the projection angles were all multiples of 5°, between 0° and 60°. For the second part, the projection angles were all integers between 24° and 36°. In this analysis, SAT-training was performed with a projection angle of 30°.

5.1.3 Results

Visual analysis

The results of the visual analysis are represented in figure 5.2.

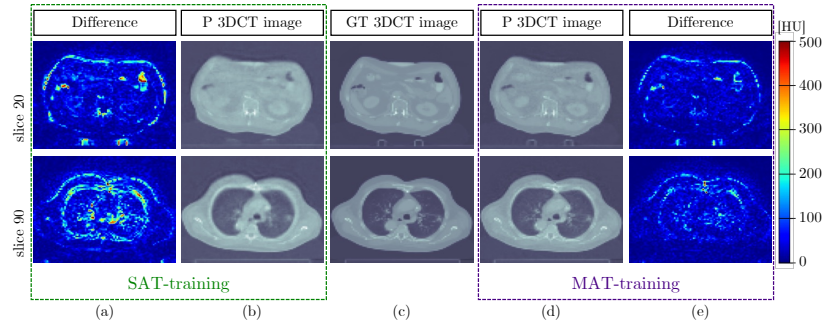


Fig. 5.2 Visualisation of two transverse slices of the (c) GT 3DCT, (b) predicted 3DCT using SAT-training and (d) predicted 3DCT using MAT-training. (a) and (e) display the absolute value of the difference (in HU) for the same slices using SAT-training and MAT-training, respectively.

A first observation is that the contours of the anatomical structures are sharper in the image predicted by the multi-angles training method (figure 5.2(d)) than by single-angle training (figure 5.2(b)). The maximum absolute value of the difference between the ground-truth 3DCT and the predicted 3DCT is greater than 1400 Hounsfield Units (HU) using single-angle training (figure 5.2(a)), while it is significantly reduced to around 600 HU using multi-angles training (figure 5.2(e)). The median of the difference using single-angle training is equal to 4.47 HU and is reduced to 1.77 HU using multi-angles training. Using the single-angle training method, 16.5% of the volume has a difference greater than 30 HU, which is decreased to only 8.9% using the multi-angles training method. Most of the large errors are located on the skin of the patient, on the table or on the belt.

Impact of the value of the projection angle used for training

The results of the first quantitative analysis are represented in figure 5.3.

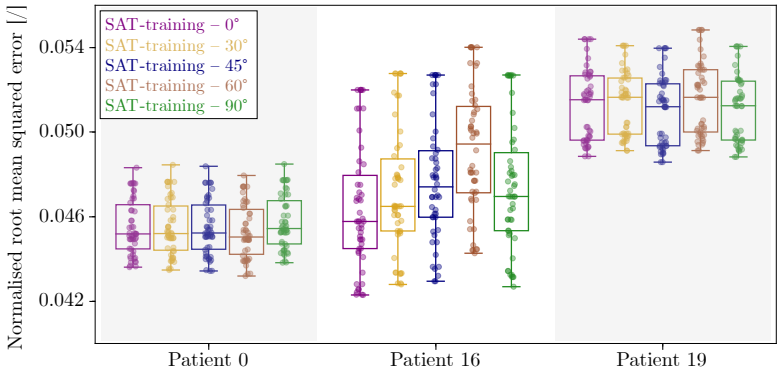


Fig. 5.3 Box plots of the NRMSE computed between the GT 3DCT and the predicted 3DCT, for all images of the test set. For each patient, the results are shown as a function of the projection angle used to train the network using SAT-training.

Figure 5.3 shows the distribution of the NRMSE for each patient, and each projection angle. The mean of the five medians is 0.045 for Patient 0, 0.047 for Patient 16 and 0.051 for Patient 19. The training of Patient 16 is more variable as the results show a larger standard deviation equal to 0.0029. Nevertheless, the analysis highlights that the performance of the network is similar whatever the value of the projection angle used to generate the DRRs given as input to the network.

Impact of the number of projection angles used for training

The results of this analysis are shown in figure 5.4 and table 5.1. Figure 5.4 displays the distribution of the NRMSE, while table 5.1 lists the means and standard deviations of the MAE, PSNR and SSIM metrics for the three patients studied in this part of the work.

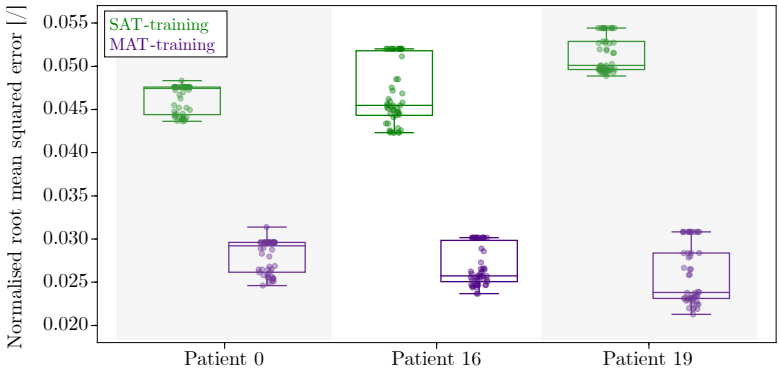


Fig. 5.4 Box plots of the NRMSE computed between the GT 3DCT and the predicted 3DCT, for all images of the test set. For each patient, the results are shown for SAT-training on the left and for MAT-training on the right.

Table 5.1 Results of the MAE, PSNR and SSIM metrics over the whole 3DCT image. Each value in the table stands for the average ± standard deviation of the metric over the 50 images in the test set.

Patient ID	SAT-training			MAT-training		
	MAE [HU]	PSNR [dB]	SSIM [/]	MAE [HU]	PSNR [dB]	SSIM [/]
Patient 0	24.80 ± 1.01	37.20 ± 0.32	0.955 ± 0.01	15.44 ± 0.92	41.58 ± 0.61	0.982 ± 0.01
Patient 16	24.39 ± 1.54	37.12 ± 0.61	0.966 ± 0.01	15.86 ± 1.33	41.96 ± 0.66	0.981 ± 0.01
Patient 19	18.08 ± 0.84	39.39 ± 0.35	0.969 ± 0.01	9.216 ± 1.05	45.42 ± 1.13	0.990 ± 0.01
Mean	**22.42 ± 1.13**	**37.90 ± 0.43**	**0.963 ± 0.01**	**13.51 ± 1.10**	**42.99 ± 0.80**	**0.984 ± 0.01**

An overall analysis shows that both training methods give a robust neural network, with very low standard deviations. The NRMSE metric has standard deviations ranging from 0.0018 for Patient 0 to 0.0033 for Patient 16 using SAT-training, and from 0.0019 for Patient 0 to 0.0034 for Patient 19 using MAT-training. The MAE metric has standard deviations, ranging from 0.84 HU for Patient 19 to 1.54 HU for Patient 16 using SAT-training, and from 0.92 HU for Patient 0 to 1.33 HU for Patient 16 using MAT-training. The standard deviations of the PSNR are between 0.32 dB

and 0.61 dB using the single-angle training method, whereas they are between 0.61 dB and 1.13 dB using the multi-angles training method. Although both training methods produce robust methods, the MAT-training method gives better results. For the three patients, the multi-angles training method obtains lower NRMSE and MAE, as well as higher PSNR and SSIM. In the order NRMSE, MAE, PSNR and SSIM, the average computed over the three patients is equal to 0.047, 22.42 HU, 37.90 dB and 0.963 using the SAT-training method, whereas it is 0.027, 13.51 HU, 42.99 dB and 0.984 using MAT-training.

Impact of a projection angle different from that used for training

The results of the third analysis are displayed in figure 5.5.

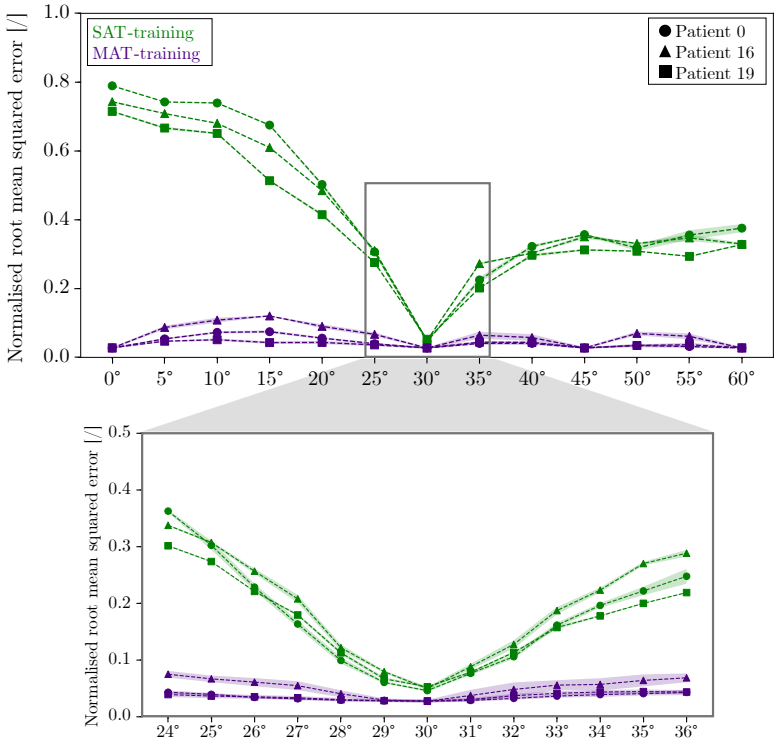


Fig. 5.5 NRMSE between the GT 3DCTs and the predicted 3DCTs of the test set. In both analyses, the predictions are made with DRRs obtained at the 13 projection angles. The results are shown for the three patients and for both training methods.

The analysis differs with the training method chosen. Using the single-angle training method, the NRMSE is minimal when the projection angle used for the prediction on the test image is the same as the training projection angle (30°). The NRMSE value rises rapidly and linearly when the angle varies. For example, using a projection angle of 25°, the mean of the NRMSE is between 0.275 and 0.310, and using a projection angle of 35°, the mean of the NRMSE is between 0.201 and 0.273, depending on the patient. Moreover, the figure shows that the behaviour of the error is not the same for angles below or above 30°. Depending on the patient, the mean of the NRMSE is between 0.715 and 0.789 for a projection angle of 0°, while it is between 0.328 and 0.376 for a projection angle of 60°. One reason might be that the DRRs generated with angles between 30° and 60° are more similar to each other, whereas DRRs generated with smaller angles are more different. Using the multi-angles training method, the mean of the NRMSE is maximum with a projection angle of 15° and is between 0.043 and 0.120. The results demonstrate that the multi-angles training method is more robust to a change of the projection angle, and an interval of 15° between two successive projection angles in the training set allows to reduce the error between the GT 3DCT and the predicted 3DCT.

5.1.4 Discussion

This work shows that the performance of the neural network trained with SAT-training is similar regardless of the projection angle used to generate the DRRs given as input to the network. Any configuration of the treatment room can then be used to obtain a 3DCT image from the x-rays projection. This means that the fluoroscopy acquisition device can either be placed on the gantry or fixed in the treatment room.

The results also show that the neural network performs better when it receives as input five digitally reconstructed radiographs for a given 3DCT. The main limitations of MAT-training are the difficulty to acquire five x-ray projections simultaneously in the treatment room as there are rarely multiple imagers, and the related increase in patient irradiation. On the other hand, the main advantage of this method is that it is able to reconstruct high-quality 3DCT images using DRRs generated with projection angles that are not used to generate the DRRs of the training set. This means that this method is robust to errors in the projection angle. If the x-rays imager is integrated into the gantry and moves during patient treatment, thereby

altering the projection angle, the MAT-training method is still able to obtain a high quality 3DCT image from fluoroscopy.

However, some limitations of this work still need to be investigated before considering a reliable clinical application.

The method proposed by this work requires a patient-specific training of a convolutional neural network, which is performed using synthetic 3DCT images created from deformations of a planning 4DCT. This supposes that inter-fractional anatomical variations such as tumor shrinkage, tumor baseline shift, or even stomach/bladder filling are not included in the training set. It would be interesting to assess whether a re-training is required for each treatment fraction or whether these variations are negligible in the reconstruction of a 3DCT as they are hardly visible on the x-rays projections.

The results presented in this section are generated from a test set composed of 50 random 3DCT images. It could be possible that images in the test set are similar to each other and to those of the training set. It would be interesting to generate a test set composed of 3DCT images depicting several extreme breathing states, e.g. deep inspiration and deep expiration, to evaluate the robustness of the neural network to situations not observed during training. This would confirm that the method is reliable for all breathing patterns, from regular breathing to irregular breathing with periods of stress, apnea or coughing.

The results presented in this section are computed on cropped images but that still leave a wide border around the patient. This border could bias the results as there is no material to reconstruct at this place. This could also explain the difference in the results between patients. A thicker patient with a narrower edge would have a higher error value, while the reconstruction within the body of this patient might be better.

The results presented in this section are based on only three patients, and the method should first be tested on a large number of patients to ensure that it works well with different anatomies.

5.2 Robustness to changes in image acquisition time

This part of the work is inspired on the journal paper submitted to Medical Physics, currently under peer review. Additional analyses are provided in this thesis.

5.2.1 Context

Radiation therapy has an important role in the treatment of lung cancer. This technique involves the precise delivery of ionising radiation to the tumor, with the aim to minimise the dose to healthy tissues and hence reduce treatment side effects. Accurate delineation of the treatment area is one of the most important steps in radiation therapy. Margins are usually added around the gross tumor volume to account for microscopic disease and setup errors. Larger safety margins are employed for lung cancer to consider the movement of the tumor as a result of breathing motion [MBvH21].

Radiation therapy is typically delivered in 30 fractions over 6 weeks for lung cancer. The movements of the patient and of its internal organs during these fractions can be divided into two categories [PLWM16]: intra-fractional anatomical variations and inter-fractional anatomical variations. Intra-fraction motion indicates changes when the patient is undergoing radiation therapy, while inter-fraction motion is the variations observed between different treatment sessions. Intra-fraction motion considers deformations related to respiratory and cardiac cycles occurring on a time scale of seconds to minutes. Inter-fraction motion covers baseline shifts and weight gain or loss, which occurs on a time scale of hours or days. It also includes changes in the position of the patient.

Image-guided radiation therapy has therefore a crucial role in identifying the anatomical changes during treatment. During the course of fractional treatments, two-dimensional x-rays projections are acquired and compared with the planning image [RLLL19b]. The projection radiography is acquired after patient positioning to determine the position error and the error of the radiation field, which are immediately corrected to obtain the appropriate position of the target area. However, this technique has two major drawbacks. One is that radiographs do not provide 3D information about the tumor volume or its spatial localisation [VDRT+08]. Another is that it requires implanted fiducial markers to visualise the tar-

get on the x-rays projections, which involves a major surgery.

Numerous studies have already been carried out on the reconstruction of a 3D volume from a 2D projection radiography. Various fields of application have been explored. Henzler et al. in [HRRR17] studied how to reconstruct 3D volumes from 2D cranial radiographs by applying deep learning, while Liang et al. in [LSY+21] developed a new model architecture to reconstruct 3D teeth from a single panoramic radiograph. Shen et al. in [SZX19] used an encoder-decoder framework to reconstruct a 3D image from a single projection view, and Montaya et al. in [MZL+21] demonstrated that is was possible to reconstruct a 3D image from two scout views with deep learning.

In this context, the aim of this part of the thesis was to study the robustness of the three-dimensional computed tomography (3DCT) reconstruction methodology to image acquisition time changes. In particular, the impact of the introduction of inter-fractional anatomical deformations in the training set is evaluated, as well as a change in the day of acquisition between training images and testing images to simulate different treatment fractions. To this end, two training strategies were explored and compared.

5.2.2 Methodology

Dataset generation

The dataset includes ten patients who received radiotherapy for lung cancer. A planning 4DCT composed of ten breathing phases evenly spread over the respiratory cycle was acquired prior to treatment. For each patient, a second 4DCT was acquired after several fractions of the treatment. For seven patients out of ten, a third 4DCT was acquired usually one week after the second. All 3DCT images have the same dimensions, which are $512 \times 512 \times 173$ with a voxel size of 1.172 mm $\times$ 1.172 mm $\times$ 2 mm. The mid-position (midp) CT image, defined as the local mean position in the breathing cycle [WSvHD08], was calculated using the mean of all velocity fields obtained by non-rigid registration between all 4DCT phases for all three 4DCTs, creating the planning midpCT, the T2 midpCT and the T3 midpCT. Figure 5.6 shows the coronal view at the center of mass of the target for the two or three midpCTs of each patient, and it also highlights the number of days between the acquisitions of the various 4DCTs. The gross tumor volume (GTV) and other surrounding organs at risk (left

and right lungs, and heart) were manually delineated by an experienced radiation oncologist on each midpCT. A binary body mask was used to exclude unnecessary background from each 3DCT image, and all images were cropped by the outer cube of the body masks in the planning midpCT with a margin of 15 mm. All images were then resized to $128 \times 128 \times 128$, and the contrast of the images was pre-processed to emphasise the different types of tissue and make the training process easier.

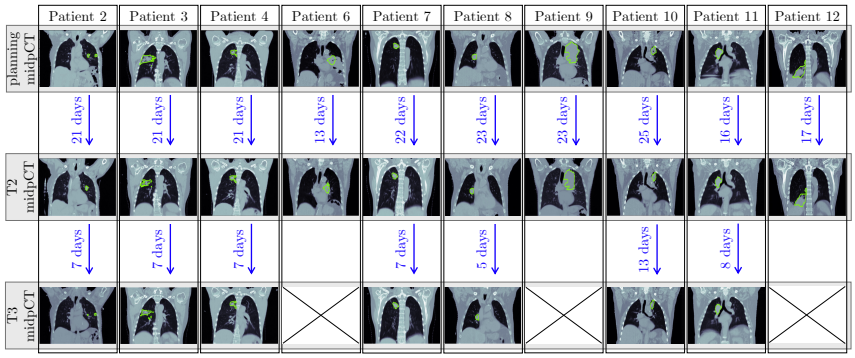


Fig. 5.6 Coronal views of pre-processed midpCT images at the center of mass of the GTV, and time gap between each acquisition. All midpCT images of the same patient are cropped with the same box to exclude unnecessary background.

An open-source data augmentation tool was implemented in OpenTPS to generate a multitude of 3DCTs of a patient, resulting from deformations of a midpCT image [WDSJ+23]. This data augmentation tool is composed of two main classes of deformations: intra-fractional changes and inter-fractional changes. The method used to create new 3DCTs including only intra-fractional changes was explained in chapter 4, it involves generating any breathing phase at a certain normalised distance from the midpCT using a linear interpolation between the two closest breathing phases and a scaling. To meet the needs of this study, the method was made more complex to take into account the inter-fractional anatomical changes that may occur between two sessions of the treatment, not just same-day variations. For example, it is possible to apply a translation and/or a rotation on the motion model to represent an error in the treatment position of the patient. It is also possible to apply shrinkage of an organ mask to represent phenomena such as tumor shrinkage, or to apply a baseline shift to the tumor mask to move it inside the lung [WDSJ+23].

The planning midpCT was used to generate 355 new inter-fractional anatomies. To best represent the changes that can truly occur during treatment, specific values for the inter-fractional changes were defined. The angle used for the rotation was a random sample from a uniform distribution $\mathcal{U}(-3,3)$ around the axes cranio-caudal and anterior-posterior, and $\mathcal{U}(-1,1)$ around the axis left-right. The value of the angle is expressed in degrees. The distance used for the translation was a random sample from a uniform distribution $\mathcal{U}(-6,6)$ in the directions cranio-caudal and left-right, and $\mathcal{U}(-1,1)$ along the direction anterior-posterior, while the distance used for the baseline shift was a random sample from a uniform distribution $\mathcal{U}(-5,5)$ in the three directions. The values of the distances are expressed in millimeters. Then, the width used for the shrinkage was a random sample from a uniform distribution $\mathcal{U}(0,3)$ in the three directions. This value is expressed in millimeters, and is then compared with the pixel spacing to determine how many voxels need to be removed from the mask.

The 355 new anatomical variations were divided into three groups of sizes 250, 65 and 40 to create intra-fractional changes. The 250, 65 and 40 anatomical models in each group were used to generate sequences of 25, 75 and 150 intra-fractional variations. In total, $25 \cdot 250 + 75 \cdot 65 + 150 \cdot 40 =$ 17.125 3DCTs were created. A DRR was derived from each of them using a projection angle of 0° along the anterior-posterior axis and the TomoPy library [GDCXJ14]. All pairs of DRR/3DCT formed the training and validation sets, with 80% for the training set and 20% for the validation set.

The performance of the network was studied on the basis of four distinct test sets: *Test Set planning* (*TSTp*), *Test Set T2* (*TST2*), *Test Set Translated T2* (*TSTT2*) and *Test Set Translated T3* (*TSTT3*) as illustrated in figure 5.7. To create these four test sets, 50 breathing amplitudes and phases were randomly selected from a normal distribution $\mathcal{N}(1,0.5)$ truncated between 0.8 and 1.5, and from a uniform distribution $\mathcal{U}(0,1)$, respectively. Each pair of amplitude and phase was given to the intra-fractional data augmentation tool to create 50 3DCTs. Each test set has its own particularities.

Test Set planning consists of 50 3DCT images derived from intra-fractional deformations of the planning midpCT.

Test Set T2 consists of 50 3DCT images derived from intra-fractional de-

formations of the T2 midpCT.

Test Set Translated T2 consists of 50 3DCT images derived from intra-fractional deformations of the T2 midpCT, which has been translated. The translation used was the result of the rigid registration between the planning midpCT and the T2 midpCT, slightly noisy. The noise was a random sample from the normal distribution $\mathcal{N}(0,1)$.

Test Set Translated T3 consists of 50 3DCT images derived from intra-fractional deformations of the T3 midpCT, which has been translated. The translation used was the result of the rigid registration between the planning midpCT and the T3 midpCT, slightly noisy. The noise was a random sample from the normal distribution $\mathcal{N}(0,1)$.

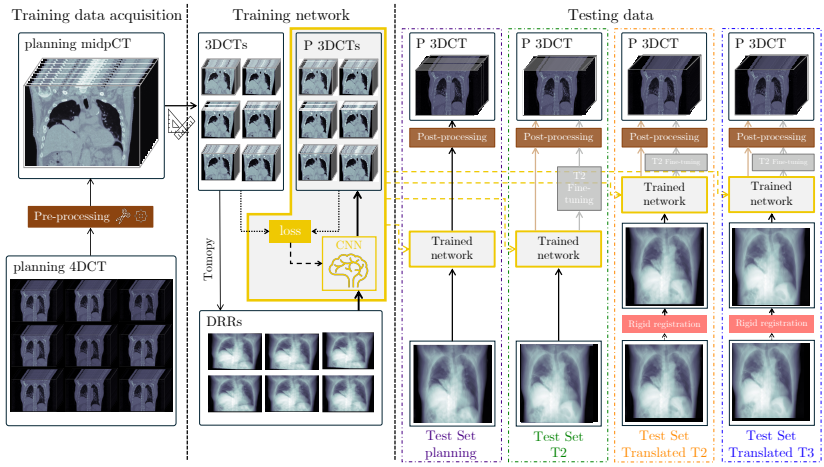


Fig. 5.7 Illustration of data partitioning and network training. The training 3DCTs are created by deforming the planning midpCT. From these 3DCT images, DRRs are generated and given as input to the convolutional neural network. The synthetic 3DCTs produced using the validation set are compared with the ground-truth 3DCTs to optimise the network hyper-parameters. Using inter-training, the DRR goes trough the trained network to obtain the corresponding predicted 3DCT (P 3DCT), whereas with T2-training a fine-tuning step on a T2 image is added. Four test sets are used: Test Set planning, Test Set T2, Test Set Translated T2 and Test Set Translated T3.

Network

The network used for the 3DCT reconstruction based on a single projection radiography is a convolutional neural network (CNN) that learns the correspondence between a 2D image and a 3D volume. This network was first proposed by Henzler et al. in [HRRR17], and we tuned the different hyper-parameters for the adaptive proton therapy application. The overall structure of the network is an encoder-decoder with skip connections. The objective of the encoder is to convert the information contained in the training data into an internal representation, which is then applied to a specific instance by the decoder. The aim of the skip connections is to share spatial details of a certain resolution at a certain level in the encoder part with the same resolution in the decoder part, converting fine details of the input 2D image into details of the output 3D volume. One special feature of this network is that it takes as input a 2D image of size 256×256 and produces as output a volume of size $128 \times 128 \times 128$, whereas encoder-decoder networks generally have input and output of the same spatial resolution.

The network implementation is based on TensorFlow, and the codes ran on a server with an NVIDIA RTX 6000 GPU. The Adam optimiser was used with an initial learning rate of 0.001 and momentum parameters $\beta_1 = 0.9$ and $\beta_2 = 0.99$. The loss function used to train the network was the mean squared error (MSE). The batch size and the number of epochs were set to 16 and 250, respectively. The training of the network was patient-specific, meaning that a new network was trained independently for each patient. The same training strategies and hyper-parameters were nevertheless used for all patients.

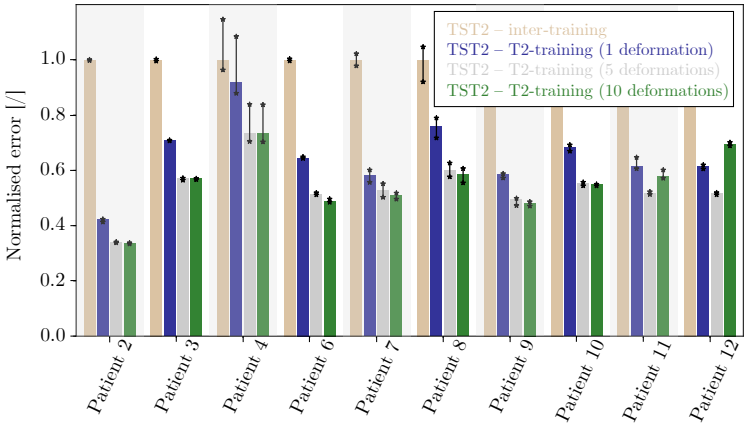
Training specifications

The neural network training was patient-specific, it used patient-specific CT data to refer to individual features. This work studied two main methods to train the model: *inter-fractional deformations training* (*inter-training*) and *fine-tuning with one T2 image training* (*T2-training*).

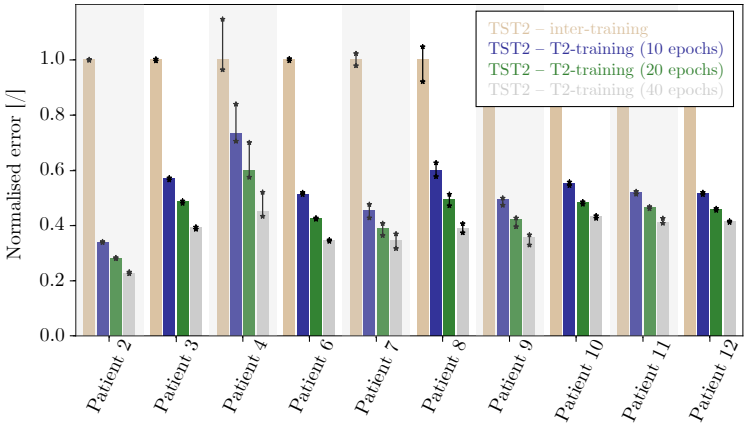
Inter-training method used as input of the network DRRs that were all generated from 3DCTs obtained using inter-fractional and intra-fractional anatomical deformations of the planning midpCT.

T2-training method was an extension of the inter-training method. It used the weights obtained at the end of inter-training as initial weights before

being fine-tuned to one image from the second acquisition time. Indeed, one pair of 3DCT/DRR generated from the T2 midpCT, called the T2-pair, was given as input to the network. The number of deformations, which only consider rotations and translations of the T2-pair, as well as the number of epochs have been optimised.



(a) Influence of the number of deformations



(b) Influence of the number of epochs

Fig. 5.8 Error in various T2-training configurations normalised by the error obtained using the inter-training method. (a) The three T2-training configurations depend on the number of deformations of the T2-pair given as input to the network, (b) the three T2-training configurations depend on the number of epochs used to train the network.

Figure 5.8(a) shows the error normalised by the median of the error obtained with inter-training for three configurations of T2-training: training with 1, 5 or 10 deformations of the T2-pair. T2-training is performed with 5 deformations as figure 5.8(a) shows that this number gives a smaller error. Figure 5.8(b) shows the error normalised by the median of the error obtained with inter-training for three configurations of T2-training: training with 10, 20 or 40 epochs. T2-training is performed with 40 epochs as the error decreases when the number of epochs increases. This number of epochs is also the maximum ensuring a training time of less than 90 s, enabling the method to be used in real clinical situations.

Performance evaluation

In order to evaluate the performance of the proposed method, the 50 pairs of DRR/3DCT of the four test sets were used. The outputs of the trained convolutional neural network, which are called the predicted 3DCT images, were compared with the ground-truth 3DCT images of the test sets through various analyses.

Visual analysis was carried out to qualitatively assess the performance of the proposed method. To that end, the predicted image was displayed next to the corresponding ground-truth 3DCT, for the four test sets studied in this part of the work. The image used in this analysis was the first image of each test set of Patient 10.

Various quantitative analyses have been carried out to evaluate the results of the proposed training strategies. The quantitative analyses were based on four metrics commonly used in image processing: normalised root mean squared error, mean absolute error, peak signal-to-noise ratio and structural similarity index measurement. Each of these metrics was computed for the 50 images composing the test sets concerned by the analysis, between the ground-truth image and the corresponding predicted image. The predicted image is given either by inter-training or by T2-training. Each quantitative analysis aims to answer a specific research question.

What is the impact on the predictions quality if inter-fractional deformations are added to the training set? This analysis looks at the variation in image quality metrics results when the test and training images are both generated from the planning midpCT, but inter-fraction motion is included in the training set. In this analysis, the results on TSTp obtained

with inter-training were compared to the results achieved with the basic-training method (explored in chapter 4).

What is the impact on the predictions quality if the x-rays projections used to test the network are acquired at a different time from the images used to train it? This analysis looks at the variation in image quality metrics results when the test images are acquired on a different treatment fraction. In this analysis, two test sets were studied: TSTp and TST2.

What is the impact on the predictions quality if setup errors are reduced between two treatment fractions? This analysis looks at the variation in image quality metrics results when a rigid registration is performed to minimise errors in the position of the patient. In this analysis, two test sets were studied: TST2 and TSTT2.

What is the impact on the predictions quality if the x-rays projections used to test the network are acquired throughout the treatment of the patient? This analysis looks at the variation in image quality metrics when the test images are acquired on a subsequent treatment fraction. In this analysis, two test sets were studied: TSTT2 and TSTT3.

5.2.3 Results

Visual analysis

The results of the visual analysis are represented in figure 5.9. For each test set, this figure shows a ground-truth 3DCT compared with the predicted 3DCT given by inter-training and the predicted 3DCT given by T2-training. In the case of TSTp, the predicted 3DCT is only obtained with the inter-training method. This figure shows that the 3DCT images reconstructed using the methodology proposed in this part of the thesis appears to be of lower quality than those reconstructed using the methodologies developed in the previous contributions. Indeed, the predicted 3DCT images are noisier and the edges between the different soft tissues are less perceptible. Besides, the vertebrae of the spine appear less sharp and are not precise. This figure also suggests that the 3DCTs generated using T2-training are of better quality for all three test sets. For example, the patient is correctly positioned in the predicted 3DCT generated with T2-training, whereas the position of the patient is slightly rotated in the predicted 3DCT generated with inter-training.

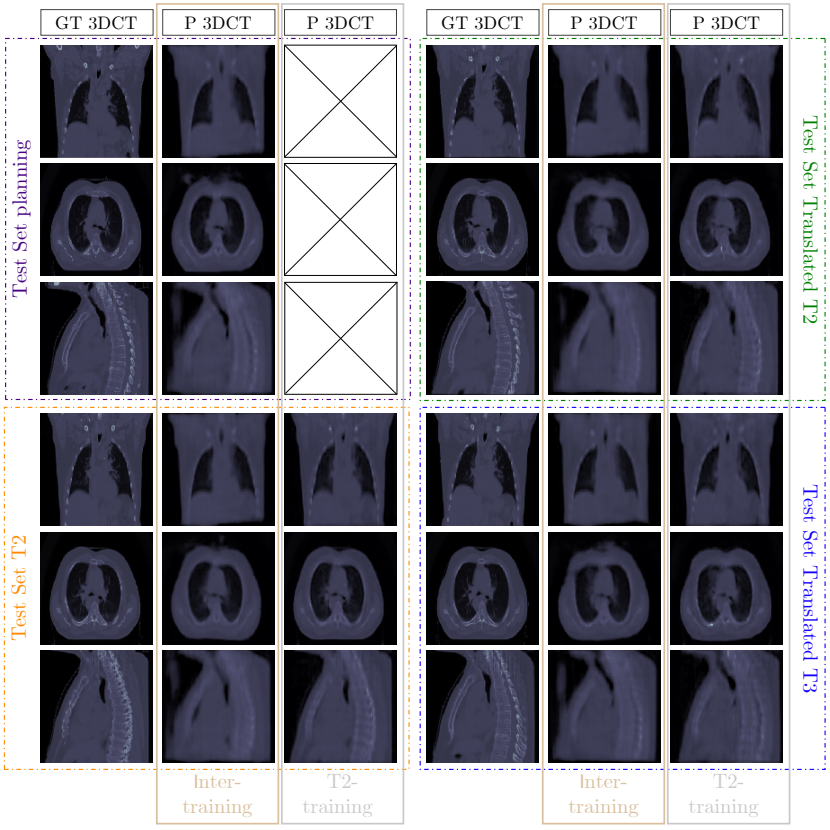


Fig. 5.9 Qualitative comparison of the 3DCT images reconstructed by the method proposed in this paper. For each test set, the GT 3DCT is compared to the predicted 3DCT using inter-training and the predicted 3DCT using T2-training. These are the results for Patient 10.

Impact of adding inter-fractional deformations in the training set

The results of the first research question are displayed in figure 5.10 and table 5.2. Figure 5.10 highlights the normalised root mean squared error computed between the ground-truth images and the predicted images using either basic-training and inter-training for the four patients common to both studies. Table 5.2 lists the results of the mean absolute error, peak signal-to-noise ratio and structural similarity index measurement metrics computed between corresponding images.

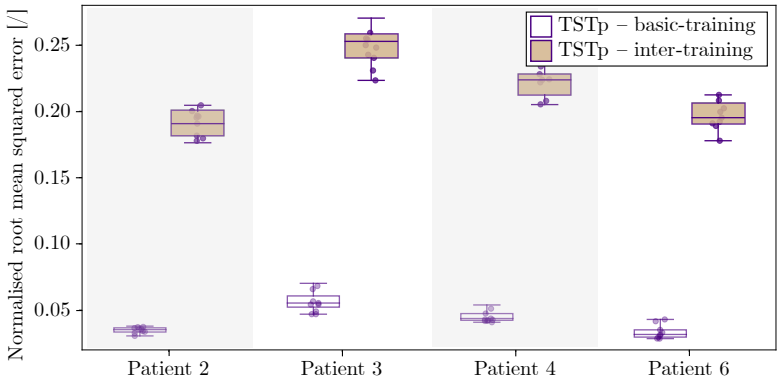


Fig. 5.10 Box plots of the NRMSE computed between the ground-truth 3DCTs and the predicted 3DCTs from the TSTp, for each patient. The predicted 3DCTs are generated using basic-training and inter-training.

Table 5.2 Results of the MAE, PSNR and SSIM metrics computed between the ground-truth 3DCTs and the predicted 3DCTs from the TSTp, for each patient. The predicted 3DCTs are generated using basic-training and inter-training. Each value in the table stands for the average ± standard deviation of the metric over the 50 images in the test set.

Patient ID	TSTp - basic-training			TSTp - inter-training		
	MAE [HU]	PSNR [dB]	SSIM [/]	MAE [HU]	PSNR [dB]	SSIM [/]
Patient 2	15.86 ± 0.71	43.78 ± 0.43	0.98 ± 0.00	69.69 ± 0.69	29.03 ± 0.23	0.81 ± 0.00
Patient 3	27.05 ± 2.46	36.50 ± 0.83	0.95 ± 0.01	89.20 ± 0.34	26.65 ± 0.08	0.73 ± 0.00
Patient 4	19.21 ± 1.16	41.81 ± 0.59	0.97 ± 0.00	82.16 ± 0.31	26.81 ± 0.03	0.77 ± 0.01
Patient 6	15.16 ± 1.27	43.58 ± 0.77	0.99 ± 0.00	70.00 ± 0.63	28.83 ± 0.17	0.79 ± 0.00
Patient 7	NA	NA	NA	80.22 ± 3.17	27.59 ± 0.20	0.76 ± 0.01
Patient 8	NA	NA	NA	80.61 ± 0.39	27.43 ± 0.09	0.76 ± 0.00
Patient 9	NA	NA	NA	63.29 ± 2.04	29.54 ± 0.25	0.82 ± 0.01
Patient 10	NA	NA	NA	85.42 ± 0.92	27.37 ± 0.12	0.75 ± 0.00
Patient 11	NA	NA	NA	73.31 ± 0.51	28.25 ± 0.06	0.79 ± 0.00
Patient 12	NA	NA	NA	65.23 ± 1.94	29.36 ± 0.41	0.82 ± 0.01
Mean	$\mathbf{19.32 \pm 1.4}$	$\mathbf{41.42 \pm 0.66}$	$\mathbf{0.97 \pm 0.00}$	$\mathbf{75.91 \pm 1.09}$	$\mathbf{28.09 \pm 0.16}$	$\mathbf{0.78 \pm 0.00}$

This analysis shows that for all patients, the error between the GT 3DCT and the predicted 3DCT is greater when the predicted 3DCT is generated with inter-training. Using basic-training, the mean of the NRMSE computed over the four patients is 0.047, whereas it is 0.223 using inter-training. This is also reflected in the other metrics. Using basic-training, the mean of MAE is 19.32 HU, of PSNR is 41.42 dB and of SSIM is 0.97, while

they are respectively 75.91 HU, 28.09 dB and 0.78 using inter-training.

Impact of an acquisition time different from that mainly used for training

The results of the second research question are displayed in figure 5.11 and table 5.3. Figure 5.11 highlights the normalised root mean squared error computed between the ground-truth images and the predicted images of both test sets TSTp and TST2. Table 5.3 lists the results of the mean absolute error, peak signal-to-noise ratio and structural similarity index measurement metrics computed between the same images.

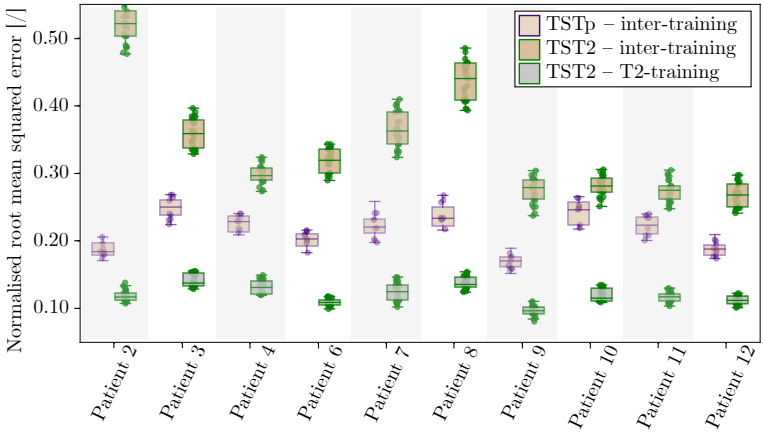


Fig. 5.11 Box plots of the NRMSE computed between the ground-truth 3DCTs and the predicted 3DCTs from TST2 and TSTp, for each patient. The predicted 3DCTs from TSTp are generated using inter-training, whereas the predicted 3DCTs from TST2 are generated using inter-training and T2-training.

The analysis differs with the training method chosen. Using the inter-training strategy, the error between the GT images and the predicted images increases when testing images come from TST2, whatever the patient. In this case, the mean of the MAE is 133.76 HU, of PSNR is 21.75 dB and of SSIM is 0.62. On the other hand, using the T2-training method, the mean of the MAE is 47.85 HU, of PSNR is 33.00 dB and of SSIM is 0.84. The quality of the predicted 3DCT images generated with the network trained using T2-training on the TST2 images is better than the quality of the predicted 3DCT images generated with the network trained using inter-training on the TSTp images, with a NRMSE reduced by more than 50%. However, this

quality is still poorer than that obtained with the basic-training method.

Table 5.3 Results of the MAE, PSNR and SSIM metrics computed between the ground-truth 3DCTs and the predicted 3DCTs from the TST2, for each patient. The predicted 3DCTs are generated using inter-training and T2-training. Each value in the table stands for the average ± standard deviation of the metric over the 50 images in the test set.

Patient ID	TST2 - inter-training			TST2 - T2-training		
	MAE [HU]	PSNR [dB]	SSIM [/]	MAE [HU]	PSNR [dB]	SSIM [/]
Patient 2	265.78 ± 1.37	19.87 ± 0.03	0.48 ± 0.00	46.86 ± 0.98	32.70 ± 0.20	0.83 ± 0.00
Patient 3	135.60 ± 1.10	23.55 ± 0.04	0.59 ± 0.00	54.08 ± 0.47	31.72 ± 0.11	0.82 ± 0.00
Patient 4	114.14 ± 1.77	24.53 ± 0.08	0.67 ± 0.01	56.60 ± 0.92	31.46 ± 0.17	0.77 ± 0.00
Patient 6	122.46 ± 0.96	24.59 ± 0.05	0.61 ± 0.00	47.02 ± 0.56	33.85 ± 0.15	0.85 ± 0.00
Patient 7	137.13 ± 4.34	23.61 ± 0.24	0.58 ± 0.01	47.64 ± 4.14	32.88 ± 0.86	0.84 ± 0.02
Patient 8	168.70 ± 5.29	22.29 ± 0.21	0.50 ± 0.00	57.63 ± 0.60	32.06 ± 0.12	0.79 ± 0.00
Patient 9	107.90 ± 2.05	25.20 ± 0.19	0.68 ± 0.00	38.13 ± 1.48	34.22 ± 0.49	0.90 ± 0.00
Patient 10	98.481 ± 1.26	25.93 ± 0.12	0.70 ± 0.00	45.42 ± 0.65	33.30 ± 0.14	0.88 ± 0.00
Patient 11	94.843 ± 0.85	26.29 ± 0.08	0.69 ± 0.00	42.22 ± 1.24	33.93 ± 0.25	0.87 ± 0.01
Patient 12	92.537 ± 0.76	26.21 ± 0.09	0.71 ± 0.00	42.87 ± 0.93	33.84 ± 0.12	0.88 ± 0.01
Mean	$\mathbf{133.76 \pm 1.98}$	$\mathbf{21.75 \pm 0.11}$	$\mathbf{0.62 \pm 0.00}$	$\mathbf{47.85 \pm 1.20}$	$\mathbf{33.00 \pm 0.26}$	$\mathbf{0.84 \pm 0.00}$

Impact of reduced setup errors

The results of the third research question are displayed in figure 5.12 and table 5.4. Figure 5.12 highlights the normalised root mean squared error computed between the ground-truth images and the predicted images of both test sets TST2 and TSTT2. Table 5.4 lists the results of the mean absolute error, peak signal-to-noise ratio and structural similarity index measurement metrics computed between the same images.

The analysis is not similar depending on the training method studied. Using the inter-training method, performing a rigid registration between the planning midpCT and the T2 midpCT to position the patient correctly positively affects the results and generally improves the quality of the reconstructed 3DCT images. Only Patient 10 and Patient 11 show an increase in error. On the other hand, all predictions on TSTT2 images using T2-training show an increase in error when the patient is returned to a position similar to the planning position. The mean over the 10 patients of MAE is 101.26 HU for inter-training and 88.09 HU for T2-training, the mean of PSNR is 25.92 dB for inter-training and 27.62 dB for T2-training, and the mean of SSIM is 0.67 for inter-training and 0.72 for T2-training. These results show that, despite an increase in error, the T2-training method still performs better than the inter-training method.

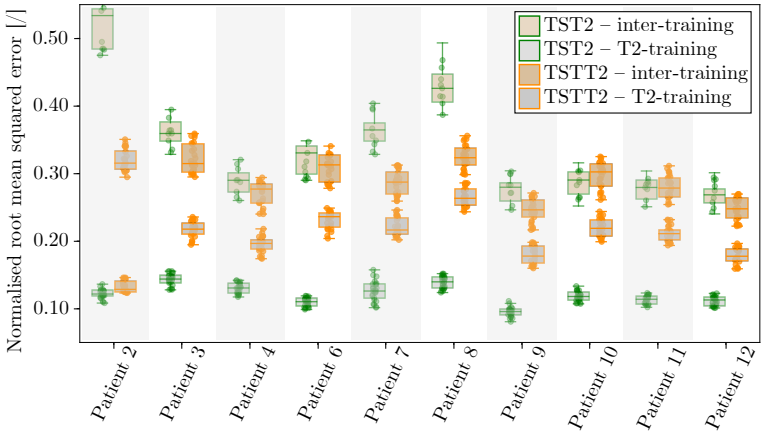


Fig. 5.12 Box plots of the NRMSE computed between the ground-truth 3DCTs and the predicted 3DCTs from TST2 and TSTT2, for each patient. The predicted 3DCTs are generated using inter-training and T2-training.

Table 5.4 Results of the MAE, PSNR and SSIM metrics computed between the ground-truth 3DCTs and the predicted 3DCTs from the TSTT2, for each patient. The predicted 3DCTs are generated using inter-training and T2-training. Each value in the table stands for the average ± standard deviation of the metric over the 50 images in the test set.

Patient ID	TSTT2 - inter-training			TSTT2 - T2-training		
	MAE [HU]	PSNR [dB]	SSIM [/]	MAE [HU]	PSNR [dB]	SSIM [/]
Patient 2	55.872 ± 0.60	30.69 ± 0.09	0.84 ± 0.00	180.7 ± 0.97	23.24 ± 0.08	0.63 ± 0.00
Patient 3	118.42 ± 0.60	24.38 ± 0.04	0.63 ± 0.00	79.84 ± 0.36	28.06 ± 0.03	0.72 ± 0.00
Patient 4	105.74 ± 2.04	25.34 ± 0.14	0.67 ± 0.01	80.05 ± 2.12	27.90 ± 0.22	0.72 ± 0.00
Patient 6	117.07 ± 1.59	24.82 ± 0.08	0.62 ± 0.00	84.55 ± 0.83	27.58 ± 0.07	0.70 ± 0.00
Patient 7	104.48 ± 1.87	25.63 ± 0.13	0.67 ± 0.01	78.06 ± 0.86	27.89 ± 0.11	0.74 ± 0.00
Patient 8	125.92 ± 0.97	24.57 ± 0.04	0.56 ± 0.00	102.6 ± 0.70	26.35 ± 0.11	0.62 ± 0.00
Patient 9	97.384 ± 0.84	26.18 ± 0.11	0.70 ± 0.00	66.50 ± 0.32	28.75 ± 0.08	0.79 ± 0.00
Patient 10	104.32 ± 1.03	25.62 ± 0.10	0.66 ± 0.00	77.39 ± 0.75	28.07 ± 0.10	0.73 ± 0.00
Patient 11	97.692 ± 0.93	26.00 ± 0.07	0.67 ± 0.00	70.81 ± 1.59	28.52 ± 0.15	0.77 ± 0.01
Patient 12	85.710 ± 0.65	26.00 ± 0.07	0.72 ± 0.00	60.43 ± 0.90	29.86 ± 0.15	0.82 ± 0.00
Mean	**101.26 ± 1.11**	**25.92 ± 0.09**	**0.67 ± 0.00**	**88.09 ± 0.94**	**27.62 ± 0.11**	**0.72 ± 0.00**

Impact of different acquisition times throughout the treatment

The results of the last research question are displayed in figure 5.13 and table 5.5. Figure 5.13 highlights the normalised root mean squared error computed between the ground-truth images and the predicted images of both test sets TSTT2 and TSTT3. Table 5.5 lists the results of the mean ab-

solute error, peak signal-to-noise ratio and structural similarity index measurement metrics computed between the same images.

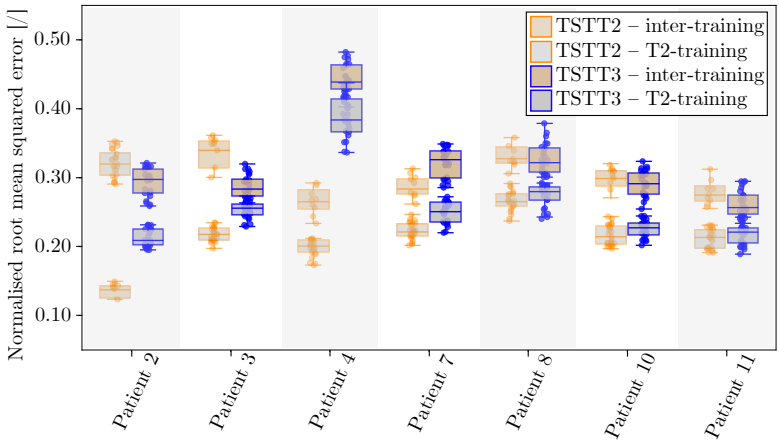


Fig. 5.13 Box plots of the NRMSE computed between the ground-truth 3DCTs and the predicted 3DCTs from TSTT2 and TSTT3, for each patient. The predicted 3DCTs are generated using inter-training and T2-training.

Table 5.5 Results of the MAE, PSNR and SSIM metrics computed between the ground-truth 3DCTs and the predicted 3DCTs from the TSTT3, for each patient. The predicted 3DCTs are generated using inter-training and T2-training. Each value in the table stands for the average ± standard deviation of the metric over the 50 images in the test set.

Patient ID	TSTT3 - inter-training			TSTT3 - T2-training		
	MAE [HU]	PSNR [dB]	SSIM [/]	MAE [HU]	PSNR [dB]	SSIM [/]
Patient 2	108.64 ± 1.12	25.07 ± 0.09	0.64 ± 0.00	81.70 ± 1.11	27.67 ± 0.14	0.70 ± 0.00
Patient 3	105.02 ± 0.88	25.26 ± 0.09	0.68 ± 0.00	91.80 ± 0.95	26.61 ± 0.10	0.69 ± 0.00
Patient 4	210.92 ± 5.42	20.71 ± 0.16	0.54 ± 0.01	164.9 ± 8.13	22.15 ± 0.36	0.57 ± 0.01
Patient 7	119.46 ± 0.72	24.56 ± 0.06	0.64 ± 0.00	89.69 ± 1.24	26.77 ± 0.14	0.71 ± 0.01
Patient 8	126.66 ± 2.77	24.32 ± 0.13	0.58 ± 0.01	102.6 ± 0.85	26.36 ± 0.05	0.61 ± 0.00
Patient 10	104.34 ± 1.96	25.56 ± 0.17	0.67 ± 0.00	81.30 ± 0.50	27.95 ± 0.06	0.72 ± 0.00
Patient 11	88.103 ± 1.87	26.69 ± 0.19	0.71 ± 0.01	70.24 ± 1.58	28.42 ± 0.19	0.78 ± 0.01
Mean	$\mathbf{123.31 \pm 2.11}$	$\mathbf{24.60 \pm 0.13}$	$\mathbf{0.64 \pm 0.00}$	$\mathbf{97.03 \pm 2.05}$	$\mathbf{25.56 \pm 0.10}$	$\mathbf{0.68 \pm 0.00}$

Regardless of the training method used to obtain the predicted images of the set TSTT3, the error obtained is within the same range of values as the error on the set TSTT2. The performance of the network is not degraded

when the images are acquired at a later stage of the treatment process. The mean over the 10 patients of MAE is 123.31 HU for inter-training and 97.03 HU for T2-training, the mean of PSNR is 24.60 dB for inter-training and 25.56 dB for T2-training, and the mean of SSIM is 0.64 for inter-training and 0.68 for T2-training.

5.2.4 Discussion

This work assesses the robustness of the method when evaluated on images of the patient acquired at different times during treatment. In order to take into account the inter-fractional anatomical deformations that can occur between two treatment sessions, the data augmentation algorithm has been made more complex, allowing to modify the size and the position of all organs. However, introducing such deformations in the training set degrades the quality of the predicted 3DCT images. The results show greater errors using the inter-training method compared with the results obtained with the basic-training method. Moreover, with the inter-training method, the error also increases when the image is acquired during another treatment session, and can reach a MAE of more than 250 HU. Future work would be to evaluate the impact of these differences on the simulation of a treatment plan delivery to validate whether the method could be used to control the dose delivered to the patient during the different treatment sessions.

The second training method studied in this work, T2-training, considerably improves the quality of the results, with MAE values generally below 100 HU. However, the main disadvantage of this training strategy is that it requires a short training on a new daily image. The number of epochs has been set to guarantee a training time of less than 90 seconds, thereby ensuring that the anatomy does not differ between daily image acquisition and treatment administration. However, acquiring a new 3DCT image increases the imaging dose delivered to the patient. To overcome this problem, unsupervised learning methods should be considered to avoid acquiring a new 3DCT image every treatment fraction. Besides, the optimal time between two 3DCT image acquisitions, and therefore between two fine-tuning of the network, should also be studied in depth. The results show that the T2-training method outperforms the inter-training method on the test set TSTT3. These results may be due to the fact that the delay between T2 and T3 images is around one week, whereas it is more than three weeks

between planning and T3 images.

One of the limitations of this work is that it is based solely on digitally reconstructed radiographs. This means that it is necessary to assess the reality gap between a x-rays projection actually acquired in the treatment room and the DRR image obtained with the TomoPy algorithm. A large difference between the two images could lead to a reduction in the quality of the results. With this in mind, future work could consider incorporating x-rays projections of the patient into the training set to ensure that the network learns to distinguish between the two types of image. It is also essential to notice that the field-of-view and the position of the patient within the image have an impact on the results. The results of Patient 2 and Patient 8 on the test set TSTT2, and the results of Patient 3 on the test set TSTT3 are poorer because of the significant differences in the position of the patient compared with the planning image.

Another limitation of this work is that it uses a fixed number of training images. Indeed, 355 inter-fractional anatomical deformations were created and used to generate sequences of 3DCT images representative of intra-fraction motion. These data present several limitations that must be considered in future work. Firstly, the realism of the 355 anatomies needs to be assessed, as the implementation of the shrinkage and baseline shift in the data augmentation tool does not reflect what actually happens over time. Secondly, different sizes of sequences were used in this work, but this choice is based on a personal idea and not on a quantitative result. Future work would involve assessing the ideal distribution between inter-fractional and intra-fractional deformations. Moreover, it is possible that the performance of the network improves if the number of training images increases. In this work, a compromise was made between the performance of the neural network, the training time needed, and the storage resources required.

Finally, the results in section 5.1 show that the patient-specific 3DCT reconstruction performs better when the neural network receives as input five digitally reconstructed radiographs for a given 3DCT. In this variant, the results are obtained by giving only one projection radiography as input to the network. Future work would involve combining the use of several angles and inter-fractional anatomical variations in the training set to try and improve the results on new treatment fractions.

5.3 Summary

This chapter studies the performance of the patient-specific 3DCT reconstruction methodology proposed in this thesis in the face of changes that may occur in the treatment room. In the first part of the chapter, these changes are caused by a new configuration of the treatment room. In particular, we study the robustness of the method to a change in the number of x-rays imaging systems available in the treatment room or to a change in the orientation of the projection angle used to obtain the digitally reconstructed radiograph, which can be caused by an incorrect alignment of the beam on the gantry. In the second part of the chapter, these changes are caused by a new treatment fraction. In particular, we study the robustness of the method to a change in the anatomy of the patient or to a change in the position of the patient. In each of these two parts, two training strategies are studied and their performances are compared according to several research questions.

6

Dosimetric evaluation of the synthetic 3DCTs

This chapter covers the assessment of the validity of the synthetic 3DCT images reconstructed by the methodology developed in this thesis for dose calculation. This chapter is inspired on the journal paper submitted to Medical Physics, currently under peer review. Additional analyses are provided in this thesis.

6.1 Context

The unique depth-dose characteristics of protons can improve tumor control while reducing toxicity. This represents a physical advantage of proton therapy over conventional radiotherapy in terms of dose compliance and sparing of normal tissue. However, this precision comes at the price of being highly vulnerable to uncertainties. Thoracic tumors are particularly concerned as the tumor motion induced by breathing can lead to density variation in the beam trajectory. This results in missing the target or shifting the expected proton range, thereby worsening the overall dose distribution. Further dose deterioration can occur due to the inference between the scanning motion of the beam and the anatomical motion, leading to hot and cold spots in the target. This phenomenon is known as interplay [SRTP09].

4D (3D + time) robust optimisation is the current state-of-the-art treatment planning approach for dealing with intra-fraction motion. It is designed to be robust against small changes in the anatomy by optimising for the worst-case scenario. In addition to the usual 3D robustness scenarios, such as range and setup errors, 4D robust optimisation is intended to withstand the multiple anatomical variations present in the planning 4D computed tomography (CT). This approach is effective and is the current best practice for the treatment of thoracic cancer [CZK+17]. However, the treatment plans obtained using 4D robust optimisation are designed to be robust only against the movements observed in the planning 4DCT and deliver a higher dose to the surrounding organs at risk (OARs). A few clinics have implemented off-line adaptive proton therapy protocols based on additional three-dimensional computed tomography (3DCT) scans acquired during treatment to better mitigate potentially adverse effects on dose distributions caused by inter-fractional changes [CLZ+14]. Although this approach can improve target coverage and OARs sparing, the adaptation process is slow and anatomical or physiological changes occurring on time scales of minutes or hours cannot be considered [PBL+17].

On-line adaptive proton therapy is seen as a promising method for minimising treatment uncertainties caused by inter-fractional changes. It consists in adapting and re-optimising the treatment plan on the basis of the daily anatomy observed in the treatment position. On-line adaptive proton therapy requires daily volumetric imaging data of the patient in the treatment position. However, this technique is not yet used clinically [AMN+20, PBSW21]. Several technological and methodological advances are required for the clinical implementation of on-line adaptive proton therapy. Over the last decade, numerous studies have investigated methods for rapid dose calculation and re-optimisation [PSP22], tools for daily segmentation of in-room images [KPL+22], concepts for on-line quality assurance [AMN+20, PBSW21] and also algorithms for synthetic CT generation [SMZS21].

The generation of a synthetic CT from an image of another modality using deep learning methods is the subject of numerous studies. In-room cone-beam CT (CBCT) is used on a daily or weekly basis during treatment for accurate patient positioning in radiation therapy, but its limited contrast resolution prevents it from performing advanced tasks. Some studies focused on improving CBCT image quality for better image-guided ra-

diation therapy [ZLD+23, CLS+20], and other studies assessed the validity of the synthetic CT for dose calculation [LCN+19, LLW+20]. The use of magnetic resonance imaging (MRI) in radiation therapy has proven its added value in delineating tumors and organs at risk thanks to its excellent soft-tissue contrast. Various methods have been proposed to convert MRI into CT-equivalent representations using deep learning [BNC+21, Han17, EDNDGH18]. However, these two imaging modalities are rarely used clinically as part of a proton therapy treatment. Only a few in-room CBCT scanners are already installed in proton therapy centers, while the clinical implementation of MR-integrated proton therapy is still in the research phase as many open questions need to be addressed beforehand, e.g. the mutual electromagnetic interactions between the MRI and proton therapy system that may degrade the quality of the MR image and the proton beam [HOM+20].

Currently, in-room image guidance in proton therapy is mainly based on 2D orthogonal x-rays imaging. This work has therefore focused on generating a synthetic 3DCT image from a single projection radiograph. In the field of radiation therapy, only a few studies have tackled this challenge. Several studies demonstrated that it was possible to reconstruct a 3DCT image from biplanar x-rays projections using a neural network [MZL+21, YGM+19], and another one used a neural network to reconstruct a 3D image from a single projection view [SZX19].

In this context, the aim of this part of the thesis was to study the validity of the 3DCT images reconstructed by our methodology for dose calculation. In particular, the impact on the estimate of the proton energy is evaluated, as well as the simulation of the delivery of one treatment plan. To this end, the different training methods and scenarios studied in the previous contributions were analysed.

6.2 Methodology

The methodology followed in this chapter is shown in figure 6.1. It highlights the two types of sequence generated, the training strategies considered, and the metrics used to assess the validity of the images for dose calculation. All this is explained in more detail in this section.

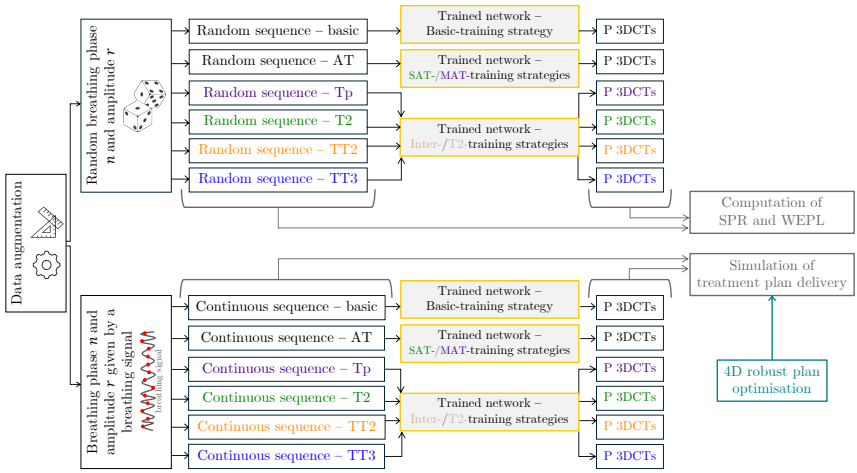


Fig. 6.1 Overview of the methodology developed in this chapter.

6.2.1 Sequences generation

The data used in this chapter come from 15 patients who were treated for lung or liver cancer by radiotherapy. For each patient, a planning 4DCT composed of 10 breathing phases was acquired before the start of treatment. For ten patients out of fifteen, a second 4DCT was acquired after several fractions of the treatment. For seven patients out of ten, a third 4DCT was acquired usually one week after the second. The mid-position (midp) CT image was calculated for all 4DCTs, creating the planning midpCT, the T2 midpCT and the T3 midpCT. A data augmentation tool, explained in detail in section 3.2.3, was used to generate two types of sequence for each patient, and each midpCT.

Random sequences

For each patient and each midpCT, a random sequence (RS) was generated. The random sequences are composed of 50 images created from random intra-fractional anatomical deformations of the midpCT image. To generate these 50 3DCT images, the normalised distances r were random samples of a normal distribution $\mathcal{N}(1, 0.5)$ truncated between 0.8 and 1.5, and the breathing phases n were random samples of a uniform distribution $\mathcal{U}(0, 1)$. The use of a normalised distance r and a breathing phase n in order to create a new 3DCT image is explained in equation 3.4.

Continuous sequences

For each patient and each midpCT, a continuous sequence (CS) was generated. The continuous sequences are composed of 120 images created from intra-fractional anatomical deformations of the midpCT image. To generate these 120 3DCT images, a 60 s synthetic respiratory signal was generated and sampled at a frequency of 2 Hz, resulting in the definition of 120 points. For each point, the two closest breathing phases of the 4DCT and the amplitude of the synthetic respiratory signal were computed and given to the intra-fractional data augmentation tool.

6.2.2 Trained neural networks

The neural network used throughout this thesis was first proposed by Henzler et al. in [HRRR17] with the aim to learn the correspondence between a 2D image and a 3D volume. This network was used in this work to obtain a 3DCT image from one or more x-rays projections. Different training strategies have been explored in the previous chapters and were analysed again in this chapter.

Basic-training strategy

The first training strategy, *basic-training*, was based on 500 3DCT images created from intra-fractional deformations of the planning midpCT. For each 3DCT, a digitally reconstructed radiograph was generated using a projection angle of 0°. In this training strategy, the patient-specific neural networks were trained for a total of 300 epochs. Basic-training was developed and explained in detail in chapter 4. For this training strategy, a random sequence, *RS-basic*, and a continuous sequence, *CS-basic*, were created from intra-fractional deformations of the planning midpCT.

SAT- and MAT-training strategies

The second and third training strategies, *SAT-training* and *MAT-training*, were based on 500 3DCT images created from intra-fractional deformations of the planning midpCT. For each 3DCT, five digitally reconstructed radiographs were generated using projection angles of 0°, 30°, 45°, 60° and 90°. In the SAT-training strategy, the network was trained independently for each angle value, whereas it was trained only once using all projection angles in the MAT-training strategy. The patient-specific neural networks were trained for a total of 500 epochs. These strategies were developed and explained in detail in section 5.1. For both of these training strategies, the same random sequence, *RS-AT*, and the same continuous sequence, *CS-AT*,

were created from intra-fractional deformations of the planning midpCT.

Inter- and T2-training strategies

The fourth and fifth strategies, *inter-training* and *T2-training*, were based on 17.125 3DCT images created from intra- and inter-fractional deformations of the planning midpCT. For each 3DCT, a digitally reconstructed radiograph was generated using a projection angle of 0°. In the inter-training strategy, the patient-specific neural networks were trained for a total of 250 epochs. In the T2-training strategy, a fine-tuning step on one intra-fractional anatomical deformation from the T2 midpCT is added. These strategies were developed and explained in detail in section 5.2. For these training strategies, four random sequences and four continuous sequences were created. *RS-Tp* and *CS-Tp* were created from intra-fractional deformations of the planning midpCT. *RS-T2* and *CS-T2* were created from intra-fractional deformations of the T2 midpCT. *RS-TT2* and *CS-TT2* were created from intra-fractional deformations of the T2 midpCT translated using the translation given by the rigid registration between planning midpCT and T2 midpCT. *RS-TT3* and *CS-TT3* were created from intra-fractional deformations of the T3 midpCT translated using the translation given by the rigid registration between planning midpCT and T3 midpCT.

6.2.3 Performance evaluation

In order to quantify the accuracy of the predicted 3DCT images in terms of dose calculation, the random and continuous sequences were used. The outputs of the convolutional neural networks trained using the different training strategies, the predicted 3DCT images, were compared with the ground-truth 3DCT images using several metrics.

Stopping power ratio (SPR) is used, in proton therapy treatment planning, for calculating the energy loss rate of protons. In this work, the SPR map was computed using the CT calibration implemented in OpenTPS [WDSJ+23]. The overall accuracy of SPR maps generated by our method was quantified by the mean absolute error (MAE), which measures the arithmetic average of the absolute errors. The MAE calculated between the ground-truth SPR map and the predicted SPR map was related to the MAE computed between the ground-truth 3DCT image and the predicted 3DCT image. This analysis was performed on the whole image and on the target volume only. In this analysis, the results were obtained using the random sequences.

Water equivalent path length (WEPL) is used, in proton therapy treatment planning, to calculate the proton energy required along the beam axis. The WEPL map is determined by the cumulative sum of the stopping power ratios in a particular orientation. Three orientations per patient were studied and correspond to the gantry angles used in the optimised treatment plan, which are listed in table 3.3. The beams were then divided into four groups according to the value of their gantry angle. For each group, the MAE was computed between the WEPL maps and related to the MAE calculated between the 3DCT images. This analysis was performed on the whole image and on a rectangular box around the GTV, determined by the size of the organ in the plane orthogonal to the direction of the beam and ranging from the first slice of the image to the furthest slice of the organ in the beam direction. In this analysis, the results were obtained using the random sequences.

Treatment plan simulation is used to verify that the treatment plan is safely delivered as prescribed. For each patient, a treatment plan was computed using 4D robust optimisation on the planning midpCT. These treatment plans were designed using RayStation v.12B [RSL, Bod18] according to the protocol used in Cliniques universitaires Saint-Luc for the treatment of lung and liver cancers. The main steps and the parameters used are described in section 3.4. The simulation of the treatment plan delivery was carried out with OpenTPS. The calculation of spot delivery times was performed with the IBA ScanAlgo simulation tool emulating delivery times on an IBA C230 cyclotron, while the simulation of dose deposition was performed with the Monte Carlo dose engine MCsquare [SGK+16]. The dose distribution on a 3DCT image was then accumulated on the midpCT using deformable image registration. The accumulated dose delivered on the ground-truth sequence was compared to the accumulated dose delivered on the predicted sequence. A visual analysis was performed to visualise where the major differences lie and a dose-volume histogram (DVH) was computed to give an indication of how the dose profiles differ between treatment plan delivery on the ground-truth sequence and on the predicted sequence. Besides, the difference between various dose quality metrics, such as D_{min}, D_{max} or D_{mean} was studied, as well as the percentage of volume of many regions of interest with an absolute difference between both doses smaller than or equal to 1 Gy and 5 Gy. In this analysis, the results were obtained using the continuous sequences.

6.3 Results

6.3.1 Stopping power ratio

Figure 6.2 shows the results of the stopping power ratio error as a function of the error between the two 3DCT images. Figure 6.2(a) shows the results calculated on the complete 3DCT image and figure 6.2(b) shows the results calculated on the tumor volume. These results show that there is a linear relationship between the error computed between the two 3DCT images and the error computed between both SPR maps. For the strategy basic-training, the coefficient of this linear regression is $0.818 \cdot 10^{-3} \pm 0.04 \cdot 10^{-3}$ when the calculation is performed on the whole volume of the image, and $0.851 \cdot 10^{-3} \pm 0.22 \cdot 10^{-3}$ when only the GTV is considered.

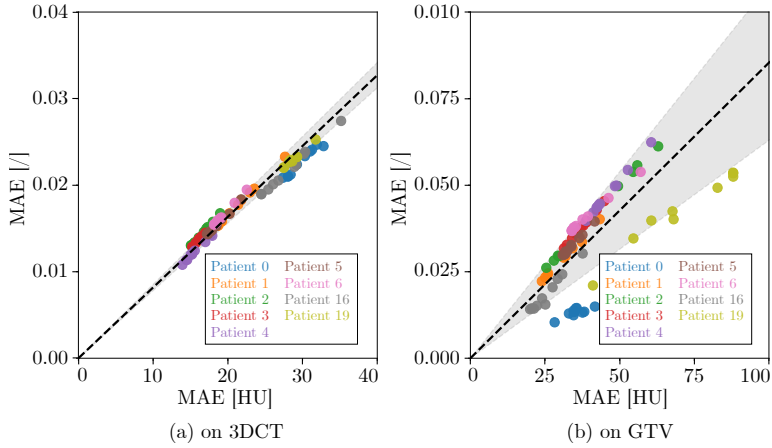


Fig. 6.2 Results of the stopping power ratio analysis for basic-training. The MAE of the SPR maps is compared with the MAE of the 3DCT images. It is computed over (a) the whole image volume and (b) the GTV volume. For ease of reading, only 10 images from the random sequence RS-basic of the nine patients are shown.

The same analysis is carried out for the different training methods. Table 6.1 shows the mean and standard deviation of the regression coefficient obtained for each training strategy. This table shows that for the different random sequences, with the exception of RS-AT, the coefficient obtained on the GTV is higher than the one computed on the whole 3DCT image. This means that for the same MAE computed between two 3DCT images, a larger SPR error is observed on the tumor level. Besides, the values of the

regression coefficients are larger when using inter-training or T2-training. This means that for the same MAE computed between two 3DCT images, the inter-training and T2-training methods always produce an image expressing a greater difference between the SPR. It can also be noted that the T2-training method used on the random sequence RS-T2 obtains results comparable to those of the other training strategies, but this is not verified on the volume of the tumor.

Table 6.1 Coefficients of the linear regression computed between the MAE of the SPR maps and the MAE of the 3DCT images for the different training strategies on the various random sequences.

Random sequence / training strategy	Coefficient of linear regression $\cdot 10^{-3}$	
	on 3DCT	on GTV
RS-basic / basic-training	0.818 ± 0.04	0.851 ± 0.22
RS-AT / SAT-training	0.755 ± 0.03	0.661 ± 0.09
RS-AT / MAT-training	0.772 ± 0.02	0.712 ± 0.11
RS-Tp / inter-training	0.919 ± 0.04	1.012 ± 0.02
RS-T2 / inter-training	0.941 ± 0.05	1.017 ± 0.09
RS-T2 / T2-training	0.872 ± 0.03	1.011 ± 0.02
RS-TT2 / inter-training	0.938 ± 0.05	0.981 ± 0.11
RS-TT2 / T2-training	0.934 ± 0.03	1.022 ± 0.02
RS-TT3 / inter-training	0.955 ± 0.02	1.019 ± 0.01
RS-TT3 / T2-training	0.936 ± 0.02	1.015 ± 0.01

6.3.2 Water equivalent path length

Figure 6.3 shows the results of the water equivalent path length error as a function of the error between the two 3DCT images. Figure 6.3(a) shows the results calculated on the complete 3DCT image and figure 6.3(b) shows the results calculated on the tumor volume. Besides, these two figures differentiate the results according to the orientation of the beam used to compute the WEPL, either the beam angle is between 0° and 90°, between 90° and 180°, between 180° and 270°, or between 270° and 360°. It can be seen that, for the basic-training method, the orientation of the beam used has almost no influence on the results of the WEPL, whereas the linear regression coefficient is generally greater when the entire volume of the image is considered. In addition, the results for the same patient are relatively similar when calculated on the whole 3DCT image, whereas they show greater variability when computed around the GTV.

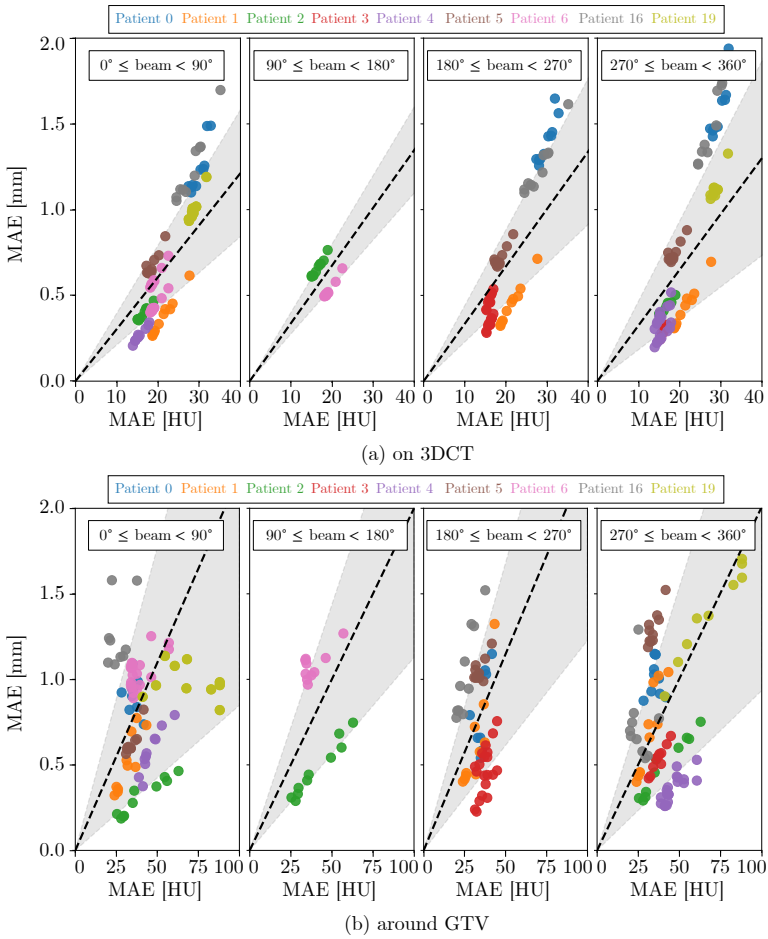


Fig. 6.3 Results of the water equivalent path length analysis for the strategy basic-training. For each patient, 3 beam angles are studied (those used in the optimised treatment plan) and the beams are then classified into 4 groups according to their gantry angle. The MAE of the WEPL maps is compared with the MAE of the 3DCT images. It is computed over (a) the whole image volume and (b) the GTV volume. For ease of reading, only 10 images from the random sequence RS-basic of the nine patients are shown.

The same analysis is carried out for the different training methods. Table 6.2 shows the mean and standard deviation of the regression coefficient obtained for each training strategy. For SAT-training and MAT-training, the coefficient values are within the same range of values regardless of the orientation of the beam, although it is observed that beams between 0° and 90° have a lower slope value, and beams between 270° and 360° have a higher slope value. The MAT-training method appears to reconstruct the images with greater confidence, since the standard deviation of this method is zero. The linear regression coefficients computed using inter-training and T2-training are higher than those obtained with the other training strategies. For these two methods, the error along a beam with a gantry angle between 0° and 90° is smaller, whereas it is higher along a beam with a gantry angle between 90° and 180°. Moreover, T2-training produces smaller coefficients than inter-training.

Table 6.2 Coefficients of the linear regression computed between the MAE of the WEPL maps and the MAE of the 3DCT images for the different training strategies on the various random sequences.

Random Sequence / training strategy	Coefficient of linear regression			
	$0° \leq$ beam $< 90°$		$90° \leq$ beam $< 180°$	
	on 3DCT	around GTV	on 3DCT	around GTV
RS-basic / basic-training	0.031 ± 0.01	0.022 ± 0.01	0.034 ± 0.01	0.020 ± 0.01
RS-AT / SAT-training	0.040 ± 0.01	0.019 ± 0.01	NA	NA
RS-AT / MAT-training	0.034 ± 0.00	0.015 ± 0.00	NA	NA
RS-Tp / inter-training	0.068 ± 0.02	0.024 ± 0.01	0.105 ± 0.03	0.040 ± 0.02
RS-T2 / inter-training	0.085 ± 0.02	0.094 ± 0.10	0.137 ± 0.02	0.126 ± 0.11
RS-T2 / T2-training	0.046 ± 0.01	0.029 ± 0.02	0.058 ± 0.00	0.029 ± 0.02
RS-TT2 / inter-training	0.083 ± 0.02	0.082 ± 0.07	0.123 ± 0.00	0.089 ± 0.06
RS-TT2 / T2-training	0.071 ± 0.02	0.049 ± 0.03	0.099 ± 0.05	0.062 ± 0.02
RS-TT3 / inter-training	0.075 ± 0.01	0.028 ± 0.02	0.115 ± 0.01	0.021 ± 0.00
RS-TT3 / T2-training	0.061 ± 0.00	0.024 ± 0.02	0.090 ± 0.00	0.021 ± 0.00

Random Sequence / training strategy	Coefficient of linear regression			
	$180° \leq$ beam $< 270°$		$270° \leq$ beam $< 360°$	
	on 3DCT	around GTV	on 3DCT	around GTV
RS-basic / basic-training	0.034 ± 0.01	0.023 ± 0.01	0.033 ± 0.01	0.020 ± 0.01
RS-AT / SAT-training	0.045 ± 0.00	0.030 ± 0.00	0.050 ± 0.01	0.023 ± 0.01
RS-AT / MAT-training	0.039 ± 0.00	0.018 ± 0.00	0.054 ± 0.02	0.012 ± 0.00
RS-Tp / inter-training	0.078 ± 0.02	0.031 ± 0.01	0.084 ± 0.02	0.022 ± 0.01
RS-T2 / inter-training	0.099 ± 0.02	0.031 ± 0.02	0.106 ± 0.03	0.036 ± 0.01
RS-T2 / T2-training	0.054 ± 0.01	0.019 ± 0.01	0.062 ± 0.02	0.021 ± 0.01
RS-TT2 / inter-training	0.097 ± 0.01	0.032 ± 0.01	0.097 ± 0.03	0.044 ± 0.02
RS-TT2 / T2-training	0.081 ± 0.01	0.027 ± 0.02	0.086 ± 0.02	0.044 ± 0.04
RS-TT3 / inter-training	0.098 ± 0.02	0.028 ± 0.01	0.101 ± 0.03	0.048 ± 0.04
RS-TT3 / T2-training	0.080 ± 0.01	0.032 ± 0.02	0.079 ± 0.02	0.035 ± 0.02

6.3.3 Treatment plan simulation

Basic-training strategy

Figure 6.4 shows the results of the accumulated dose after simulation of the treatment plan delivery on the CS-basic sequence composed of ground-truth images compared with the results of the simulation on the CS-basic sequence made of predicted images. For a human eye, the two doses appear to be relatively equivalent. However, the analysis of the difference shows that it is not zero. The largest differences are found around the tumor, while they are very small within the target volume. The zoom shows that the dose delivered on the predicted images is generally higher in front of the target relatively to the origin of the beams, and lower after it. This means that the Bragg peak on the predicted images occurs earlier in the body of the patient, so the method tends to overestimate the value of the voxels.

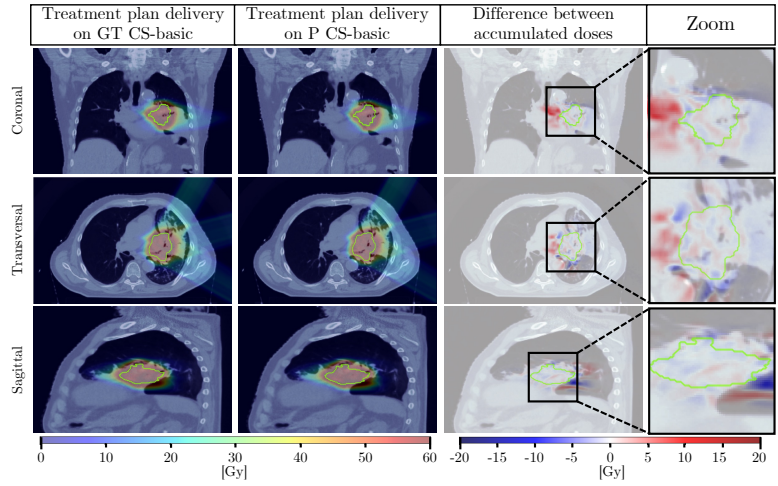


Fig. 6.4 Visualisation of three slices of the accumulated dose on the midpCT of Patient 6 after simulation of treatment plan delivery on CS-basic made of ground-truth images compared with the corresponding slices of the accumulated dose after simulation on CS-basic made of the predicted images, as well as the results of the difference between both doses and a zoom of the boxed area.

Figure 6.5 compares the DVH results obtained on the continuous sequence CS-basic composed of ground-truth or predicted images. For each patient and each organ, both the target and the organs at risk, the curve as-

sociated with the dose accumulated on the CS-basic sequence composed of predicted images is very close to the curve associated with the dose accumulated on the CS-basic sequence composed of ground-truth images. This observation is valid for patients suffering from lung cancer, with a prescribed dose of 60 Gy in 30 fractions, and also for patients suffering from liver cancer, with a prescribed dose of 52.5 Gy in 6 fractions.

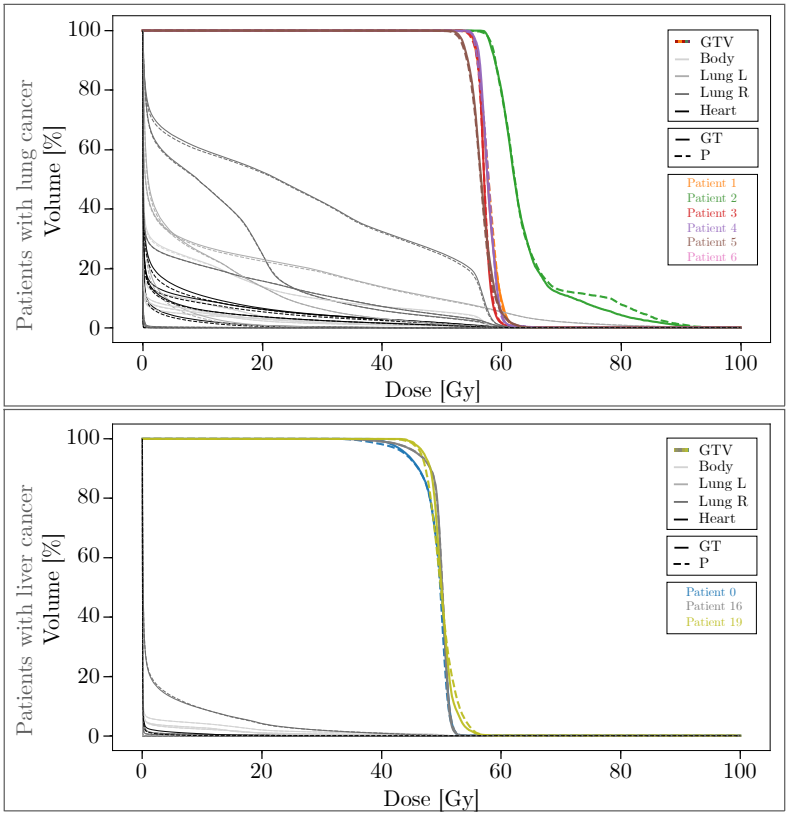


Fig. 6.5 DVH comparing accumulated dose on the ground-truth CS-basic (line) and on the predicted CS-basic (dashed line). The results for patients with lung cancer are shown at the top, those with liver cancer at the bottom. The results on the target are displayed in color, while results on organs at risk are displayed in shades of grey.

The visual analysis of the DVH can be supplemented by the results in figure 6.6. It shows the difference between the accumulated dose on the

CS-basic sequence composed of ground-truth images and the accumulated dose on the CS-basic sequence composed of predicted images, for eight metrics assessing the quality of the treatment plan. This figure shows that the dose difference is very close to 0 Gy for the different metrics: $D_{2\%}$, $D_{5\%}$, $D_{50\%}$, D_{mean}, $D_{95\%}$ and $D_{98\%}$. However, for the two metrics D_{min} and D_{max}, the difference is greater and can reach 5 Gy. It is normal to observe a greater variation in these two metrics as, unlike the others, they are assessed on a single pixel, making the results more sensitive.

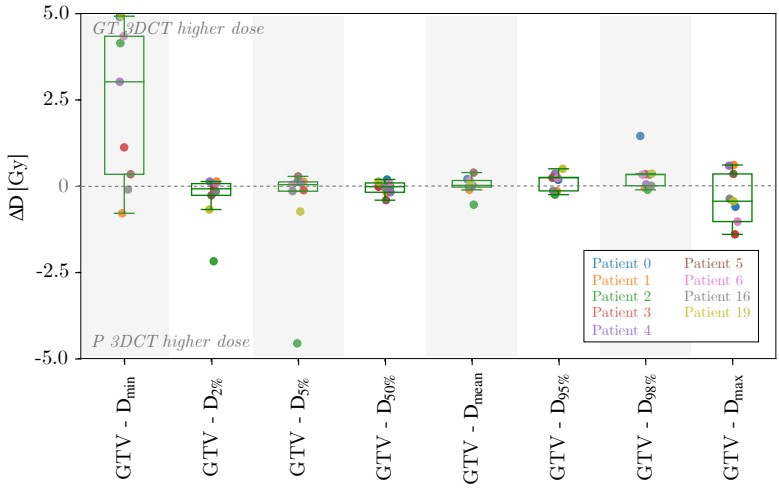


Fig. 6.6 Box plots of the difference between the dose accumulated on the midpCT after simulation of dose delivery on CS-basic made of GT 3DCTs and the dose accumulated on the same midpCT after simulation of dose delivery on CS-basic composed of P 3DCTs, for different quality metrics of a treatment plan evaluated on the target (GTV). *GTV-*$D_{x\%}$ stands for the dose delivered to at least x% of the GTV mask. Each dot represents one of the nine patients studied.

Figure 6.7 shows the percentage of volume of five different regions of interest (body, GTV, right and left lungs, and heart) with a dose difference of less than 1 Gy or less than 5 Gy. This figure shows that the GTV is the organ with the lowest percentage of volume with a difference of less than 1 Gy. For six patients over the nine studied with this training strategy, the difference between the two accumulated doses is smaller than 1 Gy for more than 75% of their target volume. However, for the other three patients, the difference between the two accumulated doses is below 1 Gy for

only 50 to 75% of the volume of their GTV. Patient 3, with a dose difference of less than 1 Gy for only 50% of his GTV volume, is also the patient with the greatest discrepancies in the DVH metrics. For the other organs studied, the difference between both doses is under 1 Gy for more than 90% of volumes, and under 5 Gy for almost all of their volumes. It should be noted that there is not necessarily a trend for these different organs, given that the different patients are relatively scattered in the metrics.

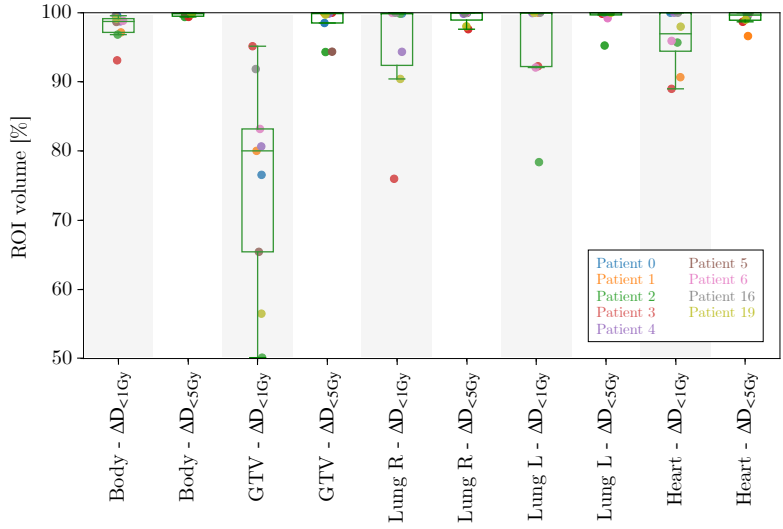


Fig. 6.7 Box plots of the percentage of the volume of a region of interest whose difference between the two accumulated doses is less than a certain threshold. The two doses studied are the doses accumulated on the midpCT after simulation of dose delivery on CS-basic made of GT 3DCTs and on CS-basic composed of P 3DCTs. *ROI*-ΔD_{xGy} stands for the difference between both accumulated doses below a certain threshold of x Gy computed over the mask of the ROI. Each dot represents one of the nine patients studied.

SAT- and MAT-training strategies

Figure 6.8 shows the results of the visual analysis after simulation of the treatment plan delivery on the CS-AT sequence. The results on CS-AT made of ground-truth images are compared with the results of the simulation on CS-AT made of predicted images. The predicted images are either obtained with SAT-training or with MAT-training. This analysis shows

that most of the significant dose differences are at the outer boundary of the target, regardless of the training method used. Besides, the dose differences are smaller after simulation on the CS-AT sequence composed of predicted images using MAT-training as there are less colored voxels. Using the MAT-training method, it can be seen that the large dose differences are mainly generated by the beam at 270°, whereas these are caused by all three orientations using the SAT-training method.

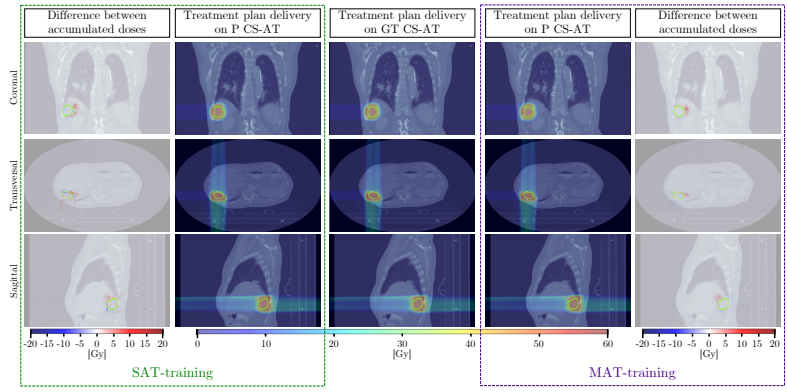


Fig. 6.8 Visualisation of three slices of the accumulated dose on the midpCT of Patient 0 after simulation of treatment plan delivery on CS-AT made of ground-truth images (middle column) compared with the corresponding slices of the accumulated dose after simulation on CS-AT made of predicted images using SAT-training (on the left) and MAT-training (on the right), as well as the results of the difference.

Figure 6.9 compares the DVH results obtained on the continuous sequence CS-AT composed of ground-truth or predicted images. The top figure shows the results when the predicted images are created using SAT-training and the bottom shows the results when the predicted images are obtained with MAT-training. This figure shows that the two training strategies produce DVH curves comparable to the DVH curves obtained using the ground-truth images, although the curves associated with the MAT-training strategy are even closer. This means that, for these two training strategies, the significant 3DCT reconstruction errors are located outside the beam trajectory or, at least, they do not affect dose delivery. In addition, the significant dose differences located outside the target, as shown in figure 6.8, do not affect the dose delivered to the organs at risk, as this

dose is also similar to that administered after treatment plan delivery on the ground-truth images.

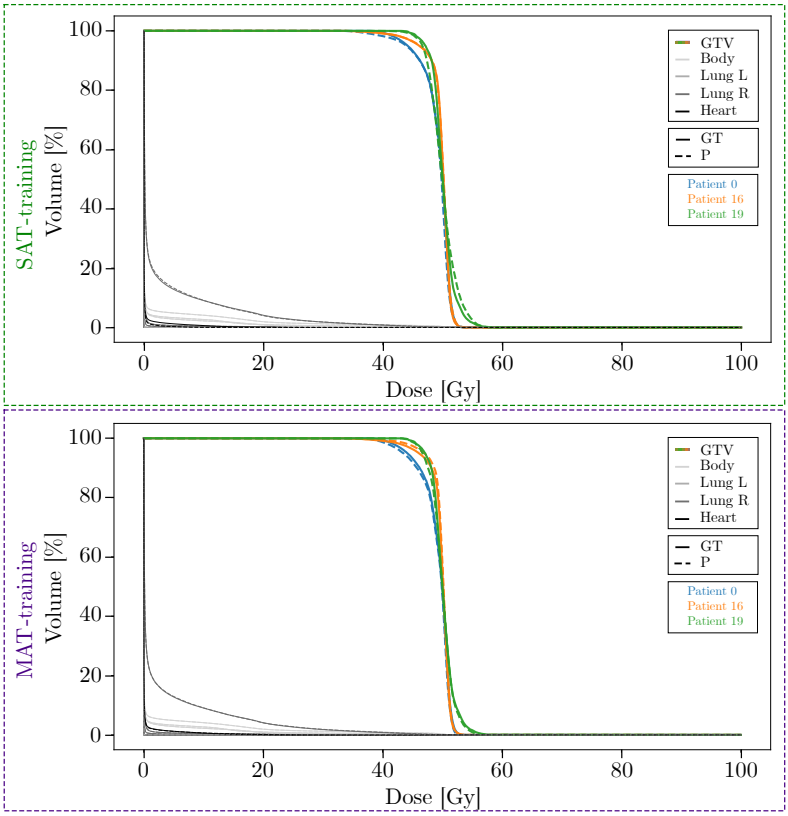


Fig. 6.9 DVH comparing accumulated dose on the ground-truth CS-AT (line) and on the predicted CS-AT (dashed line). The results for the predicted CS-AT using SAT-training are shown at the top, those for the predicted CS-AT using MAT-training at the bottom. The results on the target are displayed in color, while results on organs at risk are displayed in shades of grey.

The visual analysis of the DVH can be supplemented by the results in figure 6.10. It shows the difference between the accumulated dose on the CS-AT sequence composed of ground-truth images and the accumulated dose on predicted images using both training strategies, for eight metrics assessing the quality of the treatment plan. This figure shows that, for both training strategies, the difference is very close to 0 Gy for the differ-

ent metrics: $D_{2\%}$, $D_{5\%}$, $D_{50\%}$ and D_{mean}. However, for the four metrics D_{min}, $D_{95\%}$, $D_{98\%}$ and D_{max}, the difference is greater and can reach up to 5 Gy. This analysis also shows that, for the first time, MAT-training is not more reliable than SAT-training, given that the absolute difference in certain metrics are greater using MAT-training. However, this must be treated with caution as only three patients were studied in this part of the work. In addition, this analysis shows that, for Patient 16, the simulation on CS-AT made of predicted images tends to deliver more dose to the target than the simulation on CS-AT made of ground-truth images as the difference is always negative, whereas, for Patient 0, the opposite is observed as the simulation on CS-AT made of predicted images tends to deliver less dose to the GTV.

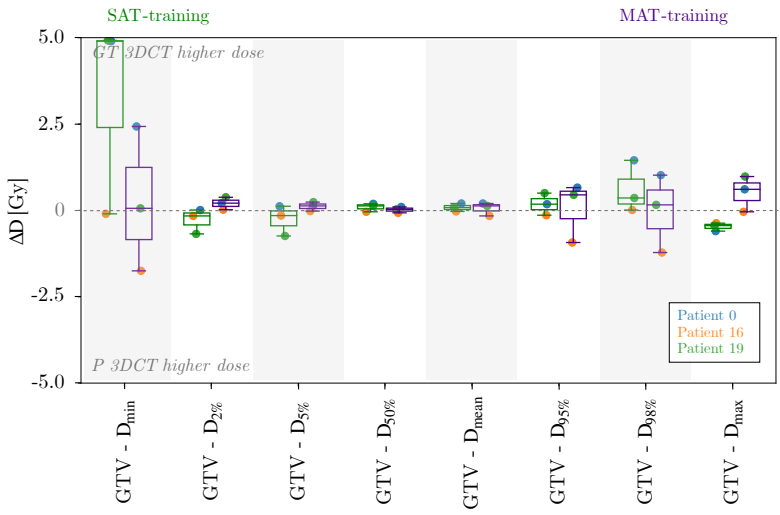


Fig. 6.10 Box plots of the difference between the dose accumulated on the midpCT after simulation of dose delivery on CS-AT made of GT 3DCTs and the dose accumulated on the same midpCT after simulation of dose delivery on CS-AT composed of P 3DCTs, for different quality metrics of a treatment plan evaluated on the target (GTV). *GTV-*$D_{x\%}$ stands for the dose delivered to at least x% of the GTV mask. Each dot represents one of the three patients studied. SAT-training results are shown in green on the left, MAT-training results in purple on the right.

Figure 6.11 shows the percentage of volume of five different regions of interest (body, GTV, right and left lungs, and heart) with a dose difference

smaller than 1 Gy and smaller than 5 Gy. The figure shows the results of the two training methods SAT-training and MAT-training. This figure shows that, whatever the training method, the GTV is the organ with the lowest percentage of volume with a difference of less than 1 Gy. However, this figure also shows that the MAT-training method considerably increases this percentage. In fact, the percentage of GTV volume with a difference below 1 Gy is between 60% and 90% using SAT-training, whereas this percentage is between 85% and 95% using MAT-training. For the other organs, the difference between the two doses is under 1 Gy for almost all their volumes, regardless of the training strategy used. The right lung is the organ at risk with the greatest variations, which is understandable since the beams pass through it, whereas the other organs are not located on the beam trajectories.

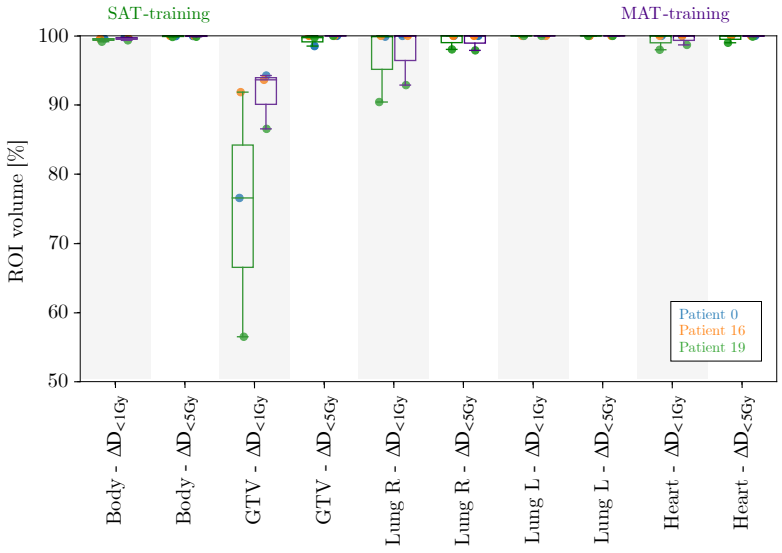


Fig. 6.11 Box plots of the percentage of the volume of a region of interest whose difference between the two accumulated doses is less than a certain threshold. The two doses studied are the doses accumulated on the midpCT after simulation of dose delivery on CS-AT made of GT 3DCTs and on CS-AT composed of P 3DCTs. $ROI\text{-}\Delta D_{xGy}$ stands for the difference between both accumulated doses below a certain threshold of x Gy computed over the mask of the ROI. Each dot represents one of the three patients studied. SAT-training results are shown in green on the left, MAT-training results in purple on the right.

Inter- and T2-training strategies

Figure 6.12 shows the results of the visual analysis after simulation of the treatment plan delivery on the four continuous sequences studied in these two training strategies.

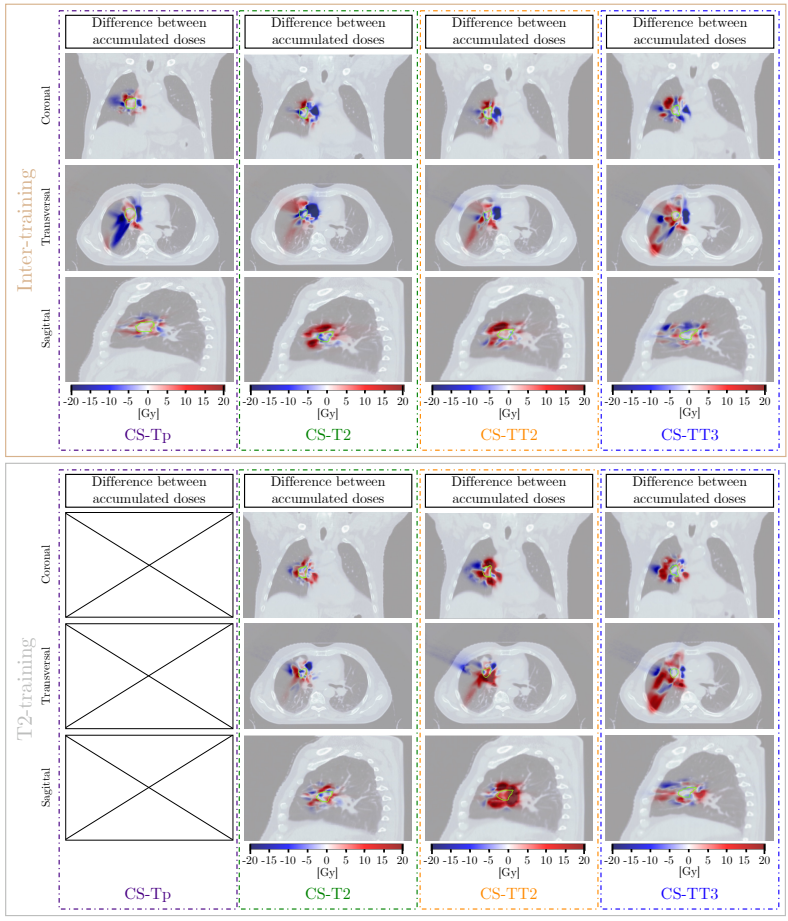


Fig. 6.12 Visualisation of three slices of the difference between the accumulated dose after simulation of treatment plan delivery on the continuous sequence made of ground-truth images and after simulation on the continuous sequence made of predicted images using either inter-training or T2-training. Four continuous sequences are considered: CS-Tp, CS-T2, CS-TT2 and CS-TT3.

The analysis differs according to the training method used. Using inter-

training, the differences are smaller for CS-Tp than for the other three continuous sequences. For this continuous sequence, the large differences are mainly located outside the target volume. For the other three sequences, large differences can also be found inside the target volume. The figure shows that correct patient positioning has a positive influence on the results as the errors are reduced on CS-TT2 compared with the errors on CS-T2. On the other hand, using T2-training, the dose is delivered more precisely on CS-T2 than on CS-TT2. In addition, the dose differences on CS-TT3 are significant but seem mainly located outside the GTV. The two training methods appear to give similar results for this continuous sequence.

In the remainder of this section, the four continuous sequences are analysed independently and successively.

Figure 6.13 compares the DVH results obtained on the continuous sequence CS-Tp composed of ground-truth or predicted images. For each patient, the curve associated with the simulation on the predicted images is similar to that associated with the simulation on the GT images, but there is generally a slight translation between the two indicating a small error in the simulation of the treatment plan delivery.

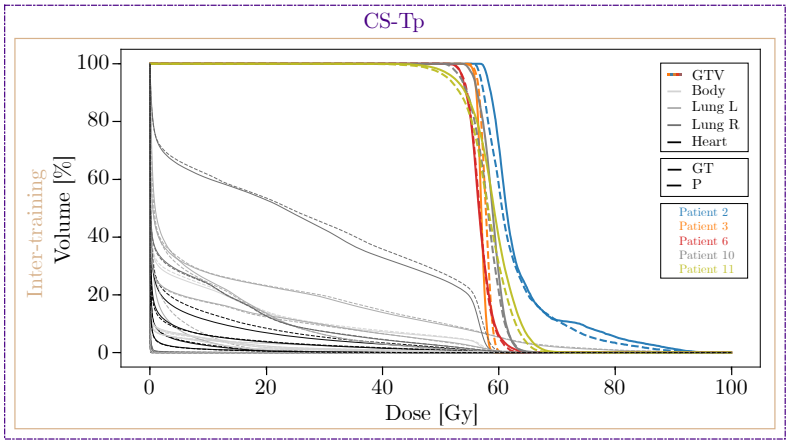


Fig. 6.13 DVH comparing accumulated dose on the ground-truth CS-Tp (line) and on the predicted CS-Tp (dashed line). The results for the predicted CS-Tp are obtained using inter-training. The results on the target are displayed in color, while results on organs at risk are displayed in shades of grey.

The visual analysis of the DVH can be supplemented by the results in figure 6.14. It shows the difference between the accumulated dose on the CS-Tp composed of ground-truth images and the accumulated dose on predicted images using inter-training, for eight metrics assessing the quality of the treatment plan. This figure shows that the accumulated dose difference is generally between −5 Gy and 5 Gy for the eight metrics. It can also be noted that the tumor does not tend to be over-dosed or under-dosed with the predicted images, as half the patients have positive differences and the other half negative differences.

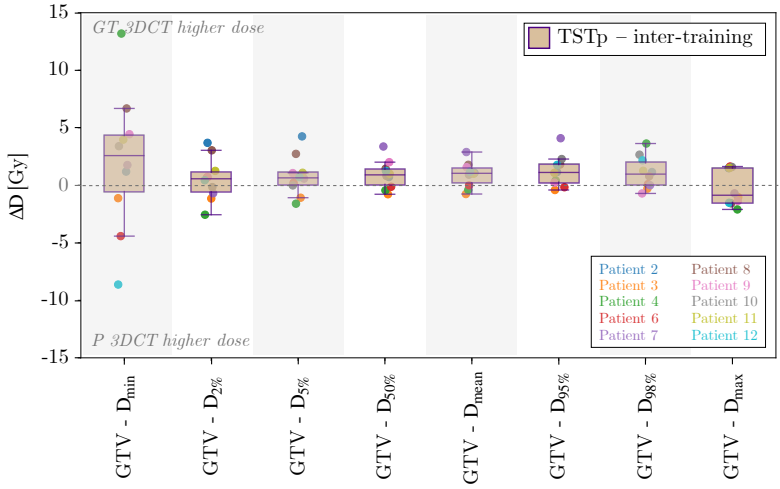


Fig. 6.14 Box plots of the difference between the dose accumulated on the midpCT after simulation of dose delivery on CS-Tp made of GT 3DCTs and the dose accumulated on the same midpCT after simulation of dose delivery on CS-Tp composed of P 3DCTs, for different quality metrics of a treatment plan evaluated on the target (GTV). *GTV-*$D_{x\%}$ stands for the dose delivered to at least x% of the GTV mask. Each dot represents one of the ten patients studied.

Figure 6.15 shows the percentage of volume of the five different regions of interest with a dose difference of less than 1 Gy and of less than 5 Gy. In the case of the simulation on CS-Tp, the GTV is the organ with the lowest percentage of volume with a difference smaller than 1 Gy, which is between 10% and 60%. The second organ with a small percentage of the volume having a difference below 1 Gy is the lung in which the tumor is located and is between 60% and 90%. The dose delivered in the other organs using

the sequence made of predicted 3DCT images is generally faithful to the dose delivered using the sequence composed of ground-truth 3DCTs as the percentage of ROI volume with a dose difference under 1 Gy is greater than 80%. Moreover, for all organs, a difference under 5 Gy is observed in more than 90% of the volumes.

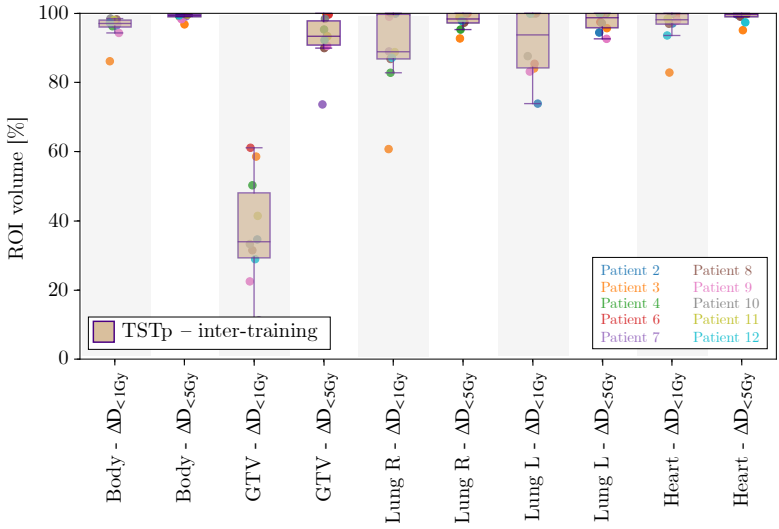


Fig. 6.15 Box plots of the percentage of the volume of a region of interest whose difference between the two accumulated doses is less than a certain threshold. The two doses studied are the doses accumulated on the associated midpCT after simulation of dose delivery on CS-Tp made of GT 3DCTs and on CS-Tp composed of P 3DCTs. *ROI*-ΔD_{xGy} stands for the difference between both accumulated doses below a certain threshold of x Gy computed over the mask of the ROI. Each dot represents one of the ten patients studied.

Figure 6.16 compares the DVH results obtained on the continuous sequence CS-T2 composed of ground-truth or predicted images. The top figure shows the results when the predicted images are created using inter-training and the bottom shows the results when the predicted images are obtained with T2-training. This figure shows that, for some patients, the anatomy of the T2 midpCT differs significantly from the planning midpCT, leading to a deterioration in the delivery of the treatment plan. For these patients, the two training strategies do not give the same results. The inter-training method shows considerable errors between the ground-truth

and predicted curves. In contrast, with T2-training, the curves remain very close to each other. For patients with greater similarity between both mid-pCTs, the two training strategies predict 3DCTs with sufficient quality so that the curve on the predicted images matches the curve on the ground-truth images.

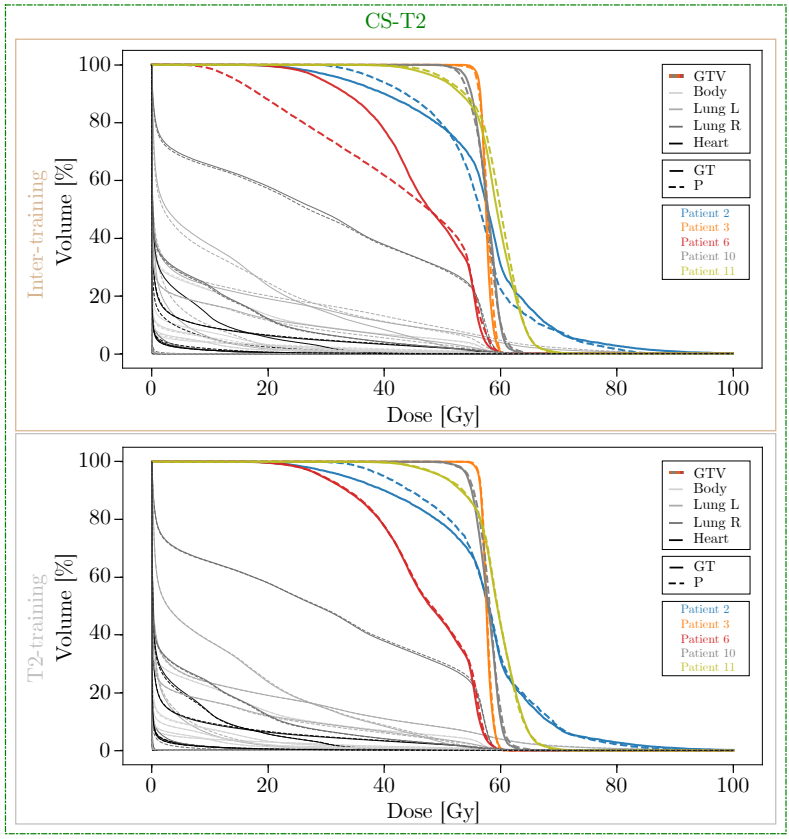


Fig. 6.16 DVH comparing accumulated dose on the ground-truth CS-T2 (line) and on the predicted CS-T2 (dashed line). The results for the predicted CS-T2 using inter-training are shown at the top, those for the predicted CS-T2 using T2-training at the bottom. The results on the target are displayed in color, while results on organs at risk are displayed in shades of grey.

The visual analysis of the DVH can be supplemented by the results in figure 6.17. It shows the difference between the accumulated dose on the

CS-T2 sequence composed of ground-truth images and the accumulated dose on predicted images using both training strategies, for eight metrics assessing the quality of the treatment plan. This figure shows that the difference between both doses assessed with the metrics $D_{2\%}$, $D_{5\%}$, $D_{50\%}$ and D_{mean} is less than 5 Gy, whatever the patient and the training method. For the other four metrics, the difference increases and can be greater than 10 Gy using inter-training. The T2-training method counters this increase and limits the absolute value of the difference to 5 Gy for these four metrics too.

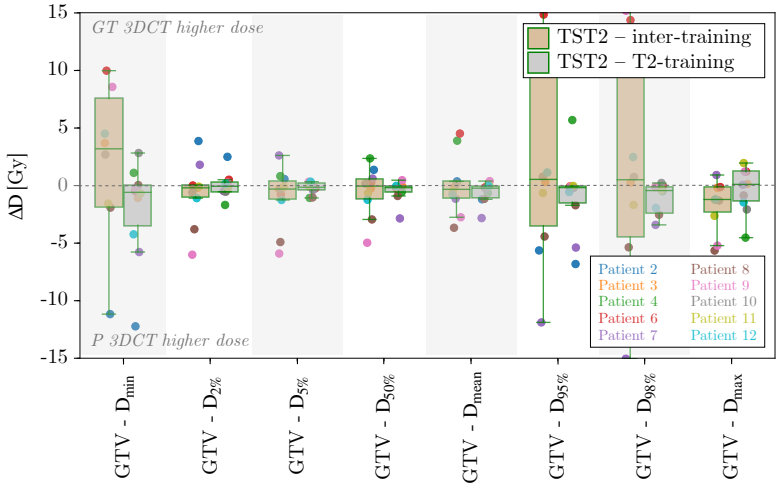


Fig. 6.17 Box plots of the difference between the dose accumulated on the midpCT after simulation of dose delivery on CS-T2 made of GT 3DCTs and the dose accumulated on the same midpCT after simulation of dose delivery on CS-T2 composed of P 3DCTs, for different quality metrics of a treatment plan evaluated on the target (GTV). *GTV-*$D_{x\%}$ stands for the dose delivered to at least x% of the GTV mask. Each dot represents one of the ten patients studied. Inter-training results are shown in brown on the left, T2-training results in grey on the right.

Figure 6.18 shows the percentage of volume of the five different regions of interest with a dose difference of less than 1 Gy and of less than 5 Gy. This figure shows the results of the two training methods inter-training and T2-training. In the case of the simulation on CS-T2, the GTV is still the organ with the lowest percentage of volume with a difference smaller than 1 Gy. However, the performance is not similar for the two training methods. The volume percentage is between 10% and 50% using inter-training, while

it is between 15% and 75% using T2-training. This percentage remains low and is between 30% and 99% for a threshold of 5 Gy using inter-training, whereas it is mainly over 80% using T2-training. For the other organs, the use of T2-training makes it possible to considerably increase the percentage of the volume with a small difference. Using this training strategy, more than 90% of the volume of the OARs shows a dose difference below 1 Gy, and more than 95% has a dose difference below 5 Gy.

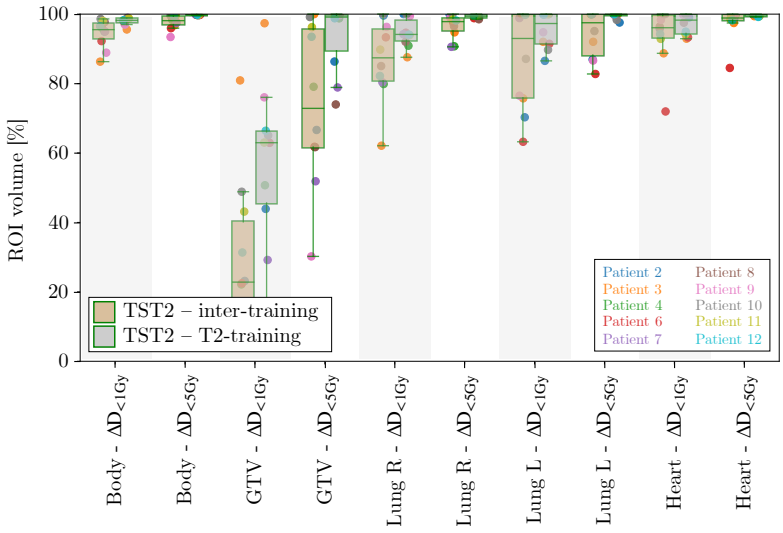


Fig. 6.18 Box plots of the percentage of the volume of a region of interest whose difference between the two accumulated doses is less than a certain threshold. The two doses studied are the doses accumulated on the associated midpCT after simulation of dose delivery on CS-T2 made of GT 3DCTs and on CS-T2 composed of P 3DCTs. *ROI-*ΔD_{xGy} stands for the difference between both accumulated doses below a certain threshold of x Gy computed over the mask of the ROI. Each dot represents one of the ten patients studied. Inter-training results are shown in brown on the left, T2-training results in grey on the right.

Figure 6.19 compares the DVH results obtained on the continuous sequence CS-TT2 composed of ground-truth or predicted images. The top figure shows the results when the predicted images are created using inter-training and the bottom shows the results when the predicted images are obtained with T2-training. This figure shows that using the inter-training method, the anatomical variations between the two treatment fractions

may be too great and therefore cannot be corrected by translation alone. At least, for these patients, the dose is not delivered on the predicted images as expected. For patients whose anatomy is more similar between the two fractions, translation can improve the results. On the other hand, using the T2-training method, the results are correct on the CS-TT2 sequence, even though the anatomy differs significantly from the planning anatomy, but are not as good as those obtained on the CS-T2 sequence.

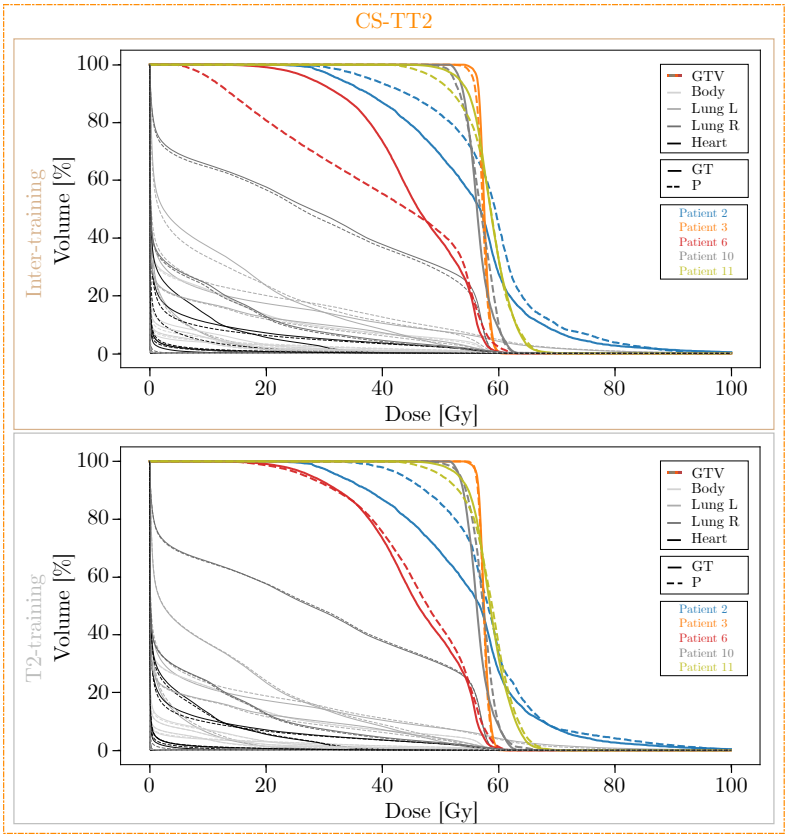


Fig. 6.19 DVH comparing accumulated dose on the ground-truth CS-TT2 (line) and on the predicted CS-TT2 (dashed line). The results for the predicted CS-TT2 using inter-training are shown at the top, those for the predicted CS-TT2 using T2-training at the bottom. The results on the target are displayed in color, while results on organs at risk are displayed in shades of grey.

The visual analysis of the DVH can be supplemented by the results in figure 6.20. It shows the difference between the accumulated dose on the CS-TT2 sequence composed of ground-truth images and the accumulated dose on predicted images using both training strategies, for eight metrics assessing the quality of the treatment plan. This figure shows similar results between both training methods for the four metrics $D_{2\%}$, $D_{5\%}$, $D_{50\%}$ and D_{mean}, with differences between -5 Gy and 5 Gy for the majority of patients. However, the increase observed on CS-T2 using inter-training is limited on this sequence CS-TT2, particularly for the both metrics $D_{95\%}$ and $D_{98\%}$. Furthermore, the T2-training method no longer offsets this increase on CS-TT2 and also shows an increase for the four metrics D_{min}, $D_{95\%}$, $D_{98\%}$ and D_{max}.

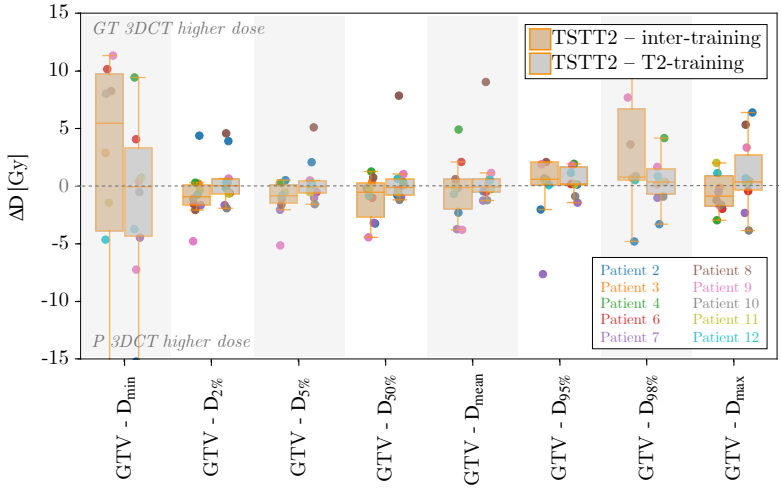


Fig. 6.20 Box plots of the difference between the dose accumulated on the midpCT after simulation of dose delivery on CS-TT2 made of GT 3DCTs and the dose accumulated on the same midpCT after simulation of dose delivery on CS-TT2 composed of P 3DCTs, for different quality metrics of a treatment plan evaluated on the target (GTV). *GTV-*$D_{x\%}$ stands for the dose delivered to at least x% of the GTV mask. Each dot represents one of the ten patients studied. Inter-training results are shown in brown on the left, T2-training results in grey on the right.

Figure 6.21 shows the percentage of volume of the five different regions of interest with a dose difference of less than 1 Gy and of less than 5 Gy. This figure shows the results of the two training methods inter-training

and T2-training. In the case of the simulation on CS-TT2, the GTV remains the organ with the lowest percentage of volume having a small difference, whereas the four other organs show large percentages. This figure shows that the T2-training method gives better results. However, it is important to note that the percentages obtained using inter-training are generally 1% to 5% higher than those obtained on CS-T2, while, using T2-training, the percentages are generally 1% to 10% lower than those obtained on CS-T2.

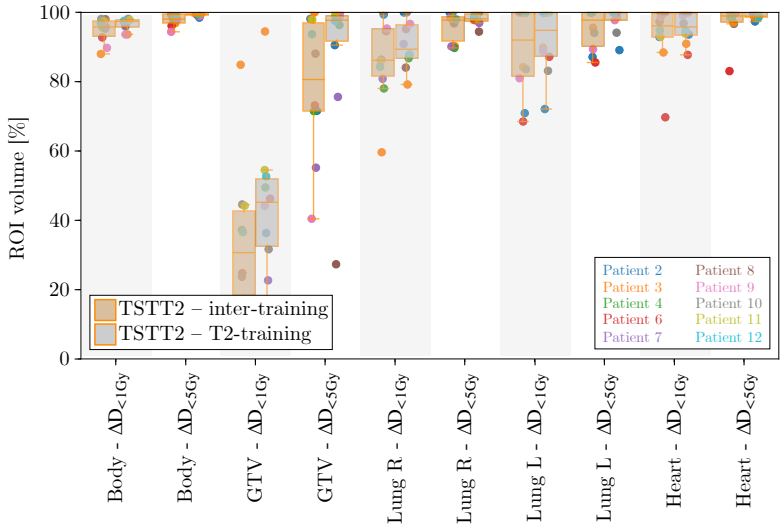


Fig. 6.21 Box plots of the percentage of the volume of a region of interest whose difference between the two accumulated doses is less than a certain threshold. The two doses studied are the doses accumulated on the associated midpCT after simulation of dose delivery on CS-TT2 made of GT 3DCTs and on CS-TT2 composed of P 3DCTs. *ROI*-ΔD_{xGy} stands for the difference between both accumulated doses below a certain threshold of x Gy computed over the mask of the ROI. Each dot represents one of the ten patients studied. Inter-training results are shown in brown on the left, T2-training results in grey on the right.

Figure 6.22 compares the DVH results obtained on the continuous sequence CS-TT3 composed of ground-truth or predicted images. The top figure shows the results when the predicted images are created using inter-training and the bottom shows the results when the predicted images are obtained with T2-training. This figure shows that the two training methods give relatively comparable results. For both training methods and the

four patients shown, the simulation of the treatment plan delivery on the predicted images gives similar results to the simulation of the treatment plan delivery on the ground-truth images.

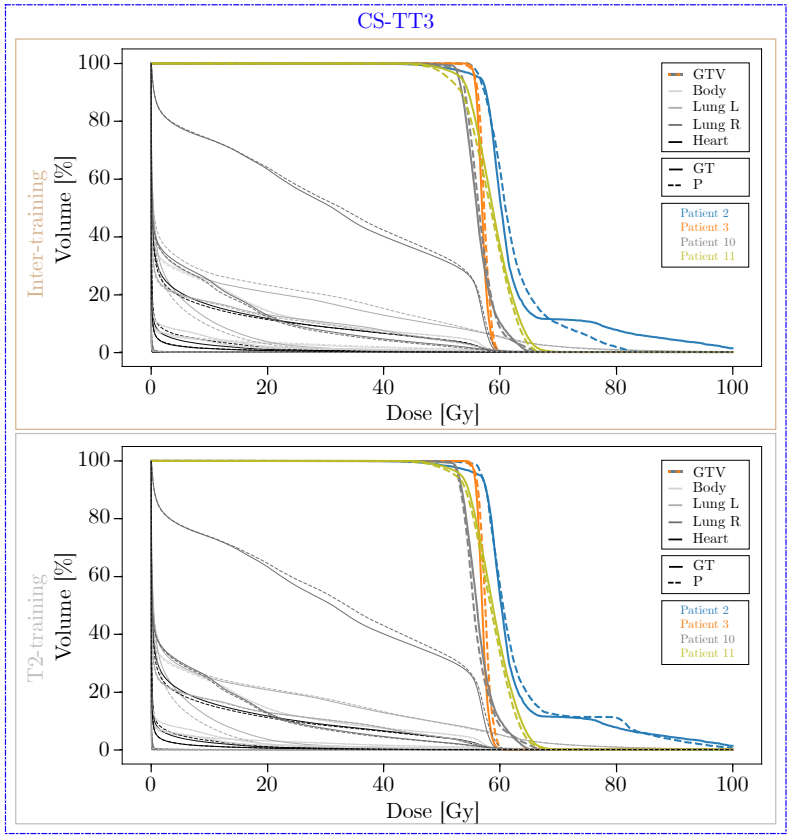


Fig. 6.22 DVH comparing accumulated dose on the ground-truth CS-TT3 (line) and on the predicted CS-TT3 (dashed line). The results for the predicted CS-TT3 using inter-training are shown at the top, those for the predicted CS-TT3 using T2-training at the bottom. The results on the target are displayed in color, while results on organs at risk are displayed in shades of grey.

The visual analysis of the DVH can be supplemented by the results in figure 6.23. It shows the difference between the accumulated dose on the CS-TT3 sequence composed of ground-truth images and the accumulated dose on predicted images using both training strategies, for eight metrics

assessing the quality of the treatment plan. This figure highlights that for the eight metrics assessing the quality of the treatment plan, the two training methods manage to limit the difference between the two doses to 5 Gy and both give similar results for these different metrics.

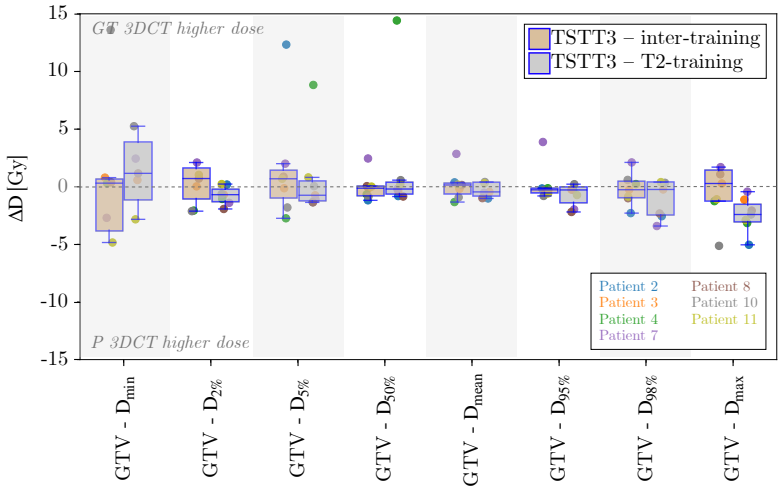


Fig. 6.23 Box plots of the difference between the dose accumulated on the midpCT after simulation of dose delivery on CS-TT3 made of GT 3DCTs and the dose accumulated on the same midpCT after simulation of dose delivery on CS-TT3 composed of P 3DCTs, for different quality metrics of a treatment plan evaluated on the target (GTV). *GTV-*$D_{x\%}$ stands for the dose delivered to at least x% of the GTV mask. Each dot represents one of the ten patients studied. Inter-training results are shown in brown on the left, T2-training results in grey on the right.

Figure 6.24 shows the percentage of volume of the five different regions of interest with a dose difference of less than 1 Gy and of less than 5 Gy. This figure shows the results of the two training methods inter-training and T2-training. In the case of the simulation on CS-TT3, the two training methods give comparable results. The GTV is the organ with the lowest percentage of volume showing a small dose difference. Between 10% and 80% of the volume of the GTV has a difference smaller than 1 Gy, and between 20% and 99% has a difference smaller than 5 Gy. According to these two metrics, the two methods give very different results depending on the patient. For the other organs, the two methods give almost identical results, with more than 60% of each volume having a dose difference below

1 Gy, and more than 80% having a dose difference below 5 Gy.

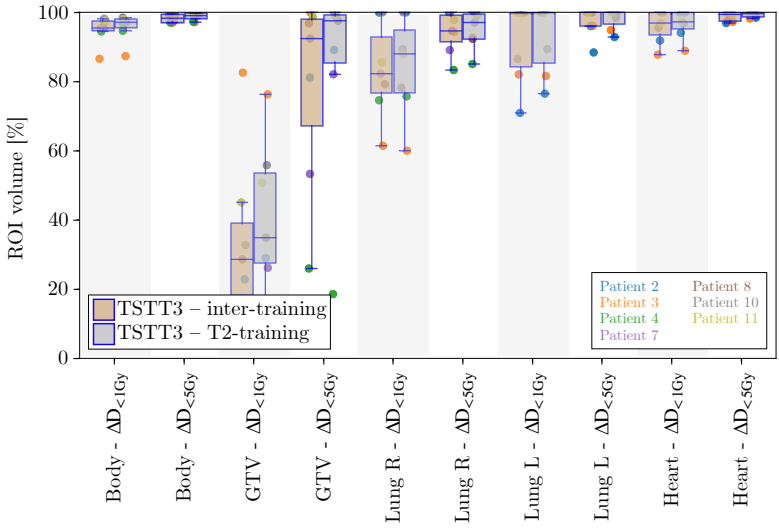


Fig. 6.24 Box plots of the percentage of the volume of a region of interest whose difference between the two accumulated doses is less than a certain threshold. The two doses studied are the doses accumulated on the associated midpCT after simulation of dose delivery on CS-TT3 made of GT 3DCTs and on CS-TT3 composed of P 3DCTs. $ROI\text{-}\Delta D_{xGy}$ stands for the difference between both accumulated doses below a certain threshold of x Gy computed over the mask of the ROI. Each dot represents one of the ten patients studied. Inter-training results are shown in brown on the left, T2-training results in grey on the right.

6.4 Discussion

This work assesses the validity of the 3DCT images reconstructed using our methodology for dose calculation. The different training strategies developed and explored in chapter 4 and chapter 5 were evaluated. The results of this chapter 6 are consistent with the results of the previous two chapters. The dose is delivered more accurately on the 3DCT images reconstructed with the basic-training, SAT-training and MAT-training methods than with inter-training or T2-training. This is the result of a better quality of the reconstruction and a smaller error between predicted and ground-truth 3DCT images.

The simulation of the treatment plan delivery on the continuous sequences associated with the basic-training, SAT-training and MAT-training methods shows very similar results for the sequences composed of predicted 3DCT images compared with the sequences composed of ground-truth 3DCT images. However, these three training methods have one major limitation. They are specific to a single moment of image acquisition and do not consider anatomical deformations between treatment fractions. This means that for each treatment fraction, these methods require the acquisition of a daily image and the training of the neural network on the basis of this new image. Unfortunately, these methods require more than eight hours of training, which makes them unsuitable for an on-line use. Instead, these different training strategies could be used to reconstruct the 3DCT image associated with the projection radiography of the previous treatment fraction in order to optimise a new treatment plan aimed at reducing inter-fraction motion occurring on a weekly or daily time scale, or to retrospectively accumulate the dose delivered at each treatment fraction.

The introduction of inter-fraction motion in the training set, as is the case with the inter-training and T2-training methods, degrades the results. Indeed, the simulation of the treatment plan delivery on the continuous sequence CS-Tp shows a notable variation between the dose delivered on the sequence composed of ground-truth 3DCT images and the sequence composed of 3DCT images predicted by inter-training. However, these two training strategies have the considerable advantage of not being specific to a single moment of image acquisition. It was then possible to simulate the treatment plan delivery on two other treatment fractions. The treatment plan delivery on both continuous sequences made of deformations of the T2 midpCT differs according to the training method and continuous sequence studied. Using the inter-training method, patient positioning and internal anatomical variations have a significant impact on the results. Correct patient positioning improves the quality of the reconstructed image and therefore the dose delivered to the patient. However, when the internal anatomical variations between the two fractions are too great, the method is less able to reconstruct the image, resulting in a deterioration in the dose delivered. Significant errors between the two doses are debatable here, as it is preferable to re-optimise the treatment plan in this type of situation. Despite these major errors, this training strategy allows to highlight the need for replanning. But, it is preferable not to use the 3DCT

image reconstructed with this method for replanning because it does not really reflect the daily anatomy. On the other hand, using the T2-training method, it is more appropriate to leave the patient in the position in which the T2 image was acquired in order to deliver the treatment plan. The results show that the fine-tuning step on the daily image makes the methodology robust to internal anatomical variations as the dose delivered using the predicted images is similar to the one delivered on the ground-truth images. This training strategy can then be used at the beginning of the treatment fraction to obtain the 3DCT image on which to optimise a new treatment plan, or to accumulate the dose delivered at each treatment fraction. Lastly, the treatment plan delivery on the continuous sequence made of deformations of the T3 midpCT is equivalent for both training strategies. Nevertheless, T2-training seems more reliable, which suggests that a gap of one week between two image acquisitions helps to reduce errors. A future work would be to study the optimal time interval between the acquisition of two images in order to provide a method that is robust to internal anatomical variations, while reducing the imaging dose.

Although this work gives a clear idea of the potential of the 3DCT images for dose calculation, it has certain limitations.

One limitation concerns the continuous sequences. The respiratory signals used to generate these sequences are regular signals, i.e. slightly noisy sine whose amplitude is close to that observed in the planning 4DCT. A future step of this work would be to perform the same treatment plan delivery simulations on continuous sequences generated from irregular signals in order to study the performance of the method in the face of sudden events, such as coughing or apnea, or to perform the simulations on continuous sequences generated from breathing signals of greater amplitudes and frequencies in order to study the performance of the method in the face of changes in breathing patterns that may occur during periods of stress. Another limitation of these continuous sequences is that they only represent one minute of breathing. Each sequence is looped until the entire treatment plan is delivered. Future work would therefore involve using a longer respiratory signal to ensure that all possible anatomical variations are considered.

Another major limitation concerns the treatment plans. The treatment plans are optimised with RayStation, but these are not treatment plans that

could be administered to patients in the clinic. In fact, due to time constraints, the treatment plans have not been completely optimised. This means that not all treatment plans meet all clinical objectives. One next step of this work would be to complete the optimisation of the various treatment plans in order to satisfy all clinical objectives and to simulate the delivery of these new plans again. The overall trend of the results should undoubtedly be the same, but this remains to be verified.

6.5 Summary

This chapter studies the performance of the patient-specific 3DCT reconstruction methodology proposed in this thesis for dose calculation and treatment plan delivery. The first part of the chapter studies the impact on the estimate of the proton energy required, while the second part of the chapter studies the impact on the treatment plan delivery if the simulation is based on images predicted by the neural network. Each part of the chapter studies the performance of the different training strategies described in the previous two chapters of this work.

7

Application scenarios and conclusions

Radiation therapy uses ionising radiations such as x-rays, gamma rays, electrons or protons to destroy cancer cells. Conventional radiotherapy has evolved significantly over the years. Adaptive radiotherapy is a modern method that continuously adapts treatment plans to account for anatomical changes, thereby improving the precision of radiation delivery, optimising therapeutic outcomes, and minimising damage to healthy tissues. Proton therapy offers a clear dosimetric advantage over conventional radiotherapy, which reduces the risk of side effects. However, proton therapy is highly vulnerable to uncertainties and lags behind photon therapy in several respects. For example, improving imaging and image-guided proton therapy could reduce some of these uncertainties. It is therefore essential to improve the quality of the treatment to fully exploit the physical advantages of protons and allow widespread adoption. This thesis focused on a major challenge for image-guided adaptive proton therapy, namely getting a 3DCT image from a projection radiography in real-time in order to obtain a three-dimensional visualisation of the anatomy of the patient and simulate the delivery of the treatment plan. Although this challenge applies to both photon therapy and proton therapy, this thesis focuses exclusively on proton therapy in order to counter its lag behind photon therapy.

The first contribution of this thesis was to develop a data augmentation tool to overcome the lack of medical data available to train and validate neural networks. The sensitive and confidential nature of medical data, and the problems involved in retrieving it from hospitals make it difficult to share. In addition, diverse principles have been defined to limit the imaging dose delivered to the patient, which prevents the acquisition of a large number of images. The data augmentation tool developed in this work can be used to simulate an infinite number of new 3DCT images. The tool can be used to modify the input 3DCT image with inter-fractional anatomical deformations such as tumor shrinkage, organ baseline shift, and image translation or rotation. Then, it is possible to create intra-fractional anatomical deformations by giving the tool a phase and an amplitude.

The second contribution of this thesis was to design a methodology for reconstructing a 3DCT image from a projection radiography. We decided to use a convolutional neural network and to train it independently for each patient using 3DCT images created with the data augmentation tool and digitally reconstructed radiographs generated from these images. We tested different training strategies and different scenarios, and evaluated the performance of our methods using image quality metrics. The results show that our methods achieve performance similar to that obtained with other current methods from the scientific literature that use other 3D imaging modalities, such as MRI or CBCT, but our methods outperform current methods that also use a projection radiography to reconstruct the 3DCT image. Table 7.1 summarises the results of some methods presented in the scientific literature, depending on the imaging modality used to generate the 3DCT image. This table should be treated with caution, as the results given are the results presented in the various contributions. This means that the data used and the network parameters have not been homogenised. It therefore does not represent a true comparison of the performance of the different methods, but it highlights general trends.

The third contribution of this work was to evaluate the quality of the 3DCT images reconstructed with our methodology in terms of dose accuracy. For example, we simulated the delivery of a treatment plan optimised with RayStation on continuous sequences made of ground-truth images and continuous sequences made of images predicted by the neural network. This analysis was carried out for the different training strategies

and different scenarios on which the methodology was tested. In particular, this enabled to simulate the delivery of the treatment plan on different treatment fractions. The difference between the accumulated doses was evaluated using DVH metrics and the results show that this difference is minimal.

Table 7.1 Quantitative analysis of the synthetic 3DCT image quality for our methods and other methods from the literature.

	Method	MAE [HU]	PSNR [dB]	SSIM [/]
MRI-to-CT	GAN [PAW+23]	80.36 ± 28.8	24.71 ± 2.97	0.800 ± 0.05
	cGAN [PAW+23]	68.28 ± 19.5	26.02 ± 2.78	0.852 ± 0.04
	IDDPM [PAW+23]	55.12 ± 9.41	28.71 ± 2.11	0.878 ± 0.04
CBCT-to-CT	pix2pix [GXW+21]	53.40 ± 9.34	26.80 ± 2.73	0.881 ± 0.07
	cycleGAN [GXW+21]	47.10 ± 6.45	28.30 ± 2.04	0.932 ± 0.04
	AGGAN [GXW+21]	43.50 ± 6.69	29.50 ± 2.36	0.937 ± 0.04
XRays-to-CT	NeRP [Caf23]	/	22.50 ± 3.20	0.290 ± 0.07
	X2CT-GAN [Caf23]	/	20.70 ± 2.40	0.570 ± 0.07
	X2Vision [Caf23]	/	23.20 ± 2.80	0.790 ± 0.09
	basic-training	21.21 ± 1.53	40.71 ± 0.64	0.968 ± 0.01
	SAT-training	22.42 ± 1.13	37.90 ± 0.43	0.963 ± 0.01
	MAT-training	13.51 ± 1.10	42.99 ± 0.80	0.984 ± 0.01
	inter-training on TSTp	75.91 ± 1.09	28.09 ± 0.16	0.780 ± 0.00
	inter-training on TST2	133.8 ± 1.98	21.75 ± 0.11	0.620 ± 0.00
	T2-training on TST2	47.85 ± 1.20	33.00 ± 0.26	0.840 ± 0.00
	inter-training on TSTT2	101.3 ± 1.11	25.92 ± 0.09	0.670 ± 0.00
	T2-training on TSTT2	88.09 ± 0.94	27.62 ± 0.11	0.720 ± 0.00
	inter-training on TSTT3	123.3 ± 2.11	24.60 ± 0.13	0.640 ± 0.00
	T2-training on TSTT3	97.03 ± 2.05	25.56 ± 0.10	0.680 ± 0.00

Although the results of our methods are promising, significant research is still needed to demonstrate the real capabilities of the whole methodology. This research should focus in particular on its main limitations.

A major limitation of this work comes from the data used. To deal with the lack of data, we use synthetic 3DCT images generated from deformations of the midpCT using a data augmentation tool. These data are there-

fore not acquired in the treatment room and may represent anatomical situations that do not actually occur. Digitally reconstructed radiographs are generated from these 3DCT images. The reality gap between a x-rays projection and the DRR generated with the TomoPy algorithm must be evaluated to guarantee the accuracy of the results obtained in this thesis. In this context, Chatzopoulos et al in [Cha22] demonstrate that it is wise to include a low-level of random noise in the training images in order to reduce the impact of fluoroscopic imaging noise on performance. Furthermore, it would be interesting to use the methodology developed in this thesis on DRRs generated with other algorithms. For example, Dhont et al. in [DVM+20] propose a novel DRR rendering framework that uses a combination of raytracing and deep learning based image-to-image translation to render highly realistic DRRs. Also, the midpCT used to create all these images comes from a 4DCT acquired with a scanner available at Cliniques universitaires Saint-Luc. This means that the network, trained and validated only on these images, works correctly on this type of data but is probably not robust for images acquired with other equipment.

Another limitation of this work comes from the number of patients used. This work uses a database composed of 15 patients, but they are never all studied in the different training strategies and different scenarios. Expanding the dataset and using a larger cohort of patients would make it possible to obtain even more reliable statistics. Furthermore, the patient-specific feature of the method, which is one of its strengths as it allows only a small sample of data from a single patient to be used, can also be seen as a limitation. In practice, this means that the neural network has to be trained several times, and therefore requires a lot of time and work. An interesting future work would be to evaluate the contribution of a multi-patients pre-training step to train the neural network for the global task of 3DCT reconstruction from a x-rays projection, and then to perfect the network only for the patient with a shorter training period requiring less data. In addition, it is possible that this pre-training step removes the need for patient-specificity if it is performed with a sufficient amount of data.

These two limitations together lead to a third limitation. Two types of tumors were studied in this work: liver tumors and lung tumors. This is a limitation of this work as the replicability of the method to other sites has not been investigated. It would be interesting to evaluate its performance

on other tumor positions and sizes.

Finally, the lack of explainability of neural networks compared with human expertise is a major obstacle to their adoption in practice. Radiation therapy and dose delivery are highly specialised subjects, and many patients and physicians may be reticent about using artificial intelligence in the treatment workflow. Not everyone is prepared to let an algorithm decide their own destiny. In recent years, various studies have been conducted to understand why and how a convolutional neural network makes a prediction. These methods are based on the Grad-CAM method that gives an activation heatmap highlighting which part of the input image the CNN focuses on when it makes the final prediction. For an in-depth explanation, the reader can refer to [SCD+17].

Despite these limitations, which are important to bear in mind, our method has the considerable advantage of being based solely on equipment already available in the treatment rooms and compatible with both proton therapy and conventional radiotherapy. This major advantage suggests a number of clinical applications for our method.

Off-line adaptive radiation therapy uses a 3DCT image acquired during the previous fraction to plan the next treatment session. The adaptive planning process is then carried out between two consecutive treatment fractions. In this case, the method developed in this thesis can be used for two purposes. The first one is to use the 3DCT image generated by our method as the new planning 3DCT to optimise the treatment plan. This has the advantage of eliminating the need to acquire a 3DCT image at every treatment fraction and using only the projection radiography acquired to position the patient, thus reducing the imaging dose delivered to the patient. The second one involves calculating and accumulating the dose delivered during each treatment fraction. This accumulated dose can then be considered for the future treatment planning.

Online daily adaptive radiation therapy requires a 3DCT image acquired at the start of the treatment fraction to optimise the new treatment plan. In this case, the method developed in this thesis can be used to provide the 3DCT image required for the optimisation of the new treatment plan, with the patient in the treatment position, and to retrospectively accumulate the dose delivered during this treatment fraction. In addition,

our method presents several advantages compared to in-room CT. The proposed method does not require a CT scanner, which saves a lot of space in the treatment room and saves the time of 3DCT image acquisition. Although it is foreseen that the extra imaging dose might become insignificant compared to the expected reduction in the integral dose achieved with online daily adaptive radiation therapy, our method reduces the extra imaging dose as it only requires the acquisition of a single x-rays projection.

Online real-time adaptive radiation therapy has been developed to account for intra-fraction motion during treatment delivery. In this case, the method developed in this thesis can be used to monitor internal motions of the patient. Using a real-time segmentation algorithm, it is possible to track the tumor in real-time on the 3DCT image and then, to adapt the delivery of the treatment plan by shifting the treatment source, shifting the beam or adjusting the patient position. A future vision for the use of our method in on-line real-time adaptive radiation therapy is to use the 3DCT image to verify the dose delivered during the treatment fraction, in order to adapt or stop the delivery of the treatment plan. This technique requires not only the acquisition of x-rays projections at optimal times during the treatment fraction to correctly track the motion, but also the simulation of the dose delivery in real-time using instantaneous dose calculation algorithms.

All these applications are promising and full of hope for improving cancer treatment, which unfortunately affects more and more people around us.

Bibliography

[ACM+97] J.R. Adler, S.D. Chang, M.J. Murphy, J. Doty, P. Geis, and S.L. Hancock. The cyberknife: a frameless robotic system for radiosurgery. *Stereotactic and Functional Neurosurgery*, 69, 1997. doi:10.1159/000099863.

[ADC+23] U. Anand, A. Dey, A. Chandel, R. Sanyal, A. Mishra, D. Pandey, V. De Falco, A. Upadhyay, R. Kandimalla, A. Chaudhary, J. Dhanjal, S. Dewanjee, J. Vallamkondu, and J. Pérez de la Lastra. Cancer chemotherapy and beyond: Current status, drug candidates, associated risks and progress in targeted therapeutics. *Genes and Diseases*, 2023. doi:10.1016/j.gendis.2022.02.007.

[AMCH99] J.R. Adler, M.J. Murphy, S.D. Chang, and S.L. Hancock. Image-guided robotic radiosurgery. *Neurosurgery*, 1999.

[AMN+20] F. Albertini, M. Matter, L. Nenoff, Y. Zhang, and A. Lomax. Online daily adaptive proton therapy. *British Journal of Radiology*, 93(1107), 2020. doi:10.1259/bjr.20190594.

[AS16] J. Abraham and J. Staffurth. Hormonal therapy for cancer. *Medicine*, 44:30–33, 2016. doi:10.1016/j.mpmed.2015.10.014.

[ASH+17] H. Arimura, Y. Shibayama, M. Haekal, Z. Jin, and K. Ikushima. Computer-assisted target volume determination. *Image-Based Computer-Assisted Radiation Therapy*, 2017. doi:10.1007/978-981-10-2945-5_5.

[atNIoH] National Cancer Institute at the National Institutes of Health. Types of cancer treatment. https://www.cancer.gov/about-cancer/treatment/types. Online; accessed November 2023.

[BAR+13] G. Bouilhol, M. Ayadi, S. Rit, S. Thengumpallil, J. Schaerer, J. Vandemeulebroucke, L. Claude, and D. Sarrut. Is abdominal compression useful in lung stereotactic body radiation therapy? a 4dct and dosimetric lobe-dependent study. *Physica Medica*, 29:333–40, 2013. doi:10.1016/j.ejmp.2012.04.006.

[BCvH19] S. Breedveld, D. Craft, R. van Haveren, and B. Heijmen. Multi-criteria optimization and decision-making in radiotherapy. *European Journal of Operational Research*, 277(1):1–19, 2019. doi:10.1016/j.ejor.2018.08.019.

[Ben14] D.J. Benjamin. The efficacy of surgical treatment of cancer - 20 years later. *Medical Hypotheses*, 2014. doi:10.1016/j.mehy.2014.01.004.

[BEP15] R. Bibb, D. Eggbeer, and A. Paterson. 2 - medical imaging. In *Medical Modelling (Second Edition)*, pages 7–34. Woodhead Publishing, 2015. doi:10.1016/B978-1-78242-300-3.00002-0.

[BHAea19] S. Baek, Y. He, B.G. Allen, and et al. Deep segmentation networks predict survival of non-small cell lung cancer. *Nature*, 2019. doi:10.1038/s41598-019-53461-2.

[BHKSC+16] J. Boda-Heggemann, A.C. Knopf, A. Simeonova-Chergou, H. Wertz, F. Stieler, A. Jahnke, L. Jahnke, J. Fleckenstein, L. Vogel, and Arns A. Deep inspiration breath hold—based radiation therapy: a clinical review. *International Journal of Radiation Oncology, Biology, Physics*, 94:478–92, 2016. doi:10.1016/j.ijrobp.2015.11.049.

[BLYY12] R. Baskar, KA Lee, R. Yeo, and KW Yeoh. Cancer and radiation therapy: Current advances and future directions. *International Journal of Medical Sciences*, 9:193–199, 2012. doi:10.7150/ijms.3635.

[BMJV+21] A Barragan-Montero, U. Javaid, G. Valdés, D. Nguyen, P. Desbordes, B. Macq, S. Willems, L. Vandewinckele, M. Holmström, F. Löfman, S. Michiels, K. Souris, E. Sterpin, and J. Lee. Artificial intelligence and machine learning for medical imaging: A technology review. *Physica Medica*, 83:242–256, 2021. doi:10.1016/j.ejmp.2021.04.016.

[BNC+21] M. Boulanger, J.C. Nunes, H. Chourak, A. Largent, S. Tahri, O. Acosta, R. De Crevoisier, C. Lafond, and A. Barateau. Deep learning methods to generate synthetic ct from mri in radiotherapy: A literature review. *Physica Medica*, 89:265–81, 2021. doi:10.1016/j.ejmp.2021.07.027.

[Bod18] D. Bodensteiner. Raystation: External beam treatment planning system. *American Association of Medical Dosimetrists*, 43:168–176, 2018. doi:10.1016/j.meddos.2018.02.013.

[Bro] J. Brownlee. Machine learning mastery. https://machinelearningmastery.com. Online; accessed January 2024.

[Bro19] Kristy K. Brock. Adaptive radiotherapy: Moving into the future. *Seminars in Radiation Oncology*, 29:181–184, 2019. doi:10.1016/j.semradonc.2019.02.011.

[Caf23] A. et al. Cafaro. X2vision: 3d ct reconstruction from biplanar x-rays with deep structure prior. *Medical Image Computing and Computer Assisted Intervention – MICCAI 2023*, 2023. doi:10.1007/978-3-031-43999-5_66.

[CCD20] C. Chen, C.-L. Chiang, and L. Dawson. Efficacy and safety of radiotherapy for primary liver cancer. *Chinese Clinical Oncology*, 10(1), 2020. doi:10.21037/cco-20-89.

[CGGS12] D. Ciresan, A. Giusti, L. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in Neural Information Processing Systems*, volume 25, 2012.

[Cha22] Edouard Chatzopoulos. Real-time tumor tracking for proton therapy using deep learning and x-ray fluoroscopy. *Mémoire UCLouvain*, 2022.

[CKC13] E.N. Chapman, A. Kaatz, and M. Carnes. Physicians and implicit bias: how doctors may unwittingly perpetuate health care disparities. *Journal of General Internal Medicine*, 28:1504–1510, 2013. doi:10.1007/s11606-013-2441-1.

[CLS+20] L. Chen, X. Liang, C. Shen, S. Jiang, and J. Wang. Synthetic ct generation from cbct images via deep learning. *Medica Physica*, 47:1115–25, 2020. doi:10.1002/mp.13978.

[CLZ+14] J.Y. Chang, H. Li, X.R. Zhu, Z. Liao, L. Zhao, A. Liu, Y. Li, N. Sahoo, F. Poenisch, D.R. Gomez, R. Wu, M. Gillin, and X. Zhang. Clinical implementation of intensity modulated proton therapy for thoracic malignancies. *International Journal of Radiation Oncology, Biology, Physics*, 90(4):809–818, 2014. doi:10.1016/j.ijrobp.2014.07.045.

[Cyb89] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Mathematics of control, signals and systems*, 2:303–14, 1989.

[CZK+17] J. Chang, X. Zhang, A. Knopf, H. Li, S. Mori, L. Dong, H-M. Lu, W. Liu, S. Badiyan, and S. Both. Consensus guidelines for implementing pencil-beam scanning proton therapy for thoracic malignancies on behalf of the ptcog thoracic and lymphoma subcommittee. *International Journal of Radiation Oncology, Biology, Physics*, 99:41–50, 2017. doi:10.1016/j.ijrobp.2017.05.014.

[DBMCV+22] C. Draguet, A. Barragán-Montero, M. Chocan Vera, M. Thomas, P. Populaire, G. Defraene, K. Haustermans, J.A. Lee, and E. Sterpin. Automated clinical decision support system with deep learning dose prediction and ntcp models to evaluate treatment complications in patients with esophageal cancer. *Radiotherapy and Oncology*, 176:101–107, 2022. doi:10.1016/j.radonc.2022.08.031.

[DGSP13] S. Dowdell, C. Grassberger, G.C. Sharp, and H. Paganetti. Interplay effects in proton scanning for lung: a 4d monte carlo study assessing the impact of tumor and beam delivery parameters. *Physics in Medicine and Biology*, 2013. doi:10.1088/0031-9155/58/12/4137.

[DHC+20] J. Dhont, S.V. Harden, L.Y.S. Chee, K. Aitken, G.G. Hanna, and J. Bertholet. Image-guided radiotherapy to manage respiratory motion: Lung and liver. *Clinical Oncology*, 32(12):792–804, 2020. doi:10.1016/j.clon.2020.09.008.

[DHS11] J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011.

[DLW+19] X. Dong, Y. Lei, T. Wang, M. Thomas, L. Tang, W.J. Curran, T. Liu, and X. Yang. Automatic multiorgan segmentation in thorax ct images using u-net-gan. *Medical Physics*, 2019. doi:10.1002/mp.13458.

[DSASM22] D. Dasnoy-Sumell, A. Aspeel, K. Souris, and B. Macq. Locally tuned deformation fields combination for 2d cine-mri-based driving of 3d motion models. *Physica Medica*, 94:8–16, 2022. doi:10.1016/j.ejmp.2021.12.010.

[Dur24] M. Durante. Current status and future trends in particle therapy – lessons from an interdisciplinary workshop. *Health Technology*, 2024. doi:10.1007/s12553-024-00855-6.

[DVB+18] J. Dhont, J. Vandemeulebroucke, M. Burghelea, K. Poels, T. Depuydt, R. Van Den Begin, C. Jaudet, C. Collen, B. Engels, T. Reynders, M. Boussaer, T. Gevaert, M. De Ridder, and D. Verellen. The long- and short-term variability of breathing induced tumor motion in lung and liver over the course of a radiotherapy treatment. *Radiotherapy and Oncology*, 126:339–346, 2018. doi:10.1016/j.radonc.2017.09.001.

[DVM+20] J. Dhont, D. Verellen, I. Mollaert, V. Vanreusel, and J. Vandemeulebroucke. Realdrr – rendering of realistic digitally reconstructed radiographs using locally trained image-to-image translation. *Radiotherapy and Oncology*, 153:213–219, 2020. doi:10.1016/j.radonc.2020.10.004.

[EC] European Commission EC. Ecis - european cancer information system. https://ecis.jrc.ec.europa.eu/. Online; accessed November 2023.

[EDNDGH18] H. Emami, M. Dong, S.P. Nejad-Davarani, and C.K. Glide-Hurst. Generating synthetic cts from magnetic resonance images using generative adversarial networks. *Medical Physics*, 45:3627–36, 2018. doi:10.1002/mp.13047.

[EFG18] E. Engwall, A. Fredriksson, and L. Glimelius. 4d robust optimization including uncertainties in time structures can reduce the interplay effect in proton pencil beam scanning radiation therapy. *Medical Physics*, 2018. doi:10.1002/mp.13094.

[EKN+17] A. Esteva, B. Kuprel, R.A. Novoa, J. Ko, S.M. Swetter, H.M. Blau, and S. Thrun. Dermatologist-level classification of skin cancer with deep neural networks. *Nature*, 2017. doi:10.1038/nature21056.

[EPF14] D. Eigen, C. Puhrsch, and R. Fergus. Depth map prediction from a single image using a multi-scale deep network. *arXiv*, 2014. doi:10.48550/arXiv.1406.2283.

[EPS+10] C.L. Eccles, R. Patel, A.K. Simeonov, G. Lockwood, M. Haider, and L.A. Dawson. Comparison of liver tumor motion with and without abdominal compression using cine-magnetic resonance imaging. *International Journal of Radiation Oncology, Biology, Physics*, 79, 2010. doi:10.1016/j.ijrobp.2010.04.028.

[FEL+] J. Ferlay, M. Ervik, F. Lam, M. Colombet, and L. Mery. Global cancer observatory: Cancer today. https://gco.iarc.fr/today. Online; accessed November 2023.

[FMD+87] B.A. Fraass, D.L. McShan, R.F. Diaz, R.K. Ten Haken, A. Aisen, S. Gebarski, G. Glazer, and A.S. Lichter. Integration of magnetic resonance imaging into radiation therapy treatment planning: I. technical considerations. *International Journal of Radiation Oncology, Biology, Physics*, 13:1897–908, 1987. doi:10.1016/0360-3016(87)90358-0.

[FQT+19] X. Feng, K. Qing, N.J. Tustison, C.H. Meyer, and Q. Chen. Deep convolutional neural network for segmentation of thoracic organs-at-risk using cropped 3d images. *Medical Physics*, 46:2169–2180, 2019. doi:10.1002/mp.13466.

[FWC+19] J. Fan, J. Wang, Z. Chen, C. Hu, Z. Zhang, and W. Hu. Automatic treatment planning based on three-dimensional dose distribution predicted from deep learning technique. *Medical Physics*, 46:370–381, 2019. doi:10.1002/mp.13271.

[GA92] B.L. Guthrie and J.R. Adler. Computer-assisted preoperative planning, interactive surgery, and frameless stereotaxy. *Clinical Neurosurgery*, 1992.

[GBCB16] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio. *Deep learning*. MIT press Cambridge, 2016.

[GDCXJ14] D. Gursoy, F. De Carlo, Xi. Xiao, and C. Jacobsen. Tomopy: a framework for the analysis of synchrotron tomographic data. *Journal of Synchrotron Radiation*, 21:1188–1193, 2014. doi:10.1107/S1600577514013939.

[GMM+16] G. Guidi, N. Maffei, B. Meduri, E. D'Angelo, G.M. Mistretta, P. Ceroni, A. Ciarmatori, A. Bernabei, S. Maggi, M. Cardinali, V.E. Morabito, F. Rosica, S. Malara, A. Savini, G. Orlandi, C. D'Ugo, F. Bunkheila, M. Bono, S. Lappi, C. Blasi, and T. Costi. A machine learning tool for re-planning and adaptive rt: A multicenter cohort investigation. *Physica Medica*, 32:1659–1666, 2016. doi:10.1016/j.ejmp.2016.10.005.

[GMO+18] A. Grimwood, H. McNair, T. O'Shea, S. Gilroy, K. Thomas, J. Bamber, A. Tree, and E. Harris. In vivo validation of elekta's clarity autoscan for ultrasound-based intrafraction motion estimation of the prostate during radiation therapy. *International Journal of Radiation Oncology, Biology, Physics*, 2018. doi:10.1016/j.ijrobp.2018.04.008.

[Goi04] M. Goitein. Organ and tumor motion: an overview. *Seminars in Radiation Oncology*, 14:2–9, 2004. doi:10.1053/j.semradonc.2003.10.007.

[GSL+14] S. Goossens, F. Senny, J.A. Lee, G. Janssens, and X. Geets. Assessment of tumor motion reproducibility with audio-visual coaching through successive 4d ct sessions. *Journal of Applied Clinical Medical Physics*, 15, 2014. doi:10.1120/jacmp.v15i1.4332.

[GXW+21] L. Gao, K. Xie, X. Wu, Z. Lu, C. Li, J. Sun, T. Lin, J. Sui, and X. Ni. Generating synthetic ct from low-dose cone-beam ct by using generative adversarial networks for adaptive radiotherapy. *Radiation Oncology*, 2021. doi:10.1186/s13014-021-01928-w.

[Han17] X. Han. Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical Physics*, 44:1408–19, 2017. doi:10.1002/mp.12155.

[HB05] C. Harvey and M. Blomley. Principles and precautions of conventional radiography. *Surgery (Oxford)*, 23(5):158–161, 2005. doi:10.1383/surg.23.5.158.65501.

[HHG+20] E. Huynh, A. Hosny, C. Guthier, D. Bitterman, S. Petit, D. Haas-Kogan, B. Kann, H. Aerts, and R. Mak. Artificial intelligence in radiation oncology. *Nature Reviews Clinical Oncology*, 17:771–781, 2020. doi:10.1038/s41571-020-0417-8.

[HMZ14] T. Hellevik and I. Martinez-Zubiaurre. Radiotherapy and the tumor stroma: the importance of dose and fractionation. *Frontiers in Oncology*, 2014. doi:10.3389/fonc.2014.00001.

[HOM+20] A. Hoffmann, B. Oborn, M. Moteabbed, S. Yan, T. Bortfeld, A. Knopf, H. Fuchs, D. Georg, J. Seco, M.F. Spadea, O. Jäkel, C. Kurz, and K. Parodi. Mr-guided proton therapy: a review and a preview. *Radiation Oncology*, 15(129), 2020. doi:10.1186/s13014-020-01571-x.

[HRRR17] P. Henzler, V. Rasche, T. Ropinski, and T. Ritschel. Single-image tomography: 3d volumes from 2d cranial x-rays. *arXiv*, 2017. doi:10.48550/ARXIV.1710.04867.

[HSG+24] W. Hartsell, C. Simone, D. Godes, J. Maggiore, M. Mehta, S. Frank, J. Metz, and I. Choi. Temporal evolution and diagnostic diversification of patients receiving proton therapy in the united states: A ten-year trend analysis (2012 to 2021) from the national association for proton therapy. *International Journal of Radiation Oncology, Biology and Physics*,

119(4):1069–1077, 2024. doi:10.1016/j.ijrobp.2023.12.041.

[HSW89] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2:359–66, 1989. doi:10.1016/0893-6080(89)90020-8.

[HWST19] R. Hirai, W. Watanabe, Y. Sakata, and A. Tanizawa. Real-time linear fiducial marker tracking in respiratory-gated radiotherapy for hepatocellular carcinoma. *International Journal of Radiation Oncology, Biology, Physics*, 105:E750–E751, 2019. doi:10.1016/j.ijrobp.2019.06.769.

[IWH+20] M. Iocolano, A.T. Wild, M. Hannum, Z. Zhang, C.B. Simone, D. Gelblum, A.J. Wu, A. Rimner, and A.F. Shepherd. Hypofractionated vs. conventional radiation therapy for stage iii non-small cell lung cancer treated without chemotherapy. *Acta Oncologica*, 59:164–170, 2020. doi:10.1080/0284186X.2019.1675907.

[JJOdX+11] G. Janssens, L. Jacques, J. Orban de Xivry, X. Geets, and B. Macq. Diffeomorphic registration of images with variable contrast enhancement. *International Journal of Biomedical Imaging*, 2011:1–16, 2011. doi:10.1155/2011/891585.

[KB14] D.P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv*, 2014. doi:10.48550/arXiv.1412.6980.

[KDK+10] W. Kilby, J.R. Dooley, G. Kuduvalli, S. Sayeh, and C.R. Maurer. The cyberknife robotic radiosurgery system in 2010. *Technology in Cancer Research and Treatment*, 9:433–452, 2010. doi:10.1177/153303461000900502.

[KMB+06] P. Keall, G.S. Mageras, J. Balter, R. Emery, K. Forster, S. Jiang, J. Kapatoes, D. Low, M. Murphy, B. Murray, C. Ramsey, M. van Herk, S. Vedam, J. Wong, and E. Yorke. The management of respiratory motion in radiation oncology report of aapm task group 76. *Medical Physics*, 33, 2006. doi:10.1118/1.2349696.

[KNO+18] P.J. Keall, D.T. Nguyen, R. O'Brien, P. Zhang, L. Happersett, J. Bertholet, and P.R. Poulsen. Review of real-time 3-dimensional image guided radiation therapy on standard-equipped cancer radiation therapy systems: Are we at the tipping point for the era of real-time radiation therapy? *International Journal of Radiation Oncology, Biology, Physics*, 102:922–931, 2018. doi:10.1016/j.ijrobp.2018.04.016.

[Kor15] S.S. Korreman. Image-guided radiotherapy and motion management in lung cancer. *The British journal of radiology*, 88, 2015. doi:10.1259/bjr.20150100.

[KPL+22] M. Kawula, D. Purice, M. Li, G. Vivar, S-A. Ahmadi, K. Parodi, C. Belka, G. Landry, and C. Kurz. Dosimetric impact of deep learning-based ct auto-segmentation on radiation therapy treatment planning for prostate cancer. *Radiation Oncology*, 17, 2022. doi:10.1186/s13014-022-01985-9.

[Kru18] J. Kruse. *Proton image guidance*. CRC Press, 2018.

[LBCE+03] Y.K. Lee, M. Bollet, G. Charles-Edwards, M.A. Flower, M.O. Leach, H. McNair, E. Moore, C. Rowbottom, and S. Webb. Radiotherapy treatment planning of prostate cancer using magnetic resonance imaging alone. *Radiotherapy and Oncology*, 66:203–16, 2003. doi:10.1016/s0167-8140(02)00440-1.

[LBD+89] Y. LeCun, B. Boser, J.S. Denker, D. Henderson, R.E. Howar, W. Hubbard, and L.D. Jackel. Backpropagation applied to handwritten zip code recognition. *Neural computation*, 1989.

[LBS+07] H.S. Lu, R. Brett, G. Sharp, S. Safai, S. Jiang, J. Flanz, and H. Kooy. A respiratory-gated treatment system for proton therapy. *Medical Physics*, 34:3273–78, 2007. doi:10.1118/1.2756602.

[LCN+19] X. Liang, L. Chen, D. Nguyen, Z. Zhou, X. Gu, M. Yang, J. Wang, and S. Jiang. Generating synthesized computed tomography (ct) from cone-beam computed tomography (cbct) using cyclegan for adaptive radiation therapy. *Physics in Medicine and Biology*, 64, 2019. doi:10.1088/1361-6560/ab22f9.

[LDSM22a] E. Loÿen, D. Dasnoy-Sumell, and B. Macq. 3dct reconstruction from a single x-ray projection using convolutional neural network. In *2022 IEEE International Conference on Image Processing (ICIP)*, pages 1111–1115, 2022. doi:10.1109/ICIP46576.2022.9897902.

[LDSM22b] E. Loÿen, D. Dasnoy-Sumell, and B. Macq. Mo-0215 3dct reconstruction from single x-ray projection using deep learning. *Radiotherapy and Oncology*, 170:S171–S172, 2022. doi:10.1016/S0167-8140(22)02317-9.

[LDSM23] E. Loÿen, D. Dasnoy-Sumell, and B. Macq. Patient-specific three-dimensional image reconstruction from a single x-ray projection using a convolutional neural network for online radiotherapy applications. *Physics and Imaging in Radiation Oncology*, 26:100444, 2023. doi:10.1016/j.phro.2023.100444.

[LKT23] G. Landry, C. Kurz, and A. Traverso. The role of artificial intelligence in radiotherapy clinical practice. *The British journal of radiology*, 18, 2023. doi:10.1259/bjro.20230030.

[LLW+20] Y. Liu, Y. Lei, T. Wang, Y. Fu, X. Tang, W.J. Curran, T. Liu, P. Patel, and X. Yang. Cbct-based synthetic ct generation using deep-attention cyclegan for pancreatic adaptive radiotherapy. *Medical Physics*, 47:2472–83, 2020. doi:10.1002/mp.14121.

[LSC+16] W. Liu, S. Schild, J. Chang, Z. Liao, Y.H.n Chang, Z. Wen, J. Shen, J.B. Stoker, X. Ding, and Y. Hu. Exploratory study of 4d versus 3d robust optimization in intensity modulated proton therapy for lung cancer. *International Journal of Radiation Oncology, Biology, Physics*, 2016. doi:10.1016/j.ijrobp.2015.11.002.

[LSJ20] J. Lee, E. Sterpin, and G. Janssens. Engineering challenges in proton therapy. *Université Catholique de Louvain*, 2020.

[LSK+17] L. Lin, K. Souris, M. Kang, A. Glick, H. Lin, S. Huang, K. Stutzer, G. Janssens, E. Sterpin, Lee J.A., T. Solberg, J. McDonough, C. Simone, and Ben-Josef E. Evaluation of motion mitigation using abdominal compression in the

clinical implemen- tation of pencil beam scanning proton therapy of liver tumors. *Medical Physics*, 44:703–12, 2017. doi:10.1002/mp.12040.

[LSY+21] Y. Liang, W. Song, J. Yang, L. Qiu, K. Wang, and L. He. X2teeth: 3d teeth reconstruction from a single panoramic radiograph. *arXiv*, 2021. doi:10.48550/arXiv.2108.13004.

[LV20] J. Lee and M. Verleysen. Machine learning: regression, deep networks and dimensionality reduction. *Université Catholique de Louvain*, 2020.

[LYZ+22] C. Liu, M. Yang, D. Zhang, M. Chen, and D. Zhu. Clinical cancer immunotherapy: Current progress and prospects. *Frontiers in Immunology*, 2022. doi:10.3389/fimmu.2022.961805.

[LZZ15] H. Li, R. Zhu, and X. Zhang. Reducing dose uncertainty for spot-scanning proton beam therapy of moving tumors by optimizing the spot delivery sequence. *International Journal of Radiation Oncology, Biology, Physics*, 93:547–56, 2015. doi:10.1016/j.ijrobp.2015.06.019.

[MBP+22] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, and D. Terzopoulos. Image segmentation using deep learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44:3523–3542, 2022. doi:10.1109/TPAMI.2021.3059968.

[MBvH21] S. Mercieca, J.S.A. Belderbos, and M. van Herk. Challenges in the target volume definition of lung cancer radiotherapy. *Translational Lung Cancer Research*, 10:1983–1998, 2021. doi:10.21037/tlcr-20-627.

[MFD+10] R. Muirhead, C. Featherstone, A. Duton, K. Moore, and S. McNee. The potential benefit of respiratory gated radiotherapy (rgrt) in non-small cell lung cancer. *Radiotherapy and Oncology*, 95:172–177, 2010. doi:10.1016/j.radonc.2010.02.002.

[MKU18] S. Mori, A.C. Knopf, and K. Umegaki. Motion management in particle therapy. *Medical Physics*, 45, 2018. doi:10.1002/mp.12679.

[MLC+16] S. Mutic, D. Low, T. Chmielewski, G. Fought, G. Gerganov, M. Hernandez, I. Kawrakow, E. Kishawi, G. Kuduvalli, A. Sharma, S. Shvartsman, J. Tabachnik, R. Kashani, O.L. Green, and J.F. Dempsey. The design and implementation of a novel compact linear accelerator–based magnetic resonance imaging–guided radiation therapy (mr-igrt) system. *International Journal of Radiation Oncology, Biology, Physics*, 2016. doi:10.1016/j.ijrobp.2016.06.2234.

[MLKP20] B. Macq, J. Lee, G. Kerckhofs, and F. Peeters. Medical imaging. *Université Catholique de Louvain*, 2020.

[MN18] K. Myeongsu and J.J. Noel. Machine learning: Fundamentals. *Prognostics and Health Management of Electronics*, 2018. doi:10.1002/9781119515326.ch4.

[MZL+21] J-C. Montoya, C Zhang, Y Li, K Li, and G-H Chen. Reconstruction of three-dimensional tomographic patient models for radiation dose modulation in ct from two scout views using deep learning. *Medical Physics*, 49:1–16, 2021. doi:10.1002/mp.15414.

[NDW00] C. Nutting, D.P. Dearnaley, and S. Webb. Intensity modulated radiation therapy: a clinical review. *The British Journal of Radiology*, 73:459–69, 2000. doi:10.1259/bjr.73.869.10884741.

[NHA+21] N. Norori, Q. Hu, F.M. Aellen, F.D. Faraci, and A. Tzovara. Addressing bias in big data and ai for health care: A call for open science. *Patterns*, 2021. doi:10.1016/j.patter.2021.100347.

[NJS+19] D. Nguyen, X. Jia, D. Sher, M-H. Lin, Z. Iqbal, H. Liu, and S. Jiang. 3d radiotherapy dose prediction on head and neck cancer patients with a hierarchically densely connected u-net deep learning architecture. *Physics in Medicine and Biology*, 2019. doi:10.1088/1361-6560/ab039b.

[NNM+09] M. Nakamura, Y. Narita, Y. Matsuo, M. Narabayashi, M. Nakata, A. Sawada, T. Mizowaki, Y. Nagata, and M. Hiraoka. Effect of audio coaching on correlation of abdominal displacement with lung tumor motion. *International Journal of Radiation Oncology, Biology, Physics*, 75:558–563, 2009. doi:10.1016/j.ijrobp.2008.11.070.

[NSJ+23] D. Niraula, W. Sun, J. Jin, I. Dinov, K. Cuneo, J. Jamaluddin, M. Matuszak, Y. Luo, T. Lawrence, S. Jolly, R. Haken, and I. Naqa. A clinical decision support system for ai-assisted decision-making in response-adaptive radiotherapy (arclids). *Nature*, 2023. doi:10.1038/s41598-023-32032-6.

[OBF+16] T. O'Shea, J. Bamber, D. Fontanarosa, S. van der Meer, F. Verhaegen, and E. Harris. Review of ultrasound image guidance in external beam radiotherapy part ii: intra-fraction motion management and novel applications. *Physics in Medicine and Biology*, 2016. doi:10.1088/0031-9155/61/8/R90.

[Ott08] K. Otto. Volumetric modulated arc therapy: Imrt in a single gantry arc. *Medical Physics*, 35:310–7, 2008. doi:10.1118/1.2818738.

[PAW+23] S. Pan, E. Abouei, J. Wynne, C.-H. Chang, T. Wang, R. Qiu, Y. Li, J. Peng, J. Roper, P. Patel, D. Yu, H. Mao, and X. Yang. Synthetic ct generation from mri using 3d transformer-based denoising diffusion model. *Medical Physics*, 51, 2023. doi:10.1002/mp.16847.

[PBL+17] L. Placidi, A. Bolsi, A.J. Lomax, R.A. Schneider, R. Malyapa, D.C. Weber, and F. Albertini. Effect of anatomic changes on pencil beam scanned proton dose distributions for cranial and extracranial tumors. *International Journal of Radiation Oncology, Biology, Physics*, 97(3):616–623, 2017. doi:10.1016/j.ijrobp.2016.11.013.

[PBPPM09] M.C. Popescu, V.E. Balas, L. Perescu-Popescu, and N.E. Mastorakis. Multilayer perceptron and neural networks. *WSEAS Transactions on Circuits and Systems 8*, 2009.

[PBSW21] H. Paganetti, P. Botas, G.C. Sharp, and B. Winey. Adaptive proton therapy. *Physics in Medicine and Biology*, 66(22), 2021. doi:10.1088/1361-6560/ac344f.

[PDR22] P.M. Putora and D.K. De Ruysscher. Is hypofractionation a good idea in radiotherapy for locally advanced nsclc? *Journal of Thoracic Oncology*, 17:487–488, 2022. doi:10.1016/j.jtho.2022.01.019.

[PHNACG18] O. Piippo-Huotari, E. Norrman, A. Anderzén-Carlsson, and H. Geijer. New patient-controlled abdominal compression method in radiography: radiation dose and image quality. *Acta Radiol Open*, 7:1–7, 2018. doi:10.1177/2058460118772863.

[PLWM16] S. Park, S.J. Lee, E. Weiss, and Y. Motai. Inter-fractional variation prediction of lung tumors using fuzzy deep learning. *IEEE Journal of Translational Engineering in Health and Medicine*, 8, 2016. doi:10.1109/JTEHM.2016.2516005.

[PSP22] O. Pastor-Serrano and Z. Perkó. Millisecond speed deep learning based proton dose calculation with monte carlo accuracy. *Physics in Medicine and Biology*, 67(10), 2022. doi:10.1088/1361-6560/ac692e.

[RB10] E. Rietzel and C. Bert. Respiratory motion management in particle therapy. *Medical Physics*, 37:449–460, 2010. doi:10.1118/1.3250856.

[RBW11] P. Risholm, J. Balter, and W.M. Wells. Estimation of delivered dose in radiotherapy: the influence of registration uncertainty. *Medical Image Computing and Computer Assisted Intervention*, pages 548–555, 2011. doi:10.1007/978-3-642-23623-5_69.

[RdHDSJM23] G. Rotsart de Hertaing, D. Dasnoy-Sumell, G. Janssens, and B. Macq. Online tumor motion forecasting using vision transformers. *4D Treatment Planning Workshop for Particle Therapy*, 2023.

[REM+16] D.J. Rezende, S.M. Eslami, S. Mohamed, P. Battaglia, M. Jaderberg, and N. Heess. Unsupervised learning of 3d

structure from images. *arXiv*, 2016. doi:10.48550/arXiv.1607.00662.

[RHW86] D. Rumelhart, G. Hinton, and R. Williams. Learning representations by back-propagating errors. *Nature*, 323:533–6, 1986. doi:10.1038/323533a0.

[RIB+18] P. Rajpurkar, J. Irvin, R.L. Ball, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C.P. Langlotz, B.N. Patel, K.W. Yeom, K. Shpanskaya, F.G. Blankenberg, J. Seekins, T.J. Amrhein, D.A. Mong, S.S. Halabi, E.J. Zucker, A.Y. Ng, and M.P. Lungren. Deep learning for chest radiograph diagnosis: A retrospective comparison of the chexnext algorithm to practicing radiologists. *PLOS Medicine*, 2018. doi:10.1371/journal.pmed.1002686.

[RLLL19a] X-C. Ren, Y-E. Liu, J. Li, and Q. Lin. Progress in image-guided radiotherapy for the treatment of non-small cell lung cancer. *World Journal of Radiology*, 11:46–54, 2019. doi:10.4329/wjr.v11.i3.46.

[RLLL19b] X.C. Ren, Y.E. Liu, J. Li, and Q. Lin. Progress in image-guided radiotherapy for the treatment of non-small cell lung cancer. *World Journal of Radiology*, 11:46–54, 2019. doi:10.4329/wjr.v11.i3.46.

[RSL] RaySearch Laboratories RSL. Raystation - comprehensive cancer treatment planning. https://www.raysearchlabs.com/raystation/. Accessed: 2024-05.

[SCD+17] R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 618–626, 2017. doi:10.1109/ICCV.2017.74.

[SGK+16] K. Souris, A. Glick, M. Kang, G. Janssens, E. Sterpin, H. Lin, J. McDonough, C. Simone, T. Solberg, and E. Ben-Josef. Abdominal compression effectively reduces the interplay effect and enables pencil beam scan- ning proton therapy of liver tumors. *Medical Physics*, 2016. doi:10.1118/1.4956257.

[Shu22] S. Shuel. Targeted cancer therapies. *Canadian Family Physician*, 2022. doi:10.46747/cfp.6807515.

[SKP10] D. Strigl, K. Kofler, and S. Podlipnig. Performance and scalability of gpu-based convolutional neural networks. *18th Euromicro Conference on Parallel, Distributed and Network-based Processing*, pages 317–24, 2010. doi:10.1109/PDP.2010.43.

[SMNJ23] RL. Siegel, KD Miller, Wagle NS., and A Jemal. Cancer statistics, 2023. *CA: A Cancer Journal for Clinicians*, 73:17–48, 2023. doi:10.3322/caac.21763.

[SMZS21] M.F. Spadea, M. Maspero, P. Zaffino, and J. Seco. Deep learning based synthetic-ct generation in radiotherapy and pet: A review. *Medical Physics*, 48(11), 2021. doi:10.1002/mp.15150.

[SRTP09] J. Seco, D. Robertson, A. Trofimov, and H. Paganetti. Breathing interplay effects during proton beam scanning: simulation and statistical analysis. *Physics in Medicine and Biology*, 2009. doi:10.1088/0031-9155/54/14/N01.

[SVM+02] G. Soete, D. Verellen, D. Michielsen, V. Vinh-Hung, J. Van de Steene, D. Van den Berge, P. De Roover, F. Keuppens, and G. Storme. Clinical use of stereoscopic x-ray positioning of patients treated with conformal radiotherapy for prostate cancer. *International Journal of Radiation Oncology, Biology, Physics*, 54:948–952, 2002. doi:10.1016/S0360-3016(02)03027-4.

[SZX19] L. Shen, W. Zhao, and L. Xing. Patient-specific reconstruction of volumetric computed tomography images from a single projection view via deep learning. *Nature Biomedical Engineering*, 3:880–888, 2019. doi:10.1038/s41551-019-0466-4.

[TEL+10] G. Tang, M.A. Earl, S. Luan, C. Wang, M.M. Mohiuddin, and C.X. Yu. Comparing radiation treatments using intensity-modulated beams, multiple arcs, and single arcs. *International Journal of Radiation Oncology, Biology, Physics*, 76:1554–62, 2010. doi:10.1016/j.ijrobp.2009.04.003.

[Top19] Eric Topol. *Deep medicine: how artificial intelligence can make healthcare human again*. Hachette UK, 2019.

[VDM86] C. Van Der Malsburg. Frank rosenblatt: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms. *Brain Theory*, 1986. doi:10.1007/978-3-642-70911-1_20.

[VDRT+08] D. Verellen, M. De Ridder, K. Tournel, M. Duchateau, T. Reynders, and T. et al. Geveart. An overview of volumetric imaging technologies and their quality assurance for igrt. *Acta Oncologica*, 47:1271–1278, 2008. doi:10.1080/02841860802244182.

[vH07] M. van Herk. Different styles of image-guided radiotherapy. *Seminars in Radiation Oncology*, 17:258–67, 2007. doi:10.1016/j.semradonc.2007.07.003.

[VODSG19] G. Van Ooteghem, D. Dasnoy-Sumell, and X. Geets. Réduction des variations pendant et entre les fractions respiro-induites: le potentiel de la ventilation mécanique orientée vers les stratégies de mitigation du mouvement pour l'irradiation des tumeurs thoraciques ou abdominales. *Cancer/Radiothérapie*, 2019. doi:10.1016/j.canrad.2019.07.029.

[VODSL+19] G. Van Ooteghem, D. Dasnoy-Sumell, M. Lambrecht, G. Reychler, G. Liistro, E. Sterpin, and X. Geets. Mechanically-assisted non-invasive ventilation: A step forward to modulate and to improve the reproducibility of breathing-related motion in radiation therapy. *Radiotherapy and Oncology*, 2019. doi:10.1016/j.radonc.2018.12.026.

[VODSLG19] G. Van Ooteghem, D. Dasnoy-Sumell, J.A. Lee, and X. Geets. Mechanically-assisted and non-invasive ventilation for radiation therapy: A safe technique to regularize and modulate internal tumour motion. *Radiotherapy and Oncology*, 141:283–291, 2019. doi:10.1016/j.radonc.2019.09.021.

[VSV+18] S. Vemprala, S. Saripalli, C. Vargas, M. Bues, Y. Hu, and J. Shen. Real-time tumor tracking for pencil beam scanning

proton therapy. *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2018. doi:10.1109/IROS.2018.8593861.

[WDSJ+23] S. Wuyckens, D. Dasnoy-Sumell, G. Janssens, V. Hamaide, M. Huet, E. Loÿen, G. Rotsart de Hertaing, B. Macq, E. Sterpin, J. Lee, K. Souris, and Deffet S. Opentps – open source treatment planning system for research in proton therapy. *arXiv preprint*, 2023. doi:10.48550/arXiv.2303.00365.

[WHOa] World Health Organization WHO. Cancer. https://www.who.int/news-room/fact-sheets/detail/cancer. Online; accessed December 2023.

[WHOb] World Health Organization WHO. Cancer. https://www.who.int/health-topics/cancer#tab=tab_1. Online; accessed November 2023.

[WPZ06] Y. Wang, V. Probin, and D. Zhou. Cancer therapy-induced residual bone marrow injury - mechanisms of induction and implication for therapy. *Current Cancer Therapy Reviews*, 2006. doi:10.2174/157339406777934717.

[WSK+15] Z. Wu, S. Song, A. Khosla, F. Yu, L. Zhang, X. Tang, and J. Xiao. 3d shapenets: A deep representation for volumetric shapes. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015. doi:10.1109/CVPR.2015.7298801.

[WSS+04] B. Warkentin, P. Stavrev, N. Stavreva, C. Field, and BG. Fallone. A tcp-ntcp estimation module using dvhs and known radiobiological models and parameter sets. *Journal of Applied Clinical Medical Physics*, 5:50–63, 2004. doi:10.1120/jacmp.v5i1.1970.

[WSvHD08] JWH. Wolthaus, JJ. Sonke, M. van Herk, and EMF Damen. Reconstruction of a time-averaged midposition ct scan for radiotherapy planning of lung cancer patients using deformable registration. *Journal of Applied Clinical Medical Physics*, 35:3998–4011, 2008. doi:10.1118/1.2966347.

[WYK+09] Q.J. Wu, S. Yoo, J.P. Kirkpatrick, D. Thongphiew, and F.F. Yin. Volumetric arc intensity-modulated therapy for spine body radiotherapy: comparison with static intensity-modulated treatment. *International Journal of Radiation Oncology, Biology, Physics*, 75:1596–604, 2009. doi:10.1016/j.ijrobp.2009.05.005.

[YGM+19] X. Ying, H. Guo, K. Ma, J. Wu, Z. Weng, and Y. Zheng. X2ct-gan: Reconstructing ct from biplanar x-rays with generative adversarial networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10611–10620, 2019. doi:10.1109/CVPR.2019.01087.

[Yin19] X. Ying. An overview of overfitting and its solutions. *Journal of Physics: Conference Series*, 2019. doi:10.1088/1742-6596/1168/2/022022.

[YND18] R. Yamashita, M. Nishio, and R.K.G. et al. Do. Convolutional neural networks: an overview and application in radiology. *Insights Imaging*, 9:611–29, 2018. doi:10.1007/s13244-018-0639-9.

[ZLD+23] Y. Zhang, C. Li, Z. Dai, L. Zhong, X. Wang, and W. Yang. Breath-hold cbct-guided cbct-to-ct synthesis via multimodal unsupervised representation disentanglement learning. *IEEE Transactions on Medical Imaging*, 42(8):2313–2324, 2023. doi:10.1109/TMI.2023.3247759.

[ZS19] R. Zaheer and H. Shaziya. A study of the optimization algorithms in deep learning. In *Third International Conference on Inventive Systems and Control (ICISC)*, pages 536–9, 2019. doi:10.1109/ICISC44355.2019.9036442.

[ZXBB22] S. Zhou, X. Xu, J. Bai, and M. Bragin. Combining multi-view ensemble and surrogate lagrangian relaxation for real-time 3d biomedical image segmentation on the edge. *Neurocomputing*, 512:466–481, 2022. doi:10.1016/j.neucom.2022.09.039.

[ZZC+19] Q. Zhang, M. Zhang, T. Chen, Z. Sun, Y. Ma, and B. Yu. Recent advances in convolutional neural network acceler-

ation. *Neurocomputing*, 323:37–51, 2019. doi:10.1016/j.neucom.2018.09.038.

[ZZQ+19] J. Zhu, J. Zhang, B. Qiu, Y. Liu, X. Liu, and L. Chen. Comparison of the automatic segmentation of multiple organs at risk in ct images of lung cancer between deep convolutional neural network-based and atlas-based techniques. *Acta Oncologica*, 58:257–264, 2019. doi:10.1080/0284186X.2018.1529421.