
POLITECNICO DI MILANO
DEPARTMENT OF ELECTRONICS INFORMATION AND BIOENGINEERING (DEIB)
DOCTORAL PROGRAMME IN BIOENGINEERING

---

# ROBOTICS AND GENERATIVE ARTIFICIAL INTELLIGENCE FOR CONE BEAM CT PROCESSING IN IMAGE-GUIDED RADIOTHERAPY

Doctoral Dissertation of:
**Matteo Rossi**

Supervisor:
**Prof. Pietro Cerveri**

Tutor:
**Prof. Raffaele Dellacá**

The Chair of the Doctoral Program:
**Prof. Gabriele Angelo Dubini**

2023 – XXXIV Cycle

*To Ilaria and Sofia,*
*your endless love and support are my guiding light.*
*With deepest gratitude,*
*Matteo*

# Acknowledgment

I would like to express my sincere gratitude to my supervisor, Prof. Pietro Cerveri, for his invaluable guidance, support, and encouragement throughout this journey. His expertise and insights have been instrumental in shaping my research and have helped me to develop as a researcher. I would also like to thank Prof. Guido Baroni for all the suggestions and coordination during the development of the project. I am grateful to Gabriele Belotti, my Ph.D. adventure companion, for never being afraid to get his hands dirty with me in long debugging sessions.
On the CNAO side, I would like to thank in particular Andrea Pella, Rosalinda Ricotti, and Barbara Tagaste, along with all the Bioengineering Unit, for their technical support during the experimental session necessary for system commissioning. Thanks also to the technical department and the radiation protection team for the installation phases and safety checks. I am grateful also to the medical physics unit for all the suggestions related to the treatment dose update.
Finally, I would like to thank my family and friends for their unwavering support and encouragement throughout my studies. Their love and encouragement have been a constant source of inspiration and motivation for me.

# Abstract

The successful utilization of advanced techniques for precise dose delivery in external beam radiotherapy relies on the combination of highly accurate in-room imaging. In image-guided radiotherapy, daily images are obtained to guide adjustments in patient position and orientation, specifically addressing issues like weight loss, tumor shrinkage, and air bubbles that can affect the reliability of the treatment plan. Achieving stable patient immobilization and precise positioning during setup is necessary for optimal results. Various factors contribute to the different designs observed in the literature, including cost-effectiveness, choice of projector (gantry vs. fixed beams), number of treatment rooms, and patient throughput optimization. Cone-beam CT (CBCT) is an imaging technique commonly employed for anatomical evaluation before fraction delivery. However, CBCT imaging introduces challenges such as scattered radiations, leading to artifacts like shading, cupping, reduced contrast, and beam-hardening. Additionally, pixel values in CBCT images may fluctuate due to these artifacts, requiring correction methods to calibrate them to CT scanner Hounsfield Unit values for dose calculation. Moreover, the limited field of view (FOV) in CBCT also presents challenges in particle therapy, affecting the measurement of the complete particle beam path. Starting in 2013, the Centro Nazionale di Adroterapia Oncologica (CNAO) employed a custom cone-beam CT (CBCT) imaging system in one of its three treatment rooms. This system involved a robotic manipulator that supported a rotating C-arm with imaging equipment and a fixed field of view (FOV). In 2019, CNAO commissioned the development of a new customized robotic CBCT imaging system with the goal of enhancing patient throughput. This manuscript presents a feasibility study exploring the potential extension of CBCT in image-guided radiotherapy, aiming to address the known challenges associated with this imaging technique. This ambition was addressed by implementing a new robotic system and developing an innovative framework for image processing based on deep learning. This framework aims to process the CBCT images by calibrating the Hounsfield Unit (HU) values, removing artifacts caused by the conical geometry acquisition, and addressing the limitations of the narrow FOV. The ultimate aim is to explore the possibility of utilizing the existing limited FOV CBCT systems not only for patient positioning but also for dosimetric evaluation without requiring additional hardware modifications. A comprehensive description of the

technological innovation of this system is given, including the design, technical commissioning, and quality assurance protocols necessary to develop a custom in-room volumetric imaging system intended for radiation treatments using accelerated particles. Regarding methodological advancements, we explored the potential of employing deep learning to overcome the intrinsic limitations of CBCT imaging. In a first exploratory study, we confirmed that deep convolutional neural networks could generate precise synthetic CT scans from CBCT images. We also provided guidelines for selecting the most suitable training technique between supervised and unsupervised. Furthermore, we introduced techniques to address the limitations of deep learning-based CBCT correction by leveraging the transfer learning paradigm in a novel two-step approach. We demonstrated that preconditioning the network with synthetic data significantly improved the quality of corrected CBCT when working with limited datasets. Additionally, employing transfer learning with real data proved effective in enhancing performance when dealing with clinical data from real practice. Moreover, we tested the deep learning framework's applicability to treatment planning updates, specifically addressing the issue of limited FOV in CBCT and evaluating the consistency of proton dosimetry compared to the original planning CT. The calibration of CBCT scans yielded promising results, with less than a 2% difference in proton dosimetry compared to the planning CT. The potential impact of organ toxicity on organs at risk decreased from approximately 50% (without calibration) to 2% (with calibration). The gamma pass rate at 3%/2mm demonstrated a substantial improvement, accurately reproducing the prescribed dose with a 37% increase before and after calibration (53.78% vs. 90.26%). The results obtained combining the technological and methodological advancements proposed in this dissertation brings the use of narrow FOV CBCT scans incrementally closer to clinical translation in proton therapy planning updates.

# Contents

# Acronyms

**A**

| | |
|---|---|
| AP | Antero-Posterior. |

**C**

| | |
|---|---|
| CBCT | Cone Beam Computed Tomography. |
| cGAN | Cycle-consistent Generative Adversarial Network. |
| CNAO | Centro Nazionale di Adroterapia Oncologica. |
| CNN | Convolutional Neural Networks. |
| CNR | Contrast-to-Noise Ratio. |
| CTV | Clinical Target Volume. |

**D**

| | |
|---|---|
| DIR | Deformable Image Registration. |
| DOF | Degrees of Freedom. |
| DPR | Dose Difference Pass Rate. |

**F**

| | |
|---|---|
| FDK | Feldkamp-Davis-Kress. |
| FF | Full Fan. |
| FOV | Field of View. |
| FPD | Flat Panel Detector. |

**G**

| | |
|---|---|
| GAN | Generative Adversarial Network. |
| GPR | Gamma Pass Rate. |
| GUI | Graphical User Interface. |

**H**

| | |
|---|---|
| HF | Half Fan. |
| HU | Hounsfield Unit. |

**I**

| | |
|---|---|
| IGRT | Image-guided Radiotherapy. |
| IQR | Inter-quartile Range. |
| Iso-I | Imaging Isocenter. |
| Iso-T | Treatment Isocenter. |
| ITK | Insight Toolkit. |

**L**

| | |
|---|---|
| LL | Latero-Lateral. |
| LOO-CV | Leave-One-Out Cross-Validation. |

**M**

| | |
|---|---|
| MAE | Mean Absolute Error. |
| MC | Monte Carlo. |

**O**

| | |
|---|---|
| OAR | Organ at Risk. |
| OTS | Optical Tracking System. |

**P**

| | |
|---|---|
| pCT | Planning CT. |
| PPS | Patient Positioning System. |
| PRD | Pelvic Reference Dataset. |
| PSNR | Peak Signal-to-Noise Ratio. |

**R**

| | |
|---|---|
| rCT | Revaluation CT. |
| ROI | Region of Interest. |
| RTK | Recontruction Toolkit. |

**S**

| | |
|---|---|
| sCBCT | Synthetic CBCT. |
| sCT | Synthetic CT. |
| sCTc | sCT corrected. |
| sCTu | sCT uncorrected. |
| SI | Superior-Inferior. |
| SSIM | Structural Similarity Index Measure. |

# List of Figures

# List of Tables

CHAPTER 1

# Introduction

*This chapter provides the reader with an introduction to the Image-guided Radiotherapy pipeline at CNAO and the proposed contribution to the field. An overview of the organization of the present dissertation is also presented.*

## 1.1 Principle of Image-guided Radiotherapy

External beam radiotherapy is a widely used and effective clinical practice for treating tumor lesions. It harnesses the power of ionizing radiation to maximize the therapeutic outcome. The main goal is to sufficiently irradiate the patient's tumor tissue, ensuring proper tumor control while minimizing radiation-induced toxicity to nearby organs. Typically, a linear accelerator delivers therapeutic X-ray doses to the tumor. However, this method also affects healthy tissues close to the target area due to how photons release their energy, which exponentially decreases as they penetrate through tissues. Tumor tissues are more susceptible to radiation damage compared to healthy tissues. This characteristic is leveraged in clinical practice through a technique called fractionated treatment. The idea is to divide the total dose into multiple fractions, creating a differential effect between healthy and tumor tissues. During the breaks between fractions, healthy tissues repair a more significant portion of the damage than the tumor. This approach allows for the optimization of therapeutic outcomes by carefully selecting doses and fractionation intervals that do not exceed a certain threshold of damage to healthy cells while still effectively sterilizing the tumor tissue. In general, the radiation dose that can be utilized is limited by the tolerance of the surrounding tissues near the tumor site [2]. However, for tumors located in anatomically sensitive regions, such as e.g. near the brainstem, cranial nerves, or hepatocellular carcinomas requiring

preservation of liver tissue, the gradual release of radiation dose along the X-ray beam poses a challenge. Excessive irradiation of these areas could result in severe harm to the patient. On the other hand, hadrontherapy, an advanced form of radiotherapy, overcomes challenges associated with dose distribution in conventional X-ray radiotherapy. This technique utilizes cyclotrons and synchrotrons to accelerate heavy charged particles, like carbon ions and protons, forming a beam that can be directed towards the tumor. The dose deposited by different types of radiation at varying depths is represented by the Bragg curve (Fig. 1.1), which shows a peak at the energy level at which the particles come to a halt [1]. Unlike photons, particle beams in proton and carbon ion therapy exploit the Bragg peak to deliver a small energy dose to superficial tissues while the particles are still moving rapidly. Slowing down, heavy ions release the maximum dose at a specific depth determined by their initial energy and the density of the tissues they pass through. Beyond that point, the particles deliver a minimal dose. An additional advantage of using ions with high atomic weight is their increased radiobiological effectiveness on tissue [3]. This effectiveness depends on the ions' charge and their acceleration speed. Consequently, fast and low atomic weight ions have a similar energy transfer to X-rays, while slower and higher atomic weight ions have relatively higher effectiveness in degrading tissues compared to photons [2].

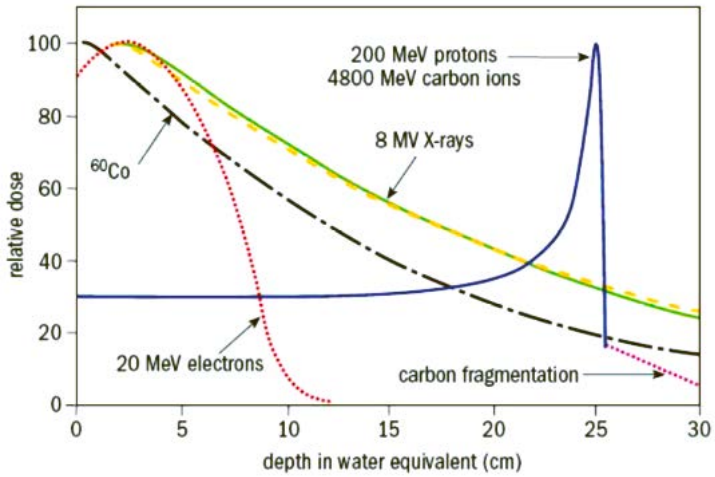


**Figure 1.1:** *The depth dependency with respect to the relative dose for various types of radiation. Due to the presence of the Bragg peak, the dose distribution could be considered as inverted compared to the nearly exponential pattern observed in high-energy photon beams. Source [1].*

Despite the significant advantages, such as reduced dose dispersion and enhanced biological effectiveness, proton and carbon ion therapy also have drawbacks. These include increased costs and complexity of the acceleration tools (cyclotrons or synchrotrons) and beam delivery. Additionally, this therapy requires techniques to reduce geometric uncertainties by accurately determining the three-dimensional position of the target tissue during treatment. Therefore, the effectiveness of advanced techniques designed for precise dose delivery in external beam radiotherapy can be successfully exploited only when combined with highly accurate in-room imaging. In an Image-guided Radiotherapy (IGRT) approach to patient treatment, daily images are acquired to guide rigid six Degrees of Freedom (DOF) setup corrections toward the planned patient po-

sition and orientation. Weight loss, tumor shrinkage, and air bubbles are well-known inter-fractional discrepancies to be accounted for in radiation therapy, especially for tumors located in the pelvic district [4]. The need for in-room imaging becomes even more crucial in the case of radiation therapy using heavy particle beams due to the inverse profile in dose deposition and the higher sensitivity to uncertainties [5–7]. Therefore, a common clinical practice involves using in-room imaging, with Cone Beam Computed Tomography (CBCT) being the most widely adopted technique for anatomical evaluation before fraction delivery [8–11]. State-of-the-art facilities are usually equipped with systems devoted to minimizing errors in patient repositioning [12–14]. In order to ensure precise alignment of the treatment target with the radiation beam, it is necessary to achieve stable immobilization of the patient and accurate positioning during setup. However, it is worth noting that while photon-based radiotherapy has widely adopted commercially available in-room imaging systems, the availability of dedicated volumetric image-guided devices for particle therapy has only recently emerged due to the complexities involved in installing such systems in a bunker room specifically designed for proton and carbon ion therapy [11, 12]. Factors such as the cost-effectiveness of particle therapy, the selected projector (gantry vs. fixed beams), the number of treatment rooms, and optimization of patient throughput (in-room vs. remote imaging) contribute to the variety of designs observed [13, 15–17]. Consequently, there is no universal image guidance solution that can be universally applied across all particle therapy centers. In the history of ion-beam therapy, patient positioning has traditionally relied on stereotactic registration using isocentric lasers and optical tracking systems. Subsequent setup corrections have been performed iteratively using orthogonal kV X-ray projections, in-room CT, or CBCT scanners [11]. Similarly, there is no dominant standard for modern heavy ion therapy centers. Various layouts have been developed, such as the use of in-room CT at the National Institute of Radiological Sciences or the customized setups at the Heidelberg Ion Therapy Center for both gantry and fixed beam lines [18, 19]. On the other hand, MedAustron recently employed the PAIR couch-mounted CBCT system [20]. In 2019 the Centro Nazionale di Adroterapia Oncologica (CNAO), a pioneering center specializing in advanced cancer treatment using precise radiation therapy with charged particles, started a project aimed at developing new solutions to improve IGRT in their clinical pipelines, commissioning the development of a new in-room CBCT system as an upgraded version of the custom robotic C-arm installed in 2013 [8]. The present dissertation describes all the experiments and activities carried out as part of this project related to CBCT imaging, resulting from the joint collaboration between CNAO and Politecnico di Milano.

## 1.2 Robotic CBCT Imaging System and Treatment Pipeline at CNAO

Since 2013, CNAO has been operating its first CBCT imaging system in one of the three available treatment rooms. These rooms have a fixed horizontal delivery line and can perform setup corrections based on bony landmarks. The one in which the first CBCT system was installed also features a vertical line [21, 22]. As a consequence, setup verification with 3D imaging was considered a key element in the final design of the imaging system [8]. Due to the simultaneous presence of the horizontal and vertical beamlines [23] and the limited space at the Treatment Isocenter (Iso-T), CNAO devel-

oped a custom solution based on a robotic manipulator mounting a rotating C-arm with imaging equipment and non-adjustable Field of View (FOV). In order to accommodate the limited space and integrate with existing equipment, the imaging system was installed in a non-treatment position within the room, thus defining a remote Imaging Isocenter (Iso-I). Such a solution allowed for obtaining adequate space for a 3D short scan (220°) imaging capabilities, as provided by the rotating C-arm. The entire treatment workflow, in terms of both Patient Positioning System (PPS) and in-room imaging, was designed to minimize the risk of collisions and patient injuries. The radiation therapy process begins by acquiring a 3D Planning CT (pCT) of the patient, enabling a precise definition of the tumor and safety margins along with the dose to be delivered. In the planning phase, clinicians also fit a personal thermoplastic mask to the patient's anatomy for immobilization, compression, and setup repeatability purposes [24]. Following the planning phase, the daily treatment routine unfolds as follows. Initially, the patient is prepared and immobilized on a carbon fiber table in a dedicated preparation area outside the treatment room (Fig. 1.2). The table is then docked onto the PPS and aligned to the nominal position at Iso-T based on the treatment plan. The alignment is initially verified in a two-step approach. First, a visual check is performed using an orthogonal laser cross superimposed on reference marks on the treatment couch. Then, the setup is refined using an optical marker-based frameless stereotactic localization method, utilizing infrared markers fixed on the thermoplastic compression mask. Subsequently, the CBCT imaging system is positioned in the acquisition pose, and the patient is automatically moved into the C-arm at Iso-I for a CBCT scan. Once the reconstruction of the 3D volume is performed, this is fed to an integrated 3D/3D registration algorithm and compared to the pCT of the patient in order to generate a correction vector with six degrees of freedom (DOF). Any rigid rototranslation applied at remote Iso-I is then propagated to Iso-T, maintaining consistency and effectively correcting the setup for treatment. In the case the clinicians decide that any of the components of the correction vector are too big, the treatment is stopped for that day, and the patients need to perform a Revaluation CT (rCT) for treatment planning updates. On the other hand, if the vector is under limits, the PPS is repositioned for treatment at Iso-T with the applied correction, and the imaging system is moved back into the parked state. A schematic of the treatment pipeline is depicted in Fig. 1.3.

## 1.3 Scientific Proposal and Work Contribution

In 2019, CNAO commissioned a new customized robotic CBCT imaging system for one of the lateral rooms with the aim of increasing patient throughput and relieving the workload for the central room. This upgrade was driven by the need to enhance in-room volumetric imaging availability. The new imaging system was designed based on the existing one but with hardware and software enhancements to address clinical requirements for faster and larger FOV CBCT. The implementation of a faster C-arm, coupled with software upgrades for reconstruction, resulted in shorter setup corrections, optimized room utilization, and reduced non-therapeutic radiation doses for patients. However, safety considerations imposed limitations on the rotation speed due to the response time of the anti-collision system installed on the robotic device. Moreover, when acquiring large anatomical districts, the main limitation of the legacy system was

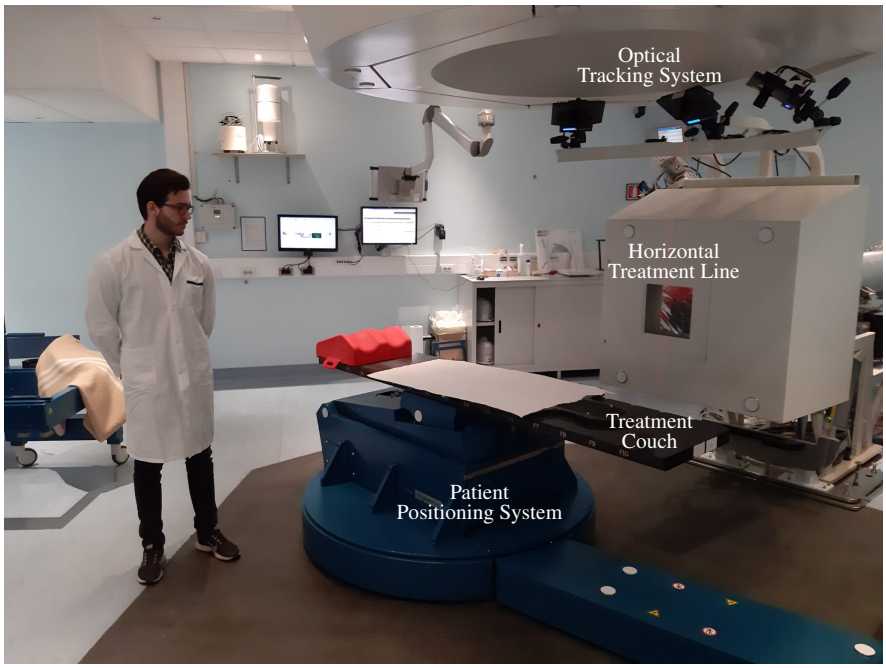


**Figure 1.2:** *Treatment Room 1 at CNAO with its main components.*

the absence of a field of view extension strategy for CBCT. With the legacy system, it was possible only to acquire CBCT with a symmetrical beam centered along the source-to-detector axis, a strategy commonly defined as Full Fan (FF), leading to truncation [25] and missing information in the axial plane of the reconstructed image. The new system aims to enlarge the FOV of the acquired CBCT by means of a strategy based on a displaced detector called Half Fan (HF) [26, 27]. The limited FOV also poses challenges in particle therapy as it hinders the measurement of the complete particle beam path. The legacy system is currently used for patient positioning correction and daily anatomical evaluation before fraction delivery. In addition, despite allowing for much faster imaging than CT, the CBCT modality introduces a significant amount of scattered radiations, resulting in image artifacts such as shading, cupping, reduced

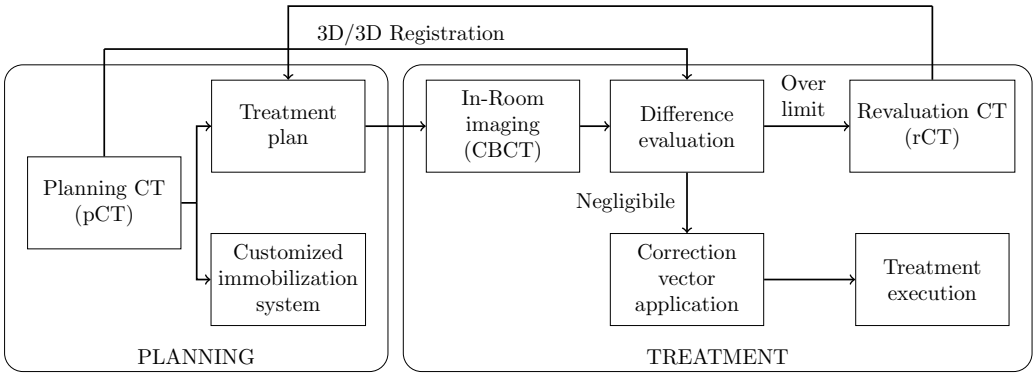


**Figure 1.3:** *Treatment pipeline at CNAO.*

image contrast, and beam-hardening [28, 29]. In particular, truncation influences scattering estimation and correction [30, 31]. Moreover, the pixel values in CBCT images may fluctuate because of these artifacts, making them not directly usable for dose calculation unless some correction methods are applied to calibrate CBCT to CT scanner Hounsfield Unit (HU) values [32–34]. The limited FOV also causes a truncation problem during reconstruction [35, 36]. Consequently, the non-uniqueness of the solution for the iterative reconstruction brings additional bright-band effects that add artifacts to the CBCT [37]. Even in the case of optimal HU calibration and scatter reduction, a CBCT acquired in a small FOV cannot be used for adaptive dose planning. Especially, narrow FOV CBCT lacks important anatomical information (e.g., the air/skin interface) necessary for properly calculating the beam path. All these issues prevent a proper dose evaluation using the CBCT imaging system installed at CNAO. At the same time, these limitations hinder the qualitative inspection of the daily anatomy, which leads to sub-optimal off-line procedures and unnecessary imaging doses delivered to the patient (i.e., the need for occasional rCT). Thus, the availability of reliable techniques for the correction of CBCT images would extend their clinical usages, such as tumor shrinkage and organ shift evaluation, and treatment planning purposes, such as soft tissue segmentation for dose delivery computation. Therefore, the implementation of the new imaging system also involves the definition of a deep-learning framework based on generative neural networks. This framework addresses the mentioned issues, such as calibrating HU, removing conic geometry acquisition artifacts, and resolving narrow FOV concerns, showcasing the potential of corrected CBCT for proton treatment planning updates. Furthermore, the framework aims to investigate the feasibility of utilizing the existing limited FOV CBCT system for both patient positioning and dosimetric evaluation without requiring any additional changes to the hardware. This way, the implemented methodologies could potentially be integrated also with the legacy system. A summary of the contributions of this work is summarized in Table 1.1

**Table 1.1:** *Summary of the main issues/needs addressed in the present dissertation, along with the proposed solution and approach used (T=Technological, M=Methodological)*

| Need/Issue | Proposed Solution | Approach | |
|---|---|---|---|
| | | T | M |
| Increase the patient throughput at CNAO | Development of the new CBCT imaging system in another treatment room | X | X |
| Increase safety | Anti-collision strategy and inhibition of simultaneous movement of room instrumentation | X | |
| CBCT scans have small FOV | Increase FOV with a displaced detector-based strategy | X | |
| CBCT scans are not calibrated to HU because of scattering artifacts | Implementation of neural networks for CBCT correction | | X |
| CBCT scans are used only for patient positioning | Evaluation of the feasibility of proton dosimetry once corrected through neural networks | | X |
| rCT introduces additional non-therapeutic doses | Corrected CBCT scans may reduce the number of rCT needed | | X |

## 1.4 Dissertation Outline

The rest of the dissertation is organized as follows. In Chapter 2, readers are introduced to the application of CBCT in image-guided radiotherapy. The initial section provides a comprehensive overview of the developed CBCT imaging system with a focus on the technological improvements introduced, while the subsequent section focuses on the challenges associated with CBCT imaging and the conventional approaches employed to address them. In Chapter 3, the use of deep neural networks for CBCT correction is examined, highlighting two primary paradigms employed for training these networks: supervised and unsupervised. The chapter presents a comparative analysis of the performance of these methods in generating synthetic CT scans from CBCT images and offers recommendations for selecting the most suitable training technique. The content of this chapter is derived from the publication by Matteo Rossi and Pietro Cerveri. "Comparison of supervised and unsupervised approaches for the generation of synthetic CT from cone-beam CT". Diagnostics (Basel, Switzerland), 11, 2021. Chapter 4 continues with an in-depth investigation of the application of deep learning-based approaches for CBCT correction, introducing a novel method to address the limitations of deep learning-based CBCT correction based on a two-step algorithm leveraging the transfer learning paradigm. This chapter draws upon the research paper by Matteo Rossi et al. "Image-based shading correction for narrow-FOV truncated pelvic CBCT with deep convolutional neural networks and transfer learning." Medical physics, 48:7112-7126, 2021. Chapter 5 provides an examination of the deep-learning framework's impact on treatment planning updates. The content of this chapter is derived from the research paper authored by Matteo Rossi et al. "Feasibility of Proton Dosimetry Overriding

Planning CT with Daily CBCT Elaborated through Generative Artificial Intelligence Tools." Preprints.org 2023, 2023040596. Finally, Chapter 6 concludes the present dissertation, reporting a summary of the aim, materials, methods, and obtained results.

CHAPTER 2

# Developing an innovative CBCT Imaging System at CNAO

*This chapter provides a detailed description of the developed CBCT imaging system, along with the technical commissioning performed to test it. The second part of this chapter presents the issues related to CBCT imaging and the traditional methods used to address them.*

## 2.1 Robotic Imaging System in Room 1

### 2.1.1 System Description

The imaging system consisted of an articulated serial manipulator with seven joints, mounting a custom-designed C-arm structure that includes a kV X-ray tube, dynamic collimator, and a displaceable Flat Panel Detector (FPD) with pulsed fluoroscopy capabilities. The Kawasaki BX300L robot was selected for installation, offering a payload capacity of 300 kg, position repeatability of 0.07 mm, and load accuracy of 0.3 mm. The development of a custom C-arm was commissioned to a third-party company. The C-arm was designed with a trussed structure using S235 JR (UNI EN 10027-1) steel to optimize the rigidity-to-weight ratio (cfr. Fig. 2.1). The geometry layout was optimized for clinical image acquisition with the following parameters: the tube-isocenter distance was fixed at 1100 mm, whereas the isocenter-to-detector was 500 mm. This resulted in an overall 1600 mm source-to-detector distance. The C-shape depth was defined as 1240 mm. A metal housing was incorporated to accommodate the FPD, equipped with a railing system that allowed for lateral displacements facilitated by a gearbox connected to a stepper motor controlled by the seventh joint of the robot (cfr. Fig. 2.2). The momentum and inertia of the C-arm were evaluated through simulations

involving realistic loads of 180 kg before the manufacturing process. The structure was intended to be connected posteriorly to the robot end effector.

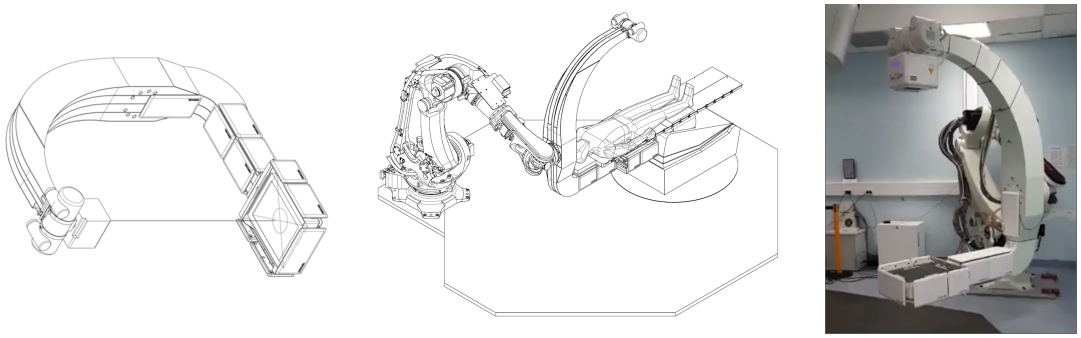

**Figure 2.1:** *Schematic of the designed C-arm where tube, collimator, and FPD are depicted (**left**); schematic of the entire system in acquisition (**center**); robot in acquisition layout after installation (**right**).*

The C-arm held the X-ray assembly (tube housing and dynamic collimator) at one extremity and the flat panel detector on the opposite side. In detail, the X-ray tube was an A292 with B130H housing (Varian Medical Systems, Palo Alto, CA, USA); its total weight, also considering the mounting trunnion ring, is approximately 30 kg. This model was chosen over the A277 of the previous system by considering the maximum anode angle guaranteed by the manufacturer (12° vs. 7°), sufficient for Half Fan (HF) acquisitions. A dynamic collimator (R221 ACS with R225 housing, Ralco, Biassono MI, Italy) was fitted on the tube housing to chase the FPD position. The flat panel was a Varian 4030D, with a sensible surface of 40×30 cm. The entire system was supplied with an HF series generator (SEDECAL, Algete MD, Spain) providing up to 150 kVp, 800 mA, 80 kW, and a heat exchanger (Varian HE-101) installed inside the treatment room, close to the robot. The resulting layout after installation is depicted in Fig. 2.1. In order to ensure consistent acquisition, the FPD and X-ray source are synchronized using a hardware connection. This synchronization allows for repeatable image acquisition. Through a dedicated Graphical User Interface (GUI), exposure parameters such as time (ms), tube voltage (kV), and tube current (mA) can be adjusted. The GUI also provides several presets tailored to specific robot poses and anatomical regions, enhancing convenience and efficiency in the clinical routine. During dynamic acquisitions, projections are generated with a pixel resolution of 0.388 mm in a matrix size of 1024×768, equivalent to 273×205 mm. Approximately 450 projections are acquired during the 220° rotation of the sixth joint of the robot as the technician presses on the dedicated pedal for continuous X-ray emission. Each acquisition takes 30 seconds, resulting in a rotation speed of 7.3° per second for the C-arm, while the FPD operates at a framerate of 15 Hz. Consequently, the angular sampling of the projections is 0.49°. It is important to remind that after completing a full-fan (FF) acquisition with a detector-centered position, the C-arm returns to its starting position within approximately 15 seconds. Conversely, when the operator requires a complementary half-fan (HF) acquisition, the robot laterally displaces the FPD by 120 mm in the right direction before scanning the patient. Following this, the first acquisition is conducted over 220° within around 30 seconds. Subsequently, the FPD is shifted to the opposite position in the left

direction to acquire the other half. The reconstruction of volumetric images is achieved through a GPU implementation of the Feldkamp-Davis-Kress (FDK) algorithm based on the Recontruction Toolkit (RTK) open-source software [38]. The final CBCT image is then reconstructed before the C-arm returns to the starting position. The image is automatically masked to the FOV pseudo cylinder size with an axial diameter of 204 and 403 mm for FF and HF modality, respectively. The length of the FOV in the superior-inferior direction is 271 mm. The resolution for the reconstructed volumes matches the pCT one ($0.98 \times 0.98 \times 1$ mm). Following the masking of CBCT reconstructed scans, they are superimposed onto the pCT and aligned through rigid registration. The registration algorithm utilized in this process relies on the Insight Toolkit (ITK) and involves three stages of minimizing normalized mutual information. An Amoeba optimizer is employed, and isotropic subsampling grids ranging from 3 mm to 1 mm are utilized. The entire registration procedure is completed in less than one minute.

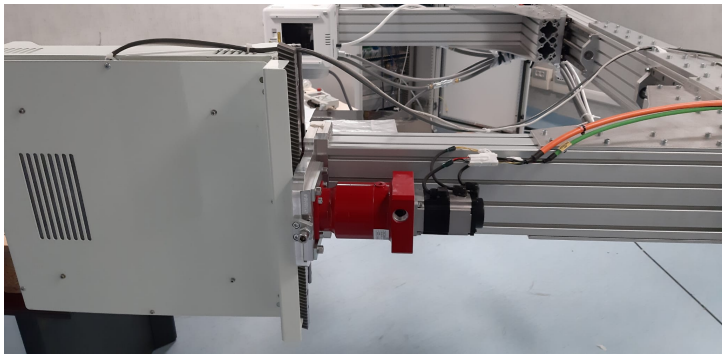

**Figure 2.2:** *FPD displacing system, involving a motor acting as an additional DOF for the robotic manipulator.*

### 2.1.2 Safety Aspects

Concerning safety aspects, an additional safety module (Kawasaki Cubic-S) was incorporated into the robot controller to ensure motion monitoring and enhance patient safety. This module adheres to various safety standards, including ISO10218-1, 13849-1 (PLd/Cat 3), and IEC61508 (SIL2). Its primary function is to monitor and halt the robot's motion in the presence of unauthorized movements or unexpected conditions. Monitoring allowed and prohibited areas effectively minimize the risk of collisions between the system and other equipment (Fig. 2.3). In the event of an emergency, the Cubic-S module facilitates fast and controlled stops, reducing the impact on ongoing tasks. Furthermore, it inhibits reciprocal motion between the robot and the PPS, preventing both systems from moving simultaneously. On the FPD side, an anti-collision system was installed to improve safety. This location is considered critical due to its proximity to the patient during treatment. The anti-collision system comprises AIRSKIN modules (Blue Danube Robotics GmbH, Vienna, Austria). These modules consist of soft pads that continuously monitor pressure using a dedicated control unit. If a force exceeding 5 N is detected, an emergency stop is triggered within a response time of 9 ms. The AIRSKIN pads comply with safety standards EN/ISO 13849-1 (PLe/Cat 3) and EN/IEC 62061 (SIL 3). At top speed during CBCT acquisition, this results in a $0.066^\circ$ delay in response and a maximum of 0.6 mm and 1.3 mm of linear travel before

stoppage at the panel and tube level, respectively.

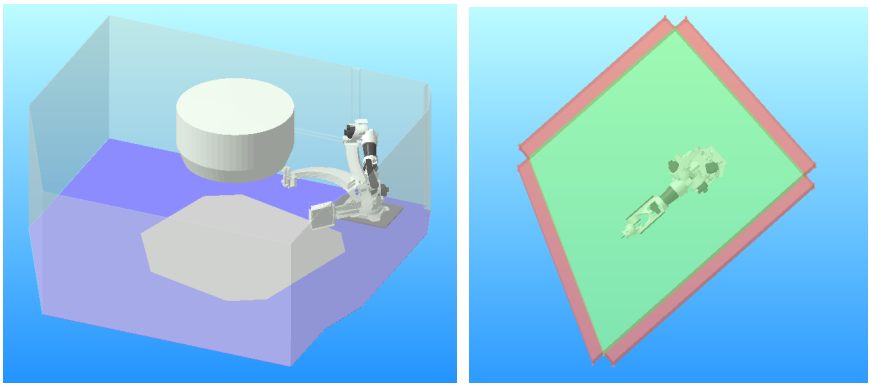

**Figure 2.3:** *(**Left**) Final design for robot installation. (**Right**) The defined permitted area (green) and prohibited ones (red). In the unexpected event that the robot collides with a red wall, an emergency stop is immediately triggered.*

### 2.1.3 Technical Commissioning

Although the single systems (PPS and robotic manipulator) offer a high level of accuracy, it was necessary to evaluate the accuracy of error propagation between Iso-I and Iso-T to guarantee precise setup correction before treatment administration. To fulfill this requirement, two sets of measurements were carried out in order to confirm the following:

1. the coherence of imaging and treatment isocenters under nominal conditions;
2. the consistency of geometry calibration across different imaging modalities.

These tests were performed using the positioning phantom P43029 (PTW-Freiburg, Germany), which comprised a PMMA cylinder phantom containing four inner reference steel rings and four spherical surface markers compatible with the Optical Tracking System (OTS) already installed in the treatment room. This phantom (referred to as the RING phantom henceforth) was acquired with the CNAO's CT system (Siemens Somaton) at the clinical resolution of $0.98 \times 0.98 \times 1$ mm. The in-room geometrical setup was replicated after the CT scan also based on the positions of surface markers. Each test was carried out following the calibration of the imaging projective parameters of the system, which was based on the Flexmap approach [39, 40]. The calibration process encompassed nine DOF and comprised a Levenberg-Marquardt optimization of back-projection errors. For this purpose, a geometric calibration X-ray phantom (Model 2008; Brandis Medizintechnik Vertriebs GmbH, Weinheim, Germany) with 37 embedded radio-opaque ball bearings (BB) was employed (Fig. 2.4).

The calibration procedure demanded the alignment of the central bead of the calibration phantom to Iso-T in relation to the room's coordinate reference system. This alignment was achieved by placing the phantom on the couch, locked by indexing bars in predetermined positions on the treatment couch. Then, the bed was moved to a predefined pose determined by laser tracking measurements of the phantom's position and orientation within the room's coordinate system. Once the phantom was aligned, the

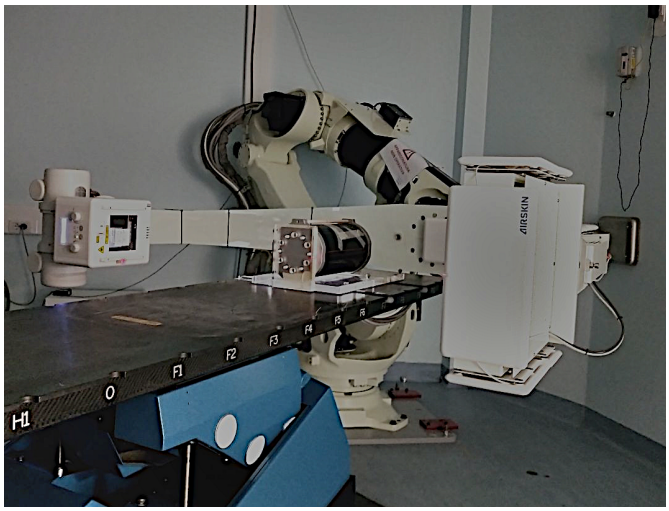


**Figure 2.4:** *Geometric calibration X-ray phantom setup at Iso-I. The locking bars that connect the phantom to the treatment couch, ensuring a repeatable pose, are also visible.*

PPS was moved to a second reference pose, also determined by laser tracker measurements. This ensured that the central axis of the couch was aligned with the rotational axis of the robot end effector and that the central beads of the phantom coincided with the center of rotation of the C-arm. A dynamic acquisition of the calibration phantom in CBCT FF modality was performed with the following exposure settings: 85 kVp, 100 mA, 8 ms. This acquisition involved a rotational range of 220° for a total of about 450 projections. A manual initialization (drag-and-drop) was performed to align the nominal BB positions with the actual projected BB, according to the angle of the end effector. This manual initialization regarded only the first acquired image, while the initialization of the subsequent projections was performed based on the final calibration parameters of the previous projection. Then, the algorithm extracted BB centroids from the acquired image, relying on a local window centered in the projected points. Finally, iterative optimization of calibration parameters minimized the back-projection errors (projected vs. extracted BB coordinates). In order to minimize unwanted drifts, per-projection parameters were subjected to smoothing using a moving average window of size ten. Regarding the HF mode, the procedure remained unchanged, but the calibration was performed independently for each of the complementary acquisitions.

#### Coherence of Geometric Calibration and Image Registration

The first test aimed to assess the quality of geometrical calibration and image registration algorithms. At first, the RING phantom was prealigned using lasers. The setup was then corrected by acquiring two orthogonal X-rays. Therefore, a second radiography set was acquired to double-check the setup correction. After that, CBCT scans were obtained to assess the expected residuals at Iso-I (cfr. Fig. 2.5). Finally, the OTS residuals at Iso-T were computed to validate the propagation of correction vectors calculated at Iso-I towards Iso-T. Note that OTS has a residual uncertainty lower than 0.3 mm [14, 23]. This test was conducted once since the RING phantom metallic inserts led to reproducible rigid alignments between different exposures.

As summarized in the left panel of Fig. 2.6, sub-millimeter and sub-degree max-

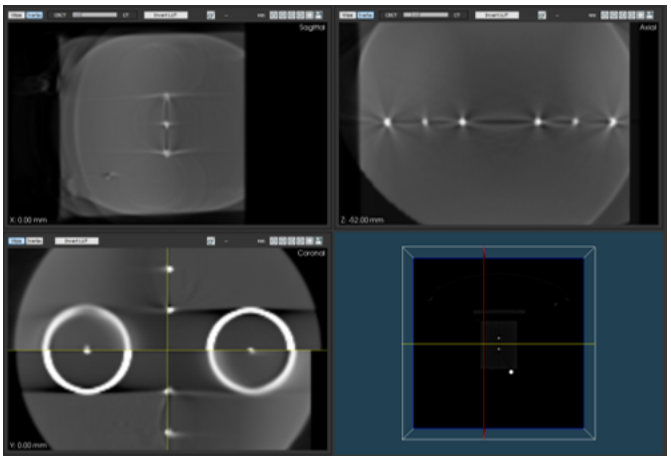

**Figure 2.5:** *Alignment of RING phantom in the CBCT modality. Sagittal (**upper left**), axial (**upper right**), and coronal (**bottom left**) views are rendered. The coronal view also shows the phantom in a chess view with the CT overlayed on the acquired in-room CBCT.*

imal discrepancies were measured when evaluating the geometrical setup error at the imaging and treatment isocenters for the anatomical axes Latero-Lateral (LL), Superior-Inferior (SI), and Antero-Posterior (AP). In general, the new scanner exhibited consistency with the OTS measures in the treatment position.

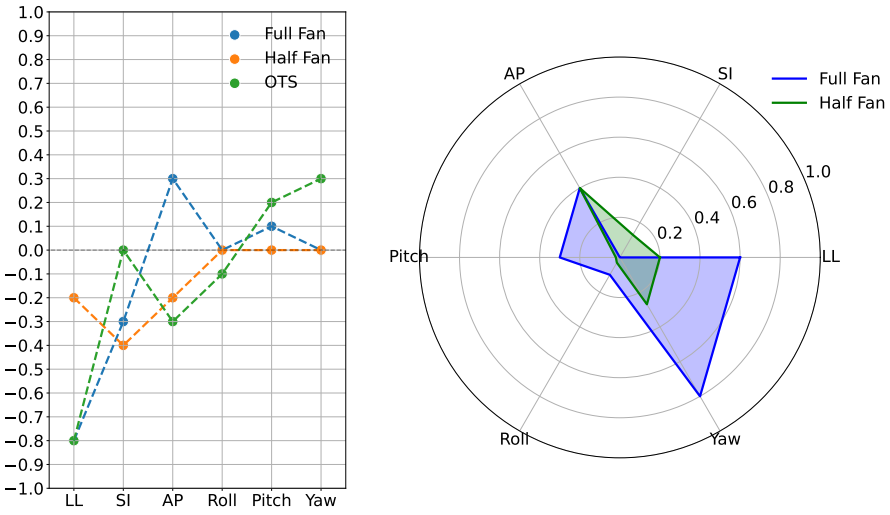


**Figure 2.6:** *Nominal accuracy of the system. Translation errors are reported in mm and rotation errors in degrees. (**Left**) Residuals of the phantom aligned to the nominal position as evaluated by the 3D/3D registrations and the OTS. (**Right**) Registration residuals after recovering the position error imposed on the PPS.*

A second test evaluated the system's capabilities in performing consistent rigid registration. The RING phantom was aligned using the isocentric lasers and OTS guidance. The PPS was driven in Iso-I, and a known composite 6 DOF error was applied to the nominal position. These errors were in the range of 3 mm for translations and 2° for rotations. CBCT scans in FF and HF modalities were then acquired, and the automatic

3D/3D registration towards the CT was applied. Finally, the absolute deviations with respect to the nominal position were collected. Residuals at Iso-I are depicted in the right panel of Fig. 2.6. FF underperformed for Yaw rotation and latero-lateral (LL) translation because the limited FOV cut some of the phantom body in the tested out-of-nominal condition. The HF modality outperformed the FF one, with residuals always under 0.4 mm.

#### Preclinical Phantom Study

The real clinical scenario cannot be well represented simply with the rigid registration of a geometrical phantom consisting of metallic rings and beads. The evaluation of overlapped sliced bony features eventually leads to more substantial residuals due to the complex and dense gradient distributions. Furthermore, the beads denoted a collection of points or fiducials resembling markers that are not present in patients. Consequently, in preparation for clinical utilization, a series of measurements were carried out on the pelvis region of the ATOM M701 adult male radio-equivalent anatomical phantom (CIRS, Norfolk, VA, USA). A CT scan was acquired with clinical resolution ($0.98\times0.98\times1$ mm). Five markers were placed for alignment purposes, as illustrated in Fig. 2.7. These measures aimed to evaluate the performance of the embedded algorithms within a large region after correcting the setup following the clinical pipeline. Consequently, the phantom underwent alignment using lasers and OTS toward Iso-I positions. Eight independent error vectors were applied to the PPS. A CBCT acquisition was then performed, and the absolute errors after 3D/3D registration were evaluated. In order to adhere to the clinical correction procedure, two acquisitions were performed per modality: one for the initial correction and another for follow-up verification and potential adjustments, considering the errors fell within the range of 10 mm for translation and 3° for rotations. Conversely, the pelvis phantom scans from the RING phantom also allowed for conducting axial FOV measurements. Mimicking the clinical conditions, the phantom study confirmed the accuracy of the registration pipeline while highlighting the advantage of volumetric imaging. Observing the maximum errors across modalities (cfr. Fig. 2.8), it can be assumed that the 3D imaging pipeline produced errors compatible with the clinical standard. However, there was a slight (0.2°) underperformance in Yaw rotation for HF modality. Despite the small size of the RING phantom, the dimensions of the pelvis phantom allowed for the appreciation of the axial FOV gain in the complementary HF strategy, as shown in Fig. 2.7. The FOV diameter in the axial plane increased almost twofold, from 204 mm in FF to 403 mm in HF, obtained with a combined 120 mm lateral displacement of the detector.

### 2.1.4 Main Contributions of the New CBCT System

The described device was installed and integrated into the IGRT workflow at CNAO. The results of commissioning activities and measurements have been presented, explicitly focusing on geometric accuracy and setup correction. The system consists of an articulated robot mounting a custom-designed C-arm structure. The choice of the robot and the geometry of the C-arm were driven by the desired functionalities of the application (patient positioning correction) in compliance with the functional requirements of the treatment room. As previously stated, the system was optimized for three-dimensional CBCT imaging with small and wide FOV. Additionally, the system needed

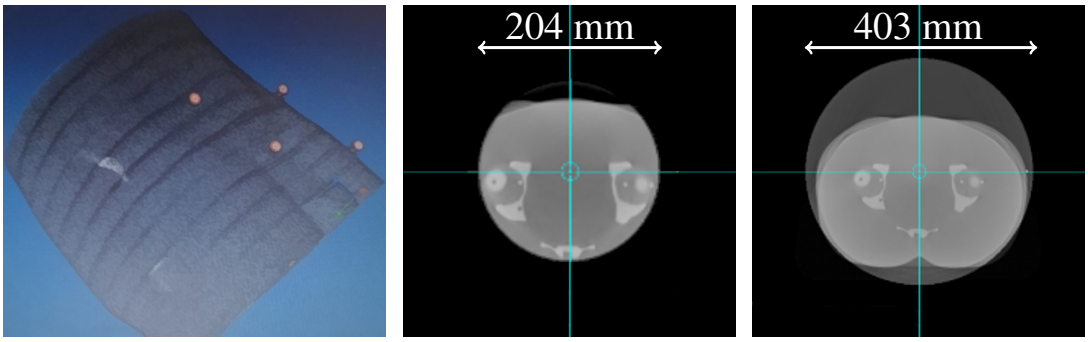


**Figure 2.7:** *Rendering of the pelvis phantom. Four out of five optical markers are easily recognizable, while the last one is not visible in this view (**left**). Axial view of a slice of the anatomical phantom acquired in FF mode (**center**) and HF one (**right**).*

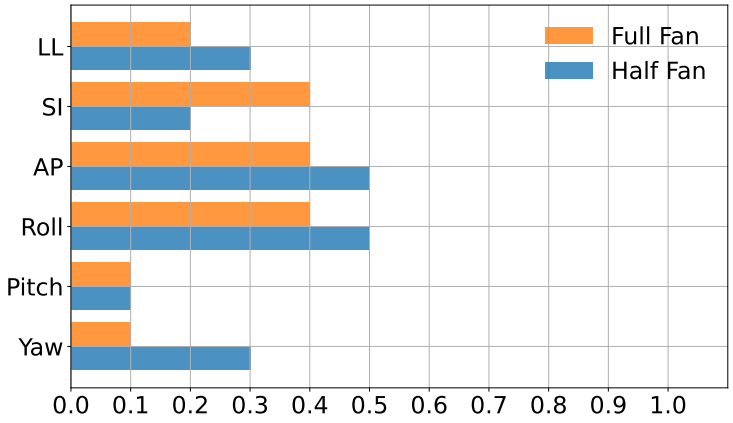


**Figure 2.8:** *Absolute residuals obtained with the 3D/3D registration after the imposition of some known vector on PPS at Iso-I for each modality. Translation errors are reported in mm and rotation ones in degree.*

sufficient flexibility to work within a treatment room with a fixed beamline, which had been operational since 2011, equipped with a pantographic robotic couch that significantly limits the floor space, thereby imposing limitations on the scanner's installation. The constraint on the treatment room's existing workflow was respected, as instrumentations already in place remained accessible and functional after the installation of the new system. Moreover, the development also focused on safety measures to prevent undesired hazardous situations, patient harm, and equipment damage. The imaging workflow was designed to avoid potential collisions as much as possible, and the simultaneous drive of the PPS and the robotic C-arm was inhibited, thus implementing motion only in controlled conditions. Additional hardware, such as a redundant controller and fast-response anti-collision cushions, was integrated to minimize damages in case of unpreventable collisions. The X-ray calibration procedure was experimentally verified using a geometric phantom with embedded metal rings, showing errors within the CT scan's voxel resolution. The accuracy of the positioning was confirmed through measurements conducted with a commercial optical tracking system (OTS).

The results demonstrated a sub-millimeter accuracy in reproducing the treatment plan position within the treatment room. Overall, the accuracy was found to be within the expected tolerance range, affirming the reliability of the positioning process. Follow-up measurements were performed on a radio-equivalent phantom to validate the correction pipeline's effectiveness with 3D/3D strategies, achieving residuals at sub-millimeter and sub-degree levels. A comparison with the existing custom system at CNAO revealed that the new system provided up to double the axial field of view (FOV), along with faster acquisition and shorter reconstruction times, without compromising accuracy (cfr. Table 2.1). This allowed for comprehensive beam path reconstruction. The limited scan in the complementary Half Fan (HF) mode was achieved by leveraging the "CBCT retur" motion. Considering the increased C-arm rotation speed, the total scan time for the FF modality was approximately 63 seconds. In contrast, a standard HF acquisition with a robot reset at the same speed would take around 75 seconds. This may seem counterintuitive, but traversing two arcs of 220° at the acquisition speed of 7.3°/sec, along with 40° at the reset speed of 14.6°/sec, was faster than traversing two 360° arcs, with the first during acquisition and the second for C-arm reset. Furthermore, the proposed C-arm rotation speed enabled acquisitions lasting 10 seconds and collecting 150 projections fewer than the legacy system [8]. Consequently, this resulted in a 25% reduction in the total imaging dose for CBCT without introducing any additional risk. In fact, the newly implemented collision detection solution, with its shorter response delay, ensured approximately 50% less residual C-arm travel, making the scanner safer compared to the previous custom system, despite rotating approximately 1.3 times faster.

**Table 2.1:** *Comparison between the CBCT imaging system installed at CNAO.*

| | **Legacy system (2013) [8]** | **New system (2023)** |
|---|---|---|
| **Acquisition protocols** | 2D, 3D FF | 2D, 3D FF, 3D HF |
| **3D/3D setup correction** | <1 mm, <1° | <1 mm, <1° |
| **FOV** | 204 mm | 204 mm (FF) – 403 mm (HF) |
| **Exposure time** | 40 s | 30 s (FF) – 30 s × 2 (HF) |
| **Acquisition Time** | 80 s | 63 s (FF) – 75 s (HF) |
| **Anticollision system** | Bumper (22 ms delay) | Air cushions (9 ms delay) |
| **Anode angle** | 7° | 12° |
| **Collimation** | Static | Dynamic |

## 2.2 CBCT Limitations

The current use of the described systems focuses on correcting patient positioning and evaluating anatomical changes before treatment. However, the CBCT modality, despite its faster imaging capabilities compared to CT, introduces image artifacts due to scattered radiations, leading to issues such as shading, reduced contrast, and beam-hardening. Additionally, the limited FOV causes truncation problems during reconstruction, resulting in bright-band effects and further artifacts in CBCT images. These limitations prevent accurate dose evaluation and hinder qualitative inspection, emphasizing the need for reliable techniques to correct CBCT images and enhance their clinical applications in tumor assessment, organ shift evaluation, and treatment planning.

### 2.2.1 Traditional Approaches for CBCT Correction

Aiming to address these issues, the limited FOV and truncation artifacts make these images not ideal for applying the virtual CT paradigm that aims to recalculate dose on a warped CT to match daily anatomy [9, 41–44]. One of the typical approaches to map the CT anatomy to that of the daily CBCT usually requires the preprocessing of the CT through Deformable Image Registration (DIR) [45–47]. However, DIR may introduce errors and require extensive validation, especially when the CBCT volume is flawed and has a narrow FOV [48]. In general, correction and translation methods for quantitative CBCT can be broadly divided into hardware-based, prior-CT-based, and model-based approaches. Hardware-based methods remove scatter based on additional equipment mounted on the acquisition system. In particular, an anti-scatter grid and partial beam blockers, along with software estimation of scattering, were proposed for scatter suppression [49–53]. While this kind of approach has shown promising results, it is not always feasible to be implemented in clinical practice due to the installation and setup of additional devices on acquisition systems. Moreover, this type of solution inherently reduces the system quantum efficiency and can degrade image quality [54]. Prior-CT-based methods leverage information obtained from high-resolution planning CT applied to up-to-date information contained in CBCT. They can be based on processing techniques like deformable registration of CT to CBCT [55], or by linearly scaling the CBCT HU values to CT ones through histogram matching [56, 57]. The shading artifacts inside CBCT can also be estimated by low-pass filtering the difference between the high-resolution CT projections and the raw CBCT ones and then subtracted from the CBCT during reconstruction [58]. In general, these methods achieved a limited solution to CBCT problems since they depend on the accuracy of the spatial alignment between CBCT and planning CT volume pairs from the same patients. This alignment is not always possible, particularly in the case of acquisition made on different days or in the presence of soft tissues. Model-based methods typically use Monte Carlo simulations to model the scatter contribution to CBCT projection data. Many researchers attempt to model the scatter distribution via some combination of analytical methods or Monte Carlo algorithms [59–62]. Model-based methods demonstrated to reproduce HU to sufficiently robust accuracy for clinical application because they rigorously simulate photon transport. However, their primary limitation comes from the physics models. Moreover, these techniques are limited by their execution time, usually in the order of minutes to hours, making them incompatible with an online application (e.g., dose evaluation or pre-treatment adaptation).

### 2.2.2 Deep Learning-based Approaches for CBCT Correction

Recently, advanced image processing techniques based on neural networks and deep learning were investigated, again leveraging the prior knowledge given by the pCT. Such methods, leveraging mainly Convolutional Neural Networks (CNN) and Generative Adversarial Network (GAN), were investigated to map the physical model of the x-ray interaction with matter disregarding the underlying complex analytics and avoiding the use of explicit statistical approaches such as Monte Carlo [63]. These methods leverage the advance in the field of image-to-image translation, making use of convolutional neural networks (CNNs) to learn how to map voxels from a source dis-

tribution (e.g., CBCT) to a target one (e.g., planning CT). Rather than truly correcting scattering or other physical effects, CNN learns this type of nonlinear mapping iteratively. Therefore, this approach does not need the complex physical model formulation required by, e.g., Monte Carlo-based methods. Moreover, although deep learning-based methods can be computationally expensive to train, once a well-trained model is developed, the image correction can be applied in seconds, making these methods compatible with clinical practice. The deep learning-based techniques can be broadly categorized into two main approaches: raw-data domain scatter correction and image domain shading correction. Regarding raw-data domain scatter correction, one study proposed to train a U-Net on CBCT projections corrected with prior information derived from deformed planning CT [46]. Other works focused instead on training a network by creating Monte Carlo (MC) simulated labels. In a study, the authors trained a deep residual CNN with CBCT scans and corresponding MC scatter-corrected CBCT scans [64]. In another proposal, instead, the authors created both the input and the label by synthetic CBCT generation from existing CT and adding MC-simulated scatter to the input volumes [31, 65]. This was first evaluated on various objects [65] and then on different anatomical regions [31] in a reasonable attempt to prove the generality of this method. Finally, the first study using MC-based CNN CBCT scatter correction on real data was trained on synthetic CBCT scan inputs with added MC-simulated scatter against unflawed CBCT scans, with particular attention to the scanner trajectory [66]. Concerning the image domain shading correction, slice-based training is also an emerging trend for the generation of synthetic CT with various types of CNN, ranging from U-Net trained with a supervised training approach [45, 47, 64, 67, 68] to the more complex Cycle-consistent Generative Adversarial Network (cGAN), based on an unsupervised training approach [69–74]. In particular, a work involved using DIR between planning CT and CBCT as a preprocessing step followed by a slice-based supervised training between them [45]. Focusing on CBCT-to-CT mapping, Xie et al. proposed a scatter artifact removal CNN based on a contextual loss function trained on the pelvis region of 11 subjects to correct the CBCT artifacts in the pelvic area [75]. Another research focused on a cGAN model to calibrate CBCT HU values in the pelvis region. The model was trained on 49 patients with unpaired data and tested on nine independent subjects, and the authors claimed the method kept the anatomical structure of CBCT images unchanged [71]. Exploring the use of deep residual neural networks in this field, a study demonstrated the capability of such architectures by proposing an iterative tuning-based training, where images with increasing resolutions are used at each step [76]. All these contributions, however, did not address the consistency of the treatment planning performed with the corrected CBCT. Conversely, Zhang et al. [77] reported the test of pelvis treatment planning in proton therapy performed on CBCT corrected with CNN. However, they summarized that the dose distribution calculated for traditional photon-based treatment outperformed the one computed for proton therapy. CBCT corrected with a cGAN was applied to evaluate the quality of the proton therapy planning in cancer treatment across different datasets with satisfactory results [47, 73]. All the mentioned research works focused on the problem of CBCT-to-CT HU conversion exploiting CBCT with a wide field of view (FOV). However, some systems present in clinical practice have a limited FOV, not sufficient to contain the entire volume of the patient, e.g., in the presence of large regions such as the pelvis or

with obese patients [11]. Considering the current use of CBCT for patient positioning purposes, small FOV CBCT systems could be preferred due to their reduced imaging dose, shorter computation time, and increased resolution over the treatment region of interest [78].

## 2.3 Conclusions

This chapter introduced the CBCT imaging system installed at CNAO and the main limitations of this modality. The main issue related to the CNAO imaging system is the limited FOV that causes truncation and increased scatter in the acquired CBCT. This prevents their use for dose computation and compromises the qualitative inspection of the daily anatomy, leading to less effective offline procedures and unnecessary exposure of the patient to imaging doses. The traditional approaches for correcting CBCT present some limitations. Hardware-based methods require the installation and setup of additional devices on acquisition systems. These devices involve additional costs and do not always guarantee high flexibility and generalization capability. For what concerns prior-CT-based methods, their most significant limitation is the dependence on planning acquisitions and the accuracy of spatial alignment with CBCT. This condition is not always easily obtainable, especially when the acquisition period between the two modalities increases. Monte Carlo-based methods are the most accurate approaches, but they suffer from two main limitations. Their execution time is usually incompatible with in-room clinical application, and they have the critical requirements of a complex physical model formulation. On the other hand, the deep-learning-based methods overcome some of these limitations. The following chapters will describe the definition of a deep-learning framework that processes the CBCT to calibrate the HU, remove artifacts due to the conic geometry acquisition, and handle narrow FOV issues to demonstrate the potential use of the corrected CBCT in the context of proton treatment planning updates. This framework aims to explore the possibility of using the in-house limited FOV CBCT system not only for patient positioning but also for dosimetric evaluation without any hardware modifications.

CHAPTER 3

# Comparison of Supervised and Unsupervised Approaches for the Generation of Synthetic CT from CBCT

*This chapter discusses the use of deep neural networks for CBCT correction and describes two main paradigms used for training such networks: supervised and unsupervised. The study compares the performance of these approaches in generating synthetic CT scans from CBCT images and provides guidelines for selecting the appropriate training technique. This chapter is based on the paper by Matteo Rossi and Pietro Cerveri. "Comparison of supervised and unsupervised approaches for the generation of synthetic CT from cone-beam CT". Diagnostics (Basel, Switzerland), 11, 2021.*

Over the past few years, evidence has been accumulating in support of the view that machine learning and artificial intelligence methods may sensibly impact diagnostic and therapeutic areas. In this regard, clinical tools have been recently proposed in oncology to speed up lesion analysis, improve tumor staging, support treatment planning, and, ultimately, contribute to the main clinical decisions [4, 79]. Such tools may incorporate deep neural networks for image analysis, within a complete end-to-end clinical pipeline, providing automatic image segmentation, image synthesis, and image-to-image translation [80–82]. As already mentioned in Chapter 2, deep learning-based methods joined traditional approaches to improve the quality of medical images. These approaches utilize advancements in image-to-image translation and employ convolutional neural networks to learn the mapping of voxels from a source distribution (e.g., CBCT) to a target one (e.g., pCT), without directly addressing physical effects like scattering, thus eliminating the need for complex physical modeling used in methods like Monte Carlo-based approaches; furthermore, while deep learning-based methods may require substantial computational resources for training, the application of well-trained models for image correction can be executed within seconds, making them compat-

ible with clinical use. The literature presents two main paradigms to train CNN for image-to-image translation: supervised and unsupervised training. Supervised training requires paired images from two dominions (e.g., CBCT and CT) for model training. The input CBCT is processed by the model generating a Synthetic CT (sCT), which is then compared to the corresponding ground-truth CT to minimize their pixel-by-pixel difference iteratively. Thanks to its neural encoding-decoding capabilities, the U-Net architecture is the most common one for this type of training [31, 45–47]. Similar to prior CT-based methods, images must be well-registered to achieve good performance when using supervised training. Since model training is based on a single pixel-wise loss function, the network could be biased and will tend to learn incorrect mapping in case of suboptimal registration. It is not always possible to achieve such good accuracy, especially in soft tissues, due to the patient's weight loss, tumor changes, or the presence of air. On the other hand, unsupervised training enabled the possibility to use unpaired data for image-to-image translation [83]. The most common architecture for this kind of training is cycle Generative Adversarial Network (cGAN) [84]. This architecture is based on two concurrent subnetworks, a generator and a discriminator, which work in opposition. Given two different datasets, the generator tries to learn the mapping to convert one dataset to the other. The generator aims to trick the discriminator. On the other hand, the discriminator's objective is to distinguish between real and synthetic images. This generator-discriminator cycle-consistent loop is designed to improve the generator's ability to produce synthetic images that are virtually indistinguishable from real ones. Cycle GAN was first proposed for natural image synthesis. Still, recently various researchers demonstrated its application for many medical image syntheses tasks, like the generation of synthetic CT from CBCT [69, 85–87], MR synthesis from CT images [88,89] or PET attenuation correction [90]. The main advantage of this method consists of the possibility to use unpaired, even unbalanced datasets, as the one-to-one correspondence between both dominions is no longer necessary. However, the computational power needed to train this architecture increases dramatically, since there are four models to be trained based on at least four distinct loss functions. In this chapter, generative deep neural networks were investigated for translating CBCT volumes into synthetic CT featuring compatible HU values with a focus on providing a fair comparison between supervised and unsupervised training paradigms. The same generative architecture was trained using supervised and unsupervised training techniques, exploiting a pelvic CT/CBCT publicly available dataset [91]. The comparison between the two obtained solutions was first quantified in terms of synthetic CBCT to original CT similarity by means of Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index Measure (SSIM). The HU agreement was then evaluated in terms of Mean Absolute Error (MAE). The deep-learning-based methods proposed here aim to overcome some of the limitations of traditional approaches for CBCT correction, as previously explained in Chapter 2. At first, it does not have any additional hardware requirements, making it of potential utility for every kind of CBCT acquisition system, e.g., robotic C-arm, gantry-mounted systems, or couch-mounted ones. Moreover, while training a deep convolutional neural network is a time-demanding task, they are very fast in generating output images in a production application once trained. The execution time for these methods is compatible with clinical use. Furthermore, the black-box nature of neural networks does not rely in general on any complex physical modeling.

In order to obtain a network with sufficient generalization capability, it is mandatory to train it with a sufficiently broad and generalized dataset. The first generative model was based on the U-Net architecture trained according to a traditional supervised pattern. The second generative model was still a U-Net-based model, which was coupled to a discriminator network and trained without supervision in a cGAN configuration. Overall, two main impacts qualify this work:

- the feasibility of CNN to generate accurate synthetic CT scans from CBCT images, which is fast and easy to use compared to traditional techniques applied in clinics;
- the proposal of guidelines to drive the selection of the better training technique, which can be shifted to a more general image-to-image translation.

## 3.1 Materials and Methods

### 3.1.1 Dataset Description

The publicly available dataset, called Pelvic Reference Dataset (PRD) [91], used in this work was obtained from the Cancer Imaging Archive. PRD included paired CT and CBCT volumes of the pelvic region of 58 subjects, spanning 28 males and 30 females. CBCT images were acquired at least 1 week after the corresponding CT volume acquisition. Moreover, PRD includes 47 and 11 volumes acquired in supine and prone positions, respectively. Two subjects (both female, one prone, and one supine) were removed from the dataset. In the first case, a metal hip prosthesis made it impossible to distinguish the anatomical structures in both CT and CBCT. In the second case, the CBCT field of view was too small to contain the entire subject's pelvic region. Therefore, a massive part of the subject present in the CT was missing in its corresponding CBCT. These two cases were considered outliers and removed from the dataset, so that 56 cases were available for this study. Both CT and CBCT volumes featured size of $512 \times 512$ pixel on the axial plane with a pixel size of $1.00 \times 1.00$ mm and a single slice thickness of 3.00 mm. The number of slices of each CBCT was 88, while in CT scans that number was variable. Along with volume files, the PRD dataset also provided a table with the $(x, y, z)$ coordinates to apply to the CBCT isocenter to align with their corresponding CT. To sum up, 4053 CT/CBCT 2D axial projections pairs were available. The entire dataset was split into train, validation, and test sets with a ratio of 80, 10, 10%, respectively, corresponding to 3243, 405, and 405 images per set.

### 3.1.2 Image Preprocessing

Preprocessing steps were performed before feeding data to the neural network. At first, a binary mask was created to separate the subject from any non-anatomical content (e.g., treatment couch). This masking procedure avoids any negative impact from these structures on training procedures. In order to prepare these masks, Otsu thresholding was applied to each volume. A max connected component analysis was then performed on each mask to remove any non-anatomical residual finely. A final binary fill holes filtering and erosion filtering were applied to enhance mask accuracy further. Finally, the mask was applied to the original volume, obtaining an isolated anatomical region.

Corresponding masked CT and CBCT were then rigidly registered, using isocenter coordinates provided with PRD Dataset. This step was necessary only for supervised training since the unsupervised one did not need paired data. However, good alignment was also required for evaluation analysis. The HU range of the grayscale values was first clipped to $[-1024, 3200]$ and then rescaled to $[0, 1]$ with a linear mapping. In order to reduce the computational cost for CNN, every axial slice was resampled to $256 \times 256$ pixel.

### 3.1.3 Deep Convolutional Neural Network Models

The two generative models were both based on the U-Net architecture trained according to two different training patterns. The supervised generative model required only this network, while the unsupervised one also required the definition of a discriminative architecture. In order to make the comparison possible, the generator architecture was precisely the same for both supervised and unsupervised procedures. The generator was implemented as an adapted version of the popular U-Net architecture. The basic U-Net is primarily used to solve pixel-by-pixel classification problems in image segmentation [92]. In the present work, U-Net was adapted to solve an image-to-image conversion task to generate sCT images from CBCT. The basic building blocks for the generator were depicted in Fig. 3.1. The first one, called the ConvBlock, was based on a modified version of the implementation of Isola et al. [83]. It was composed of a 2D convolution with a $3 \times 3$ kernel, followed by an instance normalization layer and a swish activation function. Instance normalization was demonstrated to improve the performance in image generation tasks [93]. ConvBlock was present in the network both standalone and as a part of another processing block, called the InceptionBlock. This latter block was adapted from GoogLeNet [94]. It comprised four parallel ConvBlocks, each with an increasing kernel size of dimensions $1 \times 1$, $5 \times 5$, $7 \times 7$, and $11 \times 11$. In this way, the input received by the InceptionBlock was processed simultaneously by multiple receptive fields. The output of each branch was then concatenated, and the entire stack of feature maps was returned as output. The main purpose of this processing block was to perform multi-scale feature extraction from the input image. These multi-scale extracted features, ranging from small to large receptive fields, can provide better results for image synthesis.

The overall generator structure, depicted in Fig. 3.2, was composed of a contracting and an expanding path, both based on the basic processing blocks. The first two upper processing blocks of the generator were composed of InceptionBlocks, while the deeper three were composed of ConvBlocks. In this way, the network was divided into two parts with two different functions: the inception part extracted global contextual information while the traditional part has the task to capture finer context and precise localization. The discriminator (Fig. 3.3) was a CNN that performs image classification. Its architecture was based on the PatchGAN architecture [83], considered to be the gold standard discriminator for cGAN [69, 84, 87]. It consisted of four consecutive ConvBlocks, with a $4 \times 4$ kernel size. The convolution for the first three ConvBlock was set with stride 2, giving as outputs a tensor with half the size and twice the features map. The last ConvBlock had stride one and maintained the size and the number of feature maps unchanged. A sigmoid activation function followed the last layer, producing a $32 \times 32$ map with every pixel in the $[0, 1]$ range. This output map was used for

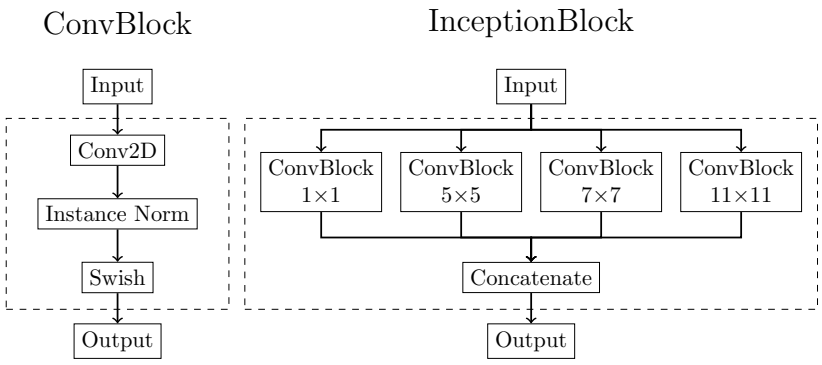


**Figure 3.1:** *Basic building block architectures. The ConvBlock (**left**) comprises a 2D convolutional step with kernel variable kernel size, followed by instance normalization and a Swish activation function. The InceptionBlock (**right**) is composed of the parallel combination of more ConvBblock with kernel dimension of* $1 \times 1$, $5 \times 5$, $7 \times 7$, *and* $11 \times 11$. *The output of each ConvBlock is then concatenated in a single output tensor.*

patch-wise classification of the input image as real or fake.

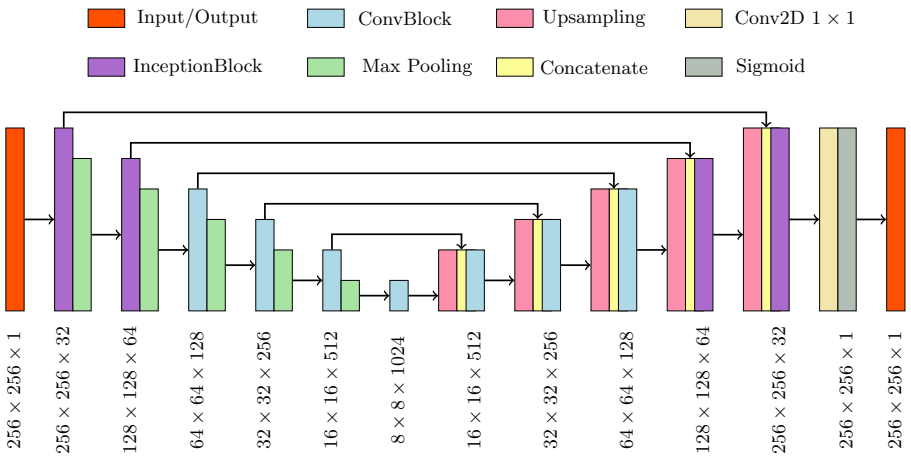


**Figure 3.2:** *Schematic of the generator network architecture.*

### 3.1.4 Training Methods

The two training routines are schematized in Fig. 3.4. For supervised training, given the original CBCT image as input to generator CT ($G_{CT}$) network, the generated sCT is compared to the corresponding ground-truth CT. During training, the network adapts its weight according to MAE loss function, expressed as:

$$\mathcal{L}_{supervised}(CT, sCT) = \frac{1}{N}\sum_{i=0}^{N} |CT_i - sCT_i| \tag{3.1}$$

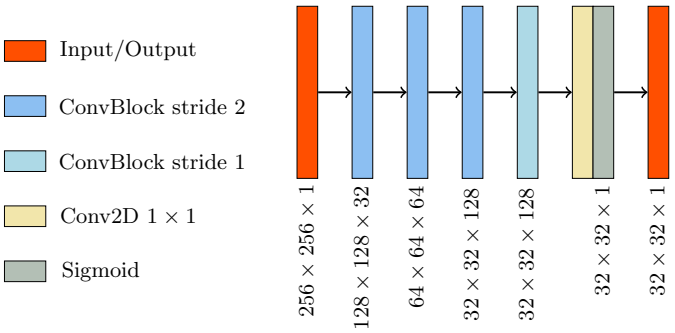


**Figure 3.3:** *Schematic of the discriminator network architecture.*

In regard to the unsupervised learning, the cGAN structure incorporated two generators and two discriminators, competing against one another, namely generator CT ($G_{CT}$), generator CBCT ($G_{CBCT}$), discriminator CT ($D_{CT}$), and discriminator CBCT ($D_{CBCT}$). $G_{CT}$ was used to generate sCT from CBCT, while $G_{CBCT}$ was used to generate Synthetic CBCT (sCBCT) from CT. On the other hand, $D_{CT}$ was used to distinguish between real CT and sCT, and $D_{CBCT}$ is used to distinguish real CBCT from sCBCT. In the first step of the training, $G_{CT}$ ($G_{CBCT}$) took CBCT (CT) as input and generated sCT (sCBCT). Then, $G_{CT}$ ($G_{CBCT}$) took sCT (sCBCT) as input and generated a cycleCBCT (cycleCT), which is supposed to be equal to the original CBCT (CT). Meantime, $D_{CT}$ ($D_{CBCT}$) tried to discriminate between real CT (CBCT), labeled as 1, and sCT (sCBCT), labeled as 0. Generator loss functions included three types of terms: adversarial loss, cycle consistency loss, and identity loss. Discriminator loss was composed only of an adversarial term. These loss functions were combined to mapping the distribution of the generated images to the distribution domain of the target images (see Section 3.1.5).

### 3.1.5 Loss Functions for Unsupervised Training

In the unsupervised training, the generator loss functions included three types of terms: adversarial loss, cycle consistency loss, and identity loss. Discriminator loss was composed only of an adversarial term. For $G_{CT}$ the adversarial loss term is the mean squared error (MSE) between 1 and $D_{CT}(G_{CT}(\text{CBCT}))$:

$$\mathcal{L}_{adv-G_{CT}}(CBCT) = \frac{1}{N}\sum_{i=0}^{N}(1 - D_{CT}(G_{CT}(CBCT_i)))^2 \tag{3.2}$$

Symmetrically, the adversarial objective for $G_{CBCT}$ is:

$$\mathcal{L}_{adv-G_{CBCT}}(CT) = \frac{1}{N}\sum_{i=0}^{N}(1 - D_{CBCT}(G_{CBCT}(CT_i)))^2 \tag{3.3}$$

The second term is represented by cycle consistency loss. It is used during training to enforce consistency of the two mappings from CT to sCBCT and CBCT to sCT [84].

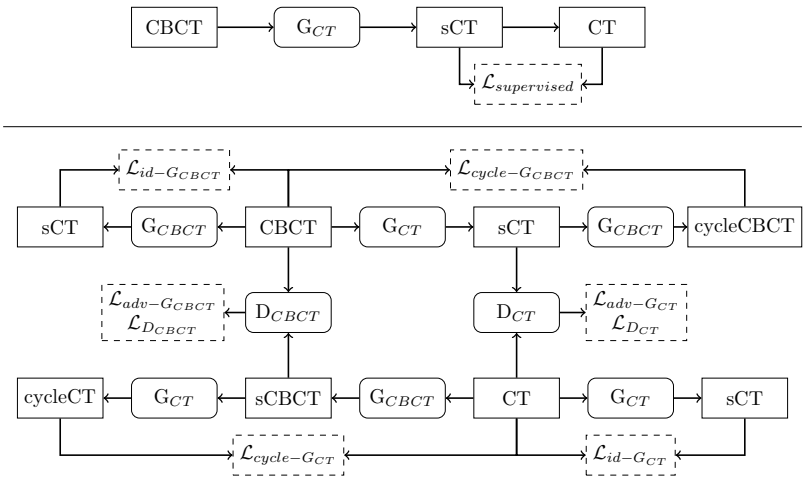


**Figure 3.4:** *Schematic flow chart of supervised (**top**) and unsupervised (**bottom**) training routines. Rectangle boxes represent images, and rounded corner boxes depict neural network models. Dashed boxes indicate a loss function. For a detailed explanation of the training routines, refer to the Section 3.1.5.*

This term is represented by the mean absolute error (MAE) between real CT (CBCT) and cycleCT (cycleCBCT), respectively:

$$\mathcal{L}_{cycle-G_{CT}}(CT) = \frac{1}{N}\sum_{i=0}^{N}|G_{CT}(G_{CBCT}(CT_i)) - CT_i| \tag{3.4}$$

$$\mathcal{L}_{cycle-G_{CBCT}}(CBCT) = \frac{1}{N}\sum_{i=0}^{N}|G_{CBCT}(G_{CT}(CBCT_i)) - CBCT_i| \tag{3.5}$$

The third term, identity mapping loss, was introduced in order to preserve HU values between real CT and sCT and between real CBCT and sCBCT. Without this term, generator $G_{CT}$ and $G_{CBCT}$ are free to change the HU values of input images even when it is not necessary. This loss acts as an identity constraint for the generator. When sCT is used as input for $G_{CT}$, it should generate CT too as output. The same applies for $G_{CBCT}$ with CBCT. Therefore, identity loss is expressed as MAE between $G_{CT}$ (CT) with respect to CT and between $G_{CBCT}$ (CBCT) against real CBCT:

$$\mathcal{L}_{id-G_{CT}}(CT) = \frac{1}{N}\sum_{i=0}^{N}|G_{CT}(CT_i) - CT_i| \tag{3.6}$$

$$\mathcal{L}_{id-G_{CBCT}}(CBCT) = \frac{1}{N}\sum_{i=0}^{N}|G_{CBCT}(CBCT_i) - CBCT_i| \tag{3.7}$$

Combining all these terms, for the overall loss of the two generators the following formula holds:

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cycle}\cdot\mathcal{L}_{cycle} + \lambda_{id}\cdot\mathcal{L}_{id} \tag{3.8}$$

where $\lambda_{cycle} = 10$ and $\lambda_{cycle} = 5$ represents the relative importance of the adversarial, cycle consistency, and identity mapping part of the objective.

The adversarial loss of the discriminator, which constrained the network to classify real images, labeled as 1, and synthetic images, labeled as 0, cab be written as:

$$\mathcal{L}_{D_{CT}}(CT, CBCT) = \frac{1}{N}\sum_{i=0}^{N}\frac{(1 - D_{CT}(CT_i) + D_{CT}(G_{CT}(CBCT_i)))^2}{2} \tag{3.9}$$

$$\mathcal{L}_{D_{CBCT}}(CT, CBCT) = \frac{1}{N}\sum_{i=0}^{N}\frac{(1 - D_{CBCT}(CBCT_i) + D_{CBCT}(G_{CBCT}(CT_i)))^2}{2} \tag{3.10}$$

### 3.1.6 Performance Metrics

In order to quantitatively evaluate network performances, three widely accepted numerical metrics were used: peak signal-to-noise ratio (PSNR), structural similarity index measure (SSIM), and mean absolute error (MAE). PSNR is computed as the ratio between the maximum possible power of a signal and the mean square error of the images being compared. Its value is measured in decibel and approaches infinity as the mean squared error between sCT and ground-truth CT approaches zero. Therefore, a higher PSNR value corresponded to higher image quality and vice-versa [95]. SSIM was first introduced in 2004 by Wang et al. [96]. It measures similarity between two images based on three factors: luminance, contrast, and structure. Compared to PSNR, SSIM provides an indication of similarity closest to that of the human visual system. The value for this metric ranges between 0 and 1, where 1 indicates the best possible level. Lastly, MAE is used to evaluate the HU accuracy between two images quantitatively. Before computing its value, the amplitude of the network output, i.e., [0,1], was scaled back to the original range, i.e., $[-1024, 3200]$. The lower the MAE, the higher the HU accuracy of the two images. Every metrics was evaluated using CT as the ground truth reference.

### 3.1.7 Cross-Validation Analysis

In order to validate the performance of the trained models, a four-fold cross-validation experiment was carried out on both methodologies. The entire dataset was divided into four subsets, each subgroup consisting of images of 14 patients. For each experiment, three subsets were used as the training set, while the remaining one was used as the test set. The cross-validation allowed a comparison of the performance of the models against the baseline results, i.e., the difference between the CT and the original CBCT images. The improvement in terms of SSIM, PSNR, and MAE was quantitatively analyzed. The statistical difference between the candidate models was also evaluated with Kruskal-Wallis non-parametric test for median differences ($p < 0.01$) and Tukey–Kramer post-hoc comparison.

### 3.1.8 Implementation Details

The Keras [97] and TensorFlow [98] Python frameworks were used to develop the network models, loss functions, metrics, and training routines. The training was carried

out in the Google Colaboratory Cuda-enabled environment, equipped with a four-core CPU, 25 GB RAM, and NVIDIA® Tesla® P100 GPU support 16 GB RAM. The number of epochs has been set to 25. When the validation set SSIM score is maximized, the training routine was configured to save the set of the best network weights. The training was optimized with ADAM (Adaptive Moment Estimation) optimizer [99], with the following parameters: learning rate $2 \times 10^{-4}$, exponential decay rate for the first moment estimates $\beta 1 = 0.5$, and exponential decay rate for the second-moment estimates $\beta 2 = 0.999$. The batch size was set at 10. In order to generalize network performance as much as possible and prevent overfitting, data augmentation was performed at run-time during training. At the beginning of each training epoch, just before giving the image as input to the network, a series of image elaborations were applied. In particular, the images were randomly rotated by multiple of 90 degrees and horizontally flipped. The same transformations were also applied to the corresponding ground truth CT in order to maintain coherence between pairs. By applying these transformations, each CBCT/CT pair featured eight different configurations (four rotations times two flips). This implied that at each iteration the network was fed with different images. For example, assuming $N$ images in the training dataset, at each iteration step $N$ images are generated from the original samples by applying the described random transformations.

## 3.2 Results

Supervised training required about 3 minutes per epoch. The only architecture used for this training pattern was the generator, corresponding to 2,554,977 trainable parameters. On the other hand, unsupervised training required about 5 minutes per epoch due to its increased complexity. This method featured two identical generators and two identical discriminators. Being the discriminator network characterized by 429,185 trainable parameters, the overall number of weights for the unsupervised model was 5,968,324. For what concern inference time, the two approaches were comparable. Generating sCT required less than 4 seconds for an entire CBCT volume (~70 slices) in both cases, computed on the same GPU environment used for training.

### 3.2.1 Performance Metrics

The cross-validation experiments confirmed that the proposed CNN models improved the metrics with respect to the baseline (Fig. 3.5), being median and interquartile results summarized in Table 3.1. In general, supervised training attained better results for every metric. Considering SSIM, supervised sCT obtained an improvement of 2.5% with respect to the baseline ($p < 0.0001$), while unsupervised sCT gained an improvement of 1.1%. PSNR also confirmed this trend, resulting in an enhancement from the baseline for supervised sCT and unsupervised sCT of 4.19 dB ($p < 0.0001$) and 2.3 dB ($p < 0.0001$), respectively. The relative gains for PSNR were 15.7% and 8.6%. Lastly, the supervised model reduced the mean absolute error between supervised sCT and CT by 58.16 HU (62.3%), while the unsupervised model resulted in a reduction of 46.92 HU (50.3%). Even in this case, the supervised model appeared significantly better than the unsupervised one ($p < 0.0001$).

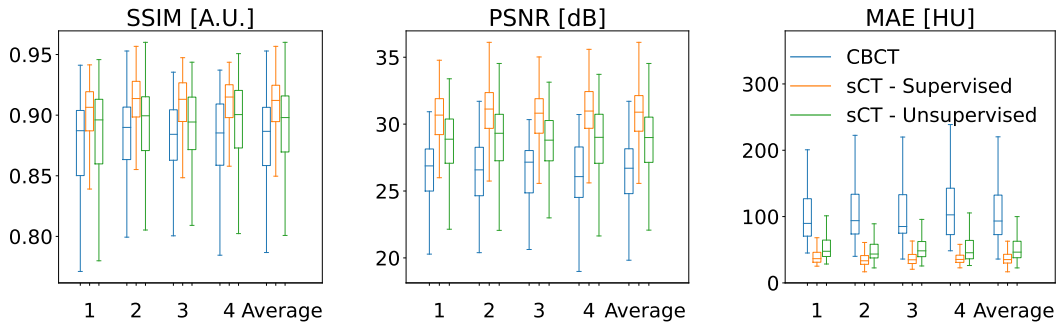


**Figure 3.5:** *Quantitative analysis of SSIM, PSNR, and SSIM values between original CBCT, sCT supervised, and sCT unsupervised against the corresponding CT, computed for each fold of the four-fold cross-validation.*

**Table 3.1:** *Performance metrics evaluated on original CBCT, sCT supervised, and sCT unsupervised. Every value is computed against the original CT, considered as the ground truth. Every value is expressed as median (interquartile range).*

| | **SSIM [A.U.]** | **PSNR [dB]** | **MAE [HU]** |
|---|---|---|---|
| CBCT | 0.887(0.048) | 26.70(3.36) | 93.30(59.60) |
| Supervised sCT | 0.912(0.030) | 30.89(2.66) | 35.14(13.19) |
| Unsupervised sCT | 0.898(0.046) | 29.00(3.38) | 46.38(24.86) |

### 3.2.2 Qualitative Comparison

Some examples of sCT generated by the supervised and unsupervised approaches are depicted in Fig. 3.6. The first column represents the ground truth CT image, while the second one contains the original CBCT input. The third and fourth columns show the generated sCT images predicted from supervised and unsupervised models, respectively. The last column compares one intensity profile (the central row depicted with a line) for every imaging modality. The same example slices are also showed as difference maps in Fig. 3.7. Every column represents the difference between the modality under evaluation (CBCT, supervised sCT, and unsupervised sCT) and the ground truth CT. The first row represents the case of a CBCT/CT pair with a similar field of view (FOV). From the intensity profile and the difference map, it can be noticed that both models enhanced overall image qualities in terms of HU mapping, with slightly better visual details for the supervised model. The second row shows an example in which CT FOV is wider than CBCT one. In this case, it can be observed that the supervised model tried to compensate for the missing structures at the border with unrealistic values. Contours of supervised sCT were blurred and unreliable, while unsupervised sCT coped better with the original CBCT boundaries. In more detail, the unsupervised model did not attempt to recreate the missing FOV, resulting in more realistic and reliable contours with respect to the original CBCT. Concerning the internal anatomical structures, both models reached good qualitative results, as demonstrated by the corresponding intensity profile and difference map. The last row in Fig. 3.6 and 3.7 shows a case in which the rectum area had an air-filled cavity with different shapes between the CT and the CBCT. The supervised model reacted by filling the air cavity with the

surrounding values, while the unsupervised model preserved its contours. This latter behavior is preferable, as it is assumed that the information content of the CBCT is more updated than that of the CT. Again, the supervised model attempted to scale the HU values without considering possible changes in the anatomical structures of interest.

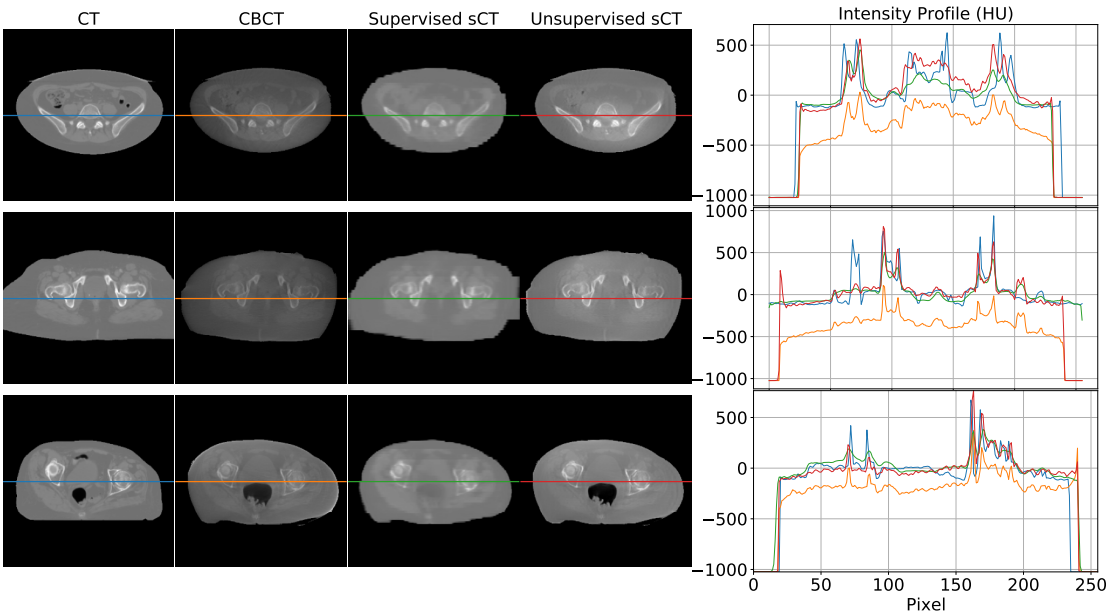


**Figure 3.6:** *Visual comparison between some axial slices depicted in every modality: CT, CBCT, sCT supervised, and sCT unsupervised. Every row corresponds to a different example taken from a different subject. The rightmost part of the figure compares the intensity profiles of the central line of the images, highlighted by the central line in the four representations. Images are displayed with Window = 2000, Level = 0.*

## 3.3 Discussion

### 3.3.1 Main Findings

This work proposed a deep-learning-based approach for generating synthetic CT images from CBCT scans, featuring HU values compatible with the traditional CT domain, providing a fair comparison between supervised and unsupervised training paradigms. As demonstrated by the cross-validation analysis, the supervised model outperformed the unsupervised ones. The supervised network guaranteed advantages in terms of computational cost, having 57% fewer weights to train and requiring 40% less training time for the same number of epochs. Moreover, it gained an increase of 12% in terms of MAE with respect to the unsupervised one, showing a superior ability in mapping HU values. On the other hand, the supervised model showed to be not always reliable, resulting in some unexpected behavior, especially in the contour regions. This kind of artifact did not appear in images generated through the unsupervised model, which proved to be more reliable in preserving the anatomical structure and the up-to-date information contained in the CBCT scan. One example of this is clearly visible in the bottom row of Fig. 3.6, where the shape of the air cavity present in CBCT is

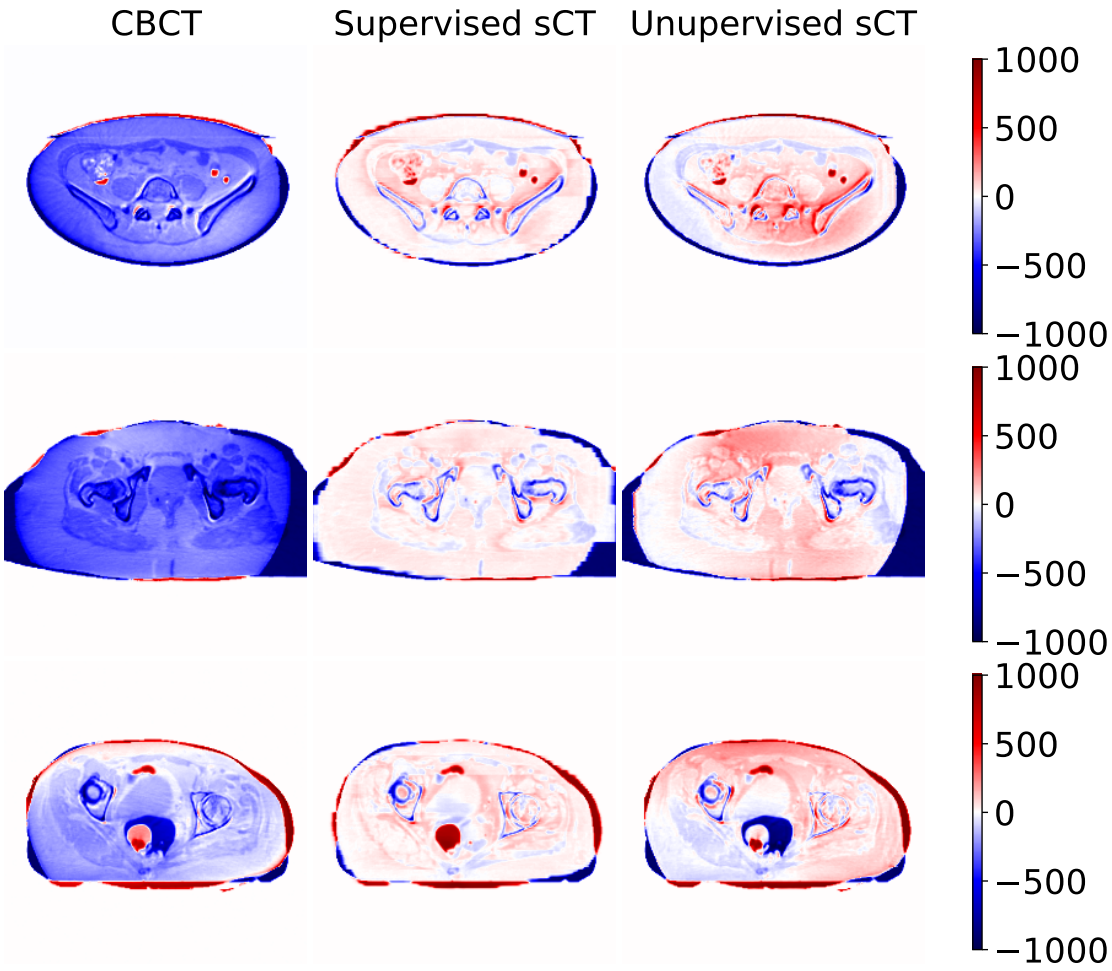


**Figure 3.7:** *Difference maps computed between original CBCT, sCT supervised, and sCT unsupervised using their corresponding CT as reference. Numeric values are in Hounsfield Units. Each row represents a different example taken from a different subject. The examples depicted in this figure correspond to the example presented in Fig. 3.6.*

preserved only for the unsupervised case, while the others tended to close the air gap. This kind of artifact could be risky as it can generate some biases in evaluating the current evolution of the patient's internal anatomy, or a wrong treatment plan evaluation in case of dose recalculation. On the other hand, the unsupervised model did not suffer from this condition due to the cyclic loss functions, consisting of three main terms, each with a different purpose. The more straightforward loss function defined for the supervised model, i.e., mean square error, specialized the network to learn mapping without adequately considering the anatomical structure.

### 3.3.2 Comparison with the Literature

To our best knowledge, just one comparison between supervised and unsupervised training of deep networks, addressing back and forth translation between CT and magnetic resonance images was proposed in the recent literature [89]. In such a work, the authors explored different network architectures proposing a modified U-Net (supervised training) and the ResNet model for the generator (unsupervised training) pro-

posed in [84]. Conversely, in this work, the same model was adopted in both the supervised and unsupervised generative parts, which smoothed out any biases in the comparison. Approaches competitive with the present work were reported in the recent literature. Using an MC-based methodology, HU correction of about 31% for five lung cancer patient images was gained [33]. Another proposal, which used histogram matching on ten prostate cancer patient scans, resulted in a 20% HU correction [57]. A phantom-based study was carried out on the Catphan 600 (The Phantom Laboratory, Salem, NY) [58] achieved an overall accuracy of 95% in HU recovery. Nonetheless, results on patient images were not presented. Concerning deep-learning-based methods, a proposal based on the U-Net trained with a supervised approach on 30 prostate cancer patients presented an HU accuracy enhancement of about 69% [46]. A comparative study explored the performances of some supervised models trained on 42 prostate cancer patients. The results obtained by the three tested models were 55, 17, 47% in terms of HU relative improvement [47]. Exploring the use of cGAN and unsupervised training, a study disclosed an enhancement of 57% with a dataset containing 120 head and neck scans [85]. Another unsupervised-based research resulted in a 16% increase for 33 prostate cancer patients [86]. The results proposed in this work are generally in line with the prior art, with an HU improvement of 62% and 50% for the supervised and unsupervised models, respectively.

### 3.3.3 Technical Challenges and Work Limitations

The use of complex models for remapping HU can be considered oversized when looking at the intensity profiles (cfr. Fig. 3.6), where linear shifts appear predominant. This would suggest that regression models may be enough to address HU mapping. However, the inter-variability between the two modalities (e.g., air gaps, morphological variations due to postures, lesion progression) may easily introduce non-linearity in the intensity profile mapping (cfr. Fig. 3.8). In the upper panel of Fig. 3.8, an air gap can be appreciated in the CBCT and in the corresponding CT. The notch in the intensity profiles is just produced by such a gap. This kind of deformation is difficult to correct by simple linear mapping. In the lower panel, relevant beam hardening of the CBCT at the image boundaries, corresponding to peripheral density appearance, is traceable in the corresponding intensity profile (orange curve) as an undue convexity at the borders. Likewise, a linear transform of the profile does not ensure accurate correction. Consequently, convolutional neural networks represent a natural approach to address as they feature input/output non-linear transforms and good generalization capabilities.

As shown, quantifying the image translation effectiveness by means of SSIM and PSNR can be inconclusive. For example, both lower SSIM and PSNR values (supervised against unsupervised) do not necessarily indicate that the images are qualitatively worse (cfr. Table 3.1). In particular, SSIM is a perception-based metric considered to be correlated with the quality perception of the human visual system, and the higher its value is, the better is the perception of similarity for human eyes [96]. Still, some studies have shown that SSIM is less effective when used to evaluate blurry and noisy images [100, 101]. This reduces the relevance of comparisons performed using such metrics. Conversely, MAE provided an objective quantification of the HU matching. As far as the selection between the two approaches is concerned, it can be argued that the supervised approach is preferable in the presence of a large amount of paired data,

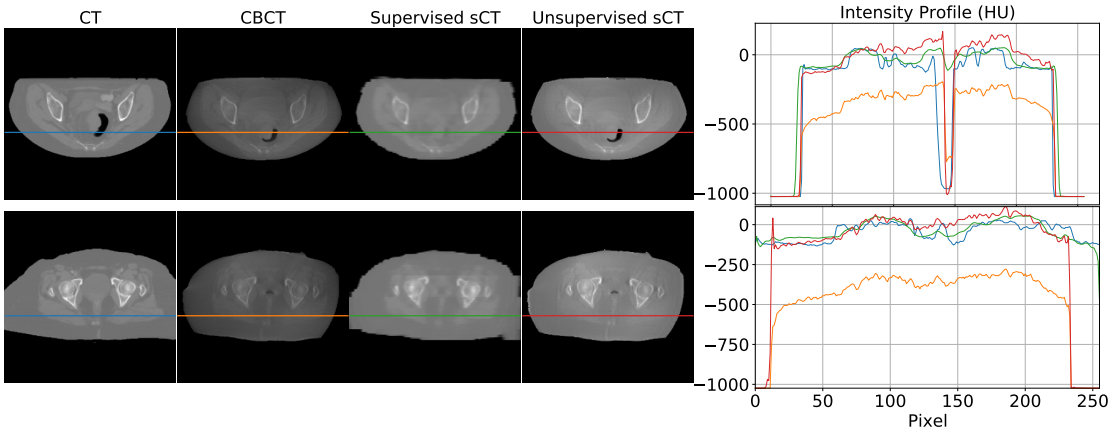


**Figure 3.8:** *Examples of non-linearity of CBCT scans. Air gap introduced a notch in the intensity profile (**top**). Beam hardening is visible as a convex shape in the orange curve of both examples (**bottom**). Images are displayed with Window = 2000, Level = 0.*

ensuring high accuracy in HU mapping. Nonetheless, gaining this advantage requires that CBCT and CT images be gathered as close as possible to minimize the anatomical difference. It is also crucial to mask images according to the smallest FOV between the two modalities in order to have comparable information content in between. In detail, masking ensures similar FOVs preventing the supervised network from attempting to reconstruct the sCT in the CBCT regions where this information is missing. In order to reduce the dependency on such constraints, the unsupervised method is suitable, avoiding pairing, requiring fewer preprocessing steps (e.g., masking) and being more conservative in terms of anatomical features. Removing the constraint on the pairing, the training effectiveness can benefit from a large amount of data, easier to collect. Nonetheless, even though obtaining unpaired images is clinically simpler, using explicitly paired data in training reduces geometric discrepancies at body boundaries. As verified in [54], unsupervised training may take advantage of paired data, allowing a better focus on soft-tissue contrast than geometric mismatches. To further validate the methods presented in this work, a dosimetric analysis would be required. However, the chosen dataset did not allow this type of study to be explored due to the lack of the original CT dose distribution plan. Moreover, the presented methodologies focused only on a single anatomic district, the pelvic one. We wanted to demonstrate the feasibility of both methods in a body region particularly subjected to inter- and intra-fractional changes. The comparison between the two training strategies was made to show their different approach to addressing image-to-image translation tasks, highlighting their strengths and limits.

## 3.4 Conclusions

The proposed methodologies allow for generating accurate synthetic CT images by automatically processing CBCT scans, featuring processing time faster than traditional Monte Carlo-based methods. The basic idea behind this approach is to correct CBCT images exploiting their potential not only in a more accurate patient positioning, but also for clinical evaluation purposes such as tumor shrinkage and organ shift. For what

concerns the integration in the clinical workflow of treatment re-planning, this method can be integrated into a tool to be deployed that elaborates the reconstructed CBCT images, acting as an image filtering process. The correction of CBCT also makes possible the direct morphological comparison with the planning CT. The following chapters will explore some techniques to further increase the ability of this framework to correct CBCT scans.

CHAPTER 4

# Limited FOV CBCT Correction with Deep Convolutional Neural Networks and Transfer Learning

*This chapter further explores the use of deep learning-based methods for CBCT correction, introducing a novel technique to address the limitations of these methods based on a two-step algorithm leveraging the transfer learning paradigm. This chapter is based on the paper by Matteo Rossi et al. "Image-based shading correction for narrow-FOV truncated pelvic CBCT with deep convolutional neural networks and transfer learning." Medical physics, 48:7112-7126, 2021*

Following what we described in Chapter 3, the supervised model outperformed the unsupervised ones in terms of both quantitative image quality metrics and computational cost. However, the supervised-trained model exhibited occasional unreliability, leading to unexpected behavior, particularly in contour regions. In contrast, the unsupervised model consistently avoided such artifacts and demonstrated greater reliability in preserving the anatomical structure and up-to-date information present in the CBCT scan. The more straightforward loss function defined for the supervised model compared to the unsupervised one, i.e., mean square error, specialized the network to learn mapping without adequately considering the anatomical structure. This chapter presents some techniques to address the limitations of deep learning-based CBCT correction. In particular, we describe a method for CBCT axial slice processing by using the U-Net architecture and leveraging the transfer learning paradigm. It has been shown that pre-training a network with synthetic data could be an effective initialization technique for many complex models, providing better performance when the network is then fine-tuned with real data [102, 103]. The clinical dataset used for this work was retrospectively available at the CNAO facility, described in Section 4.1.2. The available data was augmented through a transfer learning approach by generating a synthetic

CBCT dataset from a publicly available pelvic CT repository [91], completely avoiding using DIR. In synthesis, the aim of the present work is to provide a U-Net-based image domain shading correction for the recovery of CT-compatible HU values from narrow FOV CBCT scans. The main novelty aspect of the present work can be found in the neural network supervised training strategy, which was developed according to the transfer learning paradigm, following an innovative two-step approach described in Section 4.1.5. This allowed splitting the learning of the anatomical features from the learning of CBCT/CT shading differences. The performances obtained by the fine-tuned U-Net were then compared with a network trained only with the retrospective clinical dataset.

## 4.1 Materials and Methods

### 4.1.1 Imaging instrumentation

In this work, we exploit images directly acquired with the legacy CBCT imaging system installed at CNAO (cfr. Chapter 2) [8]. As a reminder, CBCT is used at CNAO for evaluating anatomopathological variations between the planning CT and the daily anatomy. The device was primarily intended for rigid setup correction and is not fitted with a moving flat panel or an adjustable collimator. Therefore, it cannot produce a sufficiently broad FOV for a complete anatomical description of larger districts (i.e., pelvis and thorax). Missing information about the periphery of a large anatomical district does hinder adaptive approaches based on such images. Clinical practice at CNAO reduces inter/intra fractional motion with thermoplastic fixation masks [23]. However, particle therapy is more susceptible to air cavities with respect to photon radiotherapy. Here a clinical instrument such as the limited FOV CBCT is used to qualitatively evaluate air in the bowels (e.g. when dealing with the pelvis). Moreover, the limited FOV also causes truncation, especially in the case of large anatomical districts such as the pelvis. Consequently, bright-band effects appear on the border of the FOV along with the non-uniqueness of the interior problem, as stated by Clackdoyle et al. [37] The issue of truncation artifact correction is attenuated by image processing techniques, such as the Ohnesorge filter [25], which excessively lowers the image's grayscale intensity. As such, strategies to correct shading due to different components of artifacts are required to improve the qualitative clinical procedure.

### 4.1.2 Datasets Description

For the realization of this work, two datasets were used. The first dataset, denominated $D_r$, included 18 retrospective pairs of CT/CBCT pelvis acquisitions obtained from six oncological patients (3 Male/3 Female). Clinical Target Volume (CTV) segmentation was also recovered for each patient. These data were acquired at the CNAO facility during daily pelvic district treatments. The CT scans were acquired with a Siemens Somatom Sensation Open Bore scanner at 100 kV, while the CBCT scans were acquired with a Varian A-277 X-ray tube and a Varian PaxScan 4030D flat-panel detector at 100 kV and 25 mAs, featuring a FOV of 204 mm [8] and processed by a 30% truncation correction with Ohnesorge filter [25]. In order to reduce morphological variations between corresponding CT and CBCT acquisitions, the shortest time interval determined

the pairing. In 13 cases, the pairs were acquired on the same day, while four were acquired one day apart. In a single case, the acquisitions were made two days apart. The CT scans are temporally close to the CBCT ones as they are re-evaluation CT scans. Given that CNAO employs a thermoplastic mask for patient immobilization and motion reduction [23], residual deformation between the corresponding scan was deemed negligible. Following preprocessing (highlighted in Section 4.1.3), the dataset contained 3368 CT/CBCT 2D axial slice pairs aligned and rigidly registered between the two modalities. The second dataset, termed Pelvic-Reference-Data [91], was obtained from the Cancer Imaging Archive. In this dataset, 58 pelvic CT scans, acquired in prone and supine positions, were available. Because of large organ deformation due to a support cushion placed under the pelvic area, some scans were removed from the useful set. For the remaining 46 patients (27 Male/19 Female), 8289 CT axial projections were available. The synthetic CBCT images, whose generation was described in Section 4.1.2, joined to the corresponding CT scans, were denominated $D_s$. Both datasets were divided into training, validation, and test sets with a $60/20/20\%$ ratio. The number of images in every subset was 2021/673/674 for $D_r$ and 6632/828/829 for $D_s$.

#### Synthetic Data Generation

Synthetic CBCT scans were generated from the corresponding CT scans of dataset $D_s$ using the OpenREGGUI open-source platform, written in Matlab and based on the RTK API (the Reconstruction ToolKit [104]). RTK expects intensity values to be in a $0-2^{16}$ range, with an open field signal ($I_0$) being on the right-bound and entirely blocked signal ($I_{dark}$) on the interval's left-bound. Projections are then Log-Transformed to attenuation. In order to obtain a correct projection dataset from a CT, the following formula was applied to its grayscale values:

$$CT_{\mu} = \frac{(CT_{HU} + 1000)}{2^{16}} \tag{4.1}$$

The generation of synthetic CBCT was done following the same geometry of CNAO treatment room: source-to-detector distance equal to 1672 mm, source-to-isocenter distance equal to 1172 mm, and gantry sweep of 220 degrees. A set of 500 simulated Cone Beam projections was derived, using the RTK CUDA-based forward projector [104] with a panel size of 1024x768 pixels (spacing $0.388 \times 0.388$ mm) for each considered CT. A simulation of scattering and beam hardening was then applied to the projections using the methods provided in OpenREGGUI [105]. Finally, a CBCT axial scan was reconstructed from the projection stack, using the same projective geometry defined before, resulting in a truncation of around 30% of the patient body and attenuated with the Ohnesorge filter [25]. Imaging parameters were chosen to match the ones of dataset $D_r$, with output CBCT volumes of dimensions $220 \times 220 \times 220$ pixels and spacing of $1 \times 1 \times 1$ mm. The scatter, beam hardening, and Gaussian noise factors were found empirically and set to 0.001, 1.005, and 0.001, respectively. Conversely to real CBCT, the generated synthetic CBCT resulted perfectly aligned to the corresponding CT, avoiding anatomical deformations and air pockets differences (Fig. 4.1).

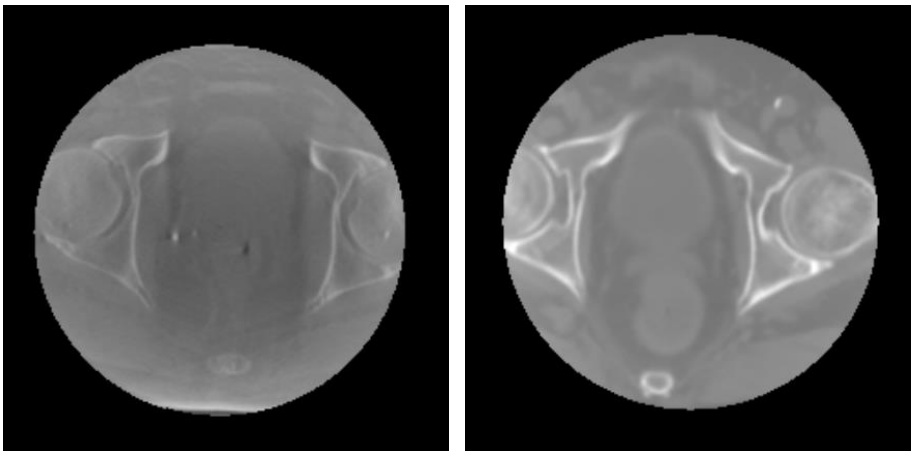

**Figure 4.1:** *Axial view of a $D_r$ CBCT (**left**) and a $D_s$ CBCT (**right**).*

### 4.1.3 Image Preprocessing

Preprocessing steps were performed before feeding data to the neural network. First, rigid registration was applied, using the ITK API (the Insight ToolKit [106]), only to $D_r$, as $D_s$ was intrinsically already aligned. A masking procedure was then applied to the CT scans of both $D_r$ and $D_s$ datasets to extract the FOV equal to the corresponding CBCT. Every CT volume was resampled to have isotropic voxels of $1 \times 1 \times 1$ mm matching CBCT voxel dimensions. The HU range of the grayscale values was first clipped to $[-1000, 3100]$ and then rescaled to $[0, 1]$ with a linear mapping. Finally, as a technical convenience in the U-Net processing, the axial slices were zero-padded to $256 \times 256$ pixels from the original size of $220 \times 220$ pixels.

### 4.1.4 Deep Convolutional Neural Network Model

The basic U-Net architecture is mainly used for image segmentation [92], solving a pixel-by-pixel classification problem. In the present work, it was adapted to solve an image-to-image translation problem to address the task of cupping removal and HU recovery from the original CBCT. In agreement with the U-Net architecture, the proposed neural network was mainly composed of a contracting and an expanding path (Fig. 4.2). In the contracting path, each processing layer (block) was constituted by two 2D convolutions, with kernel dimension $3 \times 3$, no stride, and Rectified Linear Unit (ReLU) activation function. A Batch Normalization layer followed every convolution. Between two convolutions of the same block, a Dropout layer randomly "switched off" the updating for 10% of the weights to prevent overfitting. Each block in the contracting path was connected to the next with a Max Pooling layer, giving as input for the next block a tensor with twice the feature maps and halving its size. The purpose of this contracting path was to capture the context of the input image. The expanding path had the same structure as the contracting path but with a Transpose Convolution layer instead of Max Pooling, giving an up-sampling effect to the network. Thus, each block in the expanding path outputs a tensor with half the feature maps and twice its size. Every block in this path was also connected to its corresponding block in the other paths with a connection layer. The purpose of the expanding path was to enable precise localization combined with contextual information from the contracting path. A single

feature map 2D convolution was applied in the last layer, resulting in a single image with the same dimension as the input. The last layer used a sigmoid activation function to provide values in the $[0, 1]$ range for every pixel.

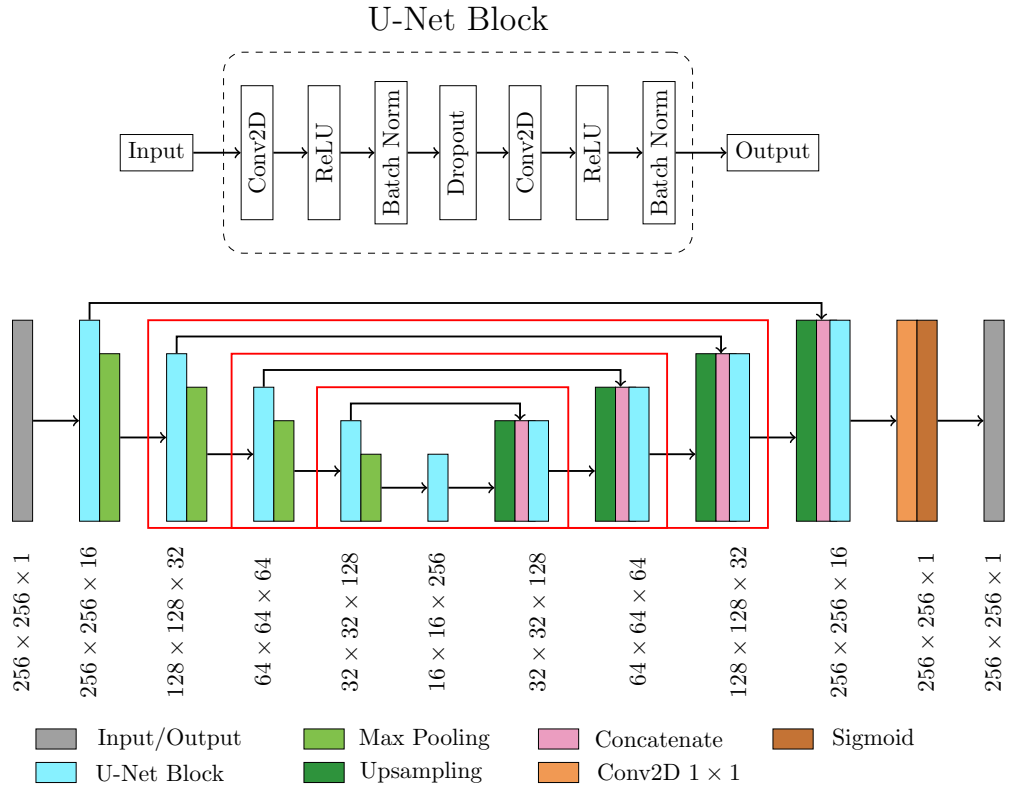


**Figure 4.2:** *Schematic of the symmetric contracting and expanding paths of the U-Net. Every U-Net processing block is composed of two convolutional, Relu, and batch normalization layers, with a dropout layer in the middle. Every block in the contracting path is followed by a Max Pooling layer, while every block in the expanding path is followed by a transpose convolution one. The last convolution is followed by a sigmoid layer. The red boxes indicate the blocks that can be retrained during each Transfer Learning experiment.*

### 4.1.5 Training of the Models

Training a U-Net requires an abundant quantity of data. In the interest of reducing potential overfitting [107], data augmentation was performed at runtime during the network training. A series of random operations were applied to the existing data (the input CBCT and its corresponding CT ground truth) so that the network was never fed twice with the same image during training. At first, random cropping of the image was performed, feeding the network with a $128 \times 128$ sub-patch of the original image. These sub-patches were also subjected to random rotations of multiples of 90 degrees and horizontal flips. Moreover, input (CBCT) and label (CT) must represent the same anatomical condition. Otherwise, the network will be forced to compensate for residual deformation. The common practice [45] requires the use of a deformable image registration algorithm to match CT to CBCT daily anatomy as a preprocessing step. This step was avoided on the basis of $D_r$ characteristics (as described in Section 4.1.2). The

synthetic $D_s$ CBCT can be perfectly superimposed on the reference CT from which they were generated, completely avoiding DIR in data preparation 4.1.2.

The models analyzed in this work were trained following two different training patterns (cfr. Fig. 4.3). The first, named the noFT model, was trained in a single step using only data from $D_r$. The second, named the $FT_x$ model, was trained using both $D_s$ and $D_r$ datasets, following a two-step approach similar to that of Gherardini et al. [108] At first, end-to-end training was performed using the synthetic data from $D_s$ on a newly created model (the Synth model). In the second step, some deeper processing blocks (one, two, or three) were fine-tuned using $D_r$, leaving the rest of the model weights fixed. Each retrained block was considered symmetrically in the contracting and expanding paths (cfr. Fig. 4.2). Depending on the number of retrained processing blocks, the model takes the name $FT_1$, $FT_2$, $FT_3$. Synthetic data are suitable for augmenting the dataset overall dimension and learning the anatomical-related features in the image. Transfer learning embeds these features in the network, giving a good initialization of the network weights. Therefore, subsequent tuning on real data is supposed to be less sensitive to residual deformation between input and ground truth labels.

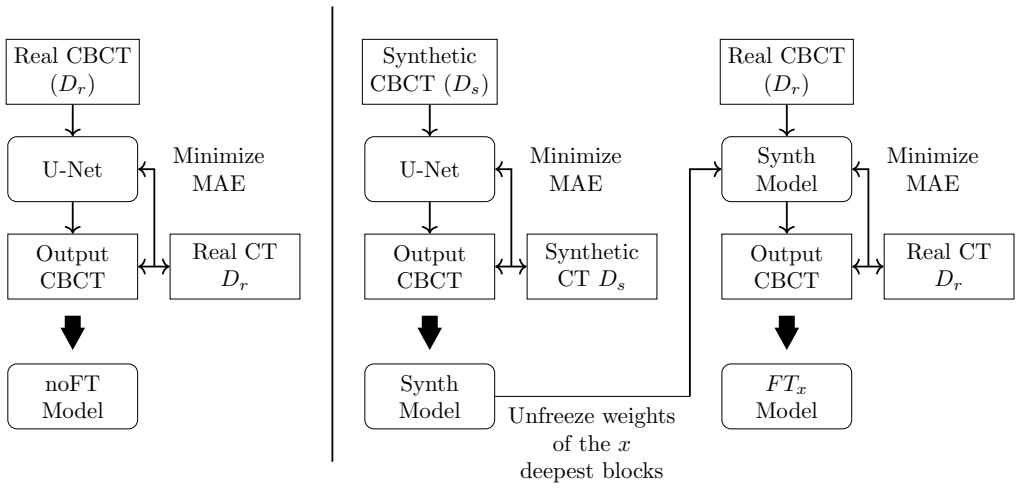


**Figure 4.3:** *Training pattern for the noFT (**left**) and $FT_x$ (**right**) models. The noFT model is trained in a single step using only data from $D_r$. The $FT_x$ model is trained in two steps. In the first one, a model is trained using only $D_s$ (Synth model), then only $x$ (1, 2 or 3) processing blocks were retrained with $D_r$ data.*

#### Loss Function and Performance Metrics

Mean absolute error (MAE) has been set as a loss function, and the training was optimized with ADAM (Adaptive Moment Estimation) [99]. ADAM has the advantage of being designed to compute individual adaptive learning rates for different parameters starting from estimates of first and second moments of the gradients. ADAM optimizer was set with the following parameters: learning rate 0.001, exponential decay rate for the first moment estimates $\beta 1 = 0.9$, and exponential decay rate for the second-moment estimates $\beta 2 = 0.999$, while the learning rate was 0.001. To evaluate the network performance, two widely used metrics were chosen, namely the peak signal-to-noise ratio

(PSNR) and the structural similarity index measure (SSIM). PSNR value approaches infinity as the mean squared error between improved CBCT and ground truth CT approaches zero. Therefore, a higher PSNR value corresponds to higher image quality and vice-versa [95]. SSIM is considered to be more correlated with the quality perception of the human visual system. The higher its value is, the better is the perception of similarity for human eyes [96].

#### Hyperparameter Tuning Experiments

In order to find the optimal U-Net architecture according to PSNR and SSIM, some experiments were conducted using exclusively $D_r$ data. These experiments aimed to find the best combination for the number of processing blocks (four or five) and the number of feature maps at the first level (16, 32, or 64), systematically varying these parameters. A deeper network with many feature maps can virtually learn more from the input data itself, but it also requires computational power accordingly. The correct trade-off between these parameters has to be found. All considered architectures were trained using dataset $D_r$ and performance results were computed in terms of median and Inter-quartile Range (IQR) ranges. The statistical difference between the candidate models was evaluated with Kruskal-Wallis non-parametric test for median differences ($p$<0.01) and posthoc comparison. If two or more models did not present significant differences, the choice fell on the lighter architecture in terms of computational cost.

#### Transfer Learning Experiments

Once the best architecture was chosen according to hyperparameter tuning experiments, some transfer learning experiments were conducted to find the optimal number of processing blocks to be retrained (one, two, or three) to increase PSNR and SSIM. For example, considering the architecture 16-32-64-128-256-128-64-32-16 (four processing blocks and sixteen feature maps in the first block), the fine-tuning of only one block meant that the retraining allowed the refinement of the weights in the 128-256-128 level. Conversely, the fine-tuning of two blocks meant that the retraining allowed the refinement of the weights in the 64-128-256-128-64 processing blocks. Even in this case, the appropriate number of blocks was chosen by evaluating the statistical differences between experimental results and the computational cost.

#### Network Implementation

The network model, loss function, metrics, and training routine were built using the Keras [97] and TensorFlow [98] frameworks in Python. The training was carried out in a Google Colaboratory Cuda-enabled environment, equipped with a 4-core CPU, 25 GB RAM, and NVIDIA® Tesla® P100 GPU support 16 GB RAM. The training routine was set to save the best weights values when the validation set SSIM score is maximized.

### 4.1.6 Cross-Validation Analysis

In order to make the network output comparable to the original volumes, the amplitude $[0, 1]$ of the network output was scaled back to HU units $[-1000, 3100]$. To reinforce

the interpretation of the results, a Leave-One-Out Cross-Validation (LOO-CV) experiment was performed. The Synth model was fine-tuned with 18 subsets of dataset $D_r$, each one excluding a single stack of slices, corresponding to unique CBCT/CT pair $P_x$. A comparison between baseline performance (Base), noFT, and $FT_x$ models was presented to show the validity of the transfer learning approach. Two analyses were conducted to assess: i) the improvement for PSNR and SSIM; ii) the reliability of HU recovery.

#### Performance Metrics

Given the small size of $D_r$ compared to $D_s$, the NoFT model risks overfitting training data. In order to validate the coherency of the performance in terms of PSNR and SSIM with respect to the parameters defined in the previous experiments, LOO-CV was used to compare the noFT and $FT_2$ model with the baseline (Base) performances. The term Base was used to name the performance computed on the dataset $D_r$ as is after rigid alignment and resampling, without any neural network elaboration. The expected result is that the $FT_x$-type models are more consistent than noFT ones. To verify this assumption, the PSNR and SSIM metrics were computed again on each fold.

#### HU Analysis and Shading Evaluation

As far as HU recovery is concerned, the difference in terms of HU between the CBCT network output versus the corresponding CT images was computed. In order to increase the consistency of the comparison, mismatching air pockets were automatically removed according to predefined HU thresholds [4] (Fig. 4.4). Additionally, an analysis based on Region of Interest (ROI) was proposed to evaluate the network improvement for different tissues. $8 \times 8 \times 8$ cubes were segmented in spongy bones, fat, muscle, and bladder (Fig. 4.5). According to the represented tissue, the HU values contained in each cube were averaged and compared between CT and CBCT. In addition, the Contrast-to-Noise Ratio (CNR) was calculated to evaluate the improvement in terms of contrast enhancement. In particular, the CNR was measured for each imaging modality by comparing the clinical target volume with the bladder, muscle, and fat ROIs, respectively. Then, the same ROIs were compared with the air region within each scan. Particular attention was given to the relation between truncation severity due to variable patients pelvic size and HU non-conformity.

## 4.2 Results

### 4.2.1 Neural Network Assessment

#### Hyperparameter Tuning Experiments

In order to find the best network architecture in terms of PSNR and SSIM, some hyperparameter tuning experiments were conducted (cfr. Table 4.1). In terms of PSNR, all the four-block architectures provided results significantly better than the five-block ones ($p$<0.0001). Considering four as the number of processing blocks, no statistical difference among 16 and 32 feature maps was found ($p$=0.46), as well as between 16 and 64 feature maps ($p$=0.74). Therefore, sixteen was chosen as it significantly reduces the number of trainable parameters, being the best trade-off between performance and

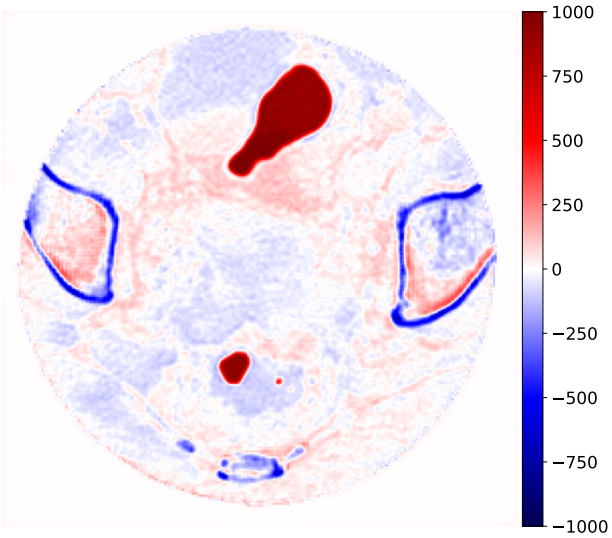


**Figure 4.4:** *Example of air pockets, visible as a red blob. Since these regions mismatch between the two images, corresponding voxels are not considered for HU difference computation.*

network complexity. The SSIM analysis further confirmed this choice. The selected U-Net configuration was 16-32-64-128-256-128-64-32-16, corresponding to 919 177 trainable parameters.

**Table 4.1:** *PSNR and SSIM performances (median and interquartile range) of the hyperparameter tuning experiments, depending on the number of processing blocks and the number of convolutional filters in the first block. Each value is computed by evaluating the dataset $D_r$ test set. The final choice for these parameters for both noFT and $FT_x$ models is highlighted.*

| Blocks # | First block filter # | PSNR (dB) | SSIM (A.U.) |
|---|---|---|---|
| **4** | **16** | **31.943 (3.261)** | **0.926 (0.030)** |
| 4 | 32 | 32.156 (2.913) | 0.918 (0.025) |
| 4 | 64 | 32.314 (2.661) | 0.925 (0.028) |
| 5 | 16 | 31.704 (2.674) | 0.928 (0.029) |
| 5 | 32 | 31.608 (2.503) | 0.918 (0.023) |
| 5 | 64 | 30.367 (2.274) | 0.926 (0.027) |

#### Transfer Learning Experiments

Considering the transfer learning experiments results (cfr. Table 4.2), retraining the two deepest blocks proved to be the best value for increasing network performance. This was confirmed by the statistical analysis of the $FT_1$, $FT_2$, and $FT_3$, considering PSNR, SSIM, and MAE. For PSNR, $FT_2$ architecture provided results significantly better than the other two $FT_1$ ($p$<0.0001) and $FT_3$ ($p$<0.0001) architectures. Considering SSIM, no statistical difference among $FT_1$ and $FT_3$ ($p$=0.78) was found, while the $FT_2$ results are significantly different from $FT_1$ ($p$=0.007) and $FT_3$ ($p$=0.0007). In addition, the three distributions were compared with the distribution obtained from the Synth model. In general, each $FT_x$ model had significantly better performance compared to the Synth model, both in terms of PSNR ($p$<0.0001), SSIM ($p$<0.0001), and MAE ($p$<0.0001). As such, Synth was not considered for further evaluation.

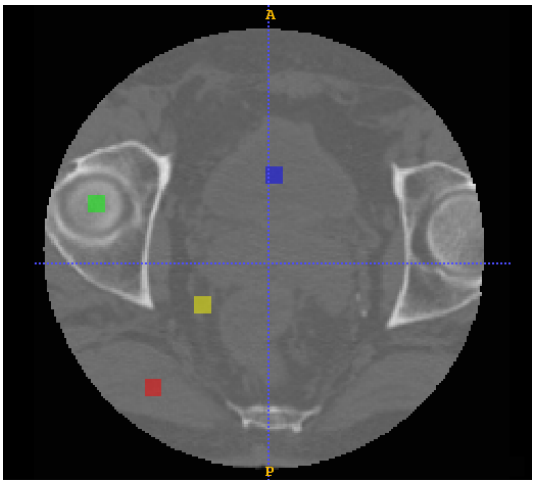


**Figure 4.5:** *Example of cubic ROIs ($8 \times 8 \times 8$ mm) extracted from a patient. Selected regions are bladder (blue), spongy bones (green), muscle (red) and fat (yellow).*

**Table 4.2:** *PSNR, SSIM, and MAE performances (median and interquartile range) of the transfer learning experiments, depending on the number of processing blocks to be retrained. Each value is computed by evaluating the dataset $D_r$ test set. The final choice for the $FT_x$ model ($FT_2$) is highlighted.*

| FT blocks # | PSNR (dB) | SSIM (A.U.) | MAE (HU) |
|---|---|---|---|
| 0 (Synth) | 26.707 (1.440) | 0.907 (0.018) | 128.184 (28.011) |
| 1 | 29.007 (2.151) | 0.918 (0.024) | 84.989 (23.187) |
| **2** | **30.799 (2.167)** | **0.921 (0.023)** | **63.067 (15.179)** |
| 3 | 29.904 (2.268) | 0.917 (0.025) | 73.187 (19.726) |

### 4.2.2 Cross-Validation Analysis

#### Performance Metrics

The LOO-CV experiments confirmed that the proposed neural network processing improved the quantification metrics with respect to baseline (Fig. 4.6). On most of the patients, the $FT_2$ model performed better than the Base and noFT in terms of PSNR and SSIM. In particular, the median PSNR for Base, noFT, and $FT_2$ were on average, over 18-fold, 26.77, 31.83 and 32.32 dB, respectively. Considering PSNR, $FT_2$ model results to be significantly better than Base ($p$<0.0001) and noFT ($p$<0.0001). The median SSIM for Base, noFT and $FT_2$ were on average 0.902, 0.915 and 0.916. Even in this case, $FT_2$ model results to be significantly better than Base ($p$<0.0001) and noFT ($p$=0.0005). Furthermore, by aggregating the training histories of the 18 LOO-CV experiments for both models and comparing the MAE for the training and validation sets, the $FT_2$ model proved to be faster in terms of convergence speed (cfr. Fig. 4.7). This is due to the $FT_2$ network weights initialization provided by the pre-training with synthetic data.

#### HU Analysis and Shading Evaluation

The ROI evaluation produced, on average, HU improvements of 177.59 for bladder, 192.2 for bone, 215.87 for muscle and 123.43 for fat using the noFT method. By

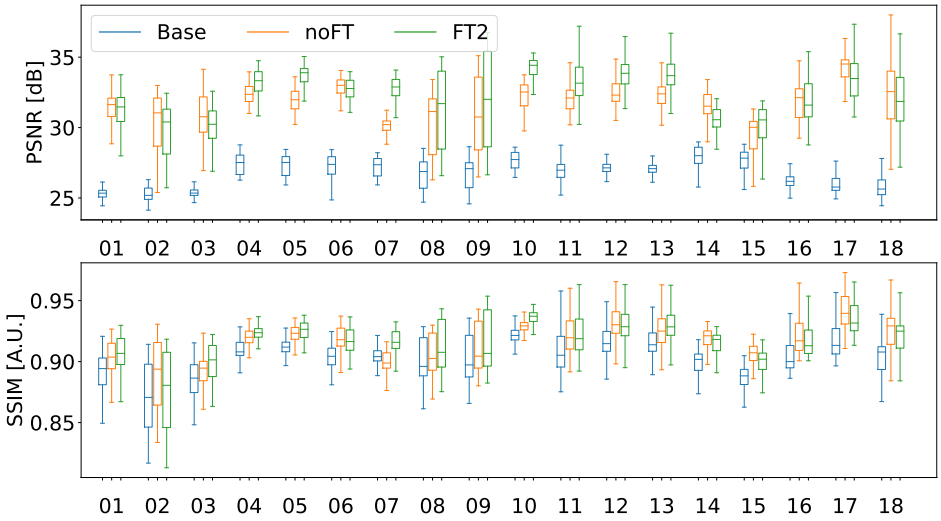


**Figure 4.6:** *Quantitative analysis of PSNR and SSIM values between every CBCT (Base, noFT, $FT_2$), computed for each fold of the Leave-One-Out Cross-Validation.*

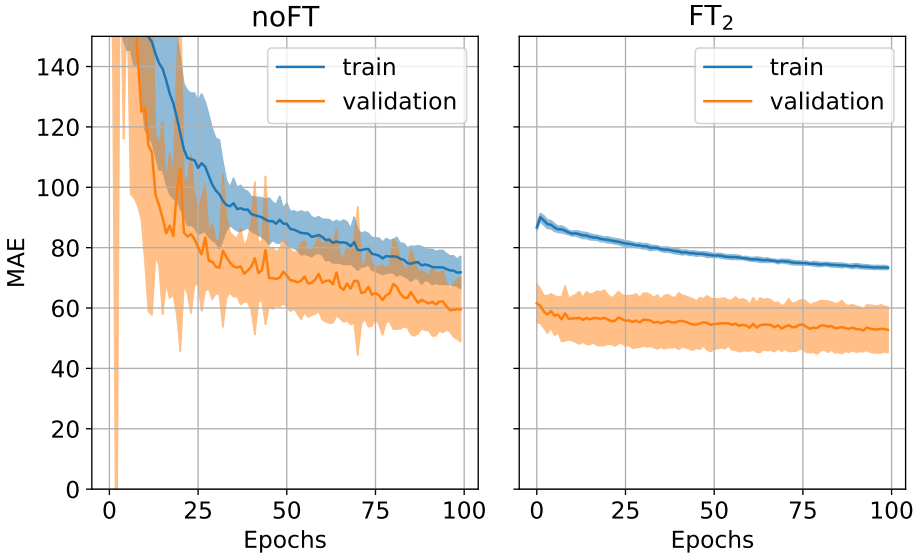


**Figure 4.7:** *Mean absolute error history for noFT (**left**) and $FT_2$ (**right**) models during training. Bold lines represent the mean values computed between each trained network in LOO-CV experiments, while the shaded region represents their standard deviation.*

contrast, the $FT_2$ method yielded an average HU improvement of 191.95 for bladder, 257.04 for bone, 206.88 for muscle and 169.73 for fat. For the whole images (Fig. 4.8), the $FT_2$ method obtained a median improvement of 111.96, while the noFT yielded 100.52. For each ROI, the comparison between noFT and $FT_2$ HU difference distribution reported a significative difference ($p$<0.0001). All values were computed across the 18-fold and are summarized in Table 4.3.

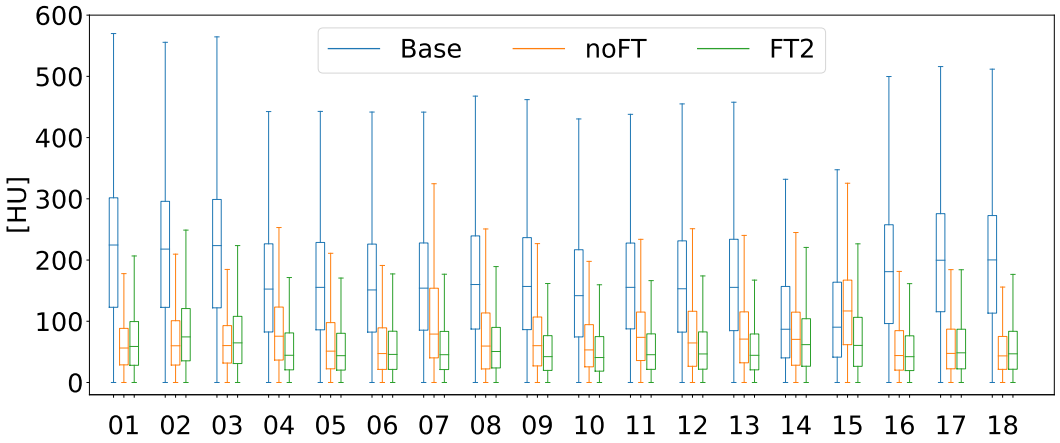


**Figure 4.8:** *Quantitative analysis of the absolute HU difference between every CBCT (Base, noFT, $FT_2$) and the corresponding ground truth CT, computed for each fold of the Leave-One-Out Cross-Validation. Both models reduce the difference in the HU ranges, with better performance for the $FT_2$ model.*

**Table 4.3:** *Absolute HU difference between every CBCT (Base, noFT, $FT_2$) and the corresponding ground truth CT, for each ROI (mean and standard deviation) and overall volumes (median and interquartile range). Values are obtained averaging between each fold of the Leave-One-Out Cross-Validation.*

| Model | Bladder | Bone | Muscle | Fat | All |
|---|---|---|---|---|---|
| Base | 213.06±47.19 | 340.79±66.06 | 249.36±20.53 | 199.27±39.46 | 161.37(162.54) |
| noFT | 35.47±25.27 | 148.59±74.25 | 33.49±28.78 | 75.84±28.02 | 60.85(80.70) |
| $FT_2$ | 23.11±20.87 | 83.75±55.41 | 42.48±28.83 | 29.54±19.40 | 49.41(66.70) |

The cupping artifact, due to scattering, clearly perceived as a darker central region in $CBCT_{Base}$, was reduced in both $CBCT_{noFT}$ and $CBCT_{FT_2}$ (Fig. 4.9). The analysis of one intensity profile (the central row depicted with a line in Fig. 4.9) confirmed that the network processing flattened the slight concavity present in $CBCT_{Base}$. The intensity profile of $CBCT_{noFT}$ appeared closer to the CT one than the corresponding profile obtained from $CBCT_{FT_2}$. At the same time, $CBCT_{noFT}$ profiles appear smoother than the CT one and its axial view is blurred. This evidence supports the quantitative results for HU in ROIs and overall image, with $FT_2$ performing better than noFT. It follows that the $FT_2$ and the noFT method both shift the range of values of CBCT closer to that of CT. The $FT_2$ model favored anatomical consistency, correcting cupping, and rescaling the intensity values. Conversely, the noFT model aggressively fits the CBCT to the reference CT, introducing blurring.

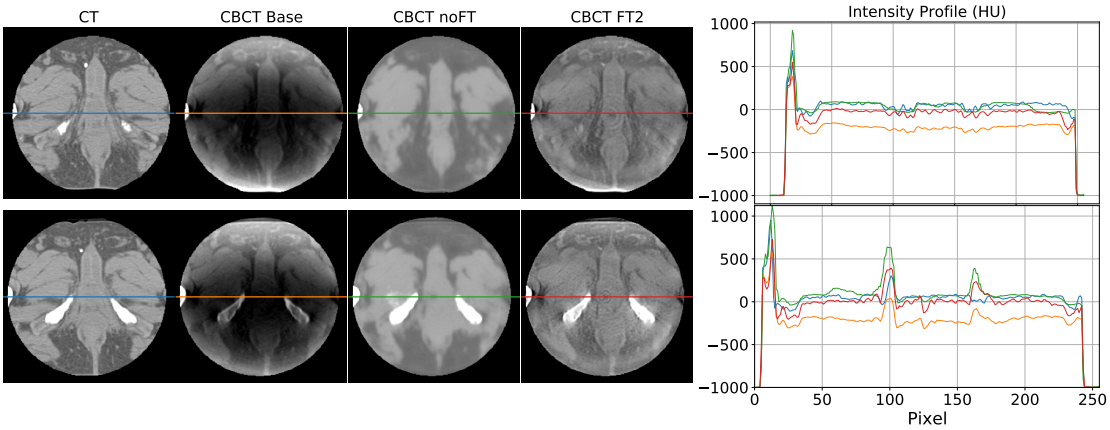


**Figure 4.9:** *Comparison between a single CBCT Base and corresponding CT axial slice with the CBCT elaborated by noFT and $FT_2$ models. The rightmost part of the figure compares the intensity profiles of the central line of the images, highlighted by the central line in the four representations. Images are displayed with Window=400, Level=20.*

The results of the tissue contrast analyses are summarized in Table 4.4. In all cases, the $FT_2$ model obtained results closest to the CT ones, considered the ground truth. These results confirmed the superior ability of this model in improving soft tissue visibility with respect to $CBCT_{Base}$ or $CBCT_{noFT}$. In particular, the comparison between CTV and soft tissues reported an improvement of about 24% for the noFT model and 67% for the $FT_2$ one, with respect to the $CBCT_{Base}$ CNR values of $CBCT_{Base}$. Concerning the comparison between soft tissues and air, the average gain was about 4% and 35% for noFT and $FT_2$, respectively.

**Table 4.4:** *CNR values for every imaging modalities (CT, $CBCT_{Base}$, $CBCT_{noFT}$, $CBCT_{FT_2}$). Values are computed between clinical target volume against every soft tissue ROI (bladder, muscle, fat). Every ROI is also evaluated against air values present in the scan. Each value is represented as median (IQR).*

| ROI | | CT | $CBCT_{Base}$ | $CBCT_{noFT}$ | $CBCT_{FT_2}$ |
|---|---|---|---|---|---|
| **Foreground** | **Background** | | | | |
| CTV | Bladder | 0.86 (1.00) | -5.24 (2.28) | -1.73 (2.35) | -1.19 (3.31) |
| | Muscle | -1.75 (1.76) | -5.12 (2.01) | -5.68 (6.27) | -1.66 (2.31) |
| | Fat | 4.68 (1.13) | 1.11 (2.31) | 2.19 (3.19) | 2.49 (2.08) |
| Bladder | Air | 42.06 (15.23) | 7.79 (2.93) | 8.66 (3.25) | 19.47 (14.05) |
| Muscle | | 43.59 (15.76) | 7.69 (2.51) | 8.86 (2.63) | 19.28 (13.49) |
| Fat | | 37.57 (13.29) | 6.82 (3.29) | 8.23 (3.02) | 18.69 (12.80) |
| CTV | | 42.48 (18.26) | 7.08 (2.90) | 8.70 (2.96) | 19.42 (13.45) |

Results regarding the relationship between MAE and pelvis width for considered cases are reported in Fig. 4.10.

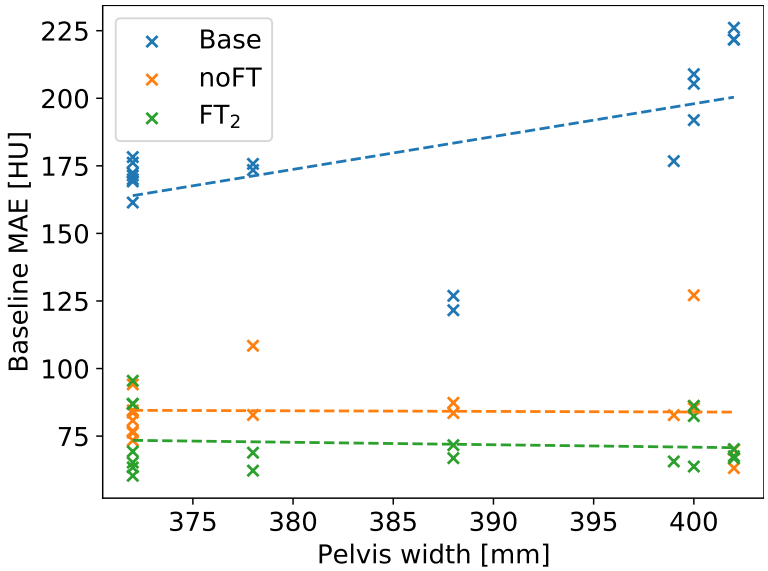


**Figure 4.10:** *Pelvis width versus MAE for every type of CBCT (Base, noFT, $FT_2$). MAE is calculated with respect to the CT ground truth.*

## 4.3 Discussion

### 4.3.1 Main Findings

In this work, we propose a U-Net-based approach to address limitations intrinsic to the limited FOV CBCT at CNAO, and to provide CBCT HU recovery, subsequently. In particular, the framework proposed focused on shading correction and soft tissue contrast enhancement. Firstly, the network improved the image intensity distribution, quantified by PSNR and SSIM, in the range of about 5 dB (noFT), 6 dB ($FT_2$), and 2% (both), respectively. These highlight global contrast and signal-to-noise ratio improvement. The relative improvement of CNR for the CTV versus various soft tissues was on average 24% (noFT) and 67% ($FT_2$). Coherently, the average CNR improvement for soft tissue with respect to the air in the CBCT scans was 4% (noFT) and 35% ($FT_2$). Secondly, the network was able to provide HU grayscale values comparable to the ground truth CT, reducing the non-linear cupping artifact and scaling the intensity values. In addition, MAE results for noFT and especially for $FT_2$ models improved both in median (62.29% and 69.38% respectively) and narrower IQR (50.35% and 58.96%, respectively) terms. The latter indicates a compensation and generalization capability of the two approaches with respect to different pelvis widths and, therefore, truncation, as shown by Fig. 4.10. Finally, both networks performed the required task. Performances were similar, but the HU comparison for ROI and the overall image showed a performance edge of the $FT_2$ model. Consequently, preconditioning the network with synthetic data proved to be an effective method for the problems addressed in this work. The two-step training allowed to split the learning of the anatomical features from the learning of CBCT/CT shading differences.

### 4.3.2 Comparison with the Literature

Existing literature using U-Net architecture aims at correcting CBCT with sufficient FOV [45–47, 64, 65] and generally does not encompass truncation artifacts, except for one work exploiting Monte Carlo simulations on synthetic data [31]. Our work evaluates a similar CNN architecture tailored to CNAO needs. The aforementioned works validate their approach by means of dosimetric accuracy. The proposed methods are tested on the overall geometrical improvement in the image Hounsfield Unit. In order to further evaluate possible clinical applications, a dosimetric study will be eventually paired with the HU analysis hereby presented. However, achieving dose calculation on narrow-FOV CBCT scans is currently out of the scope of this work. Comparing the HU difference between CT and CBCT with existing methods leveraging CT information, while most of these works do not address truncated data, we can assess the proposed network performance. For instance, using an MC-based methodology, Thing et al. reported HU correction of about 31% for five lung cancer patient images [33]. Another method based on histogram matching on ten prostate cancer patient images provided HU correction of about 20% [57]. Phantom studies have reached up to a 95% overall accuracy in the HU recovery on the Catphan 600 [58] alone, but no data was provided for patients. In comparison, our $FT_2$ and noFT models respectively obtained around 69.38% and 62.29% average improvement on the whole image in the pelvic region. Addressing the HU correction on different tissues, the improvement for $FT_2$ (noFT) can be described as 75% (56%) in the spongy bone (femoral head), 89% (83%) in bladder, 85% (62%) for fat and 83% (86%) for muscle. In particular, Kida et al. [45] investigated a U-Net approach similar to noFT with wide untruncated FOV, where authors reported improvements of 95% and 94% on the last two tissues. In another work, the authors obtained an improvement of about 90% in terms of HU accuracy. This value was computed with respect to corresponding MC-corrected CBCT scans used as the ground truth reference [64]. All data in this study had wide FOV, avoiding truncation. Concerning the contrast-to-noise improvement, one work proposed a solution based on a deep convolutional autoencoder that gained an improvement of about 42% in terms of CNR, computed evaluating muscle and fat regions from retrospective CBCT scans [66]. Overall, on real data, the results obtained in the present work align with existing methods not accounting for truncation.

### 4.3.3 Advantages of Transfer Learning

The exclusive use of dataset $D_r$ has some limitations. Firstly, this dataset has a relatively small size for a deep learning-based approach, with the risk of overfitting on this set of patients without generalization capability. Secondly, although the acquisition of CT/CBCT pairs is close in time, the pelvic district involves differences that are not always negligible between the two scans, like air bubbles in the bowel. These problems lead to a sub-optimal selection of CT/CBCT pairs and can have a non-negligible impact on the model performance even when residual deformation itself is as negligible as in our dataset. In order to overcome these issues, we performed a two-step training on $FT_2$ following the transfer learning approach, which increments the overall dataset dimensionality while also integrating exact anatomical information:

1. using synthetically-generated CBCT as input, generated from a publicly available

CT dataset, the network was trained based on perfect anatomical correspondences with ground truth CT. This step provided a suitable initialization of the image-to-image translation process and allowed mainly to learn the image filters and geometric features which identify the anatomical district of interest;

2. using the real CBCT images provided by CNAO, which were only rigidly registered to the corresponding CT scans, the network was partially retrained according to the transfer learning paradigm. In this phase, the training focused on the weights of the inner layers, potentially devoted to learning more complex characteristics related to the intrinsic quality of the image acquired with the specific technology available at the CNAO.

The action of the models (noFT and $FT_2$) can be qualitatively noted by observing the effects they have on the output images (cfr. Fig. 4.9). The noFT model aggressively fits the CBCT to the reference CT. In particular, it applies a transformation similar to an averaging filter. Instead, the $FT_2$ model outputs sharper images. The comparison of HU differences on individual ROIs shows coherent results. $FT_2$ model, with respect to the noFT model, achieves a greater improvement of 6% on the bladder, 19% on bone, 23% on fat, and in general 7% on the whole image, while underperforming only by 3% on muscle. Synthesizing, transfer learning proved to be effective for our application. However, it is important to identify the appropriate number of layers to retrain. If this number is too high (or too low), the network performance will be negatively affected. Tuning this value, we assess the correct balance between generic features like shapes, better learned from synthetic data, and artifacts better learned by real-world clinical data. In this study, we found that the optimal number of layers for retraining is two. Due to the slice-based training, artifacts are suppressed globally and without disentangling them.

### 4.3.4 Technical Challenges and Limitations

Though the results reported above were promising on clinical data, they are still preliminary. The scope of this work is, by all means, a feasibility study. We provided a shading correction method for a limited FOV CBCT scan dataset containing few patients. In order to extend the aim of this work, the dataset must be enlarged, and a dosimetric analysis has to be associated. Another limited aspect of this work is the network ability to generalize. The used datasets were tied to the pelvis, therefore, we cannot assess the impact on other large districts (e.g., lungs). In particular, variation in the patient width is correlated to resulting HU inaccuracies in the input data. However, this impact is attenuated by the proposed network. Admittedly, CBCT scans fed to the $FT_2$ network were corrected for shades but not for ring artifacts. These panel-calibration-induced ripples were already present in the original CBCT volumes, somehow hidden by the more prominent cupping artifact. The noFT model partially compensated for it through its low pass filtering action. Since those artifacts were caused by sub-optimal flat-panel calibration, this shouldn't penalize the $FT_2$ model with respect to the noFT approach. In particular, specializing the $FT_2$ model to correct this hardware issue may hinder the generalization of the method. Or rather, we suggest that this condition would be easily solved by properly calibrating the panel rather than retrospectively correcting for the subsequent artifacts. The main unaddressed limitation resides in the missing informa-

tion due to the narrow FOV of these CBCT scans. The proposed methodology does not try to provide this information alone, but further studies must address this issue. Finally, while results were promising on CNAO dataset $D_r$, they are based on the assumption that the residual deformation between input and label is negligible. Despite results suggesting that the transfer learning approach would be more robust to residual deformation, this aspect needs further investigation.

### 4.3.5 Potential Clinical Impact

Improving the images acquired through this framework had two consecutive advantages. Firstly, a CBCT without shadows and with improved visibility reduced the risk of setup errors by clinicians. Secondly, better contrast between CTV and other soft tissues was indicative of more direct visual discrimination. The CNR improvement of those tissues versus air, paired with enhanced HU adherence to CT, supported the use of CBCT scans for air cavities identification in the pelvis. Consequently, the offline clinical procedure will be more efficient and less prone to overestimation and underestimation of air presence in the bowels. Outside CNAO clinical practice, the use of a public dataset for $FT_2$ improved the repeatability of the present study. Moreover, the minimum required numerosity for the clinical dataset used for training is arguably lower than similar deep learning approaches found in the literature. Another element that may appeal to the potential clinical application is the lack of additional equipment requirements, as this would not change the already established clinical routine or require additional costs. Moreover, the proposed method has negligible execution runtime compared to image reconstruction, avoiding a bottleneck for clinical practice.

## 4.4 Conclusions

Correcting cupping and shading in pelvis CBCT using a U-Net proved flexible enough to adapt to a dataset flawed by truncation artifacts. Furthermore, the efforts to minimize prior knowledge in the network training were accomplished through the implementation of transfer learning. It was demonstrated that the recovery of CT-compatible data from limited FOV CBCT scans with shading artifacts is feasible and can be accomplished as a rapid post-processing measure.

CHAPTER 5

# A Feasibility Study on Proton Dosimetry Computation Using Corrected CBCT

*This chapter presents an analysis of the deep-learning framework in terms of treatment planning updates. The work presented here is based on the paper by Matteo Rossi et al. "Feasibility of Proton Dosimetry Overriding Planning CT with Daily CBCT Elaborated through Generative Artificial Intelligence Tools." Preprints.org 2023, 2023040596.*

Up to this point, all the analyzes conducted have been evaluated in terms of image similarity metrics (e.g., PSNR, SSIM, MAE). In this chapter, we propose a deep learning-based framework that elaborates the CBCT to calibrate the HU, remove artifacts due to the conic geometry acquisition, and handle narrow FOV issues to demonstrate the potential use of the corrected CBCT in the context of proton treatment planning updates. The present work aims to explore the possibility of using limited FOV CBCTs not only for patient positioning but also for dosimetric evaluation without hardware modifications. The deep-learning framework took its root from the CBCT-to-CT mapping model based on cGAN proposed in Chapter 3 [109] that was here extended to address the case of narrow FOV. Tests were carried out on a public dataset of planning CT scans of 40 oncological patients affected by pancreatic cancer. In a first attempt, synthetic raw CBCT volumes were properly generated from CT scans throughout the Monte Carlo simulation. This enabled us to dump anatomical variations usually present in real CBCT with respect to the corresponding planning CT. Moreover, in order to demonstrate the feasibility of the methodology also with real data, we replicated each experiment with the clinical CBCT included within the dataset. As the dataset provided annotation data about the segmented lesion and organs at risk, particle beam dosimetry was computed in the original planning CT and the corrected CBCT volume, verifying the coherency between the two dose distributions. The main contributions of this

chapter may therefore be summarized as:

- capability of the cGAN to correct CBCT (scatter reduction and HU remapping) when applied to small FOV;
- consistency of the proton dosimetry computed on corrected CBCT with respect to the original planning CT.

## 5.1 Materials and Methods

### 5.1.1 Dataset description

A publicly available dataset obtained from the Cancer Imaging Archive, called Pancreatic CT-CBCT-SEG [110], was exploited in this work. The dataset contained CT acquisition from 40 patients who received ablative radiation therapy for locally advanced pancreatic cancer at Memorial Sloan Kettering Cancer Center. Each CT was acquired during a deep inspiration breath-hold verified with an external respiratory monitor. The dataset also included manual segmentations of a region of interest (ROI), defined by expanding the dose planning target volume by 1 cm. Along with the ROI, each scan provided contours of some Organ at Risk (OAR), namely: i) the stomach with the first two segments of the duodenum, ii) the remainder of the small bowel, and iii) both lungs. The authors reported that the segmentations were performed independently by six radiation oncologists and reviewed by two trained medical physicists. The dataset also provided two CBCT scans for each subject. In the first phase, these CBCTs were not considered because they were obtained at different times with respect to the corresponding CT scan, which could lead to potential changes in patient anatomy between acquisitions. Simulated CBCT scans were considered instead by generating them directly from the corresponding planning CT (implementation is detailed below in Section 5.1.1). This way, perfect alignment and anatomical correspondence between the two volumes were both ensured, avoiding the need for additional registration steps (rigid or deformable). To summarize, using simulated CBCT scans allowed the study to focus solely on the algorithm's ability to reduce artifacts and cupping effects without considering intra-patient longitudinal variability. Each experiment was then also replicated with the real small FOV (250 mm diameter) CBCT provided with the dataset by adding an intermediate rigid registration step in the pipeline, in order to demonstrate the feasibility of the methodology also with data from the real world.

#### CBCT Simulation

Synthetic CBCTs were generated from the original available CTs following the approach documented in [111] and replicating the setup and the geometry of CNAO's CBCT acquisition system [8]. Specifically, Monte Carlo (MC) simulations were conducted to generate primary ($P_{MC}$) and scatter ($S_{MC}$) X-ray images for each CT scan. All simulations were performed using the GATE open-source software v9.2 (based on Geant4 v11) [112] with fixed forced detection, a variance reduction technique aimed at minimizing computation time. The energy-dependent detector efficiency was based on the design specifications provided by the manufacturer for the Paxscan 4030D (Varian Medical System, Palo Alto, CA). The X-ray fluence spectrum was computed using the open-source software SpekPy [113], employing 3.2 mm Al filtration at 100 kVp.

The A-277 X-ray tube (Varian Medical System, Palo Alto, CA), chosen for this work, features a 7° rhenium-tungsten molybdenum target. Images were produced according to a CBCT scan of 220° and with projection matching the size of the Paxscan 4030D detector (isometric pixel spacing 0.388 mm, detector size $768 \times 1024$ pixels). The source-to-detector and source-to-isocenter distances were set to 1600 mm and 1100 mm, respectively. A further acceleration of scatter calculation was achieved by down-sampling resolution 8-fold and simulating $S_{MC}$ at 5° steps with a statistical uncertainty $< 5\%$. $S_{MC}$ images were then upsampled and interpolated at the required points to match the corresponding $P_{MC}$ images. Lastly, the final projections $S_{MC} + P_{MC}$ were normalized by the simulated flat field image. CBCT scans were then reconstructed using open-source software RTK [104] at a $1 \times 1 \times 1$ mm resolution, with a $220 \times 220 \times 220$ size in pixel and masked to the axial field-of-view of diameter equal to 204 mm. Some examples of planning CT and simulated CBCT axial slices are visible in Fig. 5.1, along with the intensity profile of the central pixel row. The cupping effect is evident as a shaded portion in the middle of the CBCT and confirmed by the concavity in the intensity profile.

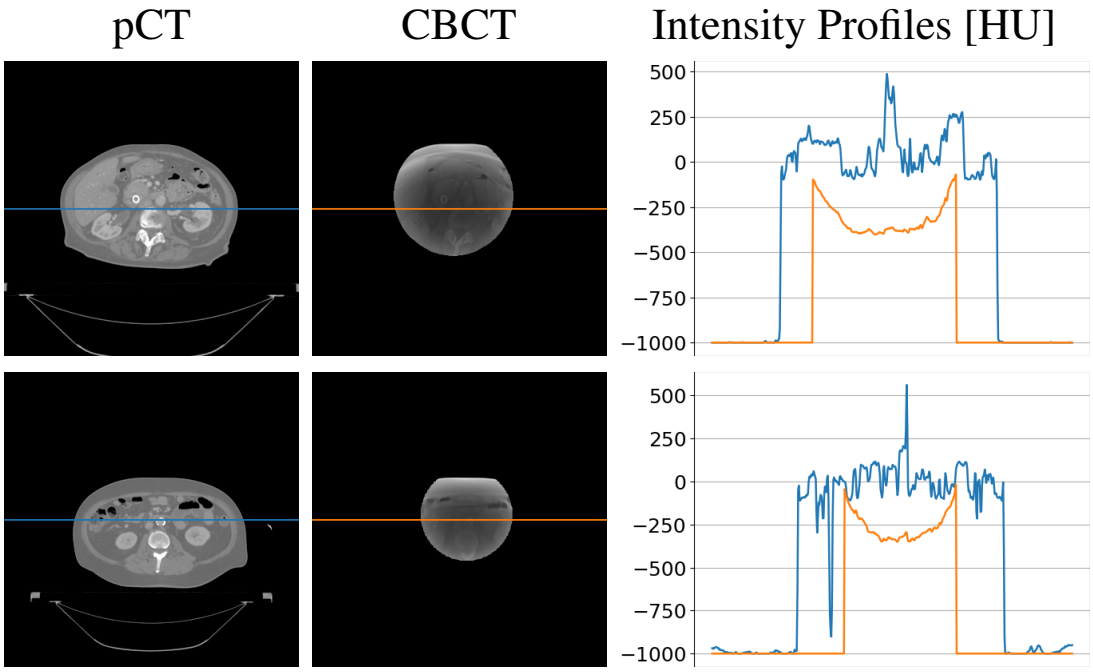


**Figure 5.1:** *Examples of two CT axial slices with their corresponding simulated CBCT. The intensity profiles of the central row (marked as a line in both images) are plotted in the right panel. Each image is displayed with Window = 1300, Level = 0.*

### 5.1.2 CBCT-to-CT Correction

#### Neural Network Architecture and Main Processing Layers

The network implemented for CBCT correction was based on the cycle Generative Adversarial Network (cGAN) [84]. This architecture is based on four concurrent sub-networks, two generators and two discriminators, which work in opposition. While the generators try to learn the mapping to convert CBCT to CT (or CT to CBCT),

the discriminator's objective is to distinguish between authentic and network-generated images. This generator-discriminator cycle-consistent loop is designed to improve the generators' ability to produce synthetic images that reproduce with high fidelity the characteristic of the destination modality (e.g., generate a calibrated synthetic CT starting from a scattered CBCT). The network's fundamental processing unit was referred to as ConvBlocks (Fig. 5.2), which were built using a 2D convolution with a $3 \times 3$ kernel, followed by an instance normalization layer and a swish activation function. Instance normalization was demonstrated to improve the performance in image generation tasks [93]. The use of swish activation was shown to combine the advantages of rectilinear and sigmoid activations. It has a smooth, differentiable form due to the sigmoid component, which can help with training stability and gradient flow [114]. The other basic processing unit for cGAN structure was the InceptionBlock, consisting of four parallel ConvBlocks, each with an increasing kernel size of dimensions $1 \times 1$, $3 \times 3$, $7 \times 7$, and $11 \times 11$, which processed the same input simultaneously with multiple receptive fields. The output of each branch of InceptionBlock was then combined, and the complete set of feature maps was produced as output. The primary objective of this processing block was to execute multi-scale feature extraction from the initial image. The extracted multi-scale features, varying from small to large receptive fields, can produce improved outcomes for image synthesis. The general design of the generator was then carried out as a modified version of the commonly used U-Net architecture. The U-Net model is usually utilized for solving pixel-by-pixel classification challenges in image segmentation [92]. Still, it can also be used to solve image-to-image conversion problems with minor changes. The overall generator structure, depicted in Fig. 5.2, was composed of a contracting and an expanding path. The upper two processing layers of the generator were based on InceptionBlocks, while the deeper two exploited ConvBlocks. Consequently, the network can be broadly top-bottom divided into two segments, each serving distinct functions: i) the inception part (upper layers) extracted global contextual information, whereas ii) the traditional 2D convolution part (bottom layers) was responsible for capturing more intricate context and precise localization. On the other hand, the CNN utilized as the discriminator was responsible for image classification and relied on the PatchGAN architecture [83]. Its architecture was based on four sequential ConvBlocks, each with a kernel size of $4 \times 4$. In the initial three ConvBlocks, the convolution was set with stride 2, leading to an output tensor with half the size and twice the features map. In contrast, the last ConvBlock had stride one and maintained the size and the number of feature maps unchanged. A sigmoid activation function was applied to the last layer, generating a $32 \times 32$ map used to classify the input image as real or fake.

#### Model Training

As formerly stated, the cGAN overall training routine employed two generators and two discriminators, competing against one another to solve the CBCT-to-CT conversion problem. The subnetworks were referred to as generator CT ($G_{CT}$), generator CBCT ($G_{CBCT}$), discriminator CT ($D_{CT}$), and discriminator CBCT ($D_{CBCT}$). $G_{CT}$ and $G_{CBCT}$ were used to produce generated CT from CBCT, and generated CBCT from CT, respectively, while $D_{CT}$ and $D_{CBCT}$ were used to distinguish the original CT and CBCT from their generated counterparts. The training routine was subdivided

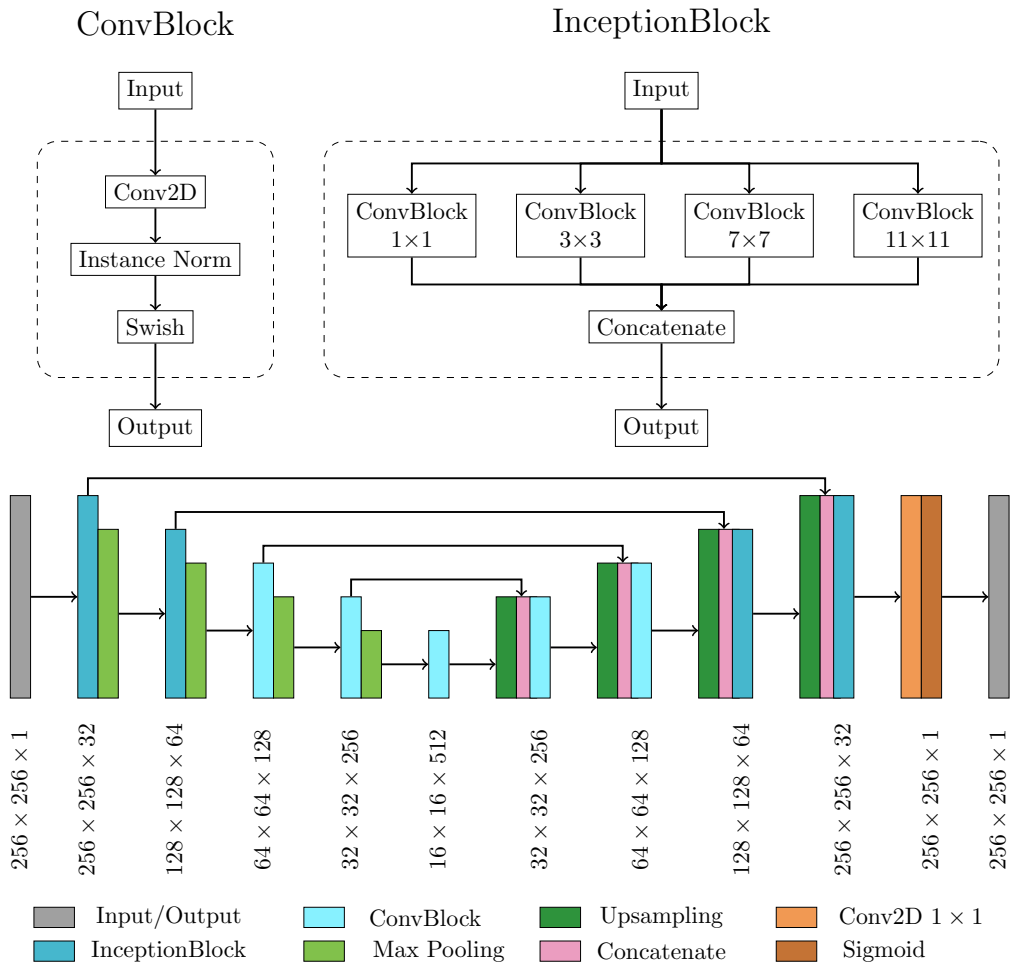


**Figure 5.2:** *Schematic of the Generator model architecture.*

into two main steps occurring simultaneously. In the first step, called the generative phase, $G_{CT}$ ($G_{CBCT}$) took a 2D axial slice of a CBCT (CT) as input and produced a generated CT (generated CBCT) as output. Then, $G_{CT}$ ($G_{CBCT}$) took the generated CT (generated CBCT) as input and produced a cyclic CBCT (cyclic CT), which was supposed to be equal to the original CBCT (CT). At the same time, during the second step, called the classification phase, $D_{CT}$ ($D_{CBCT}$) tried to distinguish between real CT (CBCT) and generated CT (generated CBCT). The entire cGAN was implemented in Python, using Keras [97] and TensorFlow [98] frameworks. All the technical details about the network implementation can be found in a previous work of our group [109]. The whole dataset, consisting of the generated CBCT-pCT (paired), was divided into training, validation, and test sets in proportions of 70%, 15%, and 15%, containing 5698, 1221, and 1221 2D axial slices, respectively.

#### Performance Metrics for Model Evaluation

The network performances were quantitatively evaluated using the original CT as the ground truth reference. In particular, the metrics evaluated were: i) peak signal-to-noise ratio (PSNR), ii) structural similarity index measure (SSIM), and iii) mean absolute er-

ror (MAE) [115]. The PSNR quantifies the quality of images by comparing the mean square error of the images being compared to the maximum possible signal power [95]. It is measured in decibels, and its value increases towards infinity as the difference between the calibrated CBCT and ground-truth CT decreases. Therefore, a larger PSNR value indicates better image quality, while a lower value indicates the opposite. SSIM evaluates the resemblance between two images by analyzing their luminance, contrast, and structure [96]. Compared to PSNR, SSIM is considered a more human-like measure of similarity. The SSIM score ranges from 0 to 1, with a value of 1 indicating the highest level of similarity between the images. MAE was used to quantitatively assess the accuracy of Hounsfield Units (HU) between the generated CBCT and the original CT. The lower value corresponds to the higher level of HU accuracy between the two images. The significance ($p < 0.05$) of the statistical difference between CBCT slices prior to and following calibration was verified using Kruskall-Wallis non-parametric test.

#### Synthetic CT Generation Pipeline

Despite the better quality of calibrated CBCT in terms of HU density values, these volumes cannot yet be used for adaptive dose planning due to their limited FOV. In fact, these CBCT acquisitions lack important anatomical information (e.g., the air/skin interface) necessary for the correct calculation of the beam path. In order to overcome this intrinsic limit, the original planning CT was used to provide the missing information. Therefore, synthetic CT (sCT) is defined in this work as an updated version of the original planning CT overridden with the calibrated voxels from daily CBCT acquired during the treatment. Starting from a scattered CBCT, the following procedure was followed in an axial slice-by-slice approach. At first, each pixel in the slice was clipped between values $[-1000; 3000]$ and then normalized in the $[0; 1]$ range. This step is fundamental because the neural network needs value in this range to operate properly. Then, the generator $G_{CT}$ processed the normalized CBCT, producing a corrected version of the same axial slice. It is important to remember that $G_{CT}$ is the only cGAN subnetwork used after completing the training. After neural network processing, the previous pixel clipping guarantees that the normalization can be reversed back to Hounsfield Unit. A rigid registration step between the corrected CBCT and the pCT followed, in order to increase the anatomical correspondence. This step was applied just for real CBCT, since the generated ones already matched the corresponding pCT anatomy. The last step involved overriding the planning CT pixels with the region acquired with the cone beam modality. Every pixel outside the CBCT FOV belonged to the original CT. The entire pipeline is summarized in Fig. 5.3. In order to evaluate the effective improvement obtained by the corrected CBCT in terms of treatment planning, two versions of sCT were generated for each subject. The first, called sCT corrected (sCTc) was obtained following the mentioned procedure, while the second, called sCT uncorrected (sCTu), was obtained simply by overriding the original CBCT volume without any kind of processing.

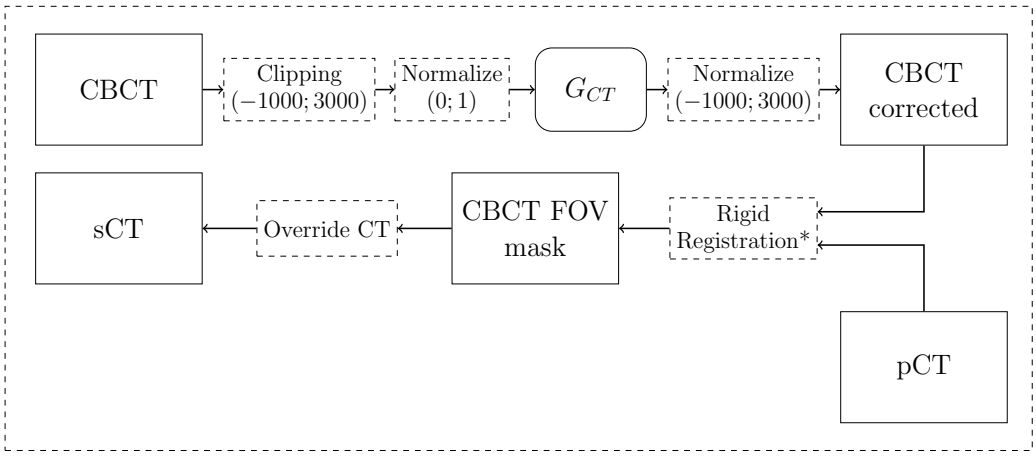


**Figure 5.3:** *Schematic of sCT generation pipeline. *Rigid registration step is applied only to real CBCT scans since generated ones are intrinsically perfectly aligned.*

### 5.1.3 Dosimetric Analysis

#### Proton-based Treatment Planning

The treatment plan for each subject was computed with the matRad package [116], an open-source radiation treatment planning toolkit written in Matlab. In particular, the planning was developed using protons as the radiation mode and optimized using the constant relative biological effectiveness times dose method, which accounts for the varying biological effectiveness of different radiation types and energies. A total of 30 fractions were scheduled for the treatment, with two beams used at gantry angles of 0 degrees (anterior direction) and 270 degrees (right lateral direction). Several constraints were chosen in the planning definition to ensure the safety and efficacy of the treatment. For the bowel and stomach regions, squared overdosing and maximum dose volume histogram constraints were used to limit the radiation dose received by these sensitive areas. The lung regions were also subject to squared overdosing constraints to limit the dose delivered to that areas. Finally, the ROI was subject to squared deviation constraints, which aim to keep the dose distribution as close as possible to the prescribed dose (30 fractions of 2 Gy equivalent) [117]. The reference dose planning was first computed directly on the pCT and used as the ground truth in further comparison. Then, this reference plan was updated, giving either corrected or uncorrected sCT as the new volume.

#### Dose Evaluation

To evaluate the suitability of corrected CBCT scans, various metrics were used, including Dose Difference Pass Rate (DPR), dose-volume histogram (DVH) metrics, and Gamma Pass Rate (GPR). The treatment dose computed for the original planning CT was considered as the prescribed ground truth dose [115]. DPR measures the percentage of pixels that meet a certain dose difference threshold, DVH compares cumulative dose to different structures in relation to volume, while GPR assesses the similarity of two dose distributions based on dose difference and distance-to-agreement criteria. The significance ($p < 0.05$) of the statistical difference in GPR distributions between sCTu

and sCTc was verified using Kruskall-Wallis non-parametric test. Mean doses, D5, and D95, measured on ROI, bowel, and stomach were considered to assess the treatment quality and the toxicity control on the organ at risk.

## 5.2 Results

### 5.2.1 Qualitative evaluation of the image translation

At a qualitative inspection, it can be seen that the darker region clearly visible in the original CBCT due to the cupping artifacts was no longer noticeable after cGAN correction (Fig. 5.4). Likewise, the corrected overwritten sCT scans were more similar to pCTs with respect to their uncorrected counterpart. (Fig. 5.5). Intensity profiles also confirmed this, showing that the concave shape observable in the unprocessed lines disappeared in the processed ones, now matching the HU values range of the reference pCT. This also confirmed that the non-linearity present in the CBCT tissue density was corrected.

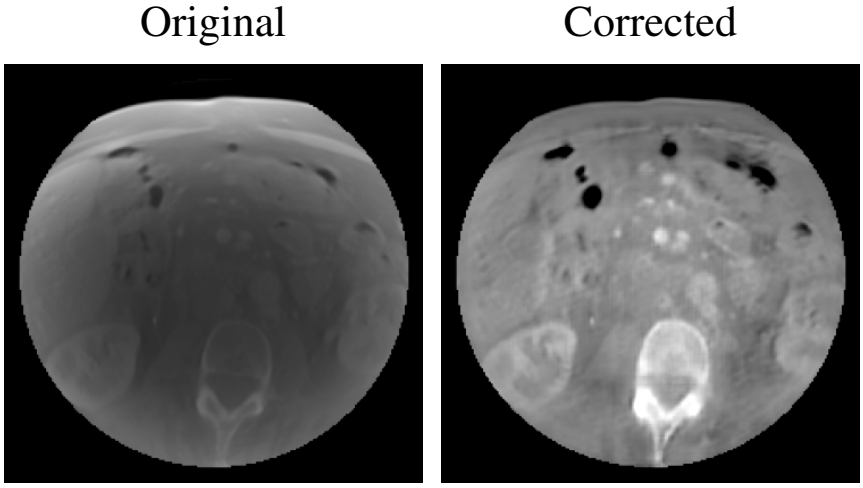


**Figure 5.4:** *Example a CBCT axial slice before (left) and after (right) cGAN correction. As it can be noticed, the cGAN was effective in the correction of the CBCT. Each image is displayed with Window = 1300, Level = 0.*

### 5.2.2 cGAN Model Evaluation

The results from evaluating the model's performance metrics (Fig. 5.6) showed promising improvements in the quality of the CBCT slices. The original images had a median PSNR of 24.60 dB (IQR 1.40 dB), while the processed images had a median PSNR of 33.41 dB (IQR 3.36 dB), resulting in a relative gain of approximately 37%. In terms of the SSIM score, the original images had a median of 0.90 (IQR 0.03), and the processed CBCTs showed a median of 0.95 (IQR 0.02), which corresponded to a relative enhancement of around 5%. Furthermore, the median MAE for the original images was 148.96 HU (IQR 31.24 HU), whereas the median MAE for the processed images was 43.47 HU (IQR 14.82 HU). These results demonstrate the effectiveness of the cGAN approach in improving CBCT image quality. For all three metrics, a statistical difference was found ($p < 0.01$).

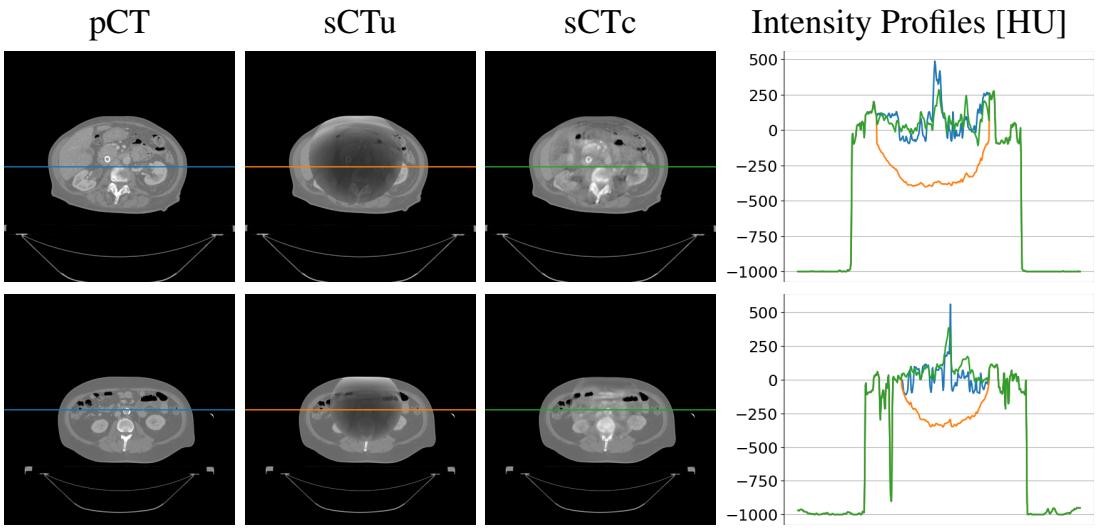


**Figure 5.5:** *Examples of two CT axial slices with their corresponding sCT generated overriding the uncorrected simulated CBCT (sCTu) and the corrected ones (sCTc). The intensity profiles of the central row (marked as a line in both images) are plotted in the right panel. Each image is displayed with Window = 1300, Level = 0.*

### 5.2.3 Treatment Planning Evaluation – Simulated Data

The qualitative comparison between the treatment plans computed for each modality confirmed an ameliorated similarity between sCTc and pCT with respect to their uncorrected counterpart. An example of this can be seen in Fig. 5.7 upper row. It is visible how the beam path computed on sCTu exceeded the ROI releasing a more significant amount of dose in the following tissues. Moreover, it could also be seen that high dose values (red pixels) break over ROI boundaries, indicating that a portion of surrounding healthy areas received overexposure to radiation. This was also confirmed by the dose difference computed with respect to the pCT reference plan (cfr. Fig. 5.7 bottom row). Conversely, the sCTc treatment plan corrected that pattern, showing a more similar beam path and a reduced difference with respect to the pCT plan. Likewise, the dose-volume histogram computed for the same test subject confirmed and enforced such a consideration (Fig. 5.8). Observing the ROI lines, the sCTc (orange dotted line) followed the profile of pCT (orange solid line) more closely than the sCTu (orange dashed line). The treatment plan calculated on sCTc and pCT showed a steep slope, indicating that 2 Gy was the dose delivered to almost all the ROI, while sCTu showed a smoother slope, a sign of overdosing in a portion of this area. Concerning the organs at risk (green and light blue lines), this consideration was even more evident, with an overdosing in the order of about two times with respect to the reference plan. This result would be incompatible with clinical practice. The GPR results for the entire dataset were computed for different gamma criteria and summarized in Table 5.1 as median (IQR). The median gamma pass rates for the 1%/1 mm, 2%/2 mm, 3%/2 mm, and 3%/3 mm criteria were consistently higher for the sCTc the sCTu, with the most significant improvement observed for the 3%/3 mm criterion (92.82% vs. 57.57%).

The GPR results for the entire dataset were computed for different gamma criteria and summarized in Table 5.1 as median (interquartile range, IQR). The median gamma

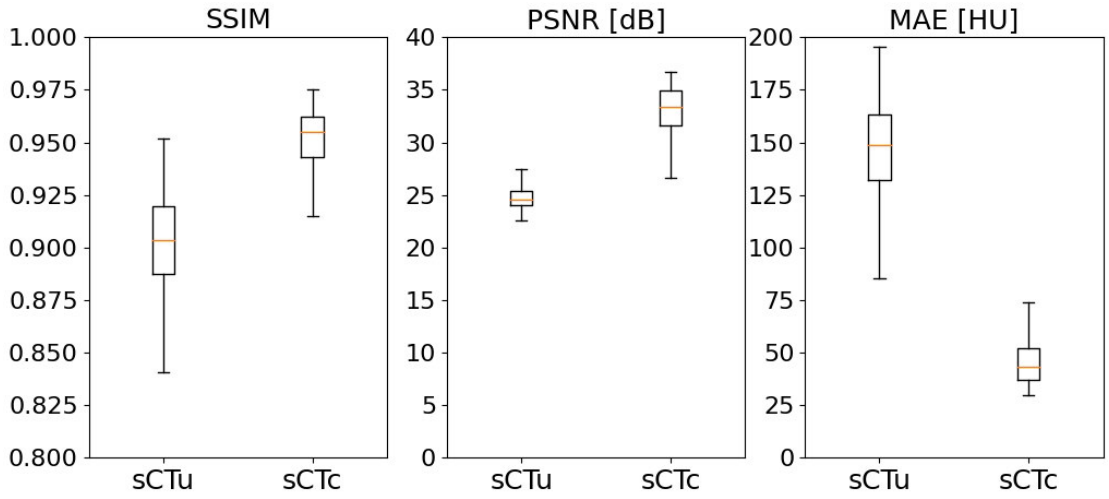


**Figure 5.6:** *Performance metrics for cGAN model evaluation, computed on the axial slices of the test set before and after model processing.*

pass rates for the 1%/1 mm, 2%/2 mm, 3%/2 mm, and 3%/3 mm criteria were consistently higher for the sCTc the sCTu, with the most significant improvement observed for the 3%/3 mm criterion (92.82% vs. 57.57%).

**Table 5.1:** *Gamma pass rate results for different criteria computed using only the simulated CBCT dataset as input to the framework. Statistical difference in gamma score distributions, between uncorrected and corrected sCT, was found ($p < 0.001$).*

| Gamma Criterion | sCTu | sCTc |
|---|---|---|
| 1%/1 mm | 44.68 (6.91) | 74.37 (4.63) |
| 2%/2 mm | 51.72 (8.15) | 87.30 (6.32) |
| 3%/2 mm | 53.78 (8.53) | 90.26 (5.70) |
| 3%/3 mm | 57.57 (7.49) | 92.82 (5.94) |

DPR at 1% was also found to be significantly higher for the processed sCT than the unprocessed ones (93.97% vs. 79.76%), indicating a 14.21% improvement in dose accuracy with the use of sCTc. In regard to mean dose distribution across the overall dataset, the advantage of sCTc was evident with respect to the overdosing of the sCTu in the ROI (Table 5.2). The relative percentage error decreased from 6% for sCTu up to 2% for sCTc. A greater advantage was achieved in terms of unwanted doses distributed at the bowel as 97% against 2%, with respect to the nominal toxicity in the pCT. Likewise, the relative toxicity in the stomach decreased from 49% up the 2%. For D5, the correction was effective in reducing the overexposure found in the unprocessed sCT. For D95, the correction underestimated the dose of about 5%. As far as OAR is concerned, the correction was again effective in ensuring a low dose, very similar to that one obtained in the planning CT, at both bowel and stomach. Interestingly, the IQR range of 0.12 Gy (D95) for the stomach potentially delivered when using the sCTu was completely zeroed by the correction.

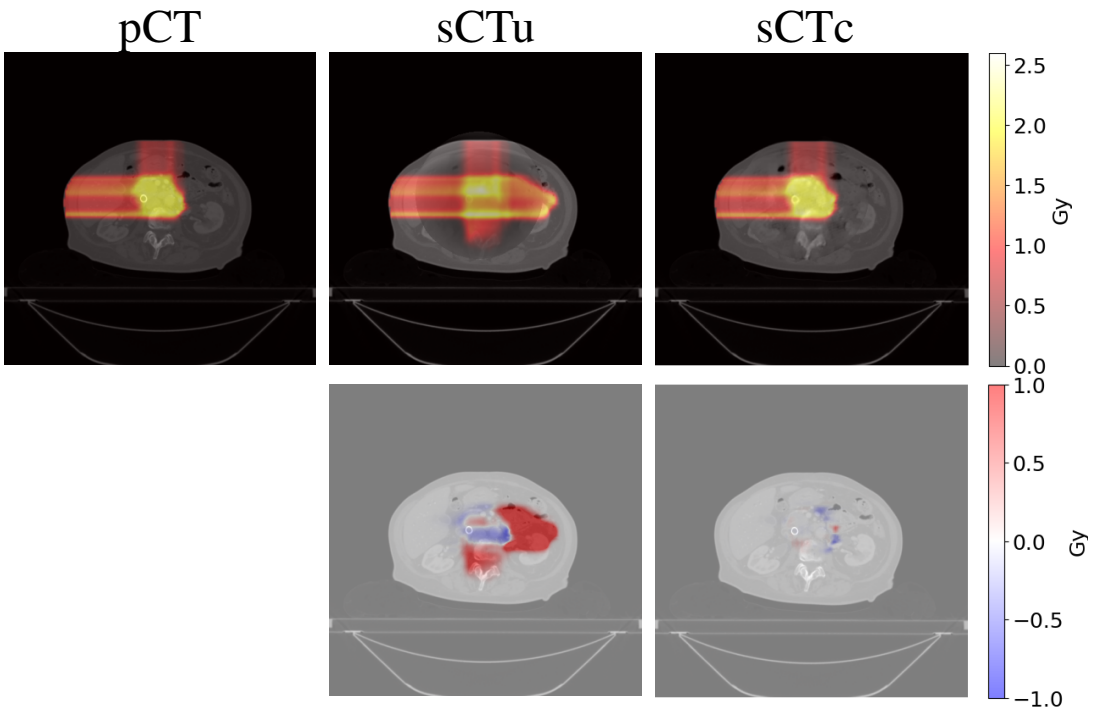


**Figure 5.7:** *Example of dose planning for an axial slice of a subject for the original pCT, sCTu, and sCTc (upper row). The difference between both sCT treatment plans with respect to the original pCT plan is shown in the bottom row. Synthetic CT scans in this figure were produced using a simulated CBCT as the input.*

### 5.2.4 Treatment Planning Evaluation – Real Data

Each result presented in previous sections referred to simulated CBCT. The following results refer to sCT generated using real CBCT as the input data in order to show the quality of the treatment planning with real-world data. As explained in Section 5.1.2, a rigid registration step was added to the pipeline just before the pCT override (Fig. 5.9). Once again, the comparison of the treatment plans generated for each method confirmed an improved similarity between sCTc and pCT with respect to their uncorrected equivalent. Fig. 5.10 depicts an example case in which it is possible to observe again how the beam path exceeded the ROI with a consequent overdosing in the adjacent healthy tissues when computed on sCTu. However, the sCTc treatment plan effectively reduced the overdosing and corrected this pattern, leading to a beam path that closely resembled the pCT plan and reducing the differences (cfr. Fig. 5.10 bottom row). The corresponding dose-volume histogram for the test subject (Fig. 5.11) further supported this observation. Looking at the ROI lines, the sCTc (orange dotted line) followed the profile of pCT (orange solid line) more closely than the sCTu (orange dashed line). Even if the corrected plan slightly overdosed about 60% of the volume, the effect is reduced compared to the uncorrected plan, which presented a smoother slope for the entirety of the ROI, giving 1.25 Gy to the 100% of the volume instead of the 2 Gy of the prescribed dose and overdosing the rest. This observation became even more apparent when considering the organs at risk (represented by the green and light blue lines). Specifically, the corrected treatment plan for the bowel closely replicated the

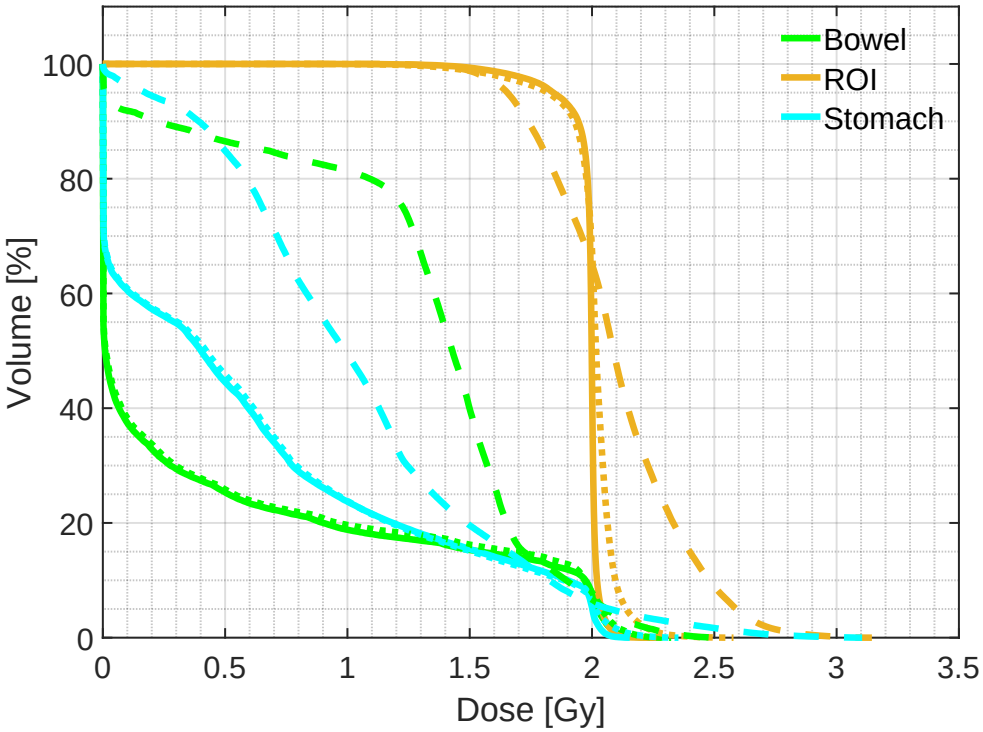


**Figure 5.8:** *Example of DVH computed for ROI and two organs at risk (bowel and stomach). Solid line: pCT, dashed line: sCTu, dotted line: sCTc. Synthetic CT scans in this figure were produced using a simulated CBCT as the input.*

prescribed one, demonstrating a good level of accuracy.

For a quantitative comparison, the GPR results computed on the entire dataset are summarized in Table 5.3 as median (IQR). Again, even in the presence of real CBCT overridden to the pCT, the median gamma pass rates for all the computed criteria were consistently higher for the sCTc, with the most significant improvement observed for the 3%/2 mm criterion (23% difference).

Regarding the average dose distribution, the superiority of sCTc over sCTu in terms of overdosing within the ROI was evident, as indicated in Table 5.4. The relative percentage error decreased from 3% for sCTu to 1% for sCTc. Furthermore, a significant advantage was also achieved in terms of undesired doses in the bowel, with error percentages of 31% (sCTu) and 13% (sCTc) in comparison to the nominal toxicity in the pCT. Similarly, the relative toxicity in the stomach decreased from 12% to 3%.

## 5.3 Discussion

This work proposed a novel image-processing framework for generating synthetic CT scans, which combines the original planning CT with routine CBCT scans, usable to update the dosimetry plan in proton therapy. The core of the framework was represented by a deep learning model, namely a cycle-consistency GAN, to correct the scatter artifacts in the CBCT images and calibrate intensity values, in the proper HU range. Especially, the framework was shown to properly address CBCT equipment scanning

**Table 5.2:** *Mean doses, D5, and D95, measured on ROI, bowel, and stomach, computed using only the simulated CBCT dataset as input to the framework. Values are expressed as median(IQR) Gy.*

| | | Mean dose | D5 | D95 |
|---|---|---|---|---|
| **ROI** | **pCT** | 1.98 (0.01) | 2.04 (0.01) | 1.86 (0.07) |
| | **sCTu** | 2.10 (0.09) | 2.61 (0.28) | 1.64 (0.10) |
| | **sCTc** | 1.93 (0.08) | 2.09 (0.05) | 1.54 (0.28) |
| **Bowel** | **pCT** | 0.47 (0.44) | 2.01 (0.04) | 0.00 (0.00) |
| | **sCTu** | 0.93 (0.52) | 2.06 (0.31) | 0.00 (0.00) |
| | **sCTc** | 0.48 (0.40) | 2.00 (0.13) | 0.00 (0.00) |
| **Stomach** | **pCT** | 0.65 (0.27) | 2.01 (0.03) | 0.00 (0.00) |
| | **sCTu** | 0.97 (0.35) | 2.14 (0.24) | 0.00 (0.12) |
| | **sCTc** | 0.64 (0.29) | 2.02 (0.11) | 0.00 (0.00) |

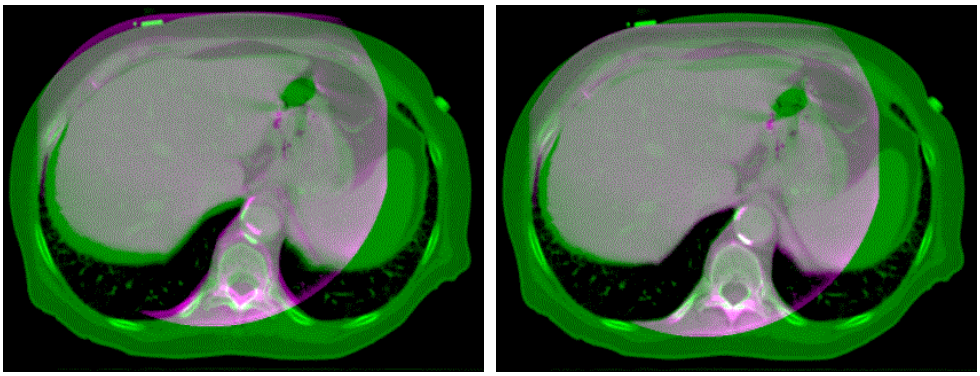

**Figure 5.9:** *Overlay of corrected CBCT (pink) on pCT (green) before (left) and after (right) rigid registration. It can be seen that the registration was effective in the alignment of the bony structures. Each image is displayed with Window = 1300, Level = 0.*

narrow FOV [115]. To the aim, the public dataset of patient-paired CT-CBCT scans, named Pancreatic-CT-CBCT-SEG [110], was exploited because of clinically consistent segmentation of ROI and OAR across all the considered patients. From the CT scans, physically consistent simulated CBCT were generated by means of the Monte Carlo algorithm, also accounting for narrow FOV. Especially, Monte Carlo parameter tuning was set according to the CBCT equipment and the acquisition setup at the CNAO. The cGAN-based correction mapped synthetic uncorrected CT into synthetic corrected CT. We remark that, as long as the dataset used to train the network is representative of the data expected to be used in clinical practice, the methodology remains robust. In terms of robustness, the applied methodology did not require retraining the model when moving from simulated to real data. Generally speaking, as soon as the predicted data are no longer satisfactory, the model can be retrained with data taken from the actual clinical setup.

The available segmentations granted replicating the particle beam planning, computed on the original CT, to the corrected sCT for straightforward comparison. The obtained results confirmed, both qualitatively and quantitatively, the capability of the cGAN-based to correct CBCT into CT-compatible images. As shown, the intensity profiles were rectified adequately thanks to the reduction of cupping and truncation artifacts

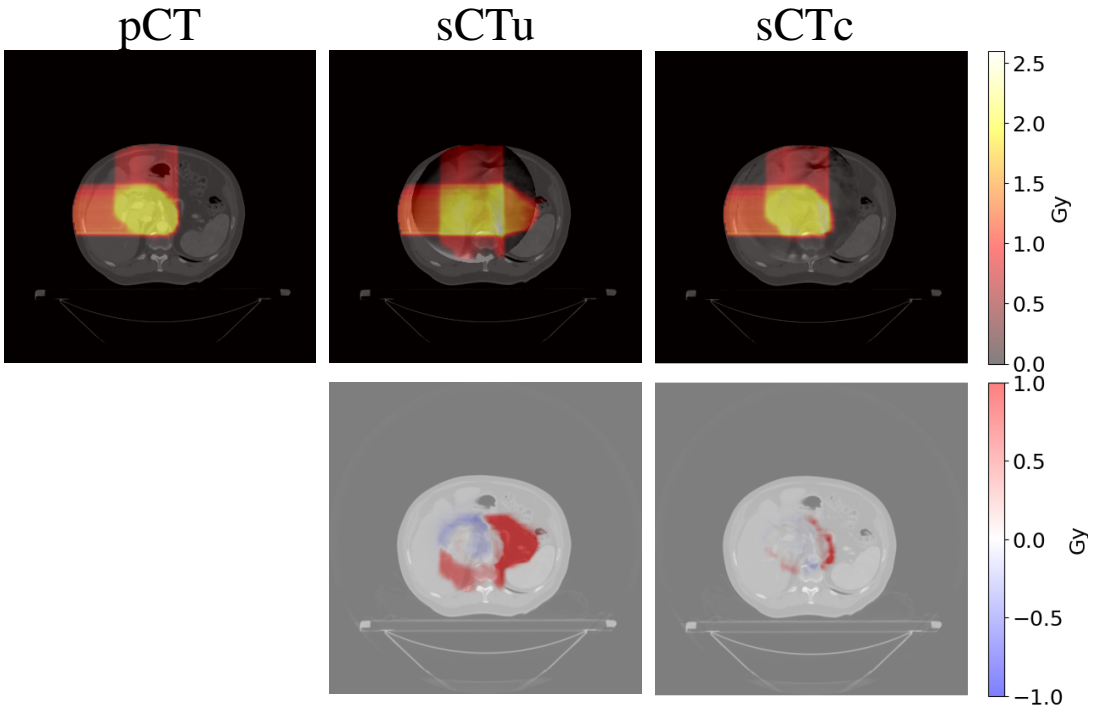


**Figure 5.10:** *Example of dose planning for an axial slice of a subject for the original pCT, sCTu, and sCTc (upper row). The difference between both sCT treatment plans with respect to the original pCT plan is shown in the bottom row. Synthetic CT scans in this figure were produced using a real CBCT as the input.*

(cfr. Fig. 5.5). A significant increase in PSNR, SSIM, and MAE metrics testified to the effectiveness of the methodology (cfr. Fig. 5.6). Dosimetry computed on uncorrected sCT featured overdosing, especially at OAR (cfr. Tables 5.1 and 5.2). An inaccurate assessment of tissue densities was made due to the lack of HU calibration. The difference in grayscale values between CBCT and CT caused a discontinuity in the volume (cfr. Fig. 5.5), leading to errors in particle beam path computation. Conversely, the dosimetric plan computed on the corrected sCT confirmed the consistency of the plan computed on the corresponding CT scan (cfr. Table 5.1). The American Association of Physicists in Medicine (AAPM) Task Group 218 defined acceptance criteria for tolerance and action limits as exceeding 95% and falling below 90%, respectively, for a 3%/2 mm GPR standard [118]. While not completely in agreement with the upper threshold, the 90.26% found in this work (cfr. Table 5.1) is to be deemed reasonable. Nonetheless, such value, overcoming the lower 90% action limit threshold, makes the methodology promising for clinical application. In order to avoid confounding factors induced by organ deformation, this work did use simulated CBCT by means of the Monte Carlo approach, so that differences between the images were exclusively due to artifacts rather than anatomy. This allowed us to use the same lesion and OAR segmentation used to compute the reference treatment plan. However, results obtained using real-world CBCT as input to the framework confirmed the feasibility of the approach (cfr. Table 5.3). Real data resulted in generally lower performances when compared to the simulated ones. This was mainly due to the anatomical changes when in the presence of volume acquired on different days during the treatment. It is important to

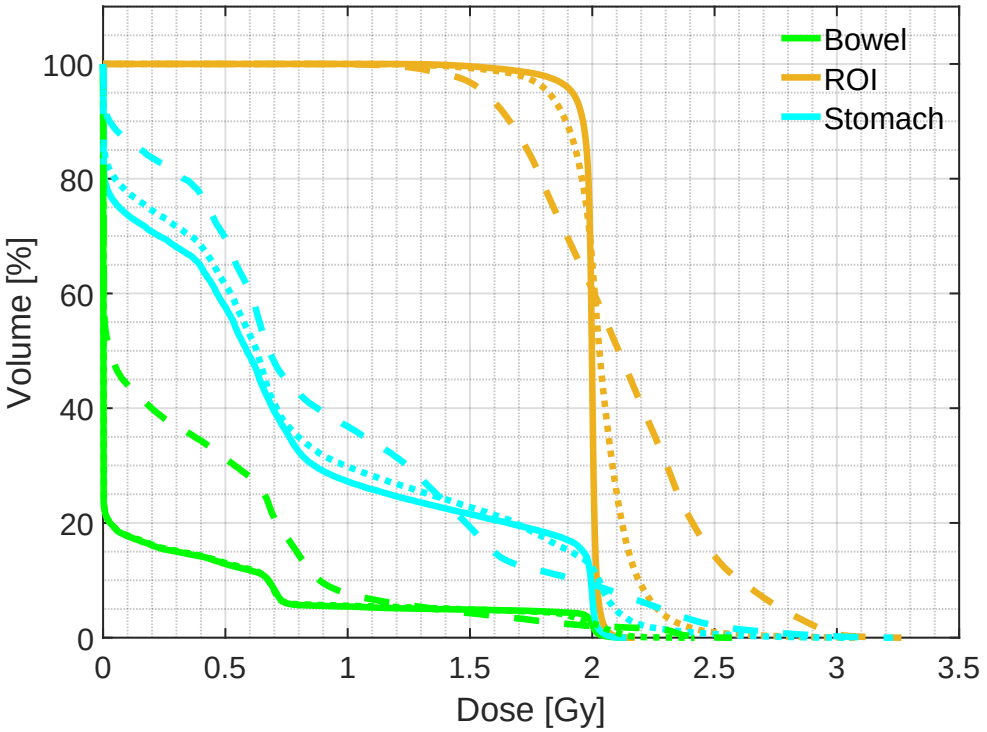


**Figure 5.11:** *Example of DVH computed for ROI and two organs at risk (bowel and stomach). Solid line: pCT, dashed line: sCTu, dotted line: sCTc. Synthetic CT scans in this figure were produced using a real CBCT as the input.*

remind that the pelvic site, the subject of this study, contains soft tissue that implies relative movement between organs and the generation of air bubbles [4, 68]. Moreover, during the treatment, the patient often loses weight, and the tumor changes shape due to the treatment itself. This makes this framework not always applicable, but a possibly useful aid when changes in internal anatomy are not so evident. For every other case, the traditional acquisition of a revaluation CT remains a better solution. Comparison with works in the literature using deep learning to correct CBCT and test proton dosimetry showcased the consistency of the methodology in terms of GPR 2%/2 mm results, even though slightly smaller values with respect to that reported in some papers (Table 5.5). Significantly, our results on simulated and real data differed less than expected. Nonetheless, we note that our study applied the correction to small FOV (204 mm diameter for simulated data and 250 mm for real data) while the mentioned works dealt with mainly wider FOV (about 480 mm diameter on average). All the works in this comparison involved large anatomical sites (e.g., pelvis, thorax, and abdomen) and the patient cohort had a similar size. Remarkably, the comparison highlighted the superiority of the generative adversarial networks [73, 86] with respect to traditional U-Net [46, 47].

Concerning the limited FOV, it is fundamental to recall that the volume of interest is entirely contained in the CBCT and that the tissues coming from the planning CT are only needed to calculate the beam path when the air/tissue interface is not already embedded in the CBCT [35, 36]. In addition, this makes it mandatory to update the

**Table 5.3:** *Gamma pass rate results for different criteria computed using real CBCT dataset as input to the framework. Statistical difference in gamma score distributions, between uncorrected and corrected sCT, was found ($p < 0.001$).*

| Gamma Criterion | sCTu | sCTc |
|---|---|---|
| 1%/1 mm | 49.73 (14.84) | 71.97 (7.01) |
| 2%/2 mm | 61.41 (14.17) | 84.37 (5.89) |
| 3%/2 mm | 65.36 (14.17) | 87.20 (5.79) |
| 3%/3 mm | 70.11 (13.66) | 89.87 (5.26) |

**Table 5.4:** *Mean doses, D5, and D95, measured on ROI, bowel, and stomach, computed using only the real CBCT dataset as input to the framework. Values are expressed as median(IQR) Gy.*

| | | Mean dose | D5 | D95 |
|---|---|---|---|---|
| **ROI** | **pCT** | 1.98 (0.01) | 2.04 (0.01) | 1.86 (0.07) |
| | **sCTu** | 2.03 (0.08) | 2.37 (0.21) | 1.73 (0.13) |
| | **sCTc** | 1.96 (0.06) | 2.11 (0.07) | 1.69 (0.18) |
| **Bowel** | **pCT** | 0.47 (0.44) | 2.01 (0.04) | 0.00 (0.00) |
| | **sCTu** | 0.62 (0.43) | 2.05 (0.22) | 0.00 (0.00) |
| | **sCTc** | 0.41 (0.29) | 2.00 (0.12) | 0.00 (0.00) |
| **Stomach** | **pCT** | 0.65 (0.27) | 2.01 (0.03) | 0.00 (0.00) |
| | **sCTu** | 0.73 (0.31) | 2.12 (0.15) | 0.00 (0.00) |
| | **sCTc** | 0.63 (0.26) | 2.02 (0.06) | 0.00 (0.00) |

segmentation mask. In this work, the FOV was particularly reduced to demonstrate the feasibility of the methodology. In general, the proposed framework can be adopted in all cases in which the district of interest is too large to fit into a single acquisition. Furthermore, the proposed method extends the use of CBCT systems currently used for patient positioning without additional hardware. Considering that daily CBCTs are acquired for patient positioning purposes, the present framework can be used in parallel with the current clinical routine. Moreover, the computation time required to produce an updated treatment plan based on the sCT starting from a CBCT is in the order of a few minutes on an average computer, making it compatible with clinical routines. However, in this work, we did not use the treatment plan computed using the CBCT scan acquired at dose delivery time as the ground truth, but the one computed on the original pCT. While we acknowledge that this can be regarded as a shortcoming towards generalization, the feasibility of the overall methodology was showcased. In future work, the method will be evaluated in a real offline therapy context at the CNAO facility [8]. Correction of CBCTs obtained day-by-day during treatment will be used to assess the evolution of the dose plan without the need to acquire additional CTs and administer further toxicity to the patient [34, 73]. Finally, no additional hardware will be needed to add to the patient positioning setup in order to increase the FOV for dosimetric evaluation.

**Table 5.5:** *Comparison of dosimetry results with literature outcomes in terms of GPR 2%/2 mm in the domain of proton therapy.*

| Work | Model | Anatomic Site | axial FOV [mm] | Patient cohort | GPR 2%/2 mm |
|---|---|---|---|---|---|
| Hansen et al. [46] | U-Net | Pelvis | 410 | 30 | 53% |
| Landry et al. [47] | U-Net | Pelvis | 410 | 42 | 85% |
| Thummerer et al. [119] | U-Net | Thorax | 500 | 33 | 90.7% |
| Kurz et al. [86] | cGAN | Pelvis | 550 | 33 | 96% |
| Uh et al. [73] | cGAN | Abdomen/Pelvis | 530 | 50 | 98.5% |
| This work - simulated | cGAN | Pelvis | 204 | 40 | 87.3% |
| This work - real | cGAN | Pelvis | 250 | 40 | 84.4% |

## 5.4 Conclusions

The present study proposed a generative artificial intelligence tool to correct CBCT scans, acquired with narrow FOV systems, enabling the reduction of scatter and the remap pixel intensity in HU. The methodology made feasible treatment planning updates, which brings the use of CBCT images incrementally closer to clinical translation in proton therapy.

CHAPTER 6

# Work Synthesis

*This chapter concludes the present dissertation. A summary of the aim, materials, methods, and results is reported.*

The present manuscript described a feasibility study for the potential extension of use for CBCT imaging in the field of image-guided radiotherapy, addressing the well-known issues of this imaging modality. This ambition was addressed by implementing a new robotic system and developing innovative methods for image processing based on deep learning. The integration of highly accurate in-room imaging is crucial for the successful use of advanced techniques in precise dose delivery for external beam radiotherapy, considering factors such as weight loss, tumor shrinkage, and air bubbles that can introduce inter-fractional discrepancies during treatment [4]. In radiation therapy using particle beams, the importance of in-room imaging becomes even more critical due to the inverse dose deposition profile and increased sensitivity to uncertainties. CBCT presents a valuable imaging technique for patient positioning and monitoring, offering quicker imaging and reducing patient exposure to non-therapeutic radiation. However, it can introduce scattering artifacts such as shading, cupping, and beam-hardening [28,29], and the presence of a limited FOV CBCT imaging system can lead to reconstruction problems [35, 36] and additional bright-band effects that introduce artifacts to the acquired scans [37].

## 6.1 Main Findings

This dissertation presented the CBCT imaging system developed and installed at CNAO and provided a description of a deep-learning framework that processes the CBCT to calibrate the HU, remove artifacts due to the conic geometry acquisition, and handle

narrow FOV issues to demonstrate the potential use of the corrected CBCT in the context of proton treatment planning updates. The framework aims to explore the possibility of using the in-house limited FOV CBCT systems not only for patient positioning but also for dosimetric evaluation without any hardware modifications. From a technological point of view, Chapter 2 detailed the design, commissioning, and quality assurance protocols undertaken to create a custom in-room volumetric imaging system for radiation treatments utilizing accelerated particles at CNAO. These tests served as essential prerequisites for the successful implementation of customized solutions for online image guidance in particle therapy. Furthermore, this chapter provided an overview of the system components and the increased FOV achieved for CBCT. It also presented the outcomes of measurement campaigns and commissioning tasks conducted to evaluate the geometric precision of the calibrated system. Additionally, the chapter examined the reliable setup correction in both nominal and clinical conditions for all available imaging modalities, validating their accuracy. Concerning methodological innovations, the subsequent chapters explored the possibility of using deep learning to address the intrinsic limitations of CBCT imaging. Given the difficulties in comparing methods from the literature, due mainly to the strong dependence on the dataset used, we started the development of this deep learning-based solution with one research question in mind: "which training paradigm is more suitable for this kind of task between supervised and unsupervised learning?". Chapter 3 tried to give an answer providing a fair comparison between the two methodologies using the same dataset, confirming the capability of CNN to generate accurate synthetic CT scans from CBCT images and also proposing some guidelines to drive the selection of the better training technique, which can also be shifted to a more general image-to-image translation. According to the obtained results, the supervised model outperformed the unsupervised ones in terms of both quantitative image quality metrics and computational cost. However, the supervised-trained model exhibited occasional unreliability, leading to unexpected behavior, particularly in contour regions. In contrast, the unsupervised model consistently avoided such artifacts and demonstrated greater reliability in preserving the anatomical structure and up-to-date information present in the CBCT scan. The more straightforward loss function defined for the supervised model compared to the unsupervised one, i.e., mean square error, specialized the network to learn mapping without adequately considering the anatomical structure. For the unsupervised training, we opted for an image-to-image approach leveraging on a cGAN because of its internal architecture and the combination of its loss functions (a detailed description of them can be found in Chapter 3.1.5). The key idea behind cGAN is the introduction of cycle consistency loss, which enforces that the translation of an image from one domain to another and then back to the original domain should ideally result in the same image. This cycle consistency constraint helps the model learn meaningful mappings between domains even without paired data. The problem presented in this thesis can be described as a "style transfer" class of problem instead of a pure pixel-to-pixel remapping. Even if the cGAN may have reduced numerical performances in terms of, e.g., pure PSNR, SSIM, or MAE when compared to other networks, its cyclic structure demonstrated to be more reliable in understanding what a CBCT "is", and what it makes different from a CT (i.e., artifacts, shading, etc.). Moreover, considering the difficulty in generating high-quality paired datasets for this particular scenario due to the relative difference

between CBCT and CT even when acquired on the same day, we considered cGAN more suitable for CBCT-to-CT translation than, e.g., pix2pix network that relies only on paired datasets. Then, Chapter 4 presented some techniques to further address the limitations of deep learning-based CBCT correction, leveraging on the transfer learning paradigm with an innovative two-step approach. We demonstrated that preconditioning the network with synthetic data is an effective method for improving the quality of corrected CBCT in the presence of a limited dataset. Then, transfer learning with real data proved to be effective for performance enhancement when dealing with data coming from real clinical practice. In particular, with this two-step approach, the model learns the balancing between generic features like shapes, better learned from synthetic data, and artifacts, better learned by real-world clinical data. Finally, the deep-learning framework was also tested in terms of treatment planning updates. In Chapter 5, we extended the analysis of CNN-based CBCT correction to the specific problem of limited FOV, evaluating the consistency of the proton dosimetry computed with respect to the original pCT. This analysis exploited a cGAN model trained through an unsupervised approach. Monte Carlo simulation was employed to generate CBCT scans, allowing for a focused evaluation of the algorithm's effectiveness in reducing artifacts and cupping effects without considering intra-patient longitudinal variability to ensure a fair comparison between planning CT and calibrated CBCT dosimetry. Then, experiments were conducted fine-tuning the model with real CBCT data to demonstrate its viability with clinical real-world data. The calibration of simulated CBCT resulted in a proton dosimetry difference of less than 2% compared to the pCT. The potential toxicity impact on organs at risk decreased from approximately 50% (without calibration) to 2% (with calibration). The gamma pass rate at 3%/2mm showed a significant improvement of around 37% in accurately reproducing the prescribed dose before and after calibration (53.78% vs. 90.26%). Real-world data confirmed these findings, albeit with slightly lower performance according to the same criteria (65.36% vs. 87.20%). These results potentially confirm that the proposed deep learning-based framework brings the use of narrow FOV CBCT scans incrementally closer to clinical translation in proton therapy planning updates.

## 6.2 Potential Clinical Impact

Improving the CBCT acquired through the developed imaging system had some main advantages regarding clinical impact. Firstly, a CBCT without shadows and with improved visibility reduces the risk of setup errors by clinicians, providing an improvement even for the primary purpose of CBCT image, i.e., the patient positioning and setup verification. Secondly, better contrast between tumors and other soft tissues is indicative of more direct visual discrimination. The contrast-to-noise improvement of those tissues versus air, paired with enhanced HU adherence to CT, supports the use of CBCT scans for, e.g., air cavity identification. Consequently, the offline clinical procedure results in more efficient and less prone to overestimating and underestimating anatomopathological changes. An additional appealing aspect regards the potential reduction in the non-therapeutic doses given to the patient related to the need for revaluation CT (rCT). With corrected CBCT, the number of unnecessary rCT could be reduced, optimizing the additional dose provided. In general, the proposed framework

can be adopted in all cases in which the district of interest is too large to fit into a single acquisition. Furthermore, the proposed method extends the use of CBCT systems currently used for patient positioning without additional hardware. We remark that, as long as the dataset used to train the network is representative of the data expected to be used in clinical practice, the methodology remains robust. In terms of robustness, the applied methodology did not require retraining the model when moving from simulated to real data. Generally speaking, as soon as the predicted data are no longer satisfactory, the model can be retrained with data taken from the actual clinical setup.

## 6.3 Conclusions

At the time of writing, the system is in the final certification stage and will soon enter clinical practice in room 1 at CNAO for patient positioning correction (Fig. 6.1). Regarding the deep learning framework and the possibility of introducing the use of CBCT for dose evaluation in clinical practice, CNAO is currently evaluating a pre-clinical experimental study with patients. Personally, I consider the production of this system a fundamental achievement of my academic career.

**Figure 6.1:** *The development of the CBCT imaging system in testing room (**left**). The final CBCT imaging system after installation (**right**)*

# Bibliography


[1] Ugo Amaldi and Gerhard Kraft. Radiotherapy with beams of carbon ions. *Reports on Progress in Physics*, 68(8):1861–1882, jul 2005.

[2] Marco Durante and Jay S. Loeffler. Charged particles in radiation oncology. *Nature Reviews Clinical Oncology*, 7(1):37–43, dec 2009.

[3] Francis A Cucinotta and Marco Durante. Cancer risk from exposure to galactic cosmic rays: implications for space exploration by human beings. *The Lancet Oncology*, 7(5):431–435, may 2006.

[4] Tianye Niu, Ahmad Al-Basheer, and Lei Zhu. Quantitative cone-beam CT imaging in radiation therapy using planning CT as a prior: First patient studies. *Medical Physics*, 39(4):1991–2000, mar 2012.

[5] S van de Water, R Kreuger, S Zenklusen, E Hug, and A J Lomax. Tumour tracking with scanned proton beams: assessing the accuracy and practicalities. *Physics in Medicine and Biology*, 54(21):6549–6563, oct 2009.

[6] Martijn Engelsman, Marco Schwarz, and Lei Dong. Physics controversies in proton therapy. *Seminars in Radiation Oncology*, 23(2):88–96, apr 2013.

[7] Antje-Christin Knopf, Dirk Boye, Antony Lomax, and Shininchiro Mori. Adequate margin definition for scanned particle therapy in the incidence of intrafractional motion. *Physics in Medicine and Biology*, 58(17):6079–6094, aug 2013.

[8] G. Fattori, M. Riboldi, A. Pella, M. Peroni, P. Cerveri, M. Desplanques, G. Fontana, B. Tagaste, F. Valvo, R. Orecchia, and G. Baroni. Image guided particle therapy in CNAO room 2: Implementation and clinical validation. *Physica Medica*, 31(1):9–15, feb 2015.

[9] Catarina Veiga, Guillaume Janssens, Ching-Ling Teng, Thomas Baudier, Lucian Hotoiu, Jamie R. McClelland, Gary Royle, Liyong Lin, Lingshu Yin, James Metz, Timothy D. Solberg, Zelig Tochner, Charles B. Simone, James McDonough, and Boon-Keng Kevin Teo. First clinical investigation of Cone Beam Computed Tomography and deformable registration for adaptive proton therapy for lung cancer. *International Journal of Radiation Oncology Biology Physics*, 95(1):549–559, may 2016.

[10] Chiaho Hua, Weiguang Yao, Takao Kidani, Kazuo Tomida, Saori Ozawa, Takenori Nishimura, Tatsuya Fujisawa, Ryousuke Shinagawa, and Thomas E. Merchant. A robotic C-arm cone beam CT system for image-guided proton therapy: design and performance. *The British journal of radiology*, 90(1079):20170266, 2017.

[11] Guillaume Landry and Chia-ho Hua. Current state and future applications of radiological image guidance for particle therapy. *Medical Physics*, 45(11):e1086–e1095, nov 2018.

[12] Salvatore Devicienti, Lidia Strigari, Marco D'Andrea, Marcello Benassi, Vincenzo Dimiccoli, and Maurizio Portaluri. Patient positioning in the proton radiotherapy era. *Journal of Experimental & Clinical Cancer Research*, 29(1), may 2010.

[13] Giovanni Fava, Lamberto Widesott, Francesco Fellin, Maurizio Amichetti, Valentina Viesi, Antony J. Lomax, Lydia Lederer, Eugen B. Hug, Claudio Fiorino, Giovannella Salvadori, Nadia Di Muzio, and Marco Schwarz. In-gantry or remote patient positioning? Monte Carlo simulations for proton therapy centers of different sizes. *Radiotherapy and Oncology*, 103(1):18–24, apr 2012.

[14] A. Pella, M. Riboldi, B. Tagaste, D. Bianculli, M. Desplanques, G. Fontana, P. Cerveri, M. Seregni, G. Fattori, R. Orecchia, and G. Baroni. Commissioning and quality assurance of an integrated system for patient positioning and setup verification in particle therapy. *Technology in Cancer Research & Treatment*, 13(4):303–314, aug 2014.

[15] Alessandra Bolsi, Antony J. Lomax, Eros Pedroni, Gudrun Goitein, and Eugen Hug. Experiences at the paul scherrer institute with a remote patient positioning procedure for high-throughput proton radiation therapy. *International Journal of Radiation Oncology Biology Physics*, 71(5):1581–1590, aug 2008.

[16] Judith van Loon, Janneke Grutters, and Fergus Macbeth. Evaluation of novel radiotherapy technologies: what evidence is needed to assess their clinical and cost effectiveness, and how should we get it? *The Lancet Oncology*, 13(4):e169–e177, apr 2012.

[17] Yolande Lievens and Madelon Pijls-Johannesma. Health economic controversy and cost-effectiveness of proton therapy. *Seminars in Radiation Oncology*, 23(2):134–141, apr 2013.

[18] Stephanie E. Combs, Oliver Jäkel, Thomas Haberer, and Jürgen Debus. Particle therapy at the Heidelberg Ion Therapy center (HIT) - integrated research-driven university-hospital-based radiation oncology service in Heidelberg, Germany. *Radiotherapy and Oncology*, 95(1):41–44, apr 2010.

[19] Abdallah Qubala, Andrea Schwahofer, Stefan Jersemann, Saleh Eskandarian, Semi Harrabi, Patrick Naumann, Marcus Winter, Malte Ellerbrock, Jehad Shafee, Samira Abtehi, Klaus Herfarth, Jürgen Debus, and Oliver Jäkel. Optimizing the patient positioning workflow of patients with pelvis, limb, and chest/spine tumors at an ion-beam gantry based on optical surface guidance. *Advances in Radiation Oncology*, 8(2):101105, mar 2023.

[20] Markus Stock, Dietmar Georg, Alexander Ableitinger, Andrea Zechner, Alexander Utz, Marta Mumot, Gabriele Kragl, Johannes Hopfgartner, Joanna Gora, Till Böhlen, Loïc Grevillot, Peter Kuess, Phil Steininger, Heinz Deutschmann, and Stanislav Vatnitsky. The technological basis for adaptive ion beam therapy at MedAustron: Status and outlook. *Zeitschrift für Medizinische Physik*, 28(3):196–210, aug 2018.

[21] S. Rossi. The status of CNAO. *The European Physical Journal Plus*, 126(8), aug 2011.

[22] Roberto Orecchia, Viviana Vitolo, Maria Rosaria Fiore, Piero Fossati, Alberto Iannalfi, Barbara Vischioni, Anurita Srivastava, Jeffrey Tuan, Mario Ciocca, Silvia Molinelli, Alfredo Mirandola, Gloria Vilches, Andrea Mairani, Barbara Tagaste, Marco Riboldi, Giulia Fontana, Guido Baroni, Sandro Rossi, and Marco Krengli. Proton beam radiotherapy: report of the first ten patients treated at the "Centro Nazionale di Adroterapia Oncologica (CNAO)" for skull base and spine tumours. *La radiologia medica*, 119(4):277–282, dec 2013.

[23] M. Desplanques, B. Tagaste, G. Fontana, A. Pella, M. Riboldi, G. Fattori, A. Donno, G. Baroni, and R. Orecchia. A comparative study between the imaging system and the optical tracking system in proton therapy at CNAO. *Journal of Radiation Research*, 54(suppl 1):i129–i135, jul 2013.

[24] M. Desplanques, S. Rossi, P. Fossati, M. Ciocca, R. Orecchia, M. Riboldi, G. Sharp, and G. Baroni. Technical and medical status of the hadrontherapy facility CNAO, sited in pavia (IT), after a three-year experience. *Physica Medica*, 31:e37, nov 2015.

[25] B. Ohnesorge, T. Flohr, K. Schwarz, J. P. Heiken, and K. T. Bae. Efficient correction for CT image artifacts caused by objects extending outside the scan field of view. *Medical Physics*, 27(1):39–46, jan 2000.

[26] Ge Wang. X-ray micro-CT with a displaced detector array. *Medical Physics*, 29(7):1634–1636, jun 2002.

[27] Paul S. Cho, Anthony D. Rudd, and Roger H. Johnson. Cone-beam CT from width-truncated projections. *Computerized Medical Imaging and Graphics*, 20(1):49–57, jan 1996.

[28] Peter M. Joseph and Robin D. Spital. The effects of scatter in x-ray computed tomography. *Medical Physics*, 9(4):464–472, jul 1982.

[29] R. Schulze, U. Heil, D. Groß, D. D. Bruellmann, E. Dranischnikow, U. Schwanecke, and E. Schoemer. Artefacts in CBCT: a review. *Dentomaxillofacial Radiology*, 40(5):265–273, jul 2011.

[30] Nadine Waltrich, Stefan Sawall, Joscha Maier, Jan Kuntz, Kai Stannigel, Kai Lindenberg, and Marc Kachelrieß. Effect of detruncation on the accuracy of Monte Carlo-based scatter estimation in truncated CBCT. *Medical Physics*, 45(8):3574–3590, jul 2018.

[31] Joscha Maier, Elias Eulig, Tim Vöth, Michael Knaup, Jan Kuntz, Stefan Sawall, and Marc Kachelrieß. Real-time scatter estimation for medical CT using the deep scatter estimation: method and robustness analysis with respect to different anatomies, dose levels, tube voltages, and data truncation. *Medical physics*, 46(1):238–249, 2019.

[32] Christopher Kurz, Florian Kamp, Yang-Kyun Park, Christoph Zöllner, Simon Rit, David Hansen, Mark Podesta, Gregory C. Sharp, Minglun Li, Michael Reiner, Jan Hofmaier, Sebastian Neppl, Christian Thieke, Reinoud Nijhuis, Ute Ganswindt, Claus Belka, Brian A. Winey, Katia Parodi, and Guillaume Landry. Investigating deformable image registration and scatter correction for CBCT-based dose calculation in adaptive IMPT. *Medical Physics*, 43(10):5635–5646, sep 2016.

[33] Rune Slot Thing, Uffe Bernchou, Ernesto Mainegra-Hing, Olfred Hansen, and Carsten Brink. Hounsfield unit recovery in clinical cone beam CT images of the thorax acquired for image guided radiation therapy. *Physics in Medicine & Biology*, 61(15):5781–5802, jul 2016.

[34] Valentina Giacometti, Alan R. Hounsell, and Conor K. McGarry. A review of dose calculation approaches with cone beam CT in photon and proton therapy. *Physica Medica*, 76:243–276, aug 2020.

[35] Victoria Y. Yu, Jani Keyrilainen, Sami Suilamo, Ilyes Beslimane, Alex Dresner, Aleksi Halkola, Uulke A. Van der Heide, and Neelam Tyagi. A multi-institutional analysis of a general pelvis continuous hounsfield unit synthetic ct software for radiotherapy. *Journal of applied clinical medical physics*, 22:207–215, Mar 2021.

[36] Christian Velten, Lee Goddard, Kyoungkeun Jeong, Madhur K. Garg, and Wolfgang A. Tomé. Clinical assessment of a novel ring gantry linear accelerator-mounted helical fan-beam kvct system. *Advances in radiation oncology*, 7:100862, Mar-Apr 2022.

[37] Rolf Clackdoyle and Michel Defrise. Tomographic reconstruction in the 21st century. *IEEE Signal Processing Magazine*, 27(4):60–80, jul 2010.

[38] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am. A*, 1(6):612–619, Jun 1984.

[39] R. Fahrig and D. W. Holdsworth. Three-dimensional computed tomographic reconstruction using a c-arm mounted XRII: Image-based correction of gantry motion nonidealities. *Medical Physics*, 27(1):30–38, jan 2000.

[40] Youngbin Cho, Douglas J. Moseley, Jeffrey H. Siewerdsen, and David A. Jaffray. Accurate technique for complete geometric calibration of cone-beam computed tomography systems. *Medical Physics*, 32(4):968–983, mar 2005.

[41] Marta Peroni, Delia Ciardo, Maria Francesca Spadea, Marco Riboldi, Stefania Comi, Daniela Alterio, Guido Baroni, and Roberto Orecchia. Automatic segmentation and online virtualCT in Head-and-Neck adaptive radiation therapy. *International Journal of Radiation Oncology* Biology* Physics*, 84(3):e427–e433, nov 2012.

[42] Catarina Veiga, Jailan Alshaikhi, Richard Amos, Ana Mónica Lourenço, Marc Modat, Sebastien Ourselin, Gary Royle, and Jamie R. McClelland. Cone-beam Computed Tomography and deformable registration-based "dose of the day" calculations for adaptive proton therapy. *International Journal of Particle Therapy*, 2(2):404–414, sep 2015.

[43] Guillaume Landry, Reinoud Nijhuis, George Dedes, Josefine Handrack, Christian Thieke, Guillaume Janssens, Jonathan Orban de Xivry, Michael Reiner, Florian Kamp, Jan J. Wilkens, Chiara Paganelli, Marco Riboldi, Guido Baroni, Ute Ganswindt, Claus Belka, and Katia Parodi. Investigating CT to CBCT image registration for head and neck proton therapy as a tool for daily dose recalculation. *Medical Physics*, 42(3):1354–1366, feb 2015.

[44] Christopher Kurz, George Dedes, Andreas Resch, Michael Reiner, Ute Ganswindt, Reinoud Nijhuis, Christian Thieke, Claus Belka, Katia Parodi, and Guillaume Landry. Comparing cone-beam CT intensity correction methods for dose recalculation in adaptive intensity-modulated photon and proton therapy for head and neck cancer. *Acta Oncologica*, 54(9):1651–1657, jul 2015.

[45] Satoshi Kida, Takahiro Nakamoto, Masahiro Nakano, Kanabu Nawa, Akihiro Haga, Jun'ichi Kotoku, Hideomi Yamashita, and Keiichi Nakagawa. Cone Beam Computed Tomography image quality improvement using a deep convolutional neural network. *Cureus*, apr 2018.

[46] David C. Hansen, Guillaume Landry, Florian Kamp, Minglun Li, Claus Belka, Katia Parodi, and Christopher Kurz. ScatterNet: A convolutional neural network for cone-beam CT intensity correction. *Medical Physics*, 45(11):4916–4926, oct 2018.

[47] Guillaume Landry, David Hansen, Florian Kamp, Minglun Li, Ben Hoyle, Jochen Weller, Katia Parodi, Claus Belka, and Christopher Kurz. Comparing unet training with three different datasets to correct CBCT images for prostate radiotherapy dose calculations. *Physics in Medicine & Biology*, 64(3):035011, jan 2019.

[48] Chiara Paganelli, Giorgia Meschini, Silvia Molinelli, Marco Riboldi, and Guido Baroni. Patient-specific validation of deformable image registration in radiation therapy: overview and caveats. *Medical physics*, 45(10):e908–e922, 2018.

[49] J. H. Siewerdsen, D. J. Moseley, B. Bakhtiar, S. Richard, and D. A. Jaffray. The influence of antiscatter grids on soft-tissue detectability in cone-beam computed tomography with flat-panel detectors. *Medical Physics*, 31(12):3506–3520, nov 2004.

[50] Lei Zhu, Yaoqin Xie, Jing Wang, and Lei Xing. Scatter correction for cone-beam CT in radiation therapy. *Medical Physics*, 36(6Part1):2258–2268, may 2009.

[51] M. Sun and J. M. Star-Lack. Improved scatter correction using adaptive scatter kernel superposition. *Physics in Medicine & Biology*, 55(22):6695–6720, oct 2010.

[52] A. Sisniega, W. Zbijewski, A. Badal, I. S. Kyprianou, J. W. Stayman, J. J. Vaquero, and J. H. Siewerdsen. Monte Carlo study of the effects of system geometry and antiscatter grids on cone-beam CT scatter distributions. *Medical Physics*, 40(5):051915, apr 2013.

[53] Uros Stankovic, Lennert S. Ploeger, Marcel van Herk, and Jan-Jakob Sonke. Optimal combination of anti-scatter grids and software correction for CBCT imaging. *Medical Physics*, 44(9):4437–4451, aug 2017.

[54] Joseph Harms, Yang Lei, Tonghe Wang, Rongxiao Zhang, Jun Zhou, Xiangyang Tang, Walter J. Curran, Tian Liu, and Xiaofeng Yang. Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography. *Medical Physics*, 46(9):3998–4009, jul 2019.

[55] Christoph Zöllner, Simon Rit, Christopher Kurz, Gloria Vilches-Freixas, Florian Kamp, George Dedes, Claus Belka, Katia Parodi, and Guillaume Landry. Decomposing a prior-CT-based cone-beam CT projection correction algorithm into scatter and beam hardening components. *Physics and Imaging in Radiation Oncology*, 3:49–52, jul 2017.

[56] Tadanori Abe, Kunihiko Tateoka, Yuichi Saito, Takuya Nakazawa, Masaki Yano, Kensei Nakata, Masanori Someya, Masakazu Hori, and Koichi Sakata. Method for converting Cone-Beam CT values into Hounsfield Units for radiation treatment planning. *International Journal of Medical Physics, Clinical Engineering and Radiation Oncology*, 06(04):361–375, 2017.

[57] Halima Saadia Kidar and Hacene Azizi. Enhancement of Hounsfield unit distribution in cone-beam CT images for adaptive radiation therapy: evaluation of a hybrid correction approach. *Physica Medica*, 69:269–274, jan 2020.

[58] Tianye Niu, Mingshan Sun, Josh Star-Lack, Hewei Gao, Qiyong Fan, and Lei Zhu. Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images. *Medical Physics*, 37(10):5395–5406, sep 2010.

[59] W. Zbijewski and F. J. Beekman. Efficient Monte Carlo based scatter artifact reduction in cone-beam micro-CT. *IEEE Transactions on Medical Imaging*, 25(7):817–827, jul 2006.

[60] G. J. Bootsma, F. Verhaegen, and D. A. Jaffray. Efficient scatter distribution estimation and correction in CBCT using concurrent Monte Carlo fitting. *Medical Physics*, 42(1):54–68, dec 2014.

[61] Yuan Xu, Ti Bai, Hao Yan, Luo Ouyang, Arnold Pompos, Jing Wang, Linghong Zhou, Steve B. Jiang, and Xun Jia. A practical cone-beam CT scatter correction method with optimized monte carlo simulations for image-guided radiation therapy. *Physics in Medicine & Biology*, 60(9):3567–3587, apr 2015.

[62] Wei Zhao, Don Vernekohl, Jun Zhu, Luyao Wang, and Lei Xing. A model-based scatter artifacts correction for cone beam CT. *Medical Physics*, 43(4):1736–1753, mar 2016.

[63] Branimir Rusanov, Ghulam Mubashar Hassan, Mark Reynolds, Mahsheed Sabet, Jake Kendrick, Pejman Rowshanfarzad, and Martin Ebert. Deep learning methods for enhancing cone-beam CT image quality toward adaptive radiation therapy: A systematic review. *Medical Physics*, 49(9):6019–6054, jul 2022.

[64] Yangkang Jiang, Chunlin Yang, Pengfei Yang, Xi Hu, Chen Luo, Yi Xue, Lei Xu, Xiuhua Hu, Luhan Zhang, Jing Wang, Ke Sheng, and Tianye Niu. Scatter correction of cone-beam CT using a deep residual convolution neural network (DRCNN). *Physics in Medicine & Biology*, 64(14):145003, jul 2019.

[65] Joscha Maier, Stefan Sawall, Michael Knaup, and Marc Kachelrieß. Deep Scatter Estimation (DSE): Accurate real-time scatter estimation for x-ray CT using a deep convolutional neural network. *Journal of Nondestructive Evaluation*, 37(3), jul 2018.

[66] Brent van der Heyden, Martin Uray, Gabriel Paiva Fonseca, Philipp Huber, Defne Us, Ivan Messner, Adam Law, Anastasiia Parii, Niklas Reisz, Ilaria Rinaldi, Gloria Vilches Freixas, Heinz Deutschmann, Frank Verhaegen, and Philipp Steininger. A Monte Carlo based scatter removal method for non-isocentric cone-beam CT acquisitions using a deep convolutional autoencoder. *Physics in Medicine & Biology*, 65(14):145002, jul 2020.

[67] Liyuan Chen, Xiao Liang, Chenyang Shen, Steve Jiang, and Jing Wang. Synthetic CT generation from CBCT images via deep learning. *Medical Physics*, 47(3):1115–1125, jan 2020.

[68] Matteo Rossi, Gabriele Belotti, Chiara Paganelli, Andrea Pella, Amelia Barcellini, Pietro Cerveri, and Guido Baroni. Image-based shading correction for narrow-FOV truncated pelvic CBCT with deep convolutional neural networks and transfer learning. *Medical physics*, 48:7112–7126, 2021.

[69] Satoshi Kida, Shizuo Kaji, Kanabu Nawa, Toshikazu Imae, Takahiro Nakamoto, Sho Ozaki, Takeshi Ohta, Yuki Nozawa, and Keiichi Nakagawa. Visual enhancement of Cone-beam CT by use of CycleGAN. *Medical Physics*, 47(3):998–1010, jan 2020.

[70] Miriam Eckl, Lea Hoppen, Gustavo R. Sarria, Judit Boda-Heggemann, Anna Simeonova-Chergou, Volker Steil, Frank A. Giordano, and Jens Fleckenstein. Evaluation of a cycle-generative adversarial network-based cone-beam CT to synthetic CT conversion algorithm for adaptive radiation therapy. *Physica Medica*, 80:308–316, dec 2020.

[71] Guoya Dong, Chenglong Zhang, Xiaokun Liang, Lei Deng, Yulin Zhu, Xuanyu Zhu, Xuanru Zhou, Liming Song, Xiang Zhao, and Yaoqin Xie. A deep unsupervised learning model for artifact correction of pelvis cone-beam CT. *Frontiers in Oncology*, 11, jul 2021.

[72] Hongfei Sun, Rongbo Fan, Chunying Li, Zhengda Lu, Kai Xie, Xinye Ni, and Jianhua Yang. Imaging study of pseudo-CT synthesized from cone-beam CT based on 3D CycleGAN in radiotherapy. *Frontiers in Oncology*, 11, mar 2021.

[73] Jinsoo Uh, Chuang Wang, Sahaja Acharya, Matthew J. Krasin, and Chia ho Hua. Training a deep neural network coping with diversities in abdominal and pelvic images of children and young adults for CBCT-based adaptive proton therapy. *Radiotherapy and Oncology*, 160:250–258, jul 2021.

[74] Jun Zhao, Zhi Chen, Jiazhou Wang, Fan Xia, Jiayuan Peng, Yiwen Hu, Weigang Hu, and Zhen Zhang. MV CBCT-based synthetic CT generation using a deep learning method for rectal cancer adaptive radiotherapy. *Frontiers in Oncology*, 11, may 2021.

[75] Shipeng Xie, Yingjuan Liang, Tao Yang, and Zhenrong Song. Contextual loss based artifact removal method on CBCT image. *Journal of Applied Clinical Medical Physics*, 21(12):166–177, nov 2020.

[76] Wangjiang Wu, Junda Qu, Jing Cai, and Ruijie Yang. Multiresolution residual deep neural network for improving pelvic CBCT image quality. *Medical Physics*, 49(3):1522–1534, jan 2022.

[77] Yang Zhang, Ning Yue, Min-Ying Su, Bo Liu, Yi Ding, Yongkang Zhou, Hao Wang, Yu Kuang, and Ke Nie. Improving CBCT quality to CT level using deep learning with generative adversarial network. *Medical Physics*, 48(6):2816–2826, may 2021.

[78] W Lu, H Yan, L Zhou, L Cervino, S Jiang, and X Jia. TU-g-141-05: Limited field-of-view cone-beam CT reconstruction for adaptive radiotherapy. *Medical Physics*, 40(6Part27):457–457, jun 2013.

[79] George X. Ding, Dennis M. Duggan, Charles W. Coffey, Matthew Deeley, Dennis E. Hallahan, Anthony Cmelak, and Arnold Malcolm. A study on adaptive IMRT treatment planning using kV cone-beam CT. *Radiotherapy and Oncology*, 85(1):116–125, oct 2007.

[80] Tao Lei, Risheng Wang, Yong Wan, Xiaogang Du, Hongying Meng, and Asoke K. Nandi. Medical image segmentation using deep learning: a survey. *arXiv preprint arXiv:2009.13120*, 2020.

[81] Tanzila Saba. Recent advancement in cancer detection using machine learning: systematic survey of decades, comparisons and challenges. *Journal of Infection and Public Health*, 13(9):1274–1289, sep 2020.

[82] Francesca Albertini, Michael Matter, Lena Nenoff, Ye Zhang, and Antony Lomax. Online daily adaptive proton therapy. *The British Journal of Radiology*, 93(1107):20190594, 2020.

[83] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, jul 2017.

[84] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *2017 IEEE International Conference on Computer Vision (ICCV)*. IEEE, oct 2017.

[85] Xiao Liang, Liyuan Chen, Dan Nguyen, Zhiguo Zhou, Xuejun Gu, Ming Yang, Jing Wang, and Steve Jiang. Generating synthesized computed tomography (CT) from cone-beam computed tomography (CBCT) using CycleGAN for adaptive radiation therapy. *Physics in Medicine & Biology*, 64(12):125002, jun 2019.

[86] Christopher Kurz, Matteo Maspero, Mark H. F. Savenije, Guillaume Landry, Florian Kamp, Marco Pinto, Minglun Li, Katia Parodi, Claus Belka, and Cornelis A. T. van den Berg. CBCT correction using a cycle-consistent generative adversarial network and unpaired training to enable photon and proton dose calculation. *Physics in Medicine & Biology*, 64(22):225004, nov 2019.

[87] Hui-Ju Tien, Hsin-Chih Yang, Pei-Wei Shueng, and Jyh-Cheng Chen. Cone-beam CT image quality improvement using cycle-deblur consistent adversarial networks (cycle-deblur GAN) for chest CT imaging in breast cancer patients. *Scientific Reports*, 11(1), jan 2021.

[88] Dong Nie, Roger Trullo, Jun Lian, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with context-aware generative adversarial networks. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2017*, pages 417–425. Springer International Publishing, 2017.

[89] Yafen Li, Wen Li, Jing Xiong, Jun Xia, and Yaoqin Xie. Comparison of supervised and unsupervised deep learning methods for medical image synthesis between Computed Tomography and Magnetic Resonance images. *BioMed Research International*, 2020:1–9, nov 2020.

[90] Xue Dong, Yang Lei, Tonghe Wang, Kristin Higgins, Tian Liu, Walter J. Curran, Hui Mao, Jonathon A. Nye, and Xiaofeng Yang. Deep learning-based attenuation correction in the absence of structural information for whole-body positron emission tomography imaging. *Physics in Medicine & Biology*, 65(5):055011, 2020.

[91] Afua A. Yorke, Gary C. McDonald, David Solis, and Thomas Guerrero. Pelvic reference data [dataset], 2019.

[92] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation, 2015.

[93] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[94] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[95] Alain Horé and Djemel Ziou. Image quality metrics: PSNR vs. SSIM. In *2010 20th International Conference on Pattern Recognition*, pages 2366–2369, 2010.

[96] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, apr 2004.

[97] François Chollet et al. Keras. `https://keras.io`, 2015.

[98] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. TensorFlow: Large-scale machine learning on heterogeneous systems, 2015. Software available from tensorflow.org.

[99] Diederik P. Kingma and Jimmy Ba. Adam: a method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[100] Guan hao Chen, Chun ling Yang, and Sheng li Xie. Gradient-based structural similarity for image quality assessment. In *2006 International Conference on Image Processing*. IEEE, oct 2006.

[101] Chaofeng Li and Alan C. Bovik. Content-partitioned structural similarity index for image quality assessment. *Signal Processing: Image Communication*, 25(7):517–526, aug 2010.

[102] Faisal Mahmood, Richard Chen, Sandra Sudarsky, Daphne Yu, and Nicholas J. Durr. Deep learning with cinematic rendering: fine-tuning deep neural networks using photorealistic medical images. *Physics in Medicine & Biology*, 63(18):185012, sep 2018.

[103] Maayan Frid-Adar, Idit Diamant, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. GAN-based synthetic medical image augmentation for increased CNN performance in liver lesion classification. *Neurocomputing*, 321:321–331, dec 2018.

[104] S. Rit, M. Vila Oliva, S. Brousmiche, R. Labarbe, D. Sarrut, and G. C. Sharp. The Reconstruction Toolkit (RTK), an open-source cone-beam CT reconstruction toolkit based on the Insight Toolkit (ITK). *Journal of Physics: Conference Series*, 489:012079, mar 2014.

[105] Yang-Kyun Park, Gregory C. Sharp, Justin Phillips, and Brian A. Winey. Proton dose calculation on scatter-corrected CBCT image: feasibility study for adaptive proton therapy. *Medical Physics*, 42(8):4449–4459, jul 2015.

[106] Matthew McCormick, Xiaoxiao Liu, Julien Jomier, Charles Marion, and Luis Ibanez. ITK: enabling reproducible research and open science. *Frontiers in Neuroinformatics*, 8, 2014.

[107] Connor Shorten and Taghi M. Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1), jul 2019.

[108] Marta Gherardini, Evangelos Mazomenos, Arianna Menciassi, and Danail Stoyanov. Catheter segmentation in x-ray fluoroscopy using synthetic data and transfer learning with light u-nets. *Computer Methods and Programs in Biomedicine*, 192:105420, aug 2020.

[109] Matteo Rossi and Pietro Cerveri. Comparison of supervised and unsupervised approaches for the generation of synthetic CT from cone-beam CT. *Diagnostics (Basel, Switzerland)*, 11, 2021.

[110] Jun Hong, Marsha Reyngold, Christopher Crane, John Cuaron, Carla Hajj, Justin Mann, Melissa Zinovoy, Ellen Yorke, Eve LoCastro, Aditya P. Apte, and Gig Mageras. Breath-hold ct and cone-beam ct images with expert manual organ-at-risk segmentations from radiation treatments of locally advanced pancreatic cancer (Pancreatic-CT-CBCT-SEG), 2021.

[111] G. Poludniowski, P. M. Evans, V. N. Hansen, and S. Webb. An efficient Monte Carlo-based algorithm for scatter correction in keV cone-beam CT. *Physics in Medicine & Biology*, 54(12):3847–3864, jun 2009.

[112] S Jan, G Santin, D Strul, S Staelens, K Assié, D Autret, S Avner, R Barbier, M Bardiès, P M Bloomfield, D Brasse, V Breton, P Bruyndonckx, I Buvat, A F Chatziioannou, Y Choi, Y H Chung, C Comtat, D Donnarieix, L Ferrer, S J Glick, C J Groiselle, D Guez, P-F Honore, S Kerhoas-Cavata, A S Kirov, V Kohli, M Koole, M Krieguer, D J van der Laan, F Lamare, G Largeron, C Lartizien, D Lazaro, M C Maas, L Maigne, F Mayet, F Melot, C Merheb, E Pennacchio, J Perez, U Pietrzyk, F R Rannou, M Rey, D R Schaart, C R Schmidtlein, L Simon, T Y Song, J-M Vieira, D Visvikis, R Van de Walle, E Wieërs, and C Morel. GATE: a simulation toolkit for PET and SPECT. *Physics in Medicine & Biology*, 49(19):4543–4561, sep 2004.

[113] Gavin Poludniowski, Artur Omar, Robert Bujila, and Pedro Andreo. Technical note: SpekPy v2.0—a software toolkit for modeling x-ray tube spectra. *Medical Physics*, 48(7):3630–3637, jun 2021.

[114] Siddharth Sharma, Simone Sharma, and Anidhya Athaiya. Activation functions in neural networks. *International Journal of Engineering Applied Sciences and Technology*, 04(12):310–316, may 2020.

[115] Maria Francesca Spadea, Matteo Maspero, Paolo Zaffino, and Joao Seco. Deep learning based synthetic-CT generation in radiotherapy and PET: A review. *Medical Physics*, 48(11):6537–6566, sep 2021.

[116] Hans-Peter Wieser, Eduardo Cisternas, Niklas Wahl, Silke Ulrich, Alexander Stadler, Henning Mescher, Lucas-Raphael Müller, Thomas Klinge, Hubert Gabrys, Lucas Burigo, Andrea Mairani, Swantje Ecker, Benjamin Ackermann, Malte Ellerbrock, Katia Parodi, Oliver Jäkel, and Mark Bangert. Development of the open-source dose calculation and optimization toolkit matRad. *Medical Physics*, 44(6):2556–2568, may 2017.

[117] Constantin Dreher, Daniel Habermehl, Swantje Ecker, Stephan Brons, Rami El-Shafie, Oliver J äkel, Jürgen Debus, and Stephanie E. Combs. Optimization of carbon ion and proton treatment plans using the raster-scanning technique for patients with unresectable pancreatic cancer. *Radiation Oncology*, 10(1), nov 2015.

[118] Moyed Miften, Arthur Olch, Dimitris Mihailidis, Jean Moran, Todd Pawlicki, Andrea Molineu, Harold Li, Krishni Wijesooriya, Jie Shi, Ping Xia, Nikos Papanikolaou, and Daniel A. Low. Tolerance limits and methodologies for IMRT measurement-based verification QA: Recommendations of AAPM task group no. 218. *Medical Physics*, 45(4), mar 2018.

[119] Adrian Thummerer, Carmen Seller Oria, Paolo Zaffino, Arturs Meijers, Gabriel Guterres Marmitt, Robin Wijsman, Joao Seco, Johannes Albertus Langendijk, Antje-Christin Knopf, Maria Francesca Spadea, and Stefan Both. Clinical suitability of deep learning based synthetic CTs for adaptive proton therapy of lung cancer. *Medical Physics*, 48(12):7673–7684, nov 2021.