Dissertation

submitted to the
Combined Faculty of Mathematics, Engineering and Natural Sciences
of Heidelberg University, Germany
for the degree of
Doctor of Natural Sciences

Put forward by

M. Sc. Fabian Jäger

Born in: Munich, Germany

Oral examination: 22nd of January, 2025.

Data-driven Image Quality Improvements for Cone-Beam Computed Tomography in Radiation Therapy

Referees: Prof. Dr. Joao Seco
Prof. Dr. Marc Kachelrieß

Datengetriebene Bildqualitätsverbesserungen für die Kegelstrahl-Computertomographie in der Strahlentherapie

Bei der Strahlentherapie werden ionisierende Strahlen eingesetzt, um den Tumor gezielt zu behandeln. Um gesundes Gewebe zu schonen, wird der Behandlungsplan anhand eines Computertomographie-Bildes (CT) optimiert. Kegelstrahl-CT-Bilder (CBCT) können aufgrund ihrer unzureichenden Bildqualität nicht für eine Planoptimierung verwendet werden. Das Ziel dieser Arbeit ist es, zwei Hauptartefakte zu reduzieren, die für den Verlust der Bildqualität verantwortlich sind: Streuung und Strahlenaufhärtung. Neuartige Deep-Learning-basierte Methoden werden für die Korrektur von Projektionen angepasst und entwickelt und mit Vergleichsmethoden moderner CBCT-Scanner verglichen. Hier wird ein deterministischer Löser der linearen Boltzmann-Gleichung verwendet, um Trainingsdaten für die Deep Scatter Estimation (DSE) zu erzeugen. Die vorgeschlagene tiefe Strahlenhärtungskorrektur wurde so konzipiert, dass sie sowohl die Beiträge von Knochen als auch von Weichgewebe berücksichtigt, was einen Vorteil gegenüber der üblicherweise angewandten Wasservorkorrekturmethode darstellt, die nur ein einzelnes Material berücksichtigt. Die DSE reduziert den mittleren absoluten Fehler in Test-Scans um ca. 96 %, übertrifft die projektionsbasierte Referenzmethode in der Bildqualität und ist über 29 Mal schneller als die Referenzmethode im Bildraum. Die vorgeschlagene Strahlenhärtungskorrektur reduziert den Fehler im Knochen deutlich und verringert den Restfehler der Wasservorkorrektur um weitere 40 %.

Data-driven Image Quality Improvements for Cone-Beam Computed Tomography in Radiation Therapy

In radiotherapy, ionizing radiation is used to accurately treat tumors. To spare healthy tissue the treatment plan is optimized on a computed tomography (CT) image. On-board cone-beam CT (CBCT) images cannot be used for a daily-updated plan because of their insufficient image quality. The aim of this thesis is to reduce two artifacts responsible for image quality loss: scatter and beam hardening. Novel deep learning-based methods are adapted and developed for the correction of projections, and are compared to correction methods of state-of-the-art CBCT scanners. Here, a deterministic solver of the linear Boltzmann equation is used to generate training data for the deep scatter estimation (DSE). The proposed deep beam hardening corrections were designed to incorporate the contributions from bones as well as soft tissue, bringing an advantage to the commonly applied water precorrection method, which only considers a single material. DSE reduces the mean absolute error in test scans by approximately 96 %, outperforming the projection-based reference method in image quality, and is over 29 times faster than the reference in image domain. The proposed beam hardening correction significantly reduces the error in bone, with the remaining error reduced by 40 % in comparison to the water precorrection.

Contents

List of Acronyms

1D . . . . . One-dimensional

2D . . . . . Two-dimensional

3D . . . . . Three-dimensional

APE . . . . . Absolute percentage error

CBCT . . . . . Cone-beam computed tomography

CNN . . . . . Convolutional neural network

CoM . . . . . Center of mass

CT . . . . . Computed tomography

CUDA . . . . . Compute unified device architecture

DBB . . . . . Deep bone blending

DBE . . . . . Deep bone extraction

DECT . . . . . Dual energy computed tomography

DKFZ . . . . . German Cancer Research Center

DL . . . . . Deep learning

DSE . . . . . Deep scatter estimation

fASKS . . . . . Fast adaptive scatter kernel superposition

FBP . . . . . Filtered backprojection

FDK . . . . . Feldkamp-David-Kress

FoM . . . . . Field of measurement

GAN . . . . . Generative adversarial network

GPU . . . . . Graphics processing unit

GT . . . . . Ground truth

HU . . . . . Hounsfield units

IGRT . . . . . Image-guided radiation therapy

LINAC . . . . . Linear accelerator

LUT . . . . . Look-up table

MAE . . . . . Mean absolute error

MAPE . . . . . Mean absolute percentage error

MC . . . . . Monte Carlo

ML . . . . . Machine learning

MLP . . . . . Multilayer perceptron

MSE . . . . . Mean squared error

NN . . . . . Neural network

ReLU . . . . . Rectified linear unit

SPMAPE . . . . . Scatter-to-primary-weighted mean absolute percentage error

SSIM . . . . . Structural similarity index metric

VMI . . . . . Virtual monochromatic image

List of Tables

List of Figures

1 Introduction

Computed tomography (CT), first used to image a patient in 1971 [1], is a powerful tool to derive cross-sectional images of the patient's anatomy. Starting as a method to image a single 2D slice of the patient, full three-dimensional (3D) body scans are nowadays possible. This distinguishes it from conventional X-ray imaging, which provides only 2D transmission images. As a non-invasive imaging method, it has improved patient diagnosis and treatment in medicine. Not only did this lead to a rapidly growing number of CTs in hospitals (today, most clinics operate at least one CT) but it also led to the award of the Nobel Prize in Physiology or Medicine to Allan M. Cormack and Godfrey N. Hounsfield in 1979.

Since 1972, there has been an ongoing effort by the scientific community and industry to improve CT systems and develop new applications. In the beginning, a small detector, consisting of only two detector elements, and a needle beam were used to raster scan the field of measurement (FoM) [2]. In contrast, modern CT systems acquire data from hundreds of thousands of detector elements simultaneously. Moreover, the source and the detector rotate to acquire hundreds of projections from different directions in a short time. State-of-the-art clinical CT scanners have a rotation time of less than 0.25 s [3], [4]. CT scanners are not only used for diagnostic but also for a wide range of other applications, such as image guidance during a stent placement [5], image-guided radiation therapy (IGRT) [6], [7], dental scans [8], and are important for the optimization process of radiation treatment plans in radiation therapy [6]. As treatment planning is based on the CT images, the quantitative accuracy is vitally important to guarantee optimal radiotherapy treatment.

There are different types of CT scanner. A key distinction between cone-beam computed tomography (CBCT) systems and clinical CT scanner is the shape of the detector and the manner of data acquisition. Clinical CTs, also called fan beam CTs, utilize a narrow, fan-shaped X-ray beam and curved detector(s) that rotate around the patient in a helical or circular pattern. In the conventional approach, the couch is advanced through the rotating detector and source in order to obtain the helical pattern. CBCTs, with a flat panel detector, on the other hand employ a cone-shaped X-ray beam that captures the entire FoM in a single rotation around the patient. Nowadays, flat detectors can be integrated on other medical systems, including radiation therapy devices [7], [9].

Over half of individuals diagnosed with cancer undergo radiation therapy as a treatment modality. In radiation therapy, the patient is irradiated with ionizing radiation to

disable cancerous cells. Given that this radiation damages healthy tissue as well, a significant amount of work is dedicated to improve the precision of the prescribed dose to the target volume. Currently, a planning CT is conducted with a clinical CT, which is subsequently utilized to generate the treatment plan. A target volume is defined and the plan is optimized with constraints regarding vital organs such as the heart or salivary glands. Commonly, the treatment plan is divided into so-called fractions, with the planned dose not applied in one session but rather spread out over several sessions and weeks. It is noteworthy that the patient anatomy changes over time. For instance, the tumor may shrink or air bubbles move in the abdomen. This can result in a discrepancy between the planning CT and the actual patient anatomy, decreasing treatment accuracy. In modern IGRT, a daily update of the anatomy is considered by acquiring a new CT image on an imaging device built into the treatment unit. It is not possible to use a clinical CT as an on-board imaging device due to mechanical restrictions. However, a CBCT can be used as an on-board system for this purpose. The image quality of a CBCT scanner is inferior to a clinical CT [10], so the update image is not used to optimize the treatment plan, but to register the previously taken clinical CT to the current anatomy [11]. Moreover, errors in the CBCT image can propagate to the registration during the registration process.
The image is distorted by artifacts, which are the result of an imperfect modeling of the physical processes during image reconstruction or mechanical constraints of the projection acquisition. Common artifacts originate from the assumption that the attenuation of the photons is energy-independent while the scanners utilize a polychromatic energy spectrum. This causes the so-called beam hardening artifacts. Another frequently occurring source of artifacts are scattered photons. These scattered photons do not travel on a straight line from the source to a detector pixel. Instead, the intensity of scattered photons is accumulated at other detector elements. Flat detectors, as used in CBCT systems, produce more severe artifacts than clinical CT systems with curved detectors, which are nearly artifact-free. Considerable effort is dedicated to the enhancement of image quality for all scanner types, through mechanical components, such as prefilter and anti-scatter grids [6] or computational post-processing [12]–[18].
In recent times, there has been considerable interest in machine learning and a plethora of deep learning applications. Large models impact daily life and are accessible to the general public. Examples of such models include ChatGPT [19], text-to-image applications [20], and image segmentation and classification tools [21], [22]. Moreover, novel deep learning algorithms have demonstrated considerable potential in a diverse array of image processing tasks. In the context of medical imaging, this led to advances in several application, such as image segmentation [22], [23], artifact correction in CT [14], [15], [24]–[26], image reconstruction [27], [28] or tool extraction in fluoroscopy [29]. Recently, deep learning-based methods found their way into commercial products in medical imaging, including artifact reduction [30] in CT image reconstruction.
This work aims at improving CBCT image quality by adapting and developing new data-driven methods to correct artifacts. It is essential to implement fast algorithms to reduce the time between image acquisition and treatment. This need is especially important for the emerging field of online treatment. Consequently, deep learning-based

corrections of the projections, which can be applied in real time [24], is a promising avenue of research. This work focuses on the two severe artifacts in CBCT, namely: beam hardening and scatter. Firstly, a deep learning-based approach to correct for scatter is adapted for an on-board CBCT scanner, the Ethos, used in radiotherapy. The Ethos is a device that combines a linear accelerator (LINAC) with an on-board CBCT scanner. It is currently in clinical use in radiotherapy facilities around the world. For this purpose, a deep neural network is trained on data not generated by Monte Carlo simulations as in many previous studies [14], [24], [25], [31] but with Acuros, a deterministic solver of the linear Boltzmann transport equation [13], [32]. To be used routinely, the proposed approach was optimized to fit various clinical settings including different anatomical regions. In a second step, the efficiency of data-driven methods to account for beam hardening are investigated. A novel approach was developed that makes use of a deep learning-based segmentation within a physically motivated correction scheme. This approach was designed, to incorporate the contributions from bones as well as soft tissue, bringing an advantage to the commonly applied water precorrection method, which only considers a single material.
The structure of this work is as follows: Chapter 2 covers the fundamentals of X-ray computed tomography. It discusses the main physical principles, image reconstruction and common artifacts. In addition, it gives a brief overview of machine learning and deep learning with neural networks (NNs). Then, Chapter 3 focuses on the improvements made to existing methods and presents novel deep learning-based artifact corrections. Chapter 4 will present the results from the proposed methods and their robustness to different clinical settings. This is followed by a summary and discussion in Chapter 5 and finally, conclusions are drawn in Chapter 6.

2 | Fundamentals

X-ray beams pass through objects or patients while being attenuated by interaction with the surrounding tissue. In CT systems, the attenuated photons are measured with a detector behind the object. Therefore, the measured data at the detector are 1D or 2D projections of the spatial attenuation distribution of the 3D object. While X-ray imaging uses a projection directly, CT reconstructs the 3D distribution from many projections. The following sections give a brief overview of the underlying physics, the mechanical components of a CT scanner and the image reconstruction. In addition, the section introduces machine learning and deep learning.

2.1 X-ray Tubes

X-rays in medical imaging applications are usually in the range of 30 keV to 150 keV [6]. Hence, only a fraction of the electromagnetic spectrum which is referred to as X-rays (100 eV to several hundred keV) is used. In CT, X-ray tubes are used as the photon source, which generate photons by accelerating and decelerating electrons. Figure 2.1 illustrates a reflection X-ray tube. The cathode is heated (≈2400 K [33]) to overcome the binding energy between filament and electrons, thereby producing free electrons. A potential difference, U_B, is applied between the cathode and a ring anode, which accelerates the electrons. After accelerating, the electrons are further focused and guided through magnetic fields towards the anode. The internal vacuum of the tube allows electrons to reach the target with a maximum kinetic energy of $T = eU_\mathrm{B}$, where e is the electron charge. Inside the target, the electrons are rapidly decelerated which heats up the target on the point of the incoming beam. To prevent a single point on the target from overheating, the target is rotated, thereby increasing the dispersion of heat and the operational time of the X-ray tube. As a consequence of the considerable heat production, the target must have a high melting point. One material that is frequently utilized is tungsten, which has the additional advantage of a high atomic number, increasing the efficiency for electron-material interactions. When electrons impact on the target, they primarily decelerate through collisions and radiative interactions. The collisions are inelastic Coulomb collisions with bound atomic electrons, which can lead to excitation and ionization. This can create “holes” in inner electron shells which are filled by outer shell electrons. Following the transition to a shell with lower potential energy, the difference is emitted as electromagnetic radiation, which is also known

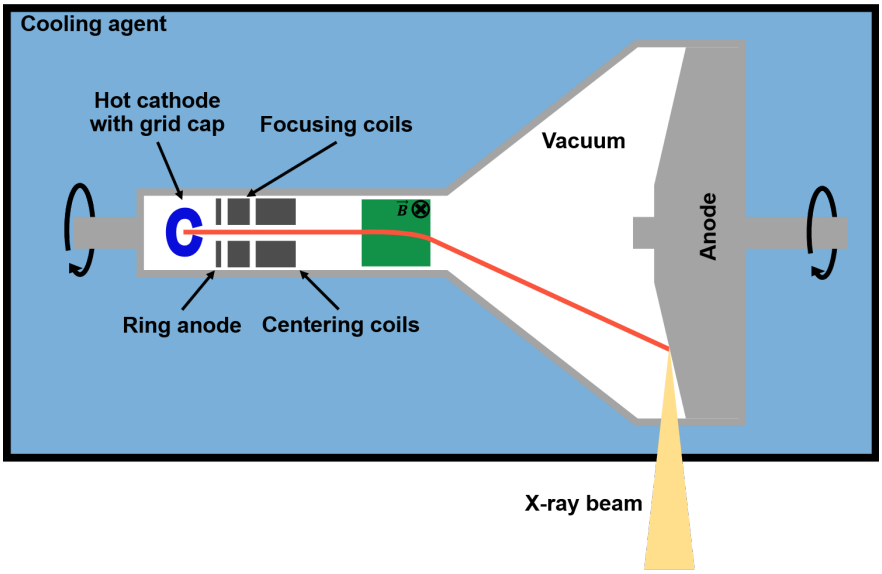

Figure 2.1: Schematic of a reflection X-ray tube in a cooling agent. In addition, the anode rotates for better heat dissipation.

as characteristic radiation due to discrete and material-dependent potential energy differences. The second kind of deceleration is caused by the Coulomb field of the atomic nuclei and is called bremsstrahlung. A continuous spectrum is emitted, whereby the maximum energy corresponds to the kinetic energy of the electrons $T = h\nu_{\max}$ before hitting the target. Considering the spectrum nature of both types of interactions, the emitted spectrum is continuous [34] with distinctive peaks at the energy corresponding to the transition energy between different atomic (sub)shells.

2.2 X-rays Matter Interaction

The process of imaging is dependent upon the attenuation of the photons inside the object or patient placed between the X-ray tube and the detector. The following subsections presents the fundamental principles governing the attenuation of a photon beam.

2.2.1 Beer-Lambert Law

For a monoenergetic X-ray beam of intensity I_0 and an object with a homogeneous attenuation, the two parameters that determine the attenuation are the thickness d of the object and the linear attenuation μ. This is described by the Beer-Lambert law based on the work of Bougert, Lambert and Beer [37]–[39]

$$I = I_0 e^{-\mu \cdot d}\,. \tag{2.1}$$

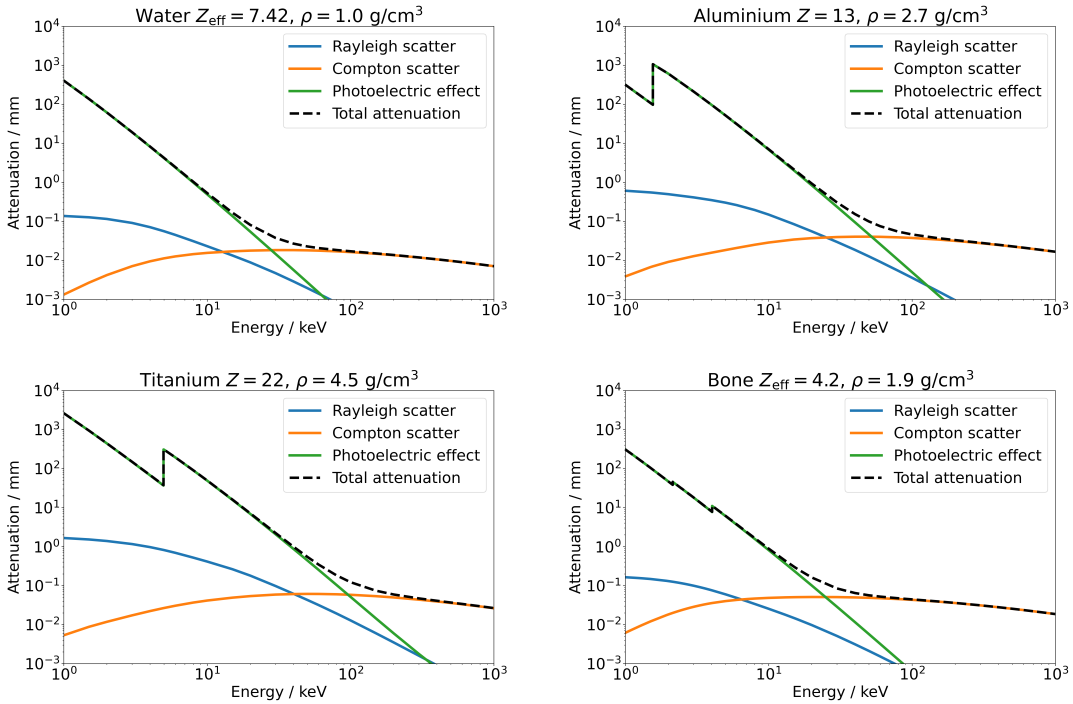

Figure 2.2: The overall attenuation and the contributions from the three primary attenuating interactions for four materials. The attenuation data were taken from [35], the bone composition and density were taken from [36].

Here, I_0 is the intensity before attenuation and I behind the object, after attenuation. It should be noted that the assumptions of a monoenergetic beam and a homogeneous attenuation coefficient do not apply to medical cases where a tube emits a continuous spectrum (see Section 2.1) and the patients consist of different types of tissue and density. Considering the spatial and energy dependence of μ, i.e. $\mu \to \mu(E, \boldsymbol{r})$, Equation (2.1) can be rewritten as

$$I(E) = I_0 e^{-\int_0^d \mu(E, \boldsymbol{s}+\lambda\boldsymbol{\Theta})d\lambda}\,, \tag{2.2}$$

incorporating that the beam starts at the source position $\boldsymbol{s}$ and travels along $\boldsymbol{\Theta}$ with a spatial and energy dependent μ.
The attenuation is the combined effect of different interactions. In the energy range of medical CT scans the total attenuation is governed by three interactions: the photoelectric effect $\mu_{\text{P}}(E)$, the Rayleigh scattering $\mu_{\text{R}}(E)$ and the Compton scattering $\mu_{\text{C}}(E)$. The total attenuation is the sum of all contributions:

$$\mu(E) = \mu_{\text{P}}(E) + \mu_{\text{R}}(E) + \mu_{\text{C}}(E)\,. \tag{2.3}$$

Above 1.022 MeV, pair production, the creation of an electron-positron pair from a

photon, becomes possible as well. For this interaction to occur, the energy of the photon must be above the aforementioned threshold, which is the combined rest mass of an electron and a positron. Therefore, it is not present in medical CT scanners which operate in an energy range far below 1 MeV. Figure 2.2 shows the total attenuation and the contribution of the three primary interactions for four distinct materials at different energies. The most important interactions and their cross-sections, which are used to describe the interaction probability, are explained in more detail in the following subsections. Moreover, the cross-section σ is linked to the attenuation by

$$\mu = \frac{\rho N_\mathrm{A}}{A}\sigma\,. \tag{2.4}$$

Here, ρ is the density, N_A is the Avogadro constant and A is the atomic mass number.

2.2.2 Photoelectric Effect

As illustrated in Figure 2.2, the photoelectric effect is the most prominent interaction for low energies, extending up to 80 keV for water. In this case, the photon is absorbed by an electron, resulting in an excited state. If the binding energy E_b of the electron is smaller than the energy of the photon E_γ, the photon can ionize the atom. The remaining energy is then emitted as kinetic energy $T = E_\gamma - E_\mathrm{b}$, and the remaining kinetic energy of the electron is transferred locally as heat [33]. There is a vacancy in the shell of the ejected electron, which is filled by an electron from an outer shell. As explained before, this emits a (characteristic) photon. The specific binding energy required to remove an electron produces edges in the attenuation that are visible in Figure 2.2 in the shown energy range for all materials except for water.
The dependency on energy and atomic number of the cross-section σ_P is given by

$$\sigma_\mathrm{P} \propto \frac{Z^4}{E^{7/2}}\,. \tag{2.5}$$

This relation remains valid for photons with energies lower than $E_\gamma/T < 0.9$ [40].

2.2.3 Rayleigh Scattering

Rayleigh scattering, also known as elastic or coherent scattering, describes an interaction between a photon and a particle whose size is much smaller than the wavelength of the photon [33]. In this scenario, the photon is deflected by a bound electron. The energy in the resting center of mass (CoM) frame remains the same and only the direction of flight of the photon is changed. Contrary to the photoelectric effect, the electron cannot be excited, because no energy is transferred.

2.2.4 Compton Scattering

In contrast to Rayleigh scattering, in Compton scattering the photon changes not only the direction of its trajectory but also its energy. In addition, the photon does not

interact with a bound but with a quasi-free electron. This, in most cases, is a valence electron in an outer shell. The new energy E'_γ of an incident photon of energy E_γ, which is inelastically scattered with an angle α, is given by:

$$E'_\gamma = \frac{E_\gamma}{1 + \frac{E_\gamma}{m_e c}\left(1 - \cos\alpha\right)}\,, \tag{2.6}$$

with m_e the electron mass. The differential cross-section of the Compton scattering is given by the Klein-Nishina equation [41]:

$$\frac{d\sigma}{d\Omega_{\text{Klein-Nishina}}} = \frac{1}{2}\frac{\alpha^2}{m_0^2}\left(\frac{E'_\gamma}{E_\gamma}\right)^2\left[\frac{E'_\gamma}{E_\gamma} + \frac{E_\gamma}{E'_\gamma} - \sin^2\alpha\right]\,. \tag{2.7}$$

The cross-section can be obtained by integrating Equation (2.7) over all angles [42].

2.3 Detection of X-Rays

After the photon beam has been attenuated by the object, the beam impinges on the detector. Commonly, an anti-scatter grid is placed before the detector to block scattered X-rays. For more information about scatter artifacts the reader is referred to Section 2.6.1. The detected signal relies on the successful detection of incoming X-rays and thus the efficiency of the detector. Most detectors are energy-integrating detectors, but in recent years a new generation of detectors, so-called photon-counting detectors have been introduced to CT systems. Figure 2.3 shows an illustration for the two detectors, both of which are equipped with an anti-scatter grid.
A variety of detector shapes have been developed, each with its own advantages and disadvantages. The first detector built consisted of only two detector pixels [2] which had to be moved with the source to acquire a single 2D projection of the object. Nowadays, detectors can have many rows and columns, with up to millions of pixels. The two most prominent designs feature either curved surfaces like a section cut off a cylinder with a highly optimized gantry for rotation or as flat plates offering an easier mechanical implementation at the cost of increased imaging artifacts. The former is used in clinical CT scanners, while the latter is characteristic for CBCT scanners. In the following, CT systems with a flat detector will be referred to as CBCT and CT scanners with a curved detector as clinical CT, if not explicitly stated otherwise. In addition to the different shapes, there is a variety of detector types, which can be attributed to differences in the materials used and the techniques employed to convert the measured intensity into an electric signal. Figure 2.3 shows the two most common detector types for converting the incoming photons into a measured signal.

2.3.1 Energy-Integrating Detectors

This detector type relies on an indirect conversion from X-rays to the measured signal. Below a thin reflective film, a scintillator crystal converts the incident X-rays into

optical photons. Typically, an inorganic crystal such as gadolinium oxysulfide (Gd_2O_2S) serves as the material of choice. The incoming X-ray excites an electron from the valence band (Compton scattering or photoelectric effect) and leaves a corresponding hole behind. Similar to the photoelectric interaction, only discrete energy transitions exist. Commonly, artificial impurities are added to the crystal to increase the number of energy levels, thereby improving the efficiency. In the case of Gd_2O_2S, this may include the doping with Pr^{4+} and Tb^{4+} [43]. The electron-hole pairs are loosely bound and wander through the lattice. They are captured by impurity centers and an optical photon is released in this process. A photo diode or photomultiplier is used to convert the optical photons to an electric signal. Given the prolonged decay time of each discrete signal, it is only possible to measure the combined signal. Thus, the measured signal represents an integral of all incoming photons over a time interval, with each photon weighted according to its respective energy. Accordingly, this detector type is referred to as energy-integrating.

2.3.2 Photon-Counting Detectors

In contrast, the newer generation of detectors is based on semiconducting sensors that measure electron-hole pairs created by the incoming X-rays. Similar to the previously described process the incident X-ray creates electron-hole pairs. Instead of creating optical photons, a potential difference is applied between the bottom and the top. This accumulates the created charges at the electrodes situated at the bottom. The collected charge is then compared to references in order to estimate the photon energy. In comparison to an energy integrating detector, the readout time is considerably faster, and the detection of energy resolved spectra are possible, offering the potential for material decomposition [44].

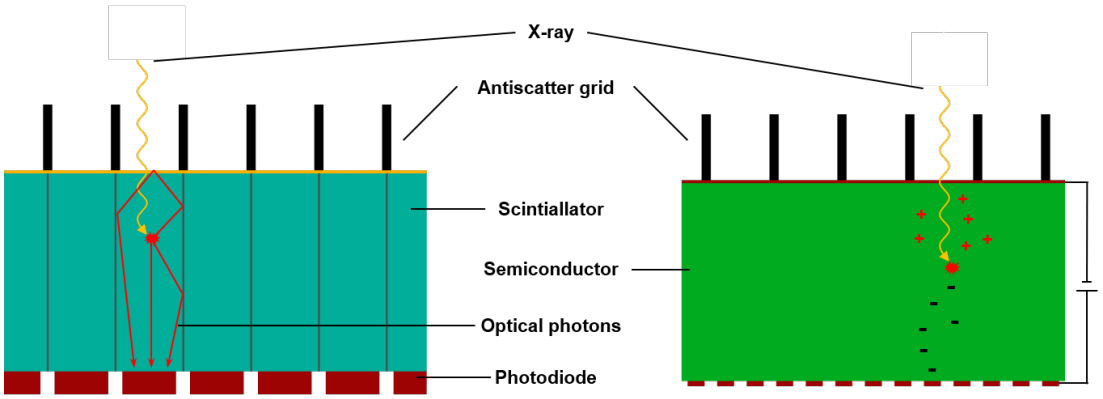

Figure 2.3: Schematic of an energy-integrating detector (left) and a photon-counting detector (right). The latter being a new technology available in clinical CT scanners in recent years increasing the spatial resolution and offering material decomposition from a single source-single detector scan. Note that the anti-scatter grid is not drawn to scale, as in real systems it has a height up to 30 times larger than a detector pixel.

2.4 Raw Data Post-processing

As described in Section 2.2.1, photons are attenuated according to the Beer-Lambert law before reaching the detector. The intensity I measured at the detector is further processed before reconstruction to gain the projection value p:

$$p = -\ln\left(\frac{I}{I_0}\right) . \tag{2.8}$$

The air norm I_0 is measured by acquiring a projection without an attenuating object between source and detector. The projection values p are stored in a sinogram for several rotation angles. It is standard practice to correct the measured intensity I before further processing. This may entail a flat field correction or a defect pixel correction.

2.5 Image Reconstruction

In order to transform the acquired raw data to slice images, an image reconstruction algorithm is necessary. This section shows the mathematical basis for a 2D CT reconstruction with parallel beam geometry as illustrated in Figure 2.4. The approach is based on the work of Johann Radon [45]. As shown in Figure 2.4, ϑ represents the angle between the ray and the x-axis, which is also referred to as the projection angle. The symbol ξ denotes the distance of the ray to the isocenter, which is defined as the origin of the coordinate system. Hence, the equation

$$x\cos\vartheta + y\sin\vartheta - \xi = 0 \tag{2.9}$$

can be formulated to describe a ray of the beam. If the energy dependency of the attenuation is ignored as well as effects such as scatter, the projection value can be related to the attenuation of the scanned object as:

$$p(\vartheta, \xi) = \int \mu(\boldsymbol{s} + \lambda \boldsymbol{\Theta})\, d\lambda \tag{2.10}$$

$$= \int \mu(x, y)\, \delta(x \cos\vartheta + y \sin\vartheta - \xi)\, dxdy \tag{2.11}$$

$$= \int \mu(\boldsymbol{r})\, \delta(\boldsymbol{r} \cdot \boldsymbol{\Theta}_n - \xi)\, d^2r \tag{2.12}$$

$$= \mathsf{R}\mu(\boldsymbol{r}) \ . \tag{2.13}$$

Here, $\delta(\cdot)$ is the Dirac delta function, $\boldsymbol{\Theta}$ the directional vector of a ray, $\boldsymbol{\Theta}_n$ the directional vector normalized to unit length and R the Radon transform operator for two dimensions. For parallel beam geometry, the angular scanning range must be at least 180°. The aim of the reconstruction is to obtain $\mu(\boldsymbol{r})$, the spatial attenuation distribution, from a set of measured projections. For this, Equation (2.10) needs to be inverted. Simplified, this translates to a "smearing" of the projections back onto the image plane. This is done by looping over all projections and then accumulating the values of all intersecting rays at each voxel. An analytical solution can be derived with the Fourier slice theorem. It states that a straight line through the origin of the 2D Fourier transform equals the 1D Fourier transform of the projection. Therefore, the Fourier transform of the projection $p(\vartheta, \xi)$ is given by

$$P(\vartheta, u) = (\mathsf{F}p)(\vartheta, u) \tag{2.14}$$

$$= \int d\xi\, p(\vartheta, \xi)\, e^{-2\pi i u \xi} \tag{2.15}$$

$$= \int dxdy\, \mu(x, y)\, e^{-2\pi i u (x \cos\vartheta + y \sin\vartheta)} \ . \tag{2.16}$$

In accordance with the Fourier theorem, this is equivalent to the 2D Fourier transform of our object

$$F(u_x, u_y) = \int dxdy\, \mu(x, y) e^{-2\pi i (u_x x + u_y y)} \ . \tag{2.17}$$

Comparing Equation (2.14) and Equation (2.17) the relation

$$P(\vartheta, u) = F(u \cos\vartheta, u \sin\vartheta) \tag{2.18}$$

can be found. Due to mechanical constraints, such as a finite pixel size on the detector, the acquired data are discrete. This presents a challenge when attempting to reconstruct using the Fourier-slice theorem, as it requires switching between polar-coordinated and a Cartesian coordinate system (see Equation (2.18)). The discrete and equidistant sampled data in Cartesian coordinates result in an increased sampling close to the origin and a reduced sampling further away in polar coordinates. Thus, information in the low

frequency domain would be lost. The relations of Equation (2.18) and Equation (2.17) result in $u_x = u\cos\vartheta$, $u_y = u\sin\vartheta$ and $du_x du_y = |u| du d\vartheta$. Then, the equation can be written as follows

$$\mu(x,y) = \int_0^\pi d\vartheta \int du\,|u|\,P(\vartheta,u)\,e^{2\pi i u(x\cos\vartheta + y\sin\vartheta)} \tag{2.19}$$

$$= \int_0^\pi d\vartheta \int du\,K(u)\,P(\vartheta,u)\,e^{2\pi i u\xi}\,. \tag{2.20}$$

By using the convolution theorem, according to which the convolution of two functions, defined as

$$(g_1 * g_2)(x) = \int dy\, g_1(y)\, g_2(x-y)\ , \tag{2.21}$$

is equal to the inverse Fourier transform of the multiplication of their respective Fourier transform

$$(g_1 * g_2)(x) = \mathsf{F}^{-1}\left((\mathsf{F}g_1)(\mathsf{F}g_2)\right)\ , \tag{2.22}$$

we can rewrite the multiplication of the Fourier-transformed projection $P(\vartheta,u)$ with $K(u)$ which gives

$$\mu(x,y) = \int_0^\pi d\vartheta\, p(\vartheta,\xi) * k(\xi)\Big|_{\xi = x\cos\vartheta + y\sin\vartheta}\ . \tag{2.23}$$

Evidently, the projections p are filtered with kernel k before being backprojected [or smeared] onto the image plane. This method is therefore called filtered backprojection (FBP). The kernel k is defined as

$$k(\xi) = \mathsf{F}^{-1}K(u) = \int du\,|u|\,e^{2\pi i u\xi} = \frac{-1}{2\pi^2\xi^2}\,. \tag{2.24}$$

The discrete form of the kernel is given for detector pixel m with size $\Delta\xi$ by

$$k_m = \begin{cases} (2\Delta\xi)^{-2} & \text{if } m = 0, \\ -(\pi m\Delta\xi)^{-2} & \text{if } m \in 2\mathbb{Z}+1, \\ 0 & \text{otherwise.} \end{cases} \tag{2.25}$$

This kernel is called Ramachandran–Lakshminarayanan-kernel or RamLak [46]. The utilized filter influences the image properties. Thus, there are many possible filters available, all with benefits as well as disadvantages and the radiologist needs to find a trade-off between sharp edges, more smoothing to reduce the noise, and other image attributes. While FBP is a standard method, available on all CT systems, there are other more sophisticated methods such as iterative reconstructions [33] or deep learning-based reconstructions [30].
In most diagnostic settings, the attenuation values were not relatable for medical

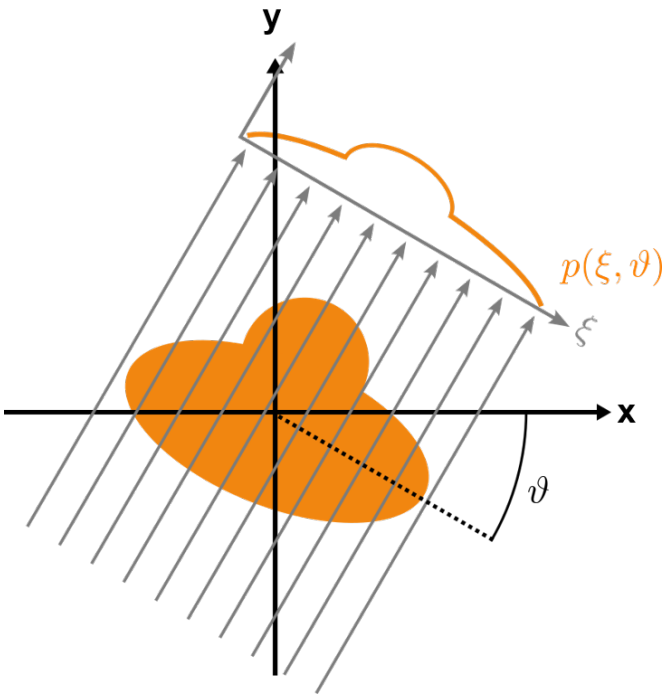

Figure 2.4: Illustration of the geometry for the 2D parallel beam geometry for a filtered backprojection.

personnel. Therefore, the CT values are displayed in Hounsfield units (HU), as defined by the following equation

$$\mathrm{CT}(\boldsymbol{x}) = \frac{\mu(\boldsymbol{x}) - \mu_{\mathrm{ref}}}{\mu_{\mathrm{ref}}} \cdot 1000\,\mathrm{HU}\,, \tag{2.26}$$

in which water at 70 keV is used as reference. Hence by definition, water is at 0 HU and air with $\mu_{\mathrm{air}} = 0\,\mathrm{mm}^{-1}$ at -1000 HU. Other common value ranges include those for fat tissue (between -100 HU and -70 HU), blood (between 30 HU and 60 HU), and cortical bone (between 350 HU and 2000 HU). The images are displayed on a gray scale, but the number of gray values are much larger than the human eye can differentiate [6]. Therefore, only a predefined window of gray values is displayed, written as $C = 0$ HU, $W = 500$ HU for a windowing with the center C and the width W. Values outside of the display window are presented as black or white, contingent upon whether they are below or above the windowing range.

For reconstruction of CBCT scans the simple 2D example has to be extended. Discussing the multitude of available reconstruction algorithms is beyond the scope of this work. The reconstructed images presented in this work are using versions of the Feldkamp-David-Kress (FDK) algorithm [47].

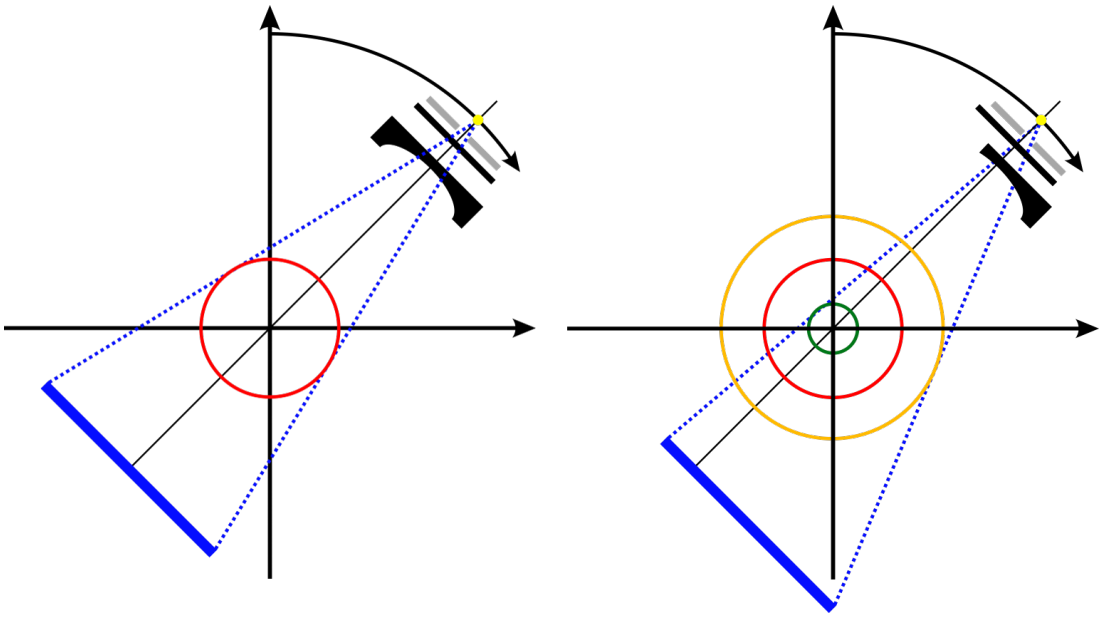

Figure 2.5: Schematic of two different CBCT scanner geometries. The right incorporates a laterally shifted detector, increasing the FoM.

2.6 CBCT Scanner Setup

A CBCT scanner is made from more components than a source and a detector. Figure 2.5 shows two possible scanner setups. The left one features a centered detector, while the right one has a lateral shift applied to its detector. Except for the shifted detector both follow the same principle. First, a collimator restricts the X-ray beam to the FoM to spare as much healthy tissue as possible from radiation. This is applied in longitudinal as well as lateral direction. Typically, different prefilters are placed between the source and the object. The first filter is a flat sheet of a high attenuating material, e.g. titanium, and the second a so-called bowtie filter. Flat prefilters are employed to adjust and shape the spectrum in a manner that is most favorable for the patient. This is accomplished by filtering out low-energy photons, which would otherwise be absorbed by the patient, thereby increasing the received dose while not contributing to the detected signal. Bowtie filters are thinner in the center and are used to modulate the intensity, attenuating more at the edges where the patient is thinner and less in the center where the patient is thicker. The two setups depicted in Figure 2.5 each possess distinct advantages and disadvantages. In both drawings, the red circle indicates the FoM of the centered detector, while the orange circle shows the FoM of the shifted detector. It is evident that the FoM of the shifted detector is increased if compared to the centered detector. Conversely, the aforementioned disadvantage becomes evident when the scan time is considered. A centered detector necessitates a scan range of $180° + \text{fan angle}$, whereas the shifted detector requires a full $360°$. A $360°$ scan requires a greater time span, with state-of-the-art scanners requiring approximately six seconds,

which can lead to motion artifacts. In comparison, clinical CT scanners are much faster, with rotation times smaller than half a second [6].

2.6.1 Artifacts

Whenever assumptions and simplifications of the physical processes are made during reconstruction, artifacts are created in the reconstructed image. A great deal of effort is expended on the correction and reduction of these artifacts, with mechanical components during the scan or with computational methods in post-processing.

i.) Beam Hardening Artifacts

This artifact can be attributed to the energy dependent attenuation. Low-energy photons have a higher attenuation, which results in an increase in the mean energy of the spectrum after passing through an object, a phenomenon known as beam hardening. Section 2.2.1 discussed the energy dependence of the attenuation and the polychromatic nature of X-ray tubes. In addition, as highlighted in Section 2.5, the energy dependency of the attenuation is neglected during the reconstruction. At a detector element the measured spectrum-dependent negative log attenuation can be expressed as

$$q(L) = -\ln \int dE\, w(E, L)\, e^{-\int_0^\infty d\lambda \mu(E, \boldsymbol{s} + \lambda \boldsymbol{\Theta}(L))}. \tag{2.27}$$

Here, L is the line corresponding to a ray starting at $\boldsymbol{s}(L)$ going along the directional vector $\boldsymbol{\Theta}(L)$. $\mu(E, \boldsymbol{r})$ is the energy-dependent spatial distribution of the linear attenuation coefficient. E is the photon energy and $w(E, L)$ is the detected spectrum normalized to unit area, which is angle-dependent, and hence depends on the direction of the line L. The spectrum includes the emitted X-ray distribution and can include attenuation aspects such as the prefiltration and the sensitivity of the detector. Note that in the following the dependence on L is dropped for the sake of readability.
The attenuation $\mu(E, \boldsymbol{r})$ of a single material can be split into a spatial and an energy-dependent part $\mu(E, \boldsymbol{r}) = \psi(E) f(\boldsymbol{r})$. Then, the line integral can be written as

$$p = \int_0^\infty d\lambda f(\boldsymbol{s} + \lambda \boldsymbol{\Theta}) := \mathsf{X} f(\boldsymbol{s} + \lambda \boldsymbol{\Theta}), \tag{2.28}$$

with X the forwards projection operator. By expressing q through p and the polychromatic Radon transform operator R_f

$$q = \mathsf{R}_f f = -\ln \int dE\, w(E) e^{-\psi(E)\, p}, \tag{2.29}$$

it is apparent that a non-linear relationship exists between q and p. If p is doubled, the measured attenuation q is not doubled, therefore the X-rays which undergo greater attenuation, for instance, when traversing the central region of the patient, are subject to underestimation in the reconstruction process, which is based on the assumption of

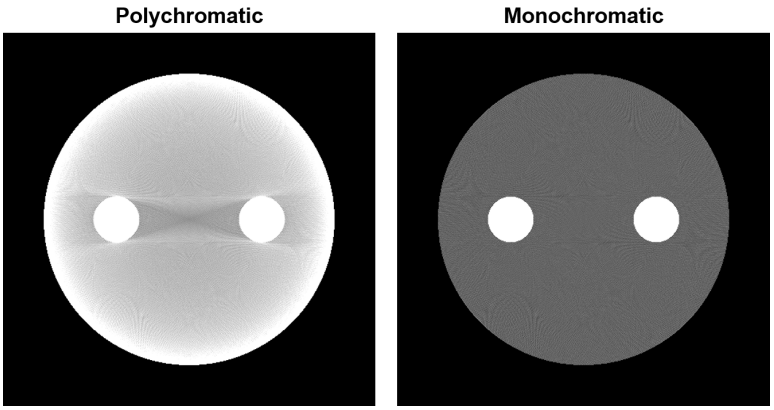

Figure 2.6: The reconstruction from a polychromatic simulation of a cylindrical water phantom with two bone inserts and the corresponding monochromatic reconstruction. The beam hardening artifacts are visible as cupping with higher values closer to the edge and as streaks originating from the bone inserts. $C = 0\,\mathrm{HU},\ W = 500\,\mathrm{HU}$

monochromatic X-ray beams. In comparison, the attenuation of X-rays that interact with the outer regions of the patient deviate less from a monochromatic beam. This causes so-called cupping artifacts, with the edge of the object appearing with a higher attenuation in the reconstructed image than the center. In addition, the difference of the energy dependency of ψ for different materials can cause streak artifacts at tissue boundaries such as soft tissue and bones. Figure 2.6 highlights both artifacts by comparing the reconstruction of polychromatic projections of a water cylinder with bone inserts to the reconstruction of monochromatic projections.

ii.) Scatter Artifacts

Scattered photons cause a significant loss of image quality. An example is displayed in Figure 2.7, where the scatter causes dark regions in the reconstructed image. The artifacts result from the wrong assumption that the measured intensity comes from photons which traveled in a straight line. As discussed in Section 2.5, the reconstruction is based on line integrals that start at the source and continue in a straight line to the detector. However, as highlighted in Section 2.2, Compton and Rayleigh scattering change the direction of flight of photons. Thus, the measured intensity I can be split into two parts

$$I = I_p + S\,, \tag{2.30}$$

the intensity of the primary intensity I_p, the contribution of non-scattered photons, and the intensity of the scattered photons S.

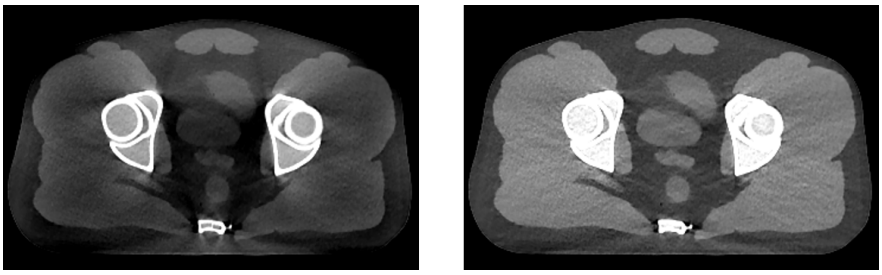

Figure 2.7: Example of a pelvis phantom scan without scatter correction (left) and with applied scatter correction (right). $C = 50\,\mathrm{HU},\ W = 500\,\mathrm{HU}$

iii.) Cone-Beam Artifacts

In contrast to the two artifacts highlighted before, the so-called cone-beam artifacts do not stem from imperfection in the physical modeling but from geometric constraints of the CBCT scanner geometry and the resulting scanning trajectory. Therefore, they are not observed in clinical CT and thus named cone-beam artifacts. They appear in reconstructed slices which do not satisfy the "Tuy-Condition" [48]. It states that a mathematically exact reconstruction is only possible for voxels for which every plane containing this voxel intersects the source trajectory. For a circular trajectory, this only applies to the central plane. The condition is violated more strongly for slices further away from the central plane, because the number of intersecting planes decreases towards the outside.

2.7 Deep Learning

Deep learning (DL) is a subset of machine learning (ML) which utilizes neural networks (NNs). It is a data driven approach which learns a representation or transformation by composing simple and non-linear modules [49]. This is in contrast to classical methods, which rely on handcrafted algorithms.

2.7.1 Artificial Neurons

The fundamental computational unit of a NN is an artificial neuron. Their setup is inspired by biological neurons. In a system of biological neurons, the neurons are connected by axons and dendrites. Signals by other neurons are received by the dendrites and if the combined signal exceeds a certain threshold the neuron "fires" a signal which is sent over its own axons to the dendrites of other neurons. In an artificial neuron the signal of the previous neurons are summed up, sometimes a bias is added (like a static potential) and passed through an activation function modeling the threshold. This can

be expressed mathematically as

$$a' = \sigma\left(\sum_k a_k w_k + b\right), \tag{2.31}$$

with a' representing the signal of the current neuron, a_k the signal of the input neuron k scaled with its respective weight w_k, the bias b and the activation function $\sigma(\cdot)$. The activation function can be used to further define the output. In instances where a probability is anticipated, the utilization of an activation function that scales the output between zero and one is a viable approach. An example for this is the softmax function [50] given by

$$\sigma(\boldsymbol{z})_i = \frac{e^{z_i}}{\sum_{k=1}^{K} e^{z_k}} \tag{2.32}$$

for output neuron i of all K outputs. Another popular activation function is the rectified linear unit (ReLU) [51], [52] function defined as

$$\sigma(x) = \max(0, x) \tag{2.33}$$

2.7.2 Neural Networks

The simplest examples of NNs are so-called multilayer perceptrons (MLPs). They consist of an input layer, any number of hidden layers and an output layer. Each layer is built from a number of artificial neurons and each neuron is connected to every neuron in the previous and the following layer. This configuration is the basis for the name fully connected neural network. A schematic for such a network is displayed in Figure 2.8 with the corresponding nomenclature. The first layer receives the input, in this case two values ($\boldsymbol{x} = (x_1, x_2)$). Then the forward pass is calculated, by propagating the input values through every layer towards the output layer. Accordingly, the calculation of the output $\boldsymbol{y}$ can be written as

$$\boldsymbol{y} = \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{a}_{\text{out}}(\boldsymbol{a}_2(\boldsymbol{a}_1(\boldsymbol{x}))), \tag{2.34}$$

with $\boldsymbol{a}_i$ being the output of the layer $i = 1, 2, \text{out}$ (for hidden layer one, two and the output). Using Equation (2.31) we can write the output of layer two as

$$\boldsymbol{a}^{(2)} = \boldsymbol{\sigma}\left(\begin{bmatrix} w_{1,1}^{(2)} & w_{1,2}^{(2)} & w_{1,3}^{(2)} & w_{1,4}^{(2)} \\ w_{2,1}^{(2)} & w_{2,2}^{(2)} & w_{2,3}^{(2)} & w_{2,4}^{(2)} \\ w_{3,1}^{(2)} & w_{3,2}^{(2)} & w_{3,3}^{(2)} & w_{3,4}^{(2)} \end{bmatrix} \cdot \begin{bmatrix} a_1^{(1)} \\ a_2^{(1)} \\ a_3^{(1)} \\ a_4^{(1)} \end{bmatrix} + \begin{bmatrix} b_1^{(2)} \\ b_2^{(2)} \\ b_3^{(2)} \end{bmatrix}\right). \tag{2.35}$$

The superscript (l) denotes the output/weights of layer l and $\boldsymbol{\sigma}$ applies the activation function to each element of the vector i.e. to each neuron. Therefore, the forward pass is a chain of matrix multiplications which can be calculated very fast through

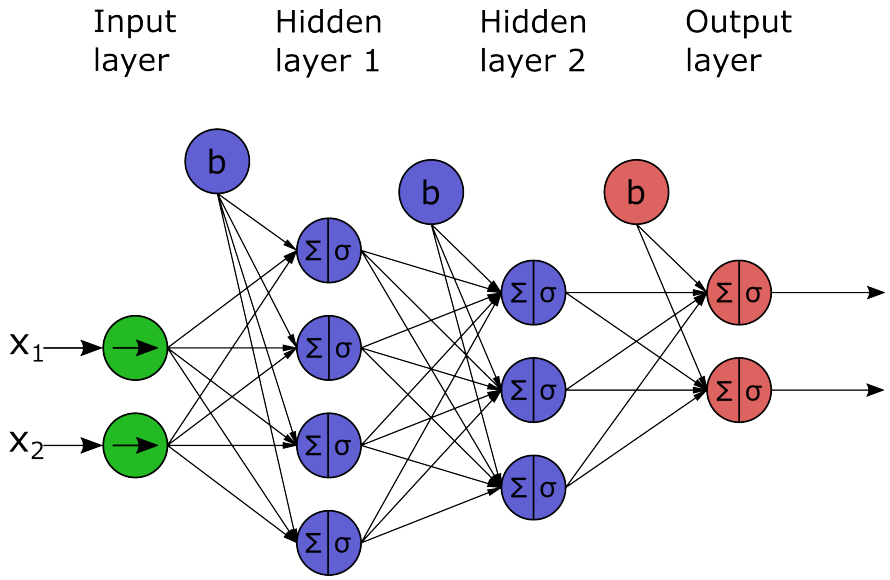

Figure 2.8: Schematic of a dense neural network with an input layer, two hidden layers each with a ReLU function as activation function and bias and an output layer.

parallelization on graphics processing units (GPUs).

2.7.3 Convolutional Neural Networks

An important subgroup of NNs are convolutional neural networks (CNNs). The number of input pixels in images is considerable, which renders conventional networks inefficient. Consequently, alternative approaches were explored, resulting in the development of convolutional neural networks, which employ standard filters. In the beginning they were primarily used in the field of pattern recognition within images and can learn to extract image specific features [53]. However, their applicability has since expanded to nearly all domains with 2D or 3D data. Similar to the previously discussed MLPs, convolutional layers are made up of artificial neurons with trainable weights. They calculate a dot-product, add a bias and pass the value through an activation function. However, the dot-product is not calculated over all input values or pixels in the case of images, but only over a small patch. Figure 2.9 shows a schematic for a 1D example. A 1D patch of size $K = 3$ is scaled with the weights and bias of the kernel and written to the output. The patch used to calculate the output is moved over the input and the output elements are created one after the other. This corresponds to a convolution of the input with a kernel. Furthermore, additional padding can be incorporated into the input to ensure that the resulting output is aligned with the input size. The convolution to calculate output O_i can be expressed as

$$\boldsymbol{a} = \boldsymbol{\sigma}\left((\boldsymbol{I} * \boldsymbol{K}) + \boldsymbol{b}\right) , \tag{2.36}$$

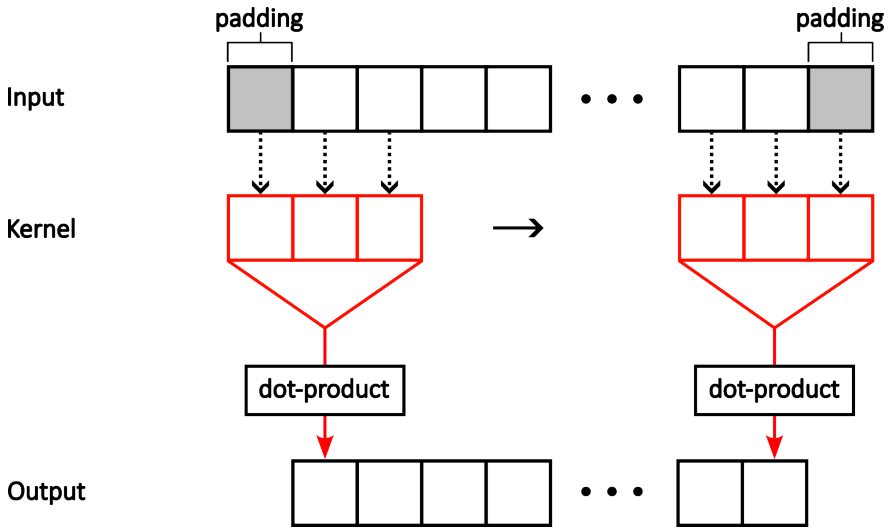

Figure 2.9: Schematic of a 1D convolutional layer with padding.

with $\boldsymbol{I}$ the input, $\boldsymbol{K}$ the kernel with the trainable weights, the bias $\boldsymbol{b}$ and the activation function $\boldsymbol{\sigma}$. Note that the process is the same with a higher dimension. Also the stride, the amount of elements the kernel is moved after it is applied, can be changed. This will reduce the size of the output, for example for a stride $s = 2$ half of the input is skipped and the output size is reduced by 50 %.
Another name for the output from a single kernel is feature map. In most cases each convolution layer does not apply a single kernel, but different kernels. The weights of each kernel are optimized during training, with the objective of enabling each kernel to extract a different meaningful feature from the input.
Pooling layers represent an additional crucial component of large convolutional networks. In this approach, patches are replaced by a statistic that summarizes the patch. In contrast to convolutional layers, these pooling layers have no trainable weights but apply mathematical operations such as maximum, minimum, or averaging. Reducing the resolution helps the network to learn low-frequency features and increases the field of perception.

2.7.4 Training of Neural Networks

It is important to have an automated method of identifying the optimal parameters in larger neural networks, which comprise millions or even billions of trainable elements. Deep neural networks are trained in a fashion that they learn a mapping from an input $\boldsymbol{X}$ to an output $\boldsymbol{Y}$, $f_{\boldsymbol{\Theta}} : \boldsymbol{X} \rightarrow \boldsymbol{Y}$. Here, $f_{\boldsymbol{\Theta}}$ represents the network f with its trainable parameters $\boldsymbol{\Theta}$. There are several different ways to find the correct parameters. Most of them can be divided into supervised and unsupervised learning approaches. In supervised learning, matching pairs of input $x \in \boldsymbol{X}$ and output $y \in \boldsymbol{Y}$, also called label, are collected. Then the predicted output of the network $\hat{y}\,(x; \boldsymbol{\Theta})$ of input x

and parameters $\boldsymbol{\Theta}$ is calculated and the error with respect to the ground truth y is evaluated. The error is also referred to as loss, and the loss function should be selected in accordance with the specific problem at hand, as it has a significant impact on the overall performance. A prominent function for a pixel-wise loss is the mean squared error (MSE) which is given by

$$\mathcal{L}_{\text{MSE}}\left(\boldsymbol{y}, \hat{\boldsymbol{y}}; \boldsymbol{\Theta}\right) = \frac{1}{N} \sum_{n=1}^{N} ||\hat{y}_n - y_n||_2^2 \,, \tag{2.37}$$

with n being a single element of the output such as a pixel of the predicted image. In general, the loss is back-propagated with a gradient-descent optimization to train the network weights. A common optimizer is the Adam optimizer [54].
The finite data, on which the parameters are optimized on, are only subsets of the infinite true set of all possible data $\boldsymbol{X}_{\text{train}} \subset \boldsymbol{X}$. Given that the trained network, $f_{\boldsymbol{\Theta}}$, is an approximation of the true function, $f : \boldsymbol{X} \rightarrow \boldsymbol{Y}$, it follows that the training data should be a good representation of the entire set, given that optimal parameters $\boldsymbol{\Theta}$ are found on this data. Commonly, the data available for training are split into three sets: training data, validation data and test data. The feed-forward computation of the training set with weight optimization is called an epoch. At the end of an epoch, the performance is evaluated on the validation set which has not been used for the optimization of the weights and biases. The comparison helps to adjust some hyperparameters, parameters which define the network structure and optimization process, and to verify that the network does not overfit on the training data. After a certain number of epochs are reached, the training is stopped and the network is applied on the test data for a final performance test. Test data, neglected during training, gives a certain amount of confidence if a suitable mapping has been learned.
While supervised training is conceptually simple, suitable matched training data are not always available since it can be very time- and resource-consuming to curate. In such cases, unsupervised learning may be a viable alternative. In that case, only the training data $x \in \boldsymbol{X}_{\text{train}}$ are available and an appropriate loss function has to be found to measure the performance without a corresponding ground truth (GT). One approach is a generative adversarial network (GAN) architecture [55] which uses another NN to evaluate the performance of the first network.
As mentioned above, one problem with training NNs is overfitting. Similar to fitting a function to data, during training the NN is fitted to the mapping of $f : \boldsymbol{X} \rightarrow \boldsymbol{Y}$. Overfitting is referred to the case, if the utilized function achieves good results on the fitting or training data while failing on data not seen during optimization. A prominent example is fitting an higher order polynomial function to data from a second order polynomial. The fitted function represents the training data perfectly while the error on test data increases. Another problem for overfitting is an inappropriately chosen training set which does not represent the whole range of possible inputs. Some strategies to improve training and to reduce overfitting are described below:

Batches: The straight forward way to calculate the loss would be to process the training

data one after one, updating the weights after each sample. This is called stochastic gradient-descent and increases the influence of some outliers in the data. Processing the whole data as a single batch would be batch gradient-descent, but is not feasible due to the limited GPU memory. The solution are mini-batches, processing a small batch of several inputs at once and calculating the mean loss. Hence, the computation of the weight optimization is optimized while averaging the loss decreases the influence of outliers in the training set.

Data Augmentation: The available data can be augmented to increase the range of "different" data included in the training so that the training data $\boldsymbol{X}_{\text{train}}$ closer represents $\boldsymbol{X}$, the set of all possible input data. This makes the network less prone to perturbations in the input and increases the training data size, possibly preventing overfitting. Typical data augmentations include some sort of transformation, such as a vertical or horizontal flipping, scaling or deformation.

Early Stopping: During training, the loss may further decrease over the training data, while the validation loss reaches a plateau or even starts to increase again. The training can be stopped after the validation loss did not improve for a set number of epochs.

Dropout: Averaging the output of differently trained methods may also increase the performance. A computationally efficient method is to use dropout layers. Here, a set percentage of neurons are dropped during training, which refers to deactivating their incoming and outgoing connections [56]. The dropped neurons are chosen independently for every forward call. For inference the probability is multiplied with the outgoing weights of each neuron in the dropout layer. In summary, a network with N neurons in a dropout layer can then be seen as 2^N networks [56].

3 | Materials & Methods

3.1 Scatter Artifact Correction

This section focused on the correction of scatter, which is a major source of image quality loss. First, the prior work is discussed, followed by a closer look into Acuros [13], [32] and deep scatter estimation (DSE) [14], [24], [31], both state-of-the-art scatter corrections.

3.1.1 Background and Prior Work

Scatter is especially problematic in CBCT because of high scatter-to-primary ratios. Typically, there are two strategies for reducing the resulting artifacts: scatter suppression and correction. While the scatter suppression is a process that occurs during the acquisition of the projections, the scatter correction is a post-processing step. The scatter can be suppressed by anti-scatter grids and collimators which reduce the number of scattered X-rays reaching the detector [57]. Scatter correction, on the other hand, tries to estimate and correct the scatter intensity in the measured data. This may be achieved through the utilization of software-based methods for physical or empirical modeling of X-ray scattering [13], [18], [58], [59], or alternatively, through the deployment of particular filters during the scanning process, such as primary modulation grids or beam blockers [60], [61].
The gold standard for the software-based scatter estimations are Monte Carlo (MC) scatter simulations. If properly implemented, they are able to represent all physical processes of the data acquisition. Thus, they can provide an accurate estimate of the scatter. A well-known drawback of MC simulations is the extensive computational effort, which makes MC simulations time-consuming. This prevents them from being applicable in real-time and making them impractical for clinical applications. A further limitation of MC simulations is that they are image-based methods. In other words, they rely on a *prior* reconstruction of the uncorrected data.
Relying on a *prior* reconstruction makes the method susceptible to errors in the reconstructed image. In addition to the uncorrected scatter artifacts, other image distortions such as truncation and movement during data acquisition may also affect the results from image-based correction methods. Other methods circumvent the problem by estimating and correcting the measured projection data in the raw data domain and therefore do not rely on a *prior* reconstruction. Common methods include kernel-based

algorithms [12] which are also in clinical use but do not perform as well as image-based correction methods [32]. The latest techniques, which are based on deep learning, demonstrate the capacity to surpass the established projection-based methods [24].

3.1.2 Acuros

An image-based method which is faster than MC simulations is AcurosCTS [13], [32] (for reasons of better readability, it will be written only as Acuros in subsequent sections). Acuros is the gold standard for scatter correction in products by Varian Medical Systems, a Siemens Healthineers Company. It solves the linear Boltzmann transport equation deterministically. Although Acuros is currently employed in clinical settings, this approach shares the same limitation as other image-based correction techniques, the necessity for a *prior* reconstruction, which is susceptible to inaccuracies and requires a full set of projections.
As described above, Acuros is the gold standard for scatter correction in clinically used Varian CBCT systems. An example for this is the Ethos, a LINAC with integrated CBCT. This subsection provides an overview of Acuros, outlining its functionality and how it addresses the linear Boltzmann transport equation. For a complete description the reader is referred to the publications [13], [32].
Acuros relies on a *prior* reconstruction. Thus, after applying some initial corrections to the measured projections, such as a darkfield correction and a first kernel-based scatter correction, a first-pass reconstruction is performed. The reconstructed images are used as input to Acuros, which first performs a segmentation into soft tissue and bone [32] to account for the different attenuation and scatter cross-section for each tissue type. Then, the linear Boltzmann transport equation is used to calculate the photon fluence of scattered and unscattered photons that reach each detector pixel.
Prior to an examination of the methodology for resolving the linear transport equation, it is necessary to provide an explanation of the equation itself. Following [13] the equation can be written as:

$$\begin{aligned} &\hat{\boldsymbol{\Omega}} \cdot \boldsymbol{\nabla}\phi\left(\boldsymbol{r}, E, \hat{\boldsymbol{\Omega}}\right) + \mu_t\left(\boldsymbol{r}, E\right)\phi\left(\boldsymbol{r}, E, \hat{\boldsymbol{\Omega}}\right) = \\ &S\left(\boldsymbol{r}, E, \hat{\boldsymbol{\Omega}}\right) + \int_0^{E_0} dE' \int_{4\pi} d\hat{\boldsymbol{\Omega}}' \left[\mu_s\left(\boldsymbol{r}, E' \to E, \hat{\boldsymbol{\Omega}}' \to \hat{\boldsymbol{\Omega}}\right)\phi\left(\boldsymbol{r}, E', \hat{\boldsymbol{\Omega}}'\right)\right] . \end{aligned} \tag{3.1}$$

The first term is the dot-product between a directional vector $\hat{\boldsymbol{\Omega}}$ and the gradient of the photon fluence ϕ for photons of energy E at spatial position $\boldsymbol{r}$. Secondly, the remaining term on the left-hand side is the total attenuation μ_t (see Section 2.2) of ϕ. On the right-hand side, the first term is the contribution from external and internal sources S, for example the X-ray tube. Finally, the increase due to the photon fluences of energy E' going in the direction $\hat{\boldsymbol{\Omega}}'$, which, after scattering with the probability given by the linear directional scattering coefficient μ_s, have the energy E and go in the direction $\hat{\boldsymbol{\Omega}}$. In summary, the left-hand side of the equation represents the photon flux emanating from position $\boldsymbol{r}$ in the direction of $\hat{\boldsymbol{\Omega}}$ and the fraction of this beam absorbed at position $\boldsymbol{r}$. The total outgoing photon fluence must therefore be in equilibrium with the angular

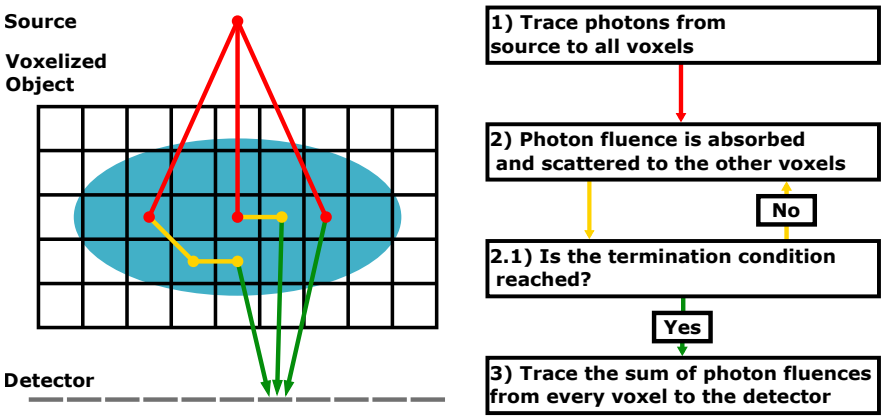

Figure 3.1: Schematic of how Acuros solves the linear Boltzmann transport equation. In a first step, the photons are traced from the source to the voxels. Then, the photon interactions are handled in an iterative fashion by calculating the contribution from each voxel to each other voxel by considering the fluence of the incoming photons and the respective linear directional scattering coefficient. When the photon fluences after the scatter interaction are too small a stopping criterion is reached, and the fluences in each voxel are summed. Then the fluences are traced to the detector pixels to sum the total fluence reaching each detector pixel from every voxel. Adapted from [13].

fluence arising from external sources along with those scattered into the point $\boldsymbol{r}$ from the specified direction $\hat{\boldsymbol{\Omega}}$. Figure 3.1 visualizes the three distinctive steps to solve the equation, which has to hold for every point.

Initially, photons are tracked from the source to the voxels. These are the fluences of the unscattered photons. To handle scattering interactions, an iterative approach is employed. The scattered photon fluence is calculated for each voxel, so in the first iteration the photons are scattered once and the contribution from each voxel to all other voxels can be determined for a single scatter contribution. Then the second iteration starts, now considering only photon fluences that have undergone a single scattering process. This iterative process continues until a stopping criterion is met, such as when the fluence from scattering drops below a predefined threshold. At this point, the total fluence in each voxel is summed and then traced to the detector pixels to determine the total fluence reaching each pixel. In the final step, attenuation of anti-scatter grids and the detector sensitivity can be taken into account by using the Beer-Lambert law (see Section 2.2.1) and the direction of incoming photons. These attenuation values can also be stored in a look-up table (LUT) for fast access [13].
It is evident that computing the photon fluence from every voxel to every other voxel is a significant computational burden. Consequently, the data are subjected to a drastic downsampling to fit on the GPU memory. The angular directions are discretized, the available energies are binned, the detector pixel size is downsampled to larger pixels, and

the voxels of the input image are downsampled as well to larger voxels with a larger edge length. In addition to memory reduction, this approach also reduces computation time. Of course, the method cannot be applied in real-time because a full set of projections is required for a *prior* reconstruction with a reasonable image quality without the introduction of excessive sparseness artifacts. Nevertheless, it is sufficiently rapid and precise to be employed in a clinical setting [32], unlike a MC simulation.

3.1.3 Deep Scatter Estimation

The DSE was introduced as a real-time scatter estimation method which uses a CNN to estimate the scatter intensity directly from a measured projection [14], [24]. The applicability of DSE has been demonstrated for both forward scatter [24] and cross-scatter in a dual source CT [31] and for proton radiation therapy treatment planning [25]. Applying a neural network holds the advantage of not relying on handcrafted kernels or features, but intrinsically learning the best fitting features and representation from training data. In [14] the DSE was presented with a network architecture similar to a U-Net, which was previously developed for image segmentation purposes [62] and has found a very broad applicability for various other problems. DSE is trained in a supervised manner with simulated data, the input being the scatter-corrupted projections and their corresponding label being the intensity of the scattered photons. In previous publications [14], [24], [25] it has been shown that DSE works well with measured data, provided that the simulation accurately reproduces the measured projections. This is an important trait, because there is not enough data of scatter-free measurements available. Thus, simulations are used to generate training data. In this work, the scatter intensities are generated by Acuros, the tool originally designed to correct for scatter. DSE used to be trained only by data generated with MC simulations, hence, the use of Acuros is a novelty. At present, DSE needs to be retrained for each individual scanner geometry. All Varian scanners are already integrated into Acuros. Therefore, if DSE can be trained with Acuros, it can be readily made available for a diverse range of scanners. Furthermore, DSE was only compared to other projection-based correction methods [14], [24], [31], whereas here it is also compared to Acuros, a state-of-the-art image-based correction method.

i.) Training Data

In order to train a CNN, a large amount of training data is required. As highlighted above, there are not enough paired data available to train a DSE network on measured data. Thus, Acuros was utilized for the simulation of data for a scanner of the Ethos product line. Here, another advantage of training the network with Acuros becomes apparent. Acuros is already in clinical use and therefore is optimized to generate a realistic scatter distribution for a wide range of scanners. In this thesis, the training data were simulated in the scanner geometry of the Ethos scanner, as this scanner is available at the German Cancer Research Center (DKFZ) for measurements. Figure 3.2 compares the workflow of Acuros when it is applied to measured data and how it was

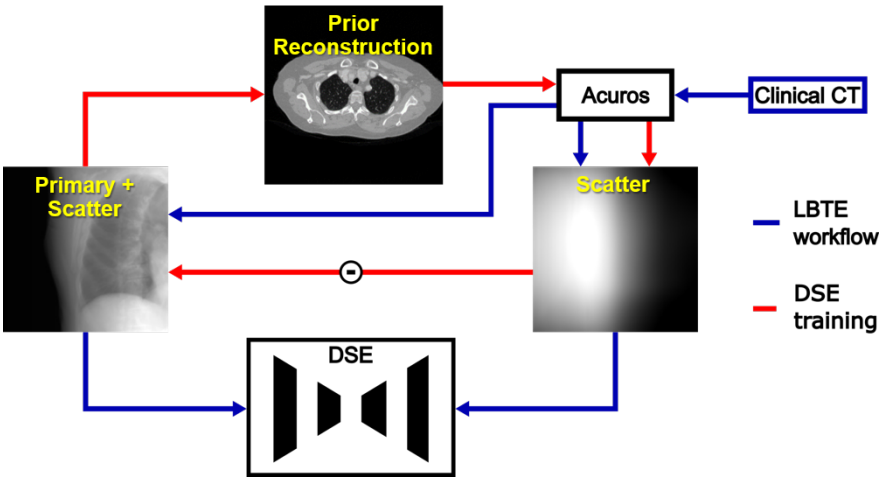

Figure 3.2: Schematic of the different Acuros pipelines. The red path shows how it is applied for measured data and the blue path how it was applied to simulate training data. Reconstructions from clinical CT scanners were used as input and Acuros simulated primary intensity and corresponding scatter distributions.

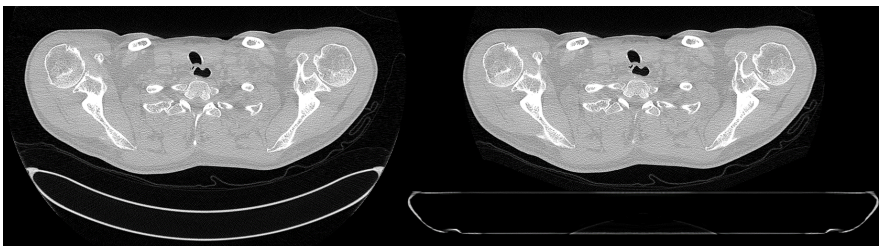

Figure 3.3: On the left, a slice of the clinical CT scan and on the right, the same slice with the Ethos couch inserted.

used for training data generation. Normally, Acuros uses a *prior* reconstruction as input. On the basis of this first-pass reconstruction the scatter is estimated and then subtracted from the measured projections. However, in order to generate training data, reconstructed images from clinical CT scanners were used as input. Due to the reduced artifacts in clinical CT, when compared to images from CBCT, they are seen as artifact-free.

A set of 24 full body clinical CT scans were collected, partly from an open source dataset which was published for the segmentation challenge [63] and the rest from cadaver measurements which have been acquired in scope of a forensic study in close collaboration with the Institute of Forensic and Traffic Medicine (Prof. Sarah Heinze), Heidelberg University, Heidelberg, Germany, after being approved by the local ethics review board (S388/2014). The clinical CT scans are pre-processed by removing the patient couch and replacing it with a couch similar to the one used in the Ethos. This

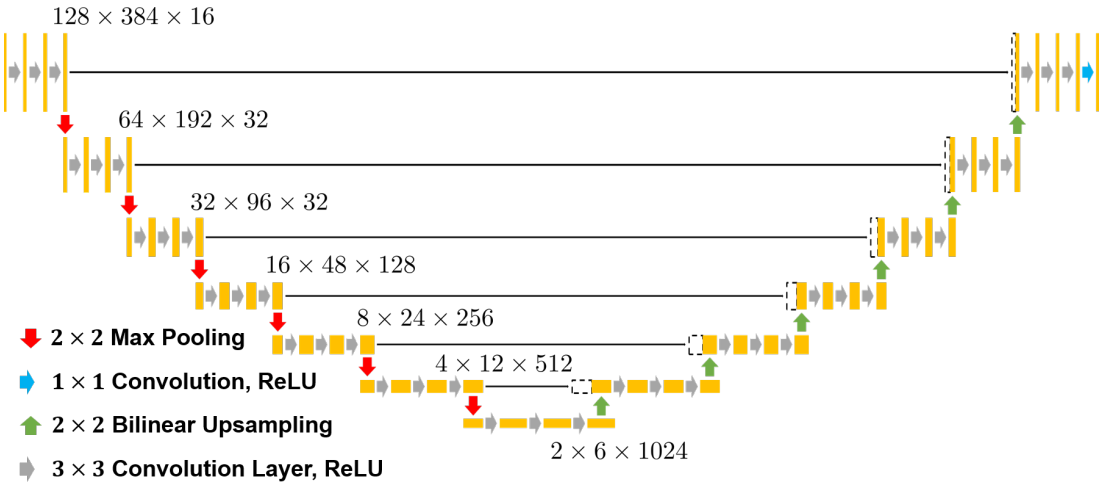

Figure 3.4: The network architecture used for the deep scatter estimation.

is necessary because the couches mounted on clinical CT systems are curved and the couches on the Ethos system are flat. Figure 3.3 shows an example slice with both couches. In addition, Wang et al. [32] found that the couch has a strong influence on the scatter distribution. Thus, a realistic couch model is required. Following the preprocessing, the patients were split into a training (17 patients), validation (2 patients) and a test set (5 patients). Then, three anatomical regions, thorax, abdomen and pelvis, were selected. For each patient and anatomical region two positions along the z-axis were randomly chosen. To ensure a wide coverage of the projection angle, 72 projections were simulated for each z-position, uniformly distributed around 360°. Furthermore, every scan was simulated twice, once with the original voxel size and once with a scaled voxel size. This is a form of data augmentation to incorporate more smaller and thicker patients. This resulted in $72 \times 2 \times 3 \times 17 \times 2 \times 2 = 14688$ projections for training and $72 \times 2 \times 3 \times 2 = 1728$ for validation. Finally, the simulated scatter S_{Acuros}, air norm I_0 and primary intensities I_{prim} are used to generate training data, with the projection calculated as:

$$p = -\ln\left(\frac{I_{\text{prim}} + S_{\text{Acuros}}}{I_0}\right), \tag{3.2}$$

and the corresponding label or ground truth as

$$y = \frac{S_{\text{Acuros}}}{I_0}. \tag{3.3}$$

ii.) Network Architecture

Figure 3.4 depicts the network architecture which follows the architecture proposed by Maier et al. [14], [24]. It resembles a U-Net, with convolutional blocks constructed from three 2D convolutional layers, each with a 3×3 kernel, a stride of $s = 1$ and followed by a

ReLU activation function. In the encoding path, downsampling is accomplished through a max pooling layer, halving the feature map size in width and height. Following each downsampling, the number of feature maps is doubled in the first convolutional layer, commencing with 16 feature maps and culminating in 1024 in the deepest convolutional block. The decoding path employs a bilinear interpolation for upsampling to double the image width and height. A skip connection concatenates the output of the convolutional blocks of the encoding path with the input of their respective convolutional blocks in the decoding path.

iii.) Training

Three different inputs are investigated to train DSE. "pDSE" is the first model which is trained on the projection p. "pepDSE" is the second and applies a transformation first. The transformation was proposed by Maier et al. [24] and is given by

$$T(p) = \text{pep} = pe^{-p}\,. \tag{3.4}$$

Lastly, the "pepDSE-Air0"-model is trained, utilizing pep as input and values in air are set to zero. This approach ensures that no information about the scatter in air is passed to the network. The pep-transform is proportional to a first order scatter estimation as it was shown in [64]. Furthermore, the input was downsampled to 128×384 pixels. Scatter distributions are known to be of low frequency, and thus, this approach allows for the preservation of all relevant features while reducing the input size. A smaller sample size decreases the computational work load and thus the inference time. Further data augmentation was applied during training by applying a 50 % probability to horizontally flipping both the input and the corresponding label.

The network was implemented with Python (v. 3.11) in the PyTorch (v. 2.3.0) and the PyTorch Lightning framework (v. 2.2.3). To accelerate the training, it was performed on an NVIDIA RTX A5000 GPU utilizing the NVIDIA API compute unified device architecture (CUDA) (v. 11.8). AdamW [65], a modified version of the Adam [54] algorithm, serves as the optimizer, with a mini-batch size of $b = 16$. The learning rate was initialized with 10^{-5} and halved whenever the validation loss did not decrease for more than 30 epochs. Ultimately, the training was set to stop after 250 epochs, with an early stopping when the validation loss did not decrease for 110 epochs. In order to enhance the network's focus on areas where the scatter is more pronounced relative to the primary signal, the loss function was set to be the scatter-to-primary-weighted mean absolute percentage error (SPMAPE) [31]. The SPMAPE for the predicted scatter S_{DSE} with corresponding ground truth S_{Acuros} and I_{prim} is given by

$$\text{SPMAPE} = \frac{1}{N}\sum \left| \frac{S_{\text{Acuros}} - S_{\text{DSE}}}{S_{\text{Acuros}}} \times \frac{S_{\text{Acuros}}}{I_{\text{prim}}} \right| \tag{3.5}$$

$$= \frac{1}{N}\sum \left| \frac{S_{\text{Acuros}} - S_{\text{DSE}}}{I_{\text{prim}}} \right| . \tag{3.6}$$

3.1.4 Evaluation

To evaluate the performance of the trained DSE, its scatter removal capabilities are compared to the image-based correction Acuros and projection-based correction fast adaptive scatter kernel superposition (fASKS) [12]. Both are currently in clinical use and represent the state-of-art of scatter correction methods by Varian. In a first step, the five test patients are used to simulate full CBCT scans with Acuros. In total, one scan per anatomical region (thorax, abdomen, pelvis) and test patient is evaluated, resulting in a total of 15 evaluated scans. fASKS is a simple correction method applied in the raw data domain and also used to correct the *prior* reconstruction for Acuros. For all three correction methods, the estimated scatter is clipped to 95 % of the measured intensity as a precautionary measure to prohibit nonphysical scatter intensities that exceed the measured signal. Finally, the reconstructions of the different scatter-corrected projections and the uncorrected projections are compared to the GT reconstruction of the scatter-free primary intensity $p_{\mathrm{GT}} = -\ln\left(\frac{I_{\mathrm{prim}}}{I_0}\right)$.
The reconstruction pipeline was created in iTools (v. 3.7.0.36), an application which incorporates the same reconstruction algorithms and correction methods as they are applied on the Ethos scanner. Thus, the default reconstruction pipelines are utilized, as they are applied to clinical cases. The *prior* image was reconstructed using the iterative algorithm iCBCT, which is the default algorithm for iterative reconstructions. All other reconstructions, as well as the Acuros scatter-corrected projections, are reconstructed with a variant of the FDK. The resulting image has $512 \times 512 \times 251$ voxels with a voxel size of $0.8\,\mathrm{mm} \times 0.8\,\mathrm{mm} \times 1.0\,\mathrm{mm}$.
The accuracy of the correction methods is evaluated through qualitative and quantitative metrics. In an initial step the skin contour of the patient is automatically segmented in the GT image. All metrics are calculated on voxels of interest, which excludes all voxels outside of the contour. For a quantitative analysis three metrics are calculated. First with the mean absolute error (MAE) given by

$$\mathrm{MAE}\,(\hat{y}, y) = \frac{1}{N}\sum_{n=1}^{N} |y_i - \hat{y}_i| \ . \tag{3.7}$$

Here, $\hat{y}$ is the reference volume, the GT, consisting of N voxels and y is the volume of the method that is to be evaluated. As the error is calculated over each voxel in relation to its corresponding voxel in the reference volume, this metric is referred to as a voxel-wise metric. Including all tissues with different Hounsfield unit ranges may bias the mean. For example, a mean difference of 10 HU may be much more severe in the kidneys which are in the range of 20 HU to 40 HU ([6]) than in cortical bone which reaches up to 2000 HU. Thus, a MAE-spectrum is calculated as well, with bins of 20 HU width from -1000 HU to 1500 HU as used in [66]. Each voxel in the ground truth reconstruction is sorted into its corresponding bin, and the MAE is calculated for each bin, with the results presented separately for each bin.
Another commonly applied metric is the structural similarity index metric (SSIM) which was developed for the evaluation of perceptual image quality on the basis of known

properties of the human visual system [67]. In this work the SSIM is used as defined in [67]:

$$\mathrm{SSIM}\left(\hat{y}, y\right) = \frac{\left(2\mu_{\hat{y}}\mu_y + C_1\right)\left(2\sigma_{\hat{y}y} + C_2\right)}{\left(\mu_{\hat{y}}^2 + \mu_y^2 + C_1\right)\left(\sigma_{\hat{y}}^2 + \sigma_y^2 + C_2\right)} \,. \tag{3.8}$$

In this notation, μ is the local mean, σ is the local standard deviation and $\sigma_{\hat{y}y}$ is the cross-covariance. C_1 and C_2 are two constants to ensure stable results if $\sigma_{\hat{y}}^2 + \sigma_y^2 \approx 0$ and $\mu_{\hat{y}}^2 + \mu_y^2 \approx 0$ [67]. The SSIM, as defined here, requires voxel values to be greater than or equal to zero. The CT value range of the reconstructed images starts at $-1000\,\mathrm{HU}$. Hence, it is rescaled to the range [0, 1]. Given the variability in windowing parameters across different anatomical regions and tasks, the proposed rescaled value, I', is dependent on these parameters

$$I' = \begin{cases} 0 & \text{if } I < C - W/2, \\ \frac{I - C + W/2}{W} & \text{if } C - W/2 \leq I \leq C + W/2, \\ 1 & \text{otherwise.} \end{cases} \tag{3.9}$$

The constants C_1 and C_2 are set to $C_1 = (0.01L)^2$ and $C_2 = (0.03L)^2$, following [68]. L is the dynamic range of the images, which is one after rescaling.
For a further comparison, all correction methods are applied to a real measurement of an anthropomorphic pelvis phantom. A proper ground truth is not available. Thus, the scans are only compared qualitatively to each other.

3.2 Beam Hardening Correction

This section presents projection-based beam hardening correction methods. First, the theoretical framework for the commonly employed water precorrection is discussed and then extended for higher-order beam hardening correction approaches. At last, novel projection-based deep beam hardening corrections are introduced, aiming to improve the water precorrection in bone tissue.

3.2.1 Water Precorrection

The water precorrection is a common beam hardening correction method available on all clinical scanners [69]. Recall from Section 2.6.1 that beam hardening comes from the energy-dependent attenuation which is neglected during reconstruction and the polychromatic nature of the X-ray tube. Following Equation (2.29) the measured attenuation q can be expressed by

$$q = -\ln \int dE w(E)\, e^{-p_0 \psi_0(E)} \quad , \tag{3.10}$$

with p_0 a line integral in the scanned object $f_0(\boldsymbol{r})$ and the energy-dependent attenuation $\psi_0(E)$. If the energy and spectrum-independent p_0 could be retrieved and reconstructed

in place of q, beam hardening would be mitigated. Assuming that the object is composed exclusively of water (i.e., $\psi_0(E) = \psi_{\text{water}}(E)$), the equation can be inverted to gain p_0. Water is commonly used due to its dominant abundance in the human body. Consequently, this approach is referred to as the "water precorrection". There are two different ways to estimate p_0. First, Equation (3.10) can be numerically inverted. For this, it is necessary to know the spectrum $w(E)$, which can be measured, and the energy dependency of water $\psi_{\text{water}}(E)$. The energy dependency is a quantity that has been the subject of experimental investigation for a long time. It has been published and discussed in publications such as [35], [42]. Consequently, it is expected to be known with a high degree of precision. Instead of an analytical inversion, a mapping from q to p_0 can be measured. For this, a scan of a water phantom of known dimensions is acquired. Consequently, for each measured value of q, the actual path length p_0 through the phantom is known. If the phantom contains a large range of line integrals, an accurate mapping from q to p_0 can be created and applied to other measurements. Some scanners store the mapping in a LUT to improve computation times.

The above method assumes that all tissues have an attenuation similar to water. However, materials such as bone or metal feature significantly different energy dependencies. Therefore, the water precorrection is unable to remove all artifacts, and some streaks persist in the image. Figure 3.5 shows this by comparing the reconstruction of a polychromatic simulated CBCT scan to the corresponding reconstruction of water-precorrected projection data. The water precorrection removes the cupping while it is unable to correct the dark streaks between the bone inserts.

3.2.2 Higher Order Beam Hardening Correction

Higher order beam hardening corrections are used to also correct for beam hardening by bones. These methods often rely on a *prior* reconstruction of the polychromatic projection. In most cases, the different tissues are segmented in the image domain. Then the forward projection for every tissue type is calculated and the resulting polychromatic projection from the *prior* reconstruction is compared to the initially measured attenuation. More complete descriptions are found in [70]–[72]. As highlighted before, the reliance on *prior* information is subject to artifacts in the reconstructed images. These are the to-be corrected beam hardening artifacts, scatter artifacts or truncation. Moreover, the calculation of multiple reconstructions and forward projections is a computationally expensive process. Furthermore, a full set of projections is necessary for reconstruction. Thus, they are not practical for fast and in real-time applicable problems, such as online imaging during treatment. This increases the need for real-time applicable methods in the raw data domain.

Equation (2.27) is re-written, similar to the water precorrection, for a higher order beam hardening correction. In this case, the patient is represented by soft tissue and bone, the two most abundant materials in the human body. Then, the attenuation q

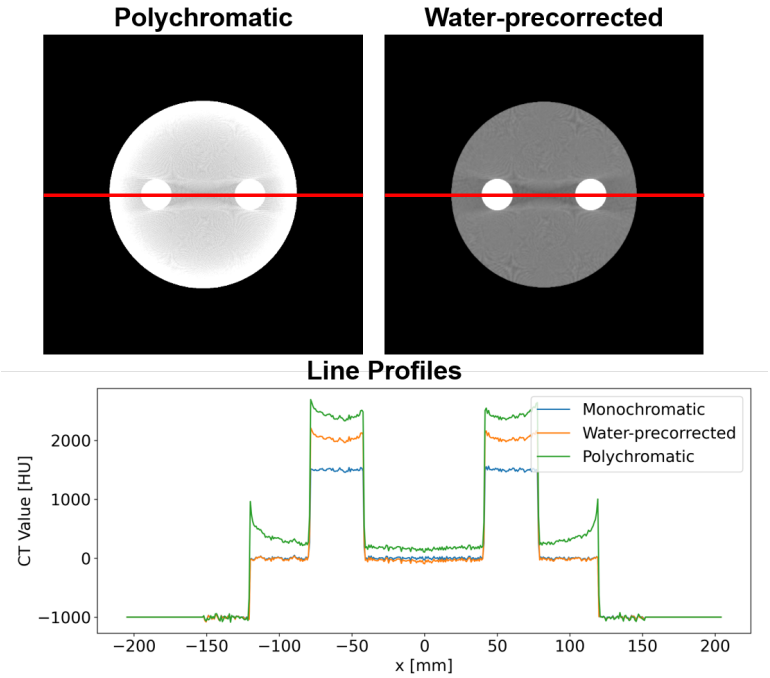

Figure 3.5: The reconstruction from a polychromatic simulation of a cylindrical water phantom with two bone inserts and the corresponding water-precorrected reconstruction. The beam hardening artifacts are visible as cupping with higher values closer to the outside and as streaks originating from the bone inserts. The water precorrection removes the cupping while the streaks persist between the bone inserts persist. The graphic below shows the line profiles for the red lines and compares them with a monochromatic reconstruction. $C = 0\,\mathrm{HU},\ W = 500\,\mathrm{HU}$

.

can be rewritten as

$$q = -\ln \int dE w(E) e^{-p_{\mathrm{soft}}\psi_{\mathrm{soft}}(E) - p_{\mathrm{bone}}\psi_{\mathrm{bone}}(E)}\,. \tag{3.11}$$

Now, the line integral is split into two parts p_{bone} and p_{soft}, each belonging to the previously mentioned materials with a distinct energy dependency, represented by ψ_{bone} and ψ_{soft}. A monochromatic projection of energy E_0 is calculated as

$$p_{\mathrm{mono}}(E_0) = p_{\mathrm{soft}}\psi_{\mathrm{soft}}(E_0) + p_{\mathrm{bone}}\psi_{\mathrm{bone}}(E_0)\,. \tag{3.12}$$

Consequently, monochromatic projections for different energies are calculable if the line integrals are known for soft tissue and bone, in addition to the respective energy-dependent attenuation. In this work, deep leaning methods are proposed to estimate the monochromatic projection or even to extract p_{soft} and p_{bone} from the measured

polychromatic projection q. Extracting p_{soft} and p_{bone} enables the calculation of virtual monochromatic images (VMIs). The advantages of virtual monochromatic reconstructions have been discussed in length for dual energy computed tomography (DECT) scanners. There, they have shown to be beneficial for non-contrast and contrast-enhanced dual energy scan routines [73].
In a preliminary investigation, a simulation study is proposed to examine the training of deep neural networks for the correction of beam hardening in the raw data domain. To this end, a variety of approaches are evaluated, with different networks trained on different labels. Subsequently, the method with the best performance is used for a robustness study in relation to the anatomical region and the X-ray tube spectrum. Table 3.1 summarizes the investigated options. Each of the proposed methods is discussed in more detail in the following subsections.

i.) Monochromatic Projection Prediction

The first method is to predict the monochromatic projection from the measured polychromatic projection. This concept of directly predicting an artifact-free projection is derived from numerous approaches that accomplish this within the image domain. There, a mapping from deteriorated images to images without artifacts is a common practice by training and applying a NN. Examples for this can be found for synthetic CT generation [74], metal artifact correction [75] or denoising [76].
Learning a direct mapping from the measured projection q to a monochromatic projection p_{mono} has the disadvantage that no additional information is gained about the bone and water content. Thus, the possibility of calculating monochromatic projections for different energies is lost. The energy of the predicted monochromatic projection is determined by the training data. To circumvent this problem, it would be possible to train multiple networks, with each network trained to map to monochromatic projections of a different energy.

ii.) Deep Bone Extraction (DBE)

The deep bone extraction (DBE) aims to enable the calculation of VMIs by estimating the line integrals in bone and soft tissue. In order to achieve this, a NN is trained to predict p_{soft} and p_{bone} directly from the polychromatic projection q. The network has two output channels, one for p_{soft} and the other for p_{bone}, which are used as labels during training. One disadvantage is the lack of data consistency. The predicted $p_{\text{soft, pred.}}$ and $p_{\text{bone, pred.}}$ are not constrained to reproduce the input q. Thus, the relation

$$q = -\ln \int dE w(E) e^{-p_{\text{soft, pred.}}\psi_{\text{soft}}(E) - p_{\text{bone, pred.}}\psi_{\text{bone}}(E)}\,, \tag{3.13}$$

may not hold for every prediction. Therefore, the reconstruction from the ground truth may not represent the ground truth anymore.

iii.) Deep Bone Blending (DBB)

Table 3.1: Summary of the evaluated network labels for a higher order beam hardening correction with their respective advantages and disadvantages.

Learned Mapping	Advantage	Disadvantage
$q \rightarrow p_{\text{mono}}$	-	· Single energy VMI
$q \rightarrow p_{\text{soft}}, p_{\text{bone}}$	· VMIs for different energies possible	· No data consistency · High frequency structures
$q \rightarrow q_{\text{water}} \rightarrow p_{\text{soft}}, p_{\text{bone}}$	· Data consistency · Less high frequency structures · VMIs for different energies possible	· Approximation of bone line integral

As seen in Figure 3.6, p_{soft} has a dark shadow at the position of the bones, which are called "holes" in the following. These holes create more high-frequency structures in p_{soft}, the correct prediction of which may be more difficult to learn. In addition, the NN has to learn a mapping from the polychromatic q to the spectrum-independent line integrals. This may limit the previous approach to the input of a polychromatic input from the characteristic tube spectrum with which the NN was trained. The third proposed method is designed to be less spectrum-dependent by predicting q_{water}. Similar to Equation (3.10), q_{water} is calculated as

$$q_{\text{water}} = -\ln \int dE w(E) e^{-p_{\text{water}} \psi_{\text{soft}}(E)} \quad . \tag{3.14}$$

p_{water} is the line integral through the patient with the bones replaced by standard water at 70 keV. An example for p_{water} can be seen in Figure 3.6. In the absence of bone tissue along the line integral, the relation $q = q_{\text{water}}$ is applicable to these line integrals. Therefore, this method can be seen as a "blending" of bones into the soft tissue and the method is called deep bone blending (DBB). Further, the polychromatic projection q is split into two parts,

$$q = q_{\text{water}} + q_{\text{bone}} \, . \tag{3.15}$$

With the assumption

$$r = \frac{p_{\text{bone}}}{p_{\text{soft}}} \approx \frac{q_{\text{bone}}}{q_{\text{water}}} \tag{3.16}$$

Equation (3.11) is rewritten into

$$q = -\ln \int dE w(E) e^{-p_{\text{soft}}\psi_{\text{soft}}(E) - r p_{\text{soft}}\psi_{\text{bone}}(E)} \,. \tag{3.17}$$

With the spectrum $w(E)$ and the energy dependencies of the soft and bone tissue Equation (3.17) can be solved with any common root finding algorithm. The spectrum can be measured, and the energy dependency can be derived from the relevant literature. After the network predicted q_{water}, the ratio is calculated and the Newton-Raphson method was utilized to solve Equation (3.17) for p_{soft}.
q_{water} has to be smaller than or equal to q, thus the post processing $q_{\text{water}} = \min(q_{\text{water}}, q)$ is done before calculating q_{bone} and the ratio r.

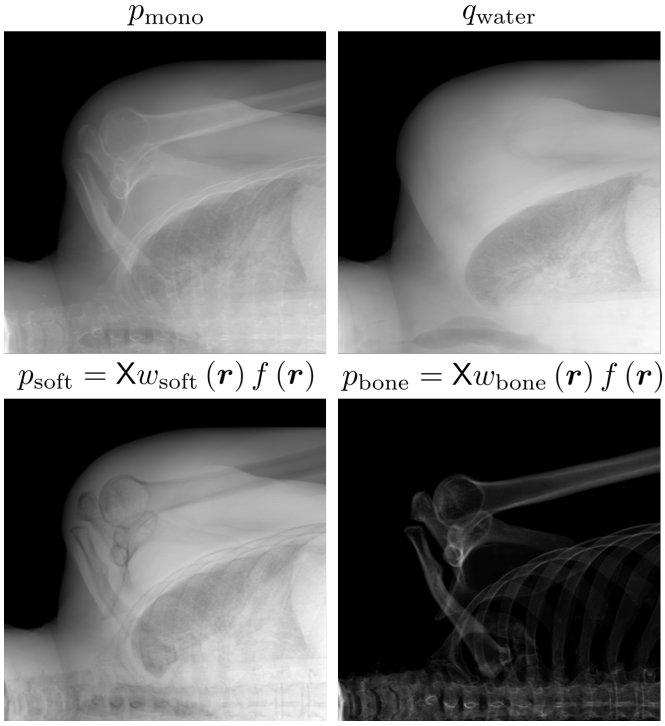

Figure 3.6: In the top row the monochromatic projection p_{mono} is shown on the left and q_{water} in the top right. The corresponding line integrals in the soft tissue p_{soft} and bone tissue p_{bone} are shown on the bottom left and right respectively. The $f(\boldsymbol{r})$ for the two line integrals can be seen in Figure 3.7.

3.2.3 2D Look-up Table (LUT)

Equation (3.17) has to be solved for every pixel individually. The p_{soft} depends only on the two parameters q and r. In addition, similar ratio and attenuation pairs are likely to appear several times across all projections of a scan. Hence, a two-dimensional LUT can be calculated once and then accessed by the given ratio and attenuation. This

gives $p_{\text{soft}}(q, r) = \text{LUT}(q, r)$. To avoid unnecessary large LUTs the ratio was limited to a maximum of 0.8, which was empirically found to be sufficiently large.

3.2.4 Training Data

In order to calculate data for training, validation and testing, clinical CT scans are prepared in the following manner. Firstly, a segmentation is created. Secondly, the NN input q and the possible labels of the networks are calculated.

i.) Segmentation

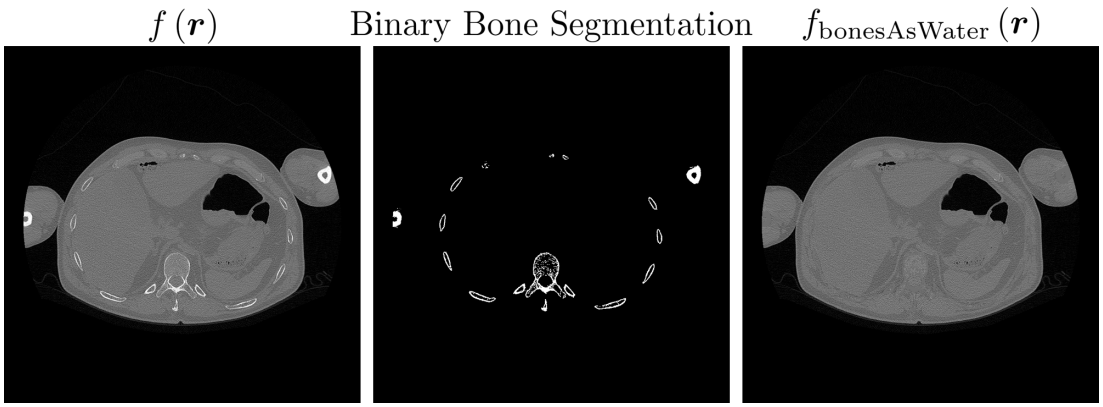

Figure 3.7: Example slice of a clinical CT scan (left) with the corresponding binary segmentation of the bone (center) and the same CT slice with the bones replaced by standard water (right). The binary segmentation was created by using a threshold of 200 HU and morphological operators to remove small voxel clusters. $C = 0\,\text{HU},\ W = 700\,\text{HU}$

Two distinct segmentation processes were conducted. The first to segment the bones to replace them with water to calculate p_{water}. For this, a binary segmentation with a threshold of 200 HU was used. In addition, morphological operations were used to eliminate singular voxels without a neighboring bone voxel. This eliminates voxels which would be classified as bone due to noise. Figure 3.7 shows an example slice of a clinical CT volume with corresponding segmentation. The segmentation is deployed to substitute bones with water with an attenuation of $\mu_{\text{water}} = 0.0192\,\text{cm}^{-1} \mathrel{\widehat{=}} 0\,\text{HU}$. Alternatively, a soft segmentation, as outlined in the following section, could be utilized for this process.
The second segmentation is necessary for the calculation of line integrals p_{soft} and p_{bone}. These line integrals are used to determine the polychromatic and monochromatic projections, as well as all other labels for training. For this, a simple image-based segmentation is utilized that assigns each voxel a material-dependent weight. Each voxel is represented as a water-bone-mixture. Hence, for each voxel of the CT reconstruction

$f(\boldsymbol{r})$ the fraction $w_{\text{water}}(\boldsymbol{r})$ and $w_{\text{bone}}(\boldsymbol{r})$ is calculated following

$$w_{\text{water}}(\boldsymbol{r}) = \begin{cases} 1 & \text{if } f(\boldsymbol{r}) < \tau_{\text{water}}, \\ \frac{\tau_{\text{bone}} - f(\boldsymbol{r})}{\tau_{\text{bone}} - \tau_{\text{water}}} & \text{if } \tau_{\text{water}} \leq f(\boldsymbol{r}) < \tau_{\text{bone}}, \\ 0 & \text{otherwise.} \end{cases} \tag{3.18}$$

and

$$w_{\text{bone}}(\boldsymbol{r}) = \begin{cases} 0 & \text{if } f(\boldsymbol{r}) < \tau_{\text{water}}, \\ 1 - w_{\text{water}}(\boldsymbol{r}) & \text{otherwise.} \end{cases} \tag{3.19}$$

In this case the thresholds were $\tau_{\text{water}} = 250\,\text{HU}$ and $\tau_{\text{bone}} = 800\,\text{HU}$. Figure 3.8 illustrates a plot of the weights. These weights are used to calculate polychromatic data as described in the next section.

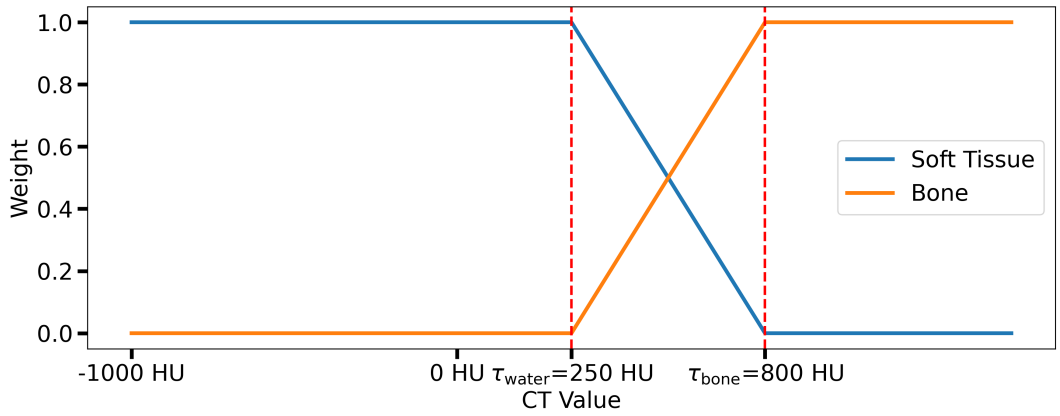

Figure 3.8: The weights for soft tissue and bone used for generating line integrals for soft tissue and bone.

ii.) Data Generation

In order to calculate q and all of the training labels, it is necessary to calculate the line integrals p_{soft}, p_{bone} and p_{water}. In addition, an X-ray spectrum and the energy dependence of the soft and the bone tissue are used to generate training data according to Equation (3.11). For the polychromatic spectrum, a semi-empirical model was used to simulate a tungsten target X-ray spectrum [77] with a tube voltage of 125 kV and no prefiltration. As previously written, the attenuation for a single material $\mu(E, \boldsymbol{r})$ can be split into the product of $f(\boldsymbol{r})$ and $\psi(E)$. The spatial function $f(\boldsymbol{r})$ is defined as

$$f(\boldsymbol{r}) = \frac{CT(\boldsymbol{r})\mu_{\text{water}}(70\,\text{keV})}{1000\,\text{HU}} + \mu_{\text{water}}(70\,\text{keV})\,. \tag{3.20}$$

$CT(\boldsymbol{r})$ are the CT values from a clinical CT used as input to simulate projections from realist patient anatomies. The clinical CT images display CT values in Hounsfield units. Therefore, $f(\boldsymbol{r})$ has the dimension length^{-1} and p_i the line integrals are unitless. The line integral p_i for tissue i is calculated using the Joseph forward projector [78] through the volume $w(\boldsymbol{r})_i \cdot f(\boldsymbol{r})$, with $w(\boldsymbol{r})_i$ the segmentation weight for tissue i. Now, $\psi(E)$ has to be unitless to as well and is defined for tissue i as

$$\psi_i(E) = \frac{\left(\frac{\mu}{\rho}\right)_i (E)}{\left(\frac{\mu}{\rho}\right)_{\text{water}} (70\,\text{keV})}\,. \tag{3.21}$$

The mass attenuation $\left(\frac{\mu}{\rho}\right)$ for various elements and tissues, along with the material compositions of different tissue types, is available in the literature [79], [80]. For soft tissue the attenuation of water is chosen as $\left(\frac{\mu}{\rho}\right)_{\text{soft}} = \left(\frac{\mu}{\rho}\right)_{\text{water}}$.
The same patients utilized for the generation of the DSE training data (see Section 3.1.3) serve as the basis for the generation of training and test data. Similar to the training in Section 3.1.3, the patients are split into a training set (17 patients), validation (two patients) and testing (five patients). For each patient, two z-positions in the pelvis region were selected for simulation of 72 projections of a CBCT scan, distributed uniformly around a 360° range. The detector consists of 1280×1280 detector elements with a pixel size of $0.336\,\text{mm} \times 0.336\,\text{mm}$. Additionally, the voxel size of the patient was once used in the original size and once multiplied by a factor of 0.9, a measure of data augmentation, in order to artificially double the number of training, validation and testing data. As a result, the training set comprised 4896 projections, while the validation set contained 576 projections.

3.2.5 Network Architecture

Similar to the previously defined architecture for the DSE, the proposed network architecture is the U-Net [62]. It consists of an encoding path and a decoding path which are connected by a skip connection. There are six stages in the encoding path, going from 32 to 1024 feature channels. Each stage applies two 3×3 convolutions, followed by a rectified linear unit (ReLU) activation function. It is to note that the network trained to directly predict p_{soft} and p_{bone} had a dying ReLU problem, which causes the neurons to only output zeros and a dying gradient. For this network, leaky ReLU activation functions were used with a slope of 0.1. The encoding path uses max pooling and the decoding path uses bilinear interpolation for upsampling to restore spatial resolution. In the final layer of the U-Net, a 1×1 convolution is applied to map the 32-component feature vector to the desired number of output channel(s). This setup ensures that the network is able to learn from the input data at various scales and spatial resolutions, capturing both high-level semantic information and low-level details.

3.2.6 Network Training

The network was initialized with the He initialization, also known as Kaiming initialization [81] and the bias with zeroes. As optimizer the AdamW [65], a modified version of the Adam [54] algorithm, was selected. The network was trained on patches of 640 px × 640 px, which were slightly larger than the receptive field of the U-Net and with a batch size of 16. Entire projections were used as input for validation and testing. The network was trained until the validation loss no longer decreased for 110 epochs with a learning rate of 10^{-5}. As for the DSE training, the learning rate was halved after 100 epochs in the absence of a reduction in the validation loss. In addition, the software and hardware used for implementing and conducting the training is the same as for the DSE training.

i.) Loss Functions

Two different loss functions were applied during training, depending on the task at hand. For the networks predicting p_{mono} or p_{soft} and p_{bone} the mean absolute error was used, evaluated at the patient position (all pixels in the projection with $q > 0$). For the network trained with q_{water} as label the MAE loss is modified. As explained before, q_{water} has to be smaller or equal than q. If $p_{\text{bone}} = 0$ for a given detector pixel $q = q_{\text{water}}$ and if the network would return $q_{\text{water}} = q$ for every detector pixel the proposed beam hardening correction using deep bone blending would be equal to a water precorrection. Therefore, in both cases, if the line integral crosses bone tissue or not, it is more advantageous for the network to predict a q_{water} higher than the label. This can be enforced by applying a weight to the lower prediction ($q_{\text{water, pred.}} < q_{\text{water, label}}$), enhancing the contribution to the loss. For a prediction y_i with GT $\hat{y}_i$ this loss is

$$\text{weightedMAE} = \frac{1}{n}\sum_{i}^{N}\begin{cases}|\hat{y}_i - y_i| & \text{if } \hat{y}_i \geq y_i \\ \alpha\,|\hat{y}_i - y_i| & \text{if } \hat{y}_i < y_i\end{cases} \tag{3.22}$$

with N pixels in the projection and weighting factor α which was empirically found to perform best for $\alpha = 2$.

3.2.7 Reference Methods and Evaluation

To evaluate the different networks, full CBCT scans with 720 projections uniformly distributed around 360° are simulated and reconstructed. A water precorrection is implemented as reference method, because it is the most common applied beam hardening correction on clinical CTs as well as the only available beam hardening correction at the Ethos CBCT. To evaluate the methods, the MAE and the MAE-spectrum are calculated as defined in Section 3.1.4.

3.2.8 Robustness Studies

The best performing network with respect to the used label is further evaluated with regards to different anatomical regions and different tube voltages.

i.) Different Tube Spectra

The selection of the appropriate tube voltage is dependent upon the specific patient in question. Factors such as patient thickness and age are crucial in determining the optimal value. In pediatric patients, lower tube voltages are used to reduce the administered dose and for thicker patients higher tube voltages are selected for a better signal. On the Ethos CBCT scanner, installed at the DKFZ, the available options are 80 kV, 100 kV, 125 kV or 140 kV. The application of different tube voltages results in the generation of distinct tube spectra, which in turn affects the beam hardening. Additionally, as outlined in Section 2.6, the use of prefilters by CBCT scanners introduces further modulation of the spectrum. To evaluate the robustness, one network is trained for each spectrum and then compared to a network trained on data from all spectra. In total four different spectra were considered during training and testing. They result from all permutations of two tube voltages (100 kV and 125 kV) and two tin prefilter thicknesses (0 mm and 1 mm). It is to note that this resulted in an unbalanced number of projections for training, because the network trained with all spectra is trained on four times as much data as the network trained on a single spectrum. There are several ways to correct for this shortcoming, for example, by decreasing the amount of update steps or using only one fourth of the data per spectrum. In this case both networks were trained until the validation loss had not decreased for more than 120 epochs and then the network with the best validation loss is used. Similar to the first study, full CBCT scans were simulated for testing, but in this case all scans were simulated once for each spectrum and both neural networks were applied for a higher order beam hardening correction. Then, the mean absolute percentage error, for N voxels of the ground truth $y_{i,\mathrm{GT}}$ and the prediction $y_{i,\mathrm{pred.}}$ defined as

$$\mathrm{MAPE} = \frac{100}{N} \sum_{n=1}^{N} \left| \frac{y_{i,\mathrm{pred.}} - y_{i,\mathrm{GT}}}{y_{i,\mathrm{GT}} + \epsilon} \right| . \tag{3.23}$$

To prevent a division by zero ϵ was set to 0.001 if not stated otherwise. The simulation parameters are summarized in Table 3.2.

ii.) Different Anatomical Regions

Applying the correction method to different anatomical regions would be practical for clinical cases. For this a simulation study was done by simulating training data not only for pelvis, but also for thorax and abdomen scans. Again, one network was trained on all anatomical regions while one was only trained for each region separately. The network with the best validation loss is used for the evaluation on the test data. All networks were evaluated on fully simulated CBCT scans on all anatomical regions by

Table 3.2: Generalizability of higher order beam hardening correction to different tube spectra: simulation study parameters.

Parameter	**Training**	**Testing**
Label	q_{water}	q_{water}
No. patients	19	5
Anatomical region	pelvis	pelvis
Source-to-isocenter distance	540 mm	540 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	0° – 360°, $\Delta\vartheta = 5°$	0° – 360°, $\Delta\vartheta = 0.5°$
Detector elements	1280 × 1280	1280 × 1280
Detector pixel size	0.4 mm × 0.4 mm	0.4 mm × 0.4 mm
Detector shift	175 mm	175 mm
Tube voltage	125 kV, 100 kV	125 kV, 100 kV
Prefilter	0 mm Sn, 1 mm Sn	0 mm Sn, 1 mm Sn
#Projection per scan	72	720
Object scaling	0.9, 1.0	1.0
Data augmentation	Horizontal flipping	-

calculating the mean absolute percentage error (MAPE) between the predictions and the ground truth. The simulation parameters are summarized in Table 3.3.

Table 3.3: Generalizability of higher order beam hardening correction to different anatomical regions: simulation study parameters.

Parameter	**Training**	**Testing**
Label	q_{water}	q_{water}
No. patients	19	5
Anatomical region	pelvis, abdomen, thorax	pelvis, abdomen, thorax
Source-to-isocenter distance	540 mm	540 mm
Source-to-detector distance	1000 mm	1000 mm
View angle	0° – 360°, $\Delta\vartheta = 5^\circ$	0° – 360°, $\Delta\vartheta = 0.5^\circ$
Detector elements	1280×1280	1280×1280
Detector pixel size	$0.4\,\text{mm} \times 0.4\,\text{mm}$	$0.4\,\text{mm} \times 0.4\,\text{mm}$
Detector shift	175 mm	175 mm
Tube voltage	125 kV	125 kV
#Projection per scan	72	720
Object scaling	0.9, 1.0	1.0
Data augmentation	Horizontal flipping	-

4 | Results

This chapter presents the results for the scatter and beam hardening correction methods. The first part compares the deep scatter estimation to gold standard methods by a quantitative evaluation with simulated data, for which a ground truth exists. Then, a phantom measurement is qualitatively evaluated. Earlier results for a scanner with a shifted detector have been presented at the conference Fully 3D 2023 [82]. The second part showcases deep beam hardening corrections in the projection domain. For this, a simulation study utilizing pelvic scans is evaluated, and the robustness with respect to different spectra and anatomical regions is subsequently analyzed. Initial results for a deep beam hardening correction have been presented at the CT Meeting 2024 [83].

4.1 Scatter Artifact Correction

4.1.1 Example Scatter Estimations

This subsection presents an analysis and evaluation of the scatter intensity estimations of Acuros, fASKS and DSE, which were introduced in Section 3.1. Figure 4.1 illustrates an uncorrected projection with the corresponding ground truth scatter distribution and example scatter estimations of the investigated methods. It is evident that all scatter estimations are consistent with the ground truth scatter distribution. Nevertheless, some differences are visible in the scatter distributions as well as the absolute percentage error (APE) which is defined for the estimated scatter S_{est} and the ground truth (GT) S_{GT} as

$$\text{APE} = 100 \left| \frac{S_{\text{est}} - S_{\text{GT}}}{S_{\text{GT}}} \right| . \tag{4.1}$$

The most significant discrepancy is evident in the scatter intensity generated with fASKS. In this scatter distribution, a decline in the predicted scatter intensity is observable in the projection's outermost area. An examination of the absolute percentage error with respect to the ground truth scatter provides further evidence of discrepancies between the scatter distribution of fASKS and the GT. In the APE image, the patient outline and anatomical features, including the lung, are clearly distinguishable. This illustrates that fASKS is not capable of adequately addressing the diverse anatomical structures present in a human body.

Furthermore, the scatter intensity of Acuros exhibits a distinctive striped pattern. This is indicated by red arrows in the scatter intensity and the corresponding APE image. It is not visible in the ground truth, which has also been generated with Acuros and is a nonphysical behavior, because scatter is expected to be a smooth distribution. This phenomenon can be attributed to the downsampling of the input volume and the detector pixel size used for the scatter correction. As discussed earlier, in order for the reconstruction to be clinically applicable, it must be fast. To reduce the computation time of the scatter correction with Acuros the *prior* reconstruction is downsampled to voxels of size $15\,\text{mm} \times 15\,\text{mm} \times 20\,\text{mm}$ and detector pixels of size $12.6\,\text{mm} \times 8.4\,\text{mm}$. For the simulation of the training and the test data for DSE, time was not a limiting factor and the full resolution has been used. Therefore, DSE shows a smooth and accurate scatter distribution in contrast to Acuros. However, the APE images of the DSE and Acuros exhibit a significantly smaller discrepancy to the ground truth in comparison to the absolute percentage error (APE) image of fASKS.

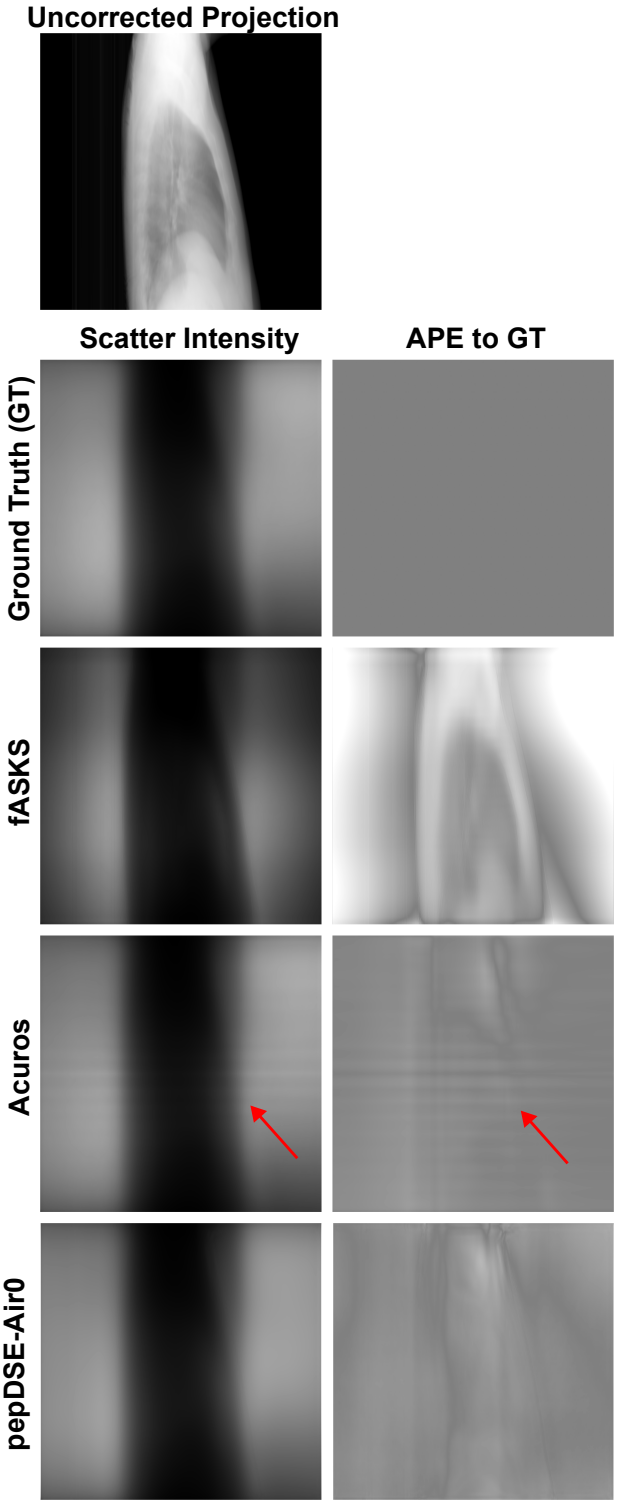

Figure 4.1: Scatter intensities from the different scatter correction methods, compared to the ground truth scatter distribution. The right column shows the absolute percentage error (APE) to the ground truth. The red arrow indicates the presence of a non-physical stripe pattern within the Acuros estimated scatter intensity. Projection: $C = 4.4$, $W = 8.8$, Scatter Intensity: $C = 0.018$, $W = 0.018$, APE: $C = 0\,\%$, $W = 50\,\%$

4.1.2 Reconstructed Slices of Scatter-corrected Scans

The following subsection examines the impact of the different scatter correction methods on the image quality of the reconstructed images. Figures 4.2 to 4.4, respectively, illustrate a slice of the thorax, abdomen, and pelvis scan of one test patient. All figures show the uncorrected scan, fASKS-, Acuros-, and all three DSE-corrected scans, as well as the difference to the scatter-free ground truth. Each of the three DSE versions are trained with different input. "pDSE" was trained on the negative log normed projection p, "pepDSE" on pep $= pe^{-p}$ and "pepDSE-Air0" was trained on pep but all values in air are set to zero. The last method has no information about the projection value in air and no features considering the scatter in air can be extracted during training. In all three example slices, the scatter artifacts are clearly evident in the uncorrected reconstruction. The bones and soft tissue exhibit considerably lower CT values and appear darkened. In the fASKS-corrected reconstruction, the artifacts are less distinct, but still present. Notably, in bones, the difference is particularly evident in the difference images for all anatomical regions. No artifacts are visible in the CT images of Acuros- and DSE-corrected scans. Upon consideration of the difference image, a slight discrepancy in performance is discernible. pDSE and Acuros demonstrate the best performance, whereas pepDSE exhibits a slight decline in performance and pepDSE-Air0 displays some artifacts in the difference images. In particular, for projections with a stronger attenuation, such as those in the pelvis region parallel to the transversal axis, a slight overestimation of the scatter of pepDSE-Air0 occurs, visible as bright streaks in the difference image. For all methods, the estimated scatter is subtracted from the scatter-corrupted signal. If the scatter is overestimated, too much signal is removed, and a stronger attenuation is assumed, resulting in the observed bright streaks in the reconstruction.
Figures 4.5 to 4.10, illustrate exemplary slices of the thorax, abdomen, and pelvis of two more test patients. Here, DSE introduces slight ring artifacts, only visible in the difference image. The rings are stronger visible in the abdomen scan (Figure 4.6) and pelvis scan (Figure 4.7) for both DSE versions using pep as input. The same observations that were made for the first test patient also apply to the other two test patients. fASKS cannot correct for scatter as well as DSE or Acuros, leaving some remaining scatter artifacts in the reconstructed image. Acuros and DSE perform on par, while the pepDSE-Air0 performs the worst of all three DSE variants, again, over predicting the scatter intensity for projections with higher attenuation.

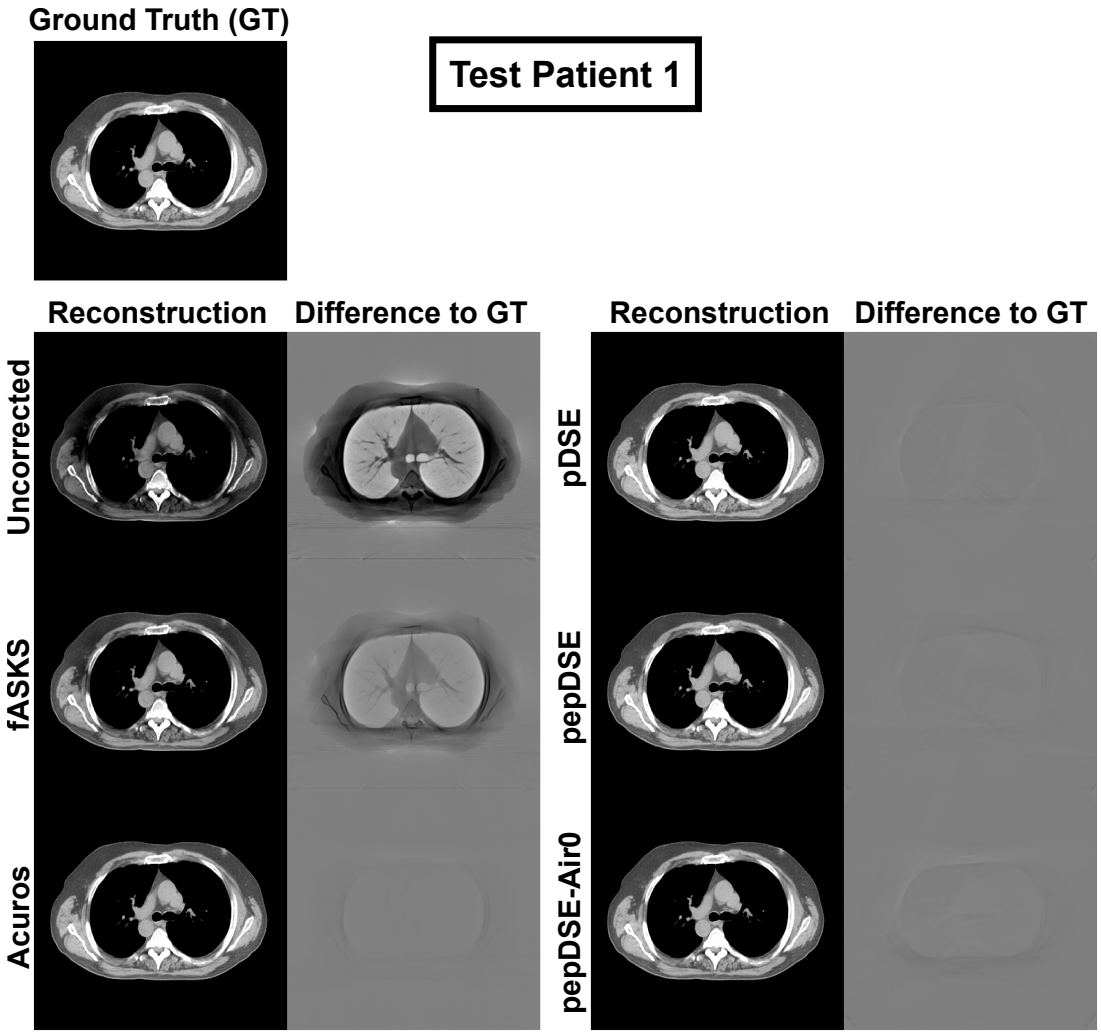

Figure 4.2: A slice of a thorax scan of test patient one, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

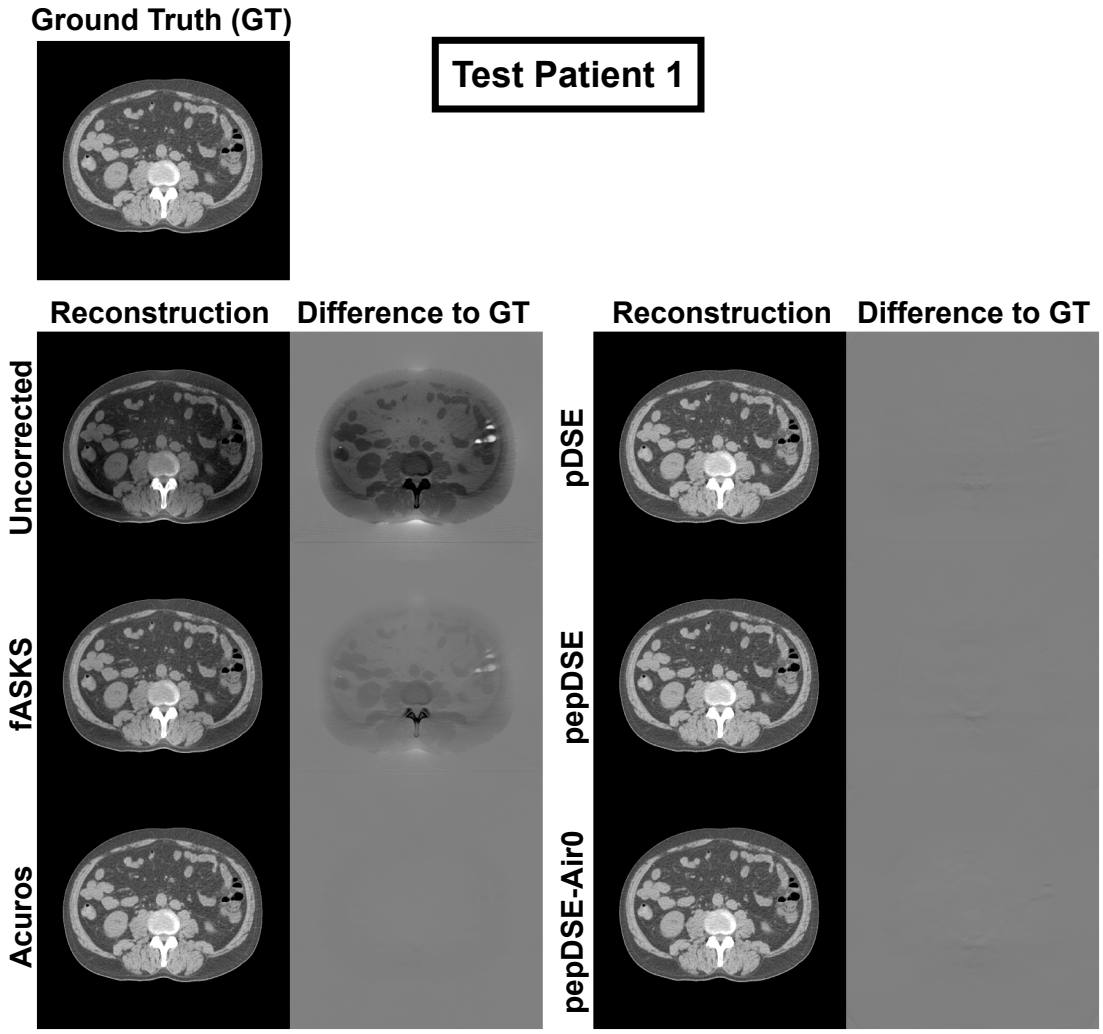

Figure 4.3: A slice of an abdomen scan of test patient one, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

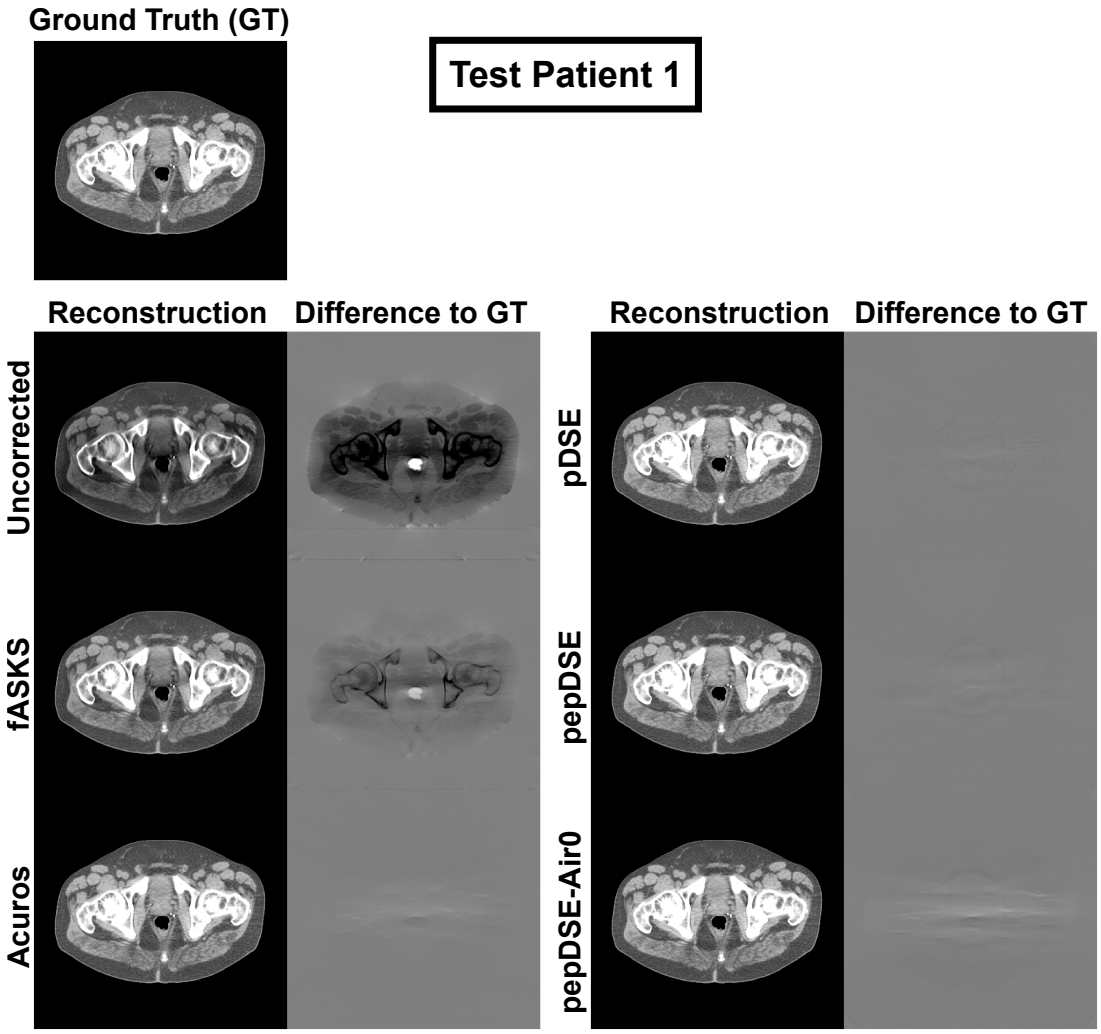

Figure 4.4: A slice of a pelvis scan of test patient one, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

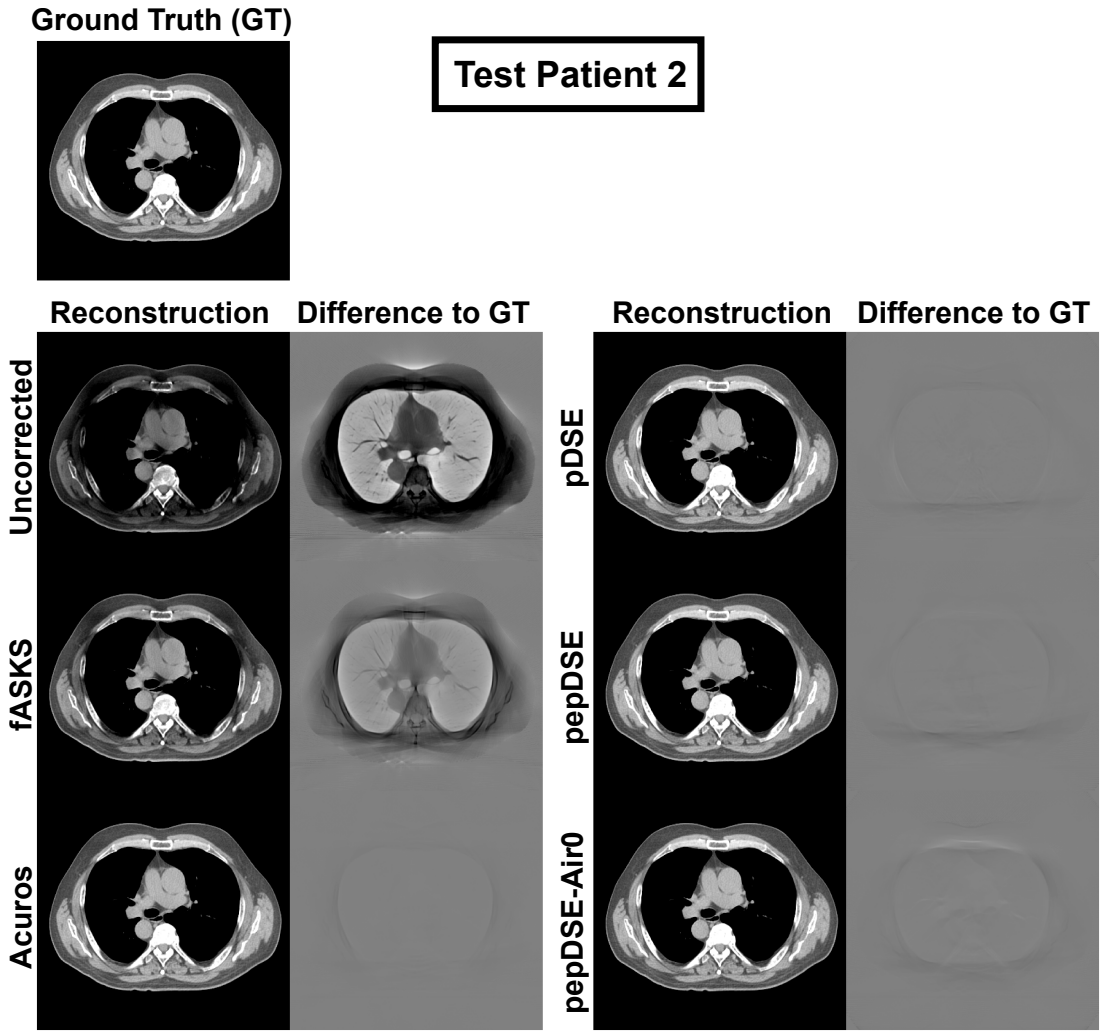

Figure 4.5: A slice of a thorax scan of test patient two, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

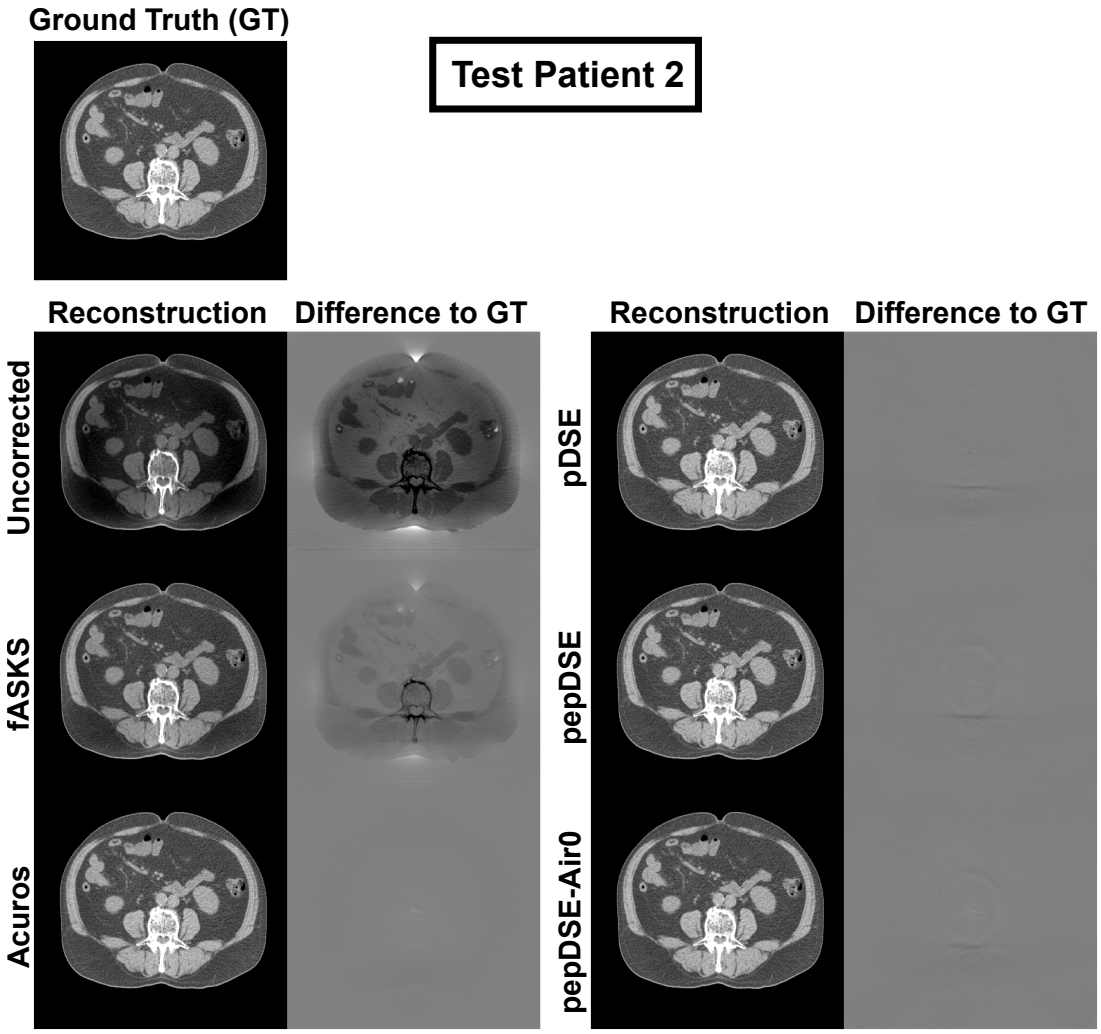

Figure 4.6: A slice of an abdomen scan of test patient two, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

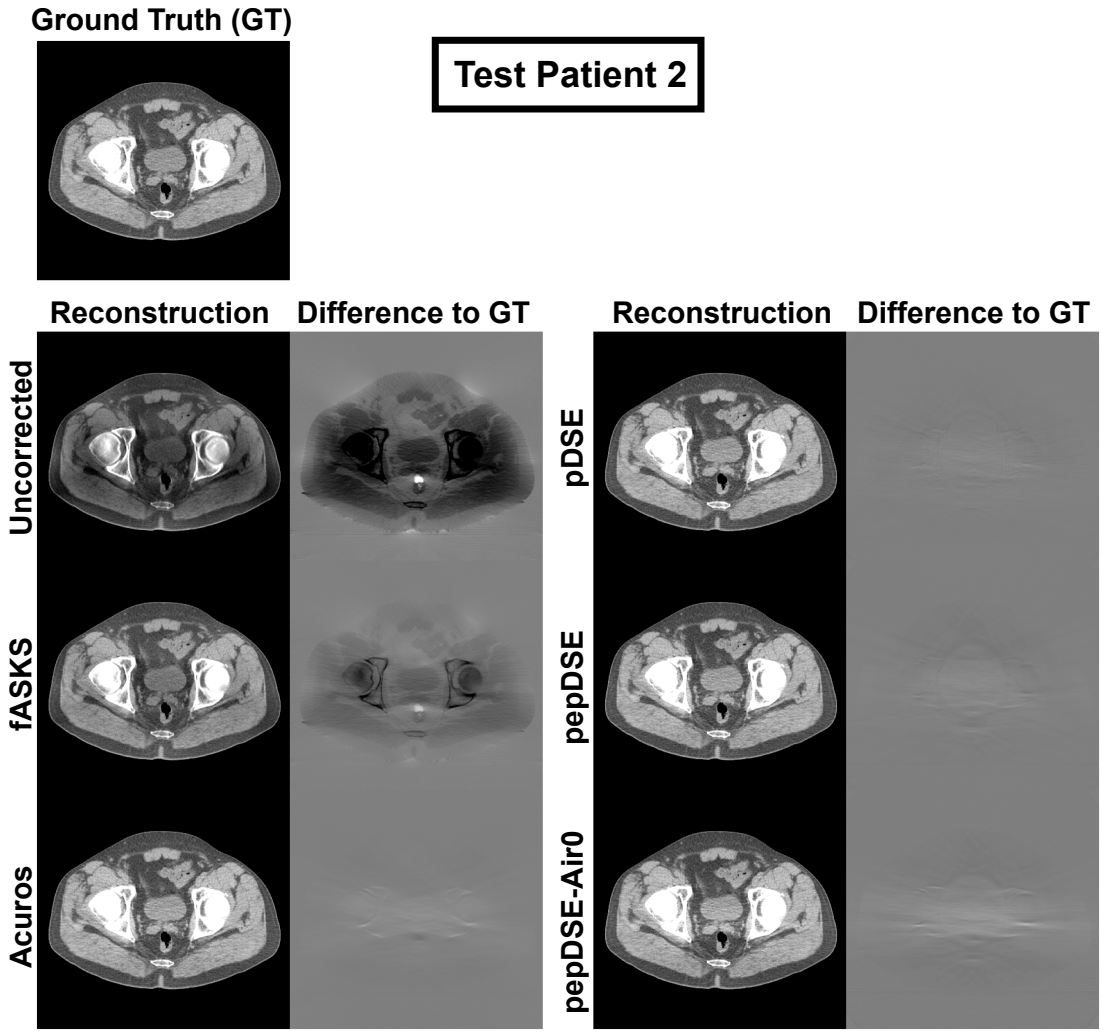

Figure 4.7: A slice of a pelvis scan of test patient two, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\text{HU},\ W = 400\,\text{HU}$ for CT and difference images.

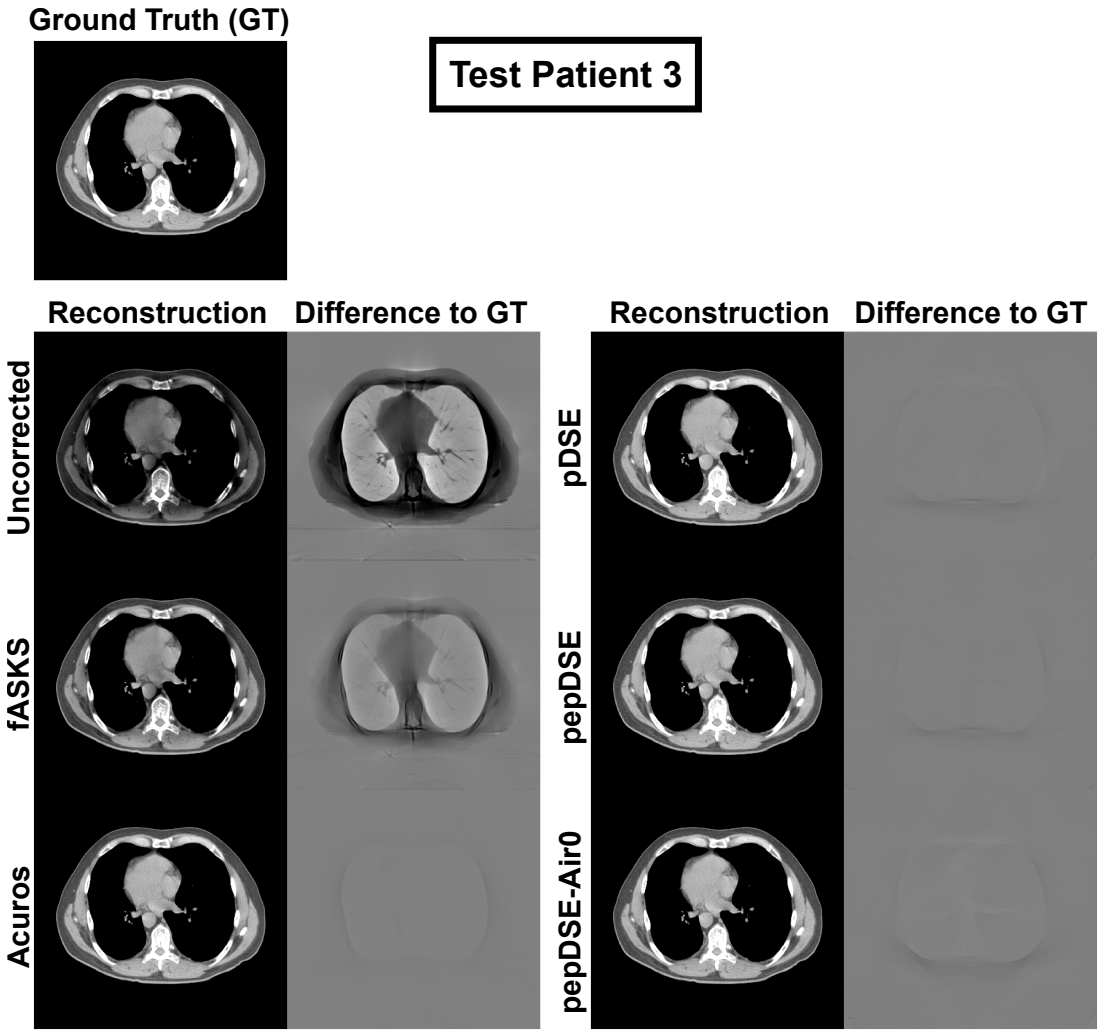

Figure 4.8: A slice of a thorax scan of test patient three, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

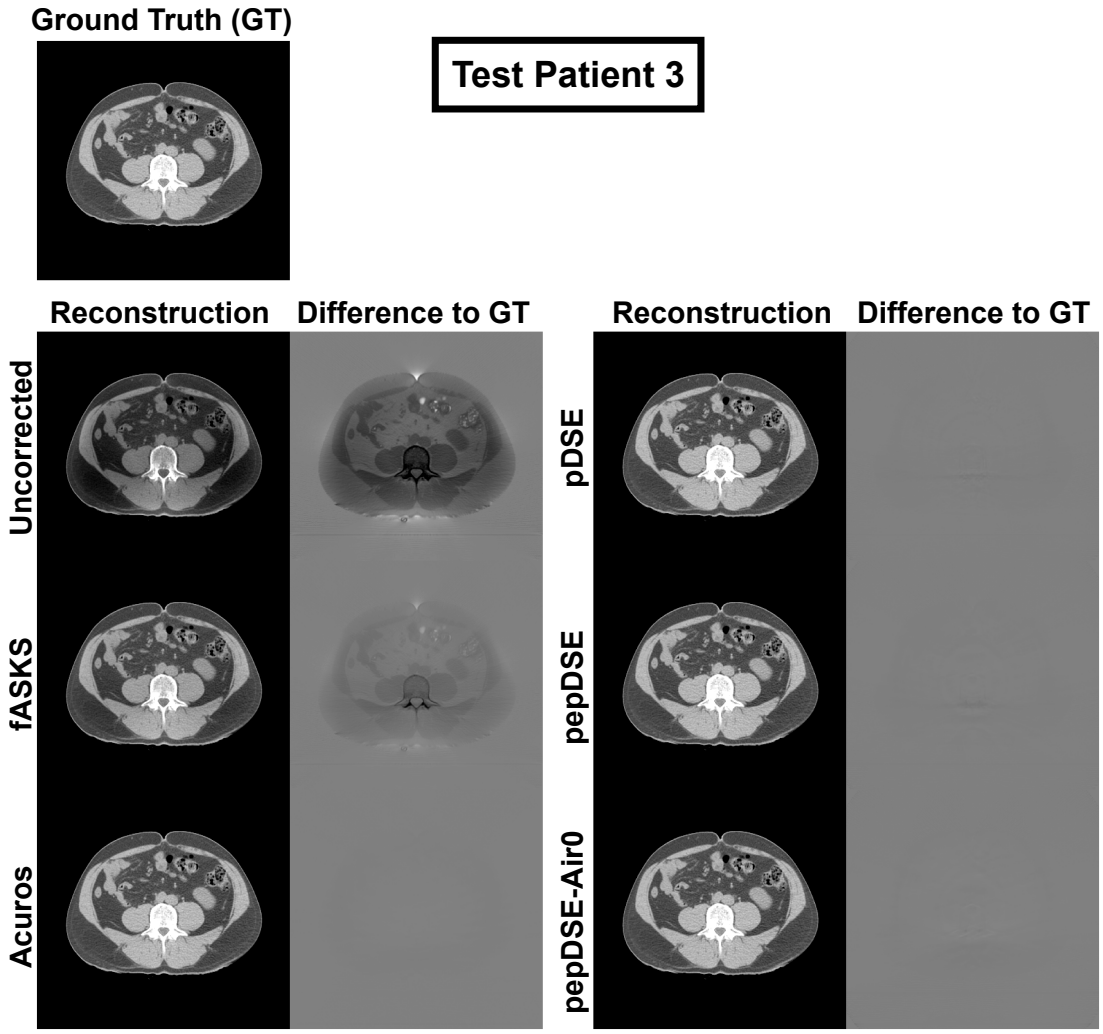

Figure 4.9: A slice of an abdomen scan of test patient three, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

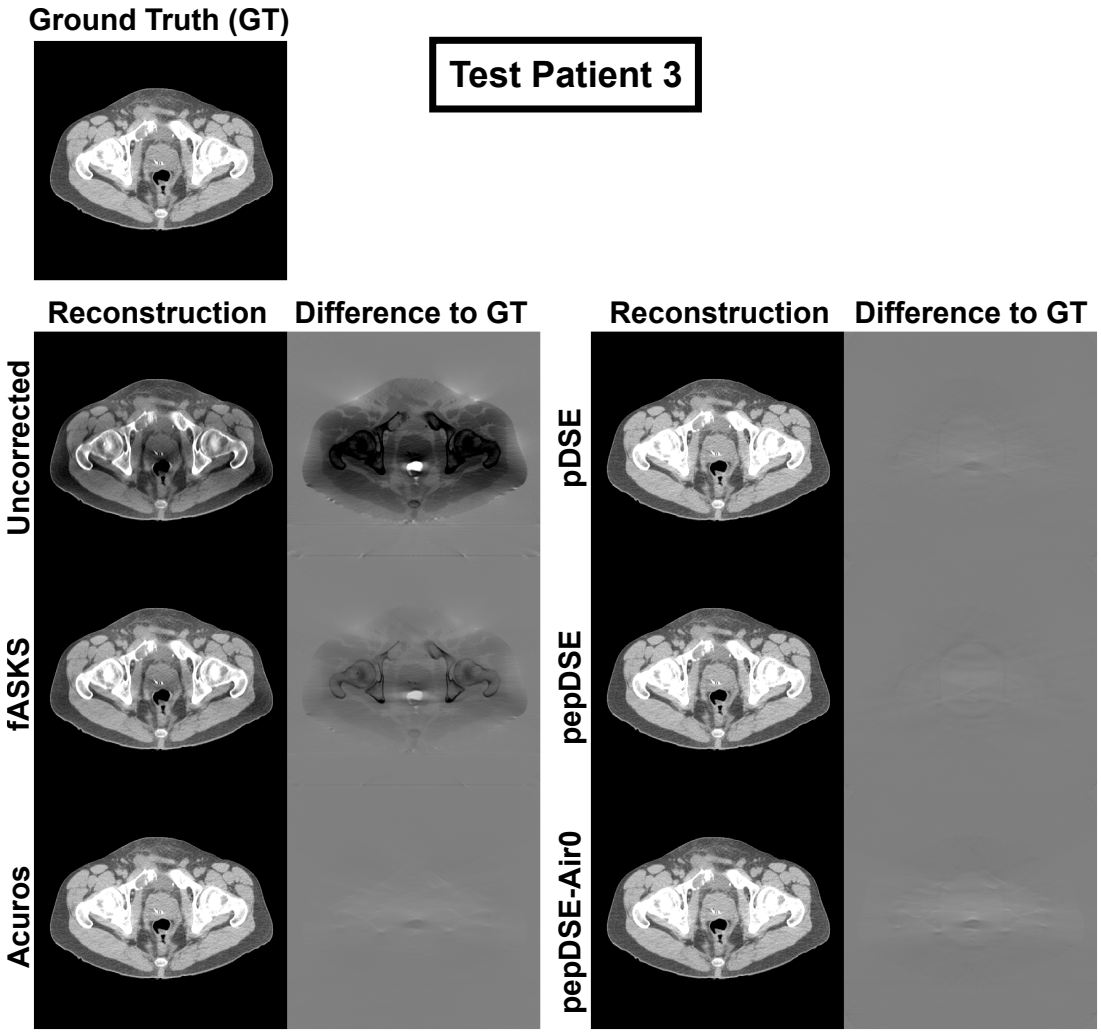

Figure 4.10: A slice of a pelvis scan of test patient three, reconstructed with different scatter correction methods with the difference to the ground truth. $C = 0\,\mathrm{HU},\ W = 400\,\mathrm{HU}$ for CT and difference images.

4.1.3 Quantitative Comparison of the Scatter Correction Methods

In addition to the qualitative comparison, a quantitative evaluation was performed. For this purpose, the MAE, MAE-spectrum and the SSIM were calculated for each test scan as described in Section 3.1.4. The MAE is limited to the voxels within the patient contour, excluding those in the surrounding air. For this, the patient was automatically segmented by applying a threshold to the CT values and only considering closed contours with a size above a given threshold. Figure 4.11 shows a box plot for the MAE across all

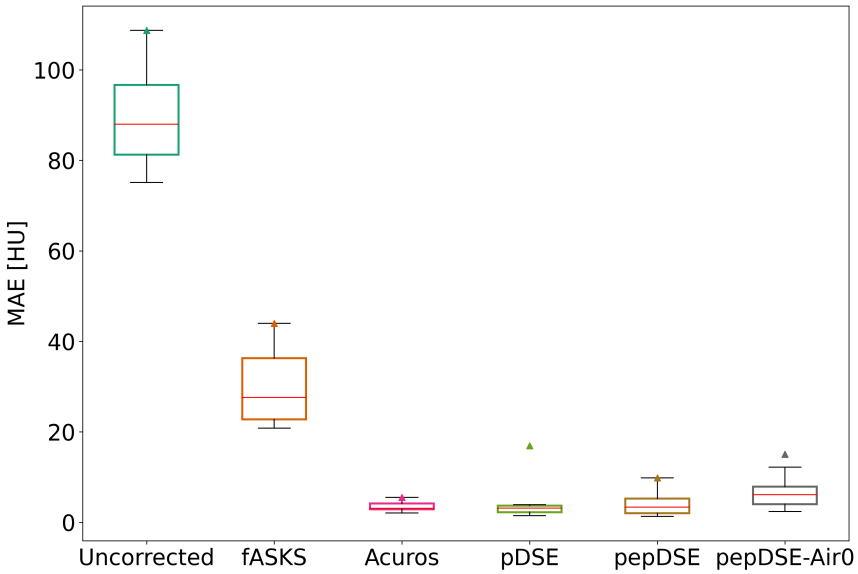

Figure 4.11: The mean absolute error (MAE) is shown to assess the image quality of the different scatter correction methods. Each box plot is drawn over the mean SSIM over all test scans and illustrates the 25^{th} and 75^{th} percentiles. The red line marks the median and the triangle the worst MAE for every box plot.

test scans for the different correction methods. The triangle marker indicates the worst performance. It is noteworthy that the least effective correction method is fASKS. While the median MAE is reduced by over 70 %, the reduction for both Acuros and DSE is over 96 %. Furthermore, the box is smaller for Acuros and all DSE variants, highlighting a better stability. The median MAEs are 88 HU, 27.6 HU, 3.1 HU, 3.2 HU, 3.4 HU, 6.1 HU, for the uncorrected, fASKS-, Acuros-, pDSE-, pepDSE- and pepDSE-Air0-corrected scans, respectively. In addition, the mean MAEs are 89.8 HU, 30.2 HU, 3.5 HU, 4.3 HU, 4.2 HU, 7.0 HU, respectively. These results indicate, that the DSE outperforms fASKS, the current gold standard projection-based scatter correction for Varian CBCT scanners. Moreover, it is comparable to Acuros, a state-of-the-art image-based scatter correction algorithm. The box plots for both pepDSE variants are slightly larger than those for pDSE. This may be attributed to a diminished capacity for generalization in networks trained with pep. It is notable that both pDSE and pepDSE-Air0 display an outlier, a

phenomenon that is discussed in greater detail in Section 4.1.4.

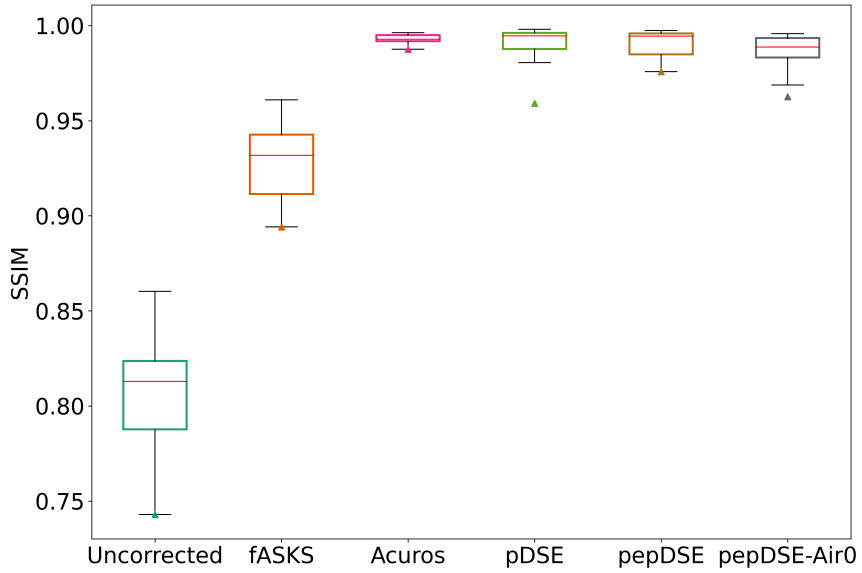

Figure 4.12: The structural similarity index (SSIM) is shown to assess the image quality of the different scatter correction methods. Each box plot is drawn over the mean SSIMs over all test scans and illustrates the 25^{th} and 75^{th} percentiles. The red line marks the median and the triangle the worst SSIM for every box plot.

As a subsequent step, the widely used image quality metric SSIM was calculated for every scan. Before calculating the SSIM on the reconstructed images, each image was scaled to values between zero and one. In order to incorporate the fact that different anatomical regions are viewed with a different windowing, the windowing settings were considered for by scaling the CT image I as

$$I_{\text{scaled}} = \begin{cases} 0, & \text{for } I < C - \frac{W}{2} \\ \frac{I-\left(C-\frac{W}{2}\right)}{W}, & \text{for } C - \frac{W}{2} \leq I \leq C + \frac{W}{2} \\ 1, & \text{for } I > C + \frac{W}{2}\,. \end{cases} \tag{4.2}$$

The chosen windowing setting were $C = 40\,\text{HU}$, $W = 400\,\text{HU}$ for pelvis, $C = 50\,\text{HU}$, $W = 350\,\text{HU}$ for abdomen and $C = -50\,\text{HU}$, $W = 400\,\text{HU}$ for thorax scans. Figure 4.12 displays the SSIMs for all test scans as box plots. The box plots show the 25^{th} and 75^{th} percentiles, with the median highlighted by a red line and the worst case marked by a triangle. Notably, the image quality improves from uncorrected to fASKS-corrected scans and even more if Acuros or DSE are used. The median SSIMs is 0.813, 0.932, 0.993, 0.995, 0.995, 0.989, for the uncorrected, fASKS-, Acuros-, pDSE-, pepDSE- and pepDSE-Air0-corrected scans, respectively. Thus, fASKS shows an improvement of 14 %, while both Acuros and DSE demonstrate an even more pronounced improvement

of approximately 22 %. Similar to the MAE, fASKS has the largest deviation between the 25$^{\text{th}}$ and 75$^{\text{th}}$ percentile. Acuros has the smallest difference, closely followed by DSE. Here, there is less difference between each DSE variant. Nonetheless, the outlier for pDSE and pepDSE-Air0 are visible. All DSE variants were found to outperform fASKS and are on a par with Acuros, with both methods achieving an almost perfect SSIM of one. This highlights the good performance of DSE which is applied in the projection domain, the same as fASKS.

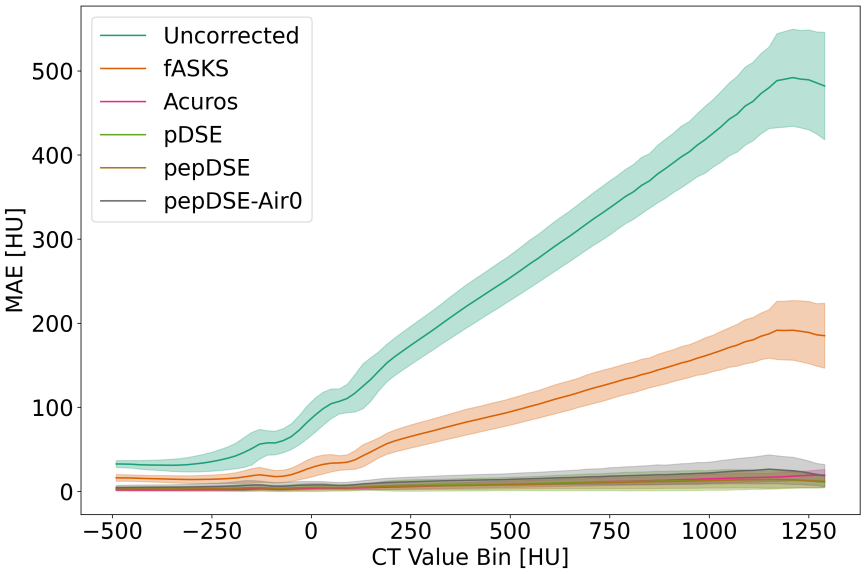

Figure 4.13: The mean absolute error spectra of the scatter-corrected reconstructions with respect to the ground truth are presented. Voxels are binned according to their CT value in bins of 20 HU width, and the MAE is calculated for each bin separately. Each line plot represents the mean over all test scans, and the colored area corresponds to one standard deviation.

Figure 4.13 depicts the final image metric, the MAE-spectrum as described in Section 3.1.4. The line plot connects the mean MAE over all patients of each bin and the colored region corresponds to one standard deviation. As with the other metrics, fASKS demonstrates a notable decline in scatter artifacts, which consequently results in a reduction of the overall MAE. Notably, scatter artifacts are stronger for voxels with higher CT values. This is a consequence of the higher attenuation by bones. On the one hand, the stronger attenuation results in a lower signal at the detector. In addition, the scatter is a smooth distribution that does not change as substantial as the primary attenuation signal. Consequently, the scatter-to-primary ratio is considerably higher for projection values from photons passing through bones, and a higher error is visible in the reconstruction. An error of over 50 HU is still present in fASKS-corrected images for CT values above 250 HU, and an error of over 100 HU is present for CT values above 750 HU. All DSE variants and Acuros demonstrate significantly reduced errors compared to

fASKS, particularly for higher CT values. The mean and standard deviation curves are overlapping for Acuros and DSE, and only a minimal increase in error is visible for higher CT values. pepDSE-Air0 again demonstrates a slightly worse performance than the other two DSE versions, as evidenced by the larger standard deviation and slightly higher mean. Nevertheless, this highlights the excellent performance of DSE.

4.1.4 Outlier Case

The quantitative evaluation illustrates an outlier for pepDSE-Air0 and pDSE, which is here presented in greater detail. Figure 4.14 shows in the top row the central transversal and coronal slice of the outlier scan for pepDSE-Air0. A red arrow highlights high attenuation streaks between the elbow bone and the upper arm bone. The pepDSE-Air0 model demonstrates an overestimation of scatter in this case, which is consistent with the findings observed in the pelvis scans, where scatter was also overestimated for high-attenuation projection values. In the pelvis reconstructions, the overestimation was less severe and not visible in the CT image, only in the difference image. One potential reason for the increased artifacts could be the patient shape. As it was shown by Maier et al. [24] the network performance drops if it is applied to projection data of anatomical regions not included in the training. In the outlier case, the pre-processing removed parts of the arm altering the anatomy. The left-over arm is indicated in Figure 4.14 by an yellow arrow. This resulted in an irregular patient shape that was not included in the training. In addition, patient scans without arms are more common in the training data, because many patients are scanned with the arms above the head to spare the healthy tissue. This is also evident in the previously shown test patients, e.g. Figures 4.2 to 4.7. Removing the arm completely reduces the artifacts and the MAE goes from 12.2 HU down to 5.8 HU, below the median of 6.1 HU. This finding backs the hypothesis, that the irregular patient shape caused by the cut-off arm which was not included in the training increases the error.

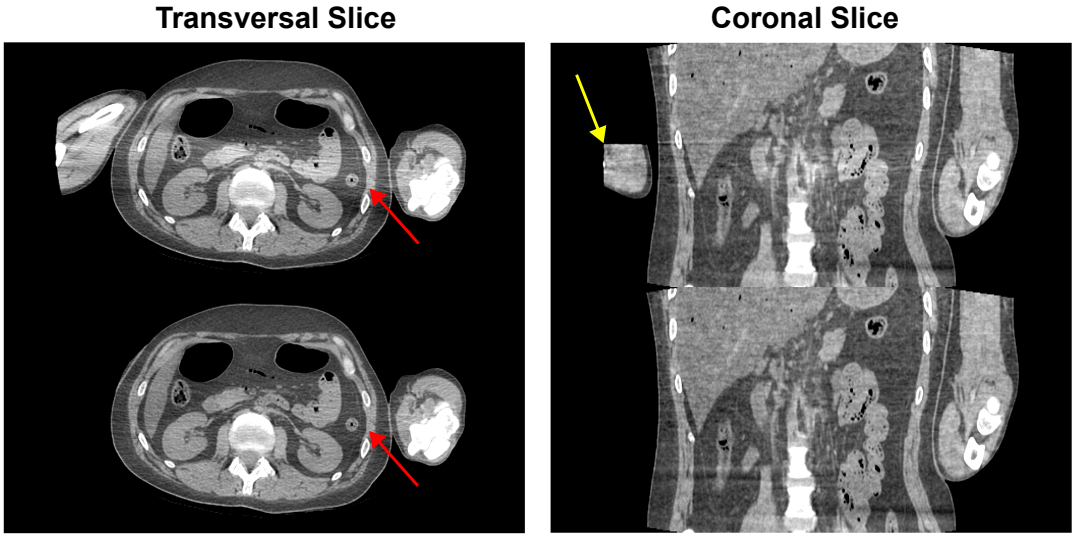

Figure 4.14: In the top row, the central transversal slice and coronal slice of the outlier scan for pepDSE-Air0. The red arrow highlights high attenuation streaks between the elbow bone and the upper arm bone. These streaks are caused by an overestimation of the scatter intensity. Furthermore, a yellow arrow marks a arm which was cut-off in a preprocessing step. The bottom shows the same scan, but before simulation the cut-off arm was completely removed, reducing the error. $C = 0\,\text{HU}$, $W = 400\,\text{HU}$.

4.1.5 Scatter Reduction for a Phantom Measurement

Following the quantitative assessment of simulated CBCT scans, the subsequent section will address the qualitative evaluation on a measured pelvis phantom scan. Figure 4.15 shows the central transversal and sagittal slice of an uncorrected, fASKS-, Acuros- and DSE-corrected scan. In the uncorrected scan, the scatter artifacts are clearly visible as dark regions. In particular, the light-gray soft tissue, which should have uniform CT values, has dark areas throughout. fASKS reduces the occurrence of scatter artifacts in the reconstruction. Nevertheless, as highlighted by the red arrow, it is unable to fully eliminate all instances of scatter artifacts, particularly in regions in close proximity to bones, where a darkening is noticeable.

The reduction in scatter artifacts is also visible in the sagittal slice. A prominent dark area is visible in the bottom part of the phantom, as indicated by the yellow arrow in the sagittal slice. In images with scatter correction, the affected area shows less darkening. Again, fASKS cannot correct for all artifacts. Acuros shows more uniform CT values than fASKS, but less brightening than pDSE and pepDSE-Air0. It is to note, that there are other artifacts, i.e. CBCT artifacts, which are uncorrected. Thus, the resulting dark areas and artifacts in the frontal slice may have a different origin than that of scatter. Therefore, remaining artifacts do not necessarily indicate a worse

performance of the scatter correction.
Notably, an artifact that is more prominent in the scatter-corrected images is noise. As illustrated by the pink arrow in Figure 4.15, thin streaks become more evident in the image after scatter correction. The visibility of these streaks is directly correlated with the reduction in scatter. They are noise artifacts, which are predominantly visible as horizontal streaks, close to bones. The attenuation is higher for these projections, resulting in a lower signal and a lower signal-to-noise ratio. Subtracting the scatter from the signal results in an additional reduction of the signal and a further decrease in the signal-to-noise ratio. This, in turn, leads to an increase in the strength of noise artifacts. In addition, DSE introduces some minor ring artifacts, which are visible in the transversal slice.
pepDSE induces strong artifacts in the reconstruction. The scatter is heavily overestimated creating bright streaks and areas in the image. For the DSE variants pDSE and pepDSE the projection values in air are not ignored. Notably, the projection values in air from the simulated projections show a deviation from the measurement. As described in Section 3.1.3 the negative values come from the scatter added to the primary signal, which has not been attenuated in air. There, more scatter is related to more negative values and a lot of information of the scatter intensity is included in air. In the measurement, other effects, such as spectrum shifts in the air norm due to the heat-up of the source, can lead to larger negative values. It appears that pepDSE takes more information from air than the other DSE versions, and overestimates the scatter significantly due to the presence of more negative values in air than during training. pDSE demonstrates greater generalizability, while pepDSE-Air0 excludes the values in air, both performing well on the measurement.
In general, the same patterns that were identified in the quantitative assessment are also evident in this qualitative analysis. fASKS is outperformed by Acuros and DSE. Furthermore, training data representing real measurements are critical for a functional deep scatter estimation, particularly for DSE trained with pep.

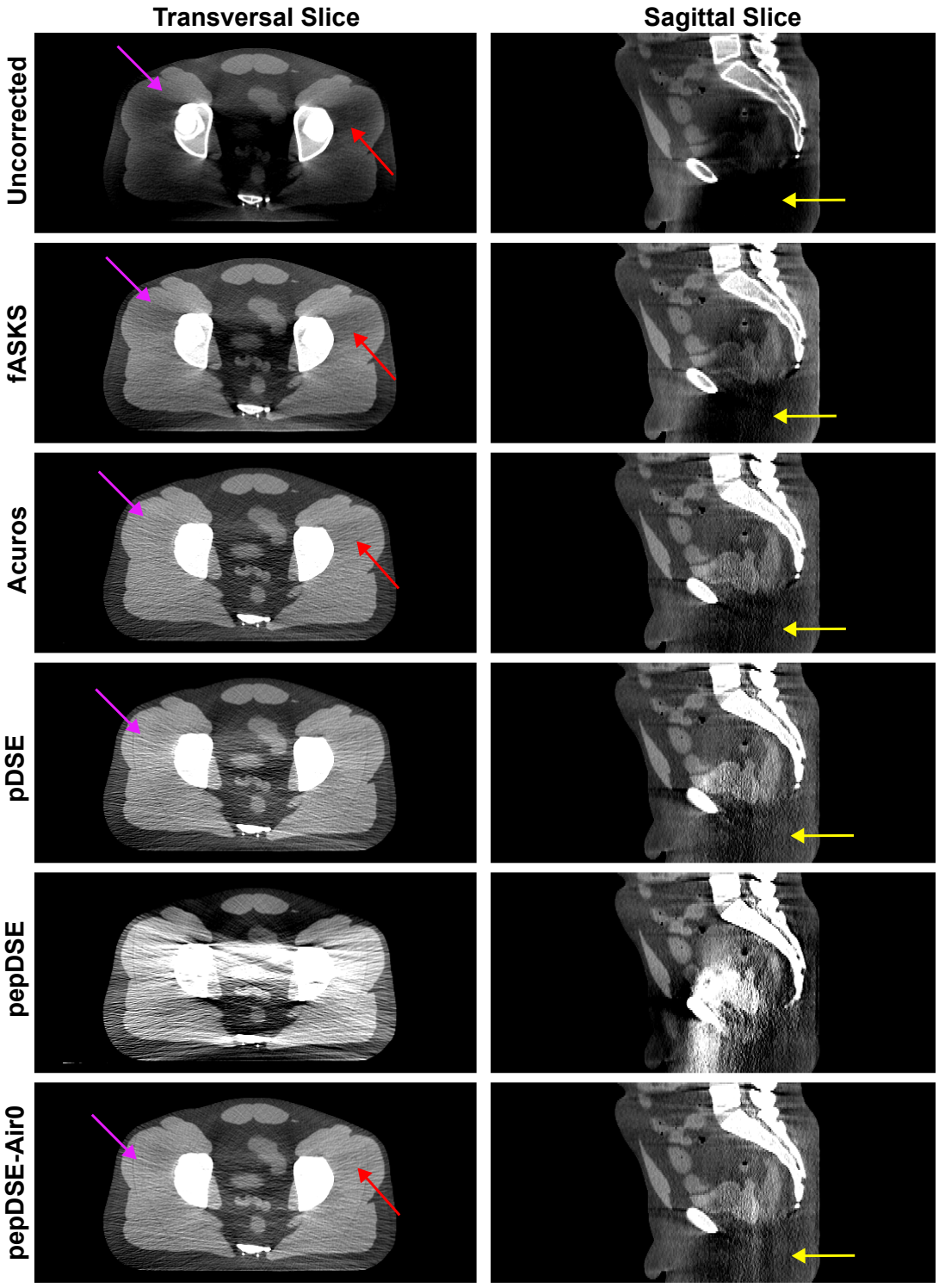

Figure 4.15: Central transversal slice and sagittal slice of a pelvis phantom measurement with different scatter correction methods applied. The red and yellow arrows highlight areas with strong scatter artifacts and the pink arrow the thin streaks caused by noise. $C = 40\,\mathrm{HU}$, $W = 400\,\mathrm{HU}$.

4.1.6 Comparison of the Computation Time

Besides the image quality, the time needed for the correction is important for radiation therapy. The time between imaging and treating the patient should be as short as possible to reduce errors caused by motion. Furthermore, for online imaging during treatment the imaging process has to be real-time applicable. Table 4.1 displays the time needed to estimate the scatter distribution for a full scan with 624 projections of size 384×3072. Acuros relies on a fASKS-corrected *prior* reconstruction. In order to reduce the computation time, the projections for DSE are downsampled to a size of 128×384 which is possible because of the low frequency nature of the scatter. In comparison to Acuros, DSE has a 34-fold increase in speed and is applied in the projection domain. Consequently, the correction can be initiated without delay upon the acquisition of the initial projection, in parallel to the acquisition of the remaining projections. The necessity for a full set of projections would exclude the use of Acuros as a potential correction method for online treatment. Note that the reconstruction tool used could not measure the time for fASKS as for the other two methods. Instead, the mean time difference of three scans with and without fASKS is displayed. Thus, the time may include other effects like a slower data loading, writing or other reconstruction processes. However, DSE is more than twice as fast as fASKS. Notably, all methods are optimized and executed on a single NVIDIA RTX A5000 GPU.

Table 4.1: The wall time for the scatter estimation for a full scan with 624 projections. The pre-processing time for Acuros contains the fASKS-corrected *prior* reconstruction. In the case of DSE, the pre-processing stage involves the downscaling of the projections, which serves to reduce the computational time and memory usage during both training and inference.

Method	Pre-processing	Single Projection	Full Scan	Total Time
Acuros	37.8 s	208 ms	129.5 s	167.3 s
DSE	0.8 s	7 ms	4.1 s	4.9 s
fASKS	-	16 ms	10.2 s	10.2 s

4.2 Projection-based Beam Hardening Correction

This section presents the evaluation of the proposed projection-based beam hardening corrections. Initially, a qualitative comparison of reconstructed pelvis scans is provided by examining an example slice of beam hardening-corrected images. Then, quantitative image quality metrics are evaluated. Finally, the results of the investigation into the robustness of DBB with respect to different tube spectra and anatomical regions are presented.

4.2.1 Example Slice of Beam Hardening-Corrected Scans

Figure 4.16 displays the central slice and the difference to the monochromatic ground truth at 70 keV. For a more realistic presentation of the uncorrected scan, it was scaled with a linear factor which was found by minimizing the MSE between the uncorrected image and the ground truth. The uncorrected reconstruction shows strong beam hardening in the soft tissue as a bright cupping. Additionally, the bones are clearly distinguishable in the difference image. Cupping artifacts are completely removed by the water precorrection, but the secondary artifacts in bone remain, as highlighted by red arrows. In addition to the error in bone, there are dark streaks in the soft tissue surrounding bones. All three proposed beam hardening correction networks can reduce the error in bone. DBE has the least error left in bone. In contrast, DBB reduces the difference in bone if compared to the water-corrected scan, but less than the other correction methods. However, it shows less errors in soft tissue. Both the direct prediction of p_{mono} and the DBE exhibit a higher difference in soft tissue than the water precorrection or DBB.

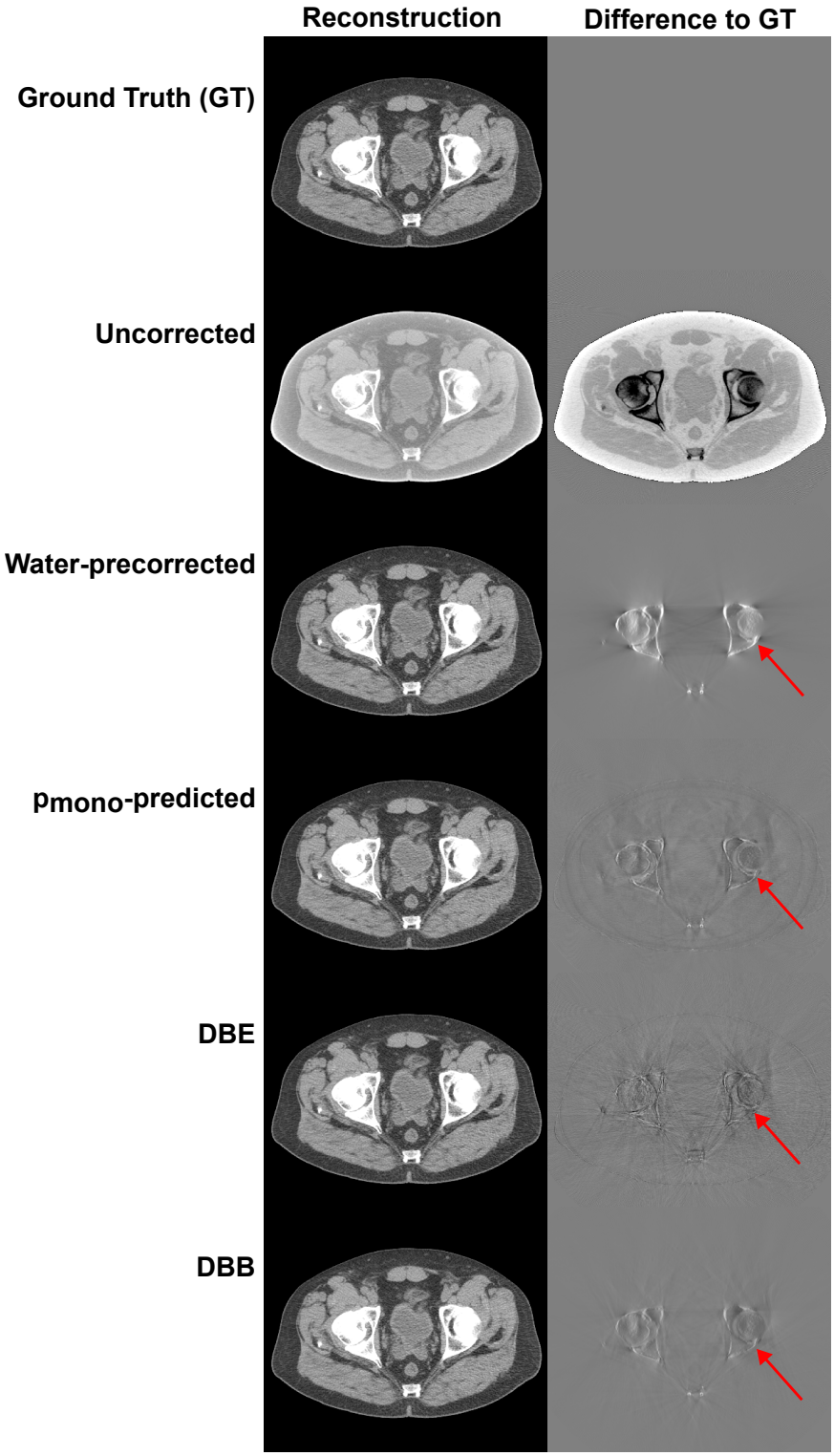

Figure 4.16: Slice of the different beam hardening correction methods. All images are calculated to show the image at 70 keV. $C = 40\,\mathrm{HU}$, $W = 400\,\mathrm{HU}$ for reconstructions and $C = 0\,\mathrm{HU}$, $W = 100\,\mathrm{HU}$ for difference images, $C = 0\,\mathrm{HU}$, $W = 400\,\mathrm{HU}$ for the difference of the uncorrected scan.

4.2.2 Quantitative Comparison of the Beam Hardening Correction Methods

After the qualitative comparison, the quantitative metrics, MAE and MAE-spectrum, are evaluated to assess the image quality. Figure 4.17 plots the MAE for all test scans as box plots. To include only important voxels, the air is always excluded from the MAE calculation. Medians are 109 HU, 5.5 HU, 3.5 HU, 4.2 HU, and 2.7 HU for the uncorrected, water-corrected, p_{mono}-predicted, DBE-, and DBB-corrected scans. Thus, DBB corrects the remaining error by 50 % if compared to the water precorrection. In addition, it has the lowest MAE error. DBE performs a little bit worse than the direct prediction of the monochromatic projection p_{mono}.

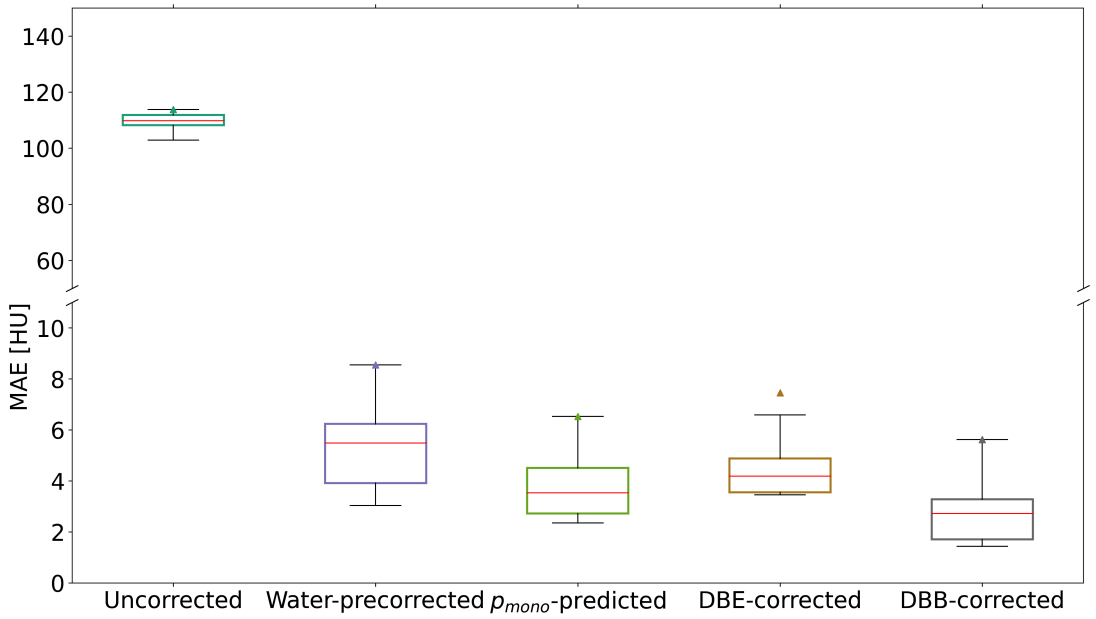

Figure 4.17: The mean absolute error (MAE) is shown to assess the image quality of the different beam hardening correction methods. Each box plot is drawn over the MAE over all test scans and illustrates the 25^{th} and 75^{th} percentiles. The red line marks the median and the triangle the worst MAE for every box plot. Note the gap and change in the y-axis scaling for better visualization due to the large errors in the uncorrected scan.

As illustrated in Figure 4.16, after a water precorrection most of the remaining error is in bones. The MAE is computed over all tissues at the same time, a drawback not shared by the MAE-spectrum. While it cannot distinguish between different tissues, binning the voxels with respect to their CT value is analogous to a simple segmentation. Figure 4.18 plots the mean MAE over all test images of each bin and the colored area corresponds to one standard deviation. There, the good performance of the water precorrection is immediately apparent. Only for bins with a CT value greater than

250 HU the error gets significant. This is to be expected, as a higher CT value in the GT indicates denser tissue, i.e. bone, and thus a greater deviation from the attenuation with respect to water. Similar to the qualitative evaluation, DBB reduces the error in bone better than the water precorrection, but it performs less optimal for higher CT values than the other two proposed methods. While the prediction of the monochromatic projection and the DBE have a better error reduction in bone, they perform slightly worse in soft tissue. This results in a better overall MAE reduction for the DBB as it is evident in Figure 4.17.

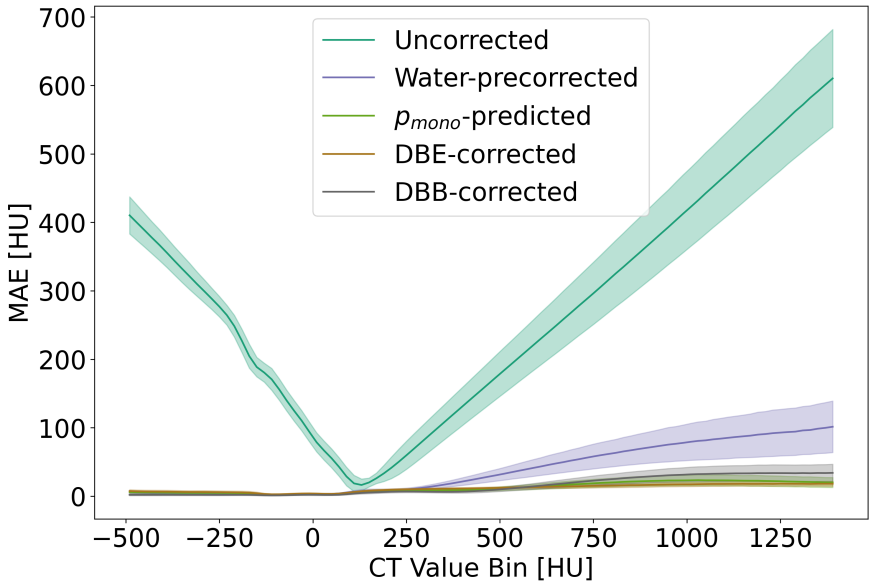

Figure 4.18: The mean absolute error spectra of the scatter-corrected reconstructions with respect to the monochromatic ground truth are presented. Voxels are binned according to their CT value in bins of 20 HU width, and the MAE is calculated for each bin separately. Each line plot represents the mean over all test scans, and the colored area corresponds to one standard deviation.

4.2.3 Generalization to Different Tube Spectra

The generalization of the DBB approach with respect to different tube spectra was evaluated as described in Section 3.2.8. Table 4.2 shows the quantitative results for DBB networks trained on four each spectrum separately and on all data at once. In general, the network trained on all spectra outperforms the networks that were only trained on a single spectrum, for all test sets. This also holds true for projections calculated from the tube spectrum used during the training of a single spectrum network. However, compared to the variation of the anatomical region (see next section), the performance reduction is rather small. For the networks trained on a single spectrum,

Table 4.2: The mean absolute percentage error of the network predictions and the ground truth label. The networks were either trained with projections from a single spectrum or with all spectra.

		Testing				
	Training	100 kV	125 kV	100 kV + Sn	125 kV + Sn	All data
DBB	100 kV	1.36 %	1.24 %	1.45 %	1.47 %	1.32 %
	125 kV	0.98 %	0.88 %	1.08 %	1.12 %	1.02 %
	100 kV + Sn	1.32 %	1.20 %	1.25 %	1.21 %	1.25 %
	125 kV + Sn	1.06 %	0.99 %	1.00 %	0.97 %	1.01 %
	All spectra	0.77 %	0.71 %	0.76 %	0.73 %	0.74 %

it is noteworthy that the networks trained on a tube voltage of 100 kV perform worse on the test data from a 100 kV tube than the networks trained with data from a 125 kV tube. In addition, the networks trained with data from a tube voltage of 100 kV show the highest MAPE for all test data sets.

4.2.4 Generalization to Different Anatomical Region

The generalization of the DBB approach with respect to different anatomical regions was evaluated as described in Section 3.2.8 by simulating projections for thorax, abdomen and pelvis scans for training and testing. Table 4.3 shows the MAPE calculated between the labels and the different DBB network predictions. Similar to the robustness study with different tube spectra, the network, trained on data including all three anatomical regions, outperforms the networks trained only on projections of a single anatomical region. This is also the case when the networks are evaluated on the anatomical region they were trained on. The MAPE increases by more than three times for the network trained on pelvis projections only, when comparing the evaluation of pelvis to the thorax projections. Figure 4.19 shows an example input, label and predictions for a thorax projection. A red arrow marks a rib bone in all projections and highlights that the DBB only trained on pelvis or abdomen projections cannot remove the ribs completely. Additionally, some residual bone structures are marked with a blue arrow. In contrast, the label, the prediction of the network, trained on all anatomical regions, and the network trained on thorax projections do not show any residual bones.

Table 4.3: The mean absolute percentage error of the network predictions and the ground truth label. One network was only trained with pelvis projections, while the other also included projections from thorax and abdomen.

		Testing			
	Training	Pelvis	Abdomen	Thorax	All data
DBB	Pelvis	0.87 %	1.75 %	2.62 %	1.75 %
	Abdomen	1.49 %	1.20 %	2.33 %	1.67 %
	Thorax	2.19 %	1.56 %	1.20 %	1.65 %
	All data	0.73 %	0.81 %	0.91 %	0.82 %

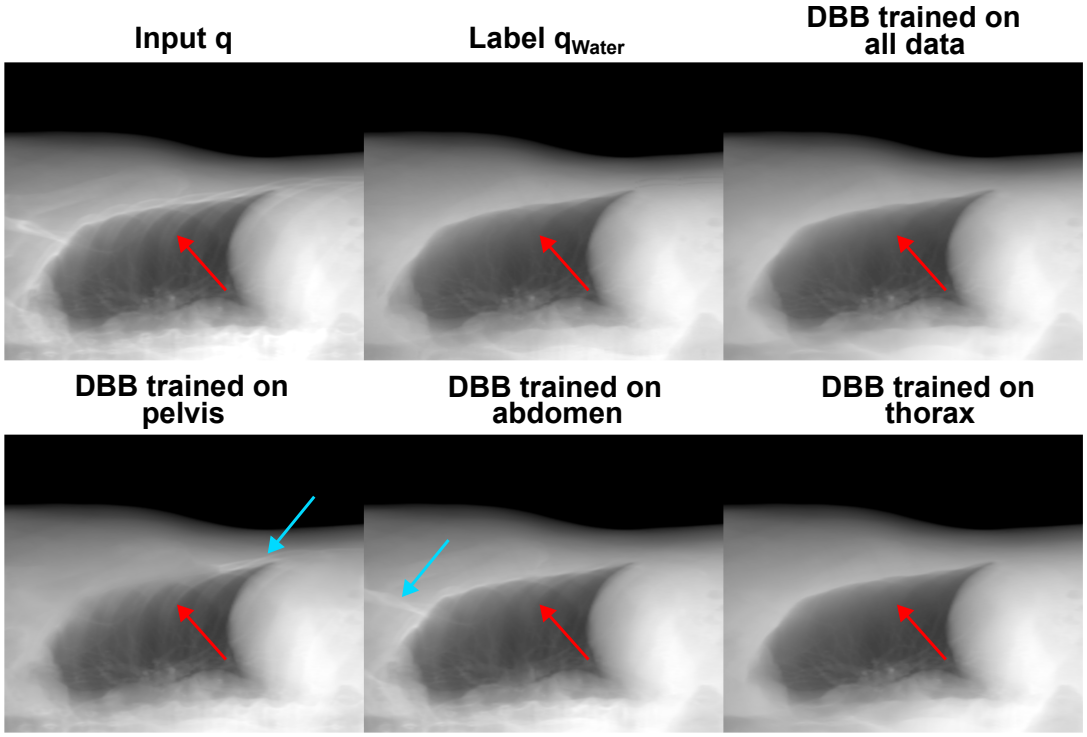

Figure 4.19: Input, label and deep bone blending predictions of DBB networks trained on different anatomical regions. The red arrows mark the position of a rib bone and the blue arrows highlights residual bones in the network predictions. $C = 3,\ W = 6$.

5 | Summary & Discussion

This thesis investigated the image quality improvements of CBCT images by addressing two major artifacts: scatter and beam hardening. For this, deep learning-based methods were developed and adapted for real-time applicable artifact corrections in the projection domain. The results of the scatter correction are discussed in the following section. Then, the results of the beam hardening correction are summarized and reviewed.

5.1 Scatter Artifact Correction

Scatter is one of the most severe artifacts in CBCT, with a significant impact on the image quality and the CT value accuracy. Modern on-board CBCT scanners, such as the Ethos [84], [85], utilize a combination of hardware, an anti-scatter grid, and software, e.g. Acuros, to reduce scatter artifacts [13], [85]. Varian treatment machines with an on-board CBCT are deployed around the world and use an image-based algorithm as state-of-the-art scatter correction method [13], [85]. Image-based methods require a *prior* reconstruction, which necessitates a good first-pass reconstruction. Moreover, a full scan has to be acquired to reduce sparseness artifacts. This thesis presents a comparative analysis of deep scatter estimation in the projection domain and existing state-of-the-art scatter correction methods, including image-based algorithms.
It is essential to utilize an appropriate simulation tool in order to generate training data from a distribution that is as close as possible to the distribution of the measured data. During the training, the neural network learns suitable features to accurately predict the scatter from simulated and labeled data. Otherwise, a network trained on simulated data demonstrates suboptimal performance when evaluated on measured data. DSE is commonly trained with MC simulations, but in this thesis a novel approach of using a deterministic solver of the linear Boltzmann transport equation was investigated. The deterministic solver used for the training data generation is Acuros, which also serves as the reference image-based correction method.
DSE trained with Acuros demonstrates clear advantages over Acuros in clinical applications. This is due to the fact that the input data needs to be downsampled for Acuros for reasons of efficiency during clinical applications, which results in the emergence of visible artifacts such as nonphysical stripe patterns in the scatter distribution. This was highlighted in the comparison of the scatter distributions displayed in Figure 2.7. In both scatter distributions, the scatter predicted by DSE and the ground truth scatter, the stripe pattern is not visible. This is due to the fact that the generation of training

data is not primarily constrained by time. Consequently, Acuros can be used with less downsampling of the input data, resulting in smooth scatter distributions for training which gives DSE an advantage over Acuros in clinical applications.
Time is especially critical for imaging in radiation therapy. The patient is in most cases placed on the treatment bench and sometimes even fixated to reduce movement. After the scan, the data obtained must be reconstructed and the treatment plan reviewed or even updated [6]. During this time, movements of the patient decrease the accuracy between actual and scanned anatomy, consequently affecting the precision of the treatment plan. In addition, motion such as gas bubbles in the abdomen cannot be controlled by the patient, which can lead to further discrepancies. Therefore, fast correction methods are essential for a more accurate treatment. Another aspect of motion are motion artifacts in the reconstructed image. They can have a strong influence on image quality [86] especially in CBCT scans with scan times of six or more seconds for state-of-the-art systems [85], [86]. While DSE requires only a single projection that is motion-free, Acuros requires a first-pass reconstruction which may contain motion artifacts due to the long scan time. In general, Acuros is much slower than DSE. Table 4.1 contains the average time to estimate the scatter intensity for a single projection. DSE needs 7 ms which makes it real-time applicable and over 29 times faster than Acuros with 208 ms. The second reference method, which is projection-based and requires 16 ms for a single projection, is approximately half as fast as DSE. A full scan including pre-processing, the full first-pass reconstruction for Acuros and the downsampling of the input projections for DSE, takes 167.3 s, 10.2 s and 4.9 s for Acuros, fASKS and DSE, respectively.
fASKS and Acuros both represent the current Varian standard of clinically utilized scatter correction methods. fASKS reduces the error in the median MAE by 70 % from 88 HU to 28 HU, while DSE and Acuros can reduce the error to 3 HU, which is a correction of about 96 %. Additionally, the assessment of the MAE-spectrum indicates that scatter results in increased error in bone, and that fASKS is unable to correct the scatter in bone as effectively as in soft tissue. The reconstructed slices in Figures 4.2 to 4.7 further highlight the increased error in the difference image, which shows a greater deviation to the ground truth in bone. Acuros and DSE only show a small increase in error for higher CT values. Overall, this demonstrates that DSE is on par with state-of-the-art image-based correction methods. Notably, all simulations do not contain any other artifacts than scatter. In a measurement, additional artifacts, such as motion or truncation, can further impair the image quality of the first-pass reconstruction, which may subsequently diminish the efficacy of Acuros. DSE, on the other hand, was already shown to perform well on truncated scans by Maier et al. [24] and, as discussed before, is less prone to motion. In summary, the DSEs trained with Acuros display an excellent performance in the quantitative as well as the qualitative evaluation.
Three DSE variants have been trained and evaluated. All variants were trained on the same data set but with a different input preprocessing. Two versions shared the same projection value transform, namely pep. pep represents a first order scatter intensity estimation and showed good performance in the initial DSE publication by Maier et al. [24]. The sole distinction is that for one pep-trained network, the negative

values in air were set to zero. Consequently, no information about the scatter in air is passed to the network, which resulted in a slight decline in performance across all image quality metrics. As evident in Figures 4.4 and 4.7 a slight overprediction of the scatter intensities is visible in the example slices. If the signal is attenuated more, the scatter-to-primary ratio increases and the estimated scatter has a larger influence on the relative signal. Thus, it is expected that a performance loss would be especially evident for projection values with a high attenuation. The decreased accuracy can be contributed to a loss of information in air. For the simulated projections the unscattered photons are not attenuated in air. Therefore, the difference to the air norm, the reason why a projection value is unequal to zero, is completely contributed to the scatter signal. Consequently, an accurate estimation of the scatter magnitude can be achieved in air. A discrepancy was identified between the simulated and measured data in air, with the measured projections showing significantly lower values. This causes the pepDSE to heavily overestimate the scatter. Interestingly, pDSE, the network trained with the projection p as input, does not perform poorly on the measurement. There are several possible reasons. First of all Maier et al. [24] investigated the robustness with respect to different anatomical regions. In their case, the network on pep also exhibits a larger error for anatomies not included in the training data than the network trained on the projection p. This leads to the assumption that networks with pep cannot generalize as well as networks trained on p. Another possibility is that during training the network trained with p learned to extract more features from inside the patient. This hypothesis would also explain the error case with the change in patient anatomy. This outlier is only visible in pDSE and the pepDSE which ignores the air.
As stated above, the simulated projections have to be very close to the ground truth. This was highlighted by the pepDSE network which created new artifacts in correction of the measurement. While this problem could be circumvented by excluding projection values in air during training, new approaches to eliminate the need of labeled data and ultimately simulated data would be an important next step. Unsupervised training would therefore be an interesting follow-up project. First experiments utilizing a cycle GAN already show promising results [87]. An alternative approach would be to leverage the fractionated treatment. In modern radiation therapy the patient is not treated a single time to apply the prescribed dose, but many times, each time applying only a fraction of the dose. Thus, a measurement is already available for the second treatment and a personalized DSE could be trained. For a more robust network, all follow-up scans could also be included in the training data.

5.2 Projection-based Beam Hardening Correction

The second part of this thesis considers beam hardening correction methods in the projection domain. Although the conventional water precorrection can greatly increase the image quality in soft tissue, it still has residual errors in and around bones. A more detailed examination of the example slice of the reconstructed images presented in Figure 4.16 revealed that the CT value in bone is higher than in the monochromatic ground truth. This would result in an overestimation of the stopping power during

treatment planning and consequently reduce the plan accuracy. All three proposed methods were able to reduce the error in bone drastically. This is evident in Figure 4.18 which shows the MAE-spectrum. As illustrated in the figure, the water precorrection drastically reduces the error, until CT values of 250 HU, after which the error increases. On one hand, DBB shows the least improvement for higher CT values of the proposed methods. On the other hand, it has the overall lowest MAE for all correction methods. In the example slice in Figure 4.16 the reason for this is apparent. While the direct prediction of the monochromatic projection and DBE show less deviation to the ground truth in the difference image in bone, more errors are visible in the soft tissue. This is expected, both methods have to correct for the projection values for soft and bone tissue, because they have to learn the mapping from the polychromatic projection q to an energy-independent variable. Slight variations in the projection can have a large influence on the reconstructed image. However, all proposed methods outperform the classic water precorrection. DBB on the other hand, only needs to correct for the contribution by bone and stays in the polychromatic domain with its prediction

In the context of this thesis, the robustness of the DBB network was investigated with respect to X-ray tube spectra and anatomical regions. The two major findings were i) it is better to train on 125 kV than on 100 kV, even if the network is applied on test data from a 100 kV tube and ii) always train on multiple anatomical regions and spectra even if it is only applied to a single spectrum or region. The latter was especially visible for the networks trained on pelvis or abdomen projections which were not able to remove the rib bones of a thorax projection.

There are currently two major limitations to the study. First of all, it does not consider a quantitative evaluation of a measurement. In standard practice, a cylinder phantom with inserts of varying electron densities is used for the evaluation. Nevertheless, as demonstrated in the robustness studies, a DBB network demonstrates reduced performance on anatomical regions not included in its training data. Therefore, the networks would need to include training examples of such electron density phantoms for a good performance. However, if the density phantom is already in the training data set it cannot be used for testing, otherwise a good performance on the electron density phantom does not allow the conclusion that the network performs well on anatomies that are not in the training data. In addition, due to the black-box nature of neural networks, the accuracy test on a cylindrical phantom may not be equivalent to the accuracy on patient scans. An option to circumvent this problem would be an anthropomorphic phantom with exact and known electron densities for the different soft tissues and bones. This phantom can be used to test the network while ensuring that this data is not included in the training set and while maintaining as close a resemblance as possible to the human anatomy. Furthermore, it is advisable to use different phantoms for each body part.

Another shortcoming of DBB is the necessity to know the spectrum. A bowtie filter will change the spectrum significantly along the u-axis of the detector or the heat up of the tube during the scan will change the emitted spectrum. It is recommended that a further robustness study should be conducted to determine the extent of error that would be caused by assuming a faulty spectrum during the calculation of p_{soft}.

Nonetheless, the spectrum is already known well enough to apply an analytical water precorrection for scanners with a bowtie filter. It can thus be assumed that the spectrum can be measured with sufficient accuracy for DBB.

6 | Conclusions

The efficiency of radiation therapy can be enhanced by modifying the treatment plan on a daily basis. At present, the on-board CBCT lacks in image quality and can only be used for registering the planning CT. This work addresses two significant error sources, scatter and beam hardening, to improve the image quality of CBCT images. In both cases, data-driven deep learning methods have been adapted, developed and compared to the correction methods currently used in state-of-the-art on-board CBCT scanners. In the case of scatter correction, the projection-based reference is outperformed in terms of image quality while the image-based reference method exhibits a similar level of image quality at a significantly slower computational speed. For beam hardening, all proposed methods decrease the error in bone tissue and show a lower deviation to the ground truth over the whole patient. Additionally, the best performing beam hardening correction was further investigated and the robustness in regards to tube spectra and anatomical region was shown if the neural network was trained on appropriate data. The proposed methods are real-time applicable and can facilitate the transition towards online treatment with on-board CBCT scanners.

Bibliography

[1] M. M. Lell, J. E. Wildberger, H. Alkadhi, J. Damilakis, and M. Kachelrieß, "Evolution in computed tomography: The battle for speed and dose," *Investigative Radiology*, volume 50, number 9, pages 629–644, 2015.

[2] G. N. Hounsfield, "Computerized transverse axial scanning (tomography). 1. description of system," *Br J Radiol*, volume 46, number 552, pages 1016–1022, 1973.

[3] T. Flohr, B. Schmidt, J. Merz, and P. Aulbach, *White paper: SOMATOM force get two steps ahead with dual source CT.*

[4] M. Lell and M. Kachelrieß, "Computed tomography 2.0: New detector technology, ai, and other developments," *Investigative Radiology*, volume 58, number 8, pages 587–601, 2023.

[5] G. Böning, W. M. Lüdemann, J. Chapiro, M. Jonczyk, B. Hamm, R. W. Günther, B. Gebauer, and F. Streitparth, "Clinical experience with real-time 3-D guidance based on C-arm-acquired cone-beam CT (CBCT) in transjugular intrahepatic portosystemic stent shunt (TIPSS) placement," *Cardiovasc Intervent Radiol*, volume 41, number 7, pages 1035–1042, 2018.

[6] W. Schlegel, C. P. Karger, and O. Jäkel, Eds., *Medizinische Physik: Grundlagen – Bildgebung – Therapie – Technik*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2018.

[7] J. G. Buckley, D. Wilkinson, A. Malaroda, and P. Metcalfe, "Investigation of the radiation dose from cone-beam CT for image-guided radiotherapy: A comparison of methodologies," *Journal of Applied Clinical Medical Physics*, volume 19, number 1, pages 174–183, 2018.

[8] T. Kaasalainen, M. Ekholm, T. Siiskonen, and M. Kortesniemi, "Dental cone beam CT: An updated review," *Med. Phys.*, volume 88, pages 193–217, 2021.

[9] T. Li, L. Xing, P. Munro, C. McGuinness, M. Chao, Y. Yang, B. Loo, and A. Koong, "Four-dimensional cone-beam computed tomography using an on-board imager," *Med. Phys.*, volume 33, number 10, pages 3825–3833, 2006.

[10] L. Lechuga and G. A. Weidlich, "Cone beam CT vs. fan beam CT: A comparison of image quality and dose delivered between two differing CT imaging modalities," *Cureus*, 2016.

[11] D. N. Stanley, J. Harms, J. A. Pogue, J.-G. Belliveau, S. R. Marcrom, A. M. McDonald, M. C. Dobelbower, D. H. Boggs, M. H. Soike, J. A. Fiveash, R. A. Popple, and C. E. Cardenas, "A roadmap for implementation of kv-cbct online adaptive radiation therapy and initial first year experiences," *Journal of Applied Clinical Medical Physics*, volume 24, number 7, e13961, 2023.

[12] M Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Phys. Med. Biol.*, volume 55, number 22, pages 6695–6720, 2010.

[13] A. Maslowski, A. Wang, M. Sun, T. Wareing, I. Davis, and J. Star-Lack, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - part I: Core algorithms and validation," *Med. Phys.*, volume 45, number 5, pages 1899–1913, 2018.

[14] J. Maier, S. Sawall, M. Knaup, and M. Kachelrieß, "Deep scatter estimation (DSE): Accurate real-time scatter estimation for X-ray CT using a deep convolutional neural network," *Journal of Nondestructive Evaluation*, volume 37, number 3, page 57, 2018.

[15] Y. Zhu, H. Zhao, T. Wang, L. Deng, Y. Yang, Y. Jiang, N. Li, Y. Chan, J. Dai, C. Zhang, Y. Li, Y. Xie, and X. Liang, "Sinogram domain metal artifact correction of CT via deep learning," *Computers in Biology and Medicine*, volume 155, page 106 710, 2023.

[16] E. Meyer, R. Raupach, M. Lell, B. Schmidt, and M. Kachelrieß, "Normalized metal artifact reduction (nmar) in computed tomography," *Med. Phys.*, volume 37, number 10, pages 5482–5493, 2010.

[17] J. Lee, J. Gu, and J. C. Ye, "Unsupervised CT metal artifact learning using attention-guided β-cycleGAN," *IEEE Transactions on Medical Imaging*, volume 40, number 12, pages 3932–3944, 2021.

[18] L. Zhu, Y. Xie, J. Wang, and L. Xing, "Scatter correction for cone-beam CT in radiation therapy," *Medical Physics*, volume 36, number 6Part1, pages 2258–2268, 2009.

[19] OpenAI, J. Achiam, S. Adler, *et al.*, *Gpt-4 technical report*, 2024.

[20] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer, "High-resolution image synthesis with latent diffusion models," in *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022, pages 10 674–10 685.

[21] Q. Li, W. Cai, X. Wang, Y. Zhou, D. D. Feng, and M. Chen, "Medical image classification with convolutional neural network," in *2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, Singapore: IEEE, 2014, pages 844–848.

[22] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Köhler, T. Norajitra, S. J. Wirkert, and K. H. Maier-Hein, "nnU-net: Self-adapting framework for U-net-based medical image segmentation," *CoRR*, volume abs/1809.10486, 2018.

[23] J. Wasserthal, H.-C. Breit, M. T. Meyer, M. Pradella, D. Hinck, A. W. Sauter, T. Heye, D. T. Boll, J. Cyriac, S. Yang, M. Bach, and M. Segeroth, "Totalsegmentator: Robust segmentation of 104 anatomic structures in CT images," *Radiology: Artificial Intelligence*, volume 5, number 5, e230024, 2023.

[24] J. Maier, E. Eulig, T. Vöth, M. Knaup, J. Kuntz, S. Sawall, and M. 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," *Med. Phys.*, volume 46, number 1, pages 238–249, 2018.

[25] A. Lalonde, B. Winey, J. Verburg, H. Paganetti, and G. C. Sharp, "Evaluation of CBCT scatter correction using deep convolutional neural networks for head and neck adaptive proton therapy," *Phys. Med. Biol.*, volume 65, number 24, 2020.

[26] M. Du, K. Liang, L. Zhang, H. Gao, Y. Liu, and Y. Xing, "Deep-learning-based metal artefact reduction with unsupervised domain adaptation regularization for practical CT images," *IEEE Transactions on Medical Imaging*, volume 42, number 8, pages 2133–2145, 2023.

[27] S. B. Park, "Advances in deep learning for computed tomography denoising," *World J. Clin. Cases*, volume 9, number 26, pages 7614–7619, 2021.

[28] M. M. Lell and M. Kachelrieß, "Recent and upcoming technological developments in computed tomography: High speed, low dose, deep learning, multienergy," *Investigative Radiology*, volume 55, number 1, pages 8–19, 2020.

[29] T. Vöth, T. Koenig, E. Eulig, M. Knaup, V. Wiesmann, K. Hörndler, and M. Kachelrieß, "Real-time 3D reconstruction of guidewires and stents using two update X-ray projections in a rotating imaging setup," *Med. Phys.*, volume 50, number 9, pages 5312–5330, 2023.

[30] J. Hsieh, E. Liu, B. Nett, J. Tang, J.-B. Thibault, and S. Sahney, "A new era of image reconstruction: TrueFidelity^{TM} technical white paper on deep learning image reconstruction," 2019.

[31] J. Erath, T. Vöth, J. Maier, E. Fournié, M. Petersilka, K. Stierstorfer, and M. Kachelrieß, "Deep learning-based forward and cross-scatter correction in dual-source CT," *Med. Phys.*, volume 48, number 9, pages 4824–4842, 2021.

[32] A. Wang, A. Maslowski, P. Messmer, M. Lehmann, A. Strzelecki, E. Yu, P. Paysan, M. Brehm, P. Munro, J. Star-Lack, and D. Seghers, "Acuros CTS: A fast, linear Boltzmann transport equation solver for computed tomography scatter - part II: System modeling, scatter correction, and optimization," *Med. Phys.*, volume 45, number 5, pages 1914–1925, 2018.

[33] T. M. Buzug, *Computed Tomography*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008.

[34] H. A. Kramers, "XCIII. on the theory of x-ray absorption and of the continuous x-ray spectrum," *Lond. Edinb. Dublin Philos. Mag. J. Sci.*, volume 46, number 275, pages 836–871, 1923.

[35] J. Hubbell and S. Seltzer, *Tables of x-ray mass attenuation coefficients and mass energy-absorption coefficients*, `http://physics.nist.gov/xaamdi`, version 1.4, Accessed: 2024-03-25, Gaithersburg, MD, 2004.

[36] H. Q. Woodard and D. R. White, "The composition of body tissues," *Br J Radiol*, volume 59, number 708, pages 1209–1218, 1986.

[37] "Essai d'optique sur la gradation de la lumière," *Nature*, volume 111, number 2784, pages 320–320, 1923.

[38] J. H. Lambert and E Anding, *Lamberts Photometrie: (Photometria, sive De mensura et gradibus luminis, colorum et umbrae) (1760)*. Leipzig: W. Engelmann, 1892.

[39] Beer, "Bestimmung der Absorption des rothen Lichts in farbigen Flüssigkeiten," *Annalen der Physik*, volume 162, pages 78–88,

[40] K. W. Fornalski, "Simple empirical correction functions to cross sections of the photoelectric effect, Compton scattering, pair and triplet production for carbon radiation shields for intermediate and high photon energies," *Journal of Physics Communications*, volume 2, number 3, page 035 038, 2018.

[41] O. Klein and Y. Nishina, "Über die streuung von Strahlung durch freie Elektronen nach der neuen relativistischen Quantendynamik von Dirac," *Zeitschrift für Physik*, volume 52, number 11, pages 853–868, 1929.

[42] J. H. Hubbell, H. A. Gimm, and I. O/verbo/, "Pair, triplet, and total atomic cross sections (and mass attenuation coefficients) for 1 MeV-100 GeV photons in elements Z=1 to 100," *Journal of Physical and Chemical Reference Data*, volume 9, number 4, pages 1023–1148, 1980.

[43] P. Lecoq, "Development of new scintillators for medical applications," *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, volume 809, pages 130–139, 2016, Advances in detectors and applications for medicine.

[44] Y. Wu, Z. Ye, J. Chen, L. Deng, and B. Song, "Photon counting CT: Technical principles, clinical applications, and future prospects," *Academic Radiology*, volume 30, number 10, pages 2362–2382, 2023.

[45] J. Radon, "Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten," *Verh. Sachs. Akad. Wiss. Leipzig, Math Phys Klass*, volume 69, 1917.

[46] G. N. Ramachandran and A. V. Lakshminarayanan, "Three-dimensional reconstruction from radiographs and electron micrographs: Application of convolutions instead of fourier transforms," *Proc Natl Acad Sci U S A*, volume 68, number 9, pages 2236–2240, 1971.

[47] L. Feldkamp, L. C. Davis, and J. Kress, "Practical cone-beam algorithm," *J. Opt. Soc. Am*, volume 1, pages 612–619, 1984.

[48] H. K. Tuy, "An inversion formula for cone-beam reconstruction," *SIAM Journal on Applied Mathematics*, volume 43, number 3, pages 546–552, 1983.

[49] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, volume 521, number 7553, pages 436–444, 2015.

[50] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, `http://www.deeplearningbook.org`.

[51] M. Zeiler, M. Ranzato, R. Monga, M. Mao, K. Yang, Q. Le, P. Nguyen, A. Senior, V. Vanhoucke, J. Dean, and G. Hinton, "On rectified linear units for speech processing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pages 3517–3521.

[52] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *International Conference on Artificial Intelligence and Statistics*, 2011.

[53] K. O'Shea and R. Nash, *An introduction to convolutional neural networks*, 2015.

[54] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014.

[55] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, *Generative adversarial networks*, 2014.

[56] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, volume 15, number 56, pages 1929–1958, 2014.

[57] E.-P. Rührnschopf and K. Klingenbeck, "A general framework and review of scatter correction methods in x-ray cone-beam computerized tomography. part 1: Scatter compensation approaches," *Med. Phys.*, volume 38, number 7, pages 4296–4311, 2011.

[58] Y. Zhang, Y. Chen, A. Zhong, X. Jia, S. Wu, H. Qi, L. Zhou, and Y. Xu, "Scatter correction based on adaptive photon path-based Monte Carlo simulation method in multi-GPU platform," *Computer Methods and Programs in Biomedicine*, volume 194, page 105 487, 2020.

[59] M. Baer and M. Kachelrieß, "Hybrid scatter correction for CT imaging," *Phys. Med. Biol.*, volume 57, number 21, page 6849, 2012.

[60] R. Ning, X. Tang, and D. Conover, "X-ray scatter correction algorithm for cone beam CT imaging," *Med. Phys.*, volume 31, number 5, pages 1195–1202, 2004.

[61] H. Gao, R. Fahrig, N. R. Bennett, M. Sun, J. Star-Lack, and L. Zhu, "Scatter correction method for x-ray CT using primary modulation: Phantom studies," *Med. Phys.*, volume 37, number 2, pages 934–946, 2010.

[62] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," volume 9351, Springer Verlag, 2015, pages 234–241.

[63] A. Hanbury, H. Müller, and G. Langs, Eds., *Cloud-Based Benchmarking of Medical Image Analysis*, 1st edition. Cham: Springer Cham, 2017, pages XVIII, 254.

[64] B Ohnesorge, T Flohr, and K Klingenbeck-Regn, "Efficient object scatter correction algorithm for third and fourth generation CT scanners," *European Radiology*, volume 9, number 3, pages 563–569, 1999.

[65] I. Loshchilov and F. Hutter, "Decoupled weight decay regularization," 2017.

[66] A. Thummerer, P. Zaffino, A. Meijers, G. G. Marmitt, J. Seco, R. J. H. M. Steenbakkers, J. A. Langendijk, S. Both, M. F. Spadea, and A. C. Knopf, "Comparison of CBCT based synthetic CT methods suitable for proton dose calculations in adaptive proton therapy," *Phys. Med. Biol.*, volume 65, number 9, page 095 002, 2020.

[67] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Transactions on Image Processing*, volume 13, number 4, pages 600–612, 2004.

[68] D. Lustermans, G. P. Fonseca, V. T. Taasti, A. van de Schoot, S. Petit, W. van Elmpt, and F. Verhaegen, "Image quality evaluation of a new high-performance ring-gantry cone-beam computed tomography imager," *Phys. Med. Biol.*, volume 69, number 10, page 105 018, 2024.

[69] Y. Kyriakou, E. Meyer, D. Prell, and M. Kachelrieß, "Empirical beam hardening correction (EBHC) for CT," *Med. Phys.*, volume 37, pages 5179–5187, 10 2010.

[70] M. Kachelrieb and W. Kalender, "Improving PET/CT attenuation correction with iterative CT beam hardening correction," in *IEEE Nuclear Science Symposium Conference Record, 2005*, volume 4, 2005, 5 pp.–1909.

[71] D. Oh, S. Kim, D. Park, S. Choi, H. Song, Y. Choi, S. Moon, J. Baek, and D. Hwang, "Correction of severe beam-hardening artifacts via a high-order linearization function using a prior-image-based parameter selection method," *Med. Phys.*, volume 45, number 9, pages 4133–4144, 2018.

[72] G. Van Gompel, K. Van Slambrouck, M. Defrise, K. J. Batenburg, J. de Mey, J. Sijbers, and J. Nuyts, "Iterative correction of beam hardening artifacts in CT," *Med. Phys.*, volume 38, number S1, S36–S49, 2011.

[73] T. D'Angelo, G. Cicero, S. Mazziotti, G. Ascenti, M. H. Albrecht, S. S. Martin, A. E. Othman, T. J. Vogl, and J. L. Wichmann, "Dual energy computed tomography virtual monoenergetic imaging: Technique and clinical applications," *Br J Radiol*, volume 92, number 1098, page 20 180 546, 2019.

[74] L. Chen, X. Liang, C. Shen, S. Jiang, and J. Wang, "Synthetic CT generation from CBCT images via deep learning," *Med. Phys.*, volume 47, number 3, pages 1115–1125, 2020.

[75] W. Cao, A. Parvinian, D. Adamo, B. Welch, M. Callstrom, L. Ren, A. Missert, and C. P. Favazza, "Deep convolutional-neural-network-based metal artifact reduction for CT-guided interventional oncology procedures (MARIO)," *Med. Phys.*, volume 51, number 6,

[76] R. Zeng, C. Y. Lin, Q. Li, L. Jiang, M. Skopec, J. A. Fessler, and K. J. Myers, "Performance of a deep learning-based CT image denoising method: Generalizability over dose, reconstruction kernel, and slice thickness," *Med. Phys.*, volume 49, number 2, pages 836–853, 2022.

[77] D. M. Tucker, G. T. Barnes, and D. P. Chakraborty, "Semiempirical model for generating tungsten target x-ray spectra," *Med. Phys.*, volume 18, number 2, pages 211–218, 1991.

[78] P. M. Joseph, "An improved algorithm for reprojecting rays through pixel images," *IEEE Transactions on Medical Imaging*, volume 1, number 3, pages 192–196, 1982.

[79] D. Cullen, J. Hubbell, and L. Kissel, "EPDL97: The evaluated photon data library," University of California Lawrence Livermore National Laboratory, Tech. Rep., 1997.

[80] D. R. White, J. Booz, R. V. Griffith, J. J. Spokas, and I. J. Wilson, "4. The composition of body tissues," *Reports of the International Commission on Radiation Units and Measurements*, volume os-23, number 1, pages 20–23, 1989.

[81] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," 2015.

[82] F. Jäger, J. Maier, P. Paysan, M. Walczak, and M. Kachelrieß, "Learning ct scatter estimation without labeled data: A feasibility study," in *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023.

[83] F. Jäger, P. Paysan, M. Walczak, and M. Kachelrieß, "Rawdata-based higher order deep beam hardening correction (DBHC)," in *Proceedings of the 8th International Conference on Image Formation in X-Ray Computed Tomography*, 2024, pages 75–78.

[84] A. J. van de Schoot, D. Hoffmans, K. M. van Ingen, M. J. Simons, and J. Wiersma, "Characterization of Ethos therapy systems for adaptive radiation therapy: A multi-machine comparison," *Journal of Applied Clinical Medical Physics*, volume 24, number 5, e13905, 2023.

[85] J. L. Robar, A. Cherpak, R. L. MacDonald, A. Yashayaeva, D. McAloney, N. McMaster, K. Zhan, S. Cwajna, N. Patil, and H. Dahn, "Novel technology allowing cone beam computed tomography in 6 seconds: A patient study of comparative image quality," *Practical Radiation Oncology*, volume 14, number 3, pages 277–286, 2024.

[86] C. Nardi, R. Molteni, C. Lorini, G. G. Taliani, B. Matteuzzi, E. Mazzoni, and S. Colagrande, "Motion artefacts in cone beam CT: An in vitro study about the effects on the images," *The British Journal of Radiology*, volume 89, number 1058, page 20 150 687, 2016.

[87] J. Maier, L. Jordan, E. Eulig, F. Jäger, S. Sawall, M. Knaup, and M. Kachelrieß, "Learning ct scatter estimation without labeled data: A feasibility study," page 56, 2022.

List of Publications

Parts of this thesis have been published in the following conference contributions: Fully 3D 2023 [4] and CT Meeting 2024 [2]. Furthermore, additional publications in academic journals and conferences are included.

Journal Articles

[1] E. Eulig, F. Jäger, J. Maier, B. Ommer, and M. Kachelrieß, "Reconstructing and analyzing the invariances of low-dose ct image denoising networks," *Med. Phys.*, volume n/a, number n/a,

[2] M Marot, F Jäger, S Greilich, C. P. Karger, O Jäkel, and L. N. Burigo, "Monte carlo simulation for proton dosimetry in magnetic fields: Fano test and magnetic field correction factors kb for farmer-type ionization chambers," *Phys. Med. Biol.*, volume 68, number 17, page 175 037, 2023.

Conference Contributions

[1] J. Bollig, F. Jäger, and M. Kachelries, "Deep bone extraction on radiographs," in *Abstracts of the 5th European Congress of Medical Physics*, volume 125, 2024, page 103 562.

[2] F. Jäger, P. Paysan, M. Walczak, and M. Kachelrieß, "Rawdata-based higher order deep beam hardening correction (DBHC)," in *Proceedings of the 8th International Conference on Image Formation in X-Ray Computed Tomography*, 2024, pages 75–78.

[3] J. Bollig, F. Jäger, and M. Kachelrieß, "Deep bone extraction in x-ray projection domain," in *Proceedings of the 8th International Conference on Image Formation in X-Ray Computed Tomography*, 2024, pages 210–213.

[4] F. Jäger, J. Maier, P. Paysan, M. Walczak, and M. Kachelrieß, "Learning ct scatter estimation without labeled data: A feasibility study," in *Proceedings of the 17th International Meeting on Fully 3D Image Reconstruction in Radiology and Nuclear Medicine*, 2023.

[5] J. Maier, L. Jordan, E. Eulig, F. Jäger, S. Sawall, M. Knaup, and M. Kachelrieß, "Learning ct scatter estimation without labeled data: A feasibility study," in *Medical Imaging 2022: Physics of Medical Imaging*, 2022, page 56.

Acknowledgments

I would like to express my gratitude to everybody who has helped and supported my research:

Prof. Dr. Joao Seco for his role as first examiner, his guidance in the context of the thesis advisory committee and during my Bachelor and Master studies.
Prof. Dr. Marc Kachelrieß for providing me with the opportunity to pursue my doctoral studies in his department and for his guidance and supervision throughout my research. His insights, expertise, and suggestions have been invaluable.
Prof. Dr. Johanna Stachel and Prof. Dr. Tilman Plehn for agreeing to serve as examiners for the oral defense.
Dr. Pascal Paysan and Dr. Michał Walczak for their support and collaborative approach. Many thanks for providing the pelvis phantom measurement.
All current and former colleagues in E025, especially Elias Eulig, Dr. Joscha Maier, Dr. Stefan Sawall and Dr. Carlo Amato for their assistance and valuable input.
I would like to express my appreciation to my parents and brother for their encouragement and support since my high school days.
Proofreading my work is not always an easy task. Thank you to all who helped me with their comments and suggestions, as well as DeepL.
I'm thankful and grateful to Tim and Elias for all the great discussions, the amazing support, the coffee breaks, and for showing me the wonderful sport of bouldering.
I would like to thank the many people I have met along the way, at school, during my studies, and in other places. I'm so grateful for all the time we spent together and for our vacations.
Finally, I want to express my deepest gratitude to Achim, Hott, Cornelius, and especially Tim and Selina, Stephen and Pia, and Vanessa, for their unwavering faith in me and their limitless support. Without you, this work may not have been possible. I am thankful from the bottom of my heart that I was fortunate to meet you.

The research presented in this thesis was funded by Varian Medical Systems, a Siemens Healthineers Company.
Parts of the CT scans are from cadaver measurements and have been acquired in scope of a forensic study in close collaboration with the Institute of Forensic and Traffic Medicine (Prof. Dr. Sarah Heinze), Heidelberg University, Heidelberg, Germany, after being approved by the local ethics review board (S388/2014).

Parts of the reconstruction and simulation software were provided by RayConStruct® GmbH, Nürnberg, Germany.

I hereby assure, that I composed this work by myself and did not use any other than the listed resources.

Heidelberg, November 6, 2024