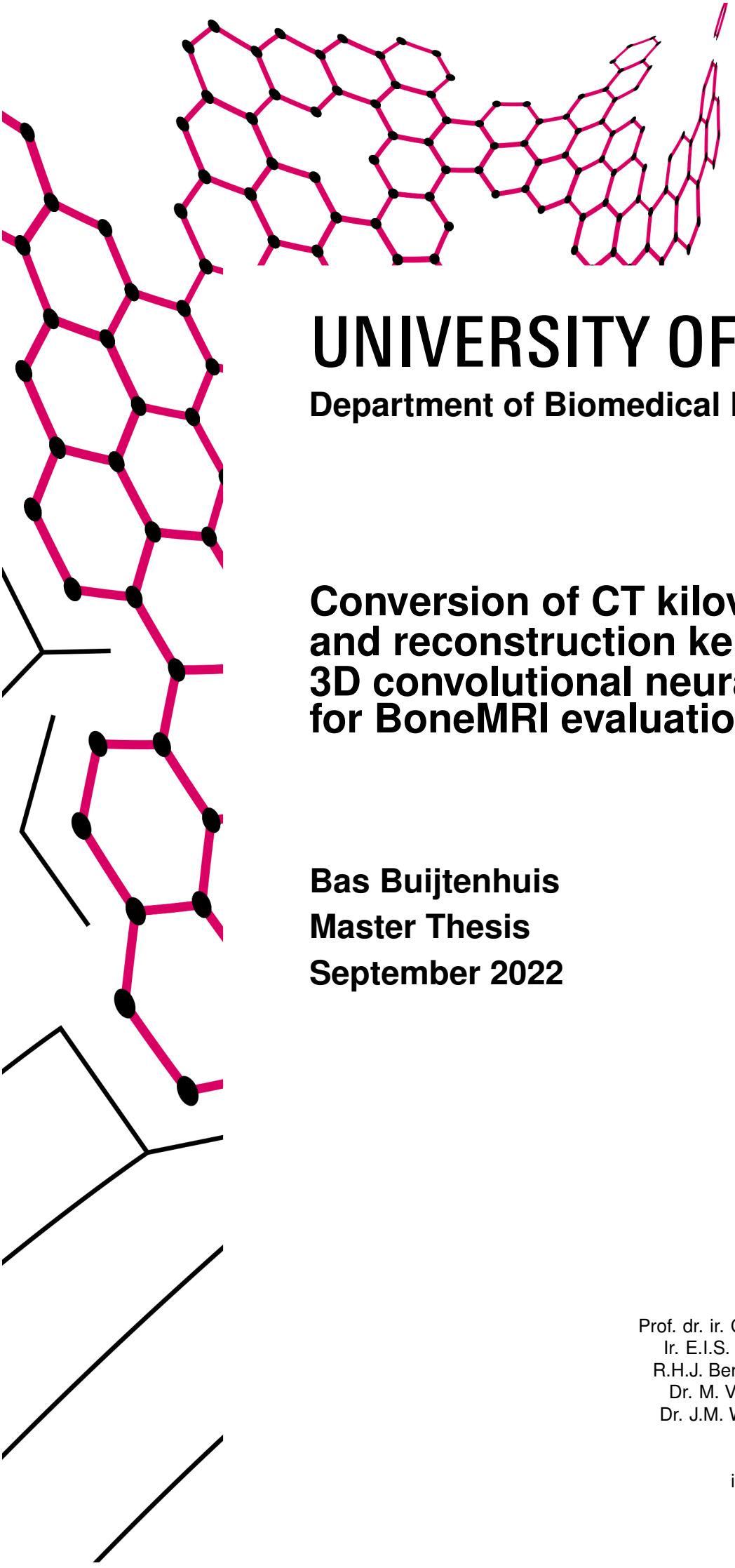

UNIVERSITY OF TWENTE.

Department of Biomedical Engineering

Conversion of CT kilovolt peak and reconstruction kernel using a 3D convolutional neural network for BoneMRI evaluation

Bas Buijtenhuis

Master Thesis

September 2022

GRADUATION COMMITTEE:
Prof. dr. ir. C.H. Slump (Cees) - Committee chair
Ir. E.I.S. Hofmeijer (Elfi) - Committee member
R.H.J. Bergmans (Rick) MSc - Daily supervisor
Dr. M. Van Stralen (Marijn) - Daily supervisor
Dr. J.M. Wolterink (Jelmer) - External member

Robotics and Mechatronics group
in cooperation with MRIguidance B.V.

I. Preface

I proudly present the thesis I have been working on for the past nine months. The completion of this research brings my student days to an end. During my years in Enschede, I gained an enormous amount of academic knowledge and at the same time was able to experience everything the city has to offer.

I would like to thank MRIguidance for the opportunity to conduct my graduation thesis at their company. The company's motivating and positive atmosphere ensured that I enjoyed working on my assignment every day. I really appreciate that I was able to get an insight into the company's activities. It was also great to be involved in all the social activities of the company. Sailing with the team, barbecuing, doing a cooking class, going for a lunch run or getting an ice cream from Roberto's. It is these kinds of activities that make MRIguidance unique and keep everyone in the company motivated and connected.

Of course, I would like to thank Rick Bergmans for all the support during my period at MRIguidance. The never-ending feedback, support and recommendations made sure I kept my final goal in mind. I would also like to thank Marijn van Stralen for being an excellent supervisor. Our meetings together with Rick were always very useful and guided me to make the right choices. I always left your office with fresh inspiration to continue my research.

I would also like to thank Cees Slump for the monthly meetings and feedback on my thesis. Your feedback made me dive into books about computed tomography that five years ago I never thought I would understand. In addition, I would like to thank Elfi Hofmeijer and Jelmer Wolterink for their feedback on my thesis and their willingness to be part of the graduation committee. Thanks to you all, I was able to do a graduation thesis that is completely in line with my interests.

II. Abstract

The combined use of CT and MRI allows for excellent diagnostic applications but requires an expensive and complex multimodal imaging workflow. MRIguidance has developed a deep learning approach to generate synthetic CT (BoneMRI) images using MRI data, which eliminates the need for a CT scan. Generation and evaluation of BoneMRI requires ground truth CT data. Evaluation metrics of BoneMRI should reflect its quality and should not depend on the quality or characteristics of the ground truth CT images. However, it was found that the kilovolt peak (kVp) and reconstruction kernel have an influence on the BoneMRI evaluation metrics. The goal of this research was therefore to standardize CT data characteristics in terms of kVp and reconstruction kernel, by developing image-to-image conversion methods.

Two methods were developed to convert CT images from a low (100) kVp or high (140) kVp to a target kVp of 120. The first method was a piecewise linear fit, while the second method used a convolutional neural network (CNN). A separate CNN was trained to convert images from a soft reconstruction kernel to a sharp reconstruction kernel.

The kVp conversion methods were used to convert the CT data being used for BoneMRI evaluation. The evaluation metrics of BoneMRI predictions were calculated with- and without using the developed conversion methods.

Both the piecewise linear method as well as the CNN were able to convert CT images from a low (100) kVp to the target (120) kVp. The remaining mean error in bone voxels was 4 ±1 HU using the piecewise linear method or 1 ±2 HU using the CNN. A mean error of 18 ±2 HU or 17 ±3 HU remained in voxels containing bone after the conversion from a high (140) kVp to the target (120) kVp using the piecewise linear or CNN method, respectively. Both kVp conversion methods improved the BoneMRI evaluation metrics. For the soft to sharp kernel conversion method a mean error of 5 ±1 HU remained in the bone. The mean absolute surface distance between bone segmentations of predicted and target images was 0.02 ±0.01 mm.

The study revealed that the conversion of CT images from one kVp to another kVp is possible using a piecewise linear fit or a CNN if paired data is available. The evaluation metrics of BoneMRI improved by applying the kVp conversion methods on the ground truth CT data. A CNN can also be used to convert CT images that were reconstructed with a particular reconstruction kernel to another reconstruction kernel. The conversion methods did not generalize to CT-data that was different than the CT-data that was present within the training data of the conversion methods.

III. Samenvatting

Het gecombineerde gebruik van CT en MRI maakt hoogwaardige diagnostische toepassingen mogelijk, maar vereist een dure en complexe multimodale workflow. MRIguidance heeft een deep learning methode ontwikkeld om synthetische CT-beelden (BoneMRI) te genereren met behulp van MRI-data, waardoor een CT-scan niet meer nodig is. Voor het genereren en evalueren van BoneMRI is 'ground truth' CT-data nodig. Evaluatie metingen van de BoneMRI moet de kwaliteit ervan weerspiegelen en mag niet afhankelijk zijn van de kwaliteit of kernmerken van de CT-data. Er is echter gebleken dat de gebruikte kilovoltpiek (kVp) en de reconstructiekernel van invloed zijn op de BoneMRI evaluatie metingen. Het doel van dit onderzoek was daarom het standaardiseren van CT-datakarakteristieken in termen van kVp en reconstructiekernel, door beeld-naar-beeld conversiemethoden te ontwikkelen.

Er werden twee methoden ontwikkeld om CT-beelden te converteren van een lage (100) kVp of hoge (140) kVp naar een beoogde kVp van 120. De eerste methode was een 'piecewise' lineaire aanpassing, terwijl de tweede methode gebruik maakte van een convolutioneel neuraal netwerk (CNN). Een afzonderlijk CNN werd getraind om beelden om te zetten van een zachte naar een scherpe reconstructiekernel.

De kVp-omzettingsmethoden werden gebruikt om de CT-beelden om te zetten die voor de BoneMRI-evaluatie werden gebruikt. De evaluatie van BoneMRI-beelden werd uitgevoerd met en zonder gebruik van de ontwikkelde conversiemethoden.

Zowel de piecewise lineaire methode als het CNN waren in staat CT-beelden van een lage (100) kVp om te zetten naar de beoogde (120) kVp. De resterende gemiddelde fout in bot voxels was 4 ±1 HU bij gebruik van de lineaire methode, of 1 ±2 HU bij gebruik van het CNN. Een gemiddelde fout van 18 ±2 HU of 17 ±3 HU bleef in bot over na de conversie van een hoge (140) kVp naar de (120) kVp met respectievelijk de lineaire of de CNN-methode. Beide kVp-omzettingsmethoden verbeterden de BoneMRI evaluatie metingen. Een gemiddelde fout van 5 ±1 HU bleef in de bot voxels over na de conversie van zachte reconstructies naar scherpe reconstructies met behulp van een CNN. De resterende gemiddelde absolute oppervlakte-afstand tussen de segmentaties van de voorspelling en het doel was 0.02 ±0.01 mm.

De studie heeft aangetoond dat de conversie van CT-beelden van een kVp naar een andere kVp mogelijk is met behulp van een piecewise lineaire fit of een CNN indien gepaarde data beschikbaar is. De evaluatie metingen van BoneMRI verbeterde door toepassing van de kVp conversiemethoden op de ground truth CT-data. Een CNN kan ook worden gebruikt om CT-beelden die zijn gereconstrueerd met een bepaalde reconstructiekernel om te zetten naar een andere reconstructiekernel. De conversiemethoden generaliseerden niet naar CT-data die anders was dan de CT-data die aanwezig was binnen de trainingsdata van de conversiemethoden.

IV. Abbreviations

Abbreviation	Meaning
3D-CNN	Three-dimensional convolutional neural network
AAPM	American Association of Physicists in Medicine
CNN	Convolutional Neural Network
CT	Computed Tomography
DECT	Dual-Energy Computed Tomography
GAN	Generative Adversarial Network
HU	Hounsfield Units
kVp	kilovolt peak
MAE	Mean absolute error
ME	Mean error
MRI	Magnetic Resonance Imaging
NIST	National Institute of Standards and Technology
ReLU	Rectified linear unit
sCT	Synthetic CT
SWF	System weighting function

Table of Contents

1 Introduction

1.1 Motivation

Medical images play an important role during clinical diagnosis. Common medical image modalities include Computed Tomography (CT), Magnetic Resonance Imaging (MRI), Ultrasound (US) and X-Ray.

Both MRI and CT images are often acquired for applications that affect both soft and bone tissues [21]. These include skull [30], spine [2, 36] and joint disorders [60]. MRI allows the acquisition of three-dimensional images with a high soft-tissue contrast [21]. The number of MRI scans made worldwide has increased considerably, reaching 95 million a year [32]. A major advantage of MRI is the lack of radiation exposure [20]. One major drawback of MRI is the inability to visualize bone structures.

CT is a widely used imaging modality which enables clear visualization of osseous structures [21]. The resulting three-dimensional images are valuable for the diagnosis and treatment of bone pathologies. The modality has a high spatial resolution and good reproducibility [20]. It provides information about electron densities which can be used for radiotherapy purposes [20]. CT imaging has the drawback of using radiation to acquire an image of the patient.

The combined use of CT and MRI is also beneficial for radiation therapy planning [20] and orthopaedics [63]. Such a multimodal imaging workflow allows for excellent diagnostics applications, but increases patient burden, healthcare costs and causes complicated workflows in the hospital. In addition, the radiation induced by CT imaging is undesirable, especially for young patients [28].

These difficulties have driven scientists to find ways to improve the bone contrast of MRI. This may lead to an MRI-only workflow, which would also lead to lower costs. On top of that, it eliminates the ionizing radiation induced to the patient.

Various MRI sequences exist to improve the bone contrast. These include spin-echo, gradient-echo and dedicated ultrashort and zero echo time imaging sequences [21].

An alternative method to enhance bone visualization is the use of image processing to convert MR images to CT Hounsfield Units (HU). This creates synthetic CT (sCT) images, with quantitative values. The most promising image processing methods to create sCT images make use of deep learning-based models [21].

MRIguidance (MRIguidance B.V., Utrecht, The Netherlands) [11] has developed a deep learning approach to convert three-dimensional MR images to sCT images using a 3D T1-weighted radiofrequency spoiled multiple gradient-echo (T1MGE) sequence [31]. This deep learning method (called BoneMRI) uses a patch-based convolutional neural network, similar to a U-net. The model was trained on paired MRI and CT data [31]. This method was validated for the pelvis and cervical- and lumbar spine regions [22, 57].

An example of such a BoneMRI image can be seen in Figure 1. The BoneMRI image (middle image) generated with the deep learning method closely resembles the ground truth CT (right image). The synthetic image is calculated based on MRI data, and thus eliminates the need for a CT scan, while providing an image with excellent bone contrast.

Figure 1: Left: T2w MRI. Middle: synthetic CT (BoneMRI). Right: ground truth CT. Image adopted from [58].

The quality of the generated BoneMRI images is assessed by comparing the generated BoneMRI to the ground truth CT image acquired using a CT scanner. Quantitative measures like the mean error (ME), mean absolute error (MAE), root-mean-square error (RMSE) and surface distance are calculated during this evaluation.

The input images of the BoneMRI model are all acquired using the same (standardized) MRI-acquisition sequence and are processed using the same model. The images are, however, compared to CT images acquired and reconstructed using a protocol that can vary depending on the clinical needs. Technologists adjust CT scanner acquisition and reconstruction parameters according to various factors, including manufacturer recommendations, institutional standards and patient characteristics [62]. These variations make medical images susceptible to inter- and intra-CT scanner model and settings variability.

While the evaluation of the BoneMRI indicates sufficient quality of the BoneMRI images, some interesting findings are found when the images are compared to ground truth CT images. As shown in Figure 2, the mean error between the BoneMRI predictions and the ground truth CT scans change when they are compared to CT scan ‘a’ or CT scan ‘b’. This difference in evaluation metrics is not desirable. The BoneMRI evaluation metrics should reflect the quality of the BoneMRI prediction and should not depend on the characteristics of the ground truth CT image. It is therefore important to investigate and minimize the influence of using different CT acquisition and reconstruction parameters on the evaluation of BoneMRI images.

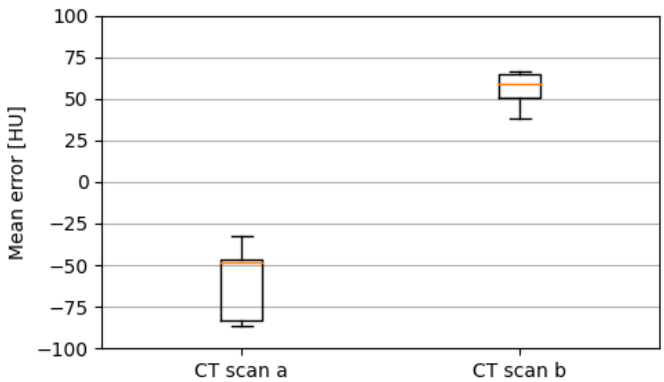

Figure 2: Boxplot of the mean error between BoneMRI predictions and two different CT scans.

1.2 Objective

A conversion method needs to be developed to convert CT images acquired and reconstructed with different settings. The conversion will minimize the influence of the different settings on the evaluation metrics of BoneMRI predictions. The BoneMRI workflow with the conversion step included can be seen in Figure 3. The focus hereby is to minimize differences in voxels containing bone tissue, as this is the tissue of interest within the BoneMRI images. Ultimately, the conversion will improve image consistency regarding style and appearance, which in turn improves the evaluation metrics of the BoneMRI images.

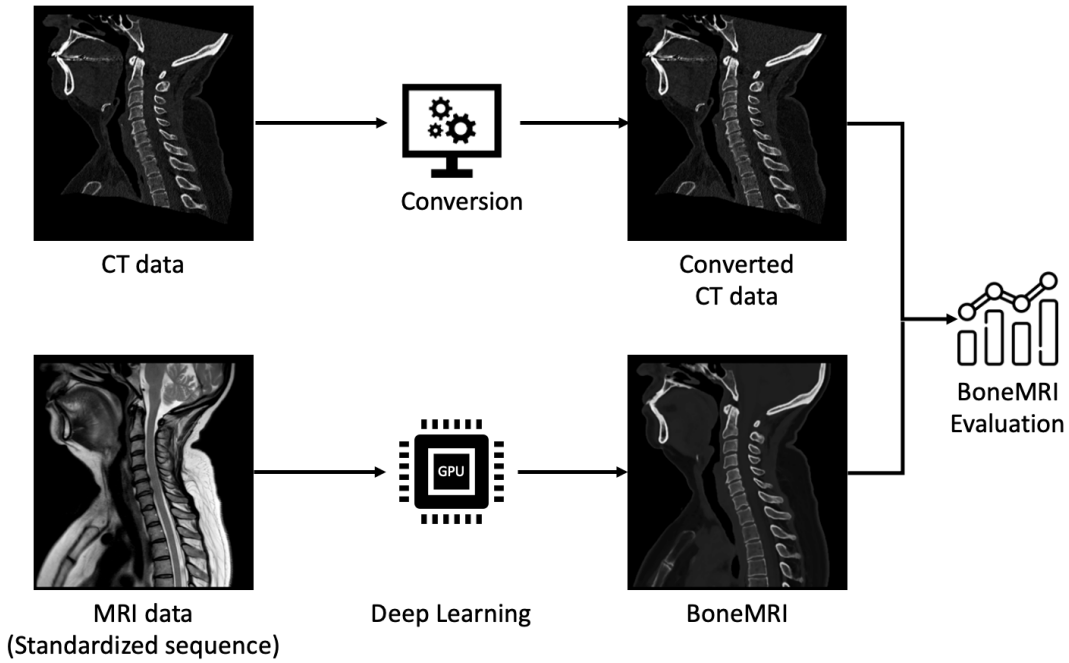

Figure 3: Workflow of the generation and evaluation of synthetic CT (BoneMRI) images. Internal data from MRIguidance.

Image acquisition parameters including tube voltage, pitch, voxel size, vendor and tube current have an influence on the characteristics of a reconstructed CT image [59]. Among these parameters, the kilovoltage peak (kVp) has a major influence on the quantitative values assigned to tissue voxels [10, 19, 48, 53]. During the image reconstruction process, the reconstruction technique and the associated reconstruction filtering kernel also have a significant impact on the values assigned to tissue voxels [37, 39, 42, 51, 59].

The research questions that need to be addressed are therefore:

- How can CT images acquired with a particular kVp be converted to another kVp?
- How can CT images reconstructed with a particular reconstruction kernel be converted to another reconstruction kernel?
- How does the use of the conversion method(s) improve the evaluation of the BoneMRI images?

Ideally, the CT-scans acquired using any kVp are converted to 120 kVp, as is recommended by the American Association of Physicists in Medicine (AAPM) for CT scans of the head, chest, abdomen and pelvis [6]. A sharp ‘bone’ reconstruction kernel will be defined as the target. A sharp kernel preserves the higher spatial frequencies, which results in a high spatial resolution, at the expense of an increase in noise [9]. The high spatial resolution is desirable to examine bone structures [56].

1.3 Layout of document

This research report is divided into eight chapters. The current chapter describes the research motivation and objectives. An explanation of the related background information is given in chapter 2. This includes general information about the acquisition and reconstruction of CT images and includes an explanation of the major sources of variation in reconstructed CT images. The chapter includes an overview of the available research to standardize or convert CT images. A gentle introduction to deep learning is given, as it was used within this research to convert CT images. The third, fourth, fifth and sixth chapters contain experiments related to the research questions, as stated in the chapter about the objective. Each chapter contains a discussion about the findings of the experiment. Chapter 7 contains a general discussion about the study findings and its limitations. The chapter also contains recommendations for future work. The final chapter covers the conclusions that are a result of this study. An additional chapter containing appendices is included at the end of the document, which contains supporting information.

2 Background

This chapter includes an introduction to CT and deep learning. The basic principles of CT image acquisition and reconstruction are explained, after which the major sources of variation within CT images are explained. A brief introduction to deep learning is included as it will be used as a method to convert CT images.

2.1 Computed Tomography

CT is a diagnostic imaging method developed by Sir Godfrey Hounsfield in the late 1960s [8]. The most important parts of the scanner are an X-ray tube, a gantry with a ring of X-ray sensitive detectors and a computer. Inside the X-ray tube, X-rays are generated by electrons that collide with an anode. The generated X-rays pass through the object to be examined and get attenuated by the materials present inside the object. The remaining X-ray photons are then measured by the detectors. The measured data by the detectors is sent to the computer that processes the data for further usage.

The measurement of attenuation at a row of detectors is called a projection (see equation 1). By acquiring projection measurements at many angles, a so called 'sinogram' can be obtained (see Figure 5b). Mathematically, every projection (g_{fan}) of a fan-beam geometry can be represented as a function using:

$$g_{fan}(\alpha,\gamma) = \iint_{-\infty}^{\infty} \delta(\mathrm{xcos}(\gamma+\alpha) + \mathrm{ysin}(\gamma+\alpha) - \mathrm{Rsin}\,\gamma) f(x,y)dxdy, \qquad (1)$$

where x and y are the position in the image domain. α is the projection angle and γ is the angular deviation of the ray with respect to the central direction of the beam (see Figure 4). The distance of the source to the origin is denoted by R [52].

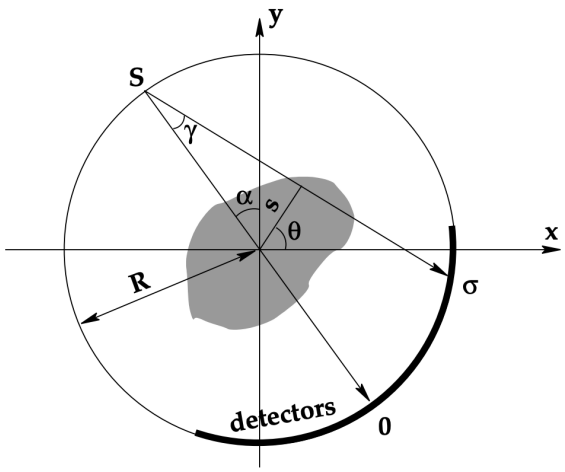

Figure 4: Schematic view of the fan-beam geometry. Figure adopted from [52].

Once the projection data $g_{fan}(\alpha,\gamma)$ is acquired, the inverse problem needs to be solved to find the original image $f(x,y)$. One method to find the original image is to apply the inverse Radon transform. Applying this transformation results in a blurry image (see Figure 5c). Therefore, a frequency filter is used to obtain a filtered back projection (FBP). An example of an image reconstructed with a ramp filter can be seen in Figure 6. The ramp filter corrects the image by reducing blurring. On the downside, the filter enhances high frequencies which can cause unwanted noise.

Two other methods to reconstruct an image from projection data are iterative reconstruction and deep learning-based reconstruction. With iterative reconstruction, an image is assumed, after which the assumption is compared to the measured values. Corrections are then made to the assumption. This

process is repeated until the assumed and measured values are within acceptable limits [53]. Deep learning-based reconstruction is a relatively new method for image reconstruction. This method incorporates neural networks into the image reconstruction process to improve spatial resolution while using data acquired at a low dose [47].

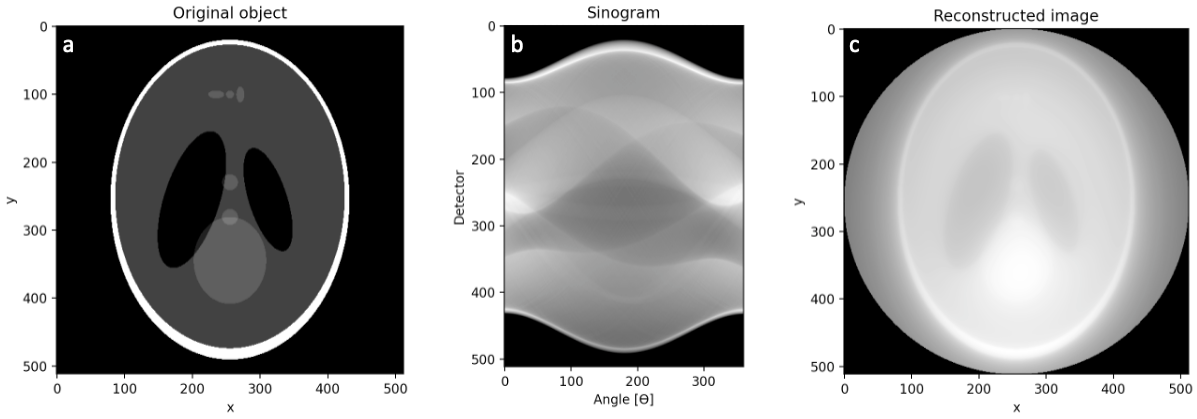

Figure 5: (a)Shepp-Logan phantom. (b) Sinogram after applying the radon transform. (c) Reconstructed image after applying back projection (no filter), 360 projections.

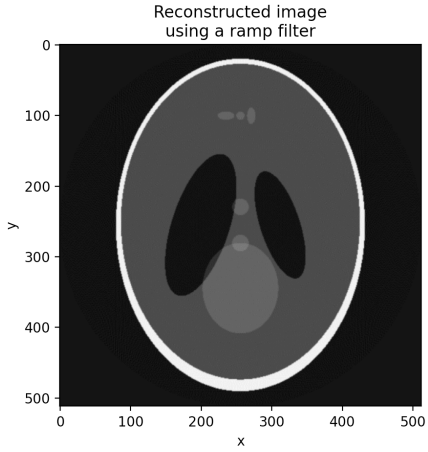

Figure 6: Reconstruction of the image using filtered back projection with a ramp filter (360 projections).

During a CT acquisition, many parameters have an influence on the resulting projections. These parameters include the number of projections, the number of detectors, the size of the detectors, the tube voltage, the tube current, slice thickness and the acquisition mode. Additionally, the reconstruction technique and reconstruction kernel influence the appearance of the reconstructed image. The appearance of the final image is therefore depending on many parameters. All these parameters can be chosen or influenced. A CT protocol needs to be chosen which results in an image with sufficient quality, while taking the dose induced to the patient into account.

2.2 CT variability

Variation in the reconstructed image depends on many parameters (see section 2.1). The variation in the appearance of an image can be divided into two groups: CT number accuracy and CT number noise. The parameters that influence the CT number accuracy include the tube voltage, the reconstruction kernel and image artifacts. Since CT image artifacts are not the focus of this research, no explanation of the influence of image artifacts on the CT numbers is included. The influence of the tube voltage and the reconstruction kernel on the CT number accuracy are explained in the sections below.

2.2.1 Tube voltage

In CT, pixels in a reconstructed image are assigned a CT number, which is also referred to as Hounsfield Unit (HU). This number is calculated using:

$$CT\ number = \frac{\mu_x(E) - \mu_w(E)}{\mu_w(E)} \times 1000, \tag{2}$$

where $\mu_x(E)$ and $\mu_w(E)$ are the energy dependent linear attenuation coefficients of the tissue of interest and water, respectively. Using this scale, water is defined to be zero HU and air is equal to -1000 HU [14].

The effective attenuation coefficient $\bar{\mu}(\vec{r})$ can be calculated using:

$$\bar{\mu}(\vec{r}) = \int_E w(E)\mu(E,\vec{r})dE, \tag{3}$$

where $w(E)$ is the system weighting function (SWF), $\mu(E,\vec{r})$ is the ground truth attenuation of the material at position $\vec{r}$ for energy E [26]. The weighted attenuation thus depends on the SWF which can be calculated using:

$$w(E) = \frac{S(E)D(E)}{\int_E S(E')D(E')dE'}, \tag{4}$$

where $S(E)$ is the emitted spectrum by the x-ray source, and $D(E)$ is the energy-dependent detector responsivity [26]. An example of emitted x-ray photon spectra for different tube voltages can be seen in Figure 7. An example of an energy dependent Gadolinium Oxysulphide (GdOS) detector responsivity is shown in Figure 8. Equations 3 and 4 indicate that a change in the material composition, emitted spectrum or detector responsivity will change the measured attenuation coefficient.

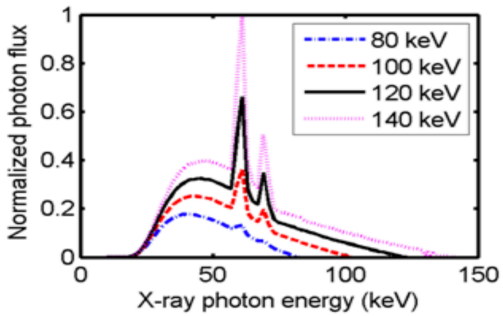

Figure 7: Example of simulated x-ray photon energy spectra S(E) with a 2.5 mm aluminum filter. Figure adopted from [68].

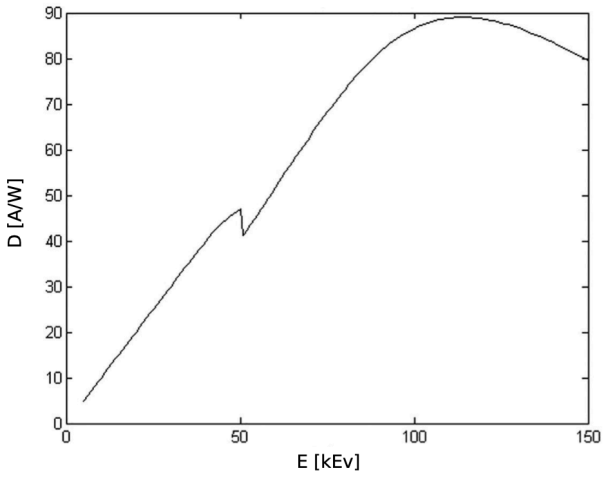

Figure 8: Detector responsivity D(E) for a 1.4 mm Gadolinium Oxysulphide scintillator. Figure adopted from [24].

The true (linear) attenuation coefficient of a material $\mu(E)$ can be calculated using

$$\mu(E) = \rho\left(\frac{\mu(E)}{\rho}\right), \tag{5}$$

where ρ is the material density in g/cm^3, and $\frac{\mu}{\rho}$ is the mass attenuation coefficient in cm^2/g [26]. The mass attenuation coefficient depends on the chemical elements that are present in the material. These energy dependent mass attenuation coefficients are available from literature sources like the National Institute of Standards and Technology (NIST) [29]. An example of the attenuation coefficients for air, water and cortical bone can be seen in Figure 9. The energy of the photons is expressed in electron volts (eV). One eV is defined as the energy acquired by an electron as it traverses an electrical potential difference of one volt in a vacuum [15].

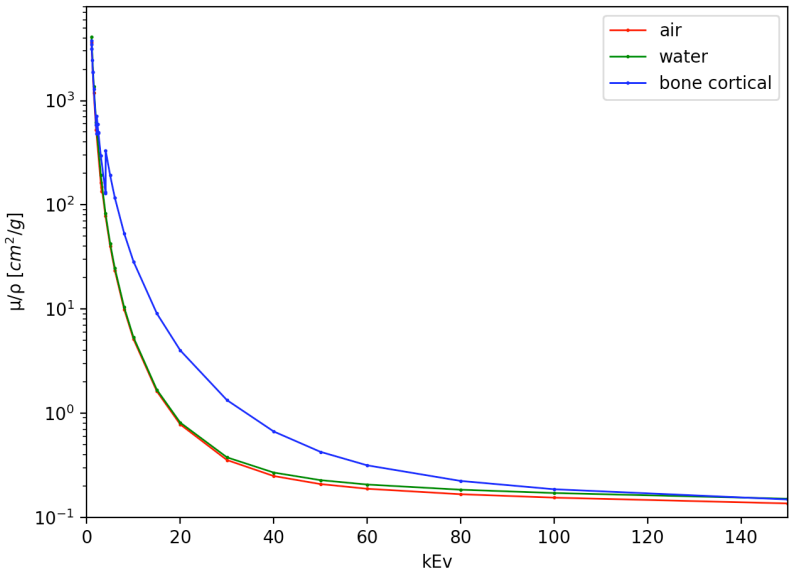

Figure 9: Energy dependent mass attenuation coefficients for air, water and cortical bone. The values are determined by the National Institute of Standards and Technology [29].

The HU value that is assigned to a specific material is thus dependent on the attenuation of the material of interest as well as the attenuation of water for the range of energies being emitted by the x-ray source.

Three interaction techniques are dominant in the energy range used for CT imaging. These are Rayleigh scattering, Compton scattering and the photoelectric effect.

Rayleigh scattering occurs in the low-energy region (15 to 30 keV) [15]. This interaction technique causes an incident photon to alter its path (coherent scattering). Detection of these scattered photons will have a negative effect on the image quality [15]. Therefore, the photons in the low energy range are removed from the emission spectrum using filtering. Filtering refers to the removal of x-rays by passing the beam of photons through a material. This material (typically aluminum or tin) attenuates photons in the low-energetic region.

The second interaction technique is called Compton scattering. When Compton scattering occurs, a photon will transfer part of its energy to an electron. The remaining energy will travel as a photon in a scattered direction [12]. This interaction technique is dominant at diagnostic photon energies in materials of low atomic number such as tissue and air [15]. The probability of Compton scattering is approximately proportional to the density of the material.

The third interaction technique is photoelectric absorption, which is the dominant interaction technique for materials with high atomic number such as bone and contrast materials [15]. The probability of photoelectric absorption is approximately proportional to Z^3/E^3, where Z is the atomic number and E is the energy of the incident photon.

Thus, different interaction techniques are dominant for soft- and bone like tissue. This causes the attenuation coefficient of bone to decrease more drastically compared to soft tissue when the energy of the incident photons is increased. This effect can be seen in Figure 9.

This unequal decrease in attenuation in combination with the simple equation (see equation 2) makes the assigned HU value of dense materials dependent on the x-ray spectrum.

On most CT scanners, the tube voltage can be set to various settings. The kVp indicates the maximum amount of energy an emitted photon can have, which is equal to the tube voltage being used. As can be seen in Figure 7, a change in the tube voltage will change the spectrum of the emitted spectrum. As stated earlier, this is expected to change the HU values for some materials. As can be seen in Figure 10, this is the case for materials with HU values above zero. The HU value can differ up to approximately 500 HU for cortical bone when the tube voltage is changed from 80 to 140 kVp. Materials below zero HU have a negligible change in HU value when the tube voltage is changed. The finding of the tube voltage dependent HU values is supported by other research [1, 3, 13, 19, 65]. An example image of scans acquired at two different kVp settings (100- and 140 kVp) can be seen in Figure 11. The HU values assigned to voxels can differ up to 640 for voxels containing dense material.

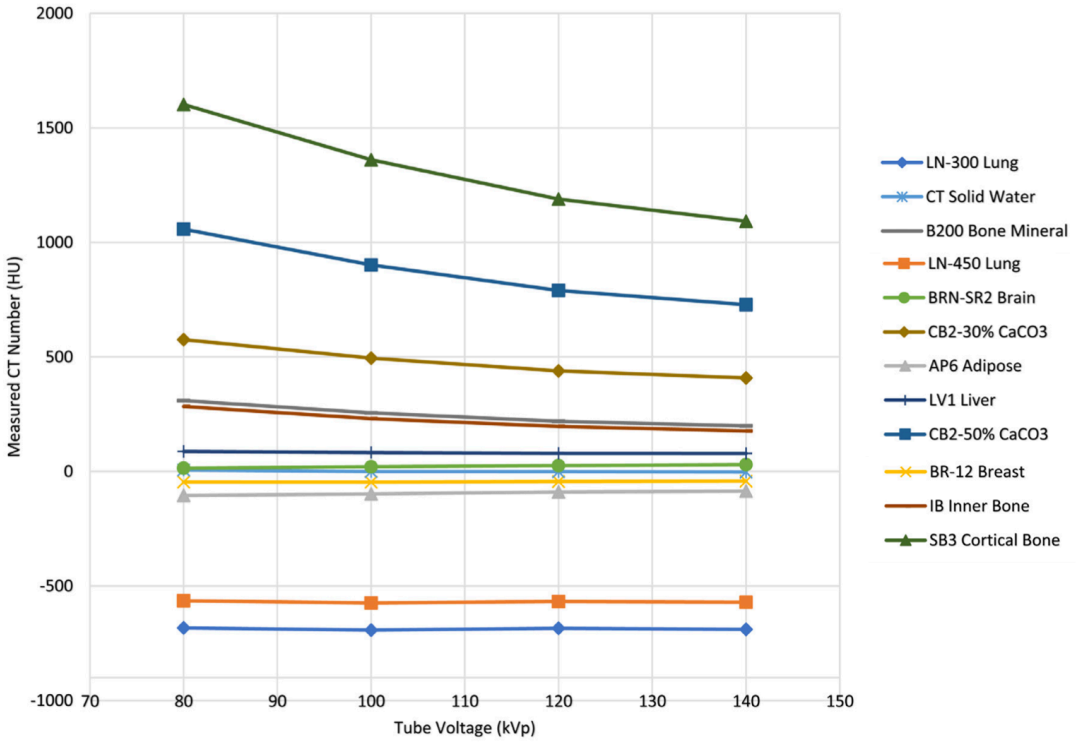

Figure 10: kVp dependence of HU values for different materials using a phantom scan. Figure adopted from [5]

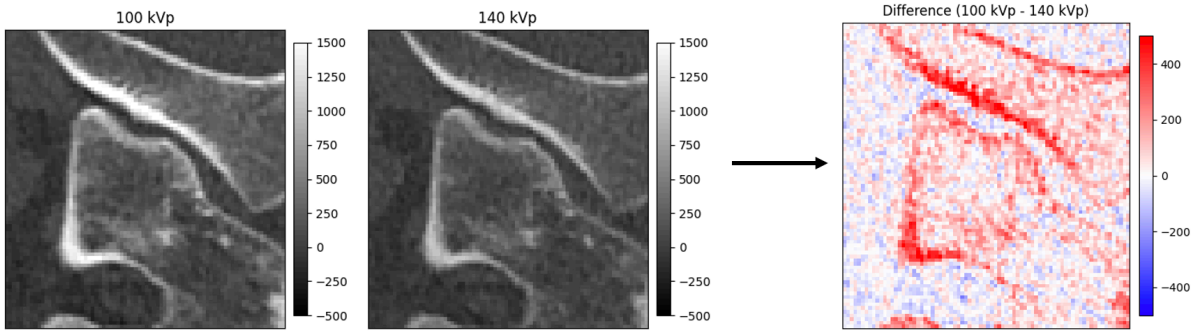

Figure 11: Images acquired with two different kVp settings. Left and middle image W/L: 2000/500 HU. Right: Difference image between the reconstructed images (W/L: 1000/0 HU). Internal data from MRIguidance.

The optimal tube voltage is a tradeoff between radiation dose and image quality [67]. The dose is proportional to the square of the kVp [53]. Decreasing the kVp from 140 to 120 reduces the dose by 28% to 40% for a standard phantom. With regards to patient safety, the dose delivered to the patients should be kept as low as reasonably achievable (ALARA). To minimize the dose, a low kVp might therefore be preferable. A lower kVp is also beneficial when iodine-based contrast materials are being used. The iodine has an increased attenuation at lower energies, which improves the conspicuity of vascular pathologies [67]. However, images obtained using lower tube voltages tend to have more noise, caused by the high absorption for low-energy photons by the patient [67]. The choice of the tube current is therefore a tradeoff between image quality and patient dose.

Larger patients attenuate more photons compared to small patients, which reduces the image quality for larger patients [44]. Therefore, an increased tube voltage might be required when scanning larger patients. While standard tube currents are defined by the AAPM [6], the CT technician can make the decision to use a different tube voltage depending on the individual patient and the diagnostic task.

2.2.2 Reconstruction kernel

As described in the section about Computed Tomography, different reconstruction kernels can be applied when reconstructing an image. A trade-off needs to be made between spatial resolution and noise when designing a reconstruction kernel.

A reconstruction kernel is often designed for a specific clinical application and should not be used indiscriminately for all applications [27]. Most vendors offer specific reconstruction kernels to reconstruct fine bone structures. These kernels are designed to enhance edges, and are not intended to maintain the CT number accuracy of small objects [27].

The used reconstruction kernel influences the HU values assigned to voxels [37, 51]. An example of two images reconstructed with different reconstruction kernels can be seen in Figure 12. The Q30f reconstruction kernel is a 'medium smooth' iterative dual-energy reconstruction kernel, while the I50f is a sharp iterative reconstruction kernel [25]. Both reconstruction kernels are developed by Siemens Healthineers.

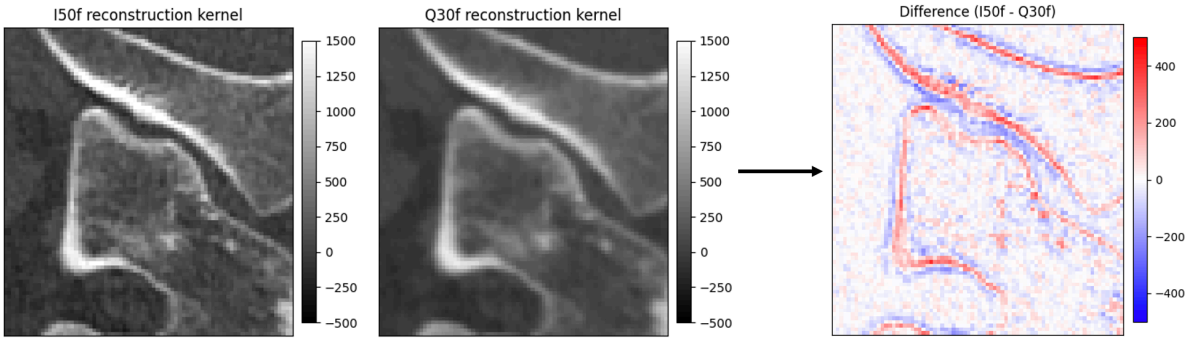

Figure 12: Images reconstructed with two different reconstruction methods. Left and middle image W/L: 2000/500 HU. Right: Difference image between the reconstructed images (W/L 1000/0 HU). Internal data from MRIguidance.

As can be seen in Figure 12, the HU value assigned to pixels differs between the two reconstructions. This is most prevalent on the edge of bone structures, where the HU values of neighboring pixels have a large difference. The difference in HU values assigned to a voxel can differ up to 400.

2.3 Standardizing CT

Various sources alter the CT image appearance, as described in the section about CT variability. The difference in quantitative HU values is not an issue from a clinical point of view. It is however problematic when the images are used for image processing purposes. For this reason, several studies related to CT image conversion have already been carried out. Related work on the topic of CT image standardization and conversion is summarized in the section below.

2.3.1 Related work

Several researchers have investigated deep learning methods to reduce CT variability induced by acquisition- or reconstruction parameters. Three articles were found that convert CT images between different tube voltages [41, 43, 69]. With regards to the reconstruction process, six studies used a deep learning approach to convert between reconstruction kernels [16, 37, 38, 54, 55, 66]. Two out of nine articles used a convolutional neural network (CNN) architecture [16, 43], while another two used a U-Net architecture [41, 69]. The remaining five articles used a Generative Adversarial Network (GAN) [37, 38, 54, 55, 66].

2.3.2 Remaining work

Six studies specifically excluded regions containing bone or did not compute quality metrics on regions containing bone [16, 37, 38, 41, 54, 55]. One of the remaining three articles admitted to be susceptible to image artifacts [66], while another article shows over-smoothened generated images [69]. The remaining article makes use of projection data during the conversion process [43].

The aim of this research is to develop a conversion method that is specifically aimed at the HU value of bone regions, since a change in kVp or reconstruction kernel creates a large difference in the HU values assigned to voxels containing bone (see section CT variability). This significant difference in voxels containing bone is of importance for the evaluation of the BoneMRI images. The conversion method should not use projection data as this is not always available. The conversion method should not be susceptible to image artifacts, as image artifacts will affect the evaluation metrics of BoneMRI images.

2.4 Deep Learning

Ever since it was possible to scan and load medical images into a computer, researchers have developed systems to automate (parts of) the medical image analysis pipeline. Medical image processing typically aims to extract features that might be difficult to assess with the naked eye [34]. The first image processing tools relied on low-level pixel processing, such as thresholding and enhancing contrast [40]. Recent advances in the development of machine learning and especially deep learning allowed the application of more advanced image processing tools to be developed for medical images. The turning point dates back to 2012, where neural networks started to outperform traditional computer vision methods [7]. Since then, deep learning research related to medical imaging has increased exponentially.

The high performance of deep learning methods was not the only reason it gained a lot of interest. Deep learning also makes problem-solving easier, by automating what used to be the most difficult step in machine learning: feature engineering [17]. Feature engineering refers to the selection and extraction of features to remove irrelevant and redundant data, in order to increase the learning accuracy [33].

Research shows that the model architecture alone is often not sufficient to obtain a well performing deep learning method [40, 49, 50]. Instead, research indicates that image preprocessing and augmentation techniques are of equal importance to optimize the performance of the deep learning method.

Deep learning has been successfully applied to problems such as classification, detection, localization, segmentation and registration problems in the medical image domain [4, 40].

To address these problems, a variety of deep learning architectures are being used. Despite extensive research in the field of deep learning, there does not (yet) exist one deep learning architecture that can solve every problem or application area [40].

2.4.1 Key elements of deep learning

The core elements of deep learning networks are ‘layers’, which are data-processing modules that extract representation out of data fed into them [17]. The stacking of multiple of these layers implements a progressive data distillation.

These layers contain weights that can be learned by training the network. When using supervised learning, the training data consists of the input data and the corresponding target data. The model tries to learn a function during training that maps the input to the target data. The difference between the prediction and the target data can be quantified using a loss function [7]. This loss function can then be used to adjust the weights of the network. Two common loss functions for regression problems are the mean squared error (MSE) and mean absolute error (MAE) [17].

When the entire dataset is too large to pass through the network at once, batch-based learning can be used. With batch-based learning, a network updates its weights based on multiple samples. After a certain number of batches, the loss is calculated on validation data. This value gives an indication if the network is over- or underfitting and can be used to tune hyperparameters.

Hyperparameters determine the network structure (like the number of layers and filters) and determine how the network is trained (for example the learning rate and optimizer).

Convolutional neural networks (CNN's) are a commonly used deep learning architecture for regression problems. The network is composed of many stacked convolutional layers. A convolutional layer applies a convolutional filter to the input image. The convolution process is suited for image related problems as it considers locally connected information [34]. Pooling layers can be used between the convolutional layers to increase the field of view of the network. The pooling layers also reduce the computational requirements of the network.

3 kVp conversion experiment

The kVp used during acquisition has a large influence on the HU values assigned to voxels in the reconstructed image, as was described in the section about CT variability. Therefore, two methods were developed to convert images from a low kVp (100) or high kVp (140) to the desired 120 kVp.

The conversion methods can then be used to convert data that is used during the evaluation of BoneMRI predictions.

3.1 Data:

The data available to develop these methods contain Dual-Energy CT (DECT) images from 26 patients (acquired at 100- and 140 kVp). The Somatom Definition Flash CT scanner from Siemens uses dual source technology to acquire DECT images. The CT scan covered the pelvic region, including part of the lumbar spine vertebrae. The data was acquired using a Siemens Somatom Definition Flash CT scanner at the Ghent University Hospital. An additional scan was available, which was the average of the CT images acquired at 100- and 140 kVp. This average image is from now on considered to be the target 120 kVp scan. The images were reconstructed using an iterative sharp (I50f) reconstruction kernel developed by Siemens Healthineers. The slice thickness of the images was 1 mm. The pixels spacing varied between 0.55 mm and 0.90 mm (depending on the patient). The pixel spacing was not rescaled. The available data was divided into training, validation, and testing datasets, according to the commonly used ratio of 60:20:20 (16:5:5 patients). An example of the data that was used can be seen in Figure 13.

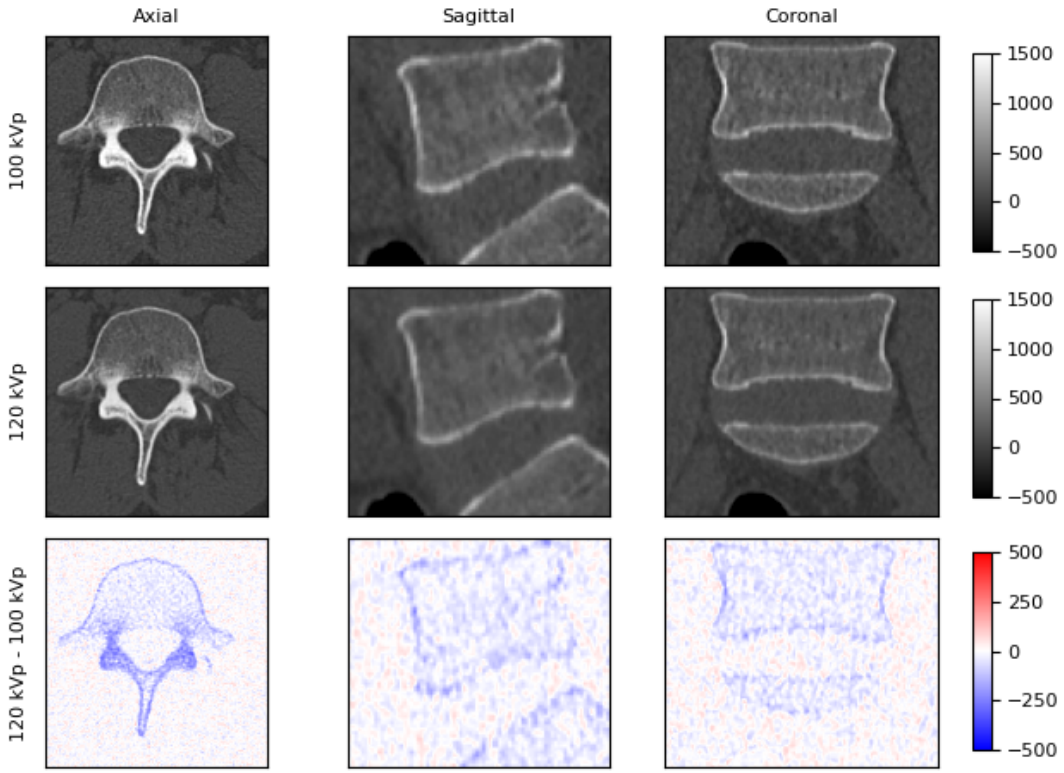

Figure 13: Example of the available data (100 kVp and 120 kVp) being used. Internal data from MRIguidance. All values are in HU.

3.2 Method

Two methods were implemented to convert the 100- or 140 kVp images to 120 kVp images. The first method was a piecewise linear fit, while the second method made use of a deep learning network.

3.2.1 Piecewise linear fit

A curve fitting method can be used to convert images from 100 kVp (or 140 kVp) to 120 kVp. Figure 14 shows a voxelwise plot of the HU values assigned to voxels from a 100- and 120 kVp scan. The data was then separated into values below- and above zero HU (based on the 100- or 140 kVp scan). The separation was done because he data below zero HU does not change between 100- and 120 kVp (as described in section 2.2.1), while the HU values above zero HU depend on the kVp.

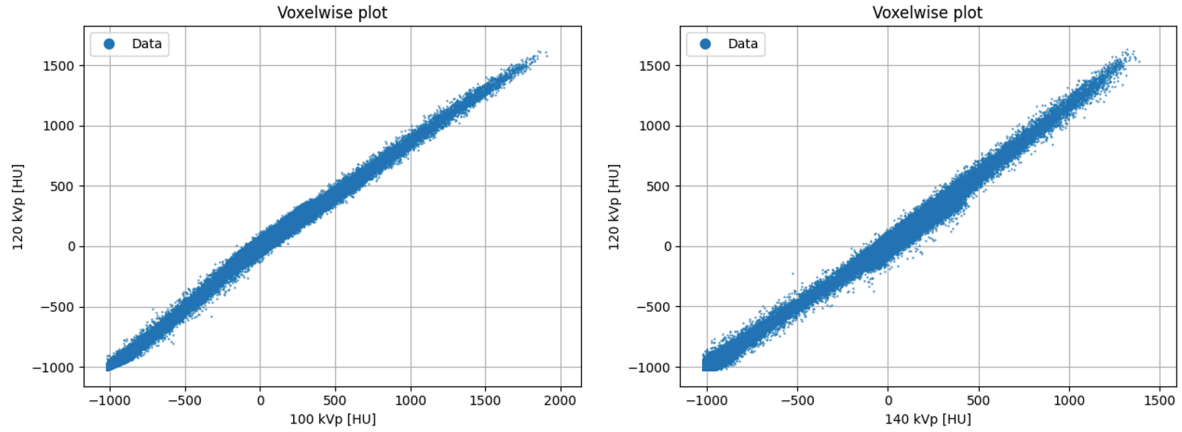

Figure 14: Voxelwise plot of the HU values assigned to voxels at 100- and 120 kVp (left) or 140- and 120 kVp (right).

A linear fit was applied to both groups of datapoints separately. The optimal linear fit was calculated using Python (version 3.8) with the Scipy package (version 1.8.0). The best fit was determined using least-squares regression. The calculated best fits can be seen in Figure 15. A CT image acquired at 100 kVp can be converted to 120 kVp using this piecewise linear curve (see equation 6). A similar conversion curve was calculated to convert images from 140 kVp to 120 kVp.

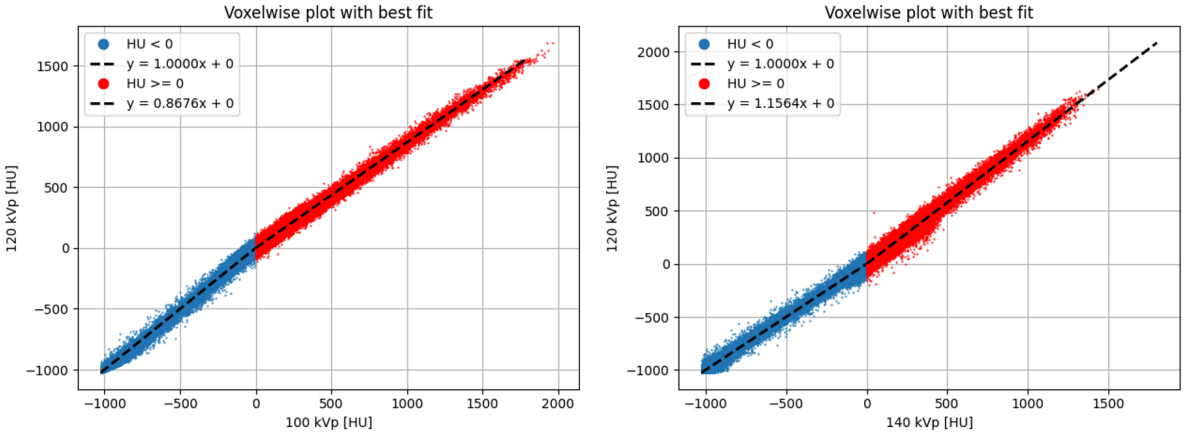

Figure 15: Voxelwise plot of the HU values assigned to voxels at 100 kVp and 120 kVp (left), and 140 kVp and 120 kVp (right). The datapoints are separated for values below- and above zero HU (based on the 100- or 140 kVp scan). An optimal linear fit was calculated.

$$CT_{120kVp}(x,y,z) = CT_{100kVp}(x,y,z) \times 0.8676 \quad if \quad CT_{100kVp} \geq 0$$
$$CT_{120kVp}(x,y,z) = CT_{100kVp}(x,y,z) \quad if \quad CT_{100kVp} < 0 \qquad (6)$$

3.2.2 Convolutional neural network

A patch-based three-dimensional convolutional neural network (3D-CNN) was developed to convert images from 100- or 140 kVp to 120 kVp. The network architecture can be seen in Figure 16. Two separate networks were trained to convert CT images from 100 kVp to 120 kVp, or from 140 kVp to 120 kVp. The network consisted of an input layer, followed by three convolutional 'blocks', which consisted out of a convolutional layer with 64 filters, followed by a rectified linear unit (ReLu) activation function and a batch normalization layer. A final convolutional layer was added with a linear activation function to get to the desired output shape of $24 \times 24 \times 24$ voxels. This is equal to the size of the input patch. All convolutional filters have a shape of $3 \times 3 \times 3$ with a stride of 1. The output of the final convolutional layer was added to the input using a skip connection. The network had a total of 225.217 trainable weights. The final model architecture can be found in Appendix A. The network was manually optimized. The optimization process can be found in Appendix B.

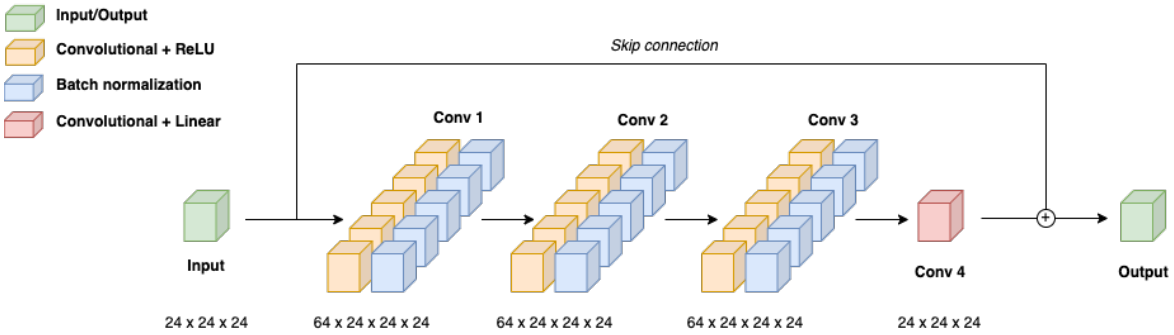

Figure 16: Schematic figure of the three-dimensional convolutional neural network used for kVp conversion.

Patches were sampled from the available data. The patches had a 50% chance to be from a region containing bone, a 40% chance to be from a region containing soft tissue and a 10% chance to be from a region containing air voxels. This oversampling of the bone was used because bone voxels were less abundant than soft and air voxels, while it is the most important tissue to predict accurately. Patches were only selected from the region inside a circular mask (see Figure 17). This mask was needed because the image acquired at 140 kVp did not contain data outside of this circular region. The acquisition method of the dual energy CT image explains why there is data missing within the 140 kVp image. The Somatom Definition Flash CT scanner from Siemens uses dual source technology to acquire DECT images. The field of view of the 140 kVp scan is smaller than the scan acquired at 100 kVp, which results in missing data for the 140 kVp scan [35].

The HU values present in the input patches were rescaled from [-1024, 3072] to [0, 1] (16-bit depth). A Nadam optimizer was used with a mean absolute error loss function and a learning rate of 1e-4. The batch size was set to 64. The network was trained for 30 epochs with 200 steps per epoch. The network architecture was implemented in Python (version 3.8) using Keras (version 2.8.0). A single NVIDIA GeForce GTX 3080 was used to train the model, which took approximately 50 minutes.

During inference, patches were fed to the model with an overlap of 12 voxels in each direction to reduce stitching artifacts.

3.3 Evaluation

The quality of the generated images was quantified using the ME and MAE. These metrics were voxel-wise calculations between the target and predicted image. The standard deviation between patients was also calculated. The metrics were calculated separately for voxels containing air, soft tissue or bone. The separation of materials was done using thresholding (see equation 7). Values of -500 and 300 were used for T_1 and T_2, respectively. An example of the masks created using thresholding can be seen in Figure 17. Next to the three material masks, an extra 'circular mask' was created. The evaluation metrics were not calculated outside of this area.

$$\begin{aligned} Air\ mask = True \quad & if\ f(x,y,x) \leq T_1 \\ Soft\ tissue\ mask = True \quad & if\ f(x,y,z) \leq T_2\ and\ f(x,y,z) > T_1 \\ Bone\ mask = True \quad & if\ f(x,y,z) > T_2 \end{aligned} \tag{7}$$

Figure 17: Example of a CT scan (W/L: 2000/500 HU) and its masks created using thresholding.

3.4 Results

The results for the conversion from 100 kVp to 120 kVp, and from 140 kVp to 120 kVp are given below.

3.4.1 Conversion from 100 kVp to 120 kVp

The ME and MAE values of the CT images compared to the target CT image before and after conversion can be seen in Figure 20. The biggest improvement can be seen in voxels containing bone material. The mean error in bone was reduced from -85 (±5) HU to 4 (±1) HU using piecewise linear conversion, and to 1 (±2) HU using the CNN. For soft tissue, the mean error increased from 1 (±2) HU to 7 (±1) HU using the piecewise linear conversion method, while the ME remained 1 (±1) HU using the CNN. The ME for air voxels remained unchanged using both conversion methods (0 ±0 HU). The large decrease in ME and MAE for bone voxels can be seen in the difference images shown in Figure 19. The predictions do not show any signs of image artifacts, as can be seen in Figure 18. The MAE in bone reduced from 85 (±5) HU to 27 (±3) HU using the piecewise linear conversion, and to 23 (±3) HU using the CNN.

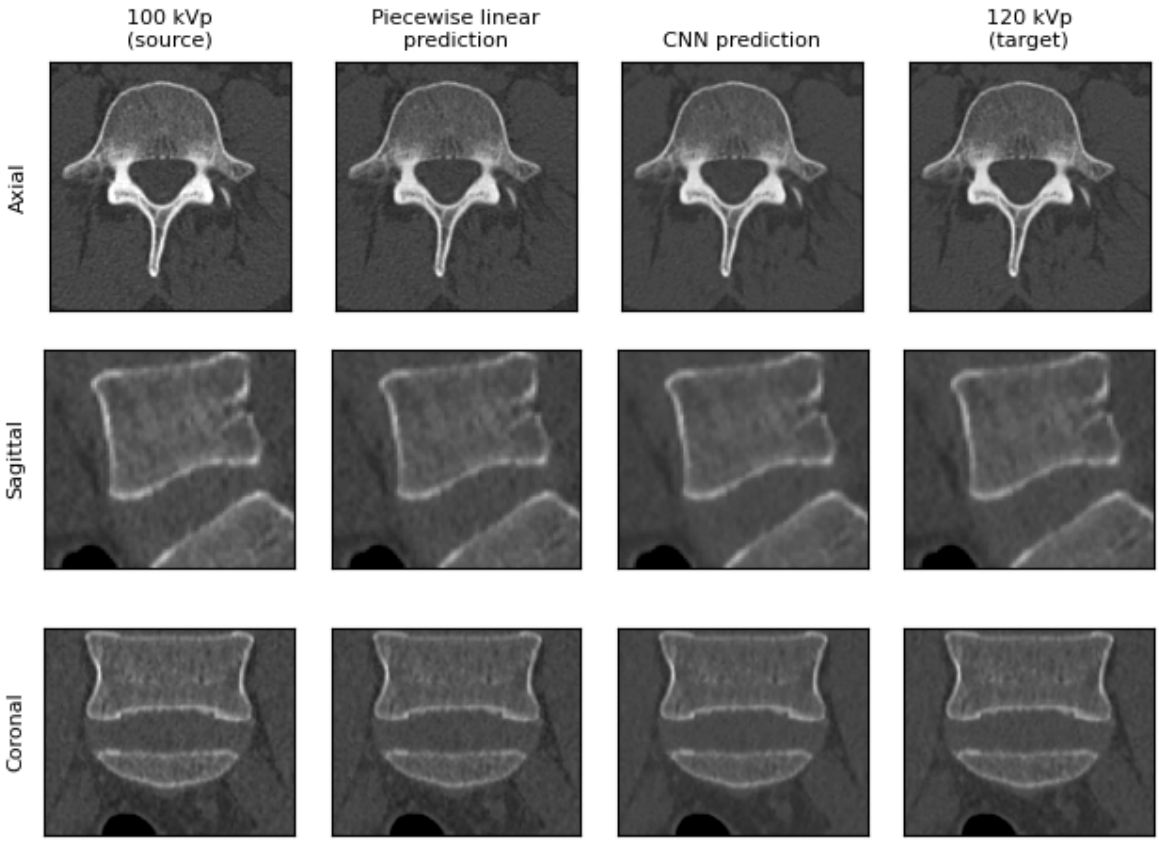

Figure 18: Illustration of the source, target and prediction data. W/L: 2000/500 HU.

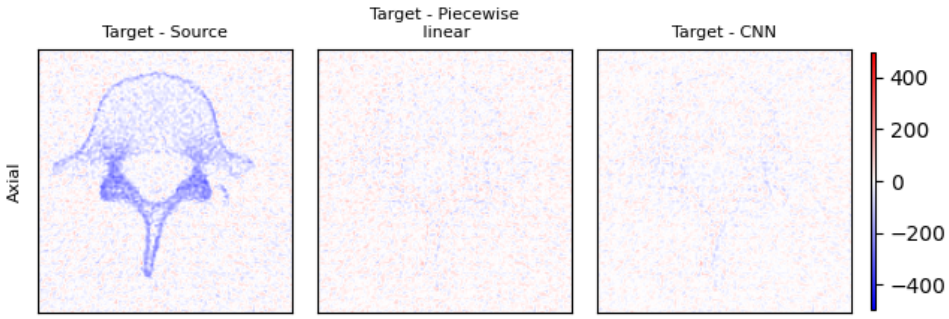

Figure 19: Difference images between the target and source CT images (left), target and piecewise linear prediction (middle) and target and CNN prediction (right). W/L: 1000/0 HU.

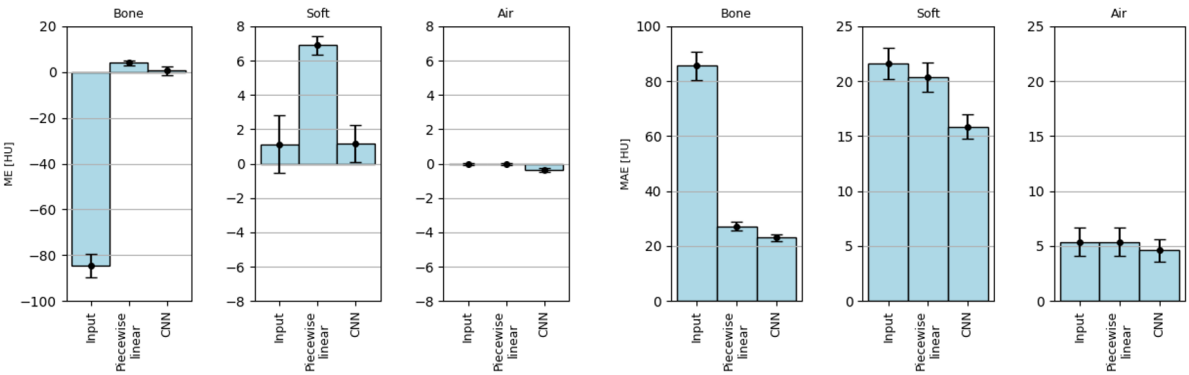

Figure 20: Mean error (left) and mean absolute error (right) values between target (120 kVp) and predictions for different segmentations. The input error is the error between the target (120 kVp) and input (100 kVp) image.

3.4.2 Conversion from 140 kVp to 120 kVp

An illustration of the source, target and predictions can be seen in Figure 21. The decrease in ME can be observed in the difference images shown in Figure 22. The ME and MAE values decreased the most for voxels containing bone, as can be seen in Figure 23. The ME value of bone voxels reduced from 85 (±5) HU to 18 (±2) HU using the piecewise linear conversion and to 17 (±3) HU using the CNN. The ME of soft tissue voxels increased using the piecewise linear conversion (from -1 ±2 HU to -6 ±1 HU), while it remained -1 ±1 HU using the CNN. The ME in voxels containing air remained comparable before or after conversion. The MAE in bone reduced from 85 (±5) HU to 39 (±3) HU using the piecewise linear conversion, and to 32 (±2) HU using the CNN.

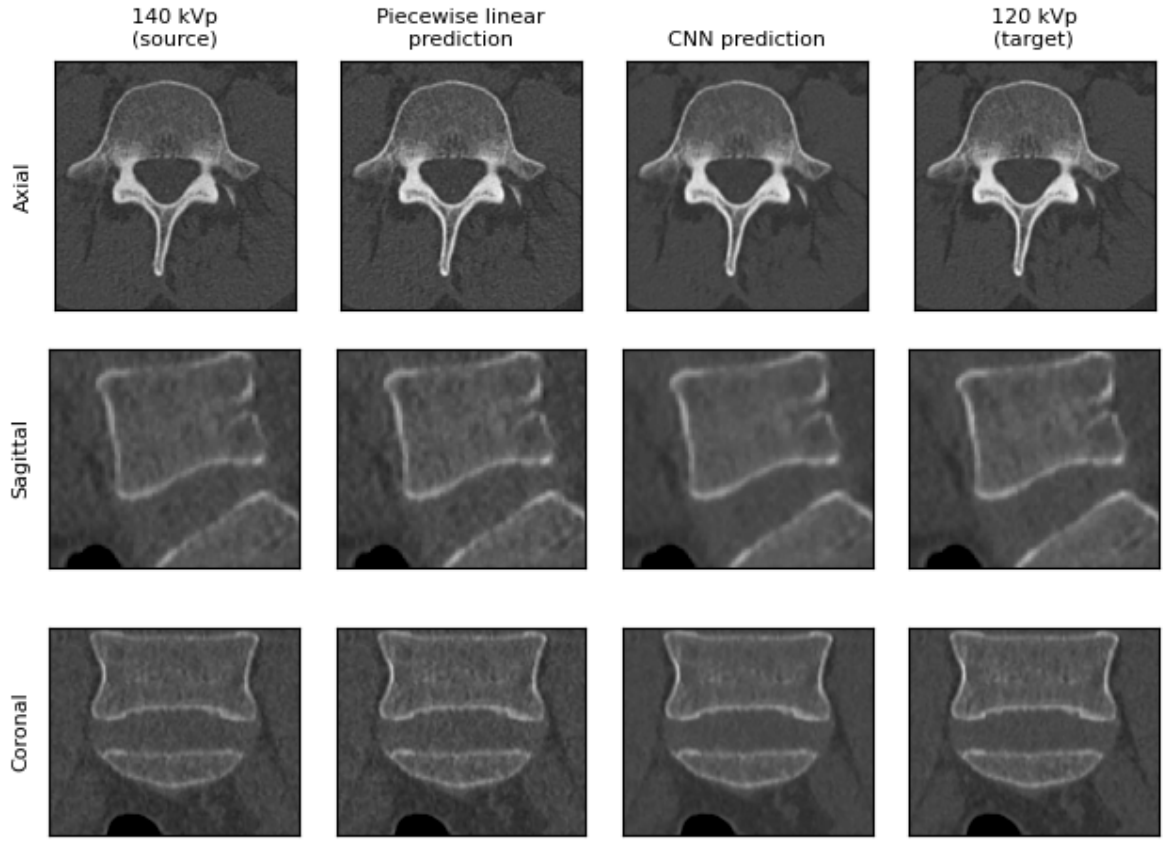

Figure 21: Example of the source, target and prediction data. W/L: 2000/500 HU.

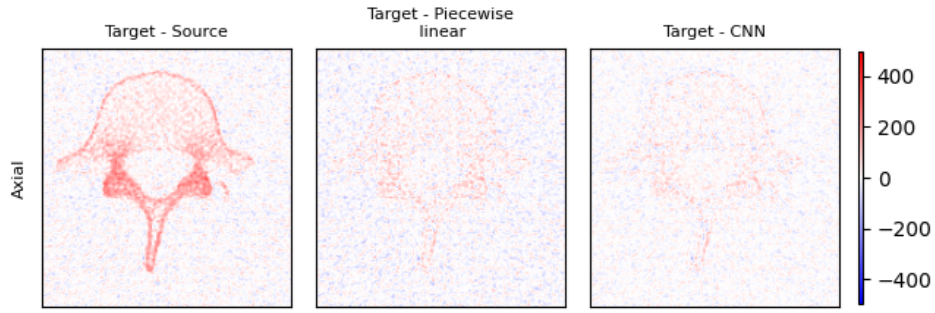

Figure 22: Difference images between the source and target CT image (left), target and piecewise linear prediction (middle) and target and CNN prediction (right). W/L: 1000/0 HU.

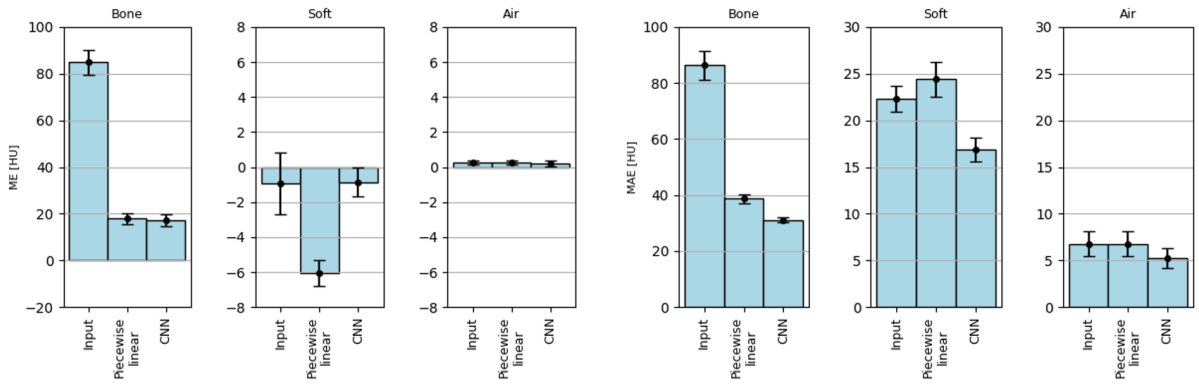

Figure 23: Mean error (left) and mean absolute error (right) values between target (120 kVp) and predictions for different segmentations. The input error is the error between the target (120 kVp) and input (140 kVp) image.

3.5 Discussion

Both the piecewise linear and the CNN conversion methods reduced the ME for bone voxels. The remaining ME for bone voxels is larger for source data of 140 kVp, than for source data of 100 kVp. A ME of 18 or 17 HU remained after conversion of 140 kVp data using the piecewise linear or CNN conversion methods, respectively. Only a ME of 4 HU remained after conversion of 100 kVp data using these methods.

The error that remains after converting the 140 kVp images might be caused by Compton scattering. The amount of Compton scattering is weakly dependent on the photon energy. The increased amount of Compton scattering might cause some artifacts within the 140 kVp CT scan, making it difficult to convert the image correctly to the target image.

The piecewise linear and CNN conversion methods show equal improvements for bone- and air voxels. The piecewise linear conversion method shows an increased ME for voxels containing soft tissue compared to the CNN. This is because soft tissue with a HU value above zero is rescaled when the piecewise linear conversion is applied, while this rescaling is not needed.

4 Applying the kVp conversion method during BoneMRI evaluation

The conversion method can be used during the evaluation of a BoneMRI model that uses data from the Ghent University hospital. The evaluation metrics are expected to degrade when the BoneMRI predictions are evaluated using CT images that are not acquired with the same kVp setting as was present within the BoneMRI training data (as explained in section 2.2.1). The CT images acquired with a different kVp can be converted to the desired kVp setting using the conversion methods described in section 3. The conversion of the CT images is expected to improve the evaluation metrics of the BoneMRI predictions.

The conversion methods were developed using data from the Ghent University hospital. The conversion methods are also applied to data from a different dataset (Jilin Hospital). The application of the conversion methods to this data will show whether the conversion methods are generalizable to data from a different dataset.

4.1 Data

Two datasets from different hospitals were used to test the kVp conversion methods during the BoneMRI evaluation phase.

The first dataset that was used consisted of CT images acquired at the Ghent University Hospital using a Siemens Somatom Definition Flash CT scanner (I50f sharp reconstruction kernel, slice thickness 1 mm). The CT scans covered the pelvis and part of the lumbar spine. For every patient, data acquired at 100 and 140 kVp was available. An additional (synthetic) 120 kVp CT scan was available which was the average of the 100- and 140 kVp scans. A total of 49 patients were available. The synthetic 120 kVp CT scans of 34 patients were used to train the BoneMRI model. The data of the remaining 15 patients were used for evaluation.

The second dataset that was available consisted of CT images acquired at a different center (Jilin Hospital). The dataset consisted of nine CT scans acquired at 120 kVp and 10 CT scans acquired at 140 kVp. The CT scans covered the pelvic region. The images were acquired using five different CT scanner models, from four different vendors. The CT scanner model name for every patient can be found in Appendix C, as well as the kVp, reconstruction kernel and slice thickness that was used. The slice thickness of the data ranged from 0.625 to 1.5 mm. The slice thickness of the data was not resampled. This dataset was used to validate the conversion methods that were trained using data from the Ghent University Hospital.

4.2 Method

BoneMRI images, generated using a research version of the BoneMRI product, were evaluated using CT data from 15 patients (from the Ghent University Hospital). The BoneMRI model was trained on 120 kVp CT images. Different combinations of kVp's were used during multiple (independent) evaluations. The combination of scans used during evaluation can be seen in Table 1. First, the predictions were compared to their corresponding 120 kVp scan (experiment 1.1). Then, the evaluation was repeated using the same BoneMRI predictions, while comparing 4 of the predictions to the 100 kVp CT scans and 5 images to the 140 kVp CT scans (experiment 1.2). The 100- and 140 kVp CT scans were then converted to a synthetic 120 kVp scan (sCT) using the convolutional neural network (experiment 1.3) or the piecewise linear conversion method (experiment 1.4).

Table 1: kVp's for every group of CT images used during the BoneMRI evaluation. The data was acquired using a Siemens Somatom Definition Flash CT scanner (I50f sharp reconstruction kernel, slice thickness 1 mm). Internal data from MRIguidance.

Experiment number	CT data used for BoneMRI evaluation	Amount of CT scans
1.1	120 kVp	15
1.2	100 kVp	4
	120 kVp	6
	140 kVp	5
1.3	Synthetic 120 kVp (converted from 100 kVp using the CNN)	4
	120 kVp	6
	Synthetic 120 kVp (converted from 140 kVp using the CNN)	5
1.4	Synthetic 120 kVp (converted from 100 kVp using the piecewise linear model)	4
	120 kVp	6
	Synthetic 120 kVp (converted from 140 kVp using the piecewise linear model)	5

The data from Jilin was used to evaluate the generalizability of the developed kVp conversion method. Four combinations of evaluation data were used to evaluate the performance of the conversion methods (see Table 2). The BoneMRI model (research version) was trained on six 120 kVp scans. The remaining three 120 kVp CT scans were used for evaluation (experiment 2.1). Ten CT scans acquired at 140 kVp were then added to the evaluation metrics (experiment 2.2). These 140 kVp CT scans were then converted to 120 kVp using the conversion methods and used for evaluation (experiments 2.3 and 2.4).

Table 2: kVp's for every group of CT images used during the BoneMRI evaluation. The data was acquired at Jilin using five different scanner models (internal data from MRIguidance).

Experiment number	CT data used for BoneMRI evaluation	Amount of CT scans
2.1	120 kVp	3
2.2	100 kVp	3
	140 kVp	10
2.3	120 kVp	3
	Synthetic 120 kVp (converted from 140 kVp using the CNN)	10
2.4	120 kVp	3
	Synthetic 120 kVp (converted from 140 kVp using the piecewise linear model)	10

4.3 Evaluation

The quality of the BoneMRI images were calculated using the ME and MAE between the BoneMRI predictions and the CT images. The ME and MAE were only calculated on segmentations of the bone.

4.4 Results

The results for the Ghent and Jilin experiments can be found in the sections below.

4.4.1 Ghent University Hospital

Figure 24 shows boxplots of the ME and MAE of the four experiments. The ME of BoneMRI images compared to 120 kVp CT images was 0 (±20) HU (experiment 1.1) When BoneMRI predictions were evaluated on mixed data (experiment 1.2), the ME increased to 4 (±47) HU. When a conversion method was applied on the evaluation data, the standard deviation decreased to 19 HU for both the convolutional neural network (experiment 1.3) as for the piecewise linear conversion method (experiment 1.4). The MAE for experiment 1.1 was 71 (±9) HU. The MAE increased to 82 (±15) HU

for experiment 1.2. Using the conversion methods, the MAE decreased to 71 (±9) HU for experiment 1.3 and to 73 (±9) HU for experiment 1.4.

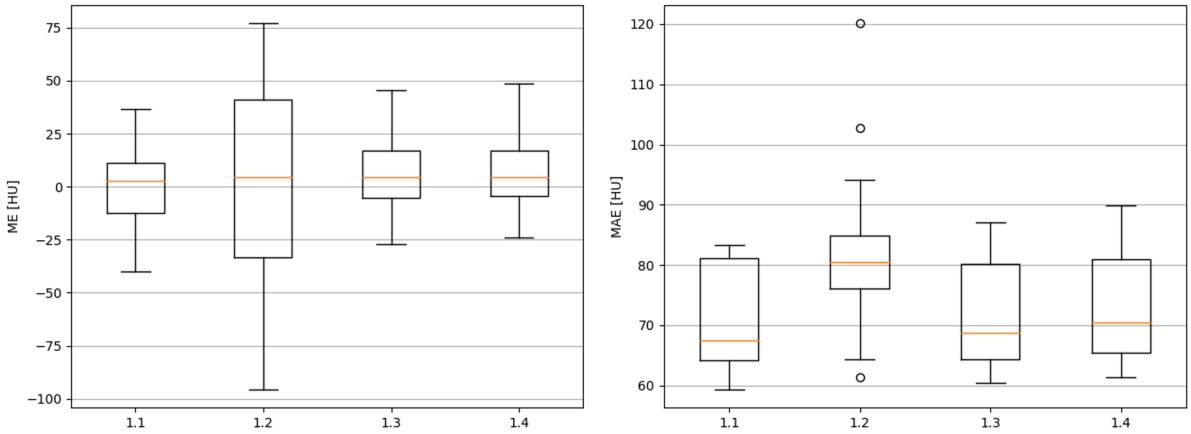

Figure 24: Boxplots of the mean error and mean absolute error of the four experiments using data from the Ghent University Hospital. The metrics are calculated on segmentations of the bone.

4.4.2 Jilin Hospital data

Boxplots of the ME and MAE of the four experiments can be seen in Figure 25. A mean ME of -17 HU (±16) HU was present when the BoneMRI predictions were evaluated on 120 kVp scans (experiment 2.1). The mean ME decreased to -14 (±25) HU when BoneMRI predictions were evaluated on CT scans with mixed kVp's (experiment 2.2). The ME increased to -61 (±41) HU when a convolutional neural network was applied on the 140 kVp CT images (experiment 2.3). The mean ME increased to -59 (±38) HU when the piecewise linear conversion method was applied on the 140 kVp CT images (experiment 2.4). The MAE for experiment 2.1 was 77 (±4) HU. The MAE increased to 91 (±13) HU for experiment 2.2. The MAE further increased to 112 (±29) HU or 112 (±27) HU when a convolutional neural network or a piecewise linear conversion was applied (experiments 2.3 and 2.4).

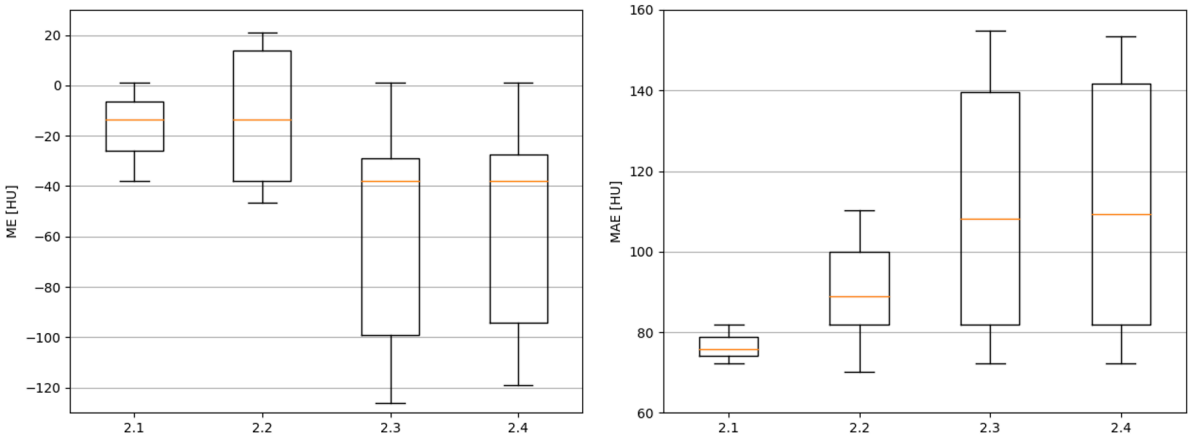

Figure 25: Boxplots of the mean error and mean absolute error of the four experiments using data from the Jilin Hospital. The metrics are calculated on segmentations of the bone.

4.5 Discussion

The results from the experiments using Ghent or Jilin data are discussed separately.

4.5.1 Ghent University Hospital

The standard deviation of the ME increased when BoneMRI images were evaluated on CT scans acquired at different kVp's than the model was originally trained on. This spread in ME values was decreased when the conversion methods were applied. The MAE increased when the BoneMRI images were evaluated on CT scans acquired with a different setting than was used within the training data (experiment 1.2). Both conversion methods were able to reduce the MAE. This shows that both conversion methods improve the evaluation consistency of BoneMRI images that are compared to CT data from the Gent University Hospital.

4.5.2 Jilin Hospital data

The conversion methods do, however, not improve the BoneMRI evaluation consistency of data from the Jilin Hospital. A ME of -17 (±16) HU is already present when the BoneMRI predictions are compared to 120 kVp images (the same as the model was trained on). This error might be due to the different CT scanner models that are present in the training and evaluation data. The model was trained on data from six patients acquired with four different CT scanner models (see Appendix C) acquired at 120 kVp. The model was then evaluated on CT data from three patients acquired with two different CT scanner (of which one scanner model was not present in the training data). This discrepancy between the CT scanner models could potentially cause a ME between the BoneMRI predictions and ground truth CT images (experiment 2.1), while the kVp was set to 120 for all scans. The quantitative HU value depends on many factors (see section 2.2.1). It is possible that the different CT scanner models use a different detector material, filter material, filter thickness or X-ray anode material. A change in one of these causes a change in the HU value assigned to bone. On top of that, a total of seven different reconstruction kernels were used to reconstruct the images. The data also included many scans that were acquired with a different slice thickness compared to the training data. This further increases the differences of the images that were present within the dataset.

The ME decreased when the evaluation of BoneMRI images on CT images acquired at 140 kVp were added (experiment 2.2). This indicates that the HU value assigned to bone material using a UIH uCT 530 CT scanner at 140 kVp is comparable to the HU value when scanning at 120 kVp using the CT models that were present within the training data.

When the conversion methods were applied, the ME and MAE increased significantly. When a CT scan needs to be converted from high (140) to a lower (120) kVp scan, the HU values need to be increased. The rate at which it should be increased was calculated based on the Ghent University Hospital data (see section 3). The Philips Somatom Definition Flash CT scanner is used within the Ghent University Hospital to acquire the DECT data. The 140 kVp scan is filtered using a tin filter. This filter increases the effective energy in the beam [35]. This increase in effective photon energy results in a lower HU value being assigned to bone when compared to a 140 kVp scan without tin filtering. This filtering is useful to increase spectral separation between the low- and high energy spectra and to increases the dose efficiency of the acquired scan [35]. It does however lead to an 'overestimate' of the difference between 120 and 140 kVp scans (when tin filtering is not applied). This results in a large error when the kVp conversion method is applied on 140 kVp data acquired without tin filtering, as was the case for the Jilin data. Additional information about the source of this error can be found in Appendix D.

This experiment therefore indicates that a kVp conversion method can only be applied on CT scans originating from a CT scanner model that is equal to the CT scanner model of the data where the conversion method was trained on. The kVp conversion methods cannot be applied to CT data from different CT scanner models, as a change in the photon spectrum will alter the required conversion.

5 Reconstruction kernel conversion experiment

The CT reconstruction kernel does also have a large influence on the HU values assigned to voxels in the reconstructed image (see section 2.2.2). Therefore, a method was developed to convert images from a soft reconstruction to a sharp reconstruction.

5.1 Data:

The data available to develop this method contained CT images from 26 patients. The data was acquired using a Siemens Somatom Definition Flash CT scanner at 100 kVp. The images were reconstructed using the Q30f (soft) and I50f (sharp) reconstruction kernels developed by Siemens Healthineers. The slice thickness of the images was 1 mm. The pixels spacing varied between 0.55 mm and 0.90 mm (depending on the patient). The available data was divided into training, validation and testing datasets, according to the commonly used ratio of 60:20:20 (16:5:5 patients). An example of the data being used can be seen in Figure 26.

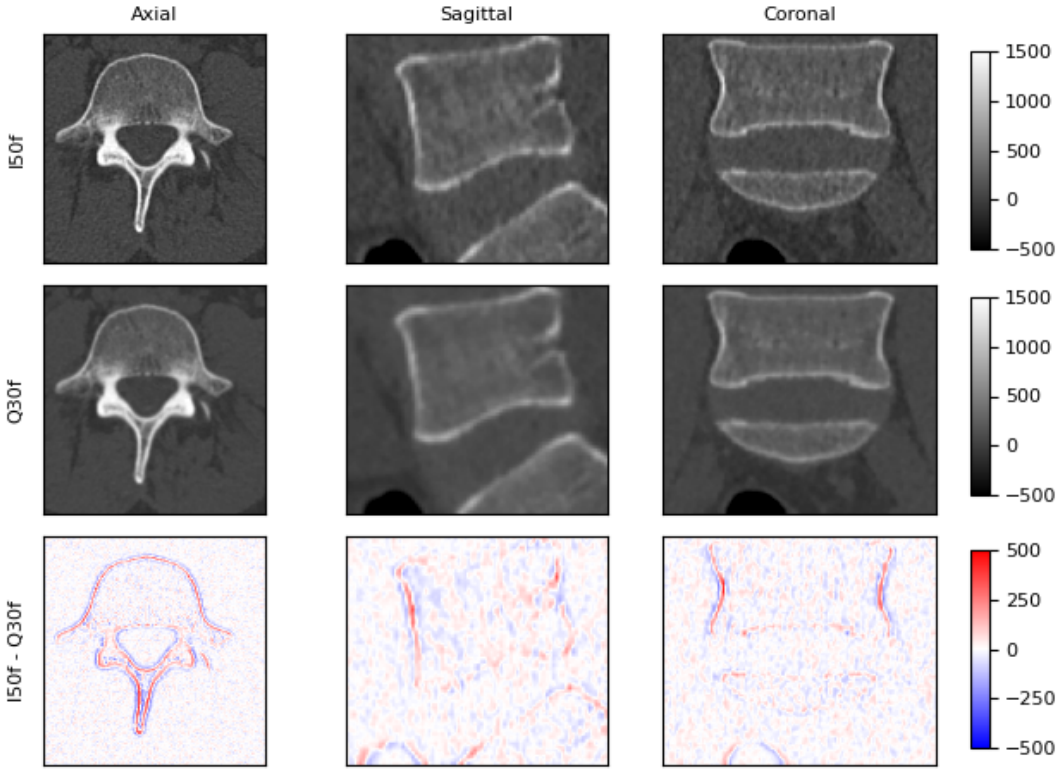

Figure 26: Illustration of the source (Q30f) and target (I50f) data and their differences. Values are in HU.

5.2 Method

A deep learning approach was implemented to convert the images from a Q30f reconstruction to a I50f reconstruction. The network was the same 3D-CNN as was used for the kVp conversion (see section 3.2.2). All hyperparameters were left unchanged.

5.3 Evaluation

The quality of the predictions was quantified using voxel-wise calculations between the target and prediction images. The ME and MAE were used to quantify the quality of the prediction. The standard deviation of the ME and MAE between patients was also calculated. The perceived image sharpness was quantified using the acutance. This metric describes the subjective perception of sharpness within an image [45]. The acutance of an image can be computed by taking the mean of a Sobel-filtered image (using a 3D Sobel-Feldman kernel). A ratio of the acutance can then be calculated by dividing the acutance of the input or predicted image by the acutance of the target image. A ratio close to one indicates a similar image sharpness.

The mean absolute surface distance was calculated between segmentations of the bone. Three separate segmentations of the bone were made of the input, target- and prediction images. The surface distance was calculated between segmentations of the target and input images. This metric was then compared to the surface distance between segmentations of the target and prediction images.

5.4 Results

The ME within bone voxels reduced from 47 (±6) HU to 5 (±1) HU after conversion, as shown in Figure 29. The MAE in bone reduced from 73 (±7) HU to 14 (±1) HU. The MAE of voxels containing soft tissue increased from 24 (±2) HU to 7 (±1) HU after conversion. The MAE in air voxels reduced from 8 (±2) HU to 2 (±0) HU. Visually, the predictions appear to be as sharp as the target image (see Figure 27). The difference images show a clear improvement after the conversion, as can be seen in Figure 28.

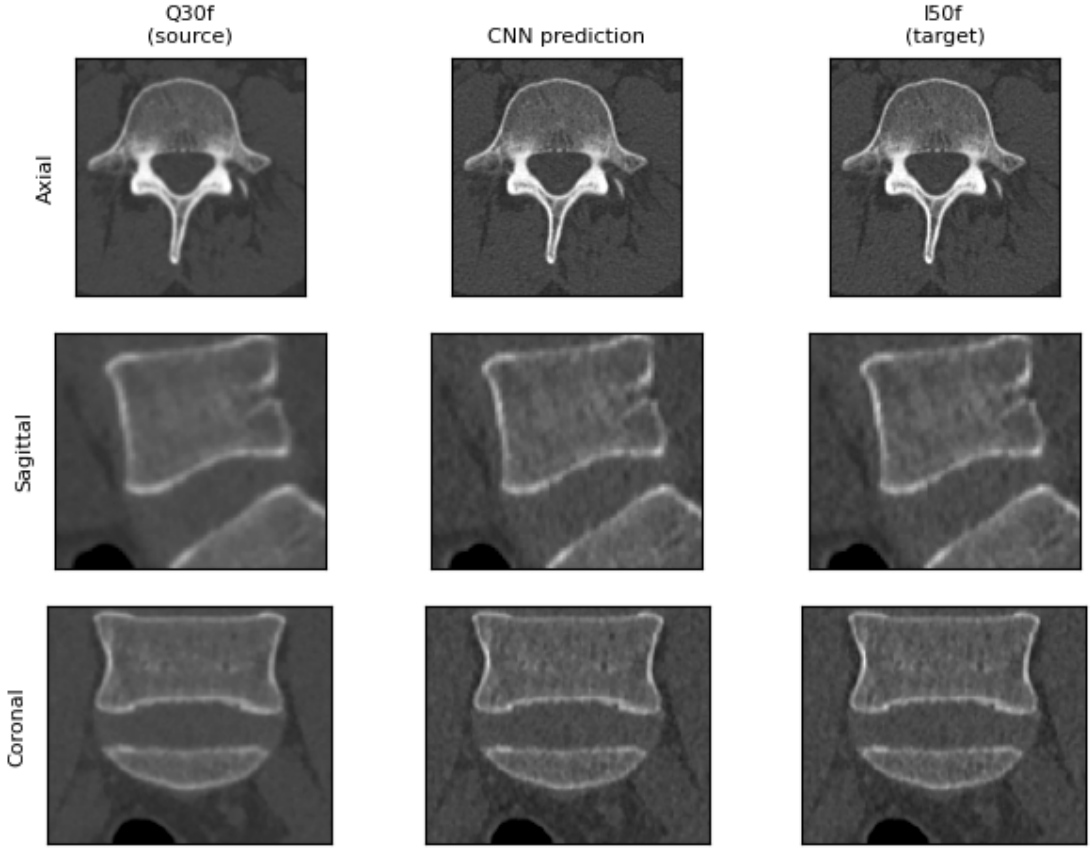

Figure 27: Example of the source, prediction and target data. W/L: 2000/500 HU.

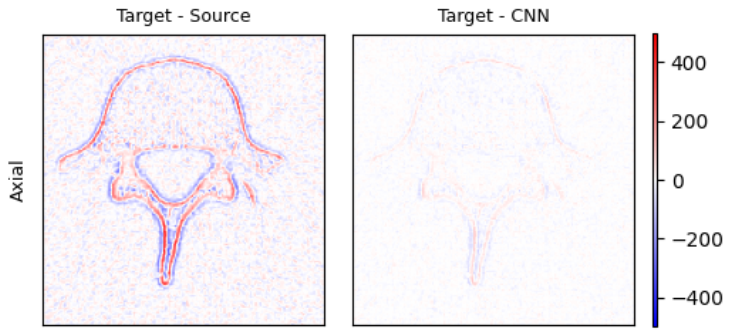

Figure 28: Voxelwise difference target and source (I50f, Q30f) and target and prediction (I50f, CNN). W/L: 1000/0 HU.

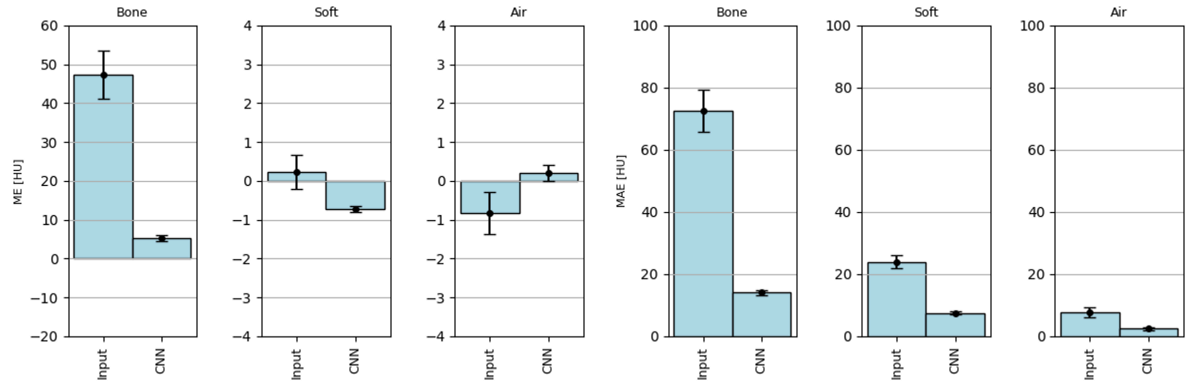

Figure 29: Mean error (left) and mean absolute error (right) values between target (sharp) and predictions for different segmentations. The input error is the error between the target (sharp) and input (soft) images.

The signal for the input (Q30f), target (I50f) and prediction along a line can be seen in Figure 30. The prediction closely resembles the target signal after the convolutional neural network has been applied to the input. The acutance ratio between the input and target images was equal to 0.68 (±0.02). The acutance ratio between the prediction and target images was 0.98 (±0.01). The mean absolute surface distance of the bone reduced from 0.17 (±0.06) mm to 0.02 (±0.01) mm after the conversion was applied. A 3D visualization of the surface distance before- and after conversion is shown in Figure 31.

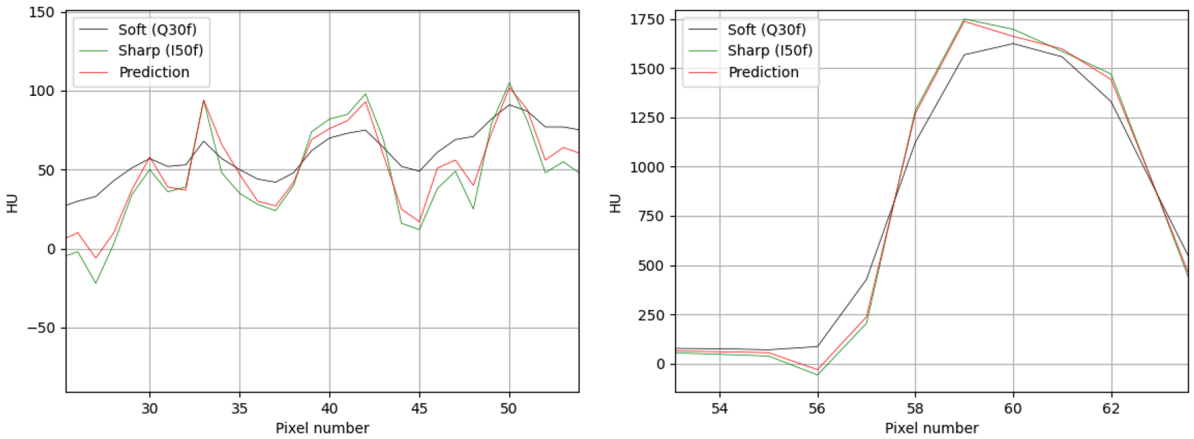

Figure 30: Plot of the signal of the input (soft), target (sharp) and prediction along a line in a CT image. The regions contain soft tissue (left) and bone tissue (right).

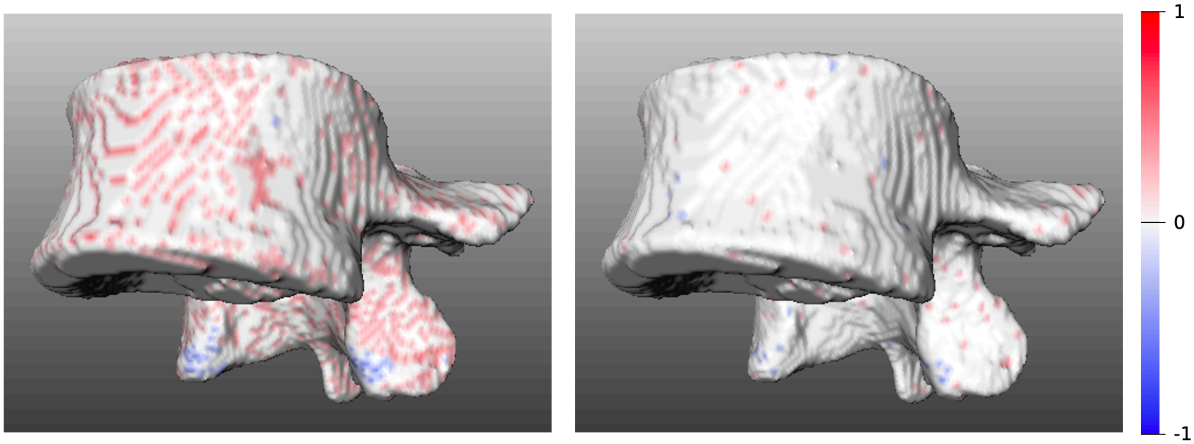

Figure 31: Segmentation of the fifth lumbar spine, with an overlayed surface distance between target and input (left) and target and prediction (right). The surface distance is in mm.

5.5 Discussion

The conversion method decreased the ME and MAE that was present between the source and target images. The difference images indicate an increased similarity after the conversion was applied. The perceived image sharpness (acutance) of the images became very close to the target acutance after the conversion method was applied. The conversion method reduced the surface distance that was present between the bone segmentations.

6 Applying the reconstruction kernel conversion method to different data

The developed reconstruction kernel conversion method of Section 5 can be applied to data from a different dataset. The method that was developed was trained to convert images from a soft (Q30f) to a sharp (I50f) reconstruction kernel. It could be useful to apply this method on CT data that was reconstructed using a different soft kernel, as it can potentially increase the image sharpness. An increased CT image sharpness will benefit the training and evaluation of BoneMRI images, as a sharp BoneMRI is a desired property.

6.1 Data

A dataset was available containing CT images from the Isala Hospital (Zwolle, the Netherlands). The CT images were acquired using a Philips Brilliance iCT 256 scanner. The images covered the neck and cervical spine region. Both soft (B kernel) and sharp (D kernel) reconstructions were available for 25 patients. These reconstruction kernels are developed by Philips. Both reconstructions were acquired at 120 kVp, with a pixel spacing of 0.49×0.49 mm. An example of the available data and the difference between the reconstructions can be seen in Figure 32.

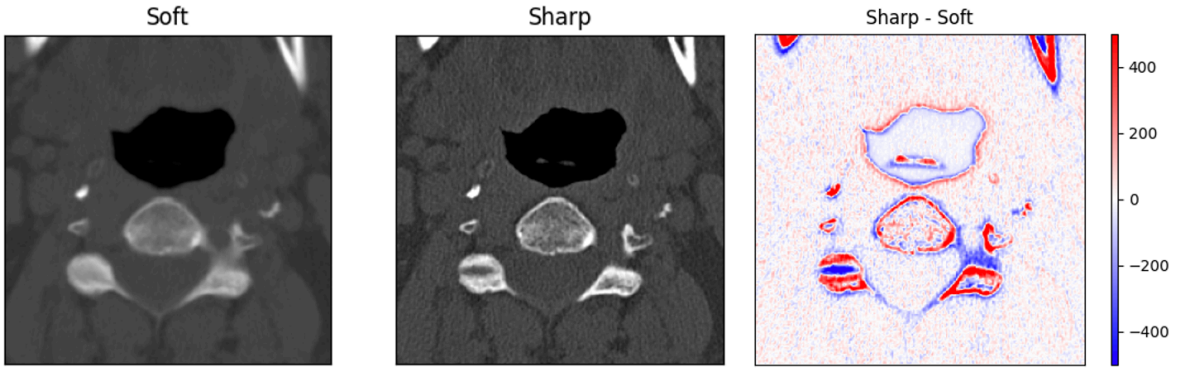

Figure 32: Example of the available soft (B kernel) and sharp (D kernel) reconstructions from the Isala Hospital. The soft reconstruction had a slice thickness of 3 mm. The sharp reconstruction was resampled from a slice thickness of 0.5 mm to a slice thickness of 3 mm. W/L left and middle images: 2000/500 HU. W/L right image: 1000/0 HU.

The image reconstructed with a D kernel (sharp) had a slice thickness of 0.5 mm, while the reconstructions with a B kernel (soft) had a slice thickness of 3.0 mm. The sharp reconstructions were therefore resampled to a slice thickness of 3.0 mm.

6.2 Method

The soft reconstruction was converted using the reconstruction kernel conversion method developed in Section 5. This conversion method was trained to convert images from a soft (Q30f) to a sharp (I50f) reconstruction kernel (acquired using a Siemens Somatom Definition Flash scanner).

6.3 Evaluation

The ME was calculated for every cervical vertebra (separately) between the soft and sharp reconstructions. Segmentations of all cervical vertebrae were made and used to calculate the surface distance between the segmentations. The acutance ratio was calculated between the soft and sharp reconstructions. The mean error, surface distances and acutance ratio were also calculated between the converted CT image and target CT image. The ME and surface distance metric values before- and after conversion were compared using a Welch's *t*-test with a significance level of $\alpha = 0.05$.

6.4 Results

The ME and surface distances before- and after conversion are shown in Figure 35. The ME did not decrease significantly for every cervical vertebra (p>0.05). An error between the converted image and target image remained for every cervical vertebra. The surface distances did not improve significantly after conversion (p>0.05). Visual inspection of the images suggests that the input (soft) images became sharper (see Figure 33 and Figure 34). The acutance ratio increased from 0.66 (±0.01) to 0.76 (±0.01) after conversion. An edge response along a line for the input, prediction and target is shown in Figure 36. The edge plot and acutance ratios confirm the visual observation that the prediction became more sharp than the input, but did not equal the sharpness of the target image.

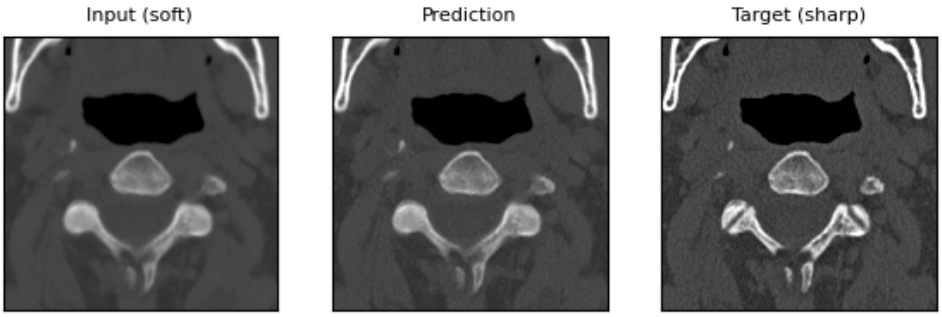

Figure 33: Example of input (soft), prediction, and target (sharp) images. W/L: 2000/500 HU.

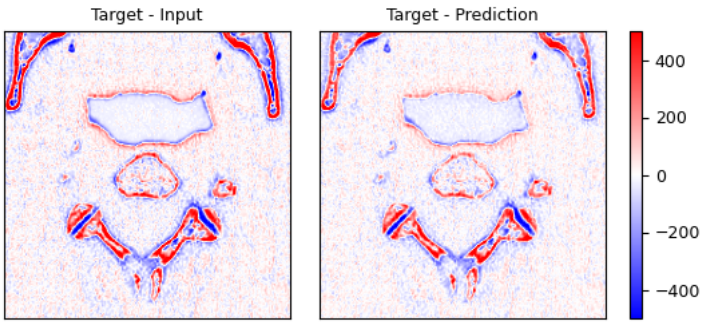

Figure 34: Difference image between target and input (left) and target and prediction (right). W/L: 1000/0 HU.

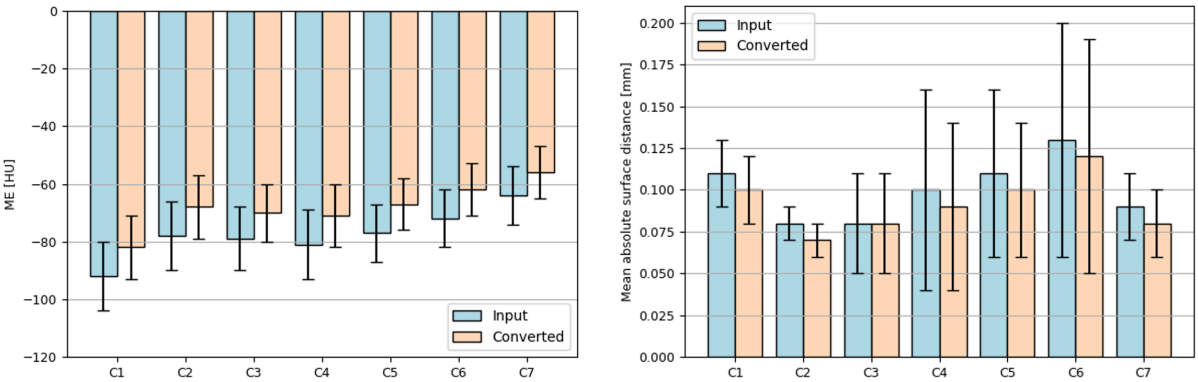

Figure 35: Left: Mean error between the target input image before- and after conversion. The values are calculated for every cervical vertebra separately. Right: The mean absolute surface distances between the segmentation of each target vertebra and the segmentation of the vertebra before- or after conversion.

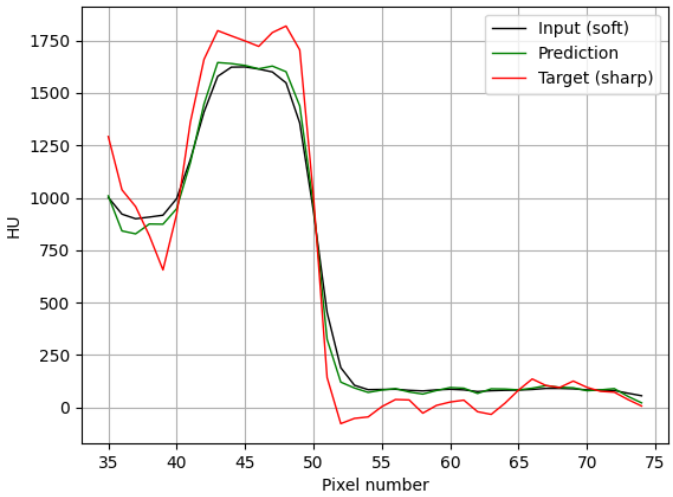

Figure 36: Edge response for input, prediction, and target.

6.5 Discussion

The resampling of the sharp reconstruction from a slice thickness of 0.5 mm to 3.0 mm was done using linear interpolation. The resampling caused smoothening of the target image. In an ideal situation, all acquisition- and reconstruction settings except from the reconstruction kernel would have been equal. Such data was unfortunately unavailable at the time.

The metrics indicate that the input image sharpness increased by applying the conversion method. The prediction does, however, not equal the sharpness of the target image (Philips D kernel). This confirms that the exact appearance of an image is reconstruction kernel specific. The difference between soft and sharp reconstructions from Siemens are not equal to the difference of soft and sharp reconstructions from Philips. The developed kernel conversion method can, however, be used to improve the sharpness of an image.

This method can be used during the development of BoneMRI images, by using the conversion method to increase the sharpness of the training CT images. This will ensure that the BoneMRI predictions have an increased sharpness, which is a desirable property of the images.

7 Discussion

The conversion from 100- or 140 kVp to 120 kVp is possible using a CNN or a piecewise linear conversion (see section 3). The CNN and piecewise linear conversion methods show an equal performance according to the evaluation metrics. One could, however, prefer to use the piecewise linear conversion method because it is clear what happens to the data, and will not induce any artifacts regardless of the input. This cannot be said about the CNN, since a small change in the data distribution can possibly have an undesired effect on the predicted image. The kVp conversion methods might not work for images containing foreign materials like metal or contrast agents. As these were not present within the training or evaluation data of the conversion methods. Still, the response of the piecewise linear conversion to a foreign object is more predictable than the response of the CNN.

One major limitation of the study is the use of synthetic 120 kVp scans. The 120 kVp scans were the average of the scans acquired at 100- and 140 kVp. A nonlinear combination of the 100- and 140 kVp scans will most likely resemble a true 120 kVp scan more closely. This makes that the developed conversion methods convert the CT images to an incorrect target distribution. A true (paired) scan acquired at 120 kVp is needed to correctly convert 100- or 140 kVp CT images to 120 kVp. It might be relatively easy to improve the piecewise linear conversion method by scanning a phantom at 100-120-, and 140 kVp. A correct conversion curve between the kVp's can then be calculated if the phantom contains air, water, and bone-like materials.

The conversion of the kVp of images is beneficial for the evaluation of BoneMRI images. The kVp conversion of images is, however, depending on the exact photon spectra of the two CT scans. This makes the kVp conversion method CT scanner model specific, as the spectrum is depending on the characteristics of the CT scanner. Additional paired data is needed if one wants to convert the kVp of an image acquired with a different CT scanner model. It is, however, difficult to get paired patient data acquired at multiple kVp settings, as every CT scan adds radiation to the patient. To avoid this problem, one could use phantoms to make multiple scans acquired with varying acquisition settings. It does, however, still cost time and money to acquire multiple CT scans. Furthermore, it is difficult to obtain data using multiple scanner models, as each hospital only has one (or a few) CT scanner models.

The reconstruction kernel conversion method can be used to sharpen CT images. Paired data was needed to convert CT images from a specific reconstruction kernel to another. The acquisition of paired data reconstructed with different kernels is relatively easy to obtain, as it does not need an additional CT scan. It can simply be reconstructed twice using the same raw sinogram data. The reconstruction of an image does take some time. For this reason, only one reconstruction is usually calculated by the hospital. A conversion method can be trained to convert the images from one reconstruction kernel to another if desired. Paired data of a single patient is sufficient to train such a kernel conversion method (see Appendix E).

In general, paired data was needed to convert between two image distributions using a convolutional neural network. Several deep learning architectures have been developed that do not need paired data to convert images from one domain to another. An example of such an architecture is the GAN. A GAN can be trained to convert input images to have a similar appearance as images from a target domain [23]. The use of such a network seems promising to convert CT data [24]. It is, however, not advised to use it to synthesize medical images, as it can lead to image artifacts [18]. Since image artifacts are undesired in converted images, one will always need paired data to verify the absence of image artifacts. This makes it impossible to develop a conversion method without paired data.

A U-Net could have been used to convert the CT images from one domain to another, as was done by other researchers [41, 69]. However, it was decided to use a CNN because a U-Net quickly contains over a million trainable parameters. The developed CNN had only 225.217 trainable parameters, making it less prone to overfitting when a small amount of data is used.

7.1 Future work and recommendations

The kVp and kernel conversion methods can be compared to methods developed by other researchers (see Section 2.3.1). This was difficult to carry out since the code and data of the researchers were not publicly available.

Other image quality metrics can be calculated to check image similarity. The structural similarity (SSIM) can be calculated to assess the visual image similarity more thoroughly [61]. It was however chosen not to include this, as this metric cannot be computed on segmented images. Next to quantitative measurements, a radiologist can be asked to assess the image similarity between the prediction and target images. This can potentially detect artifacts that are not discovered when one relies completely on voxelwise error metrics.

The hyperparameter optimization was done using manual tuning. More advanced tuning methods can potentially increase the model performance. An example of a hyperparameter optimization technique is the coarse-to-fine method [46]. Here, you discretize the available range of parameters into a grid of values. After selecting the value that seems most promising, you can perform a hyperparameter search on a finer grid of values. Another possibility is to use a Bayesian hyperparameter optimization [64]. Bayesian optimization uses the performance of previous sets of hyperparameters to choose the next combination of hyperparameters. This automated optimization strategy is said to achieve great accuracy within a few iterations, which makes the optimalization process faster compared to the manual search strategy [64].

8 Conclusion

CT images acquired with a particular kVp can be converted to another kVp using a piecewise linear fit or a convolutional neural network. The conversion between two kVp settings is possible if paired data is available. The conversion is CT scanner model specific and does not generalize to data acquired using different scanner models.

The kVp conversion methods improve the evaluation consistency of BoneMRI images. They can be used if there are scans included in the evaluation data that were acquired with a different kVp than was used for the training data. The conversion to a single kVp decreases the spread in BoneMRI evaluation metrics. The kVp conversion method needs to be trained on data acquired using the same CT scanner model.

CT images that were reconstructed with a particular reconstruction kernel can be converted to another reconstruction kernel using a convolutional neural network. Paired data of at least one patient is needed to train the conversion method. A CNN was successfully trained to convert soft reconstructions to a sharp reconstruction. The method can be applied to different CT data to sharpen the images.

References

1. Abella, M., Alessio, A. M., Mankoff, D. A., MacDonald, L. R., Vaquero, J. J., Desco, M., & Kinahan, P. E. (2012). Accuracy of CT-based attenuation correction in PET/CT bone imaging. *Physics in Medicine & Biology*, *57*(9), 2477.
2. Adams, J. M., Cockburn, M. I., Difazio, L. T., Garcia, F. A., Siegel, B. K., & Bilaniuk, J. W. (2006). Spinal clearance in the difficult trauma patient: a role for screening MRI of the spine. *The American surgeon*, *72*(1), 101-105.
3. Afifi, M. B., Abdelrazek, A., Deiab, N. A., Abd El-Hafez, A., & El-Farrash, A. (2020). The effects of CT x-ray tube voltage and current variations on the relative electron density (RED) and CT number conversion curves. *Journal of Radiation Research and Applied Sciences*, *13*(1), 1-11.
4. Ahishakiye, E., Bastiaan Van Gijzen, M., Tumwiine, J., Wario, R., & Obungoloch, J. (2021). A survey on deep learning in medical image reconstruction. *Intelligent Medicine*, *1*(03), 118-127.
5. Al-Hayek, Y., Spuur, K., Zheng, X., Hayre, C., & Davidson, R. (2021). The reliability of CT numbers as absolute values for diagnostic scanning, dental imaging, and radiation therapy simulation: A narrative review. *Journal of Medical Imaging and Radiation Sciences*, *53*(1), 128-146.
6. *The alliance for quality computed tomography*. American association of physisists in medicine. Retrieved in March 2022 from https://www.aapm.org/pubs/ctprotocols/?tab=5#CTabbedPanels
7. Anaya-Isaza, A., Mera-Jiménez, L., & Zequera-Diaz, M. (2021). An overview of deep learning in medical imaging. *Informatics in Medicine Unlocked*, *26*, 100723.
8. Blackham, K., & Vidal, L. (2014). Computed tomography. *Encyclopedia of the Neurological Sciences*, *2*, 848-853.
9. Boedeker, K. L., McNitt-Gray, M. F., Rogers, S. R., Truong, D. A., Brown, M. S., Gjertson, D. W., & Goldin, J. G. (2004). Emphysema: effect of reconstruction algorithm on CT imaging measures. *Radiology*, *232*(1), 295-301.
10. Bolus, D., Morgan, D., & Berland, L. (2017). Effective use of the Hounsfield unit in the age of variable energy CT. *Abdominal Radiology*, *42*(3), 766-771.
11. *BoneMRI*. MRIguidance. Retrieved in March 2022 from https://mriguidance.com/bonemri/
12. Bos, A., Draaisma, F., & Okx, W. Inleiding tot de stralingshygiëne. 2007. *Den Haag: Sdu Uitgevers bv*.
13. Brindhaban, A., & Jassim, O. (2021). Effect of X-ray beam energy and image reconstruction technique on computed tomography numbers of various tissue equivalent materials. *Radiography*, *27*(1), 95-100.
14. Bryant, J., Drage, N. A., & Richmond, S. (2012). CT number definition. *Radiation Physics and Chemistry*, *81*(4), 358-361.
15. Bushberg, J. T. (2003). The essential physics of medical imaging. *Medical Physics*, *30*(7), 1936.
16. Choe, J., Lee, S. M., Do, K.-H., Lee, G., Lee, J.-G., Lee, S. M., & Seo, J. B. (2019). Deep learning–based image conversion of CT reconstruction kernels improves radiomics reproducibility for pulmonary nodules or masses. *Radiology*, *292*(2), 365-373.
17. Chollet, F. (2021). *Deep learning with Python* (1, Ed.). Manning Publications.
18. Cohen, J. P., Luck, M., & Honari, S. (2018). Distribution matching losses can hallucinate features in medical image translation. *International conference on medical image computing and computer-assisted intervention*, 529-536.
19. Cropp, R. J., Seslija, P., Tso, D., & Thakur, Y. (2013). Scanner and kVp dependence of measured CT numbers in the ACR CT phantom. *Journal of Applied Clinical Medical Physics*, *14*(6), 338-349.
20. Dirix, P., Haustermans, K., & Vandecaveye, V. (2014). The value of magnetic resonance imaging for radiotherapy planning. *Seminars in radiation oncology*, *24*(3), 151-159.

21. Florkow, M. C., Willemsen, K., Mascarenhas, V. V., Oei, E. H., van Stralen, M., & Seevinck, P. R. (2022). Magnetic Resonance Imaging Versus Computed Tomography for Three-Dimensional Bone Imaging of Musculoskeletal Pathologies: A Review. *Journal of Magnetic Resonance Imaging*, *56*(1), 11-34.
22. Florkow, M. C., Willemsen, K., Zijlstra, F., Foppen, W., van der Wal, B. C., van der Voort van Zyp, J. R., Viergever, M. A., Castelein, R. M., Weinans, H., & van Stralen, M. (2021). MRI-based synthetic CT shows equivalence to conventional CT for the morphological assessment of the hip joint. *Journal of Orthopaedic Research*, *40*(4), 954-964.
23. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.
24. Gravina, M., Marrone, S., Docimo, L., Santini, M., Fiorelli, A., Parmeggiani, D., & Sansone, C. (2022). Leveraging CycleGAN in Lung CT Sinogram-free Kernel Conversion. International Conference on Image Analysis and Processing,
25. Haghighi, B., Noël, P. B., Cohen, E., Pantalone, L., Vachani, A., Rendle, K. A., Wainwright, J., Saia, C., Barbosa Jr, E. M., & Kontos, D. (2020). Assessment of CT image reconstruction parameters on radiomic features in a lung cancer screening cohort: the PROSPR study. *Medical Imaging 2020: Computer-Aided Diagnosis*, *11314*, 571-576.
26. Heismann, B. J., Schmidt, B. T., & Flohr, T. (2012). Spectral computed tomography. *SPIE Bellingham*, 1-132.
27. Hsieh, J. (2009). Key performance parameters of the CT scanner. *Computed Tomography: Principles, Design, Artifacts, and Recent Advances. 2nd ed. New York, NY: Wiley*.
28. Huang, R., Liu, X., He, L., & Zhou, P.-K. (2020). Radiation exposure associated with computed tomography in childhood and the subsequent risk of cancer: a meta-analysis of cohort studies. *Dose-Response*, *18*(2), 1559325820923828.
29. Hubbell, J., & Seltzer, S. (2004). NIST standard reference database 126: X-ray mass attenuation coefficients. *National Institute of Standards and Technology, editor. Gaithersburg*.
30. Hudgins, P. A., & Baugnon, K. L. (2018). Head and neck: skull base imaging. *Neurosurgery*, *82*(3), 255-267.
31. Jans, L. B., Chen, M., Elewaut, D., Van den Bosch, F., Carron, P., Jacques, P., Wittoek, R., Jaremko, J. L., & Herregods, N. (2021). MRI-based synthetic CT in the detection of structural lesions in patients with suspected sacroiliitis: comparison with MRI. *Radiology*, *298*(2), 343-349.
32. Kabil, J., Felblinger, J., Vuissoz, P. A., & Missoffe, A. (2020). Coupled transfer function model for the evaluation of implanted cables safety in MRI. *Magnetic resonance in medicine*, *84*(2), 991-999.
33. Khalid, S., Khalil, T., & Nasreen, S. (2014). A survey of feature selection and feature extraction techniques in machine learning. *2014 science and information conference*, 372-378.
34. Kim, J., Hong, J., & Park, H. (2018). Prospects of deep learning for medical imaging. *Precision and Future Medicine*, *2*(2), 37-52.
35. Krauss, B., Schmidt, B., & Flohr, T. G. (2011). Dual source CT. In *Dual energy CT in clinical practice* (pp. 11-20). First edition. Springer.
36. Lenchik, L., Rogers, L. F., Delmas, P. D., & Genant, H. K. (2004). Diagnosis of osteoporotic vertebral fractures: importance of recognition and description by radiologists. *American Journal of Roentgenology*, *183*(4), 949-958.
37. Li, T., Wei, L., & Hsu, W. (2021). A Multi-Pronged Evaluation For Image Normalization Techniques. *2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI)*, 1292-1296.

38. Liang, G., Fouladvand, S., Zhang, J., Brooks, M. A., Jacobs, N., & Chen, J. (2019). Ganai: Standardizing ct images using generative adversarial network with alternative improvement. *2019 IEEE International Conference on Healthcare Informatics (ICHI)*, 1-11.
39. Ligero, M., Jordi-Ollero, O., Bernatowicz, K., Garcia-Ruiz, A., Delgado-Muñoz, E., Leiva, D., Mast, R., Suarez, C., Sala-Llonch, R., & Calvo, N. (2021). Minimizing acquisition-related radiomics variability by image resampling and batch effect correction to allow for large-scale data analysis. *European radiology*, *31*(3), 1460-1470.
40. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., Van Der Laak, J. A., Van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. *Medical image analysis*, *42*, 60-88.
41. Liu, C.-K., Liu, C.-C., Yang, C.-H., & Huang, H.-M. (2021). Generation of Brain Dual-Energy CT from Single-Energy CT Using Deep Learning. *Journal of Digital Imaging*, *34*(1), 149-161.
42. Lu, L., Ehmke, R. C., Schwartz, L. H., & Zhao, B. (2016). Assessing agreement between radiomic features computed for multiple CT imaging settings. *Plos One*, *11*(12), e0166550.
43. Lyu, T., Zhao, W., Zhu, Y., Wu, Z., Zhang, Y., Chen, Y., Luo, L., Li, S., & Xing, L. (2021). Estimating dual-energy CT imaging from single-energy CT data with material decomposition convolutional neural network. *Medical image analysis*, *70*, 102001.
44. MacDougall, R. D., Kleinman, P. L., & Callahan, M. J. (2016). Size-based protocol optimization using automatic tube current modulation and automatic kV selection in computed tomography. *Journal of Applied Clinical Medical Physics*, *17*(1), 328-341.
45. Maître, H. (2015). Image quality. *From Photon to Pixel: The Digital Camera Handbook*, 205-251.
46. Michelucci, U. (2018). Applied Deep Learning. *A Case-Based Approach to Understanding Deep Neural Networks*.
47. Nagayama, Y., Sakabe, D., Goto, M., Emoto, T., Oda, S., Nakaura, T., Kidoh, M., Uetani, H., Funama, Y., & Hirai, T. (2021). Deep Learning–based Reconstruction for Lower-Dose Pediatric CT: Technical Principles, Image Characteristics, and Clinical Implementations. *Radiographics*, *41*(7), 1936-1953.
48. Okayama, S., Soeda, T., Takami, Y., Kawakami, R., Somekawa, S., Uemura, S., & Saito, Y. (2012). The Influence of Effective Energy on Computed Tomography Number Depends on Tissue Characteristics in Monoenergetic Cardiac Imaging. *Radiology Research and Practice*, *2012*, 150980.
49. Pal, K. K., & Sudeep, K. (2016). Preprocessing for image classification by convolutional neural networks. *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)*, 1778-1781.
50. Ranganathan, G. (2021). A study to find facts behind preprocessing on deep learning algorithms. *Journal of Innovative Image Processing (JIIP)*, *3*(01), 66-74.
51. Robins, M., Solomon, J., Hoye, J., Abadi, E., Marin, D., & Samei, E. (2019). Systematic analysis of bias and variability of texture measurements in computed tomography. *Journal of Medical Imaging*, *6*(3), 033503.
52. Schwarzband, G., & Kiryati, N. (2005). The point spread function of spiral CT. *Physics in Medicine & Biology*, *50*(22), 5307.
53. Seeram, E. (2015). *Computed Tomography-E-Book: Physical Principles, Clinical Applications, and Quality Control* (4 ed.). Elsevier Health Sciences.
54. Selim, M., Zhang, J., Fei, B., Zhang, G.-Q., & Chen, J. (2020). STAN-CT: Standardizing CT Image using Generative Adversarial Networks. *AMIA Annual Symposium Proceedings*, *2020*, 1100-1109.
55. Selim, M., Zhang, J., Fei, B., Zhang, G.-Q., & Chen, J. (2021). Ct image harmonization for enhancing radiomics studies. *arXiv preprint arXiv:2107.01337*.

56. Setiyono, P. A., Rochmayanti, D., Kurniawan, A. N., & Setiawan, A. N. (2020). The optimization of mastoid CT image using windows and kernel reconstructions. *Journal of Physics: Conference Series*, *1471*(1), 012015.
57. Staartjes, V. E., Seevinck, P. R., Vandertop, W. P., van Stralen, M., & Schröder, M. L. (2021). Magnetic resonance imaging–based synthetic computed tomography of the lumbar spine for surgical planning: a clinical proof-of-concept. *Neurosurgical Focus*, *50*(1), E13.
58. van der Kolk, B. B. Y., Slotman, D. J. J., Nijholt, I. M., van Osch, J. A., Snoeijink, T. J., Podlogar, M., van Hasselt, B. A., Boelhouwers, H. J., van Stralen, M., & Seevinck, P. R. (2022). Bone visualization of the cervical spine with deep learning-based synthetic CT compared to conventional CT: A single-center noninferiority study on image quality. *European journal of radiology*, *154*, 110414.
59. van Timmeren, J. E., Cester, D., Tanadini-Lang, S., Alkadhi, H., & Baessler, B. (2020). Radiomics in medical imaging—"How-to" guide and critical reflection. *Insights into imaging*, *11*(1), 1-16.
60. Vopat, B. G., Cai, W., Torriani, M., Vopat, M. L., Hemma, M., Harris, G. J., Schantz, K., & Provencher, M. T. (2018). Measurement of glenoid bone loss with 3-dimensional magnetic resonance imaging: a matched computed tomography analysis. *Arthroscopy: The Journal of Arthroscopic & Related Surgery*, *34*(12), 3141-3147.
61. Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), 600-612.
62. Wei, L. (2021). *Using AI to Mitigate Variability in CT Scans: Improving Consistency in Medical Image Analysis.* University of California, Los Angeles.
63. Wong, K. C. (2016). 3D-printed patient-specific applications in orthopedics. *Orthopedic research and reviews*, *8*, 57-66.
64. Wu, J., Chen, X.-Y., Zhang, H., Xiong, L.-D., Lei, H., & Deng, S.-H. (2019). Hyperparameter optimization for machine learning models based on Bayesian optimization. *Journal of Electronic Science and Technology*, *17*(1), 26-40.
65. Yagi, M., Ueguchi, T., Koizumi, M., Ogata, T., Yamada, S., Takahashi, Y., Sumida, I., Akino, Y., Konishi, K., & Isohashi, F. (2013). Gemstone spectral imaging: determination of CT to ED conversion curves for radiotherapy treatment planning. *Journal of Applied Clinical Medical Physics*, *14*(5), 173-186.
66. Yang, S., Kim, E. Y., & Ye, J. C. (2021). Continuous conversion of CT kernel using switchable CycleGAN with AdaIN. *IEEE transactions on medical imaging*, *40*(11), 3015-3029.
67. Yu, L., Li, H., Fletcher, J. G., & McCollough, C. H. (2010). Automatic selection of tube potential for radiation dose reduction in CT: a general strategy. *Medical Physics*, *37*(1), 234-243.
68. Zhao, J. (2016). Investigation of dose for the triple-source computed tomography based on Monte Carlo simulations. *J. Imaging Sci*, *1*(1), 1-7.
69. Zhao, W., Lv, T., Gao, P., Shen, L., Dai, X., Cheng, K., Jia, M., Chen, Y., & Xing, L. (2019). Dual-energy CT imaging using a single-energy CT data is feasible via deep learning. *arXiv preprint*, 1-7.

Appendices

Appendix A: Final model architecture

Table 3: Final model architecture of the convolutional neural network.

Layer type	Output shape	Number of parameters
Conv3D	(None, 64, 24, 24, 24)	1792
Batch Normalization	(None, 64, 24, 24, 24)	256
Conv3D	(None, 64, 24, 24, 24)	110656
Batch Normalization	(None, 64, 24, 24, 24)	256
Conv3D	(None, 64, 24, 24, 24)	110656
Batch Normalization	(None, 64, 24, 24, 24)	256
Conv3D	(None, 1, 24, 24, 24)	1729
Total params:		225,601
Trainable params:		225,217
Non-trainable params:		384

Appendix B: Experiments for model optimization

The model architecture and hyperparameters were tuned manually. The results of the optimization process are shown in this section.

Using batch normalization

Adding batch normalization increased the convergence speed of the network.

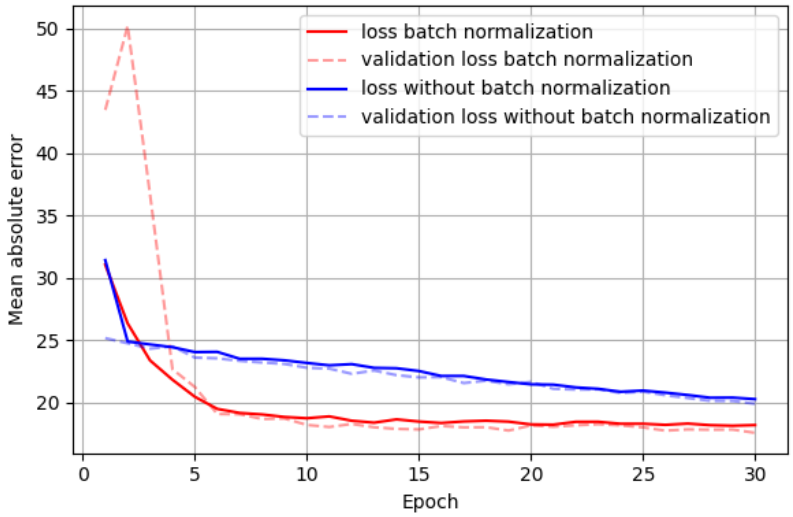

Figure 37: The training history of the convolutional neural network with- and without batch normalization after every convolutional layer.

Amount of filters

The number of filters in every convolutional layer was varied. The MAE within the training and validation data decreased by increasing the number of filters. Increasing the number of filters beyond 64 showed only minimal improvement. The number of filters was therefore chosen to be 64.

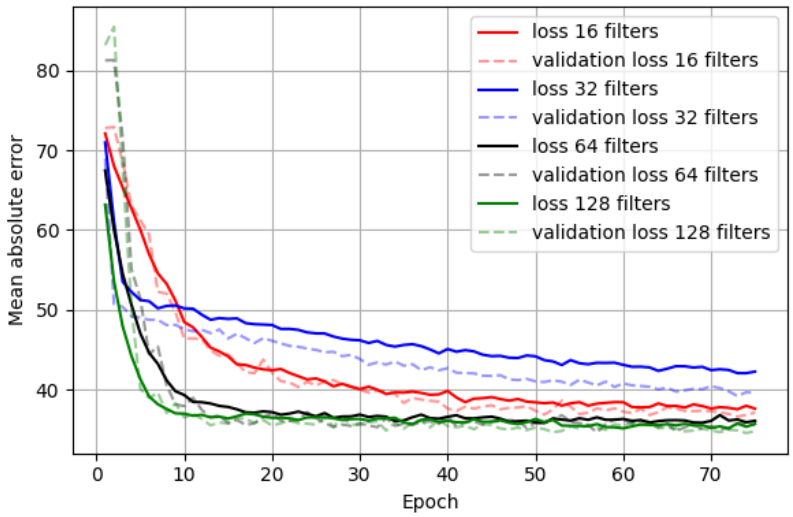

Figure 38: Training loss of the convolutional neural network with 16 (red), 32 (blue), 64 (black) and 128 (green) filters per convolutional layer.

Amount of layers

Increasing the number of convolutional layers from one to three decreased the MAE, as can be seen in Figure 39. Increasing it further to 5 layers showed minimal improvement. Therefore, the final model used three convolutional layers.

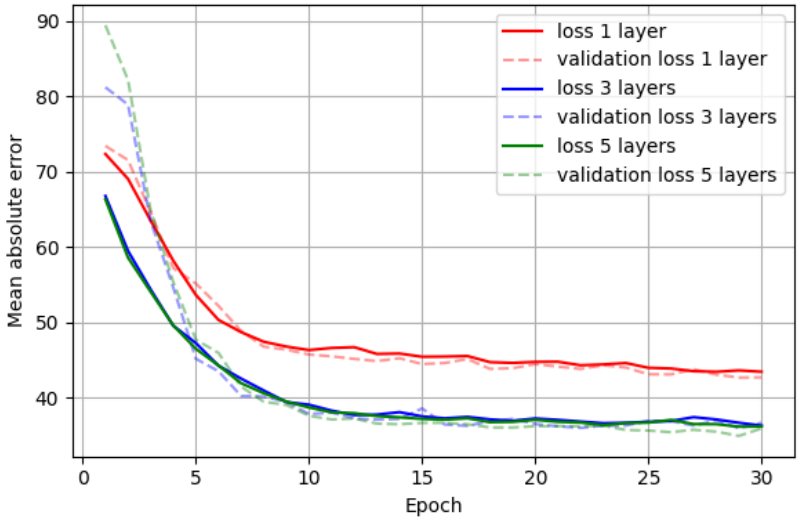

Figure 39: Training history of the convolutional neural network with one (red), three (blue) or 5 (green) convolutional layers.

Removing the skip connection

The usage of a skip connection within the network (see Figure 16) increased convergence speed, as shown in Figure 40. The skip connection also reduced the mean error of the predictions, as can be seen in Figure 41.

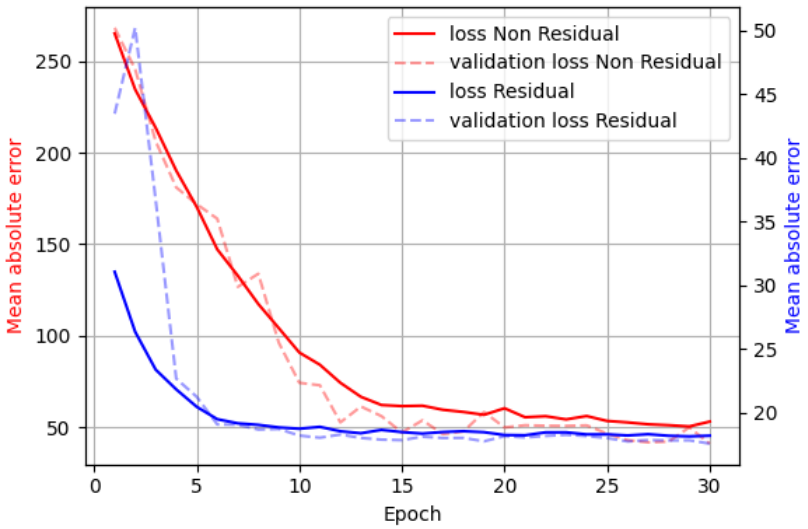

Figure 40: Training history of convolutional neural network with skip connection (residual) and without skip connection (non-residual).

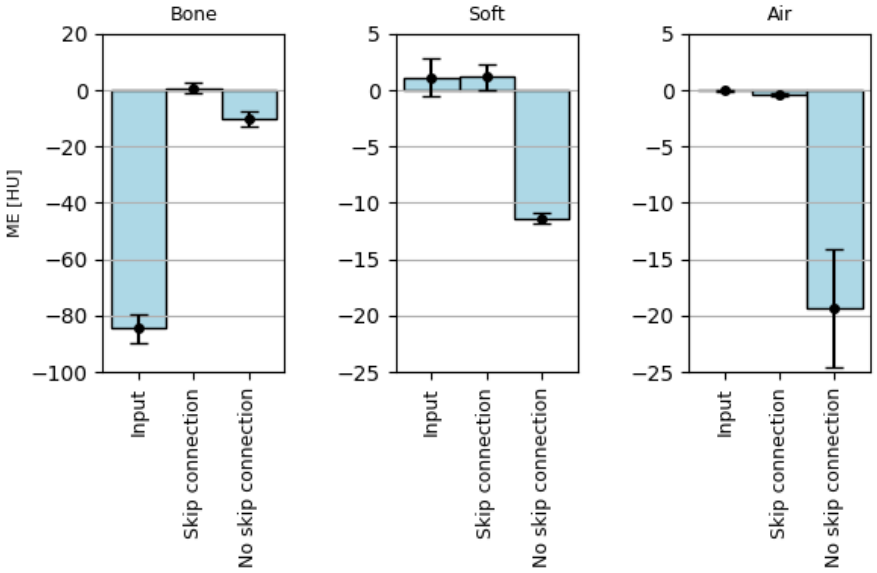

Figure 41: Mean error of segmented images between target and original data, and after conversion of the CT images using a convolutional neural network with- or without a skip connection.

Appendix C: Jilin Hospital scanner models

Table 4: CT scanner manufacturer, model and acquisition settings for every patient used during BoneMRI training.

Patient	Manufacturer	Model name	kVp	Reconstruction Kernel	Slice Thickness (mm)
1	GE MEDICAL SYSTEMS	Discovery CT750 HD	120	STANDARD	1.3
2	TOSHIBA	Aquilion ONE	120	FC 03	1.0
3	NMS	NeuViz 128	120	F20	1.0
4	NMS	NeuViz 128	120	F60	1.0
5	GE MEDICAL SYSTEMS	Revolution CT	120	BONE	1.0
6	NMS	NeuViz 128	120	F20	1.0

Table 5: CT scanner manufacturer, model and acquisition settings for every patient used during BoneMRI evaluation.

Patient	Manufacturer	Model name	kVp	Reconstruction Kernel	Slice Thickness (mm)
7	UIH	uCT 528	120	B SOFT C	1.5
8	UIH	uCT 530	120	B VSHARP B	1.5
9	TOSHIBA	Aquilion ONE	120	FC 03	1.0
10	UIH	uCT 530	140	B VSHARP C	1.5
11	UIH	uCT 530	140	B VSHARP C	1.5
12	UIH	uCT 530	140	B VSHARP C	1.0
13	UIH	uCT 530	140	B VSHARP C	1.5
14	UIH	uCT 530	140	B VSHARP C	1.5
15	GE MEDICAL SYSTEMS	Revolution CT	140	STANDARD	0.6
16	UIH	uCT 530	140	B VSHARP C	1.5
17	UIH	uCT 530	140	B VSHARP C	1.5
18	UIH	uCT 530	140	B VSHARP C	1.5
19	UIH	uCT 530	140	B VSHARP C	1.5

Appendix D: X-ray spectra

The x-ray spectra of the Siemens Somatom Definition Flash CT scanner can be seen in Figure 42. The mean energy of the 140 kV photon spectrum at the detector is 76 keV with filtering. When tin filtering is added, the mean energy of the 140 kV photon spectrum increases to 92 keV. The attenuation of bone decreases faster than the attenuation of water for increasing photon energy, which makes that the HU value of bone voxels decrease (see section 2.2.1).

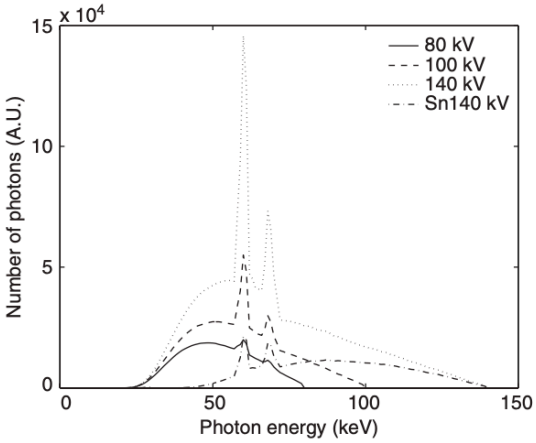

Figure 42: X-ray spectra of the Siemens Somatom Definition Flash CT scanner. Figure adopted from [35].

This means that tin filtering increases the HU difference of bone between 100- and 140 kVp scans. Histograms of different CT scanner models can be seen in Figure 43. It can be seen that the range of HU values present in the 140 kVp scans are not equal. The 140 kVp scan acquired with the Siemens Somatom Definition Flash has a maximum value of 1471 HU, while the maximum value for a CT

scan acquired with the UIH uCT 530 is equal to 2102 HU. The HU values present within a scan are thus CT scanner model dependent.

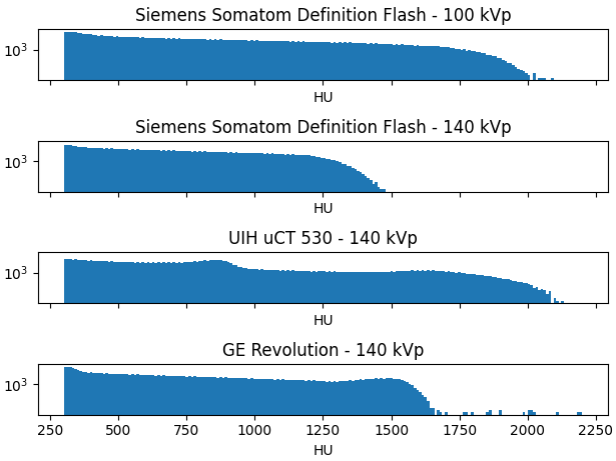

Figure 43: Histograms of CT scans acquired using different scanner models. Only HU values above 300 are shown.

Appendix E: Amount of training data

An experiment was carried out to determine the amount of data needed to train a kernel conversion CNN. The data and evaluation used were the same as in section 5. Five models were trained independently, using the data of 16, 8, 4, 2 or 1 patient(s). The models were then tested using the data of five patients. The ME and MAE were calculated for voxels containing bone. The acutance ratio was calculated between the predictions and targets. The values were compared using a Welch's *t*-test with a significance level of $\alpha = 0.05$.

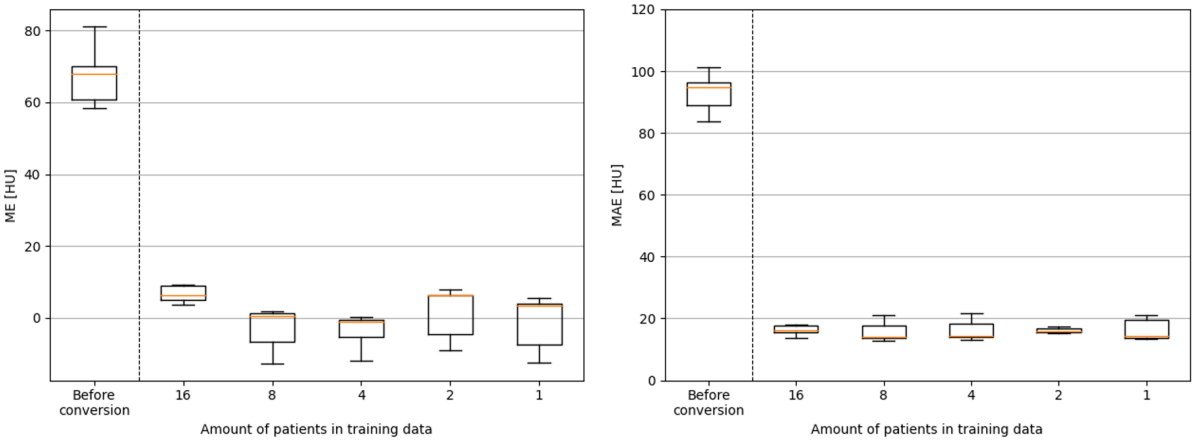

Figure 44: Boxplots of the mean error (left) and mean absolute error (right) between the prediction and target images using different amounts of training data.

The ME reduced significantly using all models when compared to the ME that was present before conversion (p<0.05). The ME decreased significantly when the amount of training patients was reduced from 16 to 8 (p=0.01). Further reduction of the amount of training data did not change the ME. The MAE did not change significantly when the amount of training data was decreased from 16 to 8 patients (p>0.05). The acutance ratio did not change by altering the amount of training data. An example of a prediction made using only one patient as training data can be seen in Figure 45.

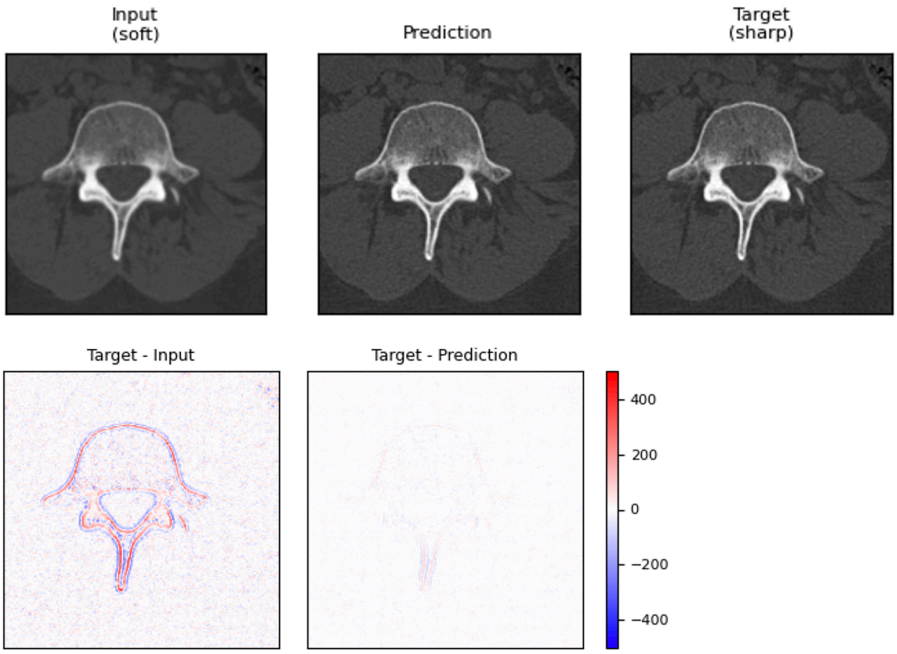

Figure 45: Input, prediction and target images of a network trained with only one patient as training data (W/L: 2000/500 HU). The difference image shows a clear improvement after the conversion (W/L: 1000/0 HU).

This experiment shows that a kernel conversion CNN can be trained using the paired data of only one patient. This makes it relatively easy to train conversions between combinations of reconstruction kernels.