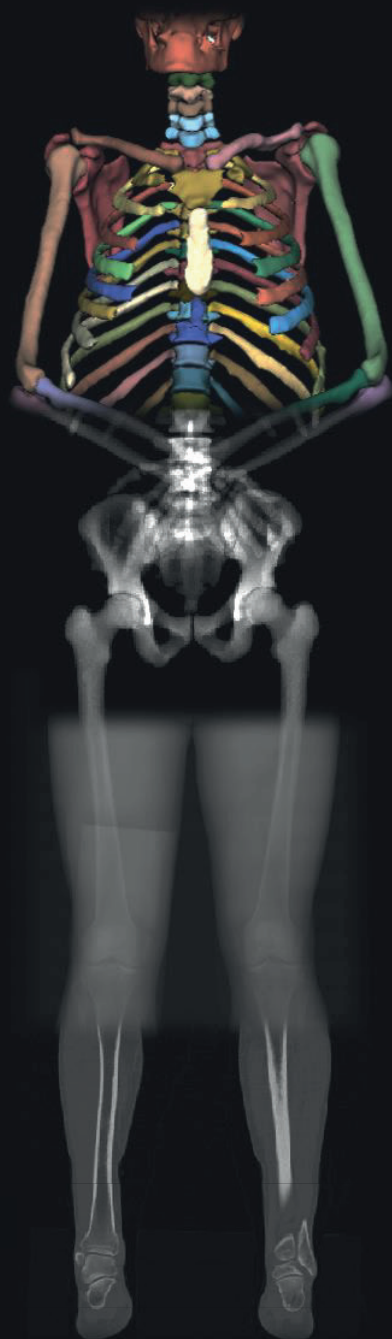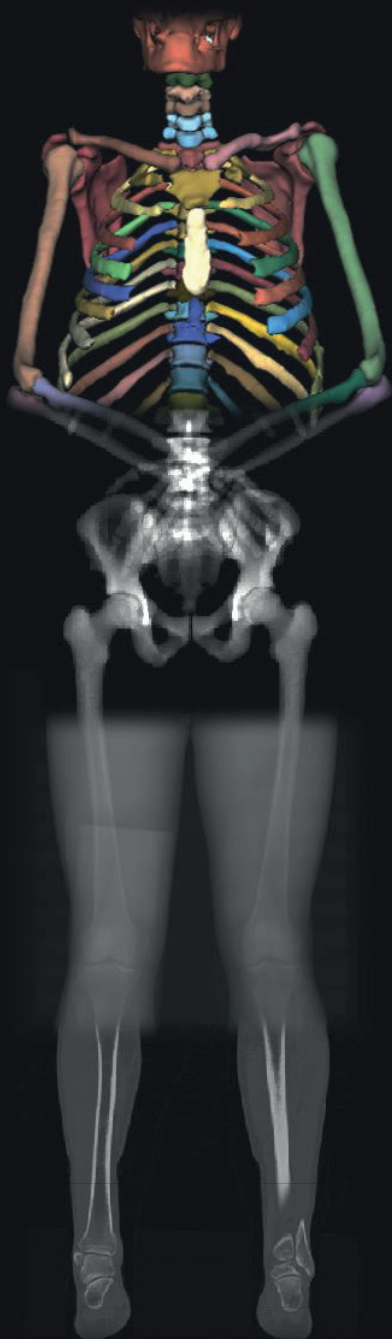# Computer-Assisted Osteotomy Planning

Ruurd Kuiper

UMC Utrecht

Computer-Assisted Osteotomy Planning

Medical Imaging Techniques for Orthopaedic Surgery

Ruurd J.A. Kuiper

Computer-Assisted Osteotomy Planning

Medical imaging techniques for orthopaedic surgery

Computerondersteunde Osteotomie Planning

Medische beeldvormende technieken voor orthopedische chirurgie

(Met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling, ingevolge het besluit van het college voor promoties in het openbaar te verdedigen op dinsdag 18 april 2023 te 12:15 uur

door

Ruurd Jan Anthonius Kuiper

geboren op 21 augustus 1992
te Eersel

Promotoren:

Prof. dr. ir. H. Weinans
Prof. dr. ir. M.A. Viergever
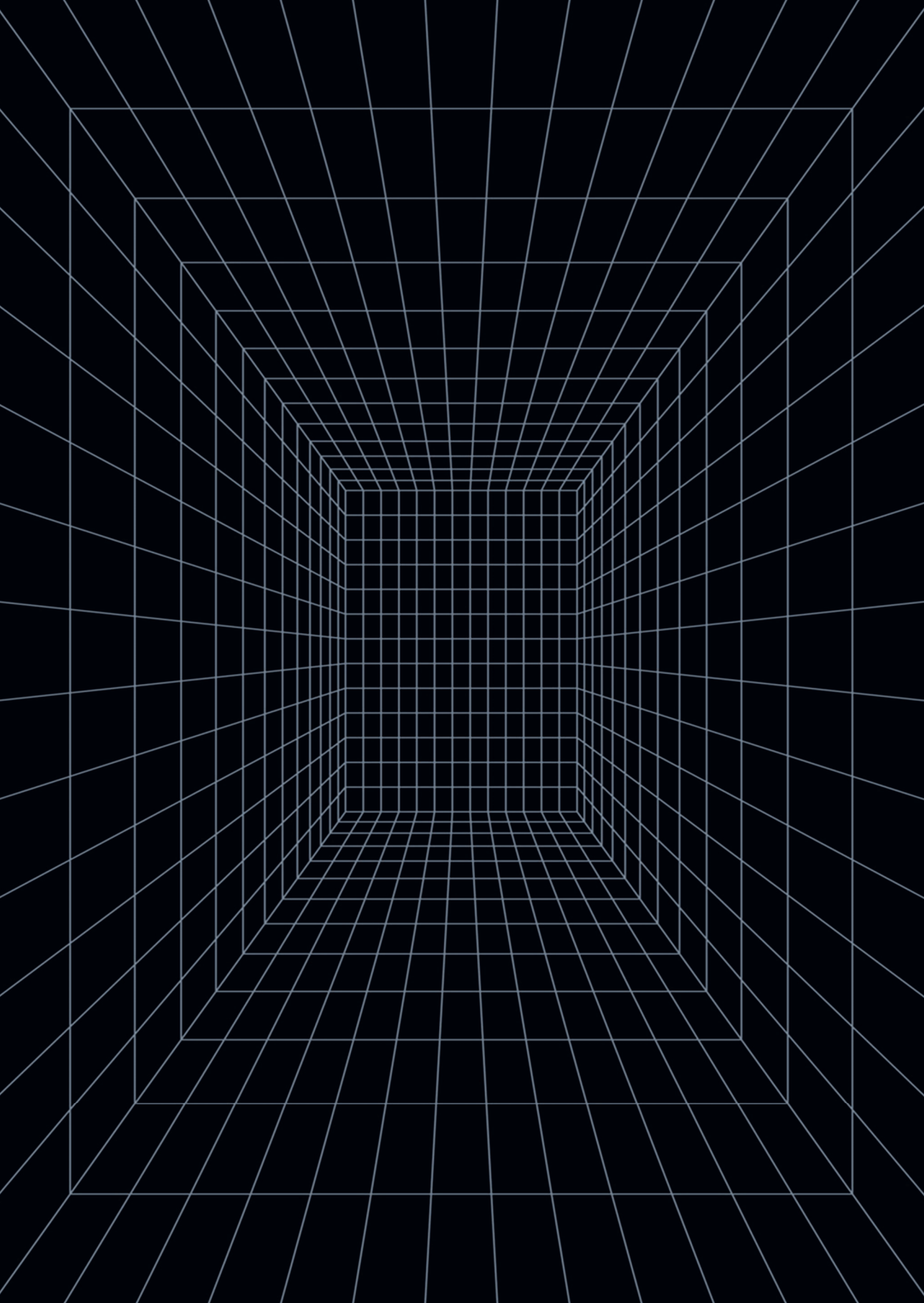
Copromotoren:

Dr. ir. P.R. Seevinck
Dr. R.J.B. Sakkers

Beoordelingscommissie:

Prof. dr. J.H. Coert
Prof. dr. J.M. van Laar (voorzitter)
Prof. dr. L. van Rhijn
Prof. dr. D. Salvatori
Prof. dr. ir. N. Verdonschot

Contents

Chapter 1

Introduction

1.1 The musculoskeletal system

The musculoskeletal (MSK) system is an organ system that consists of the bones, muscles, joints, tendons, ligaments, cartilage and other connective tissues. Its function is to allow movement, provide support and stability, and protect the rest of the organs of to the human body. Disorders that affect the musculoskeletal system reduce the effectiveness with which the MSK can provide this functionality. A 2019 study on the global burden of disease showed that approximately 1.71 billion people suffer from MSK disorders. This causes it to be the highest contributor to Years Lived with Disability (YLD), accounting for 17% of all YLDs worldwide[1]. Bone malalignment is one of these conditions that affects people worldwide, and that can be both a cause and a consequence of disorders of the MSK system.

1.2 Bone malalignment

Bone malalignment (also misalignment or malformation) occurs when the orientation or position of the joints differs from normal alignment, due to abnormal bone or cartilage morphology. It can occur in any bone in the body that interacts with the musculoskeletal system [2]. Moderate cases of bone malalignment can cause cosmetic issues or pain, for which a patient might seek treatment. In more severe cases, it can also cause deficiencies in the function of the related parts of the musculoskeletal system, ranging from reduced range of motion to complete loss of use of a limb[3]. Additionally, bone malalignment may cause degenerative disorders like osteoarthritis due to uneven loading of the articular cartilage[4].
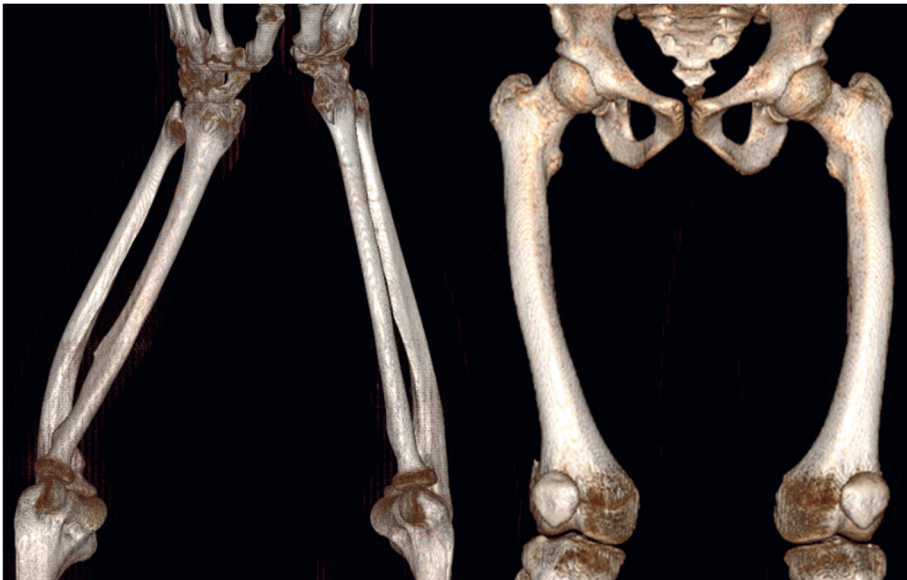
Figure 1.1: Three-dimensional rendering of bone derived from CT images, showing two examples of malaligned bones. Left: Unilateral left-sided malalignment of the radius and ulna. Right: Bilateral malalignment of the femur.

Bone malalignment can occur due to three different causes: congenital, traumatic or disease. Congenital bone malalignment is caused either by genetic defects or as a result of complications during prenatal development. Post-traumatic bone malalignment could occur either due to malunion or non-union of the bone after a broken bone, or due to damaged cartilage[5]. Different diseases can also be the source of malalignment, of which osteoarthritis is the most prevalent with over 250 million people affected worldwide[6]. Breakdown of the articular cartilage causes changes in the relative orientation of the bones inducing malalignment, which could cause further degeneration of the cartilage[4].

As bone malalignment is such a broad topic, no single source exists that summarizes the prevalence of all types of bone malalignment in all bones. However, an overview of the most prevalent causes of bone malalignment in the Netherlands could offer a general sense of the impact this disorder has on society and on the healthcare system.

In a study conducted in the northern Netherlands, congenital bone defects in the limbs occurred on average in 21.1 out of 10,000 births over 1981 to 2010. With approximately 170,000 births per year, this results in approximately 360 patients with congenital bone malalignment per year in the Netherlands[7].

In 2012, the incidence of extremity fractures in the Netherlands was 130 per 10,000 inhabitants, totalling 221,000 per year. Of these, 14% required surgical treatment[8]. An Australian study found that 8.1% of the patients with a humeral, tibial or femoral fracture that required direct surgical treatment returned within 2 years of trauma due to mal- or non-union. Combining the Dutch and Australian study results in an estimated 2500 patients with traumatic bone malalignment per year in the Netherlands.

Unicompartimental knee osteoarthritis is often treated similar to bone malalignment in patients aged under 60 years[9]. The incidence of knee osteoarthritis in the US population has been estimated to be from 0.12% for patients aged 25-34 up to 0.37% in patients aged 55-64[10]. Of these, approximately 50% is unicompartmental[11]. Combining the data from these studies with the population statistics in the Netherlands results in an approximated 8,070 new patients each year that could suffer from bone malalignment due to knee osteoarthritis in the Netherlands.

This summarizes to a total of approximately 11,000 new patients each year that could suffer from bone malalignment in the Netherlands due to the above mentioned causes. This number could be roughly extrapolated to estimate the impact that bone malalignment has on other Western societies. However, to estimate the worldwide impact of bone malalignment more data would be required, as different factors such as regional healthcare, living conditions and life expectancy might play a significant role in the prevalence of these disorders.

1.3 Diagnosis and treatment

The first step when bone malalignment is suspected in a patient is to assess the severity of the malalignment. This can be done by an external exam, but non-invasive medical imaging techniques that visualize the internal structures of the body are often used for a more accurate assessment. Different imaging modalities are available for this purpose, each with its own benefits and drawbacks. These modalities will be discussed in Section 1.4.

The choice of treatment for patients with bone malalignment depends on several factors, such as the type of malalignment, the severity, the impact on the functioning of the patient, and the availability due to cost or geographical location of the patient. In mild cases, physical therapy might be enough to alleviate the most pressing symptoms. However, due to the rigid nature of bone, this is often not enough to suppress medium to severe symptoms. Instead, the bone malalignment itself must be addressed by changing the bone morphology through surgery. This procedure is called an osteotomy and is further discussed in Section 1.5.

1.4 Medical imaging for orthopaedics

In orthopaedics, four main imaging modalities are used: Radiographs, Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and ultrasound.

Radiographs are the most often used modality for the assessment of malalignment. They are produced by emission of an X-ray (or Röntgen) beam directed through the patient. Different tissues induce different levels of absorption and scattering of the X-ray beam and the resulting intensity behind the patient can be detected using a photographic film or digital detector. Benefits of this modality are that it is fast, cost-effective, bone has a high contrast, and it can be easily taken in both standing and supine or prone position. Drawbacks are that it induces a (minor) radiation dose in the patient, and it is limited to 2D imaging. Not all malalignments are visible from these planar images alone[12].

CT is similar in methodology to radiographs, as it also uses X-ray, but in this case the beam is emitted and detected in thin slices from a multitude of directions. A computer is then used to reconstruct a 3D image of the patient. CT scans are used when the malalignment is more complex, and the assessment is not easily feasible from 2D images. Benefits of this modality are similar to the radiographs in that they have are still relatively fast, bone has a high contrast and additionally the images are in 3D. Additionally, it is quantitative by nature, which means that intra- and interpatient values can be directly compared. However, drawbacks include a larger radiation dose, and that scanners that allow for a standing patient position are rare[13].

Lower arm   **Knee**
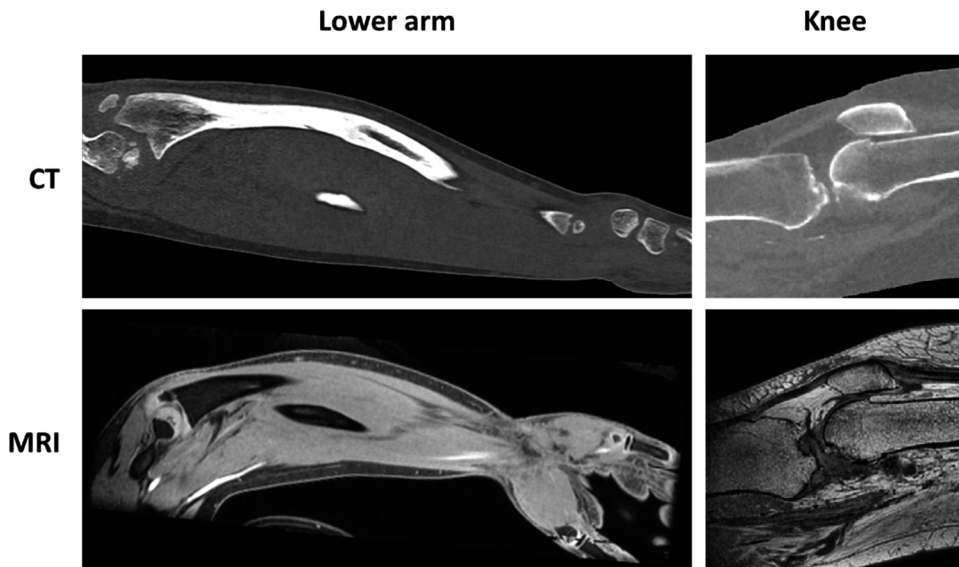
CT

MRI

Figure 1.2: Comparison of CT and MRI for the imaging of the lower arm and knee. Each image shows a single slice of a larger 3D scanned volume. Note that the MRI scans have different contrasts.

MRI uses a combination of strong magnetic fields and radio waves to induce evolving nuclear spin polarization that can be detected using radio frequency coils. Many different MRI sequences are possible by changing the parameters with which the scans are produced, which translate to different contrasts highlighting structures such as the muscles, fat, tendons or bone[14]. Benefits of this modality include its ability to produce 3D images without a radiation dose, and its ability to differentiate between soft tissues with varying proton densities. Drawbacks are that scan times are often longer than for CT, and that are currently no contrasts available that offer a unique intensity level to bone. This makes it harder to differentiate bone from the surrounding tissue. Additionally, metallic implants in the patient might make scanning unsafe or the results unclear[15].

Ultrasound images are formed by a device that emits soundwaves and measures the intensity and time delay of the echoes produced by different tissues in the body. It is an inexpensive method of imaging that can acquire images in real time, while not using harmful radiation. However, in orthopaedics it is mostly used as a diagnostic tool, as it is difficult to reconstruct full 2D or 3D bone representations from ultrasound images. This is due to its poor field of view and penetration depth, and its inability to penetrate bone. It is therefore less well suited for pre-operative surgical planning[16].

In this thesis, we have focused on CT and MRI, as these were the only modalities that enable 3D planning. Examples of CT and MR images of the lower arm and knee are shown in Figure 1.2. In Chapter 2 and Chapter 4 we further discuss and investigate the differences between these two modalities when applied to orthopaedic planning.

Figure 1.3: General workflow for automated osteotomy planning.

1.5 Osteotomy

Osteotomy (from the greek words ὀστέον, "bone", and τέμνω, "I cut") is defined as: "The surgical cutting or dividing of bone, usually to correct a deformity"[17]. This treatment option for bone malalignment is used when conservative treatment has been shown to be insufficient to alleviate the symptoms experienced by the patient. During the procedure, the surgeon first makes a cut through the skin to expose the deformed bone. Next, the location and orientation of the osteotomy are indicated, usually using guide wires or more recently by using 3D-printed patient-specific cutting guides. The surgeon then cuts the bone at the desired location(s), and if applicable removes part of the bone.

Because osteotomies are complex surgical procedures, pre-operative planning is often required. The first step is to define the deformity by acquiring two orthogonal planar radiographs or one CT scan of the malaligned bone. Images are also acquired of the healthy contralateral counterpart, if it is available. Conventionally, the osteotomy is then planned by overlay drafting, where the images of the bone are divided into fragments and overlaid by hand until the desired bone configuration is achieved[18].However, in complex cases the malalignment cannot be adequately characterized by planar imaging and planning. Therefore, orthopaedists have recently turned to three-dimensional (3D) planning based on CT or MR imaging.

To accurately plan an osteotomy for malalignment correction in three dimensions (3D), it is primarily important to know what the correct alignment of the bone should be. There are therefore two main ways to plan an osteotomy in 3D, depending on which information is available. In case the malalignment is unilateral, the healthy contralateral counterpart can be used as an example, which is often the case if the malalignment is caused by trauma. In case of congenital malalignment or malalignment caused by a medical condition (e.g. osteoarthritis), the malalignment is often bilateral. The correct alignment must then be estimated by comparing the bone morphology with that of the healthy population. This can be done based either on measurements such as length and certain angles between landmarks[2], or based on mean bone models constructed from larger population samples[19].

Pre-operative planning of osteotomies in 3D is a complex computational procedure. Conventionally, multiple manual steps are required that are time consuming and prone to inter- and intrapersonal variation[20]. In this thesis, we have developed methods to automate

the steps of this process, shown in Figure 1.3, to enable a fully automatic workflow for the quantification of bone malalignment and reconstruction of the osteotomized bone.

1.6 Image processing techniques

To assess bone malalignment and plan surgical procedures several different image processing techniques are employed and studied in this thesis. A short overview of these techniques is given in the following sub-sections.

1.6.1 Segmentation

Image segmentation is the process of delineating certain regions in a 2D or 3D image based on a common characteristic of those regions. Common characteristics in medical imaging could be the colour or intensity of the pixels or voxels (3D pixels). More clinically meaningful characteristics include the delineation of certain types of tissue such as bones, cartilage, muscles, lesions or different organs[21]. In each of the studies in this thesis, image segmentation was a critical first step of the process, on which the accuracy all subsequent steps depended. Considerable time and effort was therefore spend to ensure accurate manual segmentation, and also towards the development of reliable automated solutions to reduce the time required in the future (Chapter 2).

1.6.2 Registration

Registration is used to transform separate sets of spatial data with corresponding features into one coordinate system. Applications include matching 2D or 3D images, but also points, surfaces or volumes with each other such that the corresponding information overlaps. The transformation that is applied to the data can range from rigid transformations where only rotation and translation of the data is possible, to non-rigid (or deformable) transformations where the transformation of each data point can be different. In this thesis, we have used registration to overlay two imaging modalities (CT and MRI, Chapter 1), and to match deformed and healthy bones by their segmentation surfaces (Chapter 3).

1.6.3 Landmark indication

Landmark indication is used to find distinct points in a medical image or on a segmented surface that fit a certain predefined description. The indicated landmarks can aid in image registration, by introducing known corresponding points between different datasets. Alternatively, they can be used to measure distances, angles and morphology parameters. These parameters can aid in the diagnosis of for example hip dysplasia (Chapter 2) or lower limb malalignment (Chapter 4).

1.7 Thesis outline

In recent years, the use of computer-assisted methods for orthopaedics has seen a sharp increase in both medical research and practice[22]. Diagnosis and pre-operative planning of

disorders of the musculoskeletal system could benefit from computer-assisted methods, by increasing the accuracy and efficiency, and by reducing interoperator variation. However, adoption in the clinic is not yet widespread, with less than 5% of orthopaedic surgeons in Asia, Europe and the USA routinely using computer-assisted technology[23]. The diagnosis of bone malalignment and pre-operative planning of osteotomies are applications particularly well suited for the use of computer-assisted methods, as digital representations of the patient are often readily available in the form of medical scans.

The main goal of this thesis is to develop automatic methods for the pre-operative diagnosis and planning of osteotomies. To this end, we investigate and develop a wide range of techniques, including image registration, segmentation, 3D modelling and optimization. The application of these tools in clinical practice could ultimately aid in the diagnosis, assessment, and pre-operative planning of orthopaedic surgery.

Chapter 2 investigates a novel method that improves the registration of CT and MR images of the knee joint, where complex deformations exist between the images due to different patient positioning. The outcome is then compared to the outcome of conventional methods for registration. These registered CT and MR images could be used in the orthopaedic practice to combine the information of the two modalities into one reference frame.

Chapter 3 describes the development and optimization of a deep learning based approach to bone segmentation in the lower extremities. Automated bone segmentation could make 3D orthopaedic planning faster, less operator dependent, and more accessible.

Chapter 4 investigates the effect of using different imaging modalities, and the inclusion of bone and/or cartilage in the segmentation, on the outcome of lower arm osteotomy planning.

Chapter 5 introduces a method for the automatic assessment of lower limb alignment from CT scans, using the segmentations from Chapter 4 and an automatic method for landmark indication using non-rigid point cloud registration.

Chapter 6 summarizes and discusses the findings of the studies in the previous chapters, and gives an outlook towards the further research necessary for implementation of these methods in the clinic.
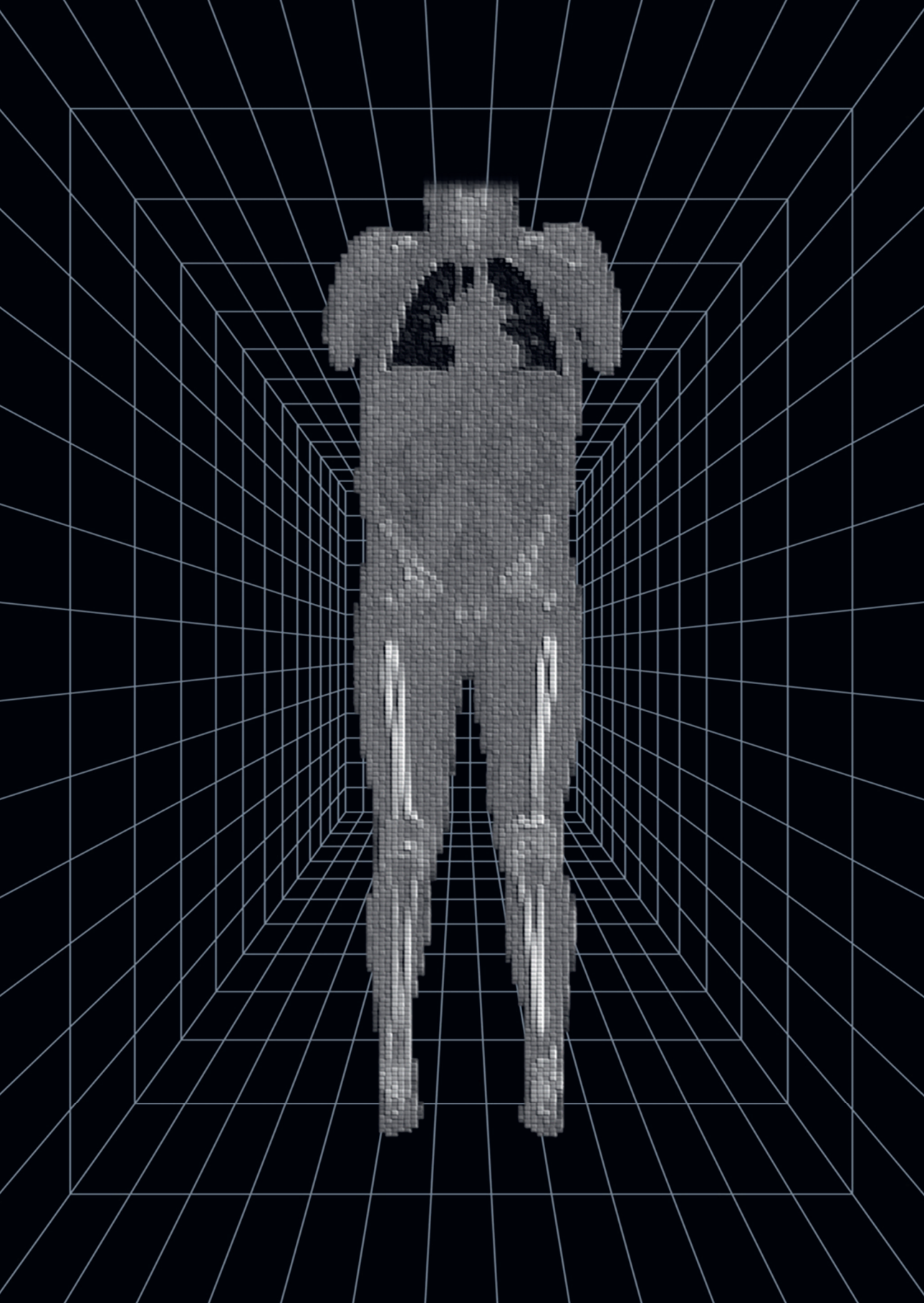
Chapter 2

CT to MR registration of complex deformations in the knee joint through dual quaternion interpolation of rigid transforms

Based on:

Abstract

Purpose: To develop a method that enables CT to MR image registration of complex deformations typically encountered in rotating joints such as the knee joint.

Methods: We propose a workflow, denoted Quaternion Interpolated Registration (QIR), consisting of three steps, which makes use of prior knowledge of tissue properties to initialise deformable registration. In the first step, the rigid skeletal components were individually registered. Next, the deformation of soft tissue was estimated using a dual quaternion-based interpolation method. In the final step, the registration was fine-tuned with a rigidity-constrained deformable registration step. The method was applied to paired, unregistered CT and MR images of the knee of 92 patients. It was compared to registration using B-Splines (BS) and B-Splines with a Rigidity Penalty (BSRP). Registration accuracy was evaluated using Mutual Information (MI), and by calculating Dice Similarity Coefficient (DSC), Mean Absolute Surface Distance (MASD) and 95th percentile Hausdorff Distance (HD95) on bone, and DSC on water and fat dominated tissue. To evaluate the rigidity of bone in the registration, the Jacobian Determinant (JD) was calculated.

Results: QIR achieved improved results with 0.86, 0.76 mm and 1.88 mm on the DSC, MASD and HD95 metrics on bone, compared to 0.84, 1.40 mm and 4.99 mm for the BS method and 0.84, 1.40 mm and 3.56 mm for the BSRP method. The average DSC of water and fat was 0.78 and 0.86 for the QIR, 0.76 and 0.84 for both BS and BSRP. Comparison of the median JD and median interquartile (IQR) ranges of the JD indicated that the QIR (1.00 median, 0.03 IQR) resulted in higher rigidity in the rigid skeletal tissues compared to the BS (0.98 median, 0.19 IQR) and BSRP (1.00 median, 0.05 IQR) methods.

Conclusion: This chapter showed that QIR could improve the outcome of complex registration problems, encountered in joints involving rigid and non-rigid bodies such as occur in the knee, as compared to a conventional registration approach.

2.1 Introduction

Diagnosis, pre-operative planning and follow up of musculoskeletal conditions heavily depend on three-dimensional (3D) imaging techniques. Computed Tomography (CT) imaging provides an accurate representation of the location and nature of deformities or lesions in the bone. In contrast, Magnetic Resonance Imaging (MRI) aids in the detection of abnormalities in the soft tissues including muscle, ligaments, nerves and cartilage[24]. The complementary information from both modalities can aid an orthopaedic surgeon during diagnosis, surgical planning or treatment monitoring. However, images of different modalities are generally not obtained with the exact same patient positioning. To combine the information of both modalities in the same reference frame, image registration is crucial.

Image registration is a generic term for all techniques that spatially transform an image such that it aligns corresponding features of the original with a target image. The principle of image registration rests on finding the optimal transform that minimizes some cost function that defines the correspondence between two images[25,26]. To find the transform that minimizes this cost function, an iterative optimization algorithm is often applied[27–29]. More recently deep learning techniques have also been employed for image registration problems[30–33], and have shown great promise in terms of computational efficiency.

In case of MRI and CT imaging of the skeletal components that we encounter in scans for orthopaedic purposes, registration can be challenging due to several reasons. Different patient positioning between scans due to rotation, flexion or extension of the joints can lead to strong local deformations in soft tissue between bone structures, whereas only rigid transformations occur between one bone structure with respect to another bone structure. These contrasting features of the transformations occurring in different tissues cause non-smooth changes in the deformation field, which are difficult to represent using currently available deformable registration techniques[34,35]. Also, rotations in the joint cause large initial misalignments
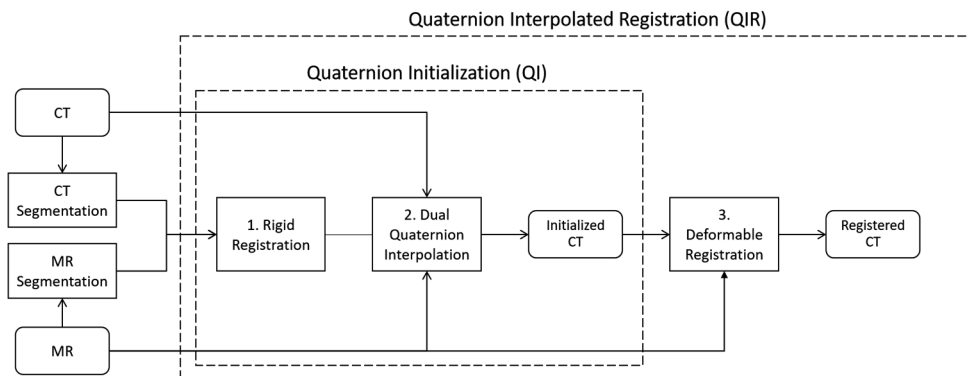
Figure 2.1: Flowchart of the Quaternion Initialization (QI) and Quaternion Interpolated Registration (QIR). QI consists of registration of the segmented rigid structures (step 1) followed by soft tissue interpolation (step 2). The output of the QI is used as input for the deformable registration (step 3).

between scans that cannot be solved with a single initial rigid registration, which makes it more difficult to match corresponding regions in subsequent steps[36]. Finally, the unique properties of the respective imaging modalities cause a nonlinear relation when comparing gray values in a voxelwise manner, which limits the availability of suitable cost functions[37] Due to these complexities, commonly used deformable registration techniques often fail to find the correct correspondence between images.

Various studies have tried to improve the registration outcome by improving initial alignment before applying the conventional deformable registration algorithms. Little et al.[38] were the
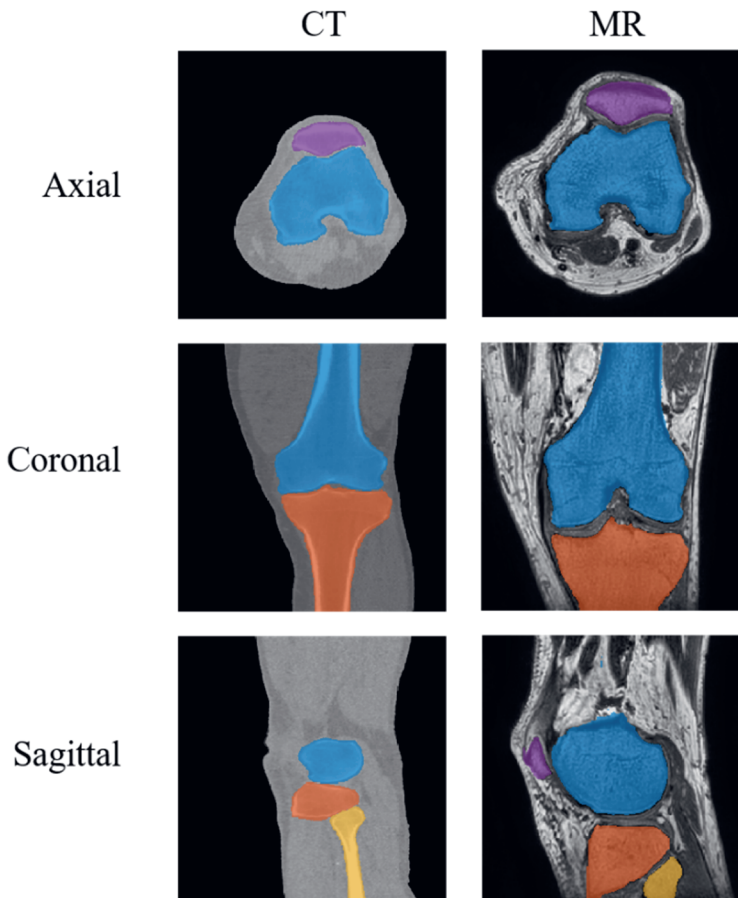
Figure 2.2: Example of the segmentations of the femur (blue), tibia (orange), patella (purple) and fibula (yellow) on CT and MR of same patient. Segmentation was performed by a combination of automatic initialization and manual refinement.
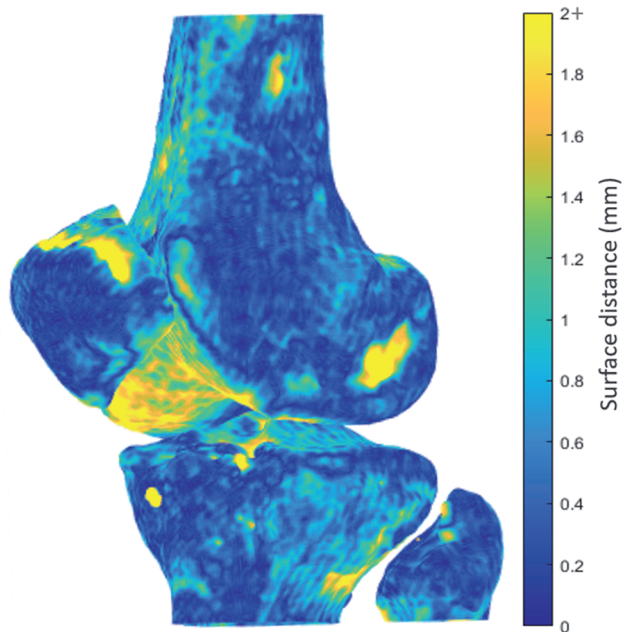
Figure 2.3: 3D rendering of bone segmentation performed on an MR image with colours indicating the point-to-surface distance to the CT bone segmentation after rigid registration using ICP. Surface distance is highest where segmentation is difficult on MR, e.g. bone-cartilage interfaces and tendon insertion sites.

first to make use of separate registration of rigid structures to aid the deformable registration of the deformed tissue. However, they only applied this to two dimensional images of the same modality. Zhang et al.[39] used segmentations of bones on CT and MR to guide their registration, but only applied these to acquire rigid transformations. Lu et al.[40] applied a similar approach for non-rigid registration of monomodal thoracic and abdominal images. They used an iterative thin plate spline model to interpolate between the rigid transformations. Walimbe and Shekhar[41] divided CT and PET images into equal sized cubes, which were registered rigidly using a multilevel approach. In order to acquire a smooth deformation field, these rigid transformations were then interpolated using dual quaternions. However, this approach did not apply any rigidity constraints to the bone and could therefore lead to unrealistic deformations. Foruzan and Motlagh[42] used Thin Plate Splines for Robust Point Matching (TPS-RPM) to align MRI and CT scans of the liver, which incorporates both rigid and non-rigid methods..

We introduce a workflow (Figure 2.1) that makes use of the known rigid properties of bone to improve the initial alignment of the images, before performing intensity-based deformable registration. The accuracy of this workflow was assessed on a dataset of paired CT and MR images of the knee. The results of the registrations were compared with a conventional workflow using the state-of-the-art registration toolbox *elastix*[27], based on voxel wise and structural correspondence measures.

2.2 Methods

2.2.1   Data

A dataset consisting of CT and MR images of 92 subjects was used. MR images were acquired in a 3T scanner using a T1-weighted dual gradient echo sequence (Ingenia, Philips Healthcare, Best, The Netherlands). Dixon reconstruction (mDIXON FFE) was performed on the scanner to obtain in-phase (IP), opposed phase (OP), fat only (F), and water only (W) images. [43,44]. The acquisition time was 3 min 21 s with TR = 7 ms, TE = 1.186 ms/2.372 ms , bandwidth = 1072 Hz/pixel, flip angle = 10°, acquired voxelsize = 0.46x0.46x0.5 mm^3. The CT images were acquired with 0.67 to 0.8 mm slice thickness and pixel spacing ranging from 0.78 to 0.87 mm in plane (IQon - Spectral CT, Philips Healthcare, Best, The Netherlands). The tube current was 115 mA with 120 kV tube voltage. The CT images were resampled to the same resolution as the MR images using tricubic interpolation. CT and MR scans were acquired on the same day, within a 4-5 hour timeframe.

All images were acquired in compliance to the regulations from the medical ethical committee, and the study is registered under www.clinicaltrials.gov number: NCT03883568.

2.2.2   Segmentation

On each CT and MR image four different bones were present: the femur, tibia, fibula and patella. The proposed QIR method requires separate segmentations of these bones. For 12 subjects, the bones were manually segmented. To reduce the time necessary for segmentation of all subjects, the twelve manual segmentations were used to train two 3D U-net neural networks for semantic bone segmentation[45,46]. After training the network was used to automatically segment the CT and MR images, and the resulting segmentations were manually checked and corrected using the open source software 3DSlicer[47]. Figure 2.2 shows typical examples of the segmentations of the bones of interest performed on CT and MR images. A detailed description of the bone segmentations can be found in Appendix A.

In order to evaluate the non-rigid component of the registration, soft tissue was roughly segmented into two components on both CT and MR. The first component was tissue that was high in water content, such as muscle tissue. The second component is the tissue with a high fat content, which in the knee joint consist mostly subcutaneous and inter- and intramuscular fat. A description of the soft tissue segmentation can be found in Appendix B. The two soft tissue components are further referred to as the water and fat segmentations.

2.2.3   Workflow

The flowchart in Figure 2.1 summarizes the registration workflow. It consists of three parts:

1. Rigid-body registration of the bones, preserving its rigidity,

2. Dual quaternion-based interpolation of the transformations in between the rigid bones, named Quaternion interpolation (QI),
3. Deformable registration of the soft tissue.

Alltogether, we reference to this workflow as Quaternion interpolated registration (QIR). In the following paragraphs we describe each of the three steps of the workflow in more detail.

2.2.4 Rigid-Body Registration

In the first step, the segmentations of the bones on the CT and MR images were used to perform rigid-body registration. The voxel representations of the segmentations were converted into 3D surfaces by using the marching cubes algorithm[48]. From these surfaces the vertices were extracted to be used for point-cloud registration. For each bone separately, the point cloud from the bone segmented from CT was registered to the point cloud from the bone segmented from MR using a multi-resolution implementation of the Iterative Closest Point (ICP) algorithm[49]. The ICP algorithm was performed three times for each bone, starting out with a sparse version and ending with a denser version of the point clouds, using 1000, 5000 and 20000 points respectively. The sparse point-clouds were acquired by mesh simplification using the built-in Matlab function *reducepatch*. The multi-resolution approach prevented the algorithm from becoming trapped in a local minimum solution. This method resulted in a separate rigid transformation for each bone, described in the transformation matrix T.

2.2.5 Dual Quaternion Interpolation between the Rigid Transformations

The second step of the workflow estimated the deformation of the soft tissue by interpolating the rigid transformations that were found. The displacement of each voxel was calculated as a weighted average of the rigid transformations using the inverse distance weighting scheme derived by Shepard[50], as shown below .

$$w_i(x) = \frac{1}{d(x, b_i)^p}$$

Equation 1

For each voxel, $w_i(\boldsymbol{x})$ defined the weight that the rigid transform of each bone i had on the transform of that voxel. $w_i(\boldsymbol{x})$ was defined by the Euclidean distance $d(\boldsymbol{x}, \boldsymbol{b}_i)$, which was the distance between the location of the voxel ($\boldsymbol{x}$) to the location of the nearest voxel $\boldsymbol{b}_i$ on the bone i. The power parameter p determined the influence of proximity on the weighting. In this study, $p = 4$ was used in order to ensure that the transform of a voxel is dominated by the bone that is closest to it.

The rigid transformations of the bones calculated using the ICP algorithm were represented as a 4 x 4 transformation matrix. However, direct interpolation of rotation matrices eliminates the orthogonality and thus the rigid-body characteristic of the transformation[41]. To overcome this issue the rotational transformations must first be rewritten to their quaternion form,

originally invented in 1844 by Sir William Rowan Hamilton[51]. Shoemake introduced quaternion interpolation in computer graphics[52]. In order to interpolate for both rotation and translation, this concept was extended further to dual quaternions, using the Dual Quaternion Linear Blending (DLB) implementation as described by Kavan et al.[53]. First, the transformation matrix T was rewritten into dual quaternion form q. Given n rigid transformations, DLB linearly interpolates the dual quaternions $(q_1, \ldots, q_n)$ using the weights $(w_1, \ldots, w_n)$ as shown in Equation 2.

$$DLB(w_1, \ldots, w_n; q_1, \ldots, q_n) = \frac{w_1 q_1 + \cdots + w_n q_n}{\|w_1 q_1 + \cdots + w_n q_n\|}$$

Equation 2

Using this method, a forward deformation field was found. Conventionally, the inverse deformation field would be used to register the moving image onto the fixed image. This is done because it guarantees that for each voxel in the fixed image reference frame, a corresponding voxel can be found in the moving image, while the use of a forward deformation field does not guarantee this. To overcome this issue, the vectors of the forward deformation field were negated and then transformed onto the fixed image reference frame using the forward deformation field itself, similar to the "Inversion Using Direct Interpolation" method described by Crum et al.[54]. As the negated vectors were not guaranteed to correspond to discrete voxel locations and some voxels in the fixed image reference frame might not be assigned a vector, Delaunay triangulation of the scattered coordinates was used to find a linear interpolation to fill the grid of the fixed reference frame[55].

2.2.6   Deformable Registration

For the third step we used an intensity-based registration[27] with a rigidity penalty (Staring *et al* 2007), preserving the rigidity of the bone specifically using the bone masks. Registration optimized normalized mutual information through a B-spline transformation model. This was implemented using *elastix* employing a multi-resolution approach and an adaptive stochastic gradient descent optimizer.

This final step corrected for the remaining soft tissue misalignments between MRI and CT after the first two steps. This was necessary because during the initialization it was assumed that only the movement of the rigid structures affects the deformation of the soft tissue. In reality, forces like muscle contraction and extension, compression and gravity influence the deformation of the soft tissue, especially further away from the rigid structures.

2.3 Experimental Setup

The QIR and QI workflows as shown in Figure 2.1 were compared to registration using only intensity-based registration with deformable B-splines. Two different workflows were used for comparison, the B-Spline (BS) and the B-Spline with Rigidity Penalty (BSRP) method.

Both of these workflows were initialized with a single rigid registration. For all deformable registrations Normalized Mutual Information (NMI) was used as the optimization metric[37,56]. The parameters used in *elastix* were optimized on a small subset of the patients, and are available as supplementary material. For both the QIR and BSRP workflow, the rigidity penalty was only applied on the bone segmentations.

2.3.1 Evaluation

The results were evaluated using five different measurements, of which the first three were based on the segmentations of the bone. The Dice Similarity Coefficient (DSC) measures the overlap of the segmentations as:

$$\text{DSC} = \frac{2|\mathbf{CT} \cap \mathbf{MR}|}{|\mathbf{CT}| + |\mathbf{MR}|}$$

Equation 3

Where **CT** and **MR** denote the segmentations of bone on CT and MR, respectively. The Mean Absolute Surface Distance (MASD) is defined as the average of the minimal absolute distance of points on the surface of the CT to the MR segmentation, and vice versa. This can be derived by calculating the minimum surface distance for each point *m* on the surface *M* of the **MR** segmentation to each point *c* on the surface *C* of the **CT** segmentation as follows:

$$D_C(m, c) = min_{c \in C} \|m - c\|$$

$$D_M(m, c) = min_{m \in M} \|c - m\|$$

Equation 4

With ‖·‖ the Euclidean norm. The MASD is then calculated as:

$$\text{MASD} = \frac{1}{|C| + |M|} \left(\sum_{c \in C} D_M(c) + \sum_{m \in M} D_C(m) \right)$$

Equation 5

The (symmetric) Hausdorff Distance (HD) is defined as the largest of the two maxima of $D_M(c)$ and $D_C(m)$, calculated as:

$$\text{HD} = max \left(max_{c \in C} (D_M(c)), max_{m \in M} (D_C(m)) \right)$$

Equation 6

In this study we used the 95th percentile Hausdorff Distance (HD95) to avoid the sensitivity HD has to small segmentation errors and better represent the gross registration errors we wanted to measure. The Normalized Mutual Information (NMI) was defined as:

$$\text{NMI}(\text{CT}, \text{MR}) = \frac{MI(\text{CT}, \text{MR})}{\sqrt{H(\text{CT})H(\text{MR})}}$$

Equation 7

Where NMI is the Normalized Mutual Information, and H denotes the entropy. It was calculated from the registered CT image to the MR image using 64 bins.

Bone was expected to stay rigid between deformations, and thus no volumetric changes should have occurred within these voxels during registration. By calculating the Jacobian Determinant (JD) we could measure the volumetric change in each voxel. The JD for a voxel at location x, y, z with deformation t_x, t_y, t_z was calculated as:

$$\text{JD}(x, y, z) = \begin{vmatrix} \dfrac{\partial t_x}{\partial x} & \dfrac{\partial t_x}{\partial y} & \dfrac{\partial t_x}{\partial z} \\ \dfrac{\partial t_y}{\partial x} & \dfrac{\partial t_y}{\partial y} & \dfrac{\partial t_y}{\partial z} \\ \dfrac{\partial t_z}{\partial x} & \dfrac{\partial t_z}{\partial y} & \dfrac{\partial t_z}{\partial z} \end{vmatrix}$$

Equation 8

A JD between 0 and 1 indicates shrinking, whereas a JD larger than 1 indicates expansion. A JD smaller than 0 indicates a physically impossible folding within the registered image, which should not occur.

The JD was calculated separately for bone and soft tissue voxels. The bone masks for each method were derived by transforming the original bone mask with the respective transformation found by each method. The masks for the soft tissue were found by thresholding the registered image using > -900 HU, after which the bone was excluded by excluding the bone mask.

2.4 Results

2.4.1   Rigid-body Registration

In Figure 2.3 a typical example of the surface distance between the CT and the MR segmentation after rigid registration using ICP is shown. Qualitative inspection of more samples indicated that larger surface distances often occurred at locations where the interface between bone and other tissues was more difficult to discern on MR, such as cartilaginous regions or tendon insertion sites. This resulted in surface distances >2mm during initialisation.

2.4.2   Qualitative Analysis

Figure 2.4 shows an example of the registrations results of one knee for all different workflows in the three orthogonal planes. Red arrows highlight points where registration errors occurred. It can be seen that after performing QI the bones were well-aligned, but no soft tissue registration had occurred. The dual quaternion interpolation did ensure that the soft tissue deformed smoothly with the bone. In the BS, most parts of the image were registered well to the MR. However, in the locations indicated by the red arrows, elastic deformation of the bone is visible. Inspection of these same locations in the registration results of the QIR showed that these elastic deformations did not occur with QIR. The BSRP registration in Figure 2.4 shows that the bone structures on CT were not well aligned to the bone structures in the MR. This was caused by the rigidity penalty, which inhibited high local deformations in the joint space. In the bottom row, these registration errors are also easily discernible from the differences between the bone, water and fat segmentations of the registered CTs and MR. An example of the segmentation differences in all three orthogonal planes is also shown in Appendix B. For this patient, only the QIR approach was able to reproduce both the rigid transformation of bone and the high local deformations in the joint space that naturally occur during bending of the knee.
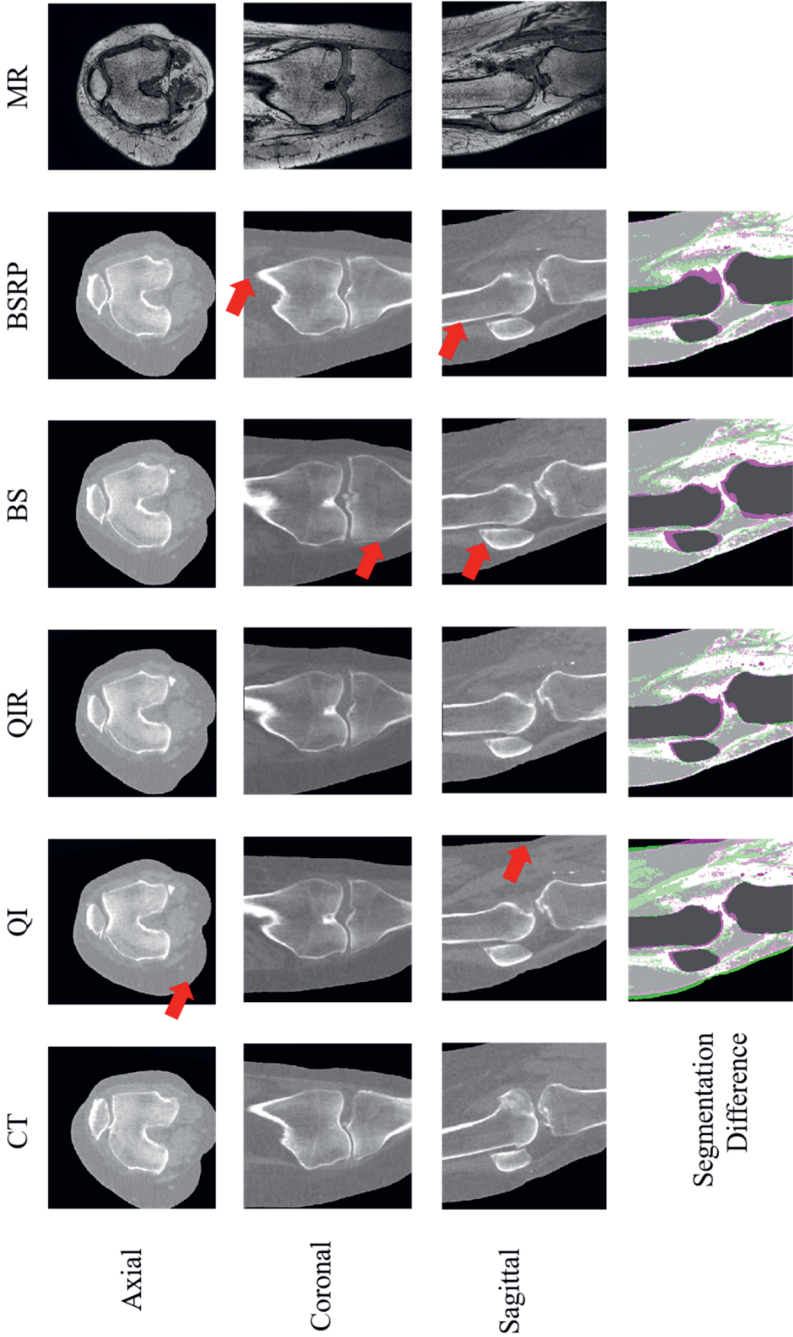
Figure 2.4: This grid shows the results of the registration using the different workflows, of one patient in three different planes. On the left the original CT is shown that needed to be registered to the MR, which is shown on the far right. Note that the original CT shown was rigidly registered to the MR using elastix. Red arrows indicate some of the locations where the registration has failed. The bottom row shows the difference in green (soft tissue difference) and pink (bone differences), between the different registered CTs and the MR based on the water, fat and bone segmentations in the sagittal plane.

2.4.3 Quantitative Analysis

In Table 2.1 the results of the registration using the different workflows are shown. The values found for the DSC, MASD, HD95 and JD metrics were not normally distributed for most of the metrics, nor were the paired differences between the methods symmetric around the median. Therefore, the *two sample paired sign test* was used to determine significant differences. Bonferroni correction was applied to correct the $p < .05$ significance threshold because multiple comparisons were done. This meant that $p < 0.0042$ was required for the results to be considered significant. In addition to this test, we also calculated the number of cases out of 92 total, in which the metrics scored better with the QIR than both the BS and BSRP methods. This was the case for all average DSC, MASD and HD95 of bone tissue. For the water and fat the DSC was highest in 78 and 65 patients respectively when using QIR.

QI achieved good results for the bone segmentation-based metrics. However, it should be noted that this method does not include a deformable registration step, and thus the soft tissue was interpolated, but not registered. This was reflected by the lower NMI and soft tissue DSC of the QI. QIR preserved bone rigidity well, while also improving the soft tissue registration, expressed by the higher MI and soft tissue DSC. Comparing the results of the complete workflows (i.e. QIR, BS and BSRP), the results in Table 2.1 showed that the QIR significantly outperformed the BS and BSRP on all metrics.

Figure 2.5 visualizes the bone segmentation and NMI metrics as boxplots to show the distribution of the obtained metrics over the patients. The distributions show that for the BS and the BSRP there were more outliers, indicative of large misregistrations of the bone segmentations. In Figure 2.6, boxplots of the DSC of water and fat tissue are shown. It was clear that water and fat are not yet well-registered after the QI. Although the improvement in water and fat registration with the QIR was not as great as the improvement in bone registration, it was still significant as is shown by p-values in Table 2.1.

2.4.4 Rigidity Evaluation

The images on the top row of Figure 2.7 show an example of the JD overlaid on the registered CT scans for each workflow. As can be seen in the QI and QIR image, strong local deformations occurred in the joint spaces, while no deformation is visible in the bone. In the BS image, it can be seen that non-rigid deformation occurred in both the tibia and the femur. The BSRP image does not show much deformation, except for a small part further away from the joint spaces and bone.
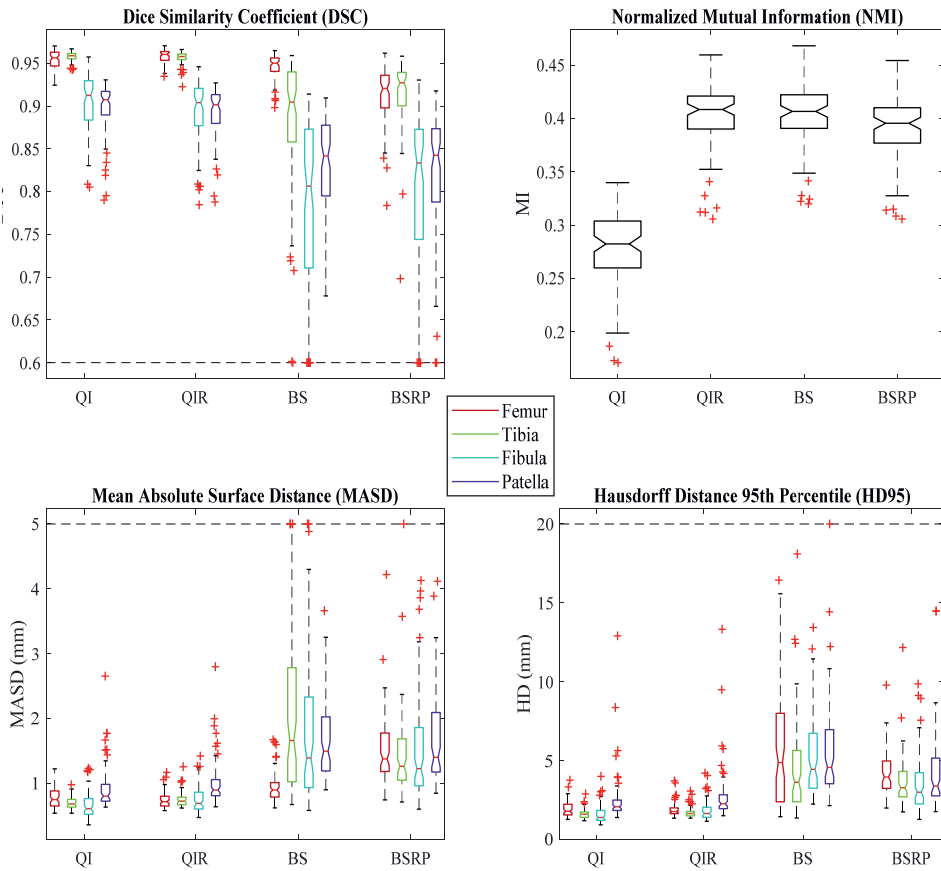
Figure 2.5: Boxplots showing the Dice Similarity Coefficient (DSC), Normalized Mutual Information (NMI), Mean Absolute Surface Distance (MASD) and 95th percentile Hausdorff Distance (HD95) for each bone after registration using the QI, QIR, BS and BSRP workflows. Outliers on the dotted lines indicate that the outlier exceeded that value, but that it was clipped for clarity.

Table 2.1: This table shows the results of the registration using the different workflows defined by their median and interquartile (IQR) ranges. Results include the Normalized Mutual Information (NMI), Dice Similarity Coefficient (DSC), Mean Absolute Surface Distance (MASD), Hausdorff Distance (HD95) and Jacobian Determinant (JD). The average value was calculated by averaging over the bones, and then taking the median over the patients. Average values indicative of best for a metric are shown in bold, with the corresponding p-value. The Median JD is calculated as the median JD in each subject, averaged over all subjects. The IQR JD is calculated as the IQR of the JD in each subject, averaged over all subjects.

|		QI		QIR			BS		BSRP	
		Median	IQR	Median	IQR		Median	IQR	Median	IQR
NMI		0.28	0.04	0.41	0.03		0.41	0.03	0.40	0.03
DSC	Femur	0.96	0.02	0.96	0.01		0.95	0.02	0.92	0.04
	Tibia	0.96	0.01	0.96	0.01		0.91	0.08	0.93	0.04
	Fibula	0.91	0.04	0.90	0.04		0.81	0.15	0.83	0.13
	Patella	0.91	0.03	0.90	0.03		0.84	0.08	0.84	0.08
	Bone Average	0.93	0.02	0.93	0.02	$p<10^{-26}$	0.87	0.05	0.87	0.06
	Water	0.64	0.08	0.77	0.06	$p<10^{-11}$	0.75	0.08	0.74	0.08
	Fat	0.69	0.12	0.86	0.07	$p<10^{-5}$	0.84	0.07	0.84	0.07
MASD (mm)	Femur	0.75	0.23	0.72	0.15		0.90	0.23	1.38	0.60
	Tibia	0.68	0.12	0.73	0.12		1.62	1.60	1.29	0.64
	Fibula	0.60	0.24	0.68	0.24		1.38	1.40	1.25	0.88
	Patella	0.81	0.26	0.90	0.26		1.53	0.80	1.41	0.91
	Average	0.72	0.18	0.76	0.19	$p<10^{-26}$	1.40	0.70	1.40	0.57
HD95 (mm)	Femur	1.79	0.68	1.77	0.36		4.88	5.55	3.96	1.72
	Tibia	1.60	0.34	1.65	0.28		3.63	3.15	3.28	1.58
	Fibula	1.39	0.60	1.65	0.62		4.45	3.46	3.00	1.96
	Patella	2.08	0.66	2.27	0.86		4.57	3.40	3.39	2.33
	Average	1.77	0.51	1.88	0.50	$p<10^{-26}$	4.99	2.83	3.56	1.40
Median JD	Bone	1.00	0.00	1.00	0.00		0.98	0.05	1.00	0.01
	Soft tissue	1.02	0.02	0.99	0.06		0.98	0.05	0.99	0.04
IQR JD	Bone	0.00	0.00	0.03	0.01	$p<10^{-19}$	0.19	0.11	0.05	0.02
	Soft tissue	0.12	0.05	0.26	0.07		0.25	0.07	0.20	0.05
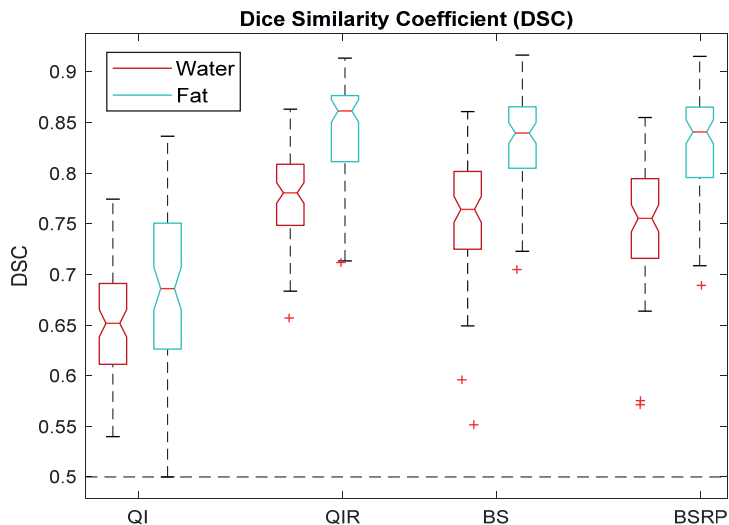
Figure 2.6: Distribution of the DSC of the automatic water and fat segmentations on CT and MR after registration using the various methods.

On the bottom row of Figure 2.7, the distributions of the JD for bone and soft tissue averaged over all scans are shown. The median JD and IQR JD are also shown in Table 2.1. The median JD was defined as the median of the JD for each patient, for which we displayed the median and interquartile range of those values over all patients. The IQR JD was defined as the interquartile range of the JD of each patient, for which we also show the median and interquartile range of each patient. As mentioned in the Methods, a JD deviating from one indicates that non-rigid deformations have occurred. Because for each method the median JD in bone is close to one, this indicates that the bones have no overall expansion or shrinkage in any method. However, the higher median IQR JD in bone in the BS and BSRP method indicate that non-rigid deformation did occur with those methods. The lower median IQR JD found in the QIR method shows less non-rigid deformation has occurred there.

The results showed that the JD within bone for the QI and QIR registered images have small interquartile ranges, and a median value of one, indicating that the bone remained rigid during registration. The higher interquartile ranges in bone for BS and BSRP showed that non-rigid deformations did occur, even for BSRP. Although BSRP resulted in higher rigidity in the bone than BS, it also caused the entire deformation to become more rigid, resulting in the registration errors evident from the lower DSC and higher MASD. The QIR, BS and BSRP all show a median JD lower than one in soft tissue, indicating shrinkage. This may have been caused by a change in tissue volume visible in the different field of views on CT and MR, due to positional changes in between scans.
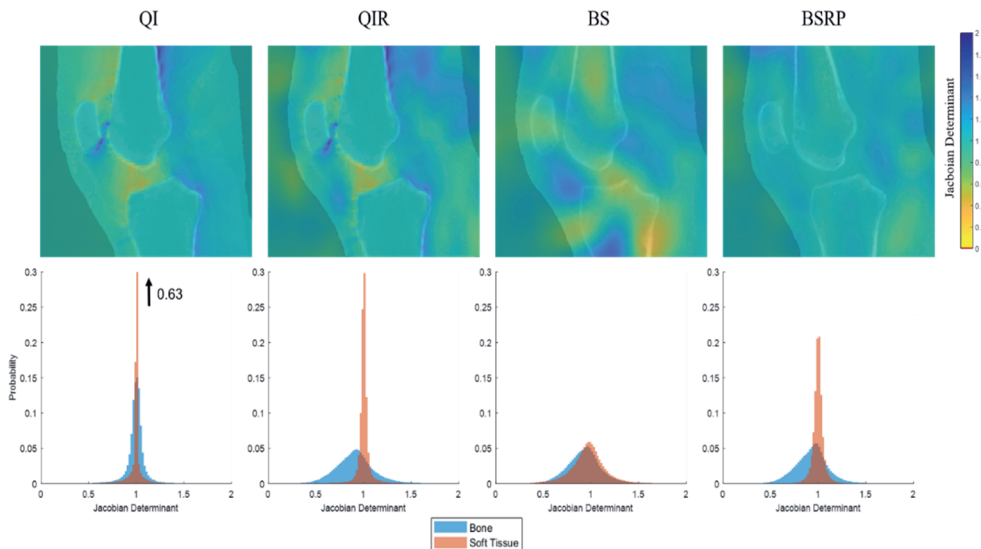
Figure 2.7: Top row: Example of a CT registered to the top image using each workflow, with a colour overlay of the Jacobian Determinant for each voxel. Bottom row: Probability histograms of the Jacobian Determinant within and outside the bone segmentation averaged over all scans. The binwidth is 0.02.

2.5 Discussion

In this chapter, we have presented a workflow for the registration of complex deformations in skeletal units between CT and MR, and compared it to two conventional workflows. Our Quaternion Interpolated Registration (QIR) algorithm exploits prior knowledge of the skeletal unit, i.e. the rigidity of bone structures. By interpolating the deformation fields of the rigidly transformed bone structures to estimate a smooth and realistic deformation for the soft tissue in proximity of the rigid bones, a good initialization could be found for a final deformable registration step.

QIR resulted in the highest scores for the bone segmentation-based metrics. When comparing the results between the BS (B-Spline) and BSRP (B-Spline with Rigidity Penalty) methods, the BS seemed to perform slightly better. However, the Jacobian Determinant (JD) interquartile ranges reported in Table 2.1 and the probability histograms in Figure 2.7 showed more non-rigid deformations were present inside the bone when applying the BS, which indicated that non-realistic deformations occurred.

The average NMI values of the three registration methods were similar, and thus gave no further indication of the quality of the registration. This is in agreement with the conclusions of Rohlfing[57], who stated that only metrics describing overlap of localized anatomical regions are able to differentiate between accurate and inaccurate registrations.

A confounding factor in this study is the quality of the bone and soft tissue segmentations. Obtaining segmentations that match exactly between the CT and MR scans is impossible, which is why this should be taken into account when interpreting the segmentation-based metrics. This means that even if the CT and MR would be perfectly registered, the DSC would not be equal to one, nor would the MASD and HD95 be zero. Besides improving the image analysis, this could potentially also be addressed by improving the bone visualisation on MRI using dedicated sequences. Recent developments in MRI-based bone visualisation that could facilitate this include ultrashort echo time or Zero-Echo time MR imaging[58,59], or machine learning based synthetic CT methods[60,61]. As both approaches aim to generate CT-like contrast, it may be expected that both segmentation and registration can be improved when using this data as input.

Intermodal registration of images of rotated joints, as encountered in our study, has shown to be challenging when using currently available state of the art methods, due to a number of different reasons. First, a high degree of deformation occurred in the joint space. As a result, highly non-rigid deformations occurred close to the bone, while the bone itself should not deform. This is difficult to capture using deformable B-Splines, as these tend to represent smooth deformations. As a countermeasure, a rigidity penalty was employed, but this inhibited large local deformations in the joint space. Secondly, because of the large deformations in the joints, it was impossible to find a good initial alignment using a single rigid transformation. In practice, the initial alignment was therefore mostly based on the alignment of the bulk of the (soft) tissue, which did not guarantee proper alignment of the bone. Lastly, in multimodal registration there is no one-to-one correspondence of voxel intensities. This is the reason NMI was used to estimate these correspondences. However, when using NMI, erroneous tissue correspondences were found when the initial alignment was not good enough, resulting in poor registration results.

In this chapter we have chosen to compare our method to algorithms available in the *elastix* toolbox because of the availability of a rigidity penalty and earlier applications to similar registration problems[60,62,63]. However, the B-Spline algorithm is limited in its ability to realistically represent transformations at sliding interfaces such as occur at the joint, which is also not solved by the dual quaternion interpolation we employ. Extension of the workflow with the option for sliding interfaces as introduced by [35] might therefore further improve results. In future research it would be interesting to include more methods in the comparison such as the diffeomorphic symmetric image normalization method (SyN) algorithm[64], or the MIND Demons registration method which proposes a different metric for inter-modality correspondence[65]. Tacchela et al.[66] suggested to establish an individualized registration strategy for each patient, where a choice was made between different algorithms based on certain criteria. This strategy could be an interesting approach to our method as well, where a decision to use the QIR or another method would be based on a quantifiable criterion.

Finally, in the past few years, a large number of papers has been published on the application of deep learning networks for image registration[67]. Both supervised models[31,32] and

unsupervised models[30,33] have shown to be an efficient alternative to traditional registration methods. However, these deep learning methods do not specifically address the problem introduced by rotated rigid structures that we address here. In the future, our method could be used to supply a deep learning network with the training data necessary to tackle this problem.

2.6 Conclusion

We developed a method to improve outcome of multimodal registration problem involving rigid and non-rigid tissues, such as encountered in the knee joint. The results showed that using rigid structures to provide a more accurate initial transform improved the registration outcome compared to two other registration workflows. Our method is most applicable to situations where conventional methods may fail to find the correct correspondence between tissues, due to a combination of rigid-body movements and large localized non-rigid deformations.

2.7 Appendix A: Automatic and Manual Bone Segmentation

2.7.1 Introduction

The Quaternion Interpolated Registration (QIR) workflow as proposed in this chapter necessitated semantic segmentation of the rigid structures in both the moving and fixed image. In the case of the knee joint, these rigid structures corresponded to four bones: the femur, tibia, patella and fibula. Segmentation of the bones on the moving (CT) and fixed (MR) image represented separate challenges.

On CT, bone segmentation is relatively simple as bone has distinct Hounsfield Unit (HU) range. However, some challenges persisted. Firstly, cancellous bone structures as well as the medullary shaft do not always exceed the threshold of 200 HU that is normally used for bone segmentation, while these structures were assumed to stay rigid between scans and should thus be included in the bone segmentation. Secondly, our method needed the bones to be separately labelled, i.e. semantically segmented. With small joint space, low tissue contrast caused by partial volume effects or spurious bone growth due to osteoarthritis, separating the different bones from each other became a non-trivial problem.

On MR, bone segmentation was an even more complex problem, as apart from the same issues faced with semantic segmentation on CT, there was also no distinct threshold value that could be reliably used to separate bone from soft tissue. Therefore, manual segmentation of MR images had to be done by a combination of thresholding and manual delineation of the bone boundaries. These boundaries are especially unclear at the tendon insertion sites, as tendons tend to have the same image intensity as bone on the MR contrasts that were used in this study.

When only a small number of CT and MR scans need to be registered, manual segmentation can be used to deal with these problems. However, when dealing with a dataset of 92 MRI and CT scans manual segmentation became prohibitively time-consuming. Therefore, we chose to initialize the segmentation process with an automatic segmentation method. In order to ensure the quality of the segmentations and the accuracy of the evaluation of our registration method, all segmentations were subsequently manually checked and corrected if needed.

2.7.2 Automatic segmentation

Automatic semantic segmentation of the bones was performed using a custom deep learning implementation. We used a Convolutional Neural Network (CNN) architecture based on the U-net and V-net architecture as proposed by Milletari et al.[46]. To deal with the large size of the images and memory restrictions of the GPU, a patch-based approach was chosen. Consequently, in order to ensure the input of the network contained enough information to discern between the different bones, a cascade-like architecture consisting of two consecutive

V-nets was used. The first V-net used low resolution, high field of view image patches of 128x128x128 voxels resampled to 32x32x32 voxels, and thus used information of a large portion of the image to produce a low-resolution segmentation. This information was passed to the second V-net together with a higher resolution, small field of view patch of 32x32x32 voxels at full resolution to segment the image in the desired resolution.

2.7.2.1   V-net

Each V-net consisted of an encoder-decoder network with a depth of 4 levels. The number of feature channels in the convolutional layers was started at 32 for the first convolution and doubled in each level. The encoder part of the network level contained a 3x3x3 convolutional layer and a rectified linear unit (ReLU). After each level the data is downsampled using a 2x2x2 convolutional layer with stride 2. The decoder part of the network was similar, but after each level the data was upsampled using 2x2x2 deconvolutional layer which halved the number of feature channels. Short-cut connections were used which connected each corresponding level of the encoder and decoder network by concatenation. At the end of the network a 1x1x1 convolutional layer was used to map the feature channels to the channels that corresponded to the segmentation of each bone. Finally, a softmax function was added to the network to produce a probability map for each bone as output.

2.7.2.2   Network Training

To train the network, 12 MR and 12 CT images were manually segmented into femur, tibia, patella and fibula channels. Separate networks were trained for CT and MR segmentation using the same architecture. All images were resampled to isotropic voxels of 0.7 mm^3. The network input consisted of two patches of Cx32x32x32 voxels, with C the number of input channels that was used. The first input patch was a Cx128x128x128 patch of the input image that was downsampled using trilinear interpolation. For CT segmentation only one channel was used, but for MR segmentation four Dixon reconstructed images were used as channels: water, fat, in phase and out of phase.

The networks were trained with a batch size of 16 per iteration. The Adam-optimizer[68] was used with a linearly decreasing learning rate from 10^{-3} at the start to 10^{-6} at the end of the training, after 200 epochs. Categorical cross-entropy loss was calculated on the output of both the first and second of the cascaded V-net. To increase the apparent size of the training set, patches were flipped along the sagittal plane with a 50% chance. The networks were implemented using PyTorch on a 24GB Quadro M6000 GPU (Nvidia, Santa Clara, California, USA).

To check the accuracy of the automatic segmentation network training was first performed as a 4-fold cross validation with a split of 9 training and 3 test images. To obtain the optimal automatic segmentations for the registration methods, both networks were trained with the full manually segmented training sets of 12 images per modality.

After running inference on the full dataset, the bone with the highest softmax prediction was chosen for each voxel. The largest connected component of each bone was used, and all smaller components were discarded. Finally the segmentations were resampled to the original image size.

2.7.3   Evaluation

The segmentation accuracy was evaluated using the same metrics as described earlier in this chapter, i.e. Dice-Sörensen Coefficient (DSC), Mean Absolute Surface Distance (MASD), Hausdorff Distance (HD) and 95th percentile Hausdorff Distance (HD95). The results of this evaluation can be found in Table 2.2. The automatic CT segmentations were generally better than the MR segmentations. The MR segmentation of the fibula performed worst. This was likely because often only a small part of the fibula was visible on the MR scans.

2.7.4   Manual corrections

To ensure correct segmentations for the registration workflow and evaluation of the other registration methods, all automatic segmentations were manually checked and corrected using 3D Slicer [47]. Checking the segmentations for errors took on average about 60 seconds per scan on both CT and MR. Manual corrections were performed in 54 CT scans and in all MR scans, out of 92 total scans for each modality. Manual corrections took on average 165 seconds for CT scans and 286 seconds for MR scans.
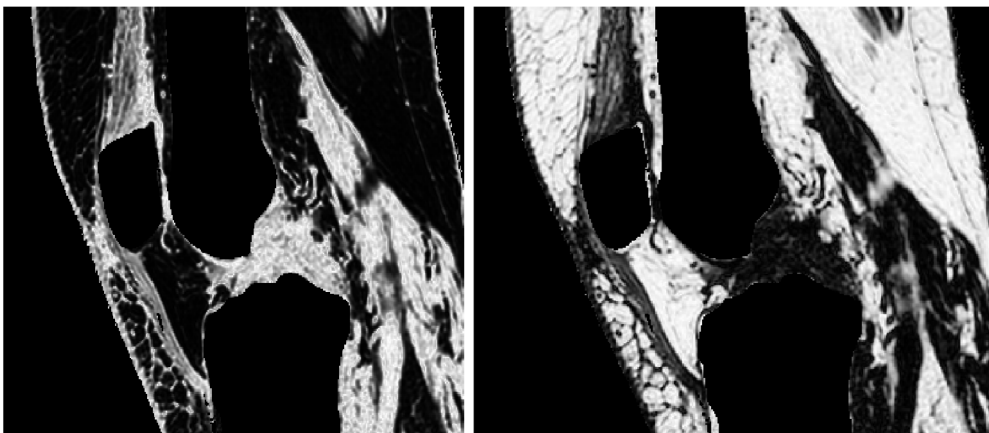

Figure 2.8: Fat fraction (left) and water fraction (right) obtained from the water and fat weighted Dixon reconstructed images.

Table 2.2: Evaluation of the cross-validated results of the automatic segmentations of the MR and CT.

		Automatic MR		Automatic CT	
		Average	Std. Dev.	Average	Std. Dev.
DSC	Femur	0.96	0.02	0.98	0.002
	Tibia	0.93	0.03	0.98	0.003
	Fibula	0.79	0.21	0.95	0.007
	Patella	0.94	0.01	0.94	0.009
	Average	0.91	0.06	0.96	0.004
MASD (mm)	Femur	0.57	0.09	0.32	0.018
	Tibia	0.65	0.33	0.35	0.042
	Fibula	1.59	1.62	0.42	0.054
	Patella	0.60	0.24	0.37	0.069
	Average	0.85	0.50	0.36	0.034
HD95 (mm)	Femur	1.64	0.37	0.80	0.051
	Tibia	2.24	1.83	0.92	0.112
	Fibula	5.64	6.23	1.10	0.303
	Patella	2.02	1.02	1.01	0.298
	Average	2.88	2.00	0.96	0.128
HD (mm)	Femur	7.66	5.67	3.89	1.780
	Tibia	4.96	2.11	4.54	2.878
	Fibula	8.89	7.47	2.90	1.125
	Patella	3.36	2.30	3.36	2.299
	Average	7.17	4.38	3.67	1.428

2.8 Appendix B: Soft Tissue Segmentation

2.8.1 Introduction

The registration workflow presented in this chapter aimed to improve the registration of both rigid (bone) tissue and the soft tissue. While evaluation of the bone registration was trivial with the bone segmentations that were already necessary for the implementation of the algorithm, evaluation of the deformable part of the registration necessitated a method to compare the soft tissue overlap after registration. Therefore, we roughly segmented the soft tissue into two parts. The first part was tissue that was high in water content, such as muscle tissue. The second part is the tissue with a high fat content, which is the adipose tissue that is mostly present around the knee joint as subcutaneous and inter- and intramuscular fat. To segment the fat and muscle tissue on CT threshold based methods were used, whereas the
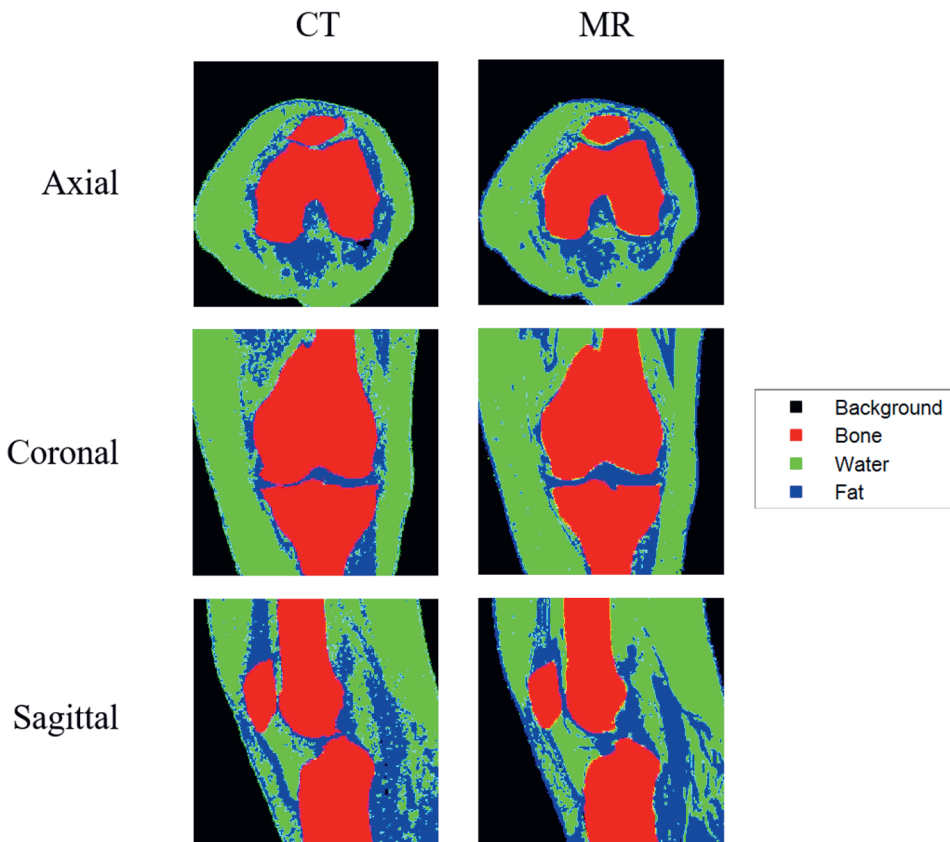
Figure 2.9: CT and MR of the knee segmented into three components: bone, water and fat.

MR was segmented using the water and fat fraction calculated using the Dixon reconstructed images.

2.8.2   Methods

2.8.2.1   CT segmentation

The thresholds for the CT were based on values found in literature that were found to correlate strongly to manual segmentations of the tissue types. For CT the threshold for fat was chosen to be between -190 up to -30 Hounsfield Units (HU)[69,70]. The threshold for muscle was chosen to be between -30 up to +150 HU[71].

2.8.2.2   MR segmentation

The water and fat components of the MR images were not calculated using thresholds, but by use of the water and fat Dixon reconstructions. The water and fat fraction of each voxel could be calculated by using the following formulas:

$$Water\ fraction\ = \frac{MR^w}{MR^{ip}}$$

$$Fat\ fraction\ = \frac{MR^f}{MR^{ip}}$$

Where MR^{ip}, MR^{water} and MR^{fat} are the in phase, water and fat reconstructed Dixon images. The foreground mask was obtained by only using voxels where either $MR^{water} > 100$ or $MR^{fat} > 100$. The water and fat segmentation were obtained by thresholding both the water and fat fraction images at > 0.5. Examples of the water and fat fraction images are shown in Figure 2.8.

The bone segmentations were used as a mask for all soft tissue segmentation such that there would be no bone and soft tissue segmentation overlap.

2.8.3   Segmentations

An example of the bone, water and fat segmenations from CT and MR are shown in Figure 2.9. In Figure 2.10 an example is shown of the segmentation overlap of CT and MR after registration with each of the registration workflows discussed in this chapter, in all three orthogonal planes. The average DSC of both water and fat after registration with the different methods are shown in Table 2.1.
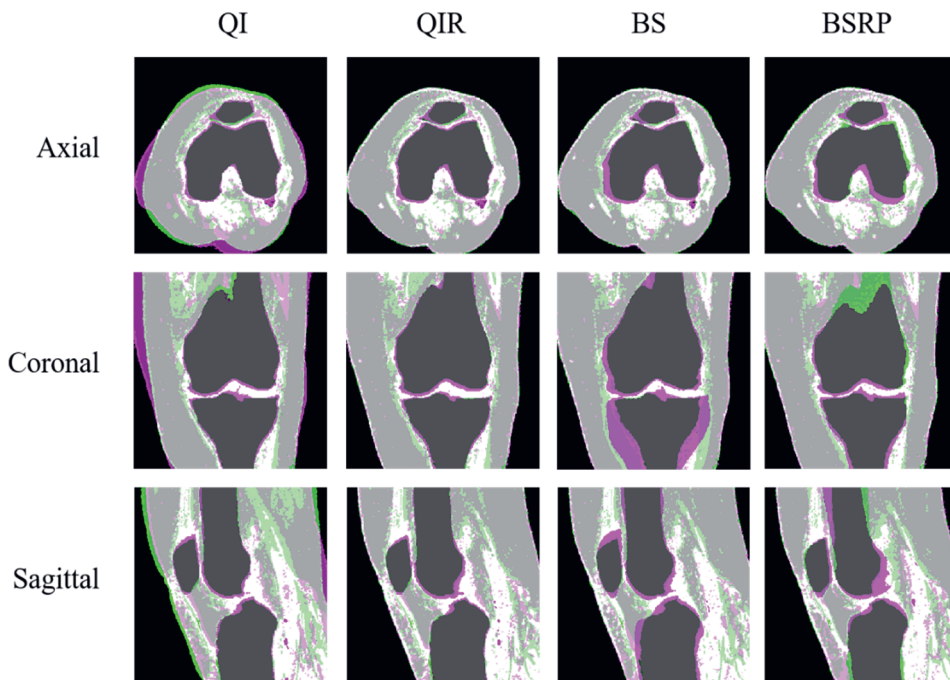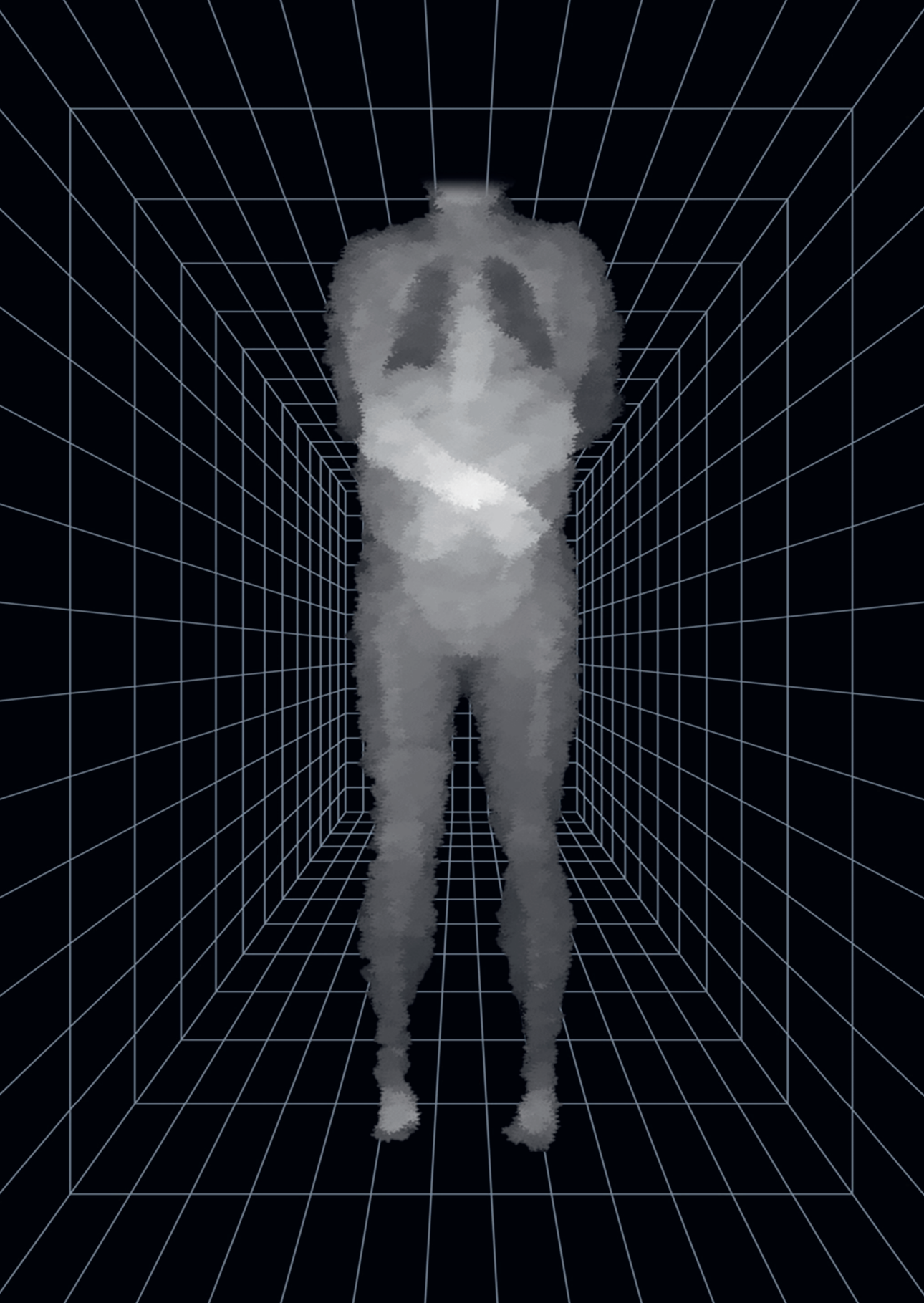

Figure 2.10: Difference images of the CT segmentations after registrations using the various workflows. Pink and green indicate segmentations of tissues that do not correctly correspond between the CT and MR images.

Chapter 3

Efficient cascaded V-net optimization for lower extremity CT segmentation validated using bone morphology assessment

Abstract

Semantic segmentation of bone from lower extremity CT scans can improve and accelerate the visualization, diagnosis and surgical planning in orthopaedics. However, the large field of view of these scans makes automatic segmentation using deep learning based methods challenging, slow and GPU memory intensive. We investigated methods to more efficiently represent anatomical context for accurate and fast segmentation and compared these with state-of-the-art methodology. Six lower extremity bones from patients of two different datasets were manually segmented from CT scans, and used to train and optimize a cascaded deep learning approach. We varied the number of resolution levels, receptive fields, patch sizes, and number of V-net blocks. The best performing network used a multi-stage, cascaded V-net approach with 128^3-64^3-32^3 voxel patches as input. The average Dice coefficient over all bones was 0.98 ± 0.01, the mean surface distance was 0.26 ± 0.12 mm and the 95th percentile Hausdorff distance 0.65 ± 0.28 mm. This was a significant improvement over the results of the state-of-the-art nnU-net, with only approximately $1/12^{th}$ of training time, $1/3^{th}$ of inference time and $1/4^{th}$ of GPU memory required. Comparison of the morphometric measurements performed on automatic and manual segmentations showed good correlation (ICC > 0.8) for the alpha angle and excellent correlation (ICC > 0.95) for the hip-knee-ankle angle, femoral inclination, femoral version, acetabular version, LCE angle, acetabular coverage. The segmentations were generally of sufficient quality for the tested clinical applications and were performed accurately and quickly compared to state-of-the-art methodology from the literature.

3.1 Introduction

Differentiating between different tissue types on radiological scans, also known as segmentation, is an important part of modern medical image analysis. In orthopaedics, it is most often used to differentiate between bones, muscle, ligaments and cartilage on X-ray, Computed Tomography (CT) and Magnetic Resonance Imaging (MRI) scans[72]. Segmentation of bones of the lower extremities has a multitude of use-cases in the clinical practice, such as the measurement of hip, ankle and knee joint range of motion[73,74], determination of anatomical and mechanical axes of the full lower limbs[3,75,76], indication of landmarks[77], and the fabrication of surgical tools[78] and medical implants[79]. However, manual segmentation is a time consuming task that requires knowledge and expertise of dedicated software[72]. Difficulties encountered in bone segmentation include differing image intensity values between cortical and cancellous bone, insufficient resolution to differentiate between bones in joint spaces and low signal-to-noise ratios in scans. Therefore, automatic segmentation methods have been developed that aim to accelerate the segmentation process and remove interobserver variability.

Initial (semi-) automatic segmentation algorithms relied mostly on intensity-based methods, such as thresholding, region growing, or edge detection. Later, more sophisticated methods such as Statistical Shape Models (SSMs) and Atlas-based methods were applied[80]. For example, Audenaert et al[81] proposed a Statistical Shape Model based pipeline for semantic segmentation of lower body CT scans. Recently, deep learning networks, and more specifically Convolutional Neural Networks (CNNs), have shown to be an effective approach for CT bone segmentation[79,82–85]. An often used network architecture for medical image segmentation is the U-net[45], which can be extended to 3D and is then called 3D U-net[86] or V-net[46]. Noguchi et al.[87] showed that the V-net architecture is suitable for binary segmentation of the complete human skeleton from CT.

The lower extremity CT scans used in this study have a large field of view and comparatively high resolution, resulting in a large amount of data to be processed. An issue encountered in CNNs when segmenting large volumes is the increase in Graphical Processing Unit (GPU) memory that is needed when the input for the network becomes larger. Patch-based networks circumvent this problem by segmenting only part of the image at a time. This reduces the amount of voxels the network needs to process each iteration, but also reduces the amount of contextual information the network has for each prediction. Additionally, due to the large number of patches per scan inference can become prohibitively slow. Multiple studies have tried to balance the amount of information available to the network and its memory consumption by implementing cascaded approaches[88–91] or by using dilated (atrous) convolutional layers[92,93]. Isensee et al.[94] introduced a (cascaded) U-net based network called nnU-net that automatically adapts its architecture and processing steps to the dataset on which it is used. However, none so far have studied the influence of varying the number of cascades by more than two. Additionally, full lower extremity scans have a larger number of voxels

than usually encountered in these studies, increasing the need for lightweight method for efficient segmentation.

Two examples of the clinical application of segmentations of the lower extremities include the Hip-Knee-Ankle angle (HKA) measurement and hip morphometry assessment. Although HKA measurements were conventionally performed manually on 2D radiographs[2], recent studies in knee realignment planning show an increased trend towards 3D CT planning and 3D measurements[95]. Hip morphometry assessment was conventionally also performed on X-rays or CT scans by manual segmentation and indication of landmarks[96–99]. Recently, Palit et al.[73] performed automatic assessment of the location of bone impingement and range of motion analysis using manual segmentations of the bone from CT. Lerch et al.[74] showed that MR-based segmentations of the hip joint could also be used. Zeng et al.[100] then showed that automatic MRI segmentation also correlated well with manual CT segmentation, removing the need for time-consuming manual segmentations.

In this chapter, we propose a deep learning based approach for automatic, accurate, fast and memory efficient segmentation of bones from the entire lower extremities. For this purpose, we introduce a new lightweight variant of the cascaded approach for the V-net architecture to reduce memory requirements and speed up training and inference. Additionally, we applied a sampling scheme that maximizes use of the available information. This approach was evaluated in four steps:

1. The optimal network architecture was studied by comparing the results with differing number of cascades and input configurations of the cascaded V-net.
2. The best performing network was trained as a five-fold cross-validation and the results were compared to existing methods found in the literature and the recent nnU-net[94].
3. The clinical applicability was evaluated by performing two different medical assessments using both the manual and automatic segmentations:
    a. The hip morphology was measured using proprietary software developed by the Move Forward service (Clinical Graphics, Zimmer Biomet, Warsaw, Indiana, USA).
    b. The Hip-Knee-Ankle (HKA) angle was measured using an automatic workflow based on the methods from Fürnstahl et al.[3]
4. The best performing network was evaluated on a dataset with different subject demographics and acquisition parameters to evaluate the robustness of the method.
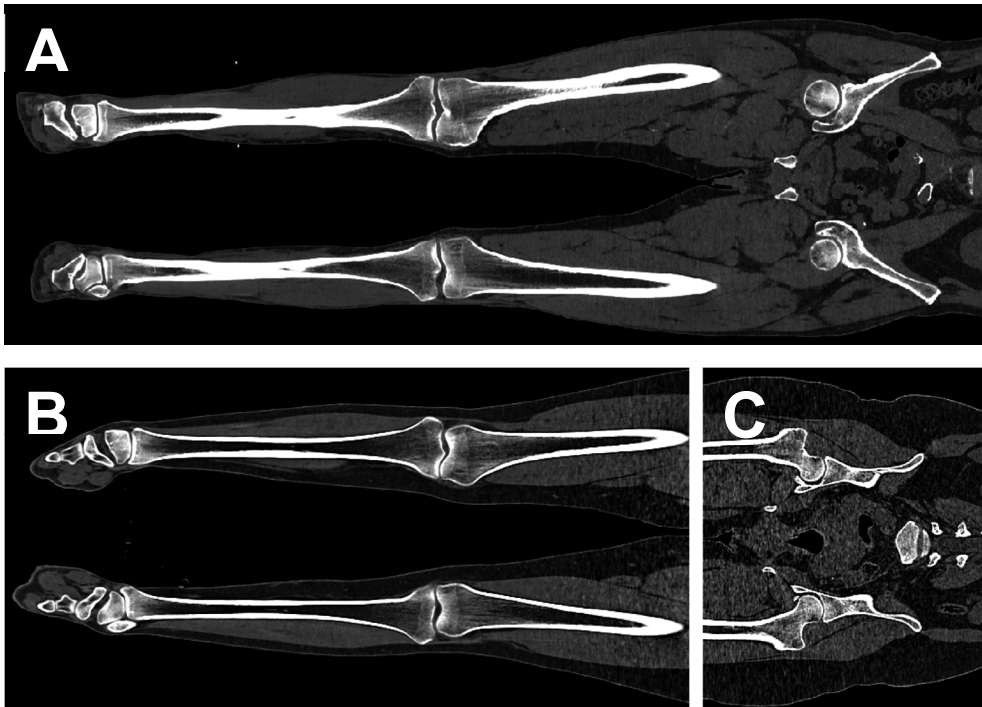
Figure 3.1: Examples of the CT scans of two patients.
A. UMCU dataset.
B. NMDID dataset, lower leg scan.
C. NMDID dataset, torso scan, cropped to include only the femora and coxae

Table 3.1: Summary of the different configurations of the cascaded V-net architecture that were compared. The receptive field denotes the shape and size in voxels, that is used for each stage of the V-net. The number of input voxels, number of trainable deep learning parameters and GPU memory consumed during inference are all a consequence of the number of V-nets and size of the receptive field. The cascaded V-net GPU memory had a base memory consumption of approximately 685 MB for each configuration, in the table the additional GPU memory used per patch during inference is shown. The 3D U-net and V-net were included as comparison.
: Not available in the study.

Network	Nr. of V-nets	Receptive field	Nr. of input voxels	Nr. of parameters	Additional GPU memory per patch (MB)
	1	32^3	3.3×10^4	1.2×10^6	51
	1	64^3	2.6×10^5	1.2×10^6	419
	1	128^3	2.1×10^6	1.2×10^6	3340
Proposed Cascaded V-net	2	$64^3\text{-}32^3$	6.6×10^4	2.4×10^6	52
	2	$128^3\text{-}32^3$	6.6×10^4	2.4×10^6	52
	2	$128^3\text{-}64^3$	5.2×10^5	2.4×10^6	421
	3	$128^3\text{-}64^3\text{-}32^3$	9.8×10^4	3.6×10^6	53
	4	$256^3\text{-}128^3\text{-}64^3\text{-}32^3$	1.3×10^5	4.8×10^6	58
3D U-net[86]	1	$132 \times 132 \times 116$	2.0×10^6	1.9×10^7	-*
V-net[46]	1	$128 \times 128 \times 64$	1.0×10^6	6.6×10^7	-*

3.2 Methods

3.2.1 Data

1.1.1.1 UMCU dataset

For the training and initial evaluation of the networks, fifty CT scans of the lower extremity were used. The anonymized data was acquired retrospectively and was judged not to be subject to the Medical Research Involving Human Subjects Act (WMO) by the Medical Ethical Committee, as described in IRB protocol number 16-612/C. Lower extremity CT scans of patients who had undergone CT scanning due to unrelated medical reasons (i.e. vascular indications) were collected from the University Medical Center Utrecht (UMCU, Utrecht, The Netherlands) and anonymized. The mean age of the male patients was 61 years (SD: 10 years) and of the female patients 53 years (SD: 15 years).

All CT scans were acquired with either the Philips iCT scanner or Philips Brilliance 64 (Philips Medical Systems, Best, The Netherlands). The CT acquisition parameters were: tube voltage = 120 kVp, tube current = 31-347 mA, effective dose = 35-150 mAs, slice thickness
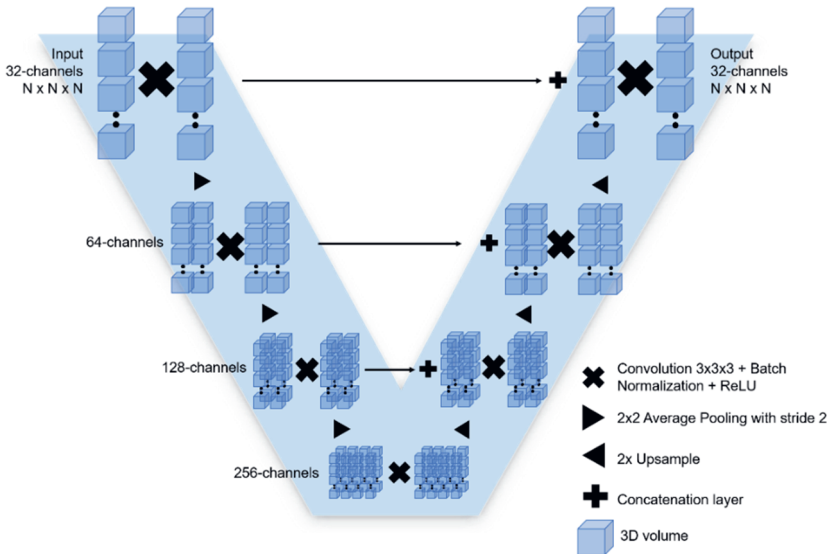
Figure 3.2: Overview of the light-weight version of the V-net [46] architecture that was used in this study. An input patch with shape NxNxN was used as input. The network consists of a downsampling (left side) and upsampling (right side) portion, where the resolution decreases and increases, respectively. In each level of the network a single convolutional block was used, of which the number of channels doubled for each level of the network.

= 1 mm, slice increment = 0.7 mm, pixel spacing = 0.63-0.98 mm, matrix size = 512 × 512 pixels. The iDOSE4 reconstruction algorithm was used.

3.2.1.1 NMDID dataset

To evaluate the robustness of the final deep learning segmentation network, a dataset of over 15000 deceased subjects with different acquisition parameters and population distribution was acquired from the New Mexico Decedent Image Database (NMDID)[101]. We selected 10 patients with death by natural cause, age under 50 and similar positioning to patients from the UMCU dataset: 6 male and 4 female patients, with a mean age of 34 years (SD: 6.4 years). For each subject, two CT scans were available which included the lower extremities. A torso scan, including the coxae and femur, and a lower leg scan, including the tibia, fibula, talus, calcaneus and femur. All CT scans were acquired on a Philips Brilliance Big Bore scanner (Philips Medical Systems, Best, The Netherlands). The CT acquisition parameters were: tube voltage = 120 kVp, tube current = 82-245 mA, effective dose = 100-301 mAs, slice thickness = 1 mm, slice increment = 0.5 mm, pixel spacing = 0.63-1.17 mm, matrix size = 512 × 512 pixels. The images were resampled using trilinear interpolation to isotropic 0.8 x 0.8 x 0.8 mm to match the voxel size of the UMCU dataset more closely. Due to high noise levels in the proximal part of the lower leg scan and in the torso scan, a Gaussian filter (sigma = 0.5, kernel size = 4) was applied before segmentation. Examples of CT scans from both datasets are shown in Figure 3.1.
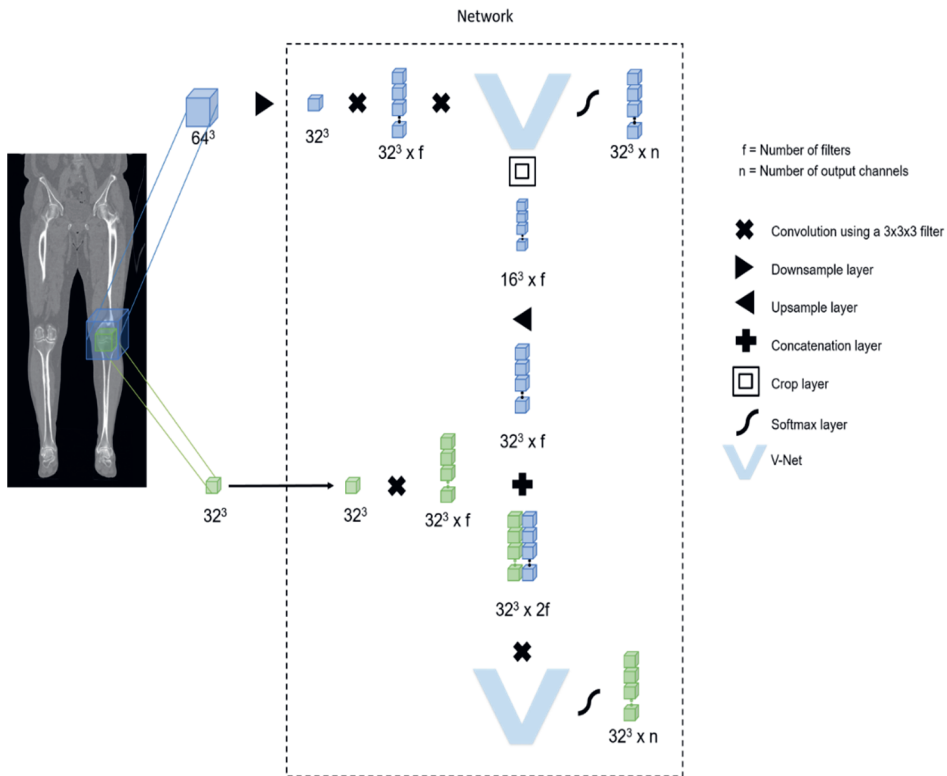
Figure 3.3: Overview of an example of the cascaded V-net architecture. In this example, two patches were chosen of 64^3 and 32^3 voxels, the smaller patch from within the larger patch. First, at the top of the network the 64^3 patch was downsampled to the same size as the smaller patch. The lightweight V-net shown in Figure 3.1 was then used to segment the patch. The output was both converted to a probability map for each segmentation using the Softmax-layer (on the right), as well as cropped to be used as additional information for the smaller, high resolution patch (downward). Next, the output of the first V-net was cropped and concatenated to the 32^3 full resolution patch. This was then used as input for the second V-net. After applying the Softmax-layer, the 32^3 full resolution probability prediction for each segmentation is the final output of the network.

3.2.1.2   Segmentation

For both datasets the bones were segmented semantically, i.e. each bone was given an individual label. The tibia, fibula, talus, calcaneus, femur and coxae were manually segmented using a combination of the CT Bone Segmentation Module and manual editing in Mimics (Mimics Medical 21.0, Materialise, Leuven, Belgium). Two operators with 2 years of experience using Mimics segmented the UMCU dataset. Both operators checked each segmentation for accuracy. A single operator with 4 years of experience using Mimics segmented the NMDID dataset. Operators reported average segmentation times of up to 20 minutes per label for each patient.

3.2.2 V-Net

The network that was used was based on the 3D U-net[86] and V-net architecture[46], which are 3D extensions of the original U-net[45]. In the remainder of this chapter, we refer to these 3D network architectures as V-net. In order to reduce the memory requirements of the network, we used a lightweight implementation with only four stages per V-net, and only a single convolutional block per stage. Each convolutional block consists of a 3x3x3 convolution, followed by a batch normalization layer and a Rectified Linear Unit (ReLU) activation layer. At the end of the V-net a final convolutional layer is appended with a voxel-wise softmax layer to transform the output to a probabilistic segmentation of the input patch. An overview of this implementation of the V-net architecture is shown in Figure 3.2.

3.2.3 Cascaded networks

The lower extremity CT-scans have a high resolution and large spatial FOV. To segment the bones semantically from these scans a deep network with a large receptive field would be preferred. However, a single V-net with these attributes would necessitate high amounts of GPU memory and could lead to slow inference. A cascaded lightweight V-net strategy was therefore employed, which reduces the amount of input voxels while maintaining a large receptive field. This could speed up inference while minimizing the memory constraints on the GPU.
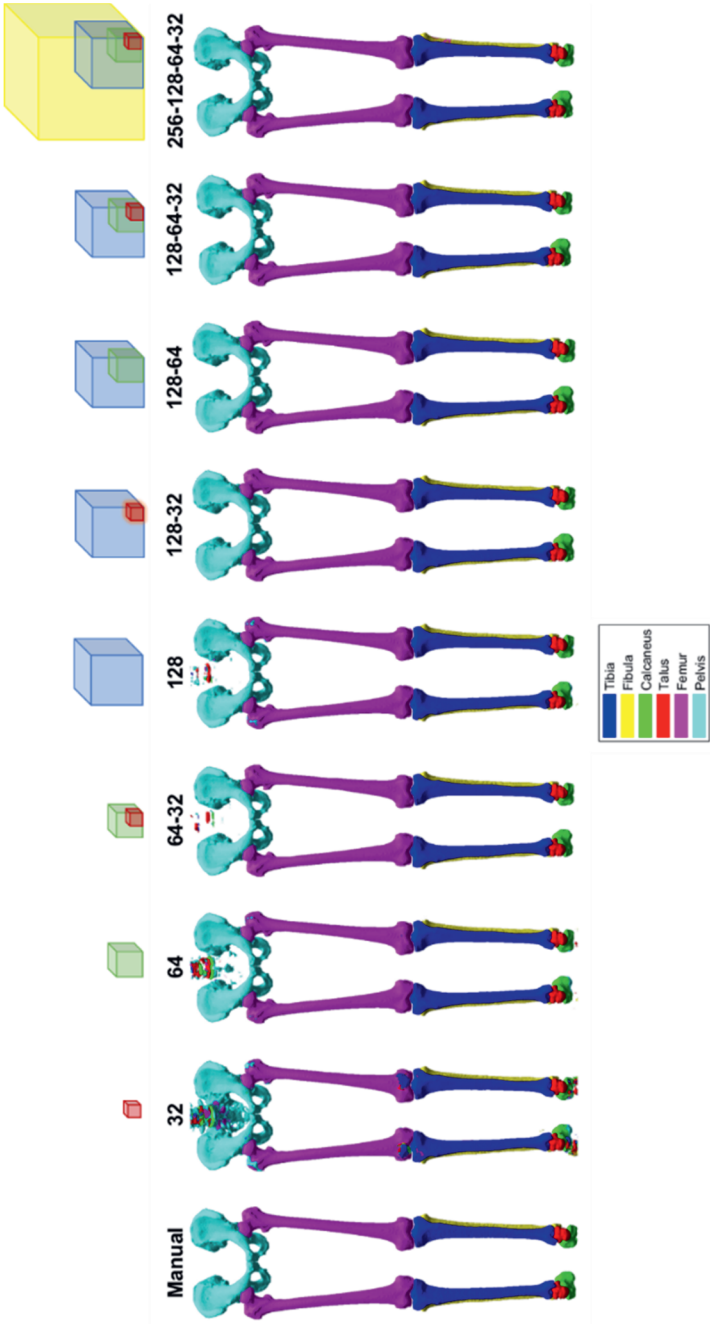
Figure 3.4: 3D visualization of the segmentation result of one subject using each input configuration, before post-processing. The cubes in the top row show the relative patch sizes, with 32^3 (red), 64^3 (green) 128^3 (blue) and 128^3 (yellow) voxels. The cubes are not to scale relative to the segmentations.
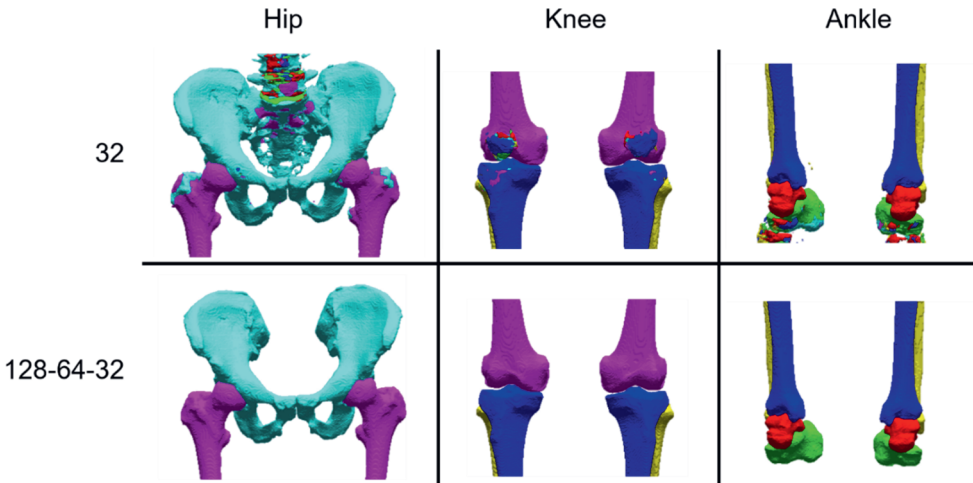
Figure 3.5: Detailed view of the differences in segmentation before post-processing when using a small, 32^3- voxel receptive field and a larger, 128^3-voxel receptive field.

The cascaded strategy used multiple consecutive V-nets. The first V-net took as input a large FOV patch from the CT that had been down-sampled, such that information from a large receptive field was available, albeit at a lower resolution. The prediction of this patch was then used by the consecutive V-net, which also used an additional CT patch at a higher resolution, but with a smaller FOV. This continued until a full resolution patch was used as input to the final V-net, after which the final full resolution patch was predicted. The network outputs a semantic segmentation prediction for each V-net, which was used during training, but only the prediction at the full resolution was used during inference. This strategy allowed the network to process information from a large FOV while maintaining low memory requirements. Figure 3.3 shows an example of a two-stage cascaded V-net. All network configurations that were studied and their metrics are summarized in Table 3.1. The network was implemented in such a way that it automatically adapted its architecture to the number and sizes of the input patches.

All patches were downsampled using trilinear interpolation to the size of the smallest patch prior to being put into the network. For example, a 128^3-64^3-32^3 configuration denotes a network that takes a 128x128x128 patch, a 64x64x64 patch and 32x32x32 patch as input, all downsampled to 32x32x32 voxels. This configuration thus has a receptive field twice larger than the V-net proposed by Çiçek et al.[86], while the number of voxels used as input to the network is ten times smaller, as shown in Table 3.1. To make optimal use of the cascaded V-net strategy an inference sampling strategy was used that performed inference on all smaller patch, instead of only the center patch. A detailed description of the training and sampling strategy is given in Supplementary Material: Appendix A.

3.2.4   Study design

The study design was divided into four stages. In stage 1, the number of cascaded networks and the input patch sizes were varied as summarized Table 3.1 to find the optimal input and network configuration. The 50 CT scans of the UMCU dataset were divided into sets of 35 scans for training, 5 scans for validation and 10 scans for testing. In stage 2, the configuration of the best performing network of stage 1 was trained such that all scans could be automatically segmented. In a five-fold cross-validation method, 40 scans were used for training and 10 for testing in each fold, as well as training of the cascaded version of nnU-net as a benchmark to compare our results. All training and inference parameters, settings and pre- and post-processing were kept the same as described in the study by Isensee et al.[94], of which the code is available at GitHub (https://github.com/MIC-DKFZ/nnUNet. In stage 3, the results of the proposed network from stage 2 were used to compare the outcome of automatic HKA measurement and hip morphometry assessment between manual and automatic segmented scans. In stage 4, the best performing network was trained on all 50 scans of the UMCU dataset using the same settings as for the five-fold cross-validation, and then evaluated on the NMDID dataset to test the robustness of the network to different subject demographics and acquisition parameters.

3.2.5   Evaluation

3.2.5.1   Segmentation analysis

Outcomes of the automatic segmentations were compared to the manual segmentations based on four commonly used metrics[21]: the Dice Similarity Coefficient (DSC), symmetric Mean Absolute Surface Distance (MASD), Hausdorff Distance (HD) and the 95th percentile of the Hausdorff Distance (HD95).

3.2.5.2   HKA and hip morphology assessment

We used both the manual and the automatic segmentations from the five-fold cross-validation to measure the HKA and the morphology of the hip joint for use in clinical practice.

The HKA was determined from the segmentations using an adapted version of the original protocol developed by Fürnstahl et al.[3] such that no manual steps were necessary. This ensured that any differences in the measurements were due to differences between the manual and automatic segmentations, and not due to intra- or interrater variability. The adjusted protocol is described in-depth in Supplementary Material: Appendix B.

A commercially available software tool, the Move Forward service (Clinical Graphics, Zimmer Biomet, Warsaw, Indiana, USA) was used to calculate the hip morphometry as defined by six different parameters: alpha angle, femoral inclination, femoral version, acetabular version, Lateral Centre-Edge (LCE) angle and acetabular coverage. An experienced user manually placed the five required landmarks medial and lateral femoral condyles, medial and lateral femoral epicondyles and the pubic tubercle by indicating the points on three orthogonal slices of the CT. As the landmarks were placed on the CT image

and not on the segmentation, the landmarks remained the same for both the manual and automatic method. A visual representation of the measurement of each of the parameters is shown in Appendix C.

Table 3.2: Image analysis metrics for the different input configurations of the network after training. The best performing input configuration is shown in bold for each metric.

Input	Dice		MASD		HD95		HD	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
32	0.958	0.024	0.97	1.31	5.64	11.04	19.5	20.2
64	0.969	0.014	0.41	0.13	1.15	0.91	11.2	8.5
128	0.967	0.015	0.46	0.17	1.38	1.37	12.7	8.2
64-32	0.968	0.014	0.40	0.09	1.00	0.26	8.1	9.7
128-32	0.970	0.014	0.37	0.05	0.90	0.12	5.5	7.1
128-64	0.970	0.015	0.38	0.10	0.95	0.33	5.9	10.0
128-64-32	**0.971**	0.013	**0.36**	0.06	**0.87**	0.12	**5.3**	8.0
256-128-64-32	0.969	0.014	0.40	0.11	0.98	0.45	7.2	10.4

3.3 Results

3.3.1 Cascaded V-net optimization

The 3D models generated from the automatic segmentations visualized in Figure 3.4 show the different segmentation results for one subject. The detailed comparison in Figure 3.5 shows that most segmentation errors appear close to the joints, and in places where unlabelled bones were present such as the spine, patella and small bones in the feet. This example showed that a small receptive field of the network negatively affected its ability to discern between bones. The DSC, MASD, HD95 and HD reported for each network configuration in Table 3.2 reflect these observations. One-way ANOVA analysis was performed and showed that only the network with a 32^3 voxel input was found to have a statistically different mean from the other configurations for any of the metrics. (DSC: $p = 8 \times 10^{-7}$, MASD: $p = 3.2 \times 10^{-16}$, HD95: $p = 3.5 \times 10^{-17}$, HD: $p = 5.2 \times 10^{-12}$).
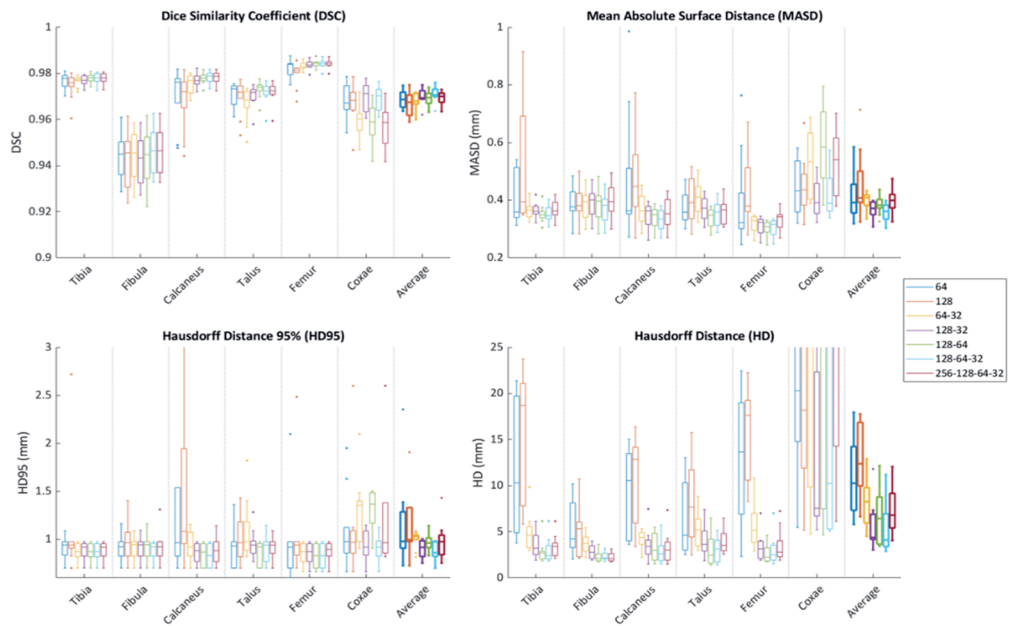
Figure 3.6: Segmentation similarity metrics calculated on the automatic segmentation versus the manual segmentation for each bone, for each network architectures. Average values indicate the metric averaged over the bones. The 32³ input configuration was left out as its results deviated too much to be able to represent them clearly in the figure.

The distribution of the segmentation metrics for each bone in Figure 3.6 show that the networks generally performed better in the larger and wider bones such as the femur and tibia, and worse in the thinner bones such as the fibula or coxae. As can be seen from the table and the figure, the input configuration with 128^3-64^3-32^3 voxel FOVs attained the highest DSC and lowest MASD, HD95 and HD, indicative of best performance. This input configuration was used to train the network in five-fold cross validation, to provide automatic segmentations for the final segmentation and hip morphometry assessment.

3.3.2   Five-fold cross-validation

The results of the five-fold cross validation were compared to the best results of other automatic segmentation methods for which the MASD or HD was reported in the literature[81,100,102–110]. The proposed method achieved lower MASD than other methods for most bones except the coxae (Table 3.3). The MASD was also lower for most bones when compared to the interobserver MASD for manual segmentations that were reported by Audenaert et al. [81] Furthermore, The HD of the proposed method was also lower than those reported in the literature, in general, except for the coxae and talus. The average DSC over all bones was 0.98 ± 0.01 and the HD95 was 0.65 ± 0.13 mm.
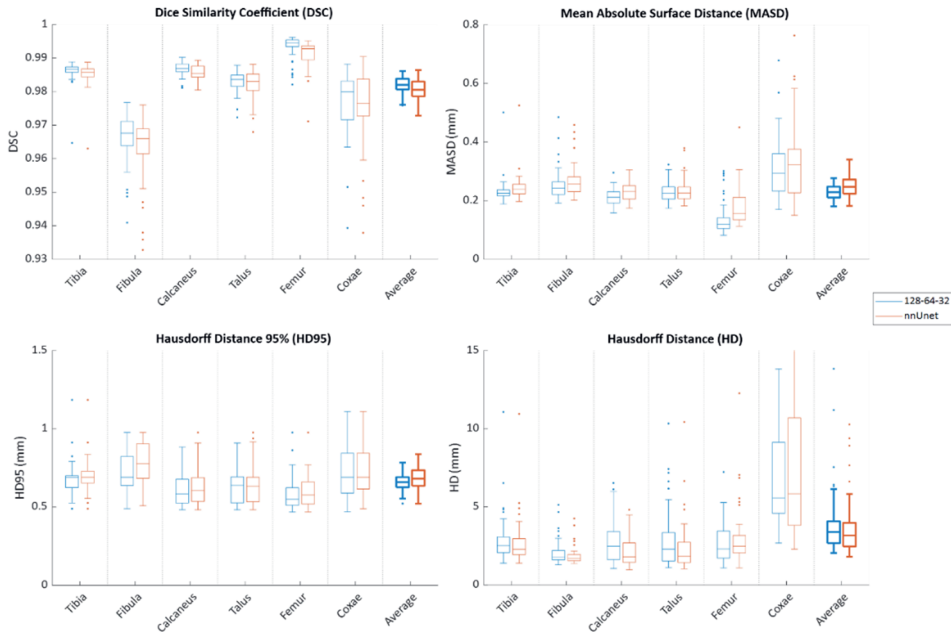
Figure 3.7: Segmentation similarity metrics calculated on the five-fold cross-validated automatic segmentation versus the manual segmentation for each bone, for the 128-64-32 implementation of the proposed network, compared with nnU-net.

3.3.3 Comparison to nnU-net

Figure 3.7 shows the results of the comparison of proposed cascaded network and the cascaded version of nnU-net run on the UMCU dataset. The two-tailed paired t-test was performed for each metric to check for significant differences between the average results of nnU-net and the proposed method. Bonferroni correction was applied to compensate for multiple testing, resulting in $p < 0.0125$ to be considered significant. The proposed method performed significantly better than nnU-net for the mean DSC ($p = 9 \times 10^{-4}$), MASD ($p = 7.8 \times 10^{-5}$) and HD95 ($p = .0048$). For the HD no significant difference was found ($p = 0.34$).

The two stage cascaded nnU-net necessitated separate training for both stages of the network. Each iteration took on average 514 seconds, for an average of 143 hours of training for each of the five folds on our system, resulting in a total of approximately 60 days of GPU time. The training of the proposed network on the same system took only 24 hours per fold, and as all stages of the cascaded V-net were imbedded in one network, the total training for all folds took approximately 5 days of GPU time. Inference using nnU-net took on average 55 minutes per scan, whereas the proposed network only needed 20 minutes per scan. For inference, nnU-net requires at least 4GB of GPU memory, while the proposed approach requires less than 1GB.

Table 3.3: Comparison of the MASD and HD between the automatic and segmentations performed on the UMCU and NMDID datasets with the best automatic segmentation methods found in the literature. Values are shown as mean ± standard deviation. Values performing better than found in the literature are shown in **bold**.

*: Mean and standard deviation of combined torso and lower leg CT.

**: Standard deviation unknown.

	Proposed Method UMCU dataset		Proposed Method NMDID dataset		Literature		Literature Interobserver[17]	
	MASD (mm)	HD (mm)	MASD (mm)	HD (mm)	MASD (mm)	HD (mm)	MASD (mm)	HD (mm)
Tibia	**0.23±0.04**	**2.91±1.53**	**0.48±0.05**	5.08±2.36	0.63±0.11[17]	4.07±2.15[17]	0.39±0.19	1.88±0.47
Fibula	**0.25±0.05**	**2.04±1.59**	**0.43±0.05**	3.96±2.17	0.76±0.76[17]	3.76±1.17[17]	0.61±0.08	2.25±0.71
Calcaneus	**0.21±0.03**	**2.73±1.32**	0.60±0.06	5.48±1.99	0.53±0.16[17]	2.90±0.77[17]	0.40±0.12	1.67±0.34
Talus	**0.23±0.03**	3.00±1.95	0.63±0.11	5.47±3.16	0.57±0.12[17]	2.97±0.59[17]	0.44±0.08	2.21±0.41
Femur	**0.14±0.06**	**3.54±5.54**	0.58±0.07*	5.03±3.20*	0.2±0.1[15]	4.34±0.86[42]	0.41±0.15	2.30±0.98
Coxae	0.31±0.10	9.22±8.56	0.56±0.13	9.17±5.43	0.30±0.5[15]	5.5[46]**	0.41±0.20	3.74±2.68

3.3.4   HKA measurement

The agreement between measurements on manual and automatic segmentations was calculated using the two-way mixed effects, absolute agreement, single measurement Intraclass Correlation Coefficient, or ICC(3,1)[111,112]. For the HKA, the ICC was 0.976, indicating excellent reliability (ICC > 0.90), and the Mean Absolute Difference (MAD) was $0.18 \pm 0.36°$. These values are comparable with the interrater agreements reported by Jud et al.[95], who reported an ICC of 0.988 and a MAD of $0.4 \pm 0.5°$. The difference in landmark placement in the hip was 0.19 ± 0.16 mm, in the knee 1.2 ± 1.6 mm and in the ankle 0.50 ± 0.49 mm.

Table 3.4: Comparison of the mean morphometric parameters found using the automatic and manual segmentations. The interobserver variability is also shown to compare the range between observers and the manual and automatic segmentations.
*: 95% confidence interval

Parameters	Manual Segmentation		Automatic Segmentation		Literature	
	Mean	SD	Mean	SD	Mean	SD
Alpha Angle (°)	46.2	13.9	46.3	14.0	45.6[99]	10.5[99]
Femoral Inclination (°)	129.5	5.2	129.7	5.4	129.2[99]	6.2[99]
Femoral Version (°)	10.2	10.1	9.7	10.1	9.7[99]	9.3[99]
Acetabular Version (°)	17.8	7.3	17.6	6.7	19[113]	4[113]
Acetabular Coverage (%)	74.9	6.2	74.5	6.1	73[96]	4[96]
LCE Angle (°)	32.0	8.5	31.3	8.2	33.6[114]	18.1–48.0*[114]

3.3.5   Hip morphology assessment

Table 3.4 shows the mean and standard deviations of all hip morphometric parameters compared to the mean and standard deviation of each parameter as they are found in the healthy adult population according to Toogood et al.[99] (alpha angle, femoral neck inclination, femoral version), Dandachli et al.[96] (acetabular coverage), Tannast et al.[98] (LCE angle) and Hingsammer et al.[113] (acetabular version). Comparison shows that the results correspond closely to the healthy population, for both the manual and automatic segmentation.

In Table 3.5 the mean and mean absolute difference of all hip parameters are shown. The ICC(3,1) for the alpha angle was 0.822 indicating good reliability (0.75 < ICC < 0.90), while the other parameters had an ICC of over 0.967, indicating excellent reliability (ICC > 0.90) [115]. Furthermore, in Table 3.5 the differences were compared to the results of the studies of Zeng et al.[100] and Chu et al.[103], who performed similar automatic segmentations for hip morphological parameter measurements. The proposed method achieved comparable or lower differences in measurement between manual and automatic workflows. Finally, comparison of the ICC with the manual interrater ICC by Harris-Hayes et al.[116] showed that the proposed method also performed comparable or better.

Table 3.5: Differences and ICC(3,1) between hip morphometry parameters when calculated from manual and automatic segmentations. The right side of the table shows the differences
*: Based on MRI instead of CT segmentations.

Parameters	Automatic and manual difference					Automatic and manual differences from literature		Manual interrater reliability
	Mean	SD	Abs. Mean	Abs. SD	ICC	Abs. Mean	Abs. SD	ICC(2,1)[53]
Alpha Angle (°)	-0.27	8.1	2.5	7.2	0.822	-	-	0.78-0.86[54]*
Femoral Inclination (°)	-0.3	1.2	0.72	0.95	0.986	2[15]*-2.1[40]	2[15]*-1.6[40]	0.96[54]*
Femoral Version (°)	-0.1	0.9	0.55	0.71	0.998	1[15]*-2.0[40]	1[15]*-1.5[40]	0.97[54]*
Acetabular Version (°)	0.02	1.8	0.95	1.5	0.983	-	-	0.94[54]*
Acetabular Coverage (%)	0.4	1.5	0.94	1.1	0.986	2[15]*-3.5[40]	1[15]*-2.3[40]	-
LCE Angle (°)	0.7	2.1	1.4	1.6	0.967	2[15]*	2[15]*	0.86[54]*

Figure 3.8 shows the Bland-Altman plots of all hip morphometry parameters. The femoral inclination and femoral version passed the Kolmogorov-Smirnov test for normality. In this case, the Limits of Agreement (LoA) were calculated as the mean +1.96 and -1.96 times the standard deviation. The alpha angle, acetabular version, acetabular coverage and LCE angles did not pass the test for normality and thus the 2.5th and 97.5th percentile were used to indicate the non-parametric limits of agreement. These plots show that although the correlation between measurements was generally high, a few large outliers did occur, especially in the alpha angle measurements.

3.3.6   NMDID dataset
Results of the best performing network retrained on all 50 scans from the UMCU dataset and evaluated on the 10 scans of the NMDID dataset are shown in Table 3.3.

3.3.7   Comparison to nnU-net
Figure 3.7 shows the results of the comparison of proposed cascaded network and the cascaded version of nnU-net run on the UMCU dataset. The two-tailed paired t-test was performed for each metric to check for significant differences between the average results of nnU-net and the proposed method. Bonferroni correction was applied to compensate for multiple testing, resulting in $p < 0.0125$ to be considered significant. The proposed method performed significantly better than nnU-net for the mean DSC ($p = 9 \times 10^{-4}$), MASD ($p = 7.8 \times 10^{-5}$) and HD95 ($p = .0048$). For the HD no significant difference was found ($p = 0.34$).

The two stage cascaded nnU-net necessitated separate training for both stages of the network. Each iteration took on average 514 seconds, for an average of 143 hours of training for each of the five folds on our system, resulting in a total of approximately 60 days of GPU time.

The training of the proposed network on the same system took only 24 hours per fold, and as all stages of the cascaded V-net were imbedded in one network, the total training for all folds took approximately 5 days of GPU time. Inference using nnU-net took on average 55 minutes per scan, whereas the proposed network only needed 20 minutes per scan. For inference, nnU-net requires at least 4GB of GPU memory, while the proposed approach requires less than 1GB.

3.4 Discussion

The aim of this chapter was to develop a fast, lightweight algorithm for semantic bone segmentation from CT images and assess its applicability to automatic 3D hip morphometry parameters. Segmentation of the bones from CT images has widespread use in the orthopaedic clinic, especially for the measurement of hip morphometry parameters. Therefore, many methods have been developed to segment the proximal femur and coxae[100,103–108,117]. However, only few studies have attempted to semantically segment larger parts of the human body with a single method[81,110], which brings unique challenges such as computing time and memory constraints. In this chapter, we have shown that by using a cascaded network with a large receptive field we were able to achieve competitive semantic bone segmentation results while minimizing the time and memory requirements.

Our lightweight implementation of the V-net, with a lower number of convolutional layers, a smaller number of filters, and less down sampling layers and thus depth of the network, resulted in strongly reduced number of parameters. A single lightweight V-net without cascades had over 15 times less parameters than a similar V-net as described by Çiçek et al.[86] This allowed the network to run training and inference quickly and with relatively low memory consumption. The implementation of cascaded V-nets allowed the volume of the receptive field of the network to increase cubically with patch size, while only increasing the number of parameters and input voxels linearly.
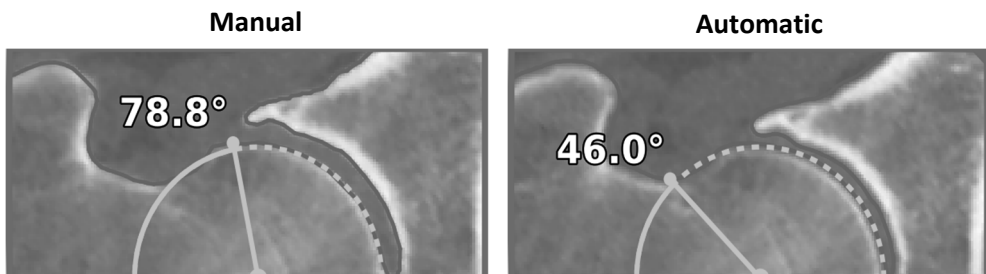

Figure 3.8: Example of 12 o' clock alpha angle calculation that is highly influenced by a small change in segmentation.

The proposed method achieved a low MASD and HD on the UMCU dataset compared to other studies, with on average sub-voxel accuracy. The proposed method did perform worse on the HD metric for the coxae. Qualitative inspection of the segmentation of the coxae showed that the HD was especially influenced by segmentation errors on the border of the image, in scans where only part of the coxae were visible. A comparison to the cascaded nnU-net applied to our dataset showed that our network was able to attain better or comparable results with lower training and inference time, and lower GPU memory requirements.

Our method performed slightly worse on the NMDID dataset, but still outperformed the MASD results found in the literature on the tibia and fibula. The decreased accuracy can probably be attributed to significant differences between the UMCU and NMDID dataset. The subjects of the NMDID dataset were purposely chosen to be different in age than the UMCU patients, to test the robustness of the algorithm to different population demographics. Additionally, UMCU patients were alive while the NMDID dataset consists of recently deceased subjects, which may have introduced post-mortem effects such as gas build-up in the lower abdomen[118]. The NMDID patients were also mostly scanned clothed, which introduced artefacts due to interference with objects on the body. Finally, the scanner and scanner settings differed. In future research, using a more diverse dataset for training might improve the robustness and thus general applicability of the network.

Excellent correspondence was found between the measurement of the HKA on the manual and automatic segmentations. A good or excellent correlation was also found for the hip morphometry parameters between the manual and automatic segmentations. Moreover, the correlation was higher than the manual interrater reliability reported by Harris-Hayes et al.[116] This is in accordance with the results of Zeng et al.[100], who also used a deep learning method to automatically segment the hip and calculate the morphometry parameters. In contrast to their study however, we segmented six bones instead of two, with a larger field of view that comprised the complete lower extremities. It should be noted that the studies by Zeng et al. and Harris-Hayes et al., with which our results were compared, were performed on MRI scans instead of CT scans. However, these represented the best results that could be found in the literature for comparison of automatic and manual hip morphometry assessment interrater reliability.

3.4.1 Limitations

While the proposed methodology, just like nnU-net, is in principle generically applicable to other segmentation tasks, it has been evaluated on the specific challenge of bone segmentation in 3D CT, albeit with a large variety of osseous structures. Future work warrants more extensive evaluations to investigate the potential benefits of this approach to other segmentation challenges in different applications and on different data. In addition, this chapter focused on optimizing the number of U-nets in a cascaded approach with a fixed

combination of other hyperparameters. An investigation of the optimal hyperparameters for each of the different network topologies might yield different results than presented here.

Although our segmentation method achieved good results relative to other state-of-the art methods in the literature, care should be taken when comparing different studies. Each study reported on separate datasets with differing parameters such as inclusion criteria, scanner settings and voxel sizes.

We assumed that the morphological parameters that were calculated using the commercial software tool gave the correct results when presented with the manual segmentation. However, in some cases the manual segmentations included small spurious bone voxels on the femoral head, which increased the alpha angle more than would be clinically expected. An example of this is shown in Figure 3.9. These erroneous morphometry measurements were a limitation caused by our automatic approach, as these would normally be corrected manually when using the commercial software.

3.5 Conclusion

In this chapter, we introduced a method for fully automatic lower extremity segmentation from CT. By using a cascaded V-net approach, it was possible to use information from a large receptive field, while maintaining a comparatively low computation time and GPU memory footprint. Comparison to state-of-the-art methods found in other studies showed that the segmentation accuracy also performed competitively. Furthermore, this study was the first to apply cascaded V-net based segmentation on lower extremity CT data, and validate the applicability to orthopaedic diagnosis using clinical implementations, i.e. HKA and hip joint measurements. Comparing the HKA and hip morphology between automatic and manual segmentations, the metrics showed good or excellent correlation, indicating that this method could be a valuable addition to many orthopaedic applications that benefit from accurate bone segmentation.

3.6 Appendix A: Training and Inference

3.6.1 Training

Each model was trained performed with 16 patches per batch, with 128 batches per epoch with the Adam optimizer[68] to minimize the cross entropy loss between the output of the network and the ground truth segmentations. Each model was trained for 1000 epochs.The loss was calculated by calculating the cross entropy loss of the output of the V-nets at each level, and then averaging over these losses. Retrospective inspection of the training and validation loss showed no increase or decrease at this point in the training for any of the input configurations. The learning rate started at 10^{-3} for the first 20 epochs and was then linearly decreased to 10^{-6}.

Data augmentation of the patches was applied by randomly resizing the input patch with a factor between 0.9 and 1.1, and to all smaller input patches correspondingly. Further augmentation was done by randomly scaling image intensity with a factor between 0.95 and 1.05. To counteract imbalance in the dataset between foreground (bone) and background patches, in each batch approximately 70% foreground and 30% background patches were selected during training. A patch was considered a foreground patch if the central voxel of the patch belonged to any bone segmentation. Finally, patches were randomly flipped along the sagittal axis. The deep learning network was implemented using PyTorch 1.7.0[119]. Training and testing of all networks was performed on a system with 256 GB RAM, 32 CPU cores and two 12 GB GeForce GTX 1080 Ti (NVIDIA, Santa Clara, CA) GPUs. All hyperparameter settings used for training and inference were based on a manual search performed with different combinations of the parameters while using the 64^3 input configuration. The same set of data was used for training and validation; the test set was kept separate.

3.6.2 Inference

The sampling strategy often used when performing inference with cascaded[88–91] or atrous[90,92] convolutional networks is to predict the centre patch of each larger FOV patch. Without image padding, this would potentially lead to part of the segmentation not being included in the prediction. If padding was used, this would lead to even larger image sizes. Additionally the available information used during prediction would not be optimally used. For example, for a predicted patch in the corner of a rectangular cuboid, only a small amount of the voxels that are used for prediction actually come from the image, with a large amount originating from the padding, which has no information that could be used for prediction.
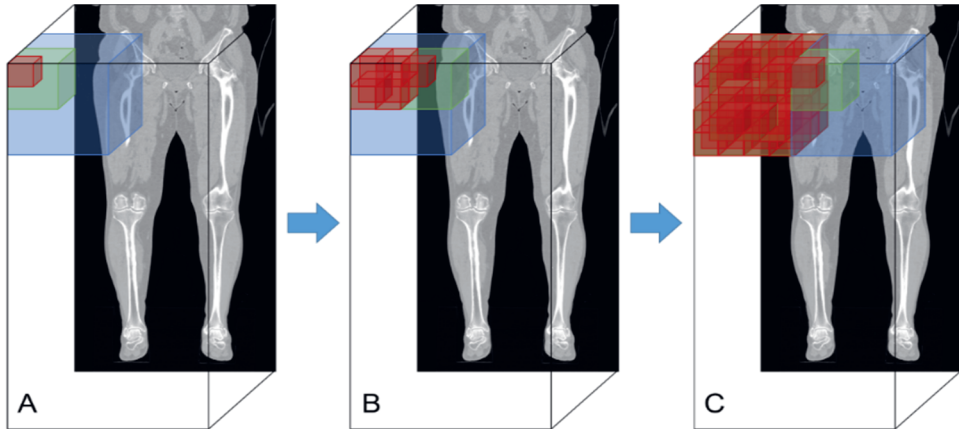
Figure 3.9: Demonstration of the inference sampling strategy for a network with input at three resolutions. During each inference step, only the volume of the small red patch is predicted. A: The first patch (red) is predicted. As long as the location of the high resolution patches remains within the boundaries of the second resolution (green), the lower resolution patches remain the same. B: After all high-resolution patches within the second resolution have been predicted, sampling moves on to the next patch of the second resolution. C: After all high-resolution patches within the third (blue) resolution have been predicted, sampling moves on to the next patch of the third resolution. This process continues until each voxel has been predicted with the smallest (red) patch.

A different sampling strategy was therefore used where no padding was required, as is demonstrated in Figure 3.10. The first V-net, with the largest FOV, predicted one cubed patch in the corner of the image. Thereafter, the consecutive V-net gave a prediction for a smaller cubed patch in the corner of the first patch. This process was repeated until the input patch with the smallest FOV was processed. The smallest FOV patch then shifted its position, with a certain stride, until all voxels in the FOV above it were predicted. Thereafter, the position of the patch above it, and the process was repeated until all voxels in the image had at least one prediction. Using this sampling strategy, the full image can be segmented, while each prediction is guaranteed to have only used information that is present in the original image.

During inference, the stride of the largest FOV was set to half the size of the FOV. The strides of the smaller FOV patches was set to the size of those patches. To reduce the amount of patches that needed to be predicted and speed up inference, patches were only used for inference if at least 2% of the full resolution patch consisted of voxels with Hounsfield Units (HU) higher than 0, the radiodensity of water. Due to this, background patches with only air were not used for prediction. As a post-processing step, all connected components in the scans smaller than 1000 voxels were removed.

3.7 Appendix B: Automatic HKA measurement

The Hip-Knee-Ankle (HKA) angle is an important metric to quantify the degree of varus or valgus alignment in the knee[2]. It measures the angle between the mechanical axes of the femur and the tibia in the coronal plane. Fürnstahl et al.[3] described a method to measure the HKA by calculating the angle between two lines drawn between three landmarks. These landmarks were the Hip Joint Center (HJC), the Knee Joint Center (KJC) and the Ankle Joint Center (AJC). The HJC was defined as the center of a sphere, fit to the femoral head using least square regression. The midpoint of the top of the two tibial intercondylar eminences was used as the KJC. The AJC was found by taking all vertices of the 3D models of the tibia and fibula that were within 3-5 mm of the talus, and then averaging their locations. The HKA was then defined as the angle between a line drawn from the KJC to the HJC, and from the KJC to the AJC, projected onto the coronal plane.

The process described by Fürnstahl et al. included three manual steps:

1. HJC: The separation of the femoral head from the femoral neck, in order to fit the sphere.
2. KJC: The top of the intercondylar eminences had to be indicated manually.
3. AJC: The distance from the talus that was used needed to be set manually to a value in the range 3-5 mm.

To be able to extract only the differences caused by segmentation and avoid differences caused by intra- or interrater variability it was necessary to use an entirely automatic method to set the landmarks. Therefore, we introduced a number of automatic steps and fixed some of the variables so no manual interaction was needed:

1. HJC: The voxelwise representation of the segmentation of each femur was reduced to a line with a width of a single voxel using the morphological skeletonization operation. The top of this line, which corresponded to the approximate center of the femoral head, was then taken as the first point. The second point was set at 20 mm further along the line. A line was drawn between the two points which was used to construct a plane perpendicular to the direction of the line, intersecting the line in the middle. This plane was then used to separate the femoral head from the femoral neck, such that the sphere could be fit using least fitting squares regression.
2. KJC: The first tibial eminence was defined as the most proximal point on the tibia. The second tibial eminence was defined as the second most proximal point, with the constraint that the distance between the points was smaller than 10 mm in the sagittal plane, and the total distance was smaller than 20 mm. Otherwise, the following most proximal point was chosen. The KJC was then chosen to be average of the two points. If there was only one distinctive eminence, this point was chosen as the KJC.

3.   AJC: The normally user-defined distance between 3-5 mm was fixed at 4 mm.

An example of this method is shown in Figure 3.11. All automatically placed landmarks were manually checked, but not altered, to ensure that their positioning was correct. It should be noted that although these methods worked well for this dataset, a difference in orientation of the patient or more severe deformations of the bone could result in erroneous positioning of the landmarks. Also, we considered the mechanical axis of the legs to be positioned aligned with the Z-axis of the coordinate system, perpendicular to the coronal plane. Finally, the CT scans were not take in a weight-bearing state, as the patients were in supine position. As Jud et al. found, this might influence the HKA of the patient[95].
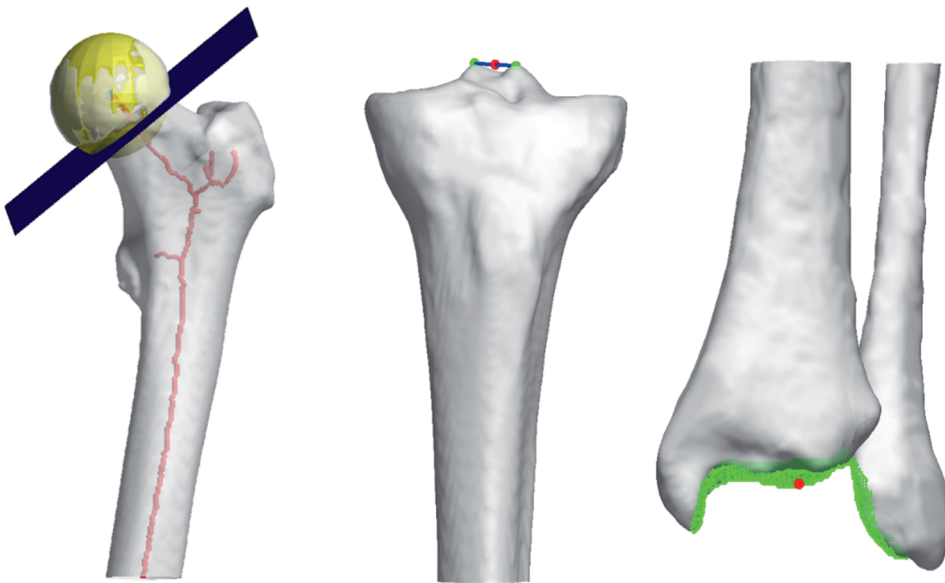
Figure 3.10: Example of the automatic determination of the center of the hip, knee and ankle joint.
Hip*: The red line indicates the skeletonized femur. The plane is drawn perpendicular to the top section of the red line, and separates the femoral head from the neck. The yellow sphere was fitted to the head using least squares regression. The HJC is located in the center of the sphere.*
Knee*: The green points indicate the tops of the tibial intercondylar eminences. The KJC is located in the middle of these two, indicated by the red point.*
Ankle*: The green points indicate vertices on the tibia and fibula. The AJC is located at the average location of the green points, indicated by the red point.*

3.8 Appendix C: Hip morphometrics
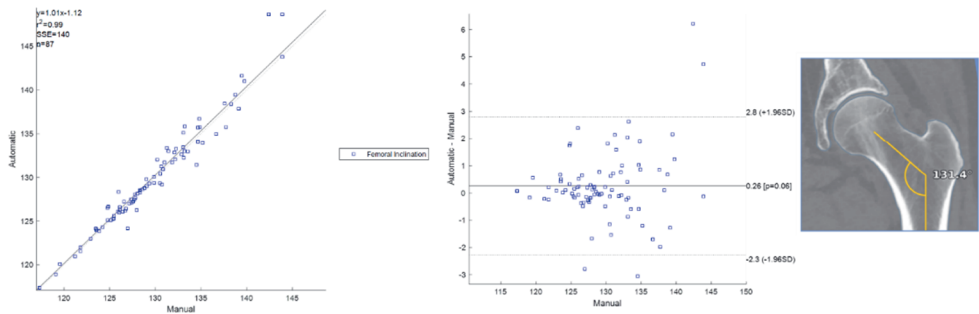
*Figure 3.11: **Femoral Inclination.** From left to right: Correlation plot. Bland-Altman plot. Visualization of an exemplary measurement of this metric.*
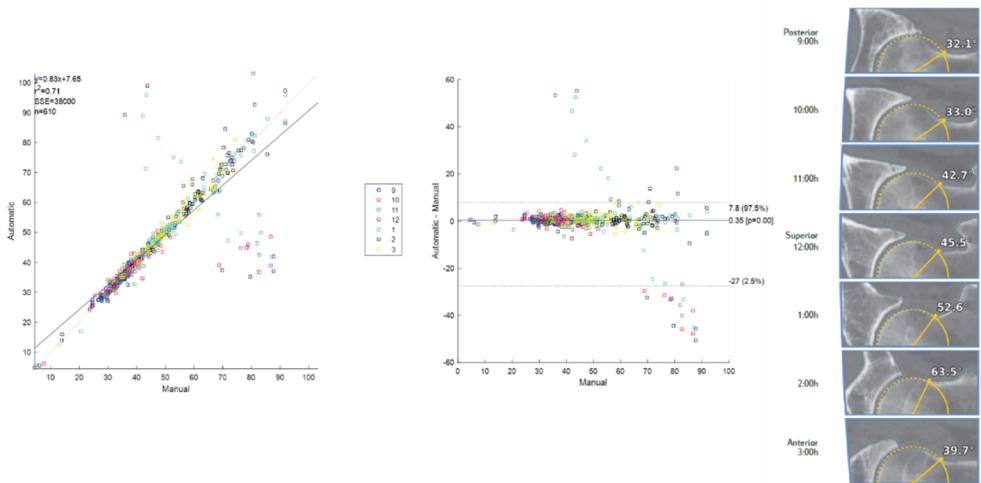
*Figure 3.12: **Alpha Angles.** From left to right: Correlation plot. Bland-Altman plot. Visualization of an exemplary measurement of this metric. Different colours in the plot indicate the different measurements by hour, as shown in the example.*
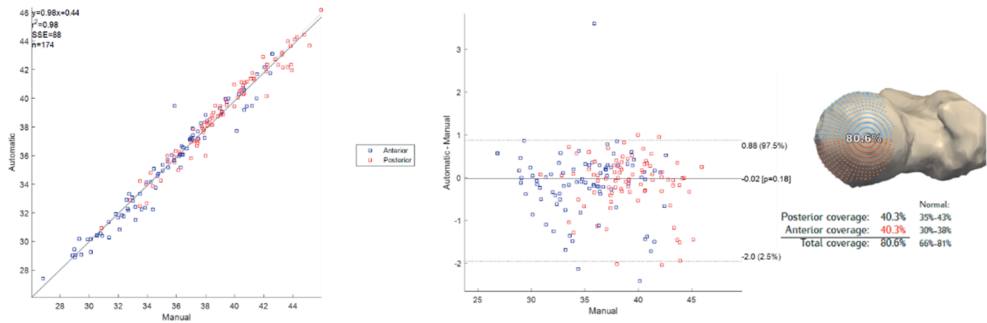
*Figure 3.13: **Acetabular Coverage.** From left to right: Correlation plot. Bland-Altman plot. Visualization of an exemplary measurement of this metric. Different colours in the plot indicate the posterior and anterior coverage, as shown in the example*
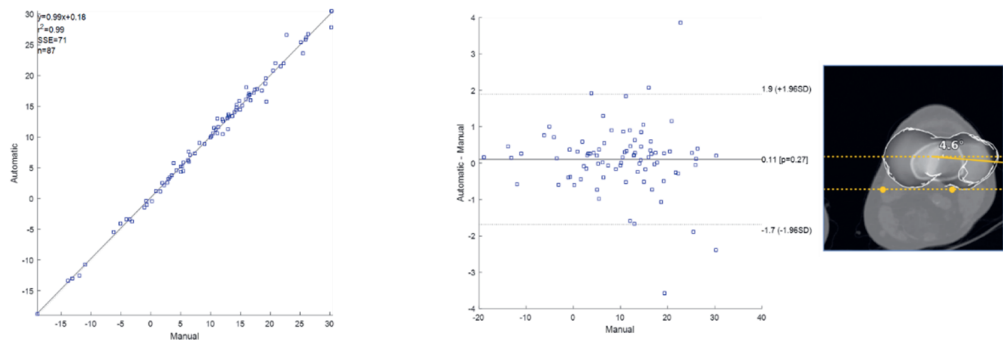
*Figure 3.14: **Acetabular Version.** From left to right: Correlation plot. Bland-Altman plot. Visualization of an exemplary measurement of this metric. Different colours in the plot indicate the acetabular version measured at the centre and at the upper part of the femur.*
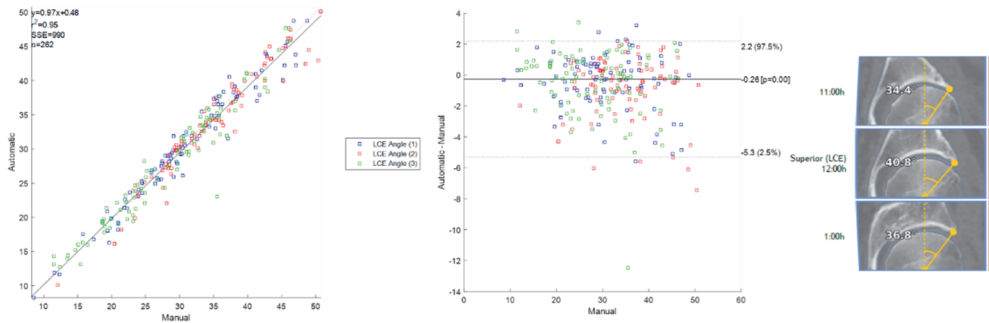
*Figure 3.15: **Lateral Centre-Edge (LCE) Angle.** From left to right: Correlation plot. Bland-Altman plot. Visualization of an exemplary measurement of this metric. Different colours in the plot indicate the different measurements by hour, as shown in the example.*
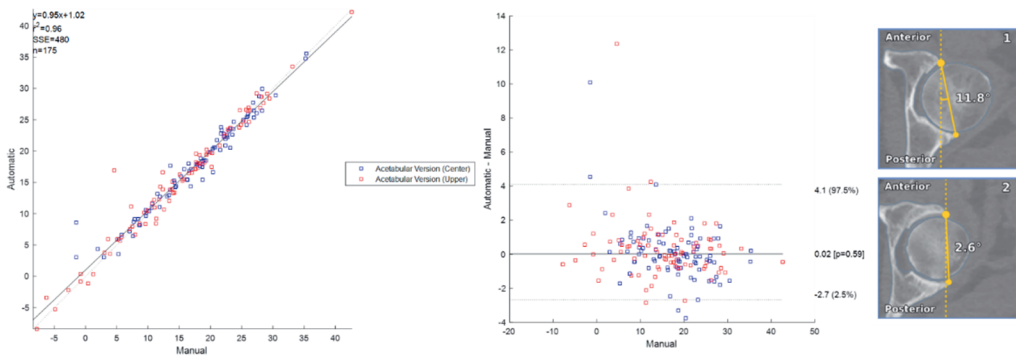
*Figure 3.16: **Femoral Version.** From left to right: Correlation plot. Bland-Altman plot. Visualization of an exemplary measurement of this metric.*
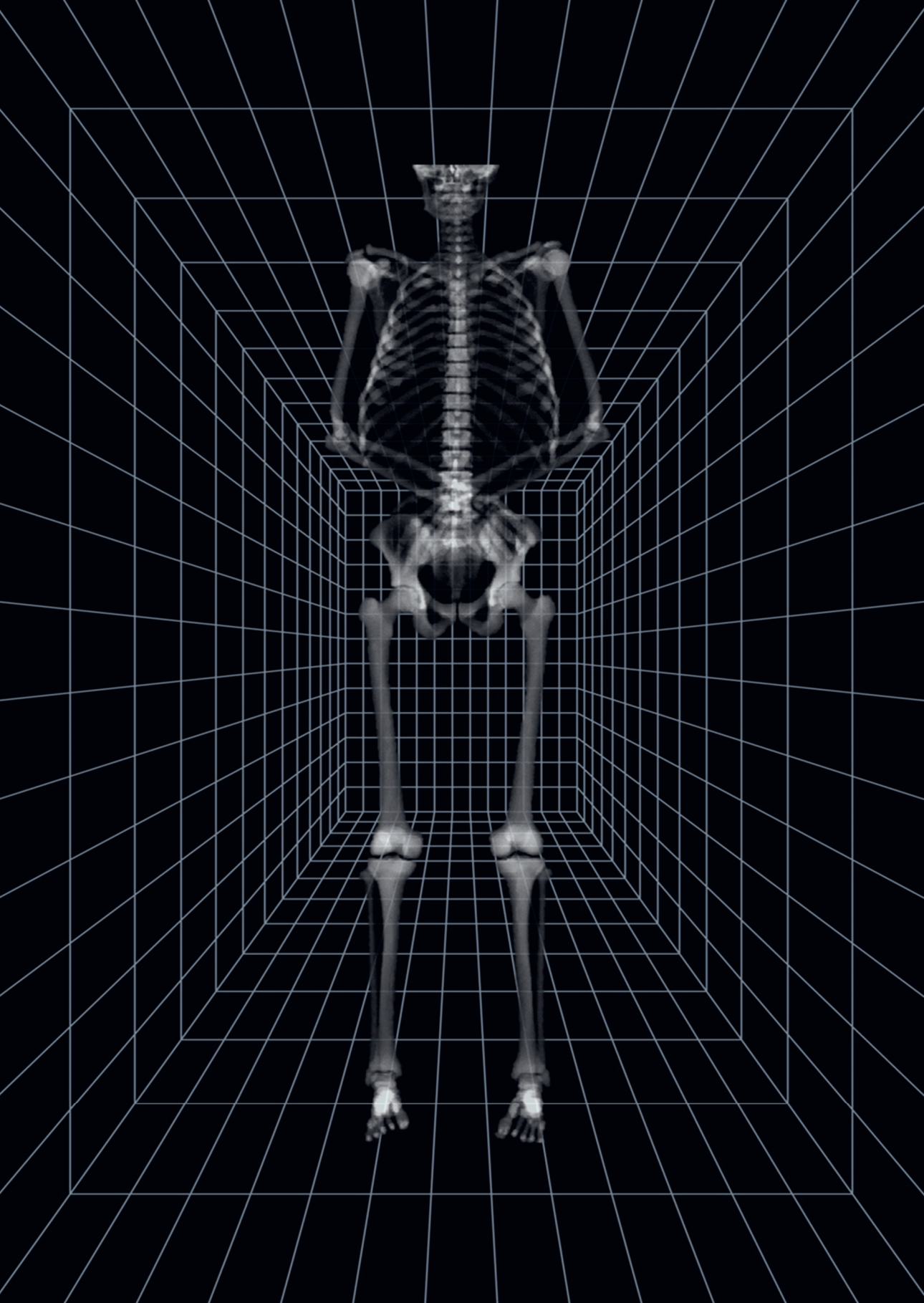
Chapter 4

Impact of bone and cartilage segmentation from CT and MRI on both bone forearm osteotomy planning

Abstract

Introduction: The use of MRI scans for pre-operative surgical planning of forearm osteotomies provides additional information of joint cartilage and soft tissue structures and reduces radiation exposure in comparison with the use of CT scans. In this study, we investigated whether using 3D information obtained from MRI with and without cartilage information leads to a different outcome of pre-operative planning.

Methods: Bilateral CT and MRI scans of the forearms of 10 adolescent and young adult patients with a unilateral bone deformation were acquired in a prospective study. The bones were segmented from CT and MRI, and cartilage only from MRI. The deformed bones were virtually reconstructed, by registering the joint ends to the healthy contralateral side. An optimal osteotomy plane was determined that minimized the distance between the resulting fragments. This process was performed in threefold: using the CT and MRI bone segmentations, and the MRI cartilage segmentations.

Results: Comparison of bone segmentation from MRI and CT scan resulted in a 0.95 ± 0.02 Dice similarity coefficient and 0.42 ± 0.07 mm mean absolute surface distance. All realignment parameters showed excellent reliability across the different segmentations. However, the mean differences in translational realignment between CT and MRI bone segmentations (4.5 ± 2.1 mm) and between MRI bone and MRI bone and cartilage segmentations (2.8 ± 2.1 mm) were shown to be clinically and statistically significant. A significant positive correlation was found between the translational realignment and the relative amount of cartilage.

Conclusion: This study indicates that although bone realignment remained largely similar when using MRI with and without cartilage information compared to using CT, the small differences in segmentation could induce statistically and clinically significant differences in the osteotomy planning. We also showed that endochondral cartilage might be a non-negligible factor when planning osteotomies for young patients.

4.1 Introduction

Radial and ulnar bone malunion after trauma can lead to reduced range of motion, chronic pain, and loss of strength[120,121]. Studies have shown that these complications can be effectively treated with realignment of the bones by a corrective osteotomy[122–124]. Corrective osteotomies are conventionally planned and assessed manually on (biplanar) two-dimensional (2D) radiographs[125]. However, recent studies have shown that computer-assisted preoperative planning based on three-dimensional (3D) Computed Tomography (CT) scans significantly improved both functional and radiographic outcome[125–128].

3D computer-assisted osteotomy planning is mostly performed using CT scans, as the unique intensity range of bone structures on this modality makes segmentation relatively straightforward[129]. However, it is difficult to distinguish between different soft tissue structures on CT, and CT scanning involves harmful radiation. In contrast, osteotomy planning using Magnetic Resonance Imaging (MRI) does not involve harmful radiation and soft tissue structures can be more clearly distinguished. Including soft tissues structures like joint cartilage might improve the outcome of preoperative planning. However, most MRI
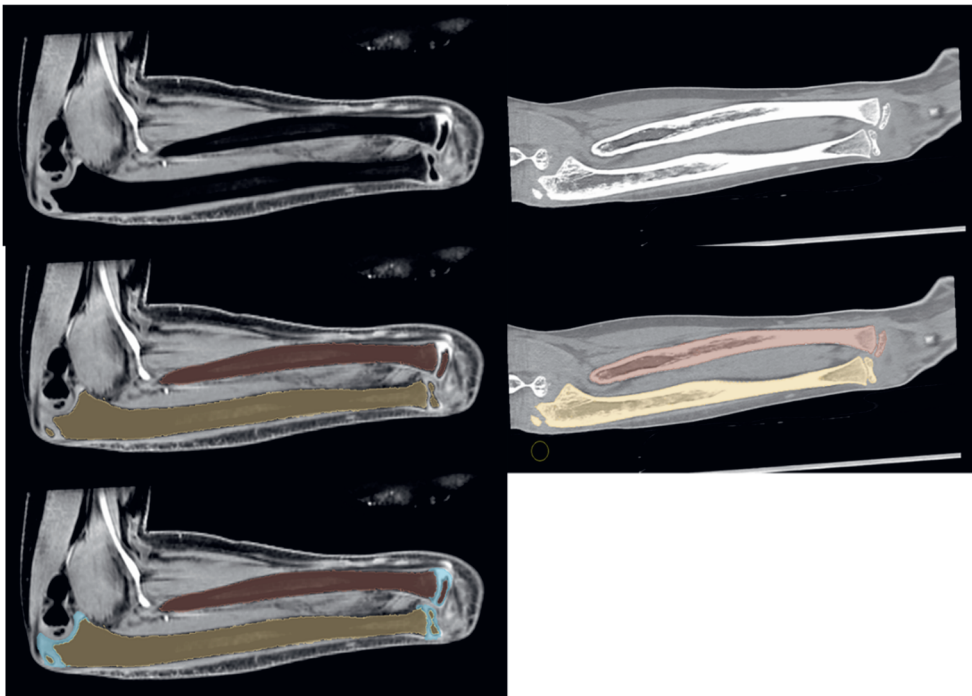
Figure 4.1: Top row: Example of the MR (left)and CT (right) images of lower right arm. Middle row: MR and CT overlaid with radius (brown) and ulna (yellow) bone segmentation. Bottom row: MR overlaid with bone and cartilage (blue) segmentation.

protocols do not result in a unique intensity range for bone and can suffer from intensity inhomogeneity, which causes segmentation to be less trivial[130].

To be able to study differences in osteotomy planning based on CT and MRI, a deterministic workflow is essential to avoid the variation in human assessment. Various authors have proposed automatic osteotomy planning methods[20,129,131,132]. Based on these studies we developed a similar deterministic, pre-operative, osteotomy planning algorithm that simulates the cutting, repositioning, and reconstruction of the deformed bone based on its healthy contralateral counterpart. The only necessary inputs were the 3D models derived from the segmentations of either the CT or MRI scans. By using an automatic, computer-assisted approach, we were able to determine the effect of different segmentations from different modalities on the outcome of the osteotomy planning to answer the following questions: (1) Does automated osteotomy planning based on bone segmentations derived from CT scans yield significantly different results when performed on bone segmentations derived from MRI scans? (2) Does automated osteotomy planning based on bone segmentations derived from MRI scans yield significantly different results when performed on combined bone and cartilage segmentations derived from MRI scans? (3) Does the amount of cartilage in the joints have a significant correlation to the differences observed between planning on bone or combined bone and cartilage segmentations?

Table 4.1: Demographics of the patients and age at time of trauma, scan and the interval between the trauma and scan.

Patient	Sex	Side trauma	Age (years)		
			Trauma	Scan	Interval
1	F	Right	7.1	20.1	13.0
2	M	Left	7.6	13.9	6.3
3	M	Left	9.2	11.9	2.7
4	F	Left	9.7	10.7	1.0
5	M	Left	17.6	18.3	0.7
6	M	Left	14.0	22.9	8.9
7	F	Left	10.7	20.1	9.4
8	M	Left	13.7	14.6	0.9
9	M	Left	7.4	12.3	4.9
10	M	Right	14.8	16.6	1.8
Mean	-	-	**11.2**	**16.1**	**5.0**
Std.	-	-	**3.4**	**3.9**	**4.1**

4.2 Methods

Automated osteotomy planning was performed in threefold for each patient, to study the effect of different scanning modalities and tissue inclusion on the outcome. The planning was performed once using bone segmentation derived from CT, once using bone segmentations derived from MRI, and once using bone and cartilage segmentations from MRI. As cartilage was not discernible on the CT scans, cartilage segmentation was not performed on CT scans.

4.2.1 Data

The data was acquired prospectively at the Erasmus MC (Erasmus University Medical Center, Rotterdam, The Netherlands) under ethical approval of the Medical Ethical Testing Committee, reference number 52987.078.15. The research protocol for data acquisition was registered in the National Trial Register under reference number 6324 and in the ICTRP portal under reference number NL8059. Written informed consent was obtained from all participants and their parents. The inclusion process and criteria were earlier described by Roth et al. [128].

Original inclusion criteria for the study consisted of the following:

- A diaphyseal both-bone forearm fracture sustained during childhood (younger than 18 years).
- Minimum age of 10 years at time of injury.
- Pronation or supination range of motion of less than 50°.
- Unsatisfactory improvement after conservative treatment.

Exclusion criteria were:

- A traumatic osseous deformity of the contralateral forearm.
- Congenital or developmental deformity of the contralateral or affected forearm.

Diaphysis was defined as the segment of the bone between 20% and 80% of its entire length[133]. Of the 18 included patients, 8 patients had to be excluded from this study due to irregularities in the data such as movement during scanning, incomplete coverage of the region of interest or implant induced artefacts. Additionally, the CT scan of Patient 6 was incomplete and thus only the MRI of this patient were used. The demographics of the patients are shown in Table 4.1.

For each patient, CT and MRI scans were taken on the same day, of both forearms. The CT scans of both arms were taken using a single scan, with the patient in prone position and the shoulders and elbows in maximal extension (Superman position). The CT scans were acquired on a dual-source scanner plane (SOMATOM Force, Siemens Healthcare AG, Erlangen, Germany) with 0.6 mm axial slice thickness and pixel spacing ranging from 0.39 to 0.87 mm. The tube current was 89 mA with 120 kV tube voltage.

MRI scans were taken separately for each arm, with the arm positioned above the head and elbow in 90-degree flexion. MRI scans were acquired in a 3T scanner using a T1-weighted Multi-echo Gradient-echo sequence (GE Healthcare, Chicago, IL, USA). Dixon reconstruction was performed to obtain in-phase (IP), opposed phase (OP), fat only (F), and water only (W) images[43,44]. The acquisition time was 3 min 21 s with TR = 6.64 ms, TE = 2.97 ms, bandwidth = 325.5 Hz/pixel, flip angle = 15°, acquired voxelsize = $0.66 \times 0.66 \times 1$ mm^3. Both CT images and MRI images were resampled to isotropic 0.5x0.5x0.5 mm^3 voxels using trilinear interpolation.

4.2.2   Segmentation

The radius and ulna were manually segmented from the CT and MRI scans to produce two sets of 3D bone models for each patient. Additionally, articular cartilage was manually segmented from the MRI images to produce 3D cartilage models. One experienced biomedical engineer (RK) performed all segmentations. In the remainder of the text, the segmentation of bone from the CT and MRI scans will be referred to as CTb and MRb, and the segmentation of both bone and cartilage from the MRI scan will be referred to as MRbc. After segmentation, all bone models were transformed from a voxelwise representation into point-and-vertex based triangulated surface mesh.

Each bone was segmented separately from the CT scans using the open source medical imaging computing platform 3D Slicer[134]. First, a threshold of 200 Hounsfield Units (HU) was applied and the largest component corresponding to the bone of interest was manually selected. Additional manual editing was performed to fill holes in the outer shell of the bone or to remove spurious voxel segmentations caused by noise, hyper-intensities due to foreign materials, or connections between bones due to partial volume effects caused by low imaging resolutions. The medullary cavity of the bone was also filled such that a solid bone segmentation was acquired.

For segmentation of the MRI scans, only the water weighted Dixon reconstruction was used on which bone is visible as a low intensity region. Bias field correction was first performed using the N4ITK filter[135], available in 3D Slicer. The parameters used for the N4ITK filter were: mask image = none, BSpline grid resolution = 5x5x5, number of iterations = 50, 40, 30, convergence threshold = 0.0001, BSpline order = 3, shrink factor = 4. The separate bones on MRI scans were then coarsely segmented by using a local inverse Otsu-threshold[136] on a manually delineated area that included bone and soft tissue, but no background. The bone segmentation was refined by manually removing parts of the segmentation that were not part of the bone. The cartilage on the joint ends of each bone was then segmented entirely by manual delineation of the cartilage outline by an experienced operator, as there was not one clear intensity threshold that separated it from the cartilage of the adjacent bones and surrounding tissue.

Examples of the CTb, MRb and MRbc segmentation are shown overlaid on the CT and MRI scans in Figure 4.1. The 3D model derived from the MRb scan is shown in Figure 4.2 – step 0.

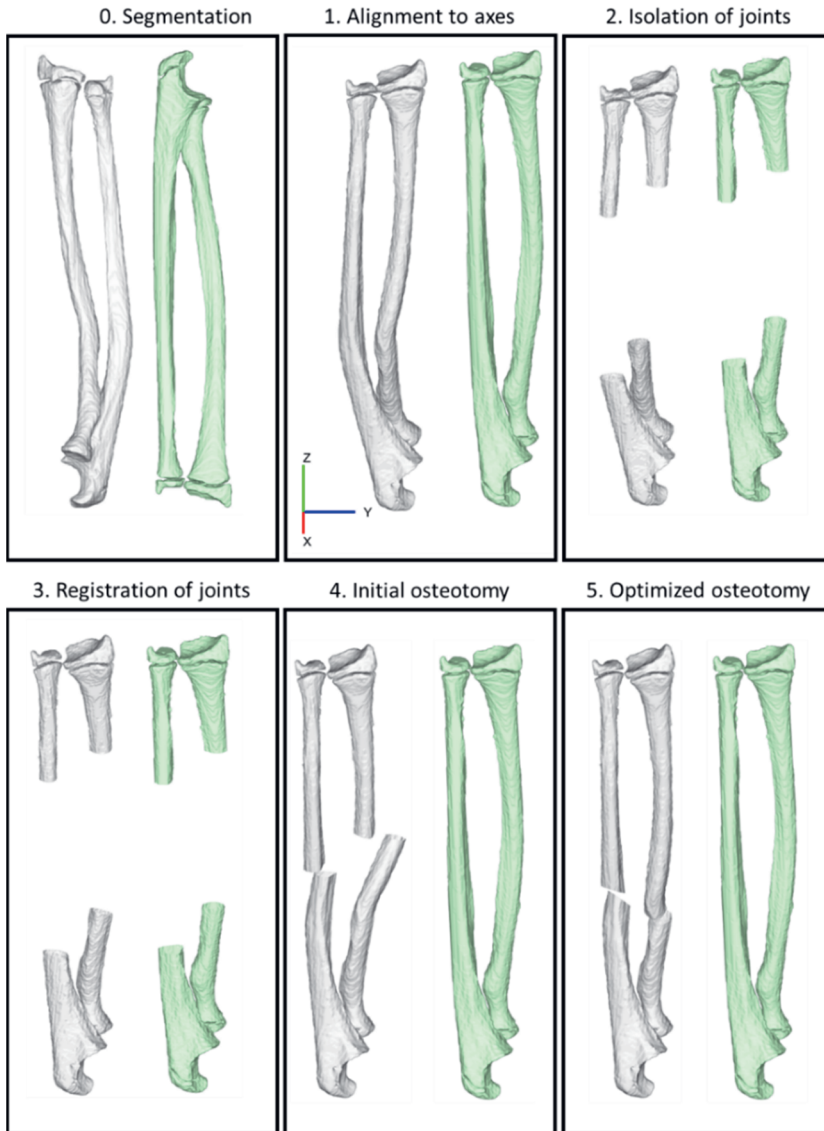

Figure 4.2: Step by step overview of the automatic osteotomy planning workflow. The deformed radius and ulna (white) were aligned to the mirrored healthy contralateral radius and ulna (green) by registering the joints and subsequently optimizing the osteotomy plane to minimize the distance between the proximal and distal bone fragments. This example used a bone segmentation derived from MR.

4.2.3 Automatic Planning

The automatic planning was performed using an in-house produced software tool written in Matlab (MATLAB 2020a, The MathWorks, Inc., Natick, Massachusetts, United States). It used the CTb, MRb and MRbc segmentations to perform the osteotomy planning in five automatic steps, which are summarized in Figure 4.2. Figure 4.2 – step 0 shows the original MRb segmentation of one of the patients before the osteotomy planning, before the deformed and healthy bone models were aligned.

First (Figure 4.2 – step 1), we aligned the CTb, MRb and MRbc models. To align the arms of all patients along approximately the same directions, the principal components of the combined point-clouds of the healthy CTb and MRb radius and ulna bone models were calculated. The healthy CTb and MRb bone models were then rotated such that the Z, X and Y axes of the patient were defined as the first, second, and third principal components respectively. The MRbc was rotated using the same transformation as the MRb, such that their relative orientation remained the same. Next, for each patient, the 3D models from the CTb radial and ulnar bones were separately registered to the same bones from the MRb using Iterative Closest Point[137] (ICP) matching. This ensured that the CTb, MRb and MRbc were all aligned similarly.

Second (Figure 4.2 – step 2), the proximal and distal part of the deformed and mirrored healthy contralateral bone were isolated. The proximal and distal joint were separated from the rest of the bone at a distance equal to 30% of the total length of each bone, measured from their respective joints. This prevented the site of the deformity to be taken into account in the reconstruction, while providing the registration algorithm with enough bone surface to converge to a realistic solution.

Third (Figure 4.2 – step 3), these deformed proximal and distal bone fragments of the radius and ulna were each registered separately to their counterparts on the healthy contralateral bone using ICP. This produced four 4x4 homogenous transformation matrices for each bone, M_{prox}^{radius} and M_{dist}^{radius}, for the distal and proximal radius, and M_{prox}^{ulna} and M_{dist}^{ulna}, for the ulna. For each reconstruction, optimal realignment of the proximal and distal joints to the mirrored contralateral side was set as a constraint. Due to this, an opening wedge osteotomy was necessary in case of bone lengthening, and a closing wedge osteotomy in case of bone shortening.

Fourth (Figure 4.2 – step 4), the original deformed bone model was osteotomized into two parts at a certain location along the bone, perpendicular to the long axis of the bone, using the plane P1. This plane was defined by six parameters: a center point P1p = (P1x, P1y, P1z) located in the center of the bone in the axial plane, and rotation of the plane around the principal axes, P1r = (P1rx, P1ry, P1rz). The saved transformations M_{prox}^{radius}, M_{dist}^{radius}, M_{prox}^{ulna} and M_{dist}^{ulna} were used to reconstruct the proximal and distal fragment into the desired positions, aligned with the mirrored healthy bone. In case bone overlap between the proximal

and distal part occurred after the digital reconstruction, the overlapping part of the bone had to be removed by a second osteotomy plane P2. The location and normal vector of this plane were found by first transforming the plane P1 forward using the transformation M_{prox}^{radius} or M_{prox}^{ulna}, and then backward using the inverse of M_{prox}^{radius} or M_{prox}^{ulna}. In case bone was present between P1 and P2 there was an overlap of bone after reconstruction. Only then this second osteotomy plane was necessary, and the bone between P1 and P2 would be removed.

Table 4.2: Average segmentation differences between CT and MR bone segmentation. DSC = Dice Similarity Coefficient, MASD = Mean Absolute Surface Distance, HD = Hausdorff Distance, HD95 = 95th percentile Hausdorff Distance.

	DSC		MASD (mm)		HD (mm)		HD95 (mm)	
Mean	0.95		0.41		4.36		0.86	
Std. Dev.	0.02		0.07		2.38		0.28	
	Ulna	Radius	Ulna	Radius	Ulna	Radius	Ulna	Radius
Mean	0.95	0.95	0.41	0.42	4.32	4.40	0.86	0.85
Std. Dev.	0.02	0.02	0.06	0.08	2.32	2.44	0.20	0.33

4.2.4 Optimization

Finally, in the fifth step (Figure 4.2 – step 5), an exhaustive search method was employed to find the optimal location and orientation of the osteotomy that minimized a predefined objective function. Although the location and orientation of the plane were defined by six parameters, only three parameters were used to search the solution space. The location of the centre point of the plane was only dependent on the location along the longitudinal direction (P1z) of the bone, as P1x and P1y were constraint to be along the centroid of the bone in the x-y plane. The normal vector was dependent only on the rotation along the directions perpendicular to the longitudinal axis of the bone (P1rx and P1ry), as possible orientations could be defined using those two rotations. The search space for the location (P1z) was constraint to the diaphysis of the bone, between 20% and 80% of the length, and between angles of -35° and +35° in both directions for the orientation (P1rx and P1ry).

The objective function was defined as the weighted (w_1, w_2) average of two metrics. The first metric (m_1) quantified the distance between the osteotomy planes after reconstruction, defined as the mean squared minimal distance between all points on the edge of the proximal osteotomy surface to the distal osteotomy surface. The second metric (m_2) quantified the degree to which bone protrudes outwards after reconstruction, defined as the minimal squared distance between all points on the reconstructed bone to the healthy contralateral template. The metrics were equally weighted and thus $w_1 = w_2 = 0.5$. The objective function can then be written as:

$$\underset{P1z, P1rx, P1ry}{\arg\min} \ (w_1 m_1 + w_1 m_2)$$

Subject to:

$$P1z \in [0.2,0.8]$$
$$P1rx \in [-35,35]$$
$$P1ry \in [-35,35]$$

Step size for the exhaustive search was 1 mm for P1z and 1 degree for P1rx and P1ry. For each iteration of the exhaustive search step four of the automatic planning was repeated with new values for P1z, P1rx and P1ry and subsequently the objective function calculated.

This automatic planning was repeated three times, for each of the different bone models sets, i.e. bone segmented from CT, bone segmented from MR, and bone with cartilage segmented from MR. The resulting transforms and reconstructed bone models were saved to evaluate differences in osteotomy planning.
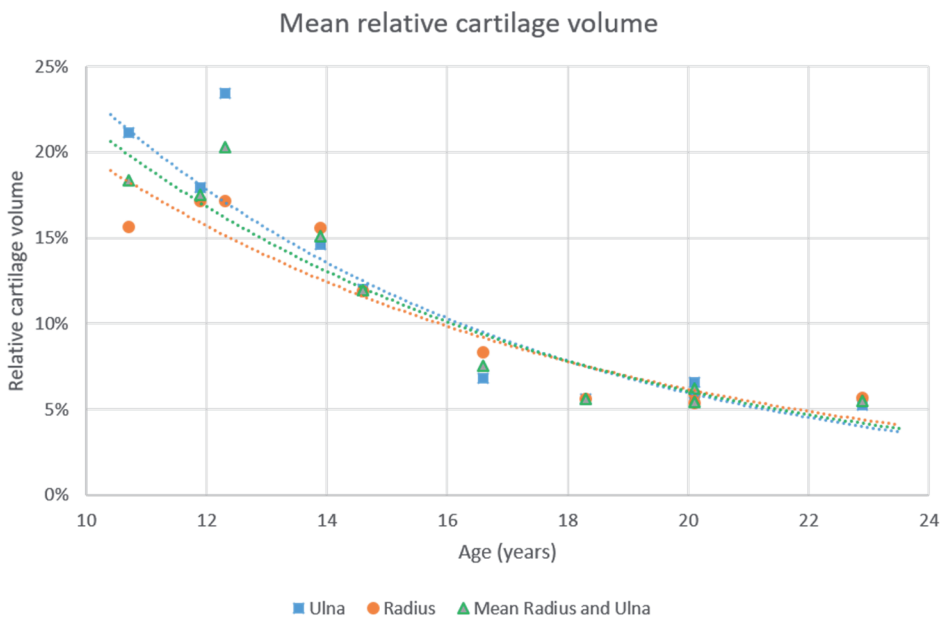
Figure 4.3: Relative cartilage volume against the age of each patient at the time of the MR scan. Both cartilage and bone segmentation were acquired from the MR scan. An exponential regression line is fitted to show the decrease in relative cartilage volume by age.

Table 4.3: Mean difference and ICC for the relative translation and rotation between the different methods of planning; on CTb, MRb and MRbc. ICC(3,1): two-way random model for absolute agreement of single measures was used [15].

CTb and MRb

	ΔT (mm)			φR (°)		
Mean	4.5			4.0		
Std. Dev.	2.1			2.4		
	Δx	Δy	Δz	φz	φy	φx
Mean	1.1	0.7	-0.3	0.2	-0.3	0.2
Std. Dev.	2.6	3.9	0.7	4.3	0.7	1.5
ICC	0.99	0.98	0.99	0.95	0.97	0.96

MRb and MRbc

	ΔT (mm)			φR (°)		
Mean	2.8			2.7		
Std. Dev.	2.1			2.6		
	Δx	Δy	Δz	φz	φy	φx
Mean	1.3	-0.1	0.2	1.0	-0.3	-0.1
Std. Dev.	2.5	1.9	0.6	3.5	0.7	0.7
ICC	0.99	0.99	0.99	0.99	0.99	1.00

Table 4.4: Mean difference in rotation and location of the osteotomy plane between the different methods of planning; on CTb, MRb and MRbc.

CTb and MRb

	ΔZ (mm)	ψY(°)	ψZ(°)
Mean	-2.5	0.5	-1.0
Std. Dev.	9.3	8.6	13.9

CTb and MRbc

	ΔZ (mm)	ψY(°)	ψZ(°)
Mean	3.8	0.4	0.6
Std. Dev.	6.6	4.4	16.4

MRb and MRbc

	ΔZ (mm)	ψY(°)	ψZ(°)
Mean	1.3	0.9	-0.5
Std. Dev.	9.5	8.8	13.6

4.2.5   Evaluation

The differences in segmentation between the CTb and MRb were quantified by four commonly used metrics, computed as described by Taha and Hanbury[21]: the Dice Similarity Coefficient (DSC), the Mean Absolute Surface Distance (MASD), the Hausdorff Distance (HD) and the 95th percentile Hausdorff Distance (HD95). The Relative Cartilage Volume (RCV) was calculate separately for the radius and ulna, and was defined as the cartilage volume divided by the bone volume as segmented from MR.

Bone realignment was defined as the rotational and translational difference between the distal bone fragment relative to the proximal bone fragment. The difference was calculated by finding the difference between the two transformation matrices that defined the bone reconstruction, using:

$$M^{radius} = M_{prox}^{radius^{-1}} \times M_{dist}^{radius}$$

$$M^{ulna} = M_{prox}^{ulna^{-1}} \times M_{dist}^{ulna}$$

Differences in relative rotational and translational realignment due to the use of differing segmentations could then be calculated as:

$$M_{CTb-MRb}^{radius} = M_{CTb}^{radius^{-1}} \times M_{MRb}^{radius}$$

$$M_{CTb-MRb}^{ulna} = M_{CTb}^{ulna^{-1}} \times M_{MRb}^{ulna}$$

$$M_{MRb-MRbc}^{radius} = M_{MRb}^{radius^{-1}} \times M_{MRbc}^{radius}$$

$$M_{MRb-MRbc}^{ulna} = M_{MRb}^{ulna^{-1}} \times M_{MRbc}^{ulna}$$

The difference between the two transformations could then be transcribed as a translation along the three orthogonal axes (Δx, Δy, Δz) and Euler angle rotation sequence (Φx, Φy, Φz). The translational differences were also evaluated by the total translation as calculated as the Euclidean distance $\Delta T = \sqrt{((\Delta x)^2 + (\Delta y)^2 + (\Delta z)^2)}$. We also calculated the total rotational distance using $\Phi R = \sqrt{((\Phi x)^2 + (\Phi y)^2 + (\Phi z)^2)}$, which is suitable for rotations where: Φx, Φz $\in [-\pi,\pi)$ and $\Phi y \in [-\pi/2,\pi/2)$ radians[138].

To estimate whether the differences in planning with the various segmentations were clinically relevant, we compared these with the residual error after surgery, as found by Vlachopoulos et al.[139]. As the definition of the axes used in this study does not directly correspond to ours, the total rotational (5.6º ± 4.2º) and translational (2.0 ± 1.4 mm) residual errors were used.

Finally, we compared the osteotomy planning by location and rotation of the cutting plane. The location was defined as the distance from the proximal end of the bone to the center of the osteotomy, along the axis of the bone, and denoted as ΔZ. The osteotomy orientation was

defined by the rotation of the osteotomy plane around the two axes ψX and ψZ, perpendicular to the longitudinal axis of the bone.

4.3 Results

4.3.1 Segmentation Accuracy

An example of the CT and MRI segmentations is shown in Figure 4.1. The difference between MRI and CT derived bone segmentation is summarized in Table 4.2. The RCV was plotted against the age of the patients in Figure 4.3. Exponential regression lines have been fitted to the data to illustrate the diminishing RCV with age.
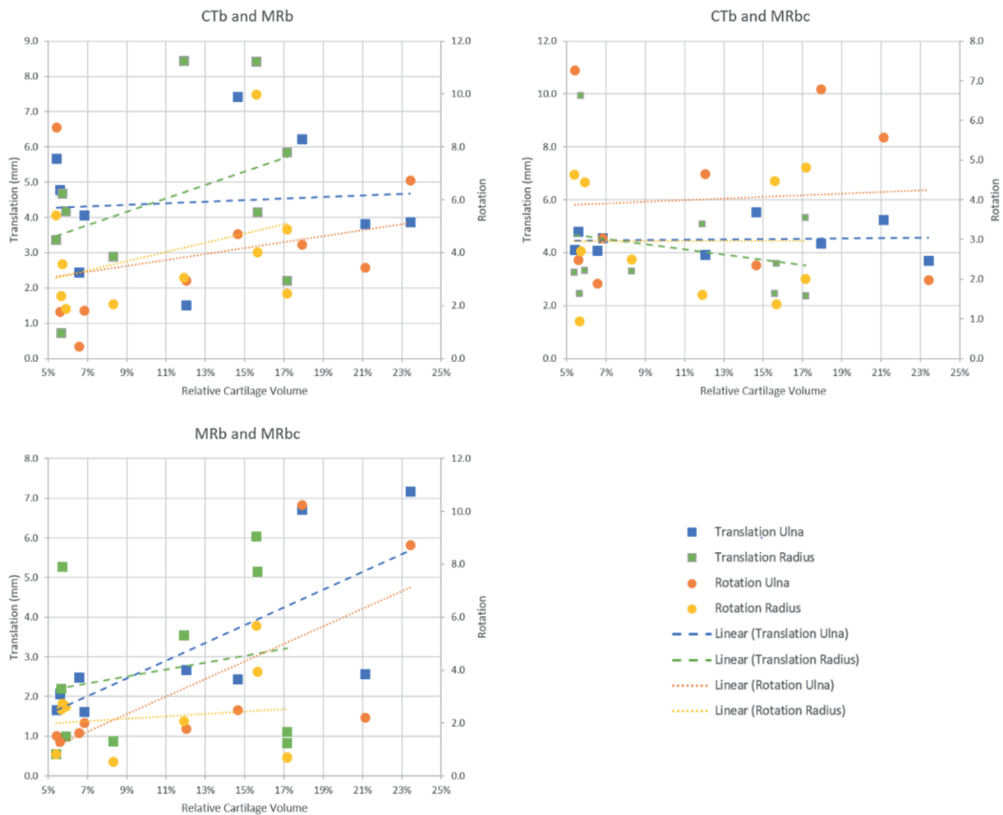
Figure 4.4: Translational and rotational realignment differences when planning the osteotomy using CTb, MRb and MRbc, plotted against the relative cartilage volumes of the patients. Linear trendlines have been fitted.
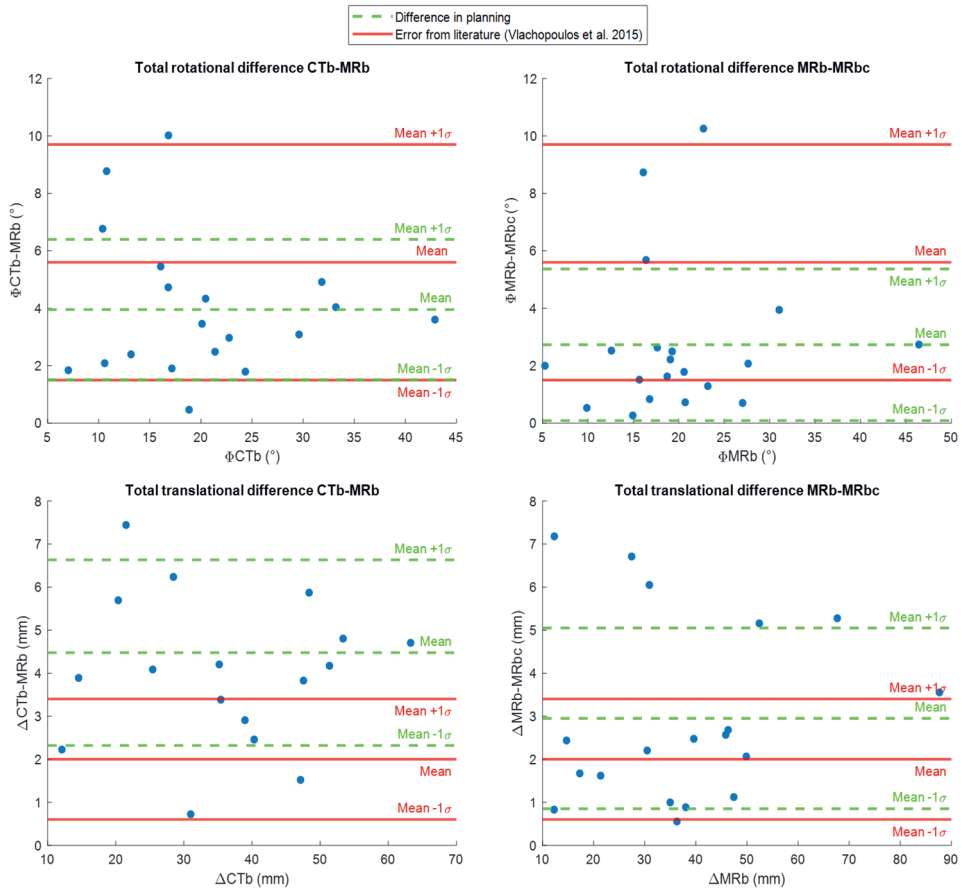
Figure 4.5: The rotational (top) and translational (bottom) differences between the CTb and MRb (left) and MRb and MRbc (right) in relative bone realignment after osteotomy are shown. The mean and standard deviation of the differences is compared to residual error after osteotomy surgery as reported by Vlachopoulos et al.[139]

4.3.2 Bone Realignment

The bone realignment was performed in threefold, with the CTb, MRb and MRbc segmentations. In Table 4.3 the mean differences in realignment when using the CTb, MRb and MRbc segmentations are shown. The Shapiro-Wilk test showed that for all methods ΔT and ΦR were normally distributed. Two-tailed paired t-tests showed that realignment performed with CTb and MRb had a significant mean difference for ΔT (p = 0.026) but did not show a significant mean difference for ΦR (p = 0.831). Comparison of realignment with MRb and MRbc again showed a significant mean difference for ΔT (p = 0.031) but not for ΦR (p = 0.173). The Intraclass Correlation Coefficient (ICC) was calculated using the two-way random model for absolute agreement of single measures, denoted as ICC(3,1). The translation and rotation for all comparisons was larger than 0.949, corresponding to excellent reliability (ICC > 0.9)[115].

In Figure 4.4 the differences in realignment when using different segmentations are shown, plotted against the RCV of the patient. Moderate to strong correlations were only found between the RCV and the difference in rotational (R = -0.68, p = 0.043) and translational (R = 0.84, p = 0.024) realignment of the ulna when using MRb versus MRbc. Only weak correlations ($|R| < 0.3$) were found between the RCV and the realignment differences between CTb versus MRb and CTb and MRbc.

The graphs in Figure 4.5 show that the rotational differences in realignment were on average smaller than the residual errors that were found after osteotomy surgery in a clinical study by Vlachopoulos et al.[139], but the translational errors were larger. Visual comparisons of the final bone reconstructions are shown in Figure 4.6.
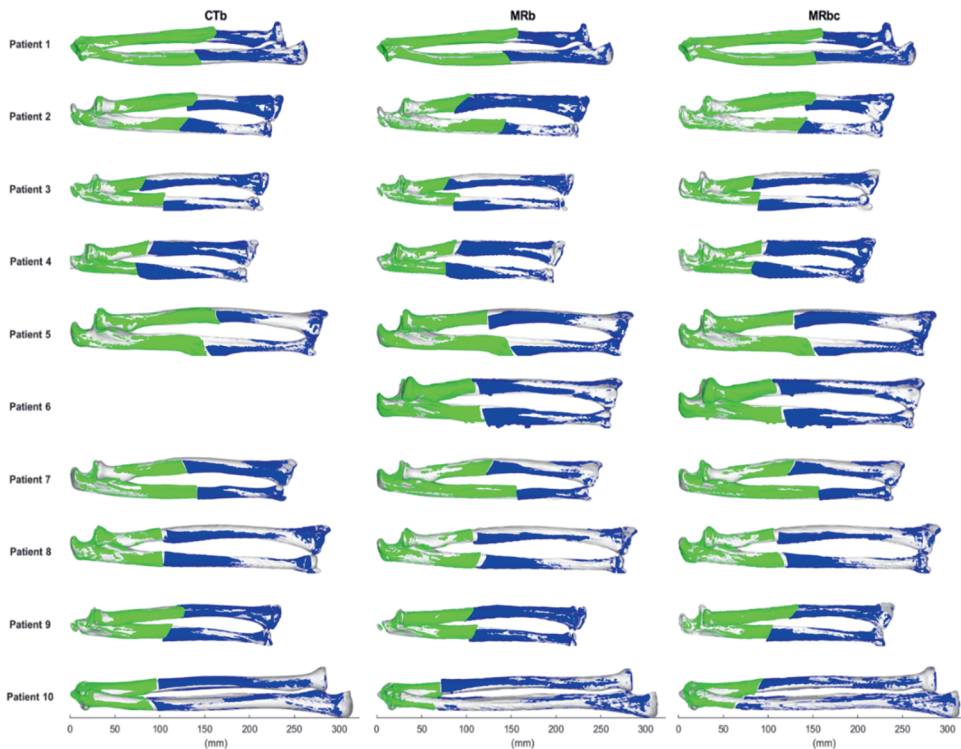
Figure 4.6: Illustration of the difference in simulated bone reconstruction in the radius and ulna between CTb, MRb and MRbc segmentations for all patients. The proximal (blue) and distal (green) fragment of the deformed bone after the planned reconstruction are overlayed on the mirrored healthy contralateral side (white). The CT scan of patient 6 was incomplete and thus not included.

4.3.3 Osteotomy Plane Optimization

A visual comparison of the osteotomy plane location and orientation on the CTb, MRb and MRbc is shown in Figure 4.7. The mean differences that arise due to the planning on different segmentations are summarized and the distribution of these metrics for both the ulna and radius is shown in Figure 4.8.

4.4 Discussion

The aim of this study was to evaluate differences in radio-ulnar osteotomy planning based on CT and MRI derived segmentations of the bone and/or cartilage. Although the realignment parameters showed excellent correlation, statistically and clinically significant mean differences were found, both due to segmentation differences of the bone between CT and MRI scans, as well as due to differences between planning on bone and bone with cartilage.
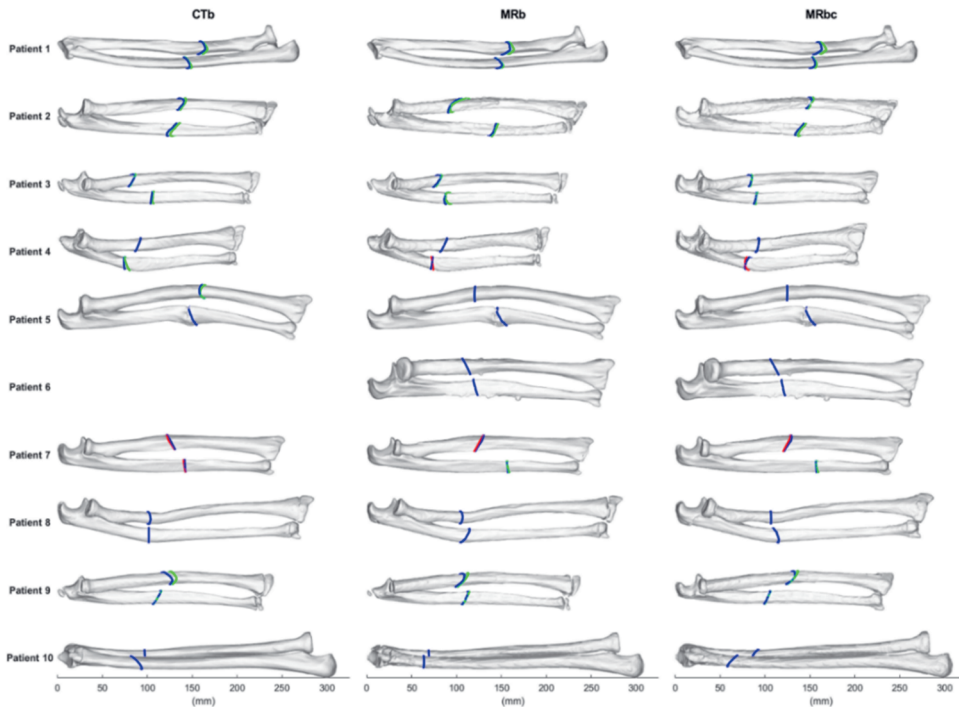
Figure 4.7: Illustration of the difference in optimal computed osteotomy location and orientation in the radius and ulna between CTb, MRb and MRbc segmentations for all patients. The blue line indicates the primary osteotomy cut. Where necessary, the secondary osteotomy cut is visible. The secondary cut is shown in green if it does not intersect the primary cut, or in red if it does overlap the primary cut. The CT scan of patient 6 was incomplete and thus not included.

A prerequisite for the evaluation of the hypotheses posed in this study was an accurate method for the segmentation of bone and cartilage from both modalities. The segmentation of CT scans was relatively straightforward, as bone is generally the only tissue with intensity values over 200 HU. However, issues like partial volume effects, small joint spaces, and artefacts due to foreign objects necessitated at least some manual correction in all CT segmentations[80].

Segmentation of bone from MRI scans was more complicated and thus time consuming, as there is no unique intensity range in bone that separates it from the other tissues. For example, in the Dixon reconstructed water images used in this study, bone, air and tendons all have approximately the same intensity range. Additionally, the resolution of the MRI scans was lower than the resolution of the CT scans. The results showed that the MASD between MRI and CT bone segmentation was approximately equal to the resampled voxelsize of the segmentations, at 0.41 mm. For comparison, the most relevant reference in the literature for comparable segmentations was Marin et al.[130], who reported an average MASD of 1 mm between registered hand and forearm bones that were manually segmented from CT and MRI.

Segmentation of the cartilage from MRI scans was entirely manual, which caused it to be a time-consuming process. This increases the risk of low intra- and interobserver agreement between segmentations, as noted by Brui et al.[140], who observed this when segmenting wrist cartilage from MRI (DSC = 0.9 intraobserver, DSC = 0.78-0.88 interobserver). For an efficient, deterministic clinical workflow, the issues with manual segmentation from MRI would need to be addressed. Methods to improve this could be to use cartilage specific MRI acquisition sequences as proposed by Dalili et al.[141], and by automating the cartilage segmentation, using for example deep learning[140].

The first hypothesis of this study was that automated osteotomy planning would yield comparable results when performed on bone segmentations derived from CT versus segmentations derived from MRI. Excellent correlation (ICC(3,1) > 0.9) was found for each of the translational (Δx, Δy, Δz) and rotational (Φx, Φy, Φz) bone realignment components. However, the differences in realignment showed a significant mean difference of the total translational component ΔT, but no significant mean difference in the total rotational component ΦR. Additionally, comparison with the residual errors found after surgery showed that the differences in planned translational realignment were larger than the accuracy with which osteotomy surgery is performed and would therefore have a clinically significant impact on the outcome. Larger differences were also seen for the osteotomy plane localization ($\Delta Z = 6.7 \pm 6.9$ mm) and orientation ($\psi Y = 5.5 \pm 6.7°$ and $\psi Z = 8.4 \pm 11.1°$). This indicated that the impact of using the CTb and MRb had a relatively small effect on the rotational realignment of the bone, but had a greater impact on the translational realignment and the position and orientation of the osteotomy plane.

The second hypothesis of this study was that the same method of osteotomy planning would yield significantly different results when cartilage was included in the segmentation. Again,
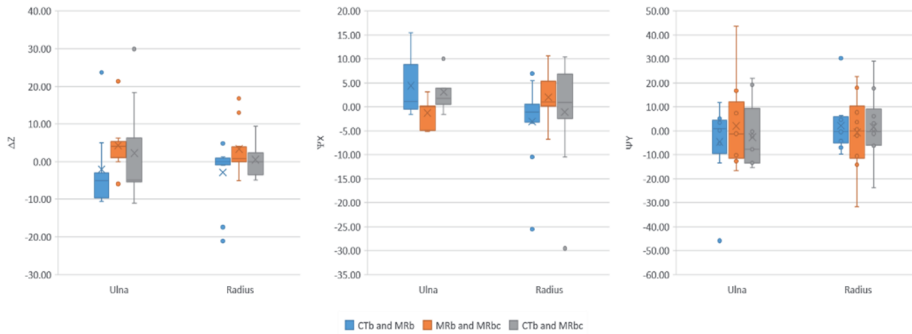
Figure 4.8: Differences in location (ΔZ) and orientation (ψX, ψZ) of the osteotomy plane between the different methods of planning; on CTb, MRb and MRbc.

the translational and rotational components of the realignment again showed excellent correlation. Significant mean differences were found for the translational, but not for the rotational realignment. The results in Figure 4.7 and Figure 4.8 showed that the differences between MRb and MRbc in osteotomy plane localization ($\Delta Z = 6.5 \pm 7.0$ mm) and orientation ($\psi Y = 5.9 \pm 6.6°$ and $\psi Z = 10.8 \pm 8.4°$) were similar to the differences between CTb and MRb. The mean difference in realignment between planning on CTb and MRb was slightly larger than the difference in planning between MRb and MRbc. This means that the impact of segmentation differences between modalities was on average larger than the impact of including cartilage.

The third hypothesis stated that the relative amount of articular cartilage would be correlated with the effect cartilage inclusion has on the osteotomy planning. Firstly, as was expected, we found that the RCV was higher in younger adolescents and decreased with age, which corresponds to the established process of endochondral ossification[142]. The graph in Figure 4.4 also showed a positive correlation between the amount of cartilage and the translational differences between MRb and MRbc for the ulna. This indicates that there is a larger impact when taking the cartilage into consideration in younger patients, who generally have a higher RCV, than in older patients. This effect was only present in the ulna and not in the radius, might be due to the differences in the developmental process of endochondral ossification for each bone[142].

4.4.1 Limitations

A limitation of this study was that the tool might not have incorporated all the criteria that an orthopaedic surgeon considers during the planning of an osteotomy. It found the optimal solution to the objective function that it was presented with, which included only two optimization targets: increase the amount of volume overlap with the healthy contralateral example and minimize the distance between the osteotomy surfaces. Other constraints that might be considered include the proximity of important tissues such as tendons, muscle and blood vessels which influence the surgical approach, the presence of earlier implants or incision scars, and the relation between the osteotomy location on the radius and ulna.

It should also be noted that the reconstruction was constrained to optimal realignment of the distal and proximal joint to the mirrored healthy contralateral side. Therefore, only two types of osteotomies were considered: either a closing or opening wedge osteotomy. Other techniques, such as the oblique single-cut rotational osteotomy[143] (OSCRO) or oblique double-cut rotational osteotomies[132] (ODCRO), might have resulted in a different planning.

We compared the differences in realignment with the postoperative residual error found by Vlachopoulos et al.[139]. Other studies reported smaller residual errors of 1.4° to 1.8°[144] or even as low as 1° and 1 mm[145]. However, these measurements were based on 2D radiographs and could therefore not be directly compared to our results, and thus the results found by Vlachopoulos et al. were deemed a more appropriate comparison.

Finally, due to the difficulty in obtaining a large dataset that adheres to the narrow inclusion criteria that were used, and to obtain both CT and MRI scans of both the deformed and healthy contralateral forearm, the dataset was relatively small. A study based on a larger dataset might have more power and could thus more accurately prove or disprove our hypotheses.

4.5 Conclusion

In this study, we performed automatic radius and ulna osteotomy planning based on bone and/or cartilage segmentations from CT and MRI. Excellent correlations were found between the realignment parameters when comparing CT and MRI segmentations, and when comparing MRI bone and bone and cartilage segmentations. However, statistically significant mean differences were found in the translational component of the bone realignment for both methods, which were larger than the differences reported in literature between planned and realized osteotomy surgery. This indicated that small differences in segmentation might have a clinically significant impact on the osteotomy planning.

When we compared planning on bone and bone with cartilage, we found a positive correlation between the relative cartilage volume of the patient and the difference in realignment of the ulna. This indicated that endochondral cartilage might also be a non-negligible factor when planning osteotomies for young patients.
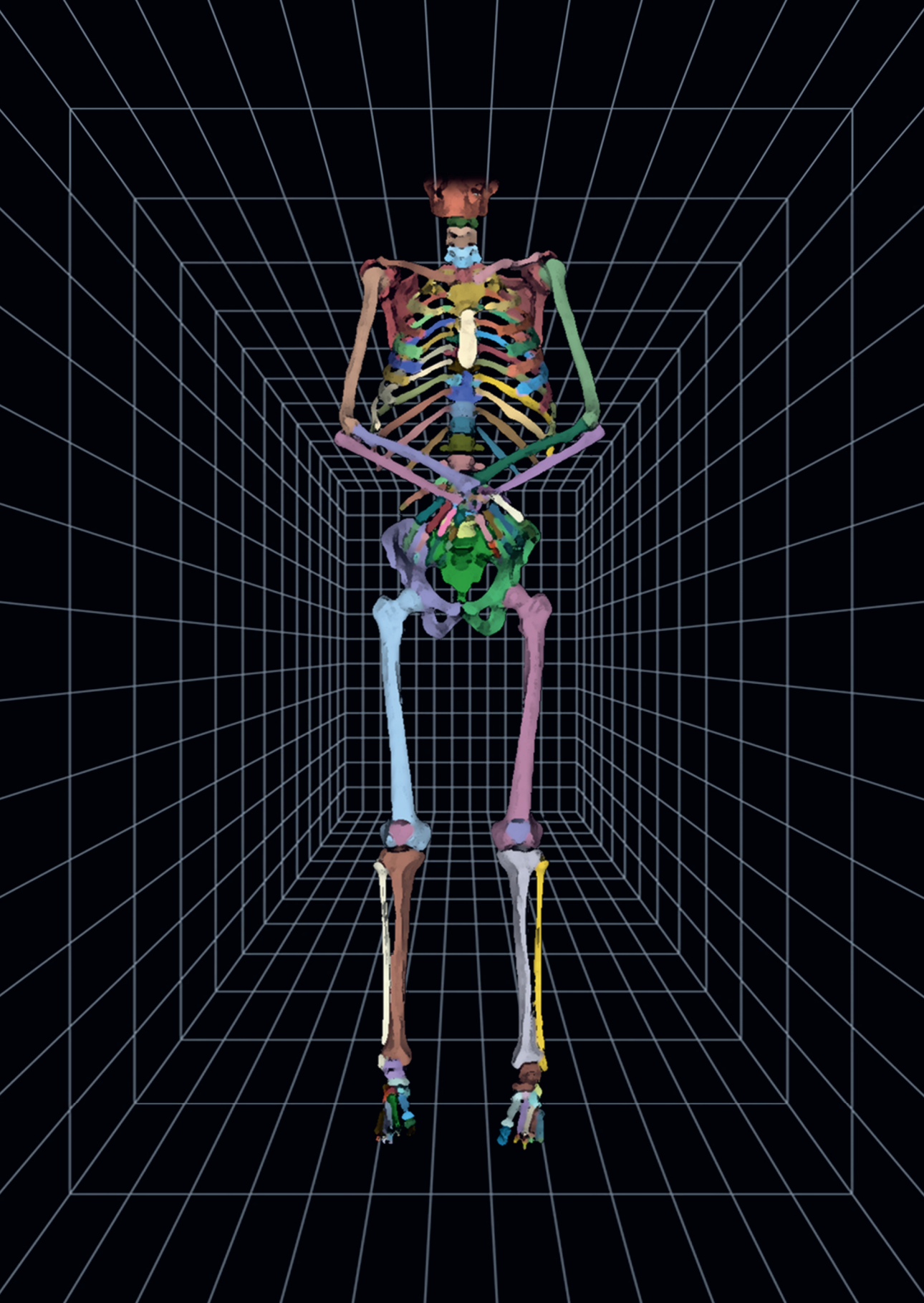
Chapter 5

Automatic assessment of lower limb alignment from CT

Abstract

Background: Pre-operative planning of lower limb realignment surgery necessitates the quantification of alignment parameters by using landmarks placed on medical scans. Conventionally, alignment measurements are performed on two-dimensional (2D) standing radiographs. To enable fast and accurate three-dimensional (3D) planning of orthopedic surgery, automatical calculation of the lower limb alignment from 3D bone models is required. The goal of this study was to develop, validate and apply a method that automatically quantifies the parameters defining lower limb alignment from CT scans.

Methods: CT scans of the lower extremities of fifty subjects were both manually and automatically segmented. Thirty-two manual landmarks were indicated twice on the bone segmentations to assess intraobserver variability on a subset of twenty subjects. The landmarks were also indicated automatically using a shape-fitting algorithm. The landmarks were then used to calculate 25 angles describing the lower limb alignment for all subjects.

Results: The mean (±standard deviation) intraobserver absolute difference was 2.01±1.64 mm and 1.05±1.48° for the landmark indication and angles respectively, whereas the mean absolute difference between the manual and fully automatic method was 2.17±1.37 mm and 1.10±1.16°. The manual method necessitated approximately 60 minutes of manual interaction compared to 12 minutes of fully automatic computation time. The intraclass correlation coefficient showed good to excellent reliability between the manual and automatic calculation for 23 out of 25 angles, which was similar to the intraobserver reliability. Of the automatically calculated angles for the fifty subjects, the mean of 18 out of the 25 angles were within the expected range.

Conclusion: We developed a method that automatically calculated a comprehensive range of 25 measurements that defined lower limb alignment in considerably less time with similar intraobserver differences compared to manual results. This method could thus be used as an efficient alternative to manual alignment assessment.

5.1 Introduction

Surgerical correction of a lower limb malalignment typically consists of an osteotomy in the femur and/or tibia. For surgical planning, the malalignment is quantified using predefined landmarks and angles, such as the center points of the femoral head, knee and ankle which define the mechanical Hip-Knee-Ankle (mHKA) angle. Paley[2] described these angles, and the methods to restore the alignment, for almost all types of malalignments.

Conventionally, these angles are measured on two-dimensional (2D) radiographs to define the malalignment in the frontal and sagittal planes. The introduction of three-dimensional (3D) imaging and planning now enables more accurate quantification of bone geometry, in particular the tibial and femoral torsion[146,147]. Additionally, it enables the use of image-based guidance technologies and personalized surgical cutting guides and implants that could increase surgical accuracy and efficiency[148–151].

Reproducible landmark placement is essential for accurate alignment assessment. Studies concerning intra- and interobserver variability showed that standardized definitions for each landmark were crucial to attain reproducible alignment parameters[76,152]. Differences were found between measurements performed on 2D and 3D images, both due to measurement modality[95] and limb loading of patients in supine or standing position[95,153]. Fürmetz et al.[76] proposed definitions for the manual placement of 24 different landmarks, such as the points on the femoral head, the femoral and tibial condyles, and the tibial plafond, to establish standardized protocols for alignment measurements.

To reduce time and interobserver variation, various methods have been proposed to automate anatomical landmark placement. Subburaj et al.[154,155] proposed methods that used the bone surface curvature. Several studies used axis of inertia alignment, model cross-sections or shape-fitting[77,156,157]. Other authors matched Statistical Shape Models (SSMs) of bones with predefined landmarks to new patients[158–161]. All of these methods required initial manual segmentation of the bones. None of these methods provided all the landmarks and angles defined by Paley.

In this study, we introduced a fully automatic method for lower limb malalignment assessment from CT. We employed this method twice: (1) To validate the automatic approach by comparing manual, semi-automatic and automatic landmark and angle calculations on twenty subjects and to compare the time needed for the different methods. (2) To automatically quantify the variation in lower limb alignment for fifty subjects, and to compare the results to the normal variation.
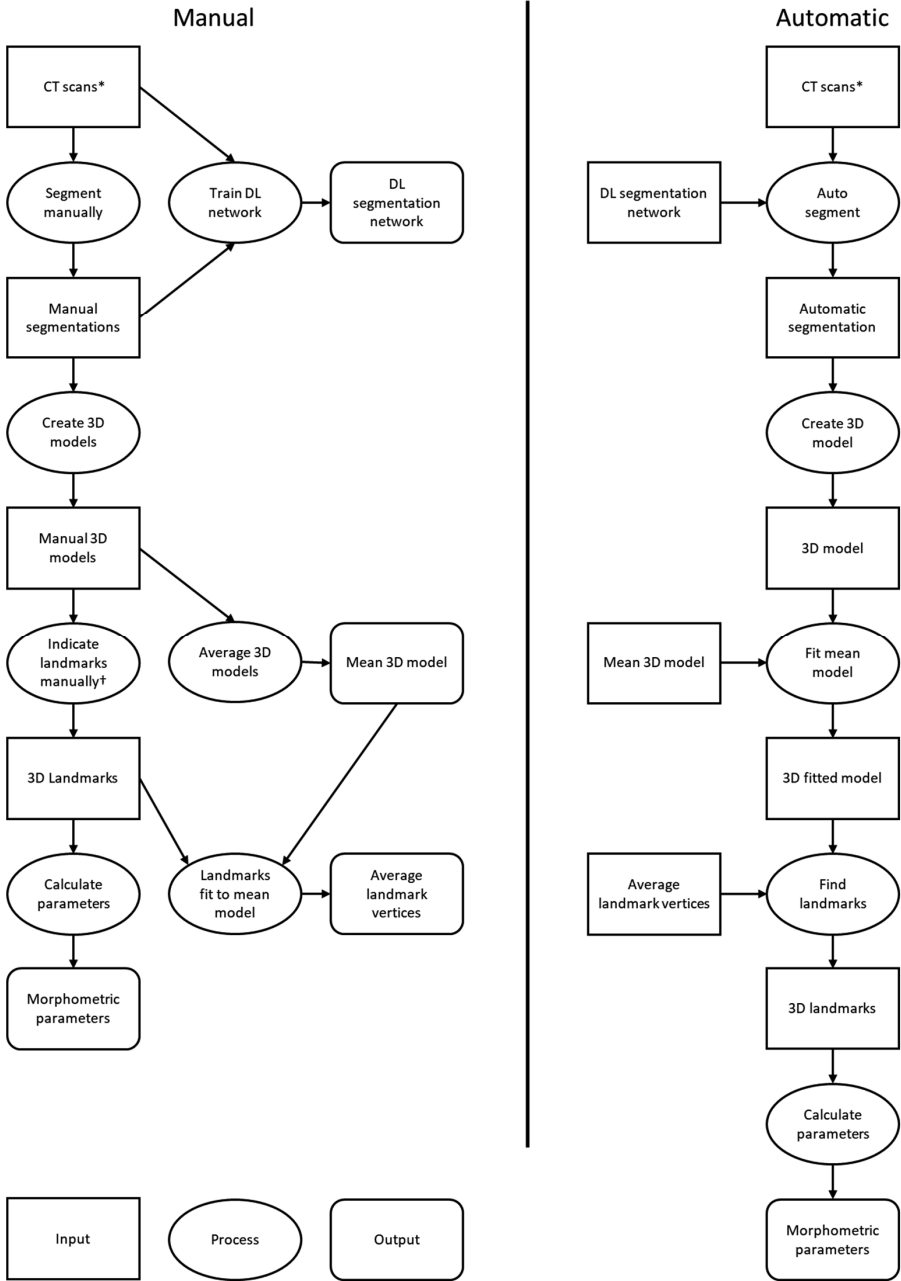
Figure 5.1: Diagrams showing the manual (left) and automatic (right) workflows that were used to calculate the morphometric parameters of the lower limb from the CT scans. The aggregated output from the network on the left was used as input for the automatic workflow on the right. The manual workflow was only performed once to obtain the necessary input for the automatic workflow, and is thereafter no longer required for new subjects.

*: 50 subjects were used. †: 20 subjects were used.

5.2   Materials and Methods

Figure 5.1 shows the step-by-step manual and automatic workflows for lower limb alignment assessment. The steps are explained in detail in the sections below.

5.2.1   Data

Bilateral, non-weight-bearing lower limb CT scans of fifty subjects in supine position were acquired retrospectively from the University Medical Center Utrecht (UMCU, Utrecht, The Netherlands) as DICOM data, and were used without reformatting. Detailed scan parameters were described by Kuiper et al.[162] The subjects had no previous clinical indication of lower limb malalignment, and had undergone a CT scan due to unrelated medical reasons (i.e. vascular indications). The study was judged not to be subject to the Medical Research Involving Human Subjects Act (WMO) by the Medical Ethical Committee, as described in IRB protocol number 16-612/C.
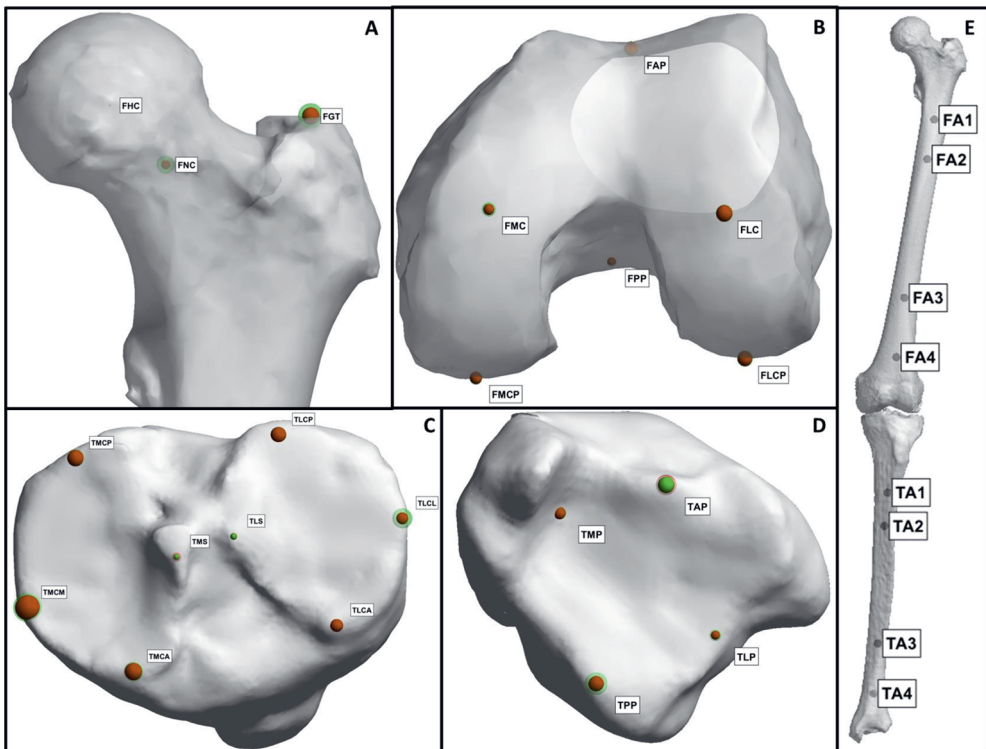
Figure 5.2: All directly indicated landmarks on the proximal femur (A), distal femur (B), proximal tibia (C) and distal tibia (D). The radius of the spheres indicate the mean intraobserver (green) and automatic intermethod (red) distance over all subjects. Landmarks along the shaft (E) were always calculated automatically and are shown in black.

5.2.2 Manual segmentation

The femur and tibia were segmented semantically from each scan, i.e. each bone was given a separate label. This was done using the CT Bone Segmentation Module and manual editing in Mimics (Mimics Medical 21.0, Materialise, Leuven, Belgium) by two operators with 2 years of experience. Each operator performed the segmentation of 25 subjects.

5.2.3 Automatic segmentation

An artificial intelligence-based workflow was used for automatic segmentation. The details of this method were described by Kuiper et al.[162] The workflow used a neural network to segment six different bones: the femur, tibia, fibula, coxae, talus and calcaneus. Only the femur and tibia segmentations were used in this study. The differences between the manual and automatic segmentations were quantified by four commonly used metrics[21]: the Dice Similarity Coefficient (DSC), the Mean Absolute Surface Distance (MASD), the Hausdorff Distance (HD) and the 95th percentile Hausdorff Distance (HD95). Both manual and automatic segmentations were converted to triangulated 3D models using the marching cubes algorithm[48].

5.2.4 Manual landmarks

Landmarks based on the methods proposed by Fürmetz et al.[76], but adapted to include all landmarks defined by Paley[2], were indicated twice on each 3D bone models of twenty randomly chosen subjects by one trained observer (RK), using the opensource software 3D Slicer[134], to establish the intraobserver variability. One week pause between the two measurements was introduced to reduce memorization bias. The acronym and definition of each landmark are summarized in Table 5.1 and they are shown in Figure 5.2. Landmarks were defined using a single point, with the exception of the following:

The centers of femoral hip (FHC) and neck (FNC) were indicated using two planes (Figure 5.3). The FNC was defined as the center of the femoral neck circumference at the intersection with the plane[159]. For the FHC, the plane separated the femoral head from the rest of the femur. The FHC was defined as the center of a sphere fitted to the femoral head[3].
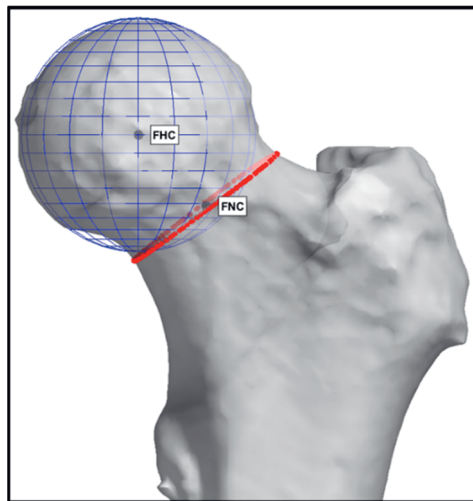
Figure 5.3: Indication of the Femoral Head Centre (FHC) as the centre of a sphere fitted to the vertices of the femoral head, and the Femoral Neck Centre (FNC) as the centre of the vertices at the smallest cross-section of the femoral neck.

Table 5.1: Definition of all landmarks that were necessary to calculate the lower limb alignment angles as defined by Paley[2]. Landmarks indicated with <u>underscore</u> were defined indirectly.

Landmark	Acronym	Definition
Proximal Femur		
<u>Femoral Hip Centre</u>	<u>FHC</u>	Midpoint of a sphere fit to the femoral head vertices
<u>Femoral Neck Centre</u>	<u>FNC</u>	Midpoint of the vertices at the smallest cross-section of the femoral neck
Femoral Greater trochanter	FGT	Most proximal point on the greater trochanter
Distal Femur		
Femoral Medial Condyle	FMC	Most distal point of the femoral medial condyle
Femoral Lateral Condyle	FLC	Most distal point of the femoral lateral condyle
<u>Femoral Centre Condyle</u>	<u>FCC</u>	Midpoint between FMC and FLC
Femoral Medial Condyle Posterior	FMCP	Most posterior point of the femoral medial condyle
Femoral Lateral Condyle Posterior	FLCP	Most posterior point of the femoral lateral condyle
Femoral Anterior Point	FAP	Anterior point on femur where the condyle meets the metaphysis
Femoral Posterior Point	FPP	Posterior point on femur where the condyle meets the metaphysis
Femoral Centre Point	FCP	2/3rd from FPP to FAP
Shaft Femur		
Femoral Axis 1-4	FA1-FA4	Midpoint of femoral transverse cross-section 15, 30, 65 and 75% along the femoral axis
Proximal Tibia		
Tibial Medial Spine	TMS	Midpoint of medial tibial spine
Tibial Lateral Spine	TLS	Midpoint of lateral tibial spine
<u>Tibial Centre Spine</u>	<u>TCS</u>	Midpoint between the TMS and TLS
Tibial Medial Condyle Medial	TMCM	Most medial point of the medial tibial condyle
Tibial Medial Condyle Anterior	TMCA	Most anterior point of the medial tibial condyle
Tibial Medial Condyle Posterior	TMCP	Most posterior point of the medial tibial condyle
<u>Tibial Medial Condyle Centre</u>	<u>TMCC</u>	Closest point on the tibial surface to the average of TMS, TMCM, TMCA and TMCP
Tibial Lateral Condyle Lateral	TLCL	Most lateral point of the lateral tibial condyle
Tibial Lateral Condyle Anterior	TLCA	Most anterior point of the lateral tibial condyle
Tibial Lateral Condyle Posterior	TLCP	Most posterior point of the lateral tibial condyle
<u>Tibial Lateral Condyle Centre</u>	<u>TLCC</u>	Closest point on the tibial surface to the average of TLS, TLCL, TLCA and TLCP
<u>Tibial Anterior Condyles</u>	<u>TAC</u>	Midpoint between TMCA and TLCA
<u>Tibial Posterior Condyles</u>	<u>TPC</u>	Midpoint between TMCP and TLCP
<u>Tibial Centre Condyles</u>	<u>TCC</u>	Midpoint between TLCC and TMCC
<u>Tibial Centre Condyles sagittal</u>	<u>TCCsag</u>	4/5th from TPC to TAC
Distal Tibia		
Tibial Anterior Plafond	TAP	Midpoint of the anterior ridge of the distal tibial joint
Tibial Posterior Plafond	TPP	Midpoint of the posterior ridge of the distal tibial joint
Tibial Medial Plafond	TMP	Midpoint of the medial plafond of the distal tibial joint
Tibial Lateral Plafond	TLP	Midpoint of the lateral plafond of the distal tibial joint
<u>Tibial Centre Plafond</u>	<u>TCP</u>	Midpoint between the TMP and TLP
<u>Tibial Centre Plafond sagittal</u>	<u>TCPsag</u>	Midpoint between the TAP and TPP
Shaft Tibia		
Tibial Axis 1-4	TA1-TA4	Midpoint of tibial transverse cross-section at 15, 30, 65 and 75% along the tibial axis
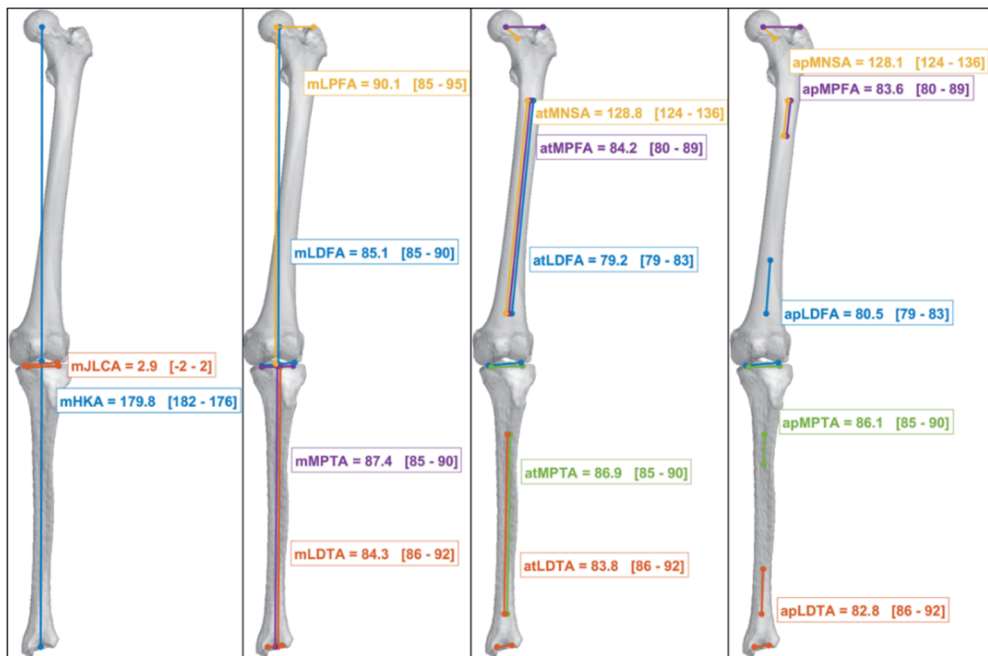
Figure 5.4: An example of the coronal alignment parameters. For each parameter, the actual angle for this patient is shown, alongside the normal range in square brackets. The colour of the textbox corresponds to the lines that were used to calculate each angle. Lines that were used to calculate more than one parameter are shown in parallel but are identical. These images were constructed using the full-auto method.

The center of the femoral and tibial condyles (FCC, TMCC, TLCC), tibial spine (TCS) and tibial ankle (TAC) were defined as the mean of several landmarks (Table 5.1). The landmarks along the medial axis of the femur (FA1-FA4) and tibia (TA1-TA4) were never placed manually, but defined as the midpoint of the femoral and tibial cross-section in the transverse plane at 15, 30, 65 and 75% bone length from the proximal side of the bone. Bone length was measured from the most proximal to the most distal point of the bone model. These dimensions were empirically found to closely correspond to the landmarks defined by Paley[2].

5.2.5　Mean bone model construction

Fifty manually segmented femurs and tibias were used to construct the mean bone models in four steps. First, one subject was randomly chosen to serve as the template. Second, the template femur and tibia were remeshed to obtain approximately isotropic meshes with 1 mm edge length. Third, for all other subjects, the mesh was rigidly registered and scaled using an Iterative Closest Point (ICP)[49] algorithm to align to the template. Fourth, the template was non-rigidly registered to the new subject using an adapted version of the non-rigid ICP algorithm[49], developed by Audenaert et al.[81], which is available online[163]. After registration, the deformed template was used as a substitute for the original mesh of that patient. The mean of the coordinates of each vertex in all transformed template meshes then defined the vertices of the mean femur and mean tibia.

5.2.6   Automatic Landmarks

Automatic landmark indication was performed by fitting the mean bone model to a new patient, using the combined rigid and non-rigid ICP algorithms. The bone models were first split into a proximal and a distal part along the transverse plane at 50% of the length, to improve registration in case of large deformations. After registration, the vertices of the mean bone model corresponded to the geometry of the bone model of the new subject. For each subject, a mean bone model based on the other 49 subjects was used to avoid bias.

Landmarks were automatically found by averaging the locations of the manual landmarks on the fitted mean bone model. In case of the FGT and FNC, the mean of the corresponding vertices was used to construct a new plane, from which the points were calculated as described in Paragraph 5.2.5. For the automatic landmarks of the twenty manually annotated subjects, only the manual landmarks of other subjects were used, to avoid bias.
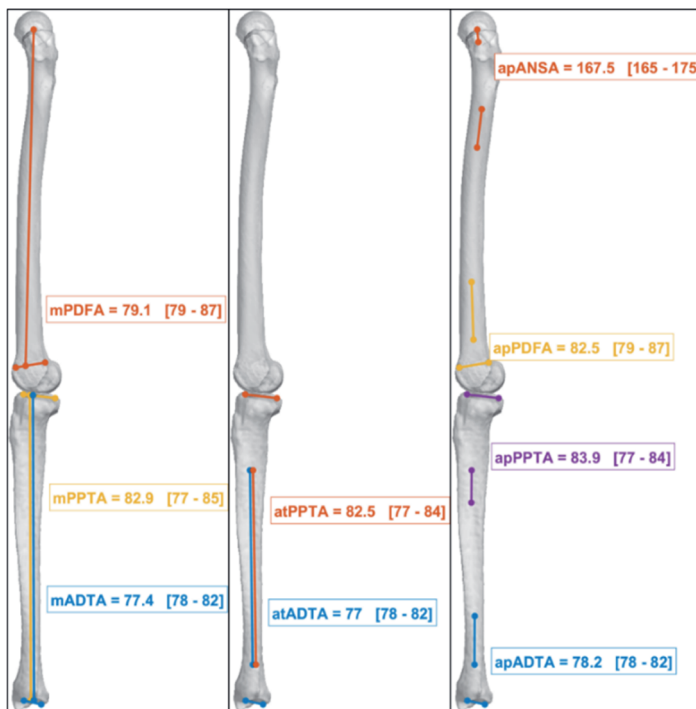
Figure 5.5: An example of the sagittal alignment parameters calculated based on the manual landmarks. For each parameter, the actual angle for this patient is shown, alongside the normal range in square brackets. The colour of the textbox corresponds to the lines that were used to calculate each angle. Lines that were used to calculate more than one parameter are shown in parallel but are identical. These images were constructed using the full-auto method.

Table 5.2: Definitions of all angles that were used to describe the lower limb alignment. Each angle is defined as the smallest angle between the two lines formed by the landmarks shown under Definition. Angles were calculated from a 2D perspective in either the coronal, sagittal or axial plane.

	Angle	Acronym	Definition		Plane	Normal Range	
			Line 1	Line 2		Min	Max
Anatomical Partial	Anterior Distal Tibial Angle	apADTA	TA4-TA3	TAP-TPP	Sagittal	78	82
	Anterior Neck Shaft Angle	apANSA	FNC-FHC	FA1-FA2	Sagittal	165	175
	Lateral Distal Femoral Angle	apLDFA	FA4-FA3	FMC-FLC	Coronal	79	83
	Lateral Distal Tibial Angle	apLDTA	TA4-TA3	TMP-TLP	Coronal	86	92
	Medial Neck Shaft Angle	apMNSA	FA1-FA2	FNC-FHC	Coronal	124	136
	Medial Proximal Femoral Angle	apMPFA	FA1-FA2	FGT-FHC	Coronal	80	89
	Medial Proximal Tibial Angle	apMPTA	TA1-TA2	TLCC-TMCC	Coronal	85	90
	Posterior Distal Femoral Angle	apPDFA	FAP-FPP	FA4-FA3	Sagittal	79	87
	Posterior Proximal Tibial Angle	apPPTA	TA1-TA2	TAC-TPC	Sagittal	77	84
Anatomical Total	Anterior Distal Tibial Angle	atADTA	TA1-TA4	TAP-TPP	Sagittal	78	82
	Lateral Distal Femoral Angle	atLDFA	FA4-FA1	FMC-FLC	Coronal	79	83
	Lateral Distal Tibial Angle	atLDTA	TA4-TA1	TMP-TLP	Coronal	86	92
	Medial Neck Shaft Angle	atMNSA	FA1-FA4	FNC-FHC	Sagittal	124	136
	Medial Proximal Femoral Angle	atMPFA	FA1-FA4	FGT-FHC	Coronal	80	89
	Medial Proximal Tibial Angle	atMPTA	TA1-TA4	TLCC-TMCC	Coronal	85	90
	Posterior Proximal Tibial Angle	atPPTA	TA1-TA4	TAC-TPC	Sagittal	77	84
Mechanical	Anterior Distal Tibial Angle	mADTA	TCCsag-TCPsag	TAP-TPP	Sagittal	78	82
	Hip Knee Ankle Angle	mHKA	TCS-TCP	TCS-FHC	Coronal	176	182
	Joint Line Convergence Angle	mJLCA	FMC-FLC	TMCC-TLCC	Coronal	-2	2
	Lateral Distal Femoral Angle	mLDFA	FCC-FHC	FMC-FLC	Coronal	85	90
	Lateral Distal Tibial Angle	mLDTA	TCP-TCC	TMP-TLP	Coronal	86	92
	Lateral Proximal Femoral Angle	mLPFA	FHC-FGT	FHC-FCC	Coronal	85	95
	Medial Proximal Tibial Angle	mMPTA	TCC-TCP	TLCC-TMCC	Coronal	85	90
	Posterior Distal Femoral Angle	mPDFA	FAP-FPP	FCP-FHC	Sagittal	79	87
	Posterior Proximal Tibial Angle	mPPTA	TCCsag-TCPsag	TAC-TPC	Sagittal	77	85

5.2.7 Alignment parameters

Both manual and automatic landmarks were used to calculate the angles described by Paley[2]. Before angle calculation the legs were first rotated to align the FHC to TCP line with the proximo-distal axis, and the FMCP to FLCP line with the medio-lateral axis.

Table 5.2 shows the angle definitions and normal variations. A visual overview of the angles is shown in Figure 5.4 and 5.5. The angles were separated into three groups, reflected by their prefix. Anatomical Partial (ap) angles were calculated using the anatomical bone axis close to the joint, Anatomical Total (at) angles using the total anatomical bone axis and Mechanical (m) angles using the mechanical axis.

5.2.8   Evaluation

We performed the first part of the study on twenty subjects using the manual, semi-automatic (semi-auto) and fully automatic (full-auto) workflow, to study the impact of the automatic steps on landmark indication and angle calculation. The manual workflow used manual segmentations and manual landmark indication. The semi-auto workflow used manual segmentations but automatic landmark indication. The full-auto workflow used automatic segmentations and landmark indication. The time needed for each method was measured in minutes.

The absolute landmark and angular intraobserver differences were defined as the distance between two corresponding manual measurements. The semi-auto and full-auto differences were compared to the mean of the two manual measurements.

The second part of the study was only performed using the full-auto workflow, to compare the parameters of the study population of fifty subjects to the normal variation as summarized in Table 5.2.

5.2.9   Statistical analysis

The Intraclass Correlation Coefficient (ICC) for the angle intraobserver difference was calculated using the two-way mixed effects, absolute agreement, single rater model. The semi-auto and full-auto ICC was calculated using the multiple measurements model[115]. The distribution of the full-auto differences was assessed using Bland-Altman analysis. The equivalence of the angles calculated using the full-auto method and the manual method was tested using paired Two One-Sided Tests (pTOST), which checks whether the mean difference falls within a predefined Limit of Agreement (LoA) margin. In this study, the intraobserver LoA was used, defined as $1.96 * \sigma_{intra}$, with σ_{intra} the intraobserver standard deviation derived from the manual measurements. As the pTOST necessitates independent data, the tests were performed on the left and right legs separately. The 25 angles with left and right separated resulted in a total of 50 comparisons. With Bonferroni correction applied, $p<0.001$ was necessary for significance.

5.2.10 Source of funding

Table 5.3: Similarity between the manual and automatic segmentations for the femur and tibia as expressed by the mean ± SD of the Dice Similarity Coefficient (DSC), Mean Absolute Surface Distance (MASD), Hausdorff Distance (HD) and 95th percentile Hausdorff Distance (HD95).

	DSC	MASD (mm)	HD (mm)	HD95 (mm)
Tibia	0.986 ± 0.003	0.23 ± 0.04	2.91 ± 1.53	0.67 ± 0.11
Femur	0.993 ± 0.003	0.14 ± 0.06	3.54 ± 5.54	0.58 ± 0.10

5.3 Results

Fifty subjects, 41 male (mean age: 61, range: 30-76 years) and 9 female (mean age: 51, range: 34-76 years) were included in this study.

5.3.1 Validation

The difference between the manual and automatic segmentations of the fifty subjects are shown in Table 5.3. Further analysis of the segmentation accuracy can be found in Kuiper et al.[162] The operators performing manual segmentation of the bones reported mean segmentation times of approximately 40 minutes per leg. The mean (±standard deviation) time for manual landmark indication was 20.3±3.2 minutes per leg. The mean time for the automatic workflow were 10.1±2.2 minutes for segmentation and 2.2±0.2 minutes for landmark indication.

Mean intraobserver and mean full-auto intermethod landmark difference are illustrated on a bone model in Figure 5.2. Figure 5.6 shows the distribution of the manual, semi-auto and full-auto landmarks differences. The mean distance over all landmarks was 2.01±1.64 mm for the intraobserver, 2.12±1.38 mm for the semi-auto and 2.17±1.37 mm for the full-auto difference. The full-auto intermethod distance ranged from 0.36±0.19 mm (FHC) to 3.63±2.21 mm (FGT).

Figure 5.7 shows the distribution of the manual, semi-auto and full-auto angle differences. The mean absolute difference in angle was 1.05±1.48° for the intraobserver, 0.98±1.15° for the semi-auto and 1.10±1.16° for the full-auto difference. The full-auto difference ranged from 0.21±0.15° (mHKA) to 2.75±1.71° (mPDFA).

5.3.2 Statistical analysis

The results of the ICC analysis on the angle measurements are shown in Table 5.4. Both intraobserver and full-auto reliability were good to excellent in 23 out of 25 measured angles. The Bland-Altman plots in Appendix A expose systematic bias in the full-auto difference of the apPDFA and mPDFA. The mean and confidence interval of the full-auto difference calculated using pTOST test are summarized in Appendix B.

5.3.3 Population variation assessment

Table 5.5 shows the alignment parameters of all fifty subjects as calculated using the full-auto method. The mean of 7 out of 25 angles fell outside of the expected normal range[2]. These were the apANSA and the three variations (ap, at and m) of the ADTA and LDTA.

*Table 5.4: Intraclass correlation coefficients with the 95% confidence interval for the measurement of the angles between the manual observers (**intraobserver**), the mean of the manual observers and the semi-auto method (**semi-auto**) and the mean of the manual observers and the full-auto method (**full-auto**). Guideline values indicating reliability: poor: <0.5, moderate: 0.5-0.75, good: 0.75-0.9, excellent: 0.9.*

		Intraobserver						Semi-auto						Full-auto				
			ICC						ICC						ICC			
	apADTA	0.77	<	0.87	<	0.93	0.87	<	0.93	<	0.96	0.87	<	0.91	<	0.95
	apANSA	0.96	<	0.98	<	0.99	0.96	<	0.98	<	0.99	0.96	<	0.97	<	0.98
	apLDFA	0.93	<	0.96	<	0.98	0.93	<	0.96	<	0.98	0.90	<	0.94	<	0.96
Anatomical Partial	apLDTA	0.62	<	0.78	<	0.88	0.81	<	0.90	<	0.95	0.76	<	0.86	<	0.92
	apMNSA	0.82	<	0.90	<	0.95	0.98	<	0.99	<	0.99	0.98	<	0.99	<	0.99
	apMPFA	0.98	<	0.99	<	0.99	0.97	<	0.99	<	0.99	0.98	<	0.99	<	0.99
	apMPTA	0.98	<	0.99	<	0.99	0.99	<	0.99	<	1.00	0.97	<	0.98	<	0.99
	apPDFA	0.13	<	0.42	<	0.64	0.00	<	0.46	<	0.71	0.10	<	0.43	<	0.63
	apPPTA	0.87	<	0.93	<	0.96	0.70	<	0.89	<	0.95	0.77	<	0.87	<	0.93
	atADTA	0.79	<	0.88	<	0.94	0.88	<	0.94	<	0.97	0.88	<	0.93	<	0.95
	atLDFA	0.95	<	0.97	<	0.99	0.96	<	0.98	<	0.99	0.94	<	0.96	<	0.98
Anatomical Total	atLDTA	0.61	<	0.78	<	0.87	0.80	<	0.89	<	0.94	0.75	<	0.85	<	0.91
	atMNSA	0.82	<	0.90	<	0.95	0.98	<	0.99	<	0.99	0.98	<	0.99	<	0.99
	atMPFA	0.98	<	0.99	<	1.00	0.98	<	0.99	<	1.00	0.98	<	0.99	<	0.99
	atMPTA	0.98	<	0.99	<	1.00	0.99	<	0.99	<	1.00	0.98	<	0.99	<	0.99
	atPPTA	0.91	<	0.95	<	0.98	0.80	<	0.93	<	0.97	0.88	<	0.93	<	0.96
	mADTA	0.77	<	0.87	<	0.93	0.87	<	0.93	<	0.96	0.87	<	0.92	<	0.95
	mHKA	0.99	<	0.99	<	1.00	0.99	<	1.00	<	1.00	1.00	<	1.00	<	1.00
	mJLCA	0.94	<	0.97	<	0.98	0.96	<	0.98	<	0.99	0.94	<	0.96	<	0.98
	mLDFA	0.92	<	0.96	<	0.98	0.94	<	0.97	<	0.98	0.92	<	0.95	<	0.97
Mechanical	mLDTA	0.58	<	0.76	<	0.86	0.76	<	0.87	<	0.93	0.72	<	0.84	<	0.90
	mLPFA	0.98	<	0.99	<	0.99	0.98	<	0.99	<	0.99	0.98	<	0.99	<	0.99
	mMPTA	0.98	<	0.99	<	0.99	0.99	<	0.99	<	1.00	0.98	<	0.98	<	0.99
	mPDFA	0.15	<	0.43	<	0.65	0.20	<	0.58	<	0.78	0.28	<	0.55	<	0.72
	mPPTA	0.87	<	0.93	<	0.96	0.76	<	0.92	<	0.97	0.86	<	0.92	<	0.96
	Mean	**0.81**	**<**	**0.89**	**<**	**0.94**	**0.84**	**<**	**0.92**	**<**	**0.96**	**0.85**	**<**	**0.91**	**<**	**0.94**

*Table 5.5: Mean of the automatically calculated angles describing lower limb alignment across 50 subjects, compared to the normal range for these angles reported by Paley[2]. Angles outside of the normal range are shown in **bold**.*

		Mean	Std.Dev.	Normal range	
				Min	Max
Anatomical Partial	**apADTA**	**85.6**	3.09	78	82
	apANSA	**161.6**	8.16	165	175
	apLDFA	80.4	1.60	79	83
	apLDTA	**84.6**	3.42	86	92
	apMNSA	128.6	5.14	124	136
	apMPFA	83.7	5.46	80	89
	apMPTA	85.0	2.41	85	90
	apPDFA	84.7	0.97	79	87
	apPPTA	82.3	2.04	77	84
Anatomical Total	**atADTA**	**83.6**	3.20	78	82
	atLDFA	80.0	1.84	79	83
	atLDTA	**85.6**	3.63	86	92
	atMNSA	127.5	4.90	124	136
	atMPFA	82.6	5.40	80	89
	atMPTA	86.0	2.97	85	90
	atPPTA	81.8	2.63	77	84
Mechanical	**mADTA**	**83.5**	3.21	78	82
	mHKA	179.0	2.64	182	176
	mJLCA	0.28	2.93	-2	2
	mLDFA	85.7	1.81	85	90
	mLDTA	**85.8**	3.50	86	92
	mLPFA	92.1	5.25	85	95
	mMPTA	86.2	2.55	85	90
	mPDFA	80.2	1.49	79	87
	mPPTA	81.7	2.28	77	85

5.4 Discussion

In this study, we developed a method that automatically calculated a comprehensive range of measurements that defined lower limb alignment. By comparing manual and automatic measurements for both landmark placement and angle calculation on twenty subjects, we found that our method achieved results that closely corresponded to intraobserver variability. Additionally, the automatic workflow required approximately 12 minutes, compared to approximately 60 minutes of manual operation. We applied the automatic method to a dataset of fifty subjects and found that the mean of most measurements fell within the expected range.

A benefit of fitting deformable bone models is that adding more landmarks is trivial, as indicating the landmark once is sufficient to find it on all subsequent subjects. This enables inclusion of landmarks without easily discernible morphological features, in contrast to methods that use specific bone morphology features[77,154–157]. Using the proposed method, further research could focus on defining better 3D measurements to assess the bone morphometry.

The mean 20.3±3.2 minutes necessary for manual landmark indication was similar to the time reported by Fürmetz et al.[76] (17.22±4.0 minutes). Additionally, intraobserver variability over all measurements were similar between our study (mean: 1.05, range:[0.15-4.68]) and Fürmetz et al.[76] (mean: 1.26, range:[0.18-4.64]).

Mean full-auto differences for landmarks and angles were 0.05 mm and 0.12° larger than semi-auto differences, indicating that automatic segmentations had only a small negative impact on the accuracy. Mean full-auto differences were also only 0.11 mm and 0.05° larger than intraobserver differences, showing that the accuracy of the manual and automatic workflow also closely corresponded.

For most angles, the intraobserver and intermethod ICC were similar, and indicative of good or excellent reliability. Only the apPDFA and mPDFA showed poor or moderate reliability for both intraobserver and full-auto differences.

The Bland-Altman analysis showed that the full-auto 95% confidence intervals are similar to the intraobserver LoA. The plots exposed no systematic bias, except for the apPDFA and mPDFA, which showed greater full-auto difference for more extreme angles. Qualitative inspection of these subjects showed that this was caused by a bias in the automatic method towards a more conservative estimate of the angle, by placing the FAP and FPP landmarks on approximately the same height along the distal femur.

Figure 5.6: Boxplots showing the intraobserver (red), semi-auto intermethod (green), and full-auto intermethod distance for each landmark (blue).

Figure 5.7: Boxplot showing the intraobserver (red), semi-auto intermethod (green), and full-auto intermethod (blue) absolute difference for each angle.
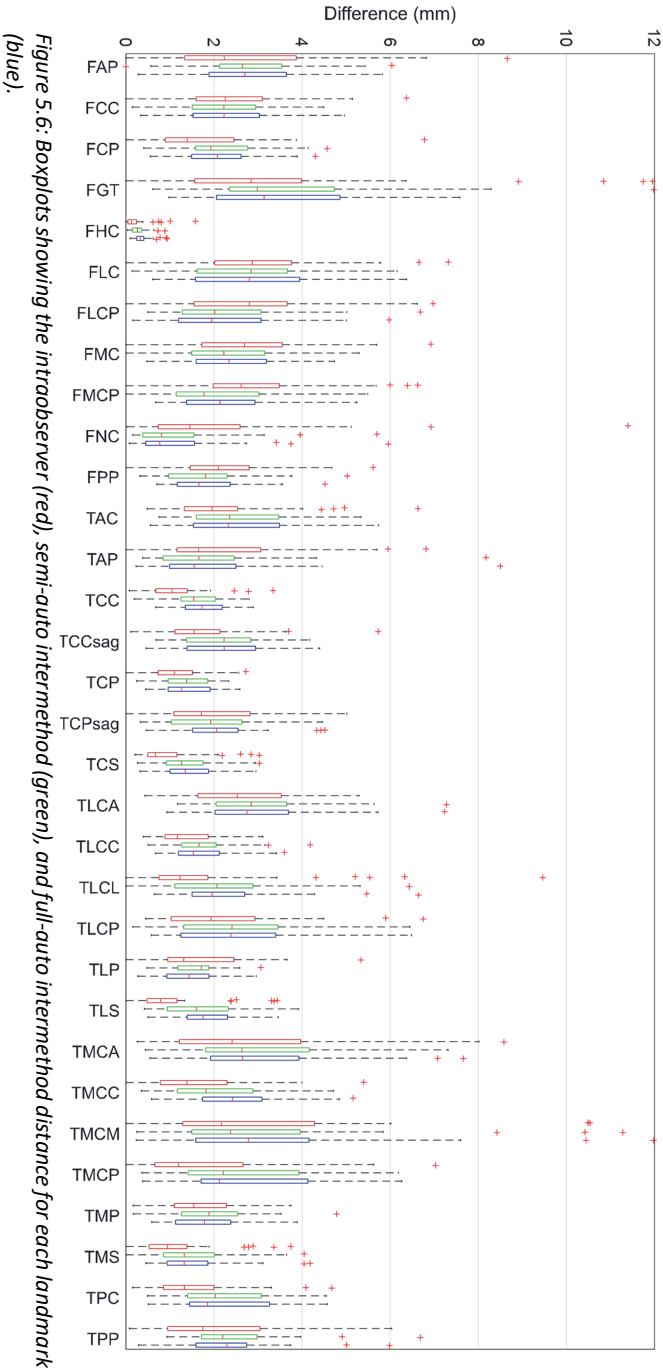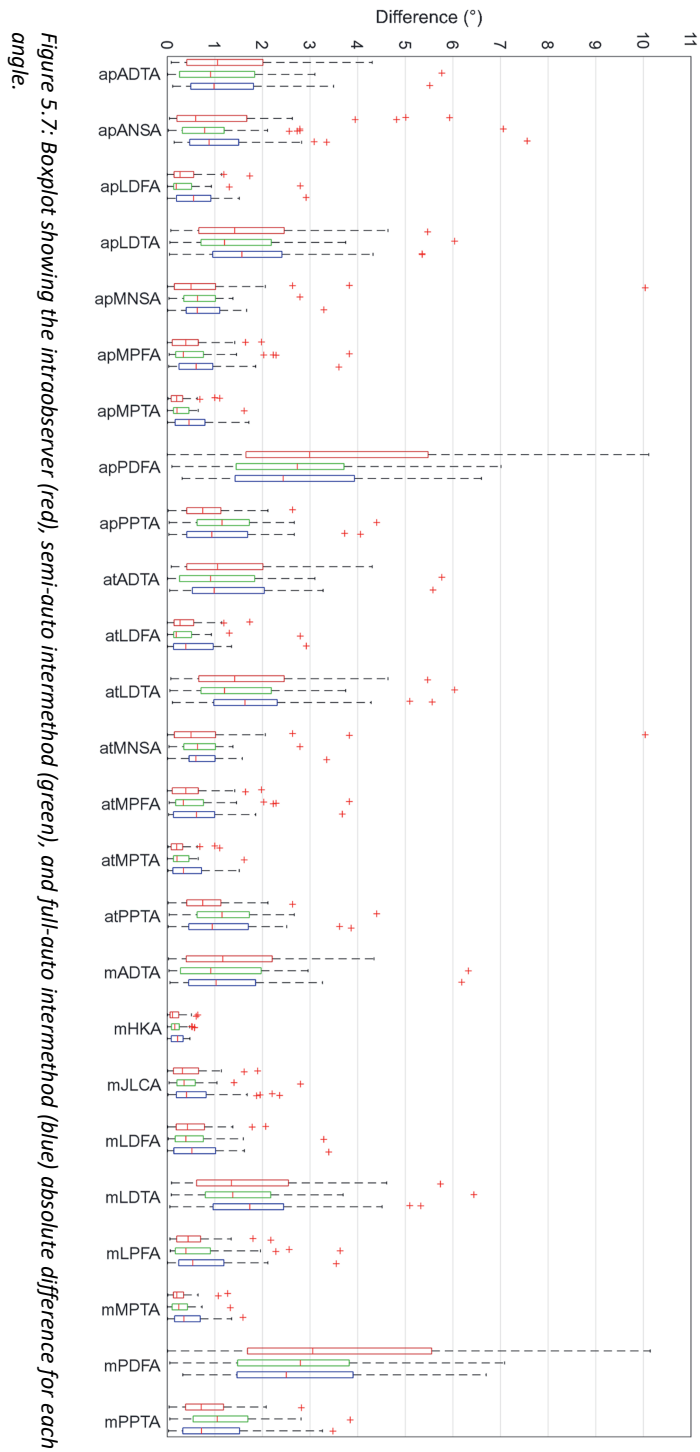
The pTOST statistical analysis showed statistical equivalence between the manual and automatic angles within the intraobserver LoA for all angles except for the apMPTA of the left leg with p=0.0018. This was slightly higher than the required p<0.001 which was required due to Bonferroni correction.

Assessing the mean angles on all fifty subjects, we found that the LDTA, ADTA and ANSA angles fell outside the expected normal ranges[2]. This could be due to differences in axial alignment between 2D and 3D scans, as this could affect alignment measurements[164]. Also, the demographics from this study population might be different, as age, gender and ethnicity could influence the mean alignment parameters[165–167].

A limitation was that the CT scans were taken in supine position, and thus not in a weight-bearing state. Roth et al.[95] noticed significant differences of 2.1±1.7° and 2.0±1.6° for the mHKA and mJLCA between supine and weight-bearing scans. However, only the mHKA and mJLCA were influenced by this effect. Solutions could be to either perform a standing CT scan[168], use an EOS scanner[169] or combine information from standing leg radiographs and CT scans to acquire a virtual standing CT scan[170].

Another limitation was that the subjects in this study did not have an indication of lower limb malalignment or pathologies such as osteoarthritis, which could influence the accuracy of both manual and automatic measurements. Inclusion of pathological and malaligned patients to study the relation between the morphometric parameters and clinical measures, and the inclusion of cadaveric samples for the direct measurement of these parameters, is therefore subject to further research.

In conclusion, the results of the proposed method closely corresponded to manual assessment, based both on segmentation accuracy, and landmark and angle calculation. It performed all steps fully automatically and within considerably less time. It would therefore be a valuable tool for fast and accurate lower limb alignment assessment.

5.5 Appendix A: Supplementary Figure
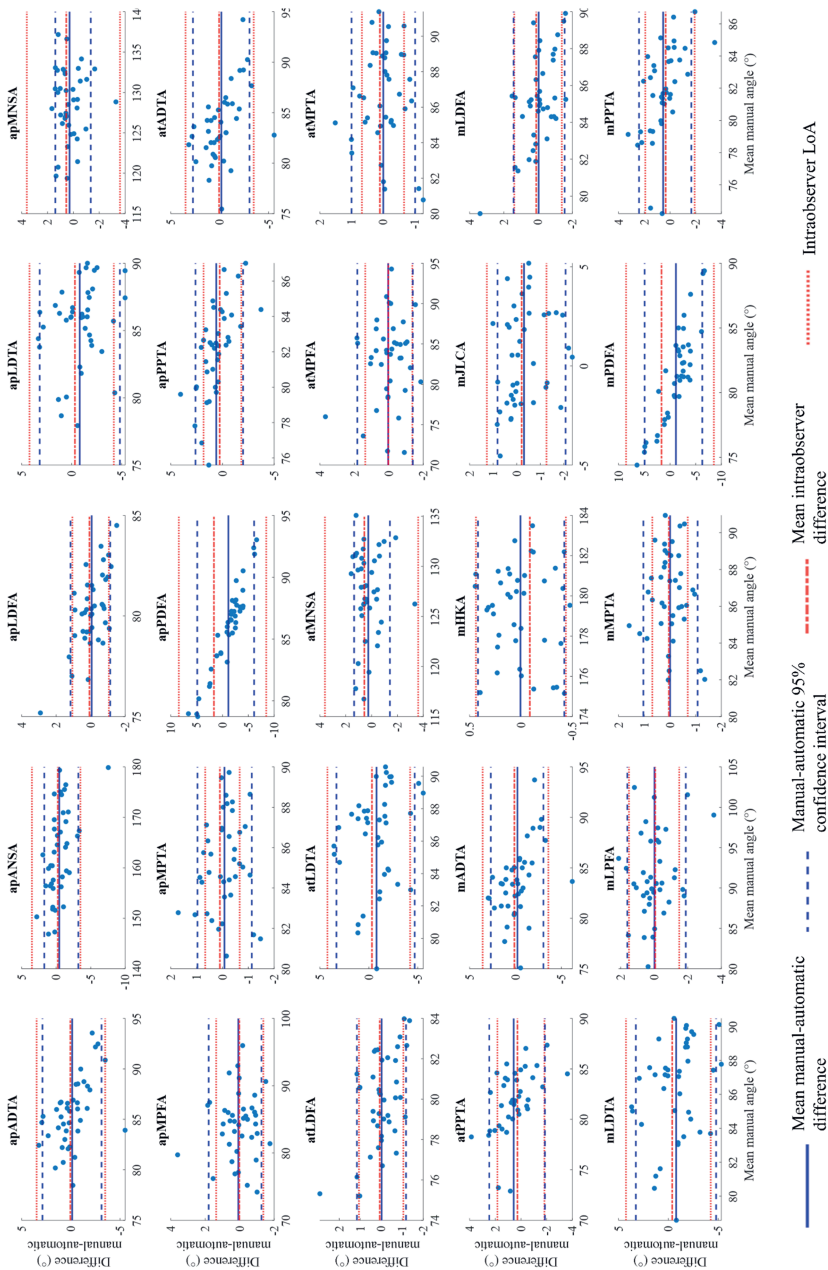
*Figure 5.8: Bland-Altman plots for each angle showing the differences between the manual and automatic angles. Striped blue lines indicate the mean and 95% Limits of Agreement between manual and automatic angles. Dotted red lines indicate the intraobserver mean and Limits of Agreement (LoA).*

5.6 Appendix B: Supplementary Table

*Table 5.6: Mean and 95% confidence interval (CI) of the mean for the difference between the manual and automatic angles calculated on the left and right legs of twenty patients, calculated using the paired Two One-Sided Test (TOST). All p-values were under 0.001 except for the angle shown in **bold**.*

		Left		95% CI		**Right**		95% CI	
		Mean	**p-value**	Lower	Upper	**Mean**	**p-value**	Lower	Upper
anatomical partial	apADTA	-0.38	2.3E-07	-0.34	1.10	0.65	5.2E-08	-1.24	-0.05
	apANSA	0.41	4.6E-09	-0.97	0.15	-0.14	7.2E-07	-1.24	0.29
	apLDFA	0.41	1.6E-04	-0.67	-0.16	-0.29	3.4E-04	-0.04	0.62
	apLDTA	0.50	3.4E-08	-1.26	0.25	1.11	3.4E-06	-1.99	-0.23
	apMNSA	-0.51	2.2E-15	0.26	0.75	-0.11	1.7E-11	-0.36	0.54
	apMPFA	-0.06	3.4E-05	-0.39	0.51	-0.10	1.5E-06	-0.24	0.43
	apMPTA	**0.17**	**1.8E-03**	**-0.40**	**0.18**	0.04	5.4E-05	-0.26	0.19
	apPDFA	1.18	2.3E-11	-2.13	-0.23	1.09	1.1E-08	-2.48	0.31
	apPPTA	-0.72	4.9E-04	0.19	1.20	-0.51	7.3E-04	-0.10	1.13
anatomical total	atADTA	-0.33	2.0E-07	-0.39	1.05	0.69	5.0E-08	-1.27	-0.10
	atLDFA	0.35	4.8E-05	-0.60	-0.10	-0.30	4.6E-04	-0.03	0.64
	atLDTA	0.45	7.1E-08	-1.26	0.36	1.14	2.7E-06	-1.99	-0.28
	atMNSA	-0.42	7.2E-16	0.19	0.66	-0.11	2.2E-11	-0.37	0.55
	atMPFA	0.02	2.7E-05	-0.47	0.43	-0.10	2.0E-06	-0.25	0.44
	atMPTA	0.10	3.1E-04	-0.31	0.22	-0.02	1.0E-05	-0.18	0.22
	atPPTA	-0.68	1.8E-04	0.19	1.13	-0.47	4.2E-04	-0.12	1.06
mechanical	mADTA	-0.26	1.9E-07	-0.49	1.03	0.69	3.1E-08	-1.27	-0.11
	mHKA	0.00	1.0E-07	-0.09	0.10	-0.01	6.1E-07	-0.10	0.11
	mJLCA	0.29	2.1E-04	-0.72	0.04	0.29	1.0E-05	-0.59	0.02
	mLDFA	0.49	1.3E-05	-0.77	-0.21	-0.41	1.3E-04	0.03	0.79
	mLDTA	0.40	6.9E-08	-1.23	0.42	1.16	2.8E-06	-2.02	-0.30
	mLPFA	-0.17	4.3E-05	-0.30	0.63	0.21	2.3E-06	-0.56	0.15
	mMPTA	0.05	2.1E-04	-0.28	0.28	0.00	2.2E-06	-0.19	0.18
	mPDFA	1.20	3.0E-11	-2.17	-0.23	1.11	9.3E-09	-2.50	0.28
	mPPTA	-0.61	1.1E-05	0.19	1.01	-0.47	1.4E-04	-0.10	1.04
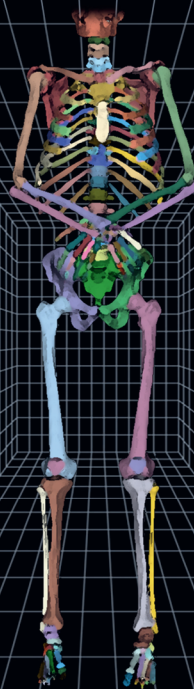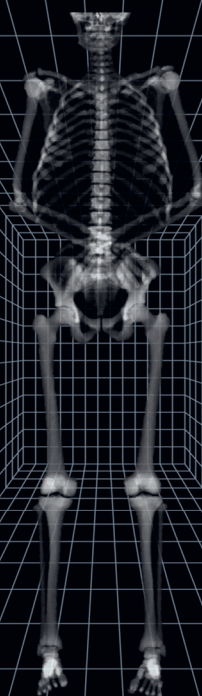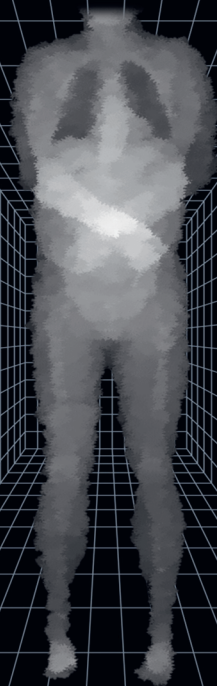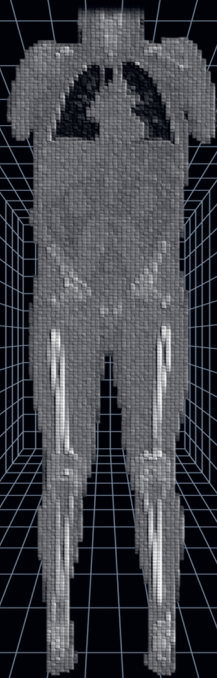
Chapter 6

Summary and Discussion

6.1 Summary

In this thesis, the ultimate goal has been to strive towards (automatic) computer assisted planning of reconstructive osteotomies. However, the quality of the output of such a workflow depends on the quality of all the intermediary steps, between the acquisition of a medical image, to the virtual reconstruction of the bone. In each chapter of this thesis we have tried to overcome some of the challenges posed by these steps, and to evaluate the intermediate results.

In **Chapter 2**, we developed a method to improve the registration of CT to MRI scans for images of the knee joint. Due to the large deformations that can arise because of the differences in patient positioning between different scanners, registration of such scans can be complex. We showed that non-rigid registration using deformable B-Splines could be improved by initializing the registration with a rigid registration of the bones, and then estimating the soft tissue deformation using a dual quaternion-based interpolation step.

In **Chapter 3**, an automatic and efficient deep learning based method for the semantic segmentation of six separate bones from lower extremity CT scans was proposed and validated. Due to the large field of view and high resolution of these scans, existing deep learning segmentation algorithms were found to be slow and memory intensive for this application. We optimized the architecture of a cascaded U-net approach, which outperformed the state-of-the-art nn-Unet[94] in memory efficiency, speed and accuracy. We then clinically validated the segmentations by performing automated morphometric measurements on the hip joint, which were found to have good to excellent correlation with the measurements derived from manual segmentations.

In **Chapter 4**, we evaluated the impact of using CT and MRI to segment bone and/or cartilage on the outcome of pre-operative osteotomy planning of the forearm in adolescent patients. To be able to directly compare the planning outcome, an automated, deterministic planning method was developed to avoid inter- or intraoperator variation in the planning. Segmentations of the bone from both CT and MRI were used to directly compare the impact of the different modalities on the planning. Additionally, segmentations of the bone and the bone combined with cartilage were derived from the MR images, to study whether inclusion of the cartilage, which forms a significant part of the joint in young patients, changed the planning outcome. The results showed that the realignment parameters were all excellently correlated between the different segmentations, although small significant differences were found in the translational part of the realignment. Furthermore, a positive correlation was found between the amount of cartilage and differences in planning on bone and bone with cartilage, indicating that for young patients it might be important not to neglect the impact of the cartilage on the shape of the joint when planning. However, due to the small sample size of ten patients, and the time-consuming and error prone manual segmentation of the CT and

MRI scans, further research is advised to further study if, and how, MRI might replace CT as the golden standard for forearm osteotomy planning.

Finally, in **Chapter 5**, we evaluated a method to automatically quantify the alignment parameters that describe the morphology of the lower extremities from 3D bone models. By using the automated segmentations studied in **Chapter 4**, we were able to develop a completely automatic pipeline to extract these parameters using a CT scan as input. The alignment parameters were calculated using landmarks proposed by earlier studies[76] and expanded upon in order to quantify all the parameters needed for lower extremity realignment planning as described by Paley[2]. The automatic method completely removed the need for manual operation while the results closely corresponded to the intrarater variability. This could pave the way for automatic realignment planning of the lower extremities, which often lacks a healthy contralateral example such as is required for the planning of unilateral bone realignment.

6.2 Discussion

6.2.1 Choosing an imaging modality

A challenge posed by 3D osteotomy planning is that accurate 3D bone models are essential. These 3D bone models in turn require 3D image acquisition. As mentioned in the Introduction, there are two feasible options for this, with their own pros and cons: CT and MRI.

When using CT, the acquisition is relatively fast, the field of view can be large, and the resolution can be high. Additionally, bone has a clearly distinguishable contrast on CT, which makes segmentation of the bone easier. However, the effective radiation dose of CT can be over two hundred times higher than in a planar radiograph[171], resulting in increased chances of radiation induced malignancy, especially in paediatric patients. In case of severe malalignment or older patients, this can be understandable, but in milder cases or younger patients, it is important to carefully evaluate if the benefits outweigh the risks of using CT[172].

In contrast, MRI is relatively slow, acquisitions with large fields of view might suffer from geometric distortion and bias fields due to magnetic field inhomogeneities, and high resolutions require even longer scanning times. In addition, bone often has a similar contrast to surrounding tissues that also have low free water content, such as tendons and ligaments. However, soft tissues are more clearly distinguishable from each other than on CT. And most importantly, the patient does not suffer from any harmful radiation[14].

The issues with both these modalities were therefore the reason to develop an improved CT to MRI registration method in **Chapter 2**. Firstly, because the combined information of the two modalities in one reference frame could aid the orthopaedic surgeon or clinical technician during diagnosis and planning, as it would be easier to visualize and segment both bone and soft tissue such as cartilage. Secondly, because the accurate registration of a large set of

orthopaedic CT and MRI could aid in the development of synthetic CT (sCT). sCT is a CT-like contrast derived from MR images, which is constructed either by atlas-based mapping of the MR image to a reference CT image, or by mapping each voxel in the MR image to an intensity in the CT range, based on the intensity of the voxel and its neighbourhood. Recently, this voxel-based mapping has been performed mostly using deep learning methods, and the proposed registration method in this thesis has already been applied in other research to produce the required training data[173].

6.2.2 Need for automated segmentation

Another challenge encountered when 3D planning osteotomies is the necessary segmentation of bone from the scans. Although CT offers a relatively clear contrast of bone compared to MRI, segmentation and separation of the different bones is still not trivial. Additionally, interoperator variation caused by manual bone segmentation, exacerbated by differences in segmentation from different modalities, might influence the outcome of osteotomy planning, as we noticed in **Chapter 4**. Furthermore, segmentation of bone might be so time-consuming that it discourages the use of 3D planning methods in practice.

Therefore, in **Chapter 3**, we proposed an automatic method for the segmentation of bone from CT, in order to alleviate some of these concerns. Ideally, however, we would like to combine the information and benefits of both MRI and CT. This could be done either by training segmentation algorithms directly on MRI, or by using the aforementioned sCT contrasts in combination with CT segmentation algorithms. It is therefore recommended that the application of our deep learning network, which is not necessarily modality specific, to bone segmentation from MRI and/or sCT, is further examined. In either case, this would combine both the bone and soft tissue information in a single, non-harmful MRI scan. The only downsides remaining are then the inherently slower speed of MRI scans, and the geometric distortions due to magnetic field inhomogeneities when scanning larger areas. In cases where only a smaller field of view is required, such as the malaligned bones in the forearms, this might not be a significant problem. In case of malalignment in larger field of view areas, such as the lower extremities, an option might be to scan the patient in multiple (overlapping) scans. Geometric distortions due to changes in patient positioning should then be avoided by carefully maintaining the same position between scans[14], or by registering overlapping parts of the scans[174].

6.2.3 Optimal realignment planning

A crucial part in automatic osteotomy planning for malaligned bones is to establish what the correct (or optimal) alignment should be. There are two main options to establish this, where the choice for either one is made based on the availability of the information:

In case of unilateral malalignment, the contralateral side could be used as an example, as was done in **Chapter 4**. This is the most often used method in case of bone malalignment due to trauma, as this often affects only one side of the body[20,129,132,143,175]. However, although some

studies have shown that bilateral difference are insignificant[176,177], others have shown that significant bilateral differences in bone morphology do occur in healthy subjects[178,179]. Additionally, after extended periods of malalignment, the contralateral, originally healthy joints may have adapted to compensate for the unilateral malalignment. In these cases, exact replication of the healthy contralateral morphology may not result in the expected functional outcome.

An alternative method to realignment based on the healthy contralateral side, is to realign the bone in order to achieve the normal biomechanical function of the bone and corresponding joints. The first step towards this method is to define the current morphology of the entire functional part of the body, as we have attempted in **Chapter 5** for the lower extremities. When the morphology of the currently dysfunctional body part is known, and the desired, correct morphology and biomechanical function has been established, it is then possible to strive to recreate an equivalently functioning unit. It is important to note that in this case we do not only look at the malaligned bone itself, but also at the morphology of the surrounding bones and joints. It has been shown that the positioning of the patient while being scanned influences the alignment of the bones with respect to each other and the mechanical axes of the body, such as the mechanical alignment of the legs when standing up or lying down[95,146]. In future studies, these changes should therefore be taken into consideration when assessing the alignment of the lower extremities for example. This could be done using CT scanners that allow for a standing patient position[180,181], 3D reconstruction from bi-planar standing radiographs[182], or by registration of CT or MR images to standing radiographs[183,184].

6.2.4 Translation of planning to patient

After diagnosis, pre-operative planning and evaluation by the surgeon, the final step that needs to be taken to remedy the malalignment is the translation of the orthopaedic planning to the patient in the surgical operating room. However, the accuracy with which this plan is implemented is dependent on several factors, such as the clarity of the instructions, the method of application, and finally the skill of the surgeon. Three main methods exist or are in development, which might aid the surgeon in making this translation.

Image guided intra-operative navigation establishes a relationship between pre-operatively acquired images and orthopaedic planning, and the intra-operatively acquired information of the patient. The correspondence between the pre-operative image and the intra-operative patient anatomy is found using corresponding fiducials or by image-to-image registration[185]. The spatial information can be acquired in various combinations of modalities, with optical systems, radiographs, fluoroscopy, CT scans, and more recently also MRI, being the most often used choices[186–188].

Another method used for intraoperative surgical guidance is the use of patient specific surgical guides. These guides are designed on computers based on the pre-operative 3D planning. They are then created using 3D printing, as this allows for the creation of complex, patient specific geometries while requiring minimal materials. Although some singular

studies have shown that the use of these guides was associated with a higher precision in replicating the planned surgery[189–192], various meta-analyses found that it was not yet proven to improve long term clinical outcome[193–195].

The most recent development in intraoperative surgical guidance has been the introduction of AR systems to the operating room[196]. It is a form of image guided intra-operative navigation, in which a 3D image with information pertaining to the operation are overlaid over the surgeon's view of the patient. This enables the surgeon to simultaneously view both the digital representation and the real world[197].

In conclusion, use of these intraoperative guidance techniques might be vital to the adoption of computer assisted planning in the clinic, as much of the precision and accuracy offered by such a planning might be lost if it cannot be ascertained that it is correctly applied in the operation room. However, although various studies have shown that the use of intra-operative navigation increases the precision with which a surgery is performed, clinical evidence of the post-operative benefits is more scarce[198]. This makes it unclear if the benefits outweigh the additional costs and time spent on each patient, and further research into the clinical benefits is therefore advised[185,196,199,200].

6.2.5   Future perspectives in automated orthopaedic planning

In this thesis, we have worked towards an automatic method for the pre-operative planning of osteotomies, in order to achieve an efficient, reproducible and accurate planning outcome. Although these advances seem promising, various factors need to be considered when trying to implement such techniques in the clinic.

A significant part of the reasoning behind the goal of automated planning was that it would reduce interoperator variability, and thus reduce the impact interpersonal differences in experience or skill of an orthopaedic surgeon on surgery outcome. Although automated planning seems promising for relatively standard cases, there always remain factors that are difficult, if not impossible, to capture in a one-size-fits-all computer algorithm. It therefore remains essential that each automated registration, segmentation, diagnosis or planning is carefully checked by an orthopaedic surgeon. These algorithms should therefore never aim to completely replace the role of the surgeon, but rather assist in the decision-making process.

This thesis is not the first work on automated orthopaedic planning, as many authors before have developed, researched and evaluated techniques to automate alignment assessment[127,154,155,157] and pre-operative planning[20,131,132,201,202]. Additionally, various computer-assisted surgical planning tools are currently commercially available that assist in different steps of the planning process[22]. However, these computer-assisted surgical planning tools still require manual input, which not only introduces interoperator variability, but often also requires expert knowledge of the software. Due to this, a completely automated planning workflow for long bone malalignment has not yet been widely adopted for use in the clinic. To attain this, we found that there are three main challenges to overcome.

Firstly, the automated planning tools need to be developed into an accessible and easy to use software package. Ideally, no prior experience with 3D modelling is required, but only orthopaedic knowledge pertaining to the correct surgical solution. In this way, the direct control over the orthopaedic planning is transferred back from the clinical technicians or companies that usually assist during orthopaedic planning, to the orthopaedic surgeon. Secondly, for a commercially available planning tool, extensive validation tests are necessary to adhere to the medical device regulations that are in place in the countries where it would be adopted, such as the FDA[203] and European Union[204] medical device regulations. Finally, the technique needs to be accepted and adopted by orthopaedic surgeons. This means that surgeons need to be correctly informed about the use of the tool and its advantages and disadvantages.

Nederlandse Samenvatting

Het bewegingsapparaat is een orgaansysteem dat bestaat uit botten, spieren, gewrichten, pezen, ligamenten, kraakbeen en ander bindweefsel. De functie van het bewegingsapparaat omvat het mogelijk maken van beweging, het bieden van ondersteuning en stabiliteit, en het beschermen van de rest van de organen van het menselijk lichaam. Wanneer aandoeningen zoals botafwijkingen het bewegingsapparaat aantasten, vermindert dit de effectiviteit waarmee het bewegingsapparaat deze functionaliteit kan bieden.

Botafwijkingen kunnen een verkeerde botuitlijning veroorzaken, waardoor de oriëntatie of positie van de gewrichten afwijkt van de normale uitlijning. Dit kan leiden tot cosmetische afwijkingen of pijn, en in ernstige gevallen kan ook functieverlies optreden, variërend van een verminderde bewegingsvrijheid tot volledig verlies van het gebruik van een ledemaat. Bovendien kunnen botafwijkingen degeneratieve aandoeningen zoals artrose veroorzaken, als gevolg van ongelijkmatige belasting van het kraakbeen. Hoewel milde gevallen nog gecorrigeerd kunnen worden met behulp van fysiotherapie, kan er bij ernstige problemen worden overgegaan tot een operatie, 'osteotomie' genaamd. Hierbij wordt het gedeformeerde bot doorgezaagd, waarna de botdelen op de correcte wijze weer aan elkaar worden gemaakt.

Osteotomieën zijn complexe chirurgische procedures, waarbij pre-operatieve planning door de orthopeed vereist is. Dit kan aan de hand van een twee dimensionale (2D) Röntgen-foto, maar om dit nauwkeuriger en drie dimensionaal (3D) te kunnen doen, wordt gebruik gemaakt van een CT of MRI scan. Als de afwijking eenzijdig (unilateraal) is, kan de gezonde contralaterale tegenhanger als voorbeeld worden gebruikt, wat vaak het geval is als de afwijking wordt veroorzaakt door trauma. Wanneer de afwijking tweezijdig (bilateraal) is, zal de juiste uitlijning moeten worden geschat door de botmorfologie te vergelijken met die van de gezonde populatie. Dit kan worden gedaan op basis van metingen zoals lengte en bepaalde hoeken tussen oriëntatiepunten, of op basis van gemiddelde botmodellen die zijn geconstrueerd op basis van metingen in een grotere populatie.

Pre-operatieve planning van osteotomieën is een complexe procedure die meerdere handmatige stappen vereist, welke tijdrovend zijn en vatbaar voor inter- en intrapersoonlijke variatie. In dit proefschrift hebben we methodes ontwikkeld, bestudeerd en gevalideerd met het doel de pre-operatieve planning van osteotomiën te verbeteren of vergemakkelijken. Dit werd bereikt door het combineren van informatie over botten en zacht weefsel, en door automatiseren van de handmatige stappen om een automatische workflow mogelijk te maken voor de diagnose, kwantificering en reconstructie van het gedeformeerde bot.

Het samenvoegen van MRI en CT beelden is onderzocht in Hoofdstuk 2, waar een methode wordt voorgesteld om de registratie van CT op MRI-scans voor beelden van het kniegewricht te verbeteren. Dit maakt het mogelijk om de orthopeed gelijktijdig informatie over het bot alsmede het zachte weefsel te verschaffen. Vanwege de grote vervormingen die kunnen ontstaan door de verschillen in patiëntpositionering tussen verschillende scanners, kan registratie van dergelijke scans complex zijn. We toonden aan dat niet-rigide registratie met behulp van vervormbare B-Splines kan worden verbeterd door de registratie te initialiseren met een rigide registratie van de botten en vervolgens de vervorming van het zachte weefsel te schatten met behulp van een op dubbele quaternionen gebaseerde interpolatiestap.

In Hoofdstuk 3 ontwikkelden en valideerden we een automatische en efficiënte op deep learning gebaseerde methode voor de segmentatie van zes afzonderlijke botten van CT-scans van de onderste extremiteit. Vanwege het grote volume en de hoge resolutie van deze scans, bleken bestaande op deep learning gebaseerde segmentatie-algoritmes traag en geheugenintensief. We hebben de architectuur van een U-net geoptimaliseerd, die beter presteerde dan een bestaande state-of-the-art methode op het gebied van geheugenefficiëntie, snelheid en nauwkeurigheid. Vervolgens hebben we de segmentaties klinisch gevalideerd door geautomatiseerde morfometrische metingen aan het heupgewricht uit te voeren, die een goede tot uitstekende correlatie bleken te hebben met de metingen afgeleid van handmatige segmentaties.

In Hoofdstuk 4 evalueerden we de impact van segmentaties van bot en/of kraakbeen afgeleid van CT en MRI op de uitkomst van preoperatieve osteotomieplanning van de onderarm bij adolescente patiënten. Om de planningsresultaten direct te kunnen vergelijken, is een geautomatiseerde, deterministische planningsmethode ontwikkeld om variaties tussen operators in de planningen te voorkomen. Segmentaties van het bot van zowel CT als MRI werden gebruikt om de impact van de verschillende modaliteiten op de planning direct te vergelijken. Bovendien werden segmentaties van het bot en het bot gecombineerd met kraakbeen afgeleid van de MR-beelden, om te onderzoeken of opname van het kraakbeen, dat een belangrijk onderdeel van het gewricht vormt bij jonge patiënten, de planningsresultaten veranderde. De resultaten toonden aan dat hoewel de reconstructie parameters uitstekend gecorreleerd waren tussen de verschillende segmentaties, er kleine maar klinisch significante verschillen werden gevonden in het translationele deel van de reconstructie. Verder werd een positieve correlatie gevonden tussen de hoeveelheid kraakbeen en verschillen in planning op bot en bot met kraakbeen, wat aangeeft dat het voor jonge patiënten belangrijk kan zijn om de impact van het kraakbeen op de vorm van het gewricht niet te verwaarlozen bij het plannen.

Ten slotte hebben we in Hoofdstuk 5 een methode geëvalueerd om automatisch de uitlijningsparameters van de onderste extremiteiten te kwantificeren op basis van 3D-botmodellen. Door gebruik te maken van de geautomatiseerde segmentaties uit Hoofdstuk 4, waren we in staat om een volledig automatische pijplijn te ontwikkelen om deze parameters te extraheren met alleen een CT-scan als invoer. De automatische methode maakte

handmatige handelingen volledig overbodig, terwijl de resultaten nauw overeenkwamen met de intrabeoordelaarsvariabiliteit. Dit zou de weg kunnen banen voor automatische osteotomie planning van de onderste ledematen, waarbij een gezond contralateraal voorbeeld vaak ontbreekt.

Samenvattend zijn er in dit proefschrift verscheidene geautomatiseerde methodes ontwikkeld die kunnen assisteren bij de diagnose van botafwijkingen en de pre-operative planning van osteotomiën. Een belangrijke beweegreden voor het ontwikkelen van deze geautomatiseerde methodes was dat het de invloed van interpersoonlijke verschillen in ervaring of vaardigheden van een orthopedisch chirurg zou verminderen. Hoewel de mogelijkheden van de moderne computeronsteunde methodes veelbelovend zijn, blijven er altijd factoren die moeilijk, zo niet onmogelijk, vast te leggen zijn in een pasklaar computeralgoritme. Het blijft daarom essentieel dat elke geautomatiseerde registratie, segmentatie, diagnose of planning zorgvuldig wordt gecontroleerd door een orthopedisch chirurg. Deze algoritmes zullen daarom de rol van de orthopeed niet snel vervangen, maar eerder een hulpmiddel vormen bij het besluitvormingsproces.

Bibliography

1. Cieza A, Causey K, Kamenov K, et al. 2021. Global estimates of the need for rehabilitation based on the Global Burden of Disease study 2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet (London, England) 396(10267):2006–2017.

2. Paley D. 2002. Principles of Deformity Correction. Springer Berlin Heidelberg.

3. Furnstahl P, Schweizer A, Graf M, et al. 2016. Surgical Treatment of Long-Bone Deformities : 3D Preoperative Planning and Patient-Specific Instrumentation. In: Computational Radiology for Orthopaedic Interventions. p 123–149.

4. Madry H, Kon E, Condello V, et al. 2016. Early osteoarthritis of the knee. Knee Surgery, Sport. Traumatol. Arthrosc. 24(6):1753–1762.

5. Santoro D, Tantavisut S, Aloj D, Karam MD. 2014. Diaphyseal osteotomy after post-traumatic malalignment. Curr. Rev. Musculoskelet. Med. 7(4):312–322.

6. Hunter DJ, Bierma-Zeinstra S. 2019. Osteoarthritis. Lancet 393(10182):1745–1759.

7. Vasluian E, Van Der Sluis CK, Van Essen AJ, et al. 2013. Birth prevalence for congenital limb defects in the northern Netherlands: a 30-year population-based study.

8. Beerekamp MSH, de Muinck Keizer RJO, Schep NWL, et al. 2017. Epidemiology of extremity fractures in the Netherlands. Injury 48(7):1355–1362.

9. Brouwer RW, Verhaar JAN. 2004. Osteotomie ter hoogte van de knie voor jonge patiënten met gonartrose. Ned. Tijdschr. Geneeskd. 148(40):1955–1960.

10. Losina E, Weinstein AM, Reichmann WM, et al. 2013. Lifetime risk and age at diagnosis of symptomatic knee osteoarthritis in the US. Arthritis Care Res. 65(5):703–711.

11. Stoddart JC, Dandridge O, Garner A, et al. 2021. The compartmental distribution of knee osteoarthritis – a systematic review and meta-analysis. Osteoarthr. Cartil. 29(4):445–455.

12. Maderbacher G, Matussek J, Greimel F, et al. 2021. Lower Limb Malrotation Is Regularly Present in Long-Leg Radiographs Resulting in Significant Measurement Errors. J. Knee Surg. 34(1):108–114.

13. Fadero PE, Shah M. 2014. Three dimensional (3D) modelling and surgical planning in trauma and orthopaedics. Surgeon 12(6):328–333.

14. Florkow MC, Willemsen K, Mascarenhas V V, et al. 2022. Magnetic Resonance

Imaging Versus Computed Tomography for Three-Dimensional Bone Imaging of Musculoskeletal Pathologies: A Review.

15. Mosher ZA, Sawyer JR, Kelly DM. 2018. MRI Safety with Orthopedic Implants. Orthop. Clin. North Am. 49(4):455–463.

16. Blankstein A. 2011. Ultrasound in the diagnosis of clinical orthopedics: The orthopedic stethoscope. World J. Orthop. 2(2):13–24.

17. 2018. Collins English Dictionary, 12th ed. Glasgow: HarperCollins Publishers.

18. Chapman MW. 2001. Chapman's orthopaedic surgery. Lippincott Williams & Wilkins.

19. Mauler F, Langguth C, Schweizer A, et al. 2017. Prediction of normal bone anatomy for the planning of corrective osteotomies of malunited forearm bones using a three-dimensional statistical shape model. J. Orthop. Res. 35(12):2630–2636.

20. Carrillo F, Roner S, von Atzigen M, et al. 2020. An automatic genetic algorithm framework for the optimization of three-dimensional surgical plans of forearm corrective osteotomies. Med. Image Anal. 60:101598.

21. Taha AA, Hanbury A. 2015. Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. BMC Med. Imaging 15(1):1–28.

22. Wang M, Li D, Shang X, Wang J. 2020. A review of computer-assisted orthopaedic surgery systems. Int J Med Robot.

23. Picard F, Deakin AH, Riches PE, et al. 2019. Computer assisted orthopaedic surgery: Past, present and future. Med. Eng. Phys. 72:55–65.

24. Oei EHG, Ginai AZ, Hunink MGM. 2007. MRI for Traumatic Knee Injury: A Review. Semin. Ultrasound, CT MRI 28(2):141–157.

25. Maintz JBA, Viergever MA. 1998. A survey of medical image registration. Med. Image Anal. 2(1):1–36.

26. Viergever MA, Maintz JBA, Klein S, et al. 2016. A survey of medical image registration – under review. Med. Image Anal. 33:140–144.

27. Klein S, Staring M, Murphy K, et al. 2010. elastix A Toolbox for Intensity-Based Medical Image Registration. IEEE Trans. Med. Imaging 29(1):196–205.

28. Avants BB, Tustison NJ, Song G, et al. 2011. A reproducible evaluation of ANTs similarity metric performance in brain image registration. Neuroimage 54(3):2033–2044.

29. Vercauteren T, Pennec X, Perchant A, Ayache N. 2009. Diffeomorphic demons: efficient non-parametric image registration. Neuroimage 45(1 Suppl):S61–S72.

30. de Vos BD, Berendsen FF, Viergever MA, et al. 2019. A deep learning framework for unsupervised affine and deformable image registration. Med. Image Anal.

52:128–143.

31.   Yang X, Kwitt R, Styner M, Niethammer M. 2017. Quicksilver: Fast predictive image registration – A deep learning approach. Neuroimage 158(March):378–396.

32.   Eppenhof KAJ, Lafarge MW, Veta M, Pluim JPW. 2019. Progressively trained convolutional neural networks for deformable image registration. IEEE Trans. Med. Imaging 39(5):1594–1604.

33.   Li H, Fan Y. 2018. Non-rigid image registration using self-supervised fully convolutional networks without training data. In: Proceedings - International Symposium on Biomedical Imaging. p 1075–1078.

34.   Staring M, Klein S, Pluim JPW. 2007. A rigidity penalty term for nonrigid registration. Med. Phys. 34(11):4098–4108.

35.   Berendsen FF, Kotte ANTJ, Viergever MA, Pluim JPW. 2014. Registration of organs with sliding interfaces and changing topologies. Proc. SPIE 9034:90340E.

36.   Lombaert H, Grady L, Pennec X, et al. 2014. Spectral log-demons: Diffeomorphic image registration with very large deformations. Int. J. Comput. Vis. 107(3):254–271.

37.   Maes F, Collignon A, Vandermeulen D, et al. 1997. Image Registration by Maximization of Mutual Information. IEEE Trans. Med. Imaging 16(2):187–197.

38.   Little JA, Hill DLG, Hawkes DJ. 1997. Deformations Incorporating Rigid Structures. Comput. Vis. Image Underst. 66(2):223–232.

39.   Zhang J, Yan CH, Chui CK, et al. 2011. Multimodal image registration system for image-guided orthopaedic surgery. Mach. Vis. Appl. 22(5):851–863.

40.   Lu ZT, Feng QJ, Yang W, Chen WF. 2011. Hybrid rigid and non-rigid registration algorithm for alignment of   intra-subject thoracic and abdominal images. Imaging Sci. J. 59(3):127–133.

41.   Walimbe V, Shekhar R. 2006. Automatic elastic image registration by interpolation of 3D rotations and translations from discrete rigid-body transformations. Med. Image Anal. 10(6):899–914.

42.   Foruzan AH, Motlagh HR. 2015. Multimodality liver registration of Open-MR and CT scans. Int. J. Comput. Assist. Radiol. Surg. 10(8):1253–1267.

43.   Dixon WT. 1984. Dixon simple proton spectroscopic imaging. Radiology 153:189–194.

44.   Eggers H, Brendel B, Duijndam A, Herigault G. 2011. Dual-echo Dixon imaging with flexible choice of echo times. Magn. Reson. Med. 65(1):96–107.

45.   Ronneberger O, Fischer P, Brox T. 2015. U-net: Convolutional networks for biomedical image segmentation. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 9351:234–241.

46. Milletari F, Navab N, Ahmadi S-AA. 2016. V-Net: Fully convolutional neural networks for volumetric medical image segmentation. Proc. - 2016 4th Int. Conf. 3D Vision, 3DV 2016 :565–571.

47. Kikinis R, Pieper SD, Vosburgh KG. 2014. 3D Slicer: A Platform for Subject-Specific Image Analysis, Visualization, and Clinical Support. In: Jolesz FA, editor. Intraoperative Imaging and Image-Guided Therapy. New York, NY: Springer New York. p 277–289.

48. Lorensen WE, Cline HE. 1987. Marching cubes: A high resolution 3D surface construction algorithm. In: Proceedings of the 14th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1987. p 163–169.

49. Besl PJ, Mckay ND. 1992. A Method for Registration of 3-D Shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14(2):239–256.

50. Shepard D. 1968. A two-dimensional interpolation function for irregularly-spaced data. In: Proceedings of the 1968 23rd ACM national conference on -. p 517–524.

51. Hamilton WR. 1844. On quaternions; or on a new system of imaginaries in algebra. London, Edinburgh, Dublin Philos. Mag. J. Sci. .

52. Shoemake K. 1985. Animating rotation with quaternion curves. ACM SIGGRAPH Comput. Graph. 19(3):245–254.

53. Kavan L, Collins S, Žára J, O'Sullivan C. 2008. Geometric skinning with approximate dual quaternion blending. ACM Trans. Graph. 27(4):1–23.

54. Crum WR, Camara O, J.Hawkes D, Hawkes DJ. 2007. Methods for Inverting Dense Displacement Fields: Evaluation in Brain Image Registration. Med. Image Comput. Comput. Assist. Interv. 10(Pt 1):659–666.

55. Amidror I. 2002. Scattered data interpolation methods for electronic imaging systems: a survey. J. Electron. Imaging 11(2):157.

56. Pluim JPW, Maintz JBA, Viergever MA. 2003. Mutual Information Based Registration of Medical Images: A Survey. IEEE Trans. Med. Imaging 22(8):986–1004.

57. Rohlfing T. 2012. Image Similarity and Tissue Overlaps as Surrogates for Image Registration Accuracy: Widely Used but Unreliable. IEEE Trans. Med. Imaging 31(2):153–163.

58. Breighner RE, Bogner EA, Lee SC, et al. 2019. Evaluation of Osseous Morphology of the Hip Using Zero Echo Time Magnetic Resonance Imaging. Am. J. Sports Med. 47(14):3460–3468.

59. Du J, Bydder GM. 2013. Qualitative and quantitative ultrashort-TE MRI of cortical bone. NMR Biomed. 26(5):489–506.

60. Florkow MC, Zijlstra F, Willemsen K, et al. 2020. Deep learning–based MR-to-CT synthesis: The influence of varying gradient echo–based MR images as input

channels. Magn. Reson. Med. 83(4):1429–1441.

61. Jans LBO, Chen M, Elewaut D, et al. 2021. MRI-based Synthetic CT in the Detection of Structural Lesions in Patients with Suspected Sacroiliitis: Comparison with MRI. Radiology 298(2):343–349.

62. Bron EE, Van Tiel J, Smit H, et al. 2013. Image registration improves human knee cartilage T1 mapping with delayed gadolinium-enhanced MRI of cartilage (dGEMRIC). Eur. Radiol. 23(1):246–252.

63. Leibfarth S, Mönnich D, Welz S, et al. 2013. A strategy for multimodal deformable image registration to integrate PET/MR into radiotherapy treatment planning. Acta Oncol. (Madr). 52(7):1353–1359.

64. Avants BB, Epstein CL, Grossman M, Gee JC. 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. 12(1):26–41.

65. Reaungamornrat S, De Silva T, Uneri A, et al. 2016. MIND Demons: Symmetric Diffeomorphic Deformable Registration of MR and CT for Image-Guided Spine Surgery. IEEE Trans. Med. Imaging 35(11):2413–2424.

66. Tacchella JM, Roullot E, Lefort M, et al. 2014. An efficient strategy based on an individualized selection of registration methods. Application to the coregistration of MR and SPECT images in neuro-oncology. Phys. Med. Biol. 59(22):6997–7011.

67. Haskins G, Kruger U, Yan P. 2020. Deep learning in medical image registration: a survey. Mach. Vis. Appl. 31(1).

68. Kingma DP, Ba J. 2015. Adam: A Method for Stochastic Optimization. In: International Conference for Learning Representations. p 1–15.

69. Kim YJ, Lee SH, Kim TY, et al. 2013. Body fat assessment method using ct images with separation mask algorithm. J. Digit. Imaging 26(2):155–162.

70. Kvist H, Chowdhury B, Grangard U, et al. 1988. Total and visceral adipose-tissue volumes derived from measurements with computed tomography in adult men and women: Predictive equations. Am. J. Clin. Nutr. 48(6):1351–1361.

71. Berg HE, Truong D, Skoglund E, et al. 2020. Threshold-automated CT measurements of muscle size and radiological attenuation in multiple lower-extremity muscles of older individuals. Clin. Physiol. Funct. Imaging 40(3):165–172.

72. Lenchik L, Heacock L, Weaver AA, et al. 2019. Automated Segmentation of Tissues Using CT and MRI: A Systematic Review. Acad. Radiol. 26(12):1695–1706.

73. Palit A, King R, Gu Y, et al. 2019. Prediction and Visualisation of Bony Impingement for Subject Specific Total Hip Arthroplasty. Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS :2127–2131.

74. Lerch TD, Degonda C, Schmaranzer F, et al. 2019. Patient-Specific 3-D Magnetic Resonance Imaging–Based Dynamic Simulation of Hip Impingement and Range of

Motion Can Replace 3-D Computed Tomography–Based Simulation for Patients With Femoroacetabular Impingement: Implications for Planning Open Hip Preserv. Am. J. Sports Med. 47(12):2966–2977.

75. Zhang YZ, Lu S, Zhang HQ, et al. 2016. Alignment of the lower extremity mechanical axis by computer-aided design and application in total knee arthroplasty. Int J CARS 11:1881–1890.

76. Fürmetz J, Sass J, Ferreira T, et al. 2019. Three-dimensional assessment of lower limb alignment: Accuracy and reliability. Knee 26(1):185–193.

77. Renault JB, Aüllo-Rasser G, Donnez M, et al. 2018. Articular-surface-based automatic anatomical coordinate systems for the knee bones. J. Biomech. 80:171–178.

78. Krishnan SP, Dawood A, Richards R, et al. 2012. A review of rapid prototyped surgical guides for patient-specific total knee replacement. J. Bone Jt. Surg. - Ser. B 94 B(11):1457–1461.

79. Minnema J, van Eijnatten M, Kouw W, et al. 2018. CT image segmentation of bone for medical additive manufacturing using a convolutional neural network. Comput. Biol. Med. 103:130–139.

80. van Eijnatten M, van Dijk R, Dobbe J, et al. 2018. CT image segmentation methods for bone used in medical additive manufacturing. Med. Eng. Phys. 51(October):6–16.

81. Audenaert EA, Van Houcke J, Almeida DF, et al. 2019. Cascaded statistical shape model based segmentation of the full lower limb in CT. Comput. Methods Biomech. Biomed. Engin. 22(6):644–657.

82. Asgari Taghanaki S, Abhishek K, Cohen JP, et al. 2020. Deep semantic segmentation of natural and medical images: a review. Springer Netherlands.

83. Presciutti SM, Karukanda T, Lee M. 2014. Management decisions for adolescent idiopathic scoliosis significantly affect patient radiation exposure. Spine J. 14(9):1984–1990.

84. Lindgren Belal S, Sadik M, Kaboteh R, et al. 2019. Deep learning for segmentation of 49 selected bones in CT scans: First step in automated PET/CT-based 3D quantification of skeletal metastases. Eur. J. Radiol. 113:89–95.

85. Zhou Z, Zhao G, Kijowski R, Liu F. 2018. Deep convolutional neural network for segmentation of knee joint anatomy. Magn Reson Med 80(6):2759–2770.

86. Çiçek Ö, Abdulkadir A, Lienkamp SS, et al. 2016. 3D U-net: Learning dense volumetric segmentation from sparse annotation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). Springer Verlag. p 424–432.

87. Noguchi S, Nishio M, Yakami M, et al. 2020. Bone segmentation on whole-body CT using convolutional neural network with novel data augmentation techniques.

Comput. Biol. Med. 121:103767.

88. Berger L, Eoin H, Cardoso MJ, Ourselin S. 2018. An adaptive sampling scheme to efficiently train fully convolutional networks for semantic segmentation. Commun. Comput. Inf. Sci. 894:277–286.

89. Kamnitsas K, Ledig C, Newcombe VFJ, et al. 2017. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Med. Image Anal. 36:61–78.

90. Roth HR, Shen C, Oda H, et al. 2018. A Multi-scale Pyramid of 3D Fully Convolutional Networks for Abdominal Multi-organ Segmentation. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). p 417–425.

91. Zhao H, Qi X, Shen X, et al. 2018. ICNet for Real-Time Semantic Segmentation on High-Resolution Images. Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics) 11207 LNCS:418–434.

92. Chen L-C, Papandreou G, Schroff F, Adam H. 2017. Rethinking Atrous Convolution for Semantic Image Segmentation.

93. Chen LC, Papandreou G, Kokkinos I, et al. 2018. DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. IEEE Trans. Pattern Anal. Mach. Intell. 40(4):834–848.

94. Isensee F, Jaeger PF, Kohl SAA, et al. 2021. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. Nat. Methods 18(2):203–211.

95. Jud L, Roth T, Roth T, et al. 2020. The impact of limb loading and the measurement modality (2D versus 3D) on the measurement of the limb loading dependent lower extremity parameters. BMC Musculoskelet. Disord. 21(1).

96. Dandachli W, Kannan V, Richards R, et al. 2008. Analysis of cover of the femoral head in normal and dysplastic hips: New CT-based technique. J. Bone Jt. Surg. - Ser. B 90(11):1428–1434.

97. Köhnlein W, Ganz R, Impellizzeri FM, Leunig M. 2009. Acetabular Morphology: Implications for Joint-preserving Surgery. Clin. Orthop. Relat. Res. 467(3):682–691.

98. Tannast M, Albers CE, Steppacher SD, Siebenrock KA. 2011. Hip Pain in the Young Adult. Eur. Instr. Lect. .

99. Toogood PA, Skalak A, Cooperman DR. 2009. Proximal femoral anatomy in the normal human population. Clin. Orthop. Relat. Res. 467(4):876–885.

100. Zeng G, Schmaranzer F, Degonda C, et al. 2021. MRI-based 3D models of the hip joint enables radiation-free computer-assisted planning of periacetabular osteotomy for treatment of hip dysplasia using deep learning for automatic segmentation. Eur. J. Radiol. Open 8(December 2020):100303.

101. Edgar H, Daneshvari Berry S, Moes E, et al. 2020. New Mexico Decedent Image

Database. Off. Med. Investig. Univ. New Mex. .

102.   Wu D, Sofka M, Birkbeck N, Zhou SK. 2014. Segmentation of multiple knee bones from CT for orthopedic knee surgery planning. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). p 372–380.

103.   Chu C, Chen C, Liu L, Zheng G. 2015. FACTS: Fully Automatic CT Segmentation of a Hip Joint. Ann. Biomed. Eng. 43(5):1247–1259.

104.   Kim JJ, Nam J, Jang IG. 2018. Fully automated segmentation of a hip joint using the patient-specific optimal thresholding and watershed algorithm. Comput. Methods Programs Biomed. 154:161–171.

105.   Almeida DF, Ruben RB, Folgado J, et al. 2016. Fully automatic segmentation of femurs with medullary canal definition in high and in low resolution CT scans. Med. Eng. Phys. 38(12):1474–1480.

106.   Chu C, Bai J, Wu X, Zheng G. 2015. MASCG: Multi-Atlas Segmentation Constrained Graph method for accurate segmentation of hip CT images. Med. Image Anal. 26(1):173–184.

107.   Chang Y, Yuan Y, Guo C, et al. 2019. Accurate pelvis and femur segmentation in hip CT with a novel patch-based refinement. IEEE J. Biomed. Heal. Informatics 23(3):1192–1204.

108.   Seim H, Kainmueller D, Heller M, et al. 2008. Automatic segmentation of the pelvic bones from CT data based on a statistical shape model. In: EG VCBM 2008 - Eurographics Workshop on Visual Computing for Biomedicine. p 93–100.

109.   Liu P, Han H, Du Y, et al. 2021. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. Int. J. Comput. Assist. Radiol. Surg. 16(5):749–756.

110.   Fu Y, Liu S, Li HH, Yang D. 2017. Automatic and hierarchical segmentation of the human skeleton in CT images. Phys. Med. Biol. 62(7):2812–2833.

111.   Mast NH, Impellizzeri F, Keller S, Leunig M. 2011. Reliability and agreement of measures used in radiographic evaluation of the adult hip. In: Clinical Orthopaedics and Related Research. p 188–199.

112.   Shrout PE, Fleiss JL. 1979. Intraclass correlations: Uses in assessing rater reliability. Psychol. Bull. 86(2):420–428.

113.   Hingsammer AM, Bixby S, Zurakowski D, et al. 2015. How Do Acetabular Version and Femoral Head Coverage Change With Skeletal Maturity? Clin. Orthop. Relat. Res. 473(4):1224–1233.

114.   Werner CML, Ramseier LE, Ruckstuhl T, et al. 2012. Normal values of Wiberg's lateral center-edge angle and Lequesne's acetabular index-a coxometric update. Skeletal Radiol. 41(10):1273–1278.

115. Koo TK, Li MY. 2016. A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. J. Chiropr. Med. 15(2):155–163.

116. Harris-Hayes M, Commean PK, Patterson JD, et al. 2014. Bony abnormalities of the hip joint: a new comprehensive, reliable and radiation-free measurement method using magnetic resonance imaging. J. Hip Preserv. Surg. 1(2):62–70.

117. Avants BB, Tustison NJ, Wu J, et al. 2011. ANTS: Advanced Open-Source Tools for Normalization And Neuroanatomy. Neuroinformatics 9(4):381–400.

118. Wagensveld IM, Blokker BM, Wielopolski PA, et al. 2017. Total-body CT and MR features of postmortem change in in-hospital deaths.

119. Paszke A, Gross S, Massa F, et al. 2019. PyTorch: An imperative style, high-performance deep learning library. Adv. Neural Inf. Process. Syst. 32(NeurIPS).

120. Richard MJ, Ruch DS, Aldridge JM. 2007. Malunions and Nonunions of the Forearm. Hand Clin. 23(2):235–243.

121. Fuller DJ, McCullough CJ. 1982. Malunited fractures of the forearm in children. J. Bone Jt. Surg. - Ser. B 64(3):364–367.

122. Mulders MAM, D'Ailly PN, Cleffken BI, Schep NWL. 2017. Corrective osteotomy is an effective method of treating distal radius malunions with good long-term functional results. Injury 48(3):731–737.

123. Wada T, Tatebe M, Ozasa Y, et al. 2011. Clinical outcomes of corrective osteotomy for distal radial malunion: A review of opening and closing-wedge techniques. J. Bone Jt. Surg. - Ser. A 93(17):1619–1626.

124. Saravi B, Lang G, Steger R, et al. 2021. Corrective Osteotomy of Upper Extremity Malunions Using Three-Dimensional Planning and Patient-Specific Surgical Guides: Recent Advances and Perspectives. Front. Surg. 8(February):1–7.

125. Bauer DE, Zimmermann S, Aichmair A, et al. 2017. Conventional Versus Computer-Assisted Corrective Osteotomy of the Forearm: a Retrospective Analysis of 56 Consecutive Cases. J. Hand Surg. Am. 42(6):447–455.

126. de Muinck Keizer RJO, Lechner KM, Mulders MAM, et al. 2017. Three-dimensional virtual planning of corrective osteotomies of distal radius malunions: a systematic review and meta-analysis. Strateg. Trauma Limb Reconstr. 12(2):77–89.

127. Roth KC, Walenkamp MMJJ, van Geenen RCII, et al. 2017. Factors determining outcome of corrective osteotomy for malunited paediatric forearm fractures: a systematic review and meta-analysis. J. Hand Surg. Eur. Vol. 42(8):810–816.

128. Roth KC, van Es EM, Kraan GA, et al. 2021. Outcomes of 3-D corrective osteotomies for paediatric malunited both-bone forearm fractures. J. Hand Surg. Eur. Vol. 47(2):164–171.

129. Caiti G, Dobbe JGG, Strackee SD, et al. 2019. Computer-Assisted Techniques in Corrective Distal Radius Osteotomy Procedures. IEEE Rev. Biomed. Eng. 13:233–

247.

130.   Marin F, Devos P, Hansen C, et al. 2019. Subject specific hand and forearm musculoskeletal 3D geometries using high-resolution MR images. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 7(1):88–95.

131.   Dobbe JGG, Strackee SD, Streekstra GJ. 2017. Minimizing the translation error in the application of an oblique single-cut rotation osteotomy: Where to cut? IEEE Trans. Biomed. Eng. 9294.

132.   Dobbe JGG, Kloen P, Strackee SD, Streekstra GJ. 2021. Comparison of an oblique single cut rotation osteotomy with a novel 3D computer-assisted oblique double cut alignment approach. Sci. Rep. 11(1):1–14.

133.   Nagy L, Jankauskas L, Dumont CE. 2008. Correction of forearm malunion guided by the preoperative complaint. Clin. Orthop. Relat. Res. 466(6):1419–1428.

134.   Fedorov A, Beichel R, Kalphaty-Cramer J, et al. 2012. 3D Slicer as an Image Computing Platform for the Quantitative Imaging Network. Magn. Reson. Imaging 30(9):1323–1341.

135.   Tustison NJ, Avants BB, Cook PA, et al. 2010. N4ITK: Improved N3 bias correction. IEEE Trans. Med. Imaging 29(6):1310–1320.

136.   Otsu N. 1979. A Threshold Selection Method from Gray-Level Histograms. IEEE Trans. Syst. Man. Cybern. 9(1):62–66.

137.   Besl PJ, McKay ND. 1992. A Method for the registration of 3-D Shapes. IEEE Trans. Pattern Anal. Mach. Intell. 14(2):239–256.

138.   Huynh DQ. 2009. Metrics for 3D Rotations: Comparison and Analysis. J Math Imaging Vis 35:155–164.

139.   Vlachopoulos L, Schweizer A, Graf M, et al. 2015. Three-dimensional postoperative accuracy of extra-articular forearm osteotomies using CT-scan based patient-specific surgical guides Orthopedics and biomechanics. BMC Musculoskelet. Disord. 16(1).

140.   Brui E, Efimtcev AY, Fokin VA, et al. 2020. Deep learning-based fully automatic segmentation of wrist cartilage in MR images. NMR Biomed. 33(8).

141.   Dalili D, Fritz J, Isaac A. 2021. 3D MRI of the Hand and Wrist: Technical Considerations and Clinical Applications. Semin. Musculoskelet. Radiol. 25(3):501–513.

142.   Rolian C. 2020. Endochondral ossification and the evolution of limb proportions. Wiley Interdiscip. Rev. Dev. Biol. 9(4).

143.   Dobbe JGG, du Pré KJ, Blankevoort L, et al. 2017. Computer-assisted oblique single-cut rotation osteotomy to reduce a multidirectional tibia deformity: case report. Strateg. Trauma Limb Reconstr. 12(2):115–120.

144.   Byrne AM, Impelmans B, Bertrand V, et al. 2017. Corrective Osteotomy for

Malunited Diaphyseal Forearm Fractures Using Preoperative 3-Dimensional Planning and Patient-Specific Surgical Guides and Implants. J. Hand Surg. Am. 42(10):836.e1-836.e12.

145. Oka K, Tanaka H, Okada K, et al. 2019. Three-Dimensional Corrective Osteotomy for Malunited Fractures of the Upper Extremity Using Patient-Matched Instruments: A Prospective, Multicenter, Open-Label, Single-Arm Trial. J. Bone Jt. Surg. - Am. Vol. 101(8):710–721.

146. Ahrend MD, Baumgartner H, Ihle C, et al. 2021. Influence of axial limb rotation on radiographic lower limb alignment: a systematic review. Arch. Orthop. Trauma Surg. (0123456789).

147. Degen N, Sass J, Jalali J, et al. 2020. Three-dimensional assessment of lower limb alignment: Reference values and sex-related differences. Knee 27(2):428–435.

148. Chaouche S, Jacquet C, Fabre-Aubrespy M, et al. 2019. Patient-specific cutting guides for open-wedge high tibial osteotomy: safety and accuracy analysis of a hundred patients continuous cohort. Int. Orthop. 43(12):2757–2765.

149. Grasso F, Martz P, Micicoi G, et al. 2022. Double level knee osteotomy using patient-specific cutting guides is accurate and provides satisfactory clinical results: a prospective analysis of a cohort of twenty-two continuous patients. Int. Orthop. 46(3):473–479.

150. Arnal-Burró J, Pérez-Mañanes R, Gallo-del-Valle E, et al. 2017. Three dimensional-printed patient-specific cutting guides for femoral varization osteotomy: Do it yourself. Knee 24(6):1359–1368.

151. Jacquet C, Sharma A, Fabre M, et al. 2020. Patient-specific high-tibial osteotomy's 'cutting-guides' decrease operating time and the number of fluoroscopic images taken after a Brief Learning Curve. Knee Surgery, Sport. Traumatol. Arthrosc. 28(9):2854–2862.

152. Kaiser P, Attal R, Kammerer M, et al. 2016. Significant differences in femoral torsion values depending on the CT measurement technique. Arch. Orthop. Trauma Surg. 136(9):1259–1264.

153. Roth T, Carrillo F, Wieczorek M, et al. 2021. Three-dimensional preoperative planning in the weight-bearing state: validation and clinical evaluation. Insights Imaging 12(1).

154. Subburaj K, Ravi B, Agarwal M. 2010. Computer-aided methods for assessing lower limb deformities in orthopaedic surgery planning. Comput. Med. Imaging Graph. 34(4):277–288.

155. Subburaj K, Ravi B, Agarwal M. 2009. Automated identification of anatomical landmarks on 3D bone models reconstructed from CT scan images. Comput. Med. Imaging Graph. 33(5):359–368.

156. Miranda DL, Rainbow MJ, Leventhal EL, et al. 2010. Automatic determination of

anatomical coordinate systems for three-dimensional bone models of the isolated human knee. J. Biomech. 43(8):1623–1626.

157. Xing Q, Theiss MM, Peng Q, et al. 2010. 3D automatic feature construction system for lower limb alignment. Proc. - 2010 Int. Conf. Cyberworlds, CW 2010 :375–382.

158. Phan CB, Koo S. 2015. Predicting anatomical landmarks and bone morphology of the femur using local region matching. Int. J. Comput. Assist. Radiol. Surg. 10(11):1711–1719.

159. Fischer MCM, Grothues SAGA, Habor J, et al. 2020. A robust method for automatic identification of femoral landmarks, axes, planes and bone coordinate systems using surface models. Sci. Rep. 10(1):1–11.

160. Porto A, Rolfe S, Maga AM. 2021. ALPACA: A fast and accurate computer vision approach for automated landmarking of three-dimensional biological structures. Methods Ecol. Evol. 12(11):2129–2144.

161. Baek SY, Wang JH, Song I, et al. 2013. Automated bone landmarks prediction on the femur using anatomical deformation technique. CAD Comput. Aided Des. 45(2):505–510.

162. Kuiper RJA, Sakkers RJB, van Stralen M, et al. 2022. Efficient cascaded V-net optimization for lower extremity CT segmentation validated using bone morphology assessment. J. Orthop. Res. (September 2021):1–14.

163. Audenaert EA. 2021. nonrigidICP. MATLAB Cent. File Exch. [cited 2022 May 4] Available from: https://www.mathworks.com/matlabcentral/fileexchange/41396-nonrigidicp.

164. Jamali AA, Meehan JP, Moroski NM, et al. 2017. Do small changes in rotation affect measurements of lower extremity limb alignment? J. Orthop. Surg. Res. 12(1):1–8.

165. Siboni R, Vialla T, Joseph E, et al. 2022. Coronal and sagittal alignment of the lower limb in Caucasians: Analysis of a 3D CT database. Orthop. Traumatol. Surg. Res. 108(3).

166. Pangaud C, Laumonerie P, Dagneaux L, et al. 2020. Measurement of the Posterior Tibial Slope Depends on Ethnicity, Sex, and Lower Limb Alignment: A Computed Tomography Analysis of 378 Healthy Participants. Orthop. J. Sport. Med. 8(1).

167. Nakano N, Matsumoto T, Hashimura M, et al. 2016. Coronal lower limb alignment in normal knees-A radiographic analysis of 797 normal knee subjects. Knee 23(2):209–213.

168. Carrino JA, Muhit A Al, Zbijewski W, et al. 2014. Dedicated cone-beam CT system for extremity imaging. Radiology 270(3):816–824.

169. Dubousset J, Charpak G, Dorion I, et al. 2005. A new 2D and 3D imaging approach to musculo-skeletal physiology and pathology with low-dose radiation and the standing position: The EOS system. Bull. Acad. Natl. Med. 189(2):287–300.

170. Kobayashi K, Sakamoto M, Tanabe Y, et al. 2009. Automated image registration for assessing three-dimensional alignment of entire lower extremity and implant position using bi-plane radiography. J. Biomech. 42(16):2818–2822.

171. Nickoloff EL, Alderson PO. 2001. Radiation exposures to patients from CT: Reality, public perception, and policy. Am. J. Roentgenol. 177(2):285–287.

172. Sawyer JR. 2021. Radiation Reduction Strategies in Pediatric Orthopaedics. J. Pediatr. Orthop. 41:S75–S79.

173. Lena B, Florkow MC, Ferrer CJ, et al. 2022. Synthetic CT for the planning of MR-HIFU treatment of bone metastases in pelvic and femoral bones: a feasibility study. Eur. Radiol. 32(7):4537–4546.

174. Rathnayaka K, Cowin G, Schuetz MA, et al. 2013. Correction of step artefact associated with MRI scanning of long bones. Med. Eng. Phys. 35(7):988–993.

175. Roner S, Vlachopoulos L, Nagy L, et al. 2017. Accuracy and Early Clinical Outcome of 3-Dimensional Planned and Guided Single-Cut Osteotomies of Malunited Forearm Bones. J. Hand Surg. Am. 42(12):1031.e1-1031.e8.

176. Radzi S, Uesugi M, Baird A, et al. 2014. Assessing the bilateral geometrical differences of the tibia - Are they same? Med. Eng. Phys. 36(12):1618–1625.

177. Hong E, Kwak DS, Kim IB. 2021. Morphological symmetry of the radius and ulna— Can contralateral forearm bones utilize as a reliable template for the opposite side? PLoS One 16(10 October).

178. Tümer N, Arbabi V, Gielis WP, et al. 2019. Three-dimensional analysis of shape variations and symmetry of the fibula, tibia, calcaneus and talus. J. Anat. 234(1):132–144.

179. Eckhoff DG, Jacofsky DJ, Springer BD, et al. 2016. Bilateral Symmetrical Comparison of Femoral and Tibial Anatomic Features. J. Arthroplasty 31(5):1083–1090.

180. Tuominen EKJ, Kankare J, Koskinen SK, Mattila KT. 2013. Weight-bearing CT imaging of the lower extremity. Am. J. Roentgenol. 200(1):146–148.

181. Hirschmann A, Buck FM, Fucentese SF, Pfirrmann CWA. 2015. Upright CT of the knee: the effect of weight-bearing on joint alignment. Eur. Radiol. 25(11):3398–3404.

182. Melhem E, Assi A, El Rachkidi R, Ghanem I. 2016. EOS{®} biplanar X-ray imaging: concept, developments, benefits, and limitations. J. Child. Orthop. 10(1):1–14.

183. Tomaževič D, Likar B, Slivnik T, Pernuš F. 2003. 3-D/2-D registration of CT and MR to X-ray images. IEEE Trans. Med. Imaging 22(11):1407–1416.

184. Väänänen SP, Isaksson H, Waarsing JH, et al. 2012. Estimation of 3D rotation of femur in 2D hip radiographs. J. Biomech. 45(13):2279–2283.

185. Karkenny AJ, Mendelis JR, Geller DS, Gomez JA. 2019. The Role of Intraoperative Navigation in Orthopaedic Surgery. J. Am. Acad. Orthop. Surg. 27(19):E849–E858.

186. Jolesz FA. 2005. Future perspectives for intraoperative MRI. Neurosurg. Clin. N. Am. 16(1):201–213.

187. Jolesz FA. 2014. Intraoperative Imaging and Image-Guided Therapy.

188. Ewurum CH, Guo Y, Pagnha S, et al. 2018. Surgical navigation in orthopedics: Workflow and system review. Adv. Exp. Med. Biol. 1093:47–63.

189. Jörgens M, Keppler AM, Ahrens P, et al. 2022. 3D osteotomies—improved accuracy with patient-specific instruments (PSI). Eur. J. Trauma Emerg. Surg. :1–8.

190. Benayoun M, Langlais T, Laurent R, et al. 2021. 3D planning and patient-specific surgical guides in forearm osteotomy in children: Radiographic accuracy and clinical morbidity. Orthop. Traumatol. Surg. Res. :102925.

191. Anderl W, Pauzenberger L, Kölblinger R, et al. 2016. Patient-specific instrumentation improved mechanical alignment, while early clinical outcome was comparable to conventional instrumentation in TKA. Knee Surgery, Sport. Traumatol. Arthrosc. 24(1):102–111.

192. Munier M, Donnez M, Ollivier M, et al. 2017. Can three-dimensional patient-specific cutting guides be used to achieve optimal correction for high tibial osteotomy? Pilot study. Orthop. Traumatol. Surg. Res. 103(2):245–250.

193. Huijbregts HJTAM, Khan RJK, Sorensen E, et al. 2016. Patient-specific instrumentation does not improve radiographic alignment or clinical outcomes after total knee arthroplasty: A meta-analysis. Acta Orthop. 87(4):386–394.

194. Kizaki K, Shanmugaraj A, Yamashita F, et al. 2019. Total knee arthroplasty using patient-specific instrumentation for osteoarthritis of the knee: A meta-analysis. BMC Musculoskelet. Disord. 20(1):1–18.

195. Cavaignac E, Pailhé R, Laumond G, et al. 2015. Evaluation of the accuracy of patient-specific cutting blocks for total knee arthroplasty: a meta-analysis. Int. Orthop. 39(8):1541–1552.

196. Goh GS, Lohre R, Parvizi J, Goel DP. 2021. Virtual and augmented reality for surgical training and simulation in knee arthroplasty. Arch. Orthop. Trauma Surg. 141(12):2303–2312.

197. Laverdière C, Corban J, Khoury J, et al. 2019. Augmented reality in orthopaedics: A systematic review and a window on future possibilities. Bone Jt. J. 101-B(12):1479–1488.

198. Desai AS, Dramis A, Kendoff D, Board TN. 2011. Critical review of the current practice for computer-assisted navigation in total knee replacement surgery: cost-effectiveness and clinical outcome. Curr. Rev. Musculoskelet. Med. 2011 41 4(1):11–15.

199. Kumar V, Baburaj V, Patel S, et al. 2021. Does the use of intraoperative CT scan improve outcomes in Orthopaedic surgery? A systematic review and meta-analysis of 871 cases. J. Clin. Orthop. Trauma 18:216–223.

200. Wong KC. 2016. 3D-printed patient-specific applications in orthopedics. Orthop. Res. Rev. 8:57.

201. Caiti G, Dobbe JGG, Loenen ACY, et al. 2018. Implementation of a semiautomatic method to design patient-specific instruments for corrective osteotomy of the radius. Int. J. Comput. Assist. Radiol. Surg. .

202. Dobbe JGG, Vroemen JC, Strackee SD, Streekstra GJ. 2013. Corrective distal radius osteotomy: Including bilateral differences in 3-D planning. Med. Biol. Eng. Comput. 51(7):791–797.

203. FDA. 2018. Clinical Decision Support. Clin. Informatics Board Rev. Self Assess. :41–69 Available from: https://www.fda.gov/regulatory-information/search-fda-guidance-documents/clinical-decision-support-software.

204. European Parliament, Council of the European Union. 2017. Regulation (EU) 2017/745 of the European Parliament and of the Council of 5 April 2017 on medical devices. Off. J. Eur. Union 60(L117):2–175 [cited 2022 Jul 29] Available from: http://data.europa.eu/eli/reg/2017/745/2020-04-24.

Acknowledgements

During this long journey that has been my thesis, there have been many people that have accompanied and helped me on my literal and figurative travels through the (academic) world.

My primary guides in the academic world were of course my supervisors and (co)promotors at the UMC Utrecht.

Prof. dr. ir. Max Viergever, you have been largely responsible for providing me with the excellent resources, work environment and amazing colleagues that form the Image Science Institute. This has made working at the UMC an absolute pleasure. You have my gratitude for still being available for guidance during the end of my PhD period, while you were already emeritus professor for quite some time.

Prof. dr. ir. Harrie Weinans, very early in my academic career, still an early Master student, you allowed me to work on scoliosis bracing in collaboration with the UMC Utrecht. Your enthusiasm on this subject, and on any scientific subject since, convinced me to continue down this path, all the way from Master thesis to this dissertation. To me, you have always been the most outstanding example of a true scientist. For reasons quite obscure to me, you also seemed to believe in my ability to bring this journey to a satisfying ending, and as always, it seems you were right in the end.

Dr. ir. Peter Seevinck, you introduced me to the ISI and the world of medical imaging, a terrain and discipline I had very little experience in. Thankfully, you have always been patient with me and keen to discuss our ideas. Your work at the UMC Utrecht and MRIGuidance have been a great example and inspiration of how to combine both academic and industry work. Most impressively, you were also still able to combine this with having fun both inside and outside of the workplace, which has obviously been an inspiration to me as well.

Dr. Ralph Sakkers, the many hours we spend discussing osteotomies, trying to get what was in your brain into my computer, were both the most interesting and most amusing hours I spend as a PhD. I am very grateful for all the time and attention you put into our work, and for giving me the opportunity to try and put the academic work, that so often remains only theoretical work on paper, into practice. Although there is always much more that can be done, I think our joined efforts have been a great start to a promising orthopaedic solution. And at the very least, it has been a lot of fun.

I would also like to thank the members of the committee who have dedicated their effort and valuable time to the evaluation of this thesis and defense.

All members of the MIRROR team have had an undeniable impact on my work. By supplying me with important insights, information and tips for my research, but also by providing me a place in a team of like-minded spirits. It was great to see this group slowly grow during my years at the UMC. From a small group of about five people, to a room packed full of people who would discuss, criticize, and sometimes praise each other's work. I am thankful for everyone who has been part of the team at some point during this time, but the following people deserve some special praise:

Dr. Mateusz, it still feels strange to call you a doctor, as you have been my roommate for the longest duration of my PhD. You were a shining example as a PhD student, but also as a colleague and friend. You were always available when I needed help with the more complicated facets of image acquisition and deep learning, and I still remember your lecture on reinforcement learning. I like to imagine we may have helped each other out, even if I may have functioned primarily as a 'rubber duck' for most of your problems.

Dr. Frank, as my first roommate, you introduced me to ISI, and to programming for medical imaging. I remember being in awe at the speed at which you were able to write a piece of Matlab software when I first started. Later on, you apparently once remarked to our supervisors that I would 'hack a solution together in a few hours' when questioned about a certain piece of programming. I still cherish that as one of the greatest compliments I have ever received.

Dr. Marijn van Stralen, your sharp wit and in-depth technical knowledge have been a great aid in forming the chapters that form this thesis. Your questions and advice during the MIRROR meetings were some of the most helpful pointers I have received. I was deeply impressed with your ability to think both in terms of the small scale, technical details and the larger, over-arching goal of a research question.

Dr. Koen, thank you for your clinical insights during the collaborations and the MIRROR meetings we had. You have produced a wonderful addition to the UMC Utrecht in the form of the 3D Lab and I am absolutely certain you have a wonderful future ahead of you, whether it will be scientific, clinical, industrial, or all of the above.

Thank you, Rick, for being the first person to actually try and apply my academic work to something practically useful, and for just being a very fun person all-round. Your presence at the UMC Utrecht had a definite joyful impact on the team.

Thank you Vahid and Saeed, it was great to work alongside you, as our similar research goals made us ideal partners for the exchange of data, information and tools. Saeed, I am especially thankful for your ever-present smile and kind words.

I would also like to thank all past and present members of ISI, who have made this place such a pleasure to work at. Both colleagues and friends, you have been an inspiration for, and a welcome distraction from, work as a PhD. Thank you, to the original summer school group

of 2017, Julia, Sanne, Mateusz, Beatrice, Steffen you have been the foundation on which most of my friendships during my PhD have grown.

Special thanks to Julia and Sanne, for you were already involved at the UMC Utrecht and you showed us newbies around. You created an amazing atmosphere in our little hallway E, the birthplace of many a young ISI PhD. Julia, you have always been able to make me laugh and I was thrilled to find a person in the academic world that shared my terrible sense of humor. Sanne, you were so kind and welcoming that you made E feel like home, and I hope you have an amazing future with Sibren and the kids. This team would of course not be complete without Matthijs, an important part of the original foundation of the ISI team, a great rock climber, and one of the nicest persons I have ever met.

And thank you Nadieh, Marielle, Max, Erik, Ishaan, Mark, Majd, Nils, Sam, Hui Shan, Hui, Mike and Wilbert, for being amazing friends and colleagues. Thank you, Bas and Hugo, for serving as the original role models for us younger PhD students. Special thanks to Maria Schreuder-Ionescu and Gerard van Hoorn, who helped me out so many times, and without whom the ISI would surely come crashing down.

Naast alle vrienden en collega's waaraan ik professioneel zoveel te danken heb, wil ik natuurlijk ook iedereen bedanken die mij op persoonlijk vlak gedurende mijn hele leven opgevoed, gesteund en vermaakt hebben.

Mijn huisgenoten zullen zeker onder diegenen zijn die het meest te verduren hebben gehad tijdens mijn PhD. Stef, Steel, Jonas, Britt, jullie hebben mij mentaal en vast ook wel eens fysiek op de been gehouden en daarvoor ben ik jullie ontzettend dankbaar. Dankzij alle fantastische feestelijkheden, tripjes, en vele uren in de zon op het terras was mijn PhD-tijd zeker niet onaangenaam.

Steel, Wolf, Stoop, Vincent, Sergei, Erik, Pater, Gijs, Toine, Stef, Jonas, jullie hilarische gezelschap tijdens en na mijn studie heeft zeker een onuitwisbare indruk achtergelaten. Wat de uitwerking daarvan precies is geweest blijft onduidelijk, maar ik ben dankbaar dat ik deze ontwikkeling van scholier tot werkende man samen met jullie heb mogen beleven. Ik heb daar dan ook enorm van genoten.

Stijn, Marijn, Britt, Laura, Niels, Jonas, het is volgens mij ontzettend bijzonder dat ik jullie na zoveel jaar nog steeds als vrienden heb, en dat ik er eigenlijk vanuit durf te gaan dat dit nog lang zo zal blijven. Jullie zijn de enige vrienden die (bijna) mijn hele ontwikkeling hebben meegemaakt, en mij daarmee ook hebben gevormd tot wie ik nu ben. Dit is dus ook een beetje jullie schuld.

Ook wil ik al mijn vrienden uit Utrecht bedanken die mij altijd met enthousiasme hebben aangemoedigd: De Douwes, Eline, Koen, Isolde, Jasper, Myrthe, Wijnand, Stefanie en Ferdi. De zondagmiddagborrels gaven mij de kracht om de week telkens weer goed te beginnen.

Ruurd Sr. en Annet, of liever, pap en mam, bedankt voor jullie oneindige steun. Zoals nu wel duidelijk moge zijn, zijn jullie altijd mijn voorbeelden geweest, soms zelfs zonder dat ik dit bewust doorhad. Zonder jullie steun was dit nooit mogelijk geweest, en ik kan geen enkele manier bedenken waarop jullie me nog beter hadden kunnen helpen. Ik hoop dat jullie heel trots zijn. Niet op mij, maar op jezelf. Jullie hebben het geweldig goed gedaan!

Lexi en Arthur, zussie en broeder, jullie ook bedankt voor mijn fantastische opvoeding en jeugd. Vanaf begins af aan zijn wij volgens mij een geweldig team. Ik zou graag nog zoveel meer met jullie samen ondernemen, en ik denk dat daar in de toekomst ook weer alle ruimte voor is. Ik ben er enorm trots op dat jullie mijn broer en zus zijn, en ook onwijs blij met de recente gezinsuitbreiding.

Lieve Valentine, dankjewel voor al je liefde en geduld in deze hele onderneming. We zijn al meer dan zeven jaar samen, wat betekent dat je het hele traject heb meegekregen. Dat je me daarbij altijd bent blijven aanmoedigen, steunen en zelfs liefhebben vind ik ongelofelijk, en ook het wetenschappelijk bewijs dat liefde blind maakt. Je hebt mij altijd opgevrolijkt als het even tegenzat, en we hebben samen zelfs de kleinste successen uitbundig gevierd. Als een ware muze wist je me te inspireren wanneer ik het even niet meer wist. Ik kan niet wachten op alle avonturen die we nog samen gaan beleven.

Publications

Publications in international journals

Kuiper RJA, Colaris JW, Stockmans F, van Es EM, Seevinck PR, Viergever MA, Weinans H, Sakkers RJB. Impact of bone and cartilage segmentation from CT and MRI on both bone forearm osteotomy planning. Submitted.

Kuiper RJA, Seevinck PR, Viergever MA, Weinans H, Sakkers RJB, Automatic Assessment of Lower Limb Alignment from CT. Accepted for publication in the Journal of Bone and Joint Surgery.

Kuiper RJA, Sakkers RJB, van Stralen M, Arbabi V, Viergever MA, Weinans H, Seevinck PR. 2022. Efficient cascaded V-net optimization for lower extremity CT segmentation validated using bone morphology assessment. Journal of Orthopaedic Research (September 2021):1–14. doi:10.1002/jor.25314.

Kuiper RJA, van Stralen M, Sakkers RJB, Bergmans R, Zijlstra F, Viergever MA, Weinans H, Seevinck PR. 2021. CT to MR registration of complex deformations in the knee joint through dual quaternion interpolation of rigid transforms. Physics in Medicine & Biology, 66(17), 175024.

Willemsen K, Ketel MHM, Zijlstra F, Florkow MC, **Kuiper RJA**, van der Wal BCH, Weinans H, Pouran B, Beekman FJ, Seevinck PR, et al. 2021. 3D-printed saw guides for lower arm osteotomy, a comparison between a synthetic CT and CT-based workflow. 3D Printing in Medicine. 7(1):13. doi:10.1186/s41205-021-00103-x.

Conference abstracts

Kuiper RJA, Willemsen K, Weinans H, Viergever MA, Seevinck PR, Sakkers RJB. 2021 Semi-Automatic Osteotomy Planning with Optimal Joint Alignment. 39th Meeting of the European Pediatric Orthopaedic Society (EPOS 2021).

Kuiper RJA, van Stralen M, Zijlstra F, Weinans H, Roth KC, Colaris JW, Sakkers RJB, Seevinck PR. 2019. A Semi-Automatic Workflow for CT-MRI Registration of Complex Deformations Induced by Interscan Radio-Ulnar Rotations. 7th Dutch Bio-Medical Engineering Conference (BME 2019).

Curriculum Vitae

Ruurd Kuiper was born on August 21th, 1992 in Eersel, The Netherlands. After secondary school, he studied Electrical Engineering at the TU Delft for one year, but changed his mind and continued his studies in the field of biochemistry under the Life, Science and Technology bachelor at the TU Delft and Leiden University. After receiving his bachelor's, he switched to the Biomedical Engineering master at the TU Delft, which he finished in 2017. His final project concerned the development of a patient specific finite element model of the human spine, in collaboration with the orthopaedics department at the UMC Utrecht. Afterwards, he continued working at the UMC Utrecht as a PhD candidate in a combined position at the orthopaedics department and the Image Science Institute. He was able to extend his research for 10 months with the help of the NWO Take-Off grant for the development of pre-operative planning software. The results of this research are presented in this thesis. The development of the methods in this thesis also form the scientific basis of the healthcare start-up UPlanner BV.