Dissertation

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

Put forward by
Cornelius Jonathan Bauer
born in Saarbrücken
Oral examination: 3rd of May 2023

Applications of a Biomechanical Patient Model for Adaptive Radiation Therapy

Referees: Prof. Dr. Joao Seco
Prof. Dr. Oliver Jäkel

Applications of a Biomechanical Patient Model for Adaptive Radiation Therapy

Biomechanical patient modeling incorporates physical knowledge of the human anatomy into the image processing that is required for tracking anatomical deformations during adaptive radiation therapy, especially particle therapy. In contrast to standard image registration, this enforces bio-fidelic image transformation. In this thesis, the potential of a kinematic skeleton model and soft tissue motion propagation are investigated for crucial image analysis steps in adaptive radiation therapy.

The first application is the integration of the kinematic model in a deformable image registration process (KinematicDIR). For monomodal CT scan pairs, the median target registration error based on skeleton landmarks, is smaller than (1.6 ± 0.2) mm. In addition, the successful transferability of this concept to otherwise challenging multimodal registration between CT and CBCT as well as CT and MRI scan pairs is shown to result in median target registration error in the order of 2 mm. This meets the accuracy requirement for adaptive radiation therapy and is especially interesting for MR-guided approaches.

Another aspect, emerging in radiotherapy, is the utilization of deep-learning-based organ segmentation. As radiotherapy-specific labeled data is scarce, the training of such methods relies heavily on augmentation techniques. In this work, the generation of synthetically but realistically deformed scans used as Bionic Augmentation in the training phase improved the predicted segmentations by up to 15% in the Dice similarity coefficient, depending on the training strategy.

Finally, it is shown that the biomechanical model can be built-up from automatic segmentations without deterioration of the KinematicDIR application. This is essential for use in a clinical workflow.

Anwendungen eines biomechanischen Patientenmodells für die adaptive Strahlentherapie

Die biomechanische Patientenmodellierung bezieht physikalische Kenntnisse der menschlichen Anatomie in die Bildverarbeitung ein, die für die Verfolgung anatomischer Verformungen während der adaptiven Strahlentherapie, insbesondere der Partikeltherapie, erforderlich ist. Im Gegensatz zur standardmäßigen Bildregistrierung wird dadurch eine bio-fidele Bildtransformation erzwungen. In dieser Arbeit werden die Vorteile eines kinematischen Skelettmodells und der Bewegungspropagation im Weichteilgewebe für entscheidende Bildanalyseschritte in der adaptiven Strahlentherapie untersucht.
Die erste Anwendung ist die Integration des kinematischen Modells in ein elastisches Bildregistrierungsverfahren (KinematicDIR). Bei monomodalen CT-Scan-Paaren ist der Median des *Target Registration Error* auf der Basis von Skelett-Landmarken kleiner als $(1,6 \pm 0,2)$ mm. Darüber hinaus wird gezeigt, dass die erfolgreiche Übertragbarkeit dieses Konzepts auf die ansonsten schwierige multimodale Registrierung zwischen CT- und CBCT- sowie CT- und MRT-Scanpaaren zu einem mittleren *Target Registration Error* in der Größenordnung von 2 mm führt. Dies genügt der Genauigkeitsanforderung für die adaptive Strahlentherapie und ist besonders für MR-geführte Ansätze interessant.
Ein weiterer Aspekt, der sich in der Strahlentherapie abzeichnet, ist die Verwendung von Deep-Learning-basierter Organsegmentierung. Da radiotherapiespezifische gelabelte Daten rar sind, ist das Training solcher Methoden stark auf Augmentationstechniken angewiesen. In dieser Arbeit verbesserte die Erzeugung synthetischer, realistisch deformierter Scans durch Bionic Augmentation in der Trainingsphase die vorhergesagten Segmentierungen um bis zu 15% im Dice-Ähnlichkeitskoeffizienten, abhängig von der Trainingsstrategie.
Schließlich wird gezeigt, dass das biomechanische Modell aus automatischen Segmentierungen aufgebaut werden kann, ohne die Genauigkeit der KinematicDIR-Anwendung zu verschlechtern. Dies ist für den Einsatz in einem klinischen Arbeitsablauf unabdingbar.

Contents

1 Introduction

Radiation therapy is one of the most important strategies for cancer treatment together with surgery, immunotherapy, and chemotherapy [1]. The goal of radiation therapy is the delivery of a prescribed therapeutic dose of ionizing radiation to maximize the tumor control probability while simultaneously minimizing the dose to healthy tissue and particularly organs at risk (OAR). Modern radiation therapy can provide highly conformal dose distribution and consequently high dose gradients at the boundaries of the target volume. This can be achieved by intensity-modulated photon therapy [2, 3] or the treatment with particles [4]. Such high-dose gradients are sensitive to anatomical changes in the patient, which can lead to under-dosing of the target region or over-dosing of healthy tissue associated with an increased risk of adverse side effects. Particle therapy is specifically affected by motion as the range of the particles can be affected by tissue heterogeneities along the beam path. Therefore, computed tomography (CT) and magnetic resonance imaging (MRI) are used to monitor any motion and compensate for or adapt to these changes.

Deformable image registration is the general concept to assess non-rigid anatomical changes within two images by finding an appropriate transformation to map one image to the other. This is essential for any adaptive radiation therapy workflow. Common intensity-based DIR algorithms provide fast image registration. They consider only changes in the intensity distribution of the image and typically do not take into account tissue properties. Therefore, they are susceptible to misregistration in homogeneous areas or in the presence of image artifacts [5]. Furthermore, implausible deformation can be the result of purely intensity-based DIR schemes.

Biomechanical models explicitly incorporate physiological and biophysical properties of the human anatomy providing a natural regularization by restricting the possible motion within the image to what is realistic human motion. This increases the bio-fidelity of the registration process. The most commonly included biomechanical models in modern model-based DIR algorithms for radiation therapy are based on the finite element method (FEM) [6, 7] that provides a motion model with explicit physical parameters. FEM can be highly complex, requires explicit tissue parameters, and can lead to long computational times. Kinematic models like an articulated skeleton model use the rigidity and articulation of

human bones to constrain skeletal motion [8] but typically lack the regularization introduced by joints and deformation of the whole image space.
The head and neck region is of special interest for the application of a kinematic model, since motion in this region is governed by complex skeletal motion with a large number of bones and joints involved. Additionally, the proximity of tumors to OARs requires high accuracy in motion monitoring. This means a DIR algorithm for the head and neck region must be accurate and robust.
Once a biomechanical model is built-up and verified to represent the human anatomy correctly, it can also be employed for the generation of synthetic image data. This artificial data can be of high interest in the field of deep learning where the availability of properly annotated training data can be a crucial prerequisite. Biomechanical models facilitate the generation of only realistic image data and therefore can improve the overall coverage of the underlying distribution in the augmented training data.
In this work, a kinematics-based model of patient motion is investigated regarding its application for monomodal and multimodal image registration and the augmentation of deep learning techniques with a novel method to generate synthetic data.
For the image registration pipeline, the kinematic model is combined with a simplex optimizer and a similarity metric specific to the registered imaging modality. The accuracy and robustness of this approach are examined for the registration of planning CTs with fraction CTs, cone beam CTs and MRI. In this process, the idea of an object-based registration is developed to facilitate multimodal registration.
A bionic augmentation scheme is developed based on the kinematic model as the second application considered in this thesis. By forward calculation of an artificial posture and consequent transformation of an original image and segmentation labels, a synthetic realistic data set is created. It is hypothesized that these artificially generated images and labels can enhance the training of deep learning-based segmentation frameworks. An approach to automate this principle with so-called generalized postures is proposed.
Finally, the potential to use automatic segmentations to build-up the kinematic model is investigated. It can be shown that the fully automated model build-up is possible without compromising the accuracy or robustness of the image registration pipeline. This provides an essential step to the use of the kinematic model in a clinical setting.

2 Background

The background of this thesis can be split in five parts. First, the general idea of radiation therapy and the issue of patient motion is discussed. Here, the motivation for the presented thesis arises. Second, the fact that modern radiation therapy relies on different imaging techniques is considered, and an overview is given regarding the particular imaging modalities that are of relevance to the research in this work. Consequently, the connection between imaging and motion is illuminated and the concept of image registration is introduced. This leads to the fundamental idea to use biomechanical models as an accurate and robust approach to solving the inverse registration problem. Here, an overview of different models and concepts is given. Finally, the basic principles of deep learning for segmentation tasks and data augmentation are introduced, since one of the applications of this work is directly related to the generation of realistic augmentation data for the training of deep learning segmentation algorithms.

2.1 Radiation Therapy and Patient Motion

The concept of radiation therapy relies on the use of ionizing radiation to induce irreparable damage in tumor tissue [9]. The common goal of any radiation therapy is to deliver the medically prescribed energy dose

$$D = \frac{\mathrm{d}E}{\mathrm{d}m}\,, \tag{2.1}$$

with the energy E and the mass m to the treatment volume while sparing healthy tissue as much as possible in particular for certain organs at risk (OARs) where explicit dose limits can be clinically required.

2.1.1 Particle and Photon Therapy

There are different types of radiation used, depending on the position and type of the tumor among other factors. Photons are by far the most utilized technique

in radiation therapy. The photon depth-dose curve can be characterized by a build-up of dose followed by a decay [10]. Protons and heavy ions on the other hand show different depth-dose curves. Their energy deposition follows the Bethe equation for the mean energy loss [11]. A particle with velocity $\beta = \frac{v}{c}$ and energy E transversing the distance x in a homogeneous medium of electron density n and mean excitation energy I, will deposit an average energy

$$\left\langle \frac{\mathrm{d}E}{\mathrm{d}x} \right\rangle = -\frac{4\pi}{m_e c^2} \cdot \frac{nz^2}{\beta^2} \cdot \left(\frac{e^2}{4\pi\varepsilon_0} \right)^2 \cdot \left[\ln\left(\frac{2m_e c^2 \beta^2}{I \cdot (1-\beta^2)} \right) - \beta^2 \right], \tag{2.2}$$

where ε_0 is the vacuum permittivity, m_e is the electron mass, e the electron charge and c the speed of light. This average energy deposition is also called the *stopping power* of the target material. The properties of this energy loss lead to the phenomenon of the so-called *Bragg peak*. After a region of constant energy deposition, a pronounced peak appears right at the end of the particle track.
Figure 2.1 displays the depth-dose curve for typical clinical energies for the treatment with photons (blue), protons (orange) and carbon ions (green). It visualizes the ability of particle treatment to grant a spatially conformal dose in the treatment volume. For carbon ions, a fractionation effect leads to a dose tail after the Bragg peak as lighter particles from the fractionation event have increased range [15].

2.1.2 Treatment Planning and Radiation Therapy Process

In modern radiation therapy (regardless of the treatment modality), digital treatment planning is used to provide standardized and patient-specific irradiation of the tumor. For this purpose, the patient is immobilized in a reproducible posture and a so-called planning CT is acquired. This represents a snapshot of the human anatomy with the special property that the Hounsfield units in such a planning CT can be mapped to the electron density (or stopping power in case of particle therapy) of the underlying tissue using a calibration curve.
In the following treatment planning step, the target volumes and organs at risk (OAR) that are sensitive to radiation damage are delineated by a radiation oncologist who also prescribes a target dose to the tumor and surrounding margins as well as limits for certain OARs following clinical guidelines [16]. Consequently, inverse planning is used to find the optimal irradiation angles and intensities to

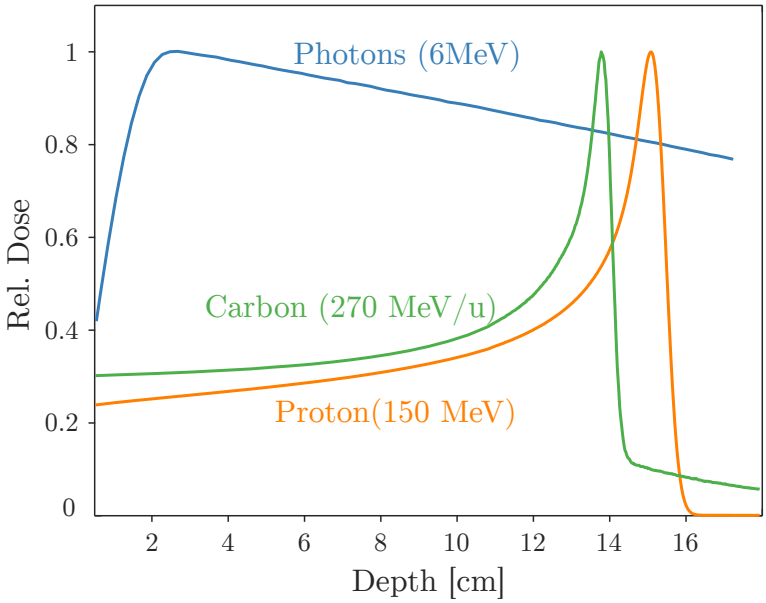

Figure 2.1: Depth dose profile of photons (blue), protons (orange) and carbon ions (green) in water. Photon dose displays the build-up of dose in the first 2 cm and then drops of slowly. Proton and carbon ions have low and constant dose deposition along most of the particle track. Shortly before they stop, most of the energy is deposited in the so called Bragg Peak. Carbon ions have a distinct dose tail after the Bragg peak caused by fragmentation. Data generated using TOPAS [12–14].

achieve the prescribed dose while sparing the OARs [17]. Typically, the dose that is prescribed to the patient is delivered in a fractionated way where the patient is irradiated only with a fraction of the dose each time. This can improve the sparing of normal tissue and OARs due to different radio-biological effects and is a standard technique in modern radiation therapy [18, 19].
This fractionated approach leads to high requirements in the precision of applied radiation. In particularly for highly conformal photon therapy or particle therapy including high gradients in the dose distribution a change in the posture can have a large effect.

2.1.3 Anatomical Changes in Radiation Therapy

Anatomical changes can occur in different ways and on different time scales during fractionated radiation therapy. While changes in the tissue due to shrinking, necrosis, or edema are not considered in this thesis, motion is another major contributor to changes in the human anatomy. Motion can occur between two

fractions (*inter-fractional* motion) when the patient's positioning is not reproduced or the patient does not lie in the original posture. Motion can also occur within a fraction (*intra-fractional* motion) caused by ,e.g., breathing, the heartbeat or abdominal organ motion. Figure 2.2 displays a planning CT and a fraction CT including the delineation of target regions (red) and OARs (different colors). In this example, inter-fractional motion has caused the spinal cord (yellow) to move from the original segmented area and is therefore no longer spared adequately in the radiation process. In addition, parts of the target volume are now outside the patient, which means the delivered dose to the tumor region is not in accordance with the original plan. To avoid this mistreatment, the occurred motion should be compensated.

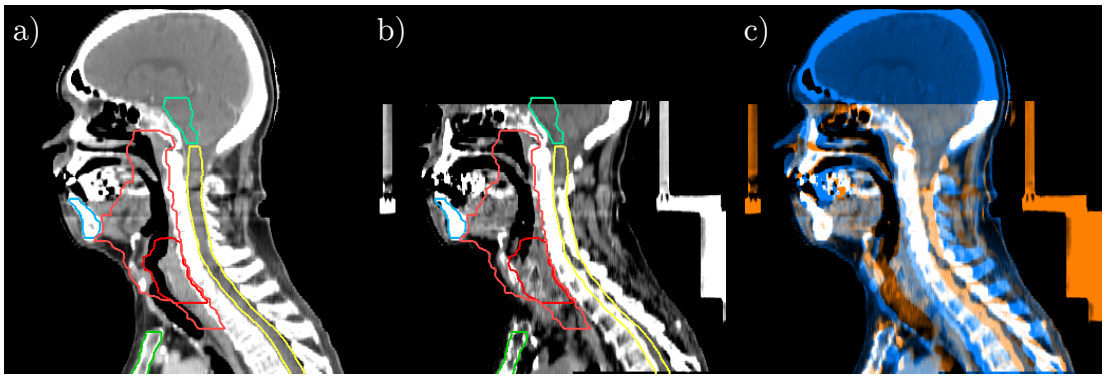

Figure 2.2: Representative image data indicating inter-fractional motion. a) Planning CT including target volumes (red) and OAR contours. The spinal cord contour (yellow) is correctly delineated b) Fraction CT with original contour overlay of planning CT. The spinal cord contour (yellow) does not match the actual spinal cord position in the image anymore. c) Image fusion of planning CT (blue) and fraction CT (orange) shows the movement in the image space (alignment would show in gray scale, see Section 2.2.1.

2.1.4 Management of Deformations

The motion present in different regions of the human anatomy can vary widely in the time scale and the range of motion. In addition, the elasticity of the tissue determines the effect of deformations. Therefore, motion management has to be specific to the tumor location being treated and the individual constitution of the patient. A first step that is always considered is the choice of the treatment posture in a reproducible and reasonably comfortable state for the patient to minimize

motion and facilitate desired dose distributions [20]. Photon irradiation in the thorax for example is typically performed in an arms-up position to avoid dose to healthy tissue. In particle therapy, the use of arms-down postures becomes feasible, which can be more comfortable for the patient and lead to a more consistent patient position during the irradiation [21].
In addition, several approaches are considered to minimize motion or the effect motion has on the treatment. A selection of these approaches is summarized in the following.

Immobilization

Immobilization describes the concept to limit the range of motion of a patient by using physical objects during the set-up. It can be used to position the patient in a reliable and reproducible manner since they are boundary conditions to the posture of the patient. This can reduce inter-fractional motion. In addition, immobilization can prevent parts of the intra-fractional motion by restricting the patient's range of motion. The choice of immobilization depends on the specific region of the human body that is irradiated.
In the head and neck region, patient-specific face masks can be used to restrict motion. They can be created either from scotch cast fitting to the patient surface [22] or from thermoplastic materials [23]. While such masks decrease the set up errors to about $3-4\,\mathrm{mm}$, they cannot guarantee full fixation [24, 25]. To include the shoulder region into the fixation, vacuum mattresses can be used that are once fitted to the patient and then provide a rigid immobilization [26].
For the abdominal region, the use of compression devices can limit the effect of breathing motion on the abdominal organs and can decrease the intra-fractional motion of e.g. kidneys, liver, or even the lung itself [27–29].

Safety Margins

The concept of safety margins compensates for motion-induced under-dosage of the tumor volume. They assure consistent target volume radiation by increasing the irradiated volume by a margin. In this approach, the so-called clinical target volume (CTV) is defined as all macroscopically visible tumor tissue and any potential microscopic infiltration. With the additional margin, the planning target volume (PTV) is defined to include any uncertainties including patient motion

and setup errors. Research in this area typically tries to decrease the margins as much as possible without losing or compromising the robustness or target coverage [30, 31]. The standard margin found in head and neck cancers, that still preserved local tumor control, was 3 mm [32].
When considering particle therapy, beam-specific margins (i.e. proximal and distal margins as compared to isotropic margins in photon therapy) are defined [33].

Image Guided Radiation Therapy

To reduce the effect of setup errors in fractionated radiation therapy, the concept of image-guided radiation therapy (IGRT) can be employed. Before each fraction, an image of the patient is acquired. A translation vector and rotation to move the treatment table (and hence the patient) is derived using rigid image registration. The most common imaging technologies for IGRT are CTs on rails [34] and cone beam CT imaging attached to the beam-delivering gantry [35].
More recently, the use of MR-IGRT has been clinically introduced for photon therapy [36] but is also proposed for proton therapy [37]. The use of MRI would introduce an excellent soft tissue contrast without any ionizing radiation and potentially enable real-time imaging during the fraction. More details on these imaging modalities can be found in Section 2.2. IGRT can typically not compensate for deformations or anatomical changes, as the original treatment plan remains unchanged.

Adaptive Radiation Therapy

Adaptive Radiation therapy (ART) is the concept of accepting the fact that human motion will always occur and needs to be compensated in the application of radiation to the patient. ART tries to adapt the radiation in a way that ensures the total dose over all fractions provides the prescribed dose to the target volume while reducing the dose to normal tissue. Two main approaches for ART can be described, the so-called plan-of-the-day and adaptive replanning. Both try to compensate for the inter-fractional motion before each treatment is delivered.
The plan-of-the-day approach relies on a library of pre-computed treatment plans for a range of anatomical variations or potential treatment scenarios. This can be done by acquiring multiple image scans of the patient on different days [38] or by employing organ-specific models to estimate the expected anatomical variations

[39]. It could be shown that even complex treatment scenarios can be handled with a plan-of-the-day approach with a good quality of the treatment plans [40]. The second approach – adaptive replanning – combines the previously applied dose distribution and the current anatomical deformation as assessed on a medical image to provide an on-the-fly calculation of an optimized new treatment plan. This requires appropriate quantification of the anatomical deformations and fast re-optimization of the new treatment plan. Additionally, quality assurance needs to be considered to comply with legal regulations and assure proper treatment of the patient.
In the head and neck region, adaptive radiation therapy is reported to improve the dose application in various studies [41–43]. With the clinical rise of MR-guided ART [44], the requirements regarding image registration (see Section 2.3) have further increased. Accurate transformation models and multimodal image registration are required to adequately quantify the anatomical deformations. Based on this quantification, the decision on whether plan adaptation is required can be made. Then the contours and dose distributions from earlier irradiation on the original planning CT can be propagated into the fraction image data. In the case of MR-guided radiation therapy, this requires a generation of synthetic electron density maps [45]. Since all these steps depend on the transformation between the planning CT and the fraction image, ART requires accurate, robust, and bio-fidelic image registration to provide beneficial treatment.

2.2 Imaging Modalities

In modern radiation therapy, various imaging techniques are used for IGRT and ART [46, 47]. X-Ray and CT imaging are of particular interest since they contain attenuation data that is required for dose calculation in photon radiation therapy and can be used to estimate stopping power for particle therapy dose calculation [48,49]. The invention of magnetic resonance imaging (MRI) introduced a novel and unique soft tissue contrast that has since been investigated intensively to be utilized in radiation therapy [50,51]. An additional benefit of MRI is the fact that no ionizing radiation is used, which results in a lower total dose to normal tissue compared to a CT approach. The fact that continuous imaging is possible with MRI also enables the vision of real-time tracking of tumor [52, 53].

2.2.1 Fundamental Properties and Representation of Medical Images

Medical images as they are considered in this work can be described as 3D arrays $X(i, j, k)$ (sometimes also called a *cube*) that are discrete in space. The singular unit of such an image array is called a volumetric pixel or *voxel*. In each voxel, an integer value describes the image intensity. In this way, all images in this thesis are in principle grayscale images and have no intrinsic color information. Figure 2.3 a) shows how grayscale images and voxel values are connected in a two-dimensional example.

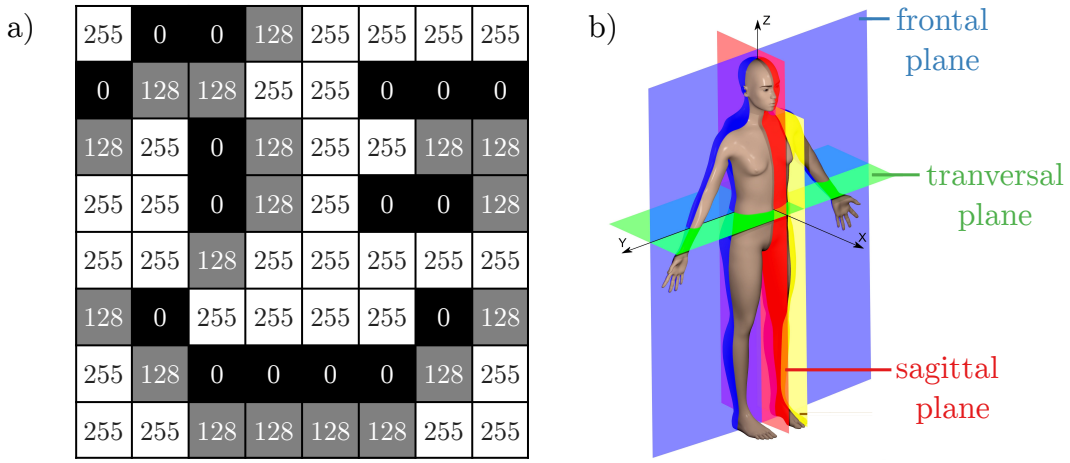

Figure 2.3: Basic conventions for medical images. a) Discretization in space and intensity of a simple 2D grayscale image (in this example as an 8-bit image). Each element – called a pixel or in the case of 3D images a voxel – displays an integer value translated to image intensity.
b) Naming convention for 2D image orientations of the human anatomy. In this thesis, transversal, sagittal, and frontal are used. Figure adapted from David Richfield and Mikael Häggström (CC BY-SA 4.0).

Figure 2.3 b) shows the common notation as frontal, sagittal and transversal orientation that has been established as the convention when displaying the human anatomy in a 2D representation (also called a *slice*). In the visualization of medical images in this thesis, a slice-wise representation of several orientations is used to give an overview of the 3D image.

When comparing images of the same human at different points in time, it is quite often desirable to show both images simultaneously and to indicate their similarity. One approach also used in this thesis is the *color fusion* where the

integer values of both images are used as the inputs for an RGB image. In this work, a complementary color approach is chosen where one image is used for the red and 50 % of the green color channel while the second image is used for the blue color channel and the remaining 50 % of the green channel. This yields a strong contrast between both images and when both images have the same intensity in a voxel, this results in a pure grayscale value that is displayed. Therefore, all differences in the image are displayed as residual color. This approach works particularly well for areas of high image intensity (as is the case for bone tissue on CT images). Figure 2.4 a) shows a typical color fusion of two CT images of the same patient. Differences are highlighted by color while an agreement is in grayscale. A second commonly used technique is *checkerboard* visualization, where images are layered in tiles, with each tile changing the view from one image to another . The tile size can be chosen to represent the size of structures that should be distinguished. The checkerboard is particularly helpful to visualize differences around edges in the image and can be used for images of the same as well as different modalities since it does not rely on intrinsic image contrast. Figure 2.4 b) shows the checkerboard visualization of 2 CTs. At the edges of the tiles, discontinuities indicate the differences between the two images. In the visualization, the tiles can often be moved interactively to yield a dynamic visualization of the differences.

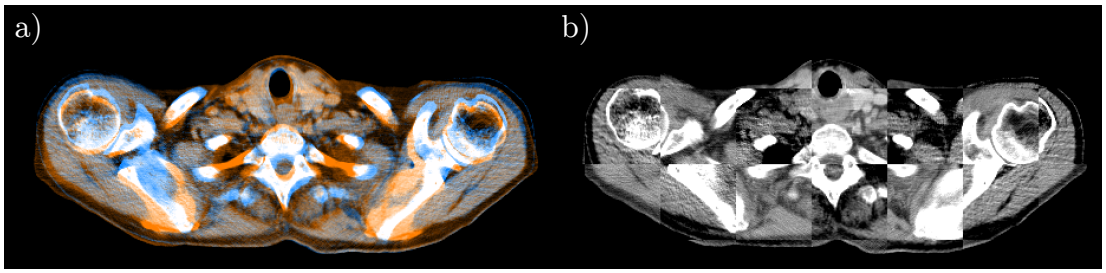

Figure 2.4: Image fusion techniques that are commonly used to visualize differences in image registration tasks. a) Color fusion using a complementary color approach. The same intensities lead to grayscale visualization while differences are shown in orange or blue. Color fusion works best for monomodal images as it relies on image contrast.
b) Checkerboard pattern displaying the two images in alternating square tiles. Differences are visible at the edges between any two tiles. The checkerboard also works for multimodal images.

2.2.2 Computed Tomography

Computed tomography relies on the attenuation (i.e. absorption of scattering) of photons in the X-ray wavelength when interacting with matter. In the simplified one-dimensional case of monochromatic photons transversing a homogeneous medium with linear attenuation coefficient $\mu(E, x)$, the intensity I of photons is given by the differential equation

$$\frac{\mathrm{d}I}{\mathrm{d}x} = -\mu(E, x)\,, \tag{2.3}$$

that describes the exponential attenuation process also known as the Beer-Lambert law [54]. For a deeper understanding of the underlying physical principles that lead to the attenuation, the interested reader is referred to the literature [55, 56]. The fundamental idea of a CT scan is to use multiple X-ray projections to reconstruct the spatial distribution of μ within the considered object as proposed and implemented by Hounsfield and Ambrose [57, 58].
The analytical and historically first approach to CT reconstruction is based on the mathematical properties of the Radon transformation [59]. For a sufficiently well-behaved function $f : \mathbb{R}^2 \to \mathbb{R}, \mathbf{x} \mapsto f(\mathbf{x})$, the Radon transform is defined as the line integral along a straight line L as

$$Rf(L) = \int_L f(\mathbf{x})\mathrm{d}\mathbf{x}\,. \tag{2.4}$$

It denotes the projection of f along the line. The analytical reconstruction of the underlying function f can be performed using *Filtered Back-Projection* in the form of

$$f(\mathbf{x}) = \int_0^{\pi} (Rf * h)\mathrm{d}\theta, \tag{2.5}$$

where the projection data is convolved with an appropriate kernel h. The details and a more rigorous mathematical derivation can be found in the literature [60]. The much more common approach to CT reconstruction is the use of iterative reconstruction approaches. The most commonly used clinical CT is arranged in a fan beam geometry to enable fast acquisition of the projection. However, to generate 3D images, a spiral acquisition is required. Here, the imaged object or patient is contentiously moved through the CT scanner [61].

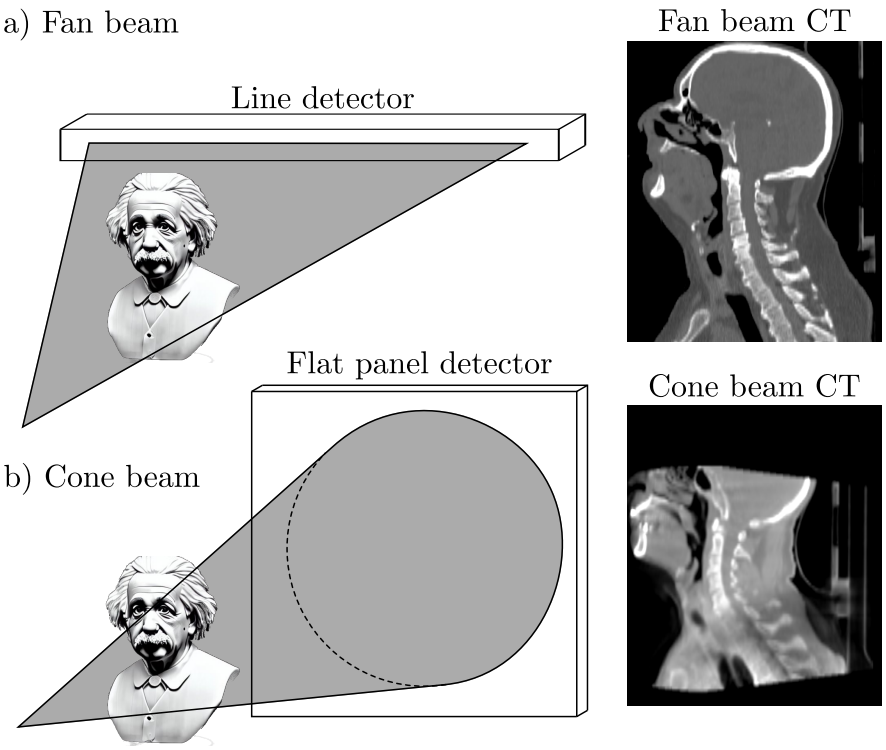

Figure 2.5: Comparison of fan beam and cone beam geometry for CT. The fan beam creates only projections in one plane, which requires longitudinal translation of the imaged object to acquire 3D information. The cone beam CT geometry uses a flat panel detector to acquire 3D information without longitudinal translation.
Representative sagittal slices for both modalities

2.2.3 Cone Beam CT

To allow for image acquisition without longitudinal translation, a so-called cone beam geometry can be used. Here, the X-ray projections are acquired from a cone-shaped beam on a flat panel detector. This means no longitudinal translation of the imaged object is required and hence image acquisition in the treatment position is feasible. As a drawback, these cone beam CTs (CBCTs) typically have inferior contrast-to-noise ratio and suffer from limitations in the field of view [62, 63].
Figure 2.5 shows the different geometries of the fan beam and cone beam approach and two representative sagittal slices of resulting images in the head and neck area. It can be seen that the cone shape of the beam leads to a limitation in the field of view. Additionally, CBCTs suffer among others from cupping artifacts that lead to inconsistent contrast in the image center when insufficient corrected [64].

2.2.4 Magnetic Resonance Imaging

Magnetic Resonance Imaging relies on the macroscopic magnetization $\vec{M}_0$ that arises when a system of N nuclei of spin $\frac{1}{2}$ and gyromagnetic ratio γ is within a magnetic field $\vec{B}$ [65]. The magnetization $\vec{M}_0$ is given by

$$\vec{M}_0 = \frac{N}{V}\frac{\gamma^2\hbar^2}{4k_BT}\vec{B}, \tag{2.6}$$

where V is the occupied volume and k_B is the Boltzmann constant. The time evolution of this macroscopic magnetization vector is given as

$$\frac{\mathrm{d}}{\mathrm{dt}}\vec{M}(t) = \vec{M}(t) \times \gamma\vec{B}(t)\,. \tag{2.7}$$

If $\vec{M}(t) \perp \vec{B}(t)$, this time evolution can be described as a precession of the magnetization with the Larmor frequency $\omega = \gamma\left|\vec{B}\right|$. In combination with the Bloch equations describing the relaxation process of non-equilibrium magnetization [66], the foundation for magnetic resonance imaging is set. A more detailed description of the physical background of MR Imaging can be found in the literature [67, 68]. In the context of image registration, it should be noted that all MR images are subject to distortions due to the acquisition technique in inhomogeneities in the magnetic field that need to be considered and corrected for [69]. Therefore, all presented MR images underwent distortion correction [70].
Different tissues in the human anatomy show different relaxation times and also different Larmor frequencies due to their chemical shift. This gives rise to a multitude of imaging sequences to manipulate and acquire the MR image with a specific contrast [71]. In this work, all shown MR images are acquired using a three-dimensional Dixon sequence [72, 73].
For the Dixon sequence, two images are acquired at different echo times (specific to the Larmor frequency of water and fat/lipids). The echo time is chosen to result in so-called in-phase and opposed-phase images. Adding these images yields the water signal while subtracting them results in the lipid signal. The resulting images are shown in figure 2.6. Due to the characteristic dark band around bones in the in-phase image, this mode is used in the presented work to identify bone tissue with the MR image.

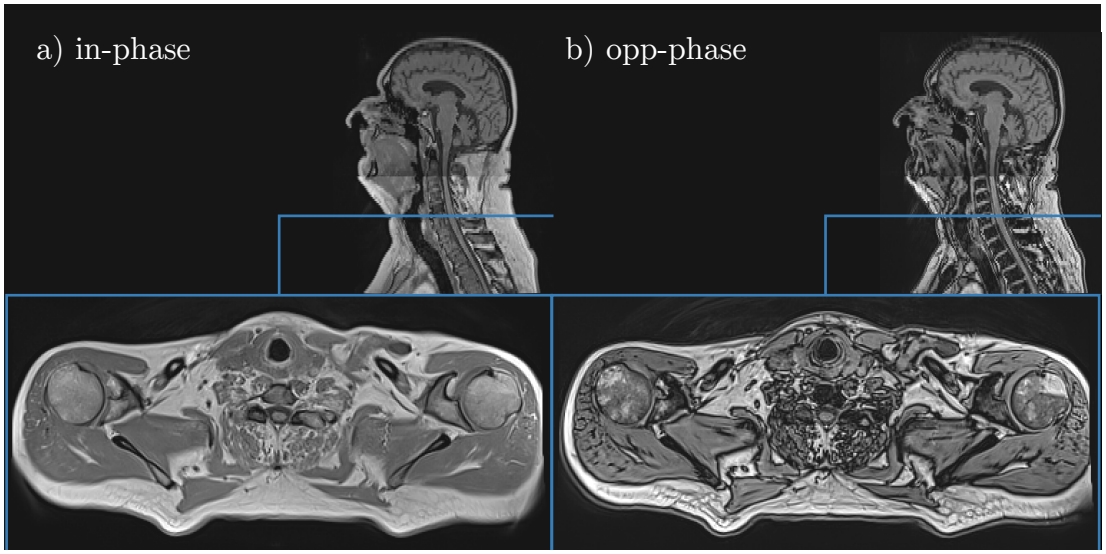

Figure 2.6: Resulting images from a Dixon sequence. a) in-phase image, b) opposed-phase image. A transverse slice and sagittal are displayed (the blue line indicates the position of the transverse slice). By adding or subtracting the images, the isolated water signal and lipid signal, respectively, can be calculated. The in-phase image shows a characteristic dark band around bones and is hence used in this thesis for bone identification in the MR image.

2.3 Image Registration

Whenever two images of the same or different modalities need to be aligned in the image space, an image registration problem arises. This class of typically inverse optimization problems tries to find the transformation that maximizes the overall alignment of image content as quantified by a similarity metric.

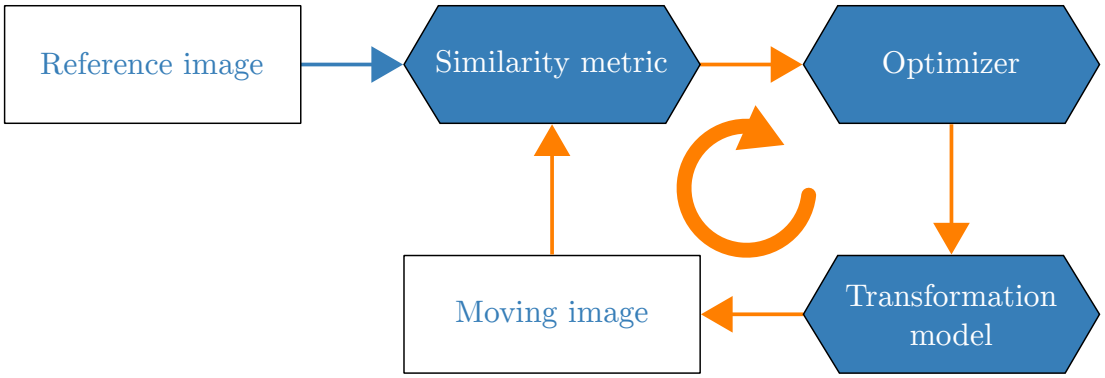

Figure 2.7: Schematic image registration process. A reference image and a moving image are compared using a similarity metric. In an iterative approach, an optimizer finds the parameters for a given transformation model such that the transformed moving image yields the maximum similarity.

The general approach to image registration is shown in Figure 2.7. When considering a set of two images, one of the images can be defined to be the reference image that is kept fixed during the optimization and is therefore called fixed image or reference image. The second image is transformed in the registration process and therefore it is typically called moving image [74]. The quantitative measure of alignment is calculated using a similarity metric (see Section 2.3.1). For this purpose, the moving image is re-sampled to the voxel grid of the reference image. To maximize the similarity, and hence register the images, an optimizer is employed in an iterative process.
In each iteration, the free parameters of a chosen transformation model (see Section 2.3.2) are changed in order to improve the similarity between the transformed moving image and the reference image.

2.3.1 Similarity Metric

The similarity metric chosen for an image registration scheme is essential since it defines how to quantify the spatial alignment of two images' content. It serves as

the objective function in the optimization procedure and represents the requirements of the aligned images.
In the case that both images are acquired with the same setting and the same imaging modality (see Section 2.2) a linear relationship between the image intensities can be assumed. This enables the use of monomodal similarity metrics. A commonly used monomodal similarity metric is the sum of squared differences (SSD) in other settings called the mean squared error (MSE) [75]. Given two image intensity arrays X and Y for the fixed and the moving image, respectively, the SSD is defined as

$$\mathrm{SSD} = \frac{1}{N}\sum_{i=1}^{N}(Y_i - X_i)^2 \,, \tag{2.8}$$

where N denotes the number of voxels in the images. The use of SSD in monomodal registration schemes is favored because it is simple to compute and simple to differentiate. This allows fast computation and use of gradient-based optimizers [76].
Cross-correlation is a second similarity metric typically used for monomodal image registration [77]. It is given as the complex conjugated convolution of both image intensity matrices

$$\begin{aligned} R_{XY}(\tau_1,\tau_2) &= X^*(\tau_1,\tau_2) * Y(\tau_1,\tau_2)) \\ &= \sum_{i=1}^{N_1}\sum_{j=1}^{N_2} X^*(i,j)Y(\tau_1 - i, \tau_2 - j)\,, \end{aligned} \tag{2.9}$$

where N_1, N_2 are the dimensions of the image.
If there is no linear relationship to be assumed between the image intensities, a multimodal similarity metric needs to be employed. The most important and widely used multi-modal similarity metric is the mutual information (MI) [78]. In the case of image registration, it can be formulated as

$$\mathrm{MI} = \sum_{i,j} p_{XY}(i,j)\log_2\frac{p_{XY}(i,j)}{p_X(i)p_Y(j)}. \tag{2.10}$$

where p_{XY} is the joint intensity distribution and p_X and p_Y are the marginal intensity distributions in both images.
One drawback of MI is the limited statistical power for lower sample sizes leading to inconclusive results [79]. For smaller sub-regions of the image, this decreases the benefit for the use in image registration.

2.3.2 Transformation Model

The transformation model is the second central property of any image registration scheme since it classifies the transformations that are permissible in the registration process. Therefore, it has a tremendous impact on the final result. Since the transformation model defines the number of free transformation parameters, i.e., the degrees of freedom (DoF) of the optimization problem, it determines the complexity of the registration scheme. This directly affects the computational time required for the registration.
The arguably most basic transformation model is a rigid body model. In the case of a three-dimensional image, it consists of three degrees of freedom for the global translation of the image and three degrees of freedom for the global rotation. Using the image center as the rotation and coordinate center, the rigid transformation from an image voxel x to the transformed voxel x' can be written as

$$x' = Rx + T\,, \tag{2.11}$$

where T is the translation vector and R denotes the rotation matrix.
If the object in the image is deformable – as is the case for most human anatomies, particularly in soft tissue – a non-rigid registration approach has to be considered to adequately represent the potential changes in the image.
A common approach to model deformations is non-linear interpolation in space between distinct, independently transformed points. In principle, control points are distributed within the image space and they undergo point-wise three translational DoF registration. The transformation of space between these points is often modeled following mathematical basis functions and in most cases does not incorporate any additional knowledge of the imaged object.
The most common type of interpolation function are splines, i.e. piece-wise polynomial functions. In the field of medical image processing, the two most important classes of basis functions are thin-plate-splines [80] and B-splines [81].
More complex models often include tissue properties or try to incorporate physical motion laws to improve the correspondence between the real motion that occurred in the image and the transformation that is found in the registration process. In this thesis, biomechanical models are considered in particular. An overview is given in Section 2.4.

2.3.3 Optimizer

The general idea of an optimization algorithm or short optimizer is to find a solution for a given optimization problem. Given a variable vector $x \in \mathbb{R}^n$ and a cost function $f : \mathbb{R}^n \to \mathbb{R}, x \mapsto f(x)$ that is supposed to be maximized[1], the optimization algorithms find the solution to

$$\arg\max_{x \in \mathbb{R}^n} f(x) := \{x \in \mathbb{R}^n : \forall \hat{x} \in \mathbb{R}^n : f(\hat{x}) \leq f(x)\} \ . \tag{2.12}$$

In the context of image registration, this means maximizing the similarity of both images. It should be noted that in most cases it is neither necessary nor realistic to find the exact global optimum. Therefore, most optimization strategies try to find a close-to-optimal solution in a time-efficient way.
Since a global maximum is often not obvious, most optimizers operate in an iterative manner. After an update of the transformation parameters, the moving image is transformed and re-sampled. Then, the similarity is calculated and checked if sufficient convergence is achieved. If convergence is not yet achieved (and if the maximum number of iterations is not reached yet), the optimizer once again updates the parameter set.

2.4 Deformation Models

Deformation models introduce non-rigid image transformation and are therefore favored the registration of medical images in the presence of motion. There are mathematical models of interpolation under smoothness constraints. In addition, there are biomechanical models. These allow the inclusion of biophysical and physiological properties into the image registration process. They provide a transformation model that is consistent with the underlying model assumptions. In this section, an overview of different models is given. The selected models are either commonly used in image registration or – in the case of chainmail and kinematic models – they are relevant to the presented thesis.
Figure 2.8 shows an overview of the relationship between complexity and computational speed for different classes of transformation models. While finite element methods (FEM) can support the highest complexity in the model they are known

[1]The cost function could as well be minimized. In terms of similarity in the image registration, the case of maximizing the cost function is described. Minimization works analogously.

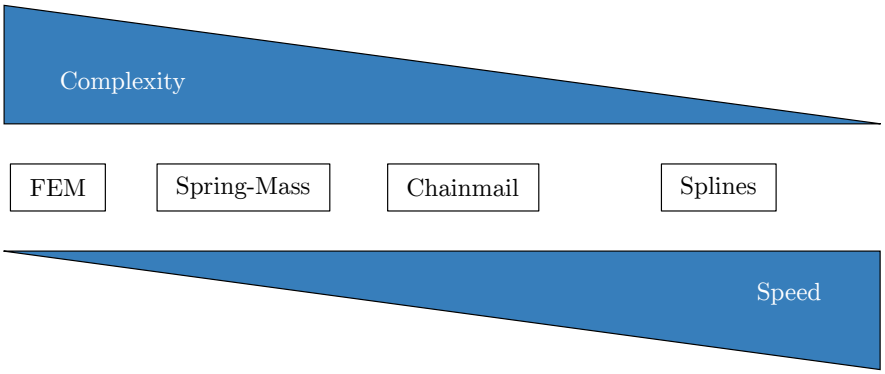

Figure 2.8: Classification of deformation models regarding speed and (computational) complexity. Splines are typically simple but fast while finite element methods can represent even complex systems at the cost of computational speed. The chainmail is a trade-off between both aims. Schematic adapted from Bartelheimer [82].

to be computationally slow. Splines on the other side of the spectrum are fast but only include interpolation that cannot model highly complex systems. The chainmail approach is a trade-off between complexity and speed. The question of which model should be chosen also depends on the required accuracy in the given application of image registration.

2.4.1 Splines

Splines are piecewise polynomial functions that are utilized frequently for interpolation. In the context of image registration, splines are used to approximate the displacement between control points for which the displacements are explicitly calculated. In the following, three representative and commonly used forms of splines are introduced.

B-Splines Basis-splines or for short *B-splines* are a very common method used in image registration since they allow an inherent degree of smoothness when combined. The name arises from the property of B-splines to form a basis of all spline functions of a given order [83].
In their application, a uniform grid of automatically selected control points ϕ_i is used to transform points x in the local environment of a control point [84]. The

transformation is given by

$$t(x) = \sum_{l=0}^{N} B_l(x)\phi_i \,. \tag{2.13}$$

This means the choice of control points is essential when working with B-splines. Rueckert et al. note that large spacing of control points is useful to identify global deformations while small spacing will yield local non-rigid transformations [84]. In the practical application, this is often solved by a step-wise registration with decreasing spacing of control points. The local transformations are then summed up to yield the total transformation.

Radial Basis Functions Radial Basis Functions (RBFs) are characterized by the radial symmetry in the distance to a given point in space. The general form of an RBF transforming points x to points $t(x)$ is given by

$$t(x) = ax + b + \sum_{i=1}^{N} c_i R(|x - x_i|) \,, \tag{2.14}$$

where a and b denote the global rotation and translation, respectively. The local deformation is given by the sum over all basis function R that only depend on the distance. These radial basis functions can take the form of Gaussians, multiquadratic or so call thin-plate spines (TPS). To determine the coefficients a,b, and c_i, by solving the equations resulting from equation 2.14 for all control points [85, 86].

Thin-Plate Splines Thin-plate splines (TPS) are one type of RBF whose basis function R is given as

$$R(|x - x_i|) = (x - x_i)^2 \ln(|x - x_i|). \tag{2.15}$$

The name of TPS arises from the original association with the bending of thin metal plates since it minimizes the bending energy given forces on fixed points [80].

2.4.2 Spring-Mass Model

In a spring-mass model, the human anatomy is simplified to a system consisting of massive particles that are connected by dampened springs. The model can be refined by topological adjustments, e.g., which particles are connected, and how coarsely the object is discretized. The one-dimensional motion of each massive particle is then described using classical mechanics in a differential equation

$$m\ddot{x} - \gamma\dot{x} - kx = 0\,, \tag{2.16}$$

where k denotes the spring constant and γ the dampening coefficient. The parametrization of these factors in the differential equation can be used to approximate the deformation of the considered tissue and hence allows tissue-specific modeling.
This model can be used for image registration [87] including several adaptations [88, 89]. Such adaptions include shear forces and dashpot dampening. While spring-mass models can be a computationally efficient way to approximate anatomical formations, they are not always convergent and can lack the required accuracy for utilization in radiation therapy [90].

2.4.3 Models Based on the Finite Element Method

Finite element models are based on the assumption that continuous mechanics for an object can be approximated by separating said object into finite elements typically in tetrahedral shape or whatever shape represents the inherent symmetries of the problem.
On the so-called mesh created by the finite elements, the partial differential equations that are typically used to describe continuum mechanics can be transformed into a coupled system of ordinary differential equations for each node (i.e. corner point of a finite element). Let $\mathbf{x}$ denote the vector containing the coordinates of all nodes, then the equations of motion can be formulated as

$$M\ddot{\mathbf{x}} + D\dot{\mathbf{x}} + K\mathbf{x} - \mathbf{f} = 0\,, \tag{2.17}$$

where M denotes the masses of each element, while D and K are the dampening and stiffness matrix, respectively. The external forces are given by $\mathbf{f}$. In practical

applications, the solution for the system of equations is found numerically.
For biological tissues, a typical assumption is that they are homogeneous (per tissue type), and isotropically linear elastic. This means they can be parameterized using two elasticity parameters. First, the Young's modulus Y relates mechanical stress σ with the strain ε as

$$Y = \frac{\sigma}{\varepsilon} \tag{2.18}$$

Secondly, the Poisson ratio ν gives the ratio between transverse strain ε_{trans} and axial strain ε_{ax}. It is defined as

$$\nu = -\frac{\mathrm{d}\varepsilon_{trans}}{\mathrm{d}\varepsilon_{ax}}\,. \tag{2.19}$$

An in-depth consideration of the different finite element approaches goes beyond the scope of this thesis, but the interested reader is referred to the literature [91,92]. It should be noted, that an essential part of any finite element method is the parametrization. These tissue properties can either be determined using cadaver experiments [93, 94], or in-vivo using MR-elastography [95]. It is not unusual, however, to use a fitting procedure to determine the patient-specific parameters [96].
Finite element models typically rely on the quality and size of the mesh and often require intensive refinement (e.g. smoothing and node reduction). Even then, they are considered computationally expensive, in particular when high precision is required.

2.4.4 Chainmail-Based Model

The concept of the 3D-Chainmail has been introduced by Sarah Gibson in 1997 [97]. The fundamental idea is to divide an object of interest into a 3D grid of interconnected elements. Transformations of these elements propagate along the connections but are restricted by geometric constraints including the permissible range of translation and shear. The chainmail concept and the effect of the geometrical constraints are illustrated in Figure 2.9. The transformation of all elements within the object describes the transformation of the whole object.
The original chainmail approach can be enhanced to include heterogeneous materials, e.g. different tissue types in the human anatomy [98]. For this, the constraint

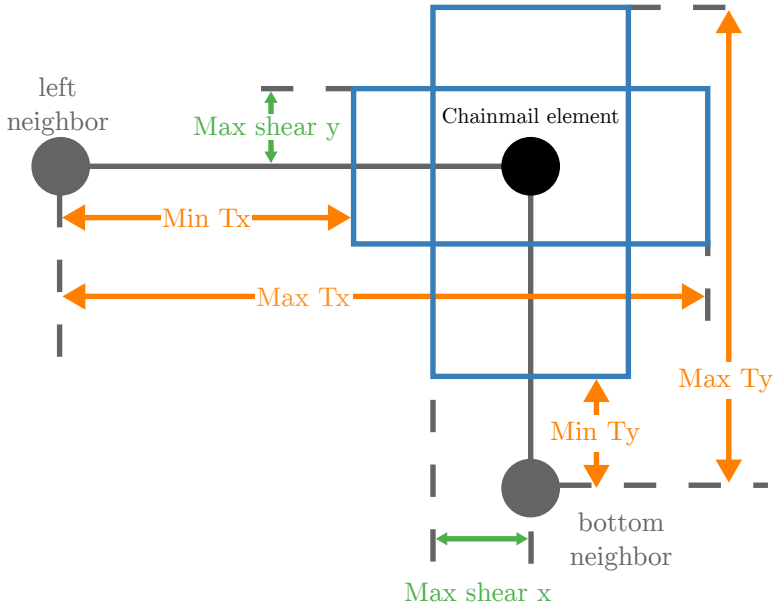

Figure 2.9: Concept of the Chainmail approach. Each point is connected to its neighbors and can move within a certain range relative to said neighbors (blue boxes). Geometrical constraints are given for the maximum shear and the translation. Figure recreated from Gibson (1997) [97].

parametrization has to be adapted depending on the tissue type. This can be done using the image information in a CT. The chainmail concept is used in this thesis for the propagation of soft tissue deformation. This is explained in more detail in Section 3.2. Further adaptations included in the utilized version of the chainmail algorithm include rotational motion propagation as well as volume preservation to account for tissue growth or shrinkage [99, 100].

2.4.5 Kinematic Model

The kinematic model is the central approach investigated in this thesis. In this section, the general idea of a kinematic model is introduced and a representative approach to solving the inverse kinematics is given. This section follows the structure of [101].
A kinematic model in this context is a multi-body system of rigid bodies that are connected via joints. The transformations in this model are described using kinematic motion laws. Figure 2.10 shows two multi-body systems and illustrates the analogy of such systems in the representative case of a robotic arm and the skeleton structure of a human arm. In principle, the rigid bodies can move following the degrees of freedom of the joints.

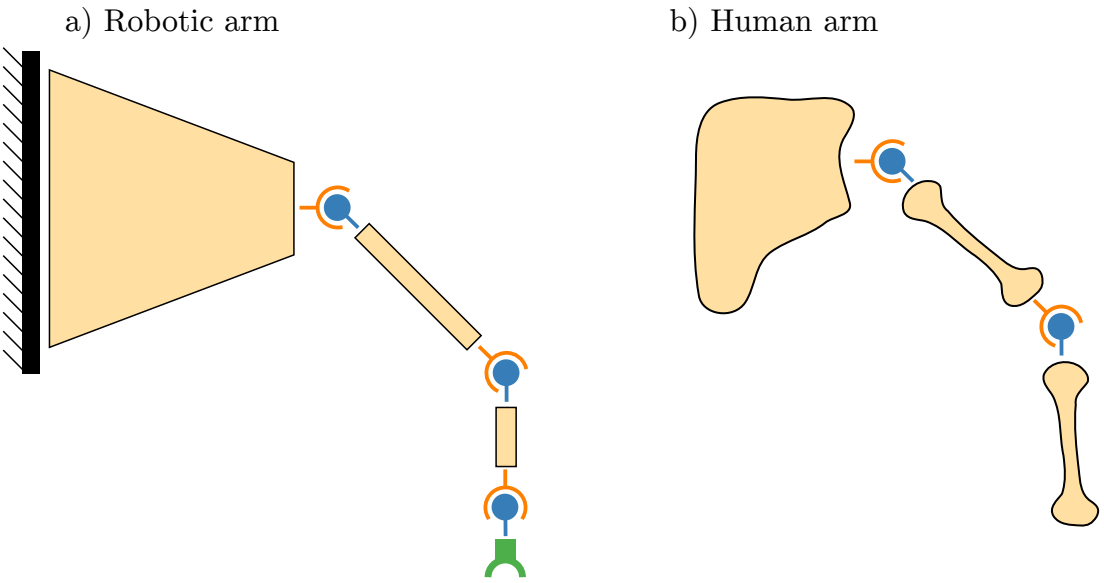

Figure 2.10: Representation of the concept of a kinematic model. a) A robot arm consisting of three rigid bodies (yellow) that are connected via joints (orange/blue). At the end of the robot arm, there is an end effector. b) Schematic of the human arm as modeled in analogy to the robot arm. The transformation of the rigid bodies is constrained by the DoF of the joints. Figure recreated from [101].

In a purely kinematic approach, there is no consideration of forces or acceleration in the system. In the case of the considered multi-body system, kinematics describes the relationship between the joint parameters (e.g. the angle in the joint) and the coordinates of the rigid bodies. In *forward kinematics* the position and orientation of the rigid bodies X can be calculated from the set of joint parameters Θ via a

function f

$$X = f(\Theta)\,. \tag{2.20}$$

To find the joint parameters given the position of rigid bodies is called *inverse kinematics*. The naive approach of an inverse function f^{-1} such that

$$\Theta = f^{-1}(X) \tag{2.21}$$

fails since in general the existence and uniqueness of this inverse problem are not given. For such an ill-posed problem, numerical optimization methods can be used [102] to find a solution with the desired properties. The most important numerical methods used for inverse kinematics are the Jacobian pseudoinverse and Jacobian transpose methods [103, 104]. However, more general numerical methods are also used to solve the inverse problem. Examples include but are not limited to damped least squares methods, cyclic coordinate descent methods, and conjugate gradient methods [105–107].
As a representative approach to how the solution of inverse kinematics can be achieved, the Jacobian pseudoinverse method is described following the works of Buss and Meredith [108, 109]. In this approach, the joint parameter vector Θ is determined using the Jacobian matrix $J(\Theta) = \left(\frac{\partial f}{\partial \Theta_j}\right)_j$ in an iterative manner. In a finite difference computation, the change in the end-effector ΔX can then be linearly approximated as

$$\Delta X \approx J \Delta \Theta. \tag{2.22}$$

This means to find the iterative update for Θ, the inverse Jacobian matrix J^{-1} could be used as

$$\Delta \Theta = J^{-1} \Delta X. \tag{2.23}$$

In this approach, ΔX can be calculated as the distance between the current and the target position. The inversion of the Jacobian is an ill-posed problem itself, which can be approximated using the Moore-Penrose pseudoinverse J^+ [110–112]. It is calculated as

$$J^+ = J^T (J J^T)^{-1} \tag{2.24}$$

The error e_{pseudo} introduced by the use of the pseudoinverse can be estimated as

$$e_{pseudo} = \left\| (I - JJ^{+})\Delta X \right\| . \tag{2.25}$$

If this error is above a threshold, a reduction in the step size can be used [109]. Finally, the joint parameters are updated via

$$\Theta_{n+1} = \Theta_n + \Delta\Theta = \Theta_n + J^{+}\Delta X \tag{2.26}$$

until sufficient convergence is achieved.
In the presented thesis, inverse kinematic calculations are performed using the Simbody library [113]. The details of the implementation are covered in details in Section 3.1.

2.5 Deep Learning and Semantic Segmentation

An essential task in many image processing applications is the segmentation of objects. In medical images, this can be done manually by delineation of the organs or regions of interest using the input of a human observer. However, these manual segmentations require knowledge and experience and even then, manual segmentation remains a time-consuming task. To automate and standardize this procedure, an artificial neural network (ANN) can provide a semantic segmentation. Given an image, a semantic segmentation classifies each voxel as part of a class (e.g. skull, lung, or background) [114].
ANNs with several hidden layers are called deep learning models and have shown the potential to solve the semantic segmentation tasks accurately [115]. The most common types of deep learning-based segmentation approaches rely on the U-net architecture or modification of it [116]. These approaches make heavy use of convolutional computation as a way to extract feature information from the image. In the following, the concept of training such a network and the consequent prediction are described. This section aims to give a short overview. For a detailed mathematical description, the interested reader is referred to the common literature [117, 118].
From a mathematical point of view, a deep learning-based ANN can be seen as a function f that maps from an input vector x to an output vector y. The function

is parametrized by so-called weights $w \in \mathbb{R}^n$ and biases $b \in \mathbb{R}^m$

$$y = f(w, b, x)\,. \tag{2.27}$$

These weights and biases are considered trainable parameters that crucially determine if the ANN can produce the desired result. Following the name, these trainable parameters are optimized in the training phase of the neural network.
For the training and evaluation of an ANN, ground truth data is required. This ground truth consists of labeled input data, i.e. input vectors x with their expected output y. In the case of semantic segmentation of medical images, the typical ground truth data consists of manually segmented images that are labeled by experts. Typically, the ground truth data is divided into the training data and the test data.
The training data is used to optimize the trainable parameters such that the ANN gives the desired outputs. A loss function $\mathscr{L}$ determines the quality of the output given the current set of trainable parameters. The final set of parameters is found by optimizing this loss function

$$\underset{w\in\mathbb{R}^n, b\in\mathbb{R}^m}{\arg\min}\ \mathscr{L}(f, x, w, b)\,. \tag{2.28}$$

In the context of deep learning, this relies on stochastic gradient descent and back-propagation [119]. Typical loss functions can be the SSD or the cross-correlation (see Section 2.3.1). To evaluate the performance of an ANN, the test data is used. All data in the test data set has not been used to optimize the wights and biases. The assumption is that a properly trained ANN can generalize the correct prediction for previously unseen data. The prediction $f(x)$ given the final set of trainable parameters is compared to the ground truth as a measure of accuracy.

3 Materials and Methods

3.1 Kinematic Model: The Puppetmaster

The kinematic skeleton model – called the *Puppetmaster* – as the central motion model of this thesis is based on the previously published work [99, 101]. It follows the principles established in Section 2.4.5 to provide a transformation model for articulated rigid structures.
The Puppetmaster as utilized in this work models the head and neck region and follows the assumption that major motion in the head and neck region is governed by skeletal motion.[2]

3.1.1 Set-up of Patient Specific Model Geometry and Joint Positioning

The generation of the Puppetmaster model is performed in two steps. First, the patient-specific geometry is set-up from a planning CT scan where all bones are individually delineated by a human observer. Using triangulation with a marching cube algorithm, a 3D representation of the skeleton is created [120, 121]. This process is illustrated in Figure 3.1. To enable the proper connection between the ribs and sternum, the costal cartilage is included in the delineation.
Second, the mobilization of this static geometry is achieved by positioning joints to connect bones. This follows a generic dependence graph (Figure 3.2 a), which is valid for healthy human (i.e. no missing or bones or malformations). This graph includes the most important synovial joints from the skull to the pelvis as long as they are relevant for the application of the model in adaptive radiation therapy. In this process, fibrous joints that do not contribute any degree of freedom to the motion are not modeled. The resulting three-dimensional model including anchor points (red) for the joints is shown in Figure 3.2 b.
All joints in the Puppetmaster are modeled as 3-degree-of-freedom ball-socket joints

[2] This is in contrast to other regions of the human anatomy – e.g. the abdominal region – where deformations in the image are driven by deformable organ motion.

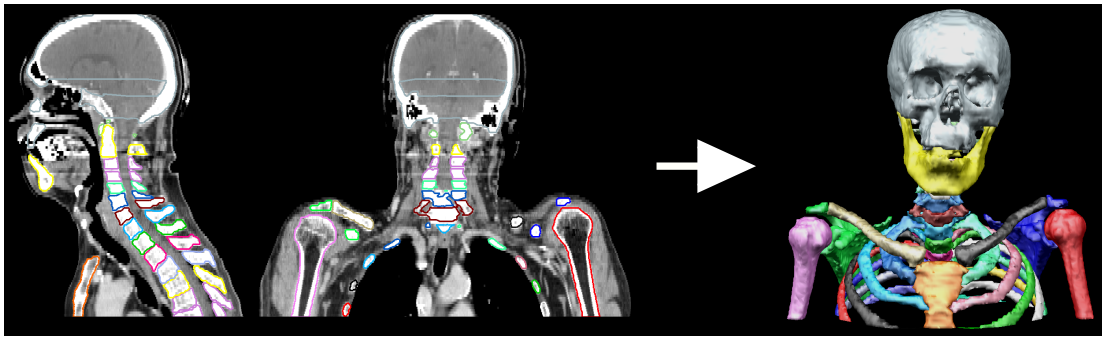

Figure 3.1: Illustration on the set-up of the skeleton geometry. From the individual delineations (representative sagittal and frontal slices), a 3D representation is calculated. This forms the static skeleton used as the foundation of the kinematic model. Recreated from [101].

without mobility restrictions. This gives the model a wider range of motion than anatomically possible. They are positioned using a rule-based system. Depending on the joint type, its position is either calculated as the nearest distance between a pair of connected bones [101] or determined using joint-specific rules [122].

3.1.2 Kinematic Tree and Inverse Kinematic Solver

To properly model skeletal motion, the Puppetmaster needs to solve inverse kinematics to find proper joint parameters for any given posture. This is done using the open-source toolkit *Simbody* [113]. It provides multi-body physics tools to model an articulated (biomechanical) system. In this sense, Simbody can simulate multi-body systems by solving kinematic as well as dynamic equations of motion [113]. To minimize the computational cost of the simulation, internal coordinate formulations are used to yield a minimal set of ordinary differential equations that can be solved efficiently [123]. Additionally, Simbody utilizes coordinate projection methods to prevent numerical drift in the constraints efficiently and robustly [124].

For the use of Simbody, it is essential that the internal coordinate formulation has to be in a kinematic chain. This means the rigid bodies need to be connected by joints in a topological tree without any loops or unconnected bodies. This enables additional simplifications of the calculations and hence an increased computational speed [125]. In the Puppetmaster, the kinematic tree starts at the skull as the so-called root element. The skull is usually well fixated and small motion can be expected. From there on, each bone is connected to exactly one unique parent

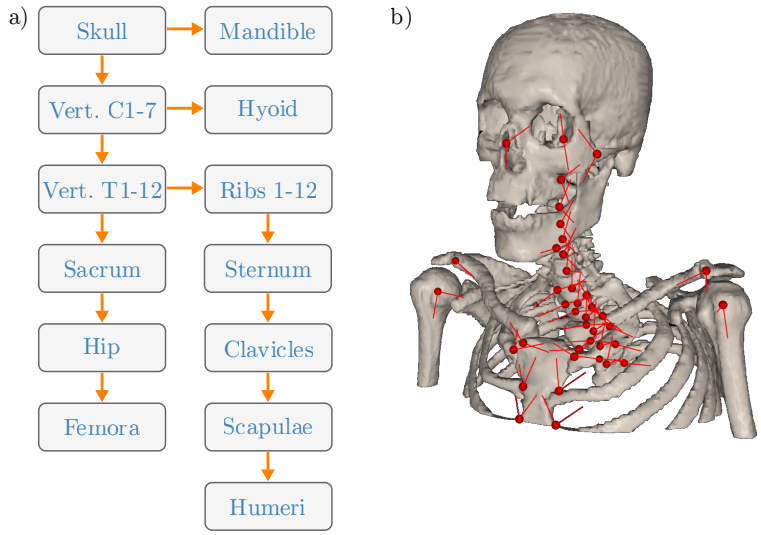

Figure 3.2: a) Dependency graph of the generic human skeleton from the top to bottom. In this thesis, the head and neck part is considered in all applications. b) Resulting joint positioning following the generic dependency graph. The joints are positioned in a nearest-neighbor connection or according to joint-specific anatomical rules.

bone. Figure 3.3 a) displays the kinematic tree for one representative patient that connects the anchor points of joints (red) and centroids of bones in a tree structure. This structure follows the generic dependency graph introduced in Figure 3.2.
In the Simbody toolkit, the permissible motion between two rigid bodies connected by a joint is described by a mobilizer [123]. Figure 3.3 b) illustrates a representative mobilizer between the scapula and the humerus. Each mobilizer adds additional degrees of freedom to the multi-body system by introducing new coordinates.
The mobilizer from a parent bone P to a child bone B can be described using the notation of Seth et al. 2016 [126]. The set of equations is given as

$$ {}^{P}X^{B} = \left[{}^{P}R^{B}(q) \quad {}^{P}p^{B}(q) \right], \tag{3.1} $$

$$ {}^{P}V^{B} = \begin{Bmatrix} {}^{P}\omega^{B}(q,u) \\ {}^{P}v^{B}(q,u) \end{Bmatrix} = {}^{P}H^{B}(q) \cdot u\,, \tag{3.2} $$

$$ {}^{P}A^{B} = {}^{P}\dot{V}^{B} = {}^{P}H^{B}(q)\dot{u} + {}^{P}\dot{H}^{B}(q)u\,, \tag{3.3} $$

$$ \dot{q} = N(q)u\,. \tag{3.4} $$

The transformation ${}^{P}X^{B}$ includes the rotation matrix R and the translation vector p. The velocity ${}^{P}V^{B}$ and acceleration ${}^{P}A^{B}$ can be described by the so-called hinge matrix H and its time derivative. It should be noted that in purely kinematic

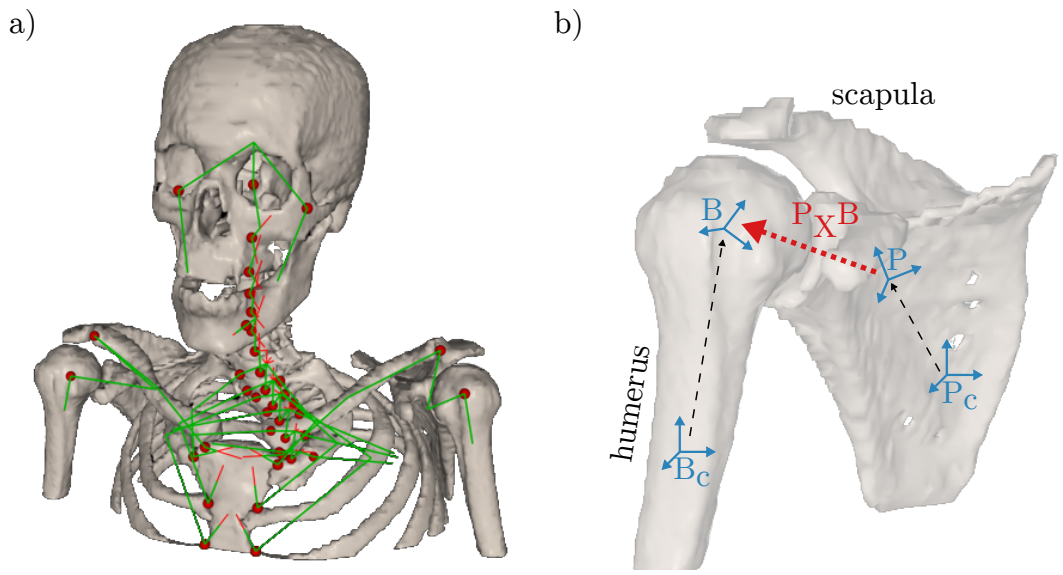

Figure 3.3: Implementation of inverse kinematics using Simbody. a) Kinematic tree (green) connecting anchor points of joints and centroids of bones. b) Representative mobilizer ${}^{P}X^{B}$red) between a parent body P and a child body B. The coordinate systems of the anchor points (blue) are given relative to the coordinate systems of the rigid bodies, defined in the centroids of the bones P_c and B_c, respectively. Relative coordinate transformations are displayed as dotted lines. Partially adapted from [101].

applications, the hinge matrix is constant, and hence $\dot{H} = 0$. The spatial velocity V is decomposed into an angular part ω and a linear part v. Finally, the kinematic coupling matrix N gives the relation between the generalized coordinates q to the generalized velocities u to maintain numerical stability [126]. By recursively applying equations (3.1) - (3.4) from the root element through the kinematic tree, the relative kinematics of each body can be determined.

The Puppetmaster uses ball-and-socket joints. Assuming no translation, this allows the transformation ${}^{P}X^{B}$ to be computed as

$$ {}^{P}X^{B} = \left[{}^{P}R^{B}(q) \;\; 0 \right], \tag{3.5} $$

using the Euler angles of rotations as the generalized coordinate $q = \{\theta_1, \theta_2, \theta_3\}$ [127]. As there are only rotational degrees of freedom, the linear component of

the velocity has to be zero. The corresponding hinge matrix is defined as

$$ {}^{P}H^{B}(q) = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix}^{T} . \tag{3.6} $$

In combination with mobilizers, kinematic constraints are required to limit the range of permissible motion within realistic boundaries. They can be applied to the generalized coordinates, speeds, or accelerations depending on the desired regularization. Any constraint can affect all rigid bodies from the initially affected body back to the root element. The most important constraint in the implementation of the Puppetmaster is the weld constraint that keeps two bodies in a fixed relation to each other. In the Puppetmaster, the sternum and the mandible are split to avoid any loop in the kinematic tree and comply with the requirements of Simbody. By welding the parts of these splits bones together, the bone is treated as one rigid object in the transformation. Typically, a weld constraint adds six constraint equations to the multi-body system [101].

To solve inverse kinematic problems as introduced in Section 2.4.5, Simbody employs an L-BFGS optimizer following the implementation by Liu and Nocedal capable of multiple end effectors [128]. This means the coordinates of several rigid bodies are input and all joint parameters are subject to optimization. In the Puppetmaster, three marker points on each bone determine its position and orientation in space. An input to the Puppetmaster can be an observation of where each of these points should be located in a new posture. The optimization in Simbody then solves the least squares problem to match the marker position $m_i(q)$ given generalized coordinates q of the multibody system to the observation o_i by

$$ \min_{q} \left(\sum_{i} w_i |o_i - m_i(q)|^2 \right) , \tag{3.7} $$

where w_i is the weighting for each pair of markers and observation. In the Puppetmaster implementation, all weights are equal. For a typical motion, Simbody can solve this optimization problem for the human anatomy in real-time as long as the motion is small. For large motion and conflicting inputs, the optimization can take seconds to compute.

3.2 Soft Tissue Propagation: The Chainmail

For the application of the previously described kinematic model (see Section 3.1) in the context of image registration and processing, it is not sufficient to consider only the transformation of the bones but the whole image space has to be transformed. This includes the soft tissue within the patient. In the presented work, an enhanced chainmail algorithm as developed by Bartelheimer et al. is employed [82, 100]. For readability, it is referred to as the Chainmail.

In the current implementation of the biomechanical model, the transformation of the image space is encoded in a so-called displacement vector field (DVF) that for each voxel contains a vector where this voxel is transformed to. For the voxel within bone tissue, this can be calculated explicitly from the rigid transformation per bone as found by the Puppetmaster (see Section 3.1). For all other voxels

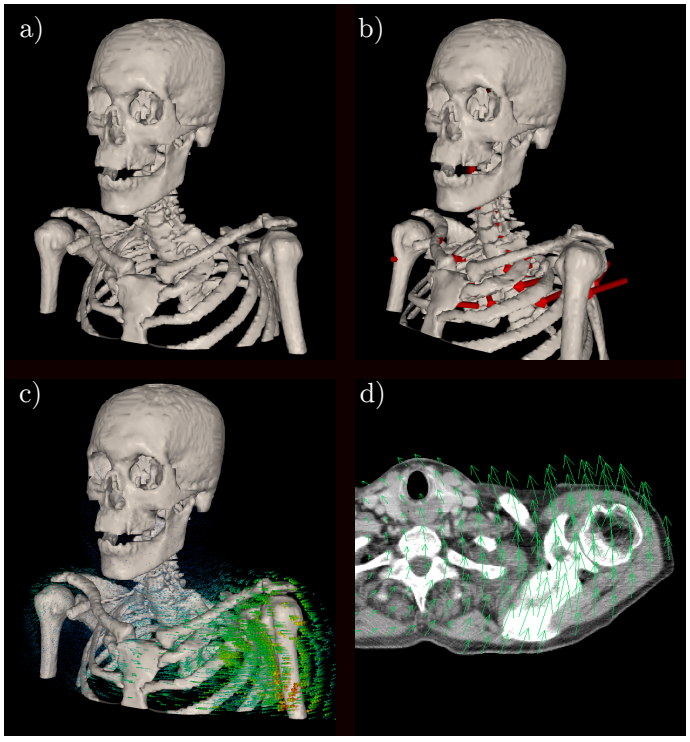

Figure 3.4: Illustration of the Chainmail generation of a displacement vector field (DVF). a) The Puppetmaster model b) Motion induced in the left shoulder. Red arrows indicate the transformation for each bone (exaggerated for visual clarity). c) 3-dimensional DVF as generated by the Chainmail algorithm. d) DVF resampled to the image. Displayed for a transversal slice of the left shoulder.

within the image, a soft tissue propagation is performed using the Chainmail algorithm. The bone transformation is considered the initiator of motion and all locally surrounding voxels are affected by the initial motion following the chainmail concept (see Section 2.4.4). The enhancement introduced by Bartelheimer is the potential of the Chainmail to include rotations in the elements. This increases the complexity of the approach but yields a more realistic description of potential motion in the human soft tissue.

Figure 3.4 illustrates the process of DVF generation as the output of the Chainmail. After the Puppetmaster has determined the bone transformation for a given posture, the Chainmail generates a 3D DVF that includes the soft tissue with motion propagation. Finally, this DVF is resampled to the discrete image cube to be applied as a voxel transformation.

An essential aspect of the Chainmail algorithm is the parametrization of the chain elements regarding their elasticity and shear parameters. The utilized Chainmail in this work takes advantage of the available Hounsfield units (HU) to generate a self-parametrization of the elastic parameters of the underlying tissue. For this purpose, the range of CT values is divided into three regimes: the fully elastic (e.g. air), the soft tissue, and the rigid tissue. Figure 3.5 shows the self-parametrization of the Chainmail in the three regimes.

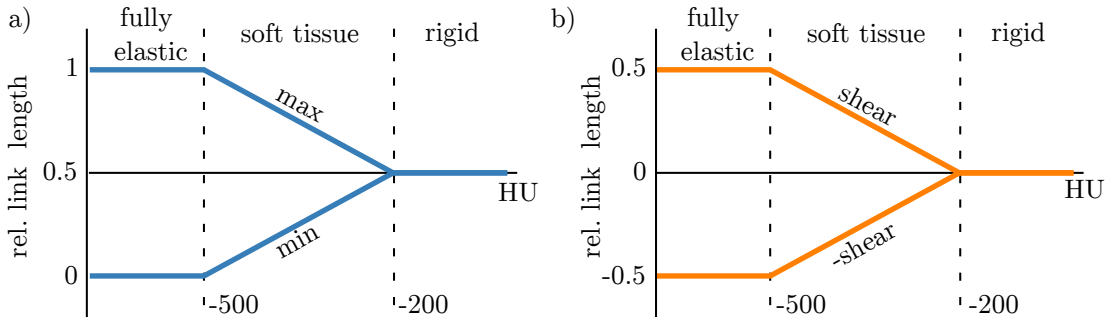

Figure 3.5: Self-parametrization of the Chainmail regarding elasticity and shear. a) The elastic relative link length can be 0% to 100% in the fully elastic regime and then decreases linearly in the soft tissue regime with increasing HU. In the rigid regime, the relative link length is fixed at 50%. b) Shear is set to be a maximum of ±50% in the elastic regime and scales linearly in the soft tissue regime to a final 0% in the rigid regime where no shear is present. Figure adapted from [82].

The elasticity is between 0 and 100% link length for the fully elastic regime spanning from -1024 HU to -500 HU. In the soft tissue regime from -500 HU to 200 HU, the elasticity linearly decreases to 50% link length where the rigid regime

starts. Here, the relative link length is fixed, since there is no elasticity in bone tissue. In the same regimes, the shear ranges from -50% to 50% in the fully elastic regime to 0% for rigid materials as they do not exhibit shear deformations. The decrease in the soft tissue region is again set to be linear. This is, of course, only a rough estimation of the elastic properties and cannot be seen as an accurate representation. The underlying assumption is once again that the skeletal motion is dominant in the head and neck region and that small discrepancies in the soft tissue model will have no relevant impact on the quality of radiation therapy.

3.3 Kinematic Deformable Image Registration

In the context of radiation therapy, and in particular in image-guided and adaptive radiation therapy, it is essential to be able to identify and quantify motion and anatomical deformations in the patient. For particle therapy, changes in the stopping power along the path of the particle beam can have a large impact on the dose deposition in the patient. For this purpose, an image registration scheme can be used. For each voxel in the image, the resulting DVF can – in principle – describe the motion and deformation. With the Puppetmaster and Chainmail model, a kinematic approach to image registration is possible.
This section describes the idea, implementation, and evaluation of the proposed kinematic deformable image registration, short *KinematicDIR*. Initially, the concept will be described for the monomodal case where a planning CT is registered to a fraction CT. The excellent bone contrast and comparable image quality between the two registered images allow for a proof-of-principle if KinematicDIR can provide an accurate registration that is robust for several patients and throughout the fractionated radiation therapy process. Then, the more realistic clinical scenario of multimodal images is considered. The current state of the art in most clinical settings is a cone beam CT that still shows correlation to the planning CT in the contrast but has a limited field of view and suffers from image artifacts. Finally, the CT-MR case is considered where the fraction image consists of a magnetic resonance image. Here, the limited bone contrast as well as the potential geometric deformation in the reconstructed MR image can be a cause for problems in the registration.

3.3.1 Monomodal KinematicDIR

The initial idea to use the Puppetmaster in monomodal CT-CT image registration was proposed by Teske and his section *Materials and Methods* should be referred to for deeper insight into the numerical aspects of KinematicDIR [101]. The enhancements and adaptations are explained in this section together with a general overview of the methodology.

KinematicDIR Pipeline

As described in Section 2.3, any registration scheme requires a transformation model, an optimizer, and a similarity metric. In KinematicDIR, the transformation model is the Puppetmaster that only considers bone tissue. Therefore, KinematicDIR includes a final step of generating a DVF for the full image space by applying chainmail-based motion propagation. This approach is visualized in the KinematicDIR pipeline in Figure 3.6. An initial pair of CT image scans (planning CT and fraction CT) that do not align in the color fusion is used as input to the Kinematic model built-up from the segmented planning CT (top left). The parameters of the kinematic model (translation and rotation of bones) are optimized to the maximum overlap of the bones in the model and the bone tissue as identified on the faction scan. From this transformation, the Chainmail algorithm is used to propagate the motion throughout the surrounding soft tissue and create a DVF. Finally, the planning CT is transformed with the DVF to align with the fraction image.

Transformation Model and Parametrization

The parametrization of bone positions in the Puppetmaster is done using up to three rotational parameters for the translation $T = [t_x, t_y, t_z]$ as well as the rotation $R = [r_x, r_y, r_z]$. Rotations are considered for all joints and are performed relative to the center of mass of the bone or relative to any joint directly connected to said bone. Table 3.1 shows the transformation parameters for all bones including rotation center and initiation range.

KinematicDIR Optimizer

The optimization of the translation and rotation parameters T and R in the transformation model is done using a hierarchical optimization scheme and a downhill simplex optimization. The simplex optimizer works on one bone at a time finding the best position and orientation for it. Following the kinematic constraints, this affects the whole skeleton. For the simplex optimization, each vertex (representing a set of parameters T and R) adopts the previous transformation changed by a random offset in the limits listed in Table 3.1. This offset is sampled from a uniform distribution. After this initialization, the simplex downhill optimization

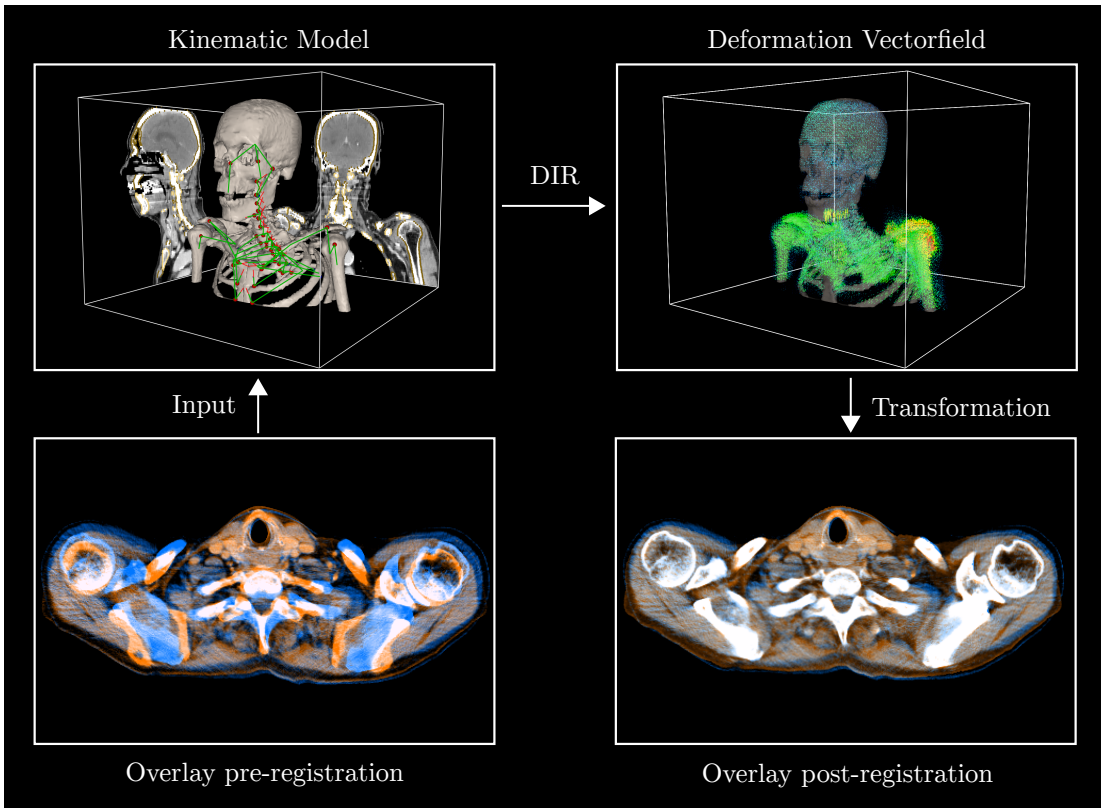

Figure 3.6: KinematicDIR pipeline. An initial pair of CT image scans (planning CT and fraction CT) that do not align in the color fusion are used as input to the Kinematic model built-up from the segmented planning CT (top left). The parameters of the kinematic model (translation and rotation of bones) are optimized to maximize the overlap of the bones in the model and the bone tissue as identified on the faction scan. From this transformation, the Chainmail algorithm is used to propagate the motion throughout the surrounding soft tissue and create a deformation vector field. Finally, the planning CT is transformed to overlay align with the fraction CT.

as introduced by Nelder and Mead is performed using the originally proposed simplex coefficients [101, 129].

To extend the optimization algorithm, the transformation model is explicitly used. After each iteration of updated transformation parameters, these are transferred to the Puppetmaster model as input and the kinematic model is solved by Simbody (see Section 3.1). The actually achieved transformation by the Puppetmaster following all kinematic constraints is then converted back to the simplex optimization replacing the originally found parameters. This guarantees that the optimization only happens within the kinematically feasible regime of

Table 3.1: Optimization parameters for KinematicDIR. For each bone, either only rotation (R) or both translation and rotation (TR) are parameters of the transformation model that are optimized. The rotation center is located in the center of mass of the bone (C) or the rotation center of the joint connected to the child bone (J_c) or connected to the parent bone (J_p). The initial range for each bone regarding translation and rotation is given to the optimizer and then sampled randomly within the interval.

Parameters	**Skull**	**Scapula**	**Vertebra**	**Sternum**	**Humerus**	**Rib**	**Mandible**	**Clavicle**	**Hyoid**
Transf.	TR	TR	TR	TR	R	R	R	R	TR
Rot. center	C	C	C	C	J_c	J_c	J_c	J_p	C
Init T [mm]	±3.0	±10.0	±3.0	±5.0	–	–	–	–	–
Init R [°deg]	±0.9	±3.0	±0.9	±1.5	±5.0	±3.0	±3.0	±3.0	±1.5

transformation. Then, the objective value for the transformed model is assigned to the corresponding vertex.

The original simplex downhill algorithm uses a heuristic search method to achieve the optimization of several parameters in non-linear problems. However, it is known to converge towards local minima. While the existence of such minima can depend on the objective function and the registration context, it has to be a considered possibility since some of the bones have intrinsic symmetries that can lead to local minima. This is prevented by the introduction of a disturbance step when an initial convergence is reached. In each disturbance step an increasing random offset is added to the parameters to improve the chances of finding the globally optimal solution. The final convergence is accepted when a fixed number of inefficient disturbance steps is detected or the maximum number of iterations is reached.

To optimize not just a single bone but also the full posture of the patient, the previously described downhill simplex optimizer is applied to all bones of the skeleton successively following a predefined hierarchy. In the implementation of KinematicDIR, the parameters of a bone are fixed once it is optimized. Hence, for each following bone, there is a reduced number of degrees of freedom and additional constraints regarding the permissible range of motion. This further improves the kinematic bio-fidelity of the optimized posture. For the application

of KinematicDIR in the head and neck region, the skull is used as the first element since it is typically fixated well and therefore can be assumed to be already in a position and orientation close to its optimum. The registration of the root element yields a rigid transformation for the whole skeleton that is used as a first estimate of the posture. From this, the hierarchy follows the kinematic tree with one exception in case the sternum is not visible in the fraction field of view. If that is the case, the lowest fully visible vertebral body is optimized early on after the clavicles and scapulae (i.e. when the rough positioning is correct) and keeps the spine and therefore, the rib cage in a more constrained position. This empirically improved the stability of the registration in this work and was necessary for several patient data sets without a visible sternum in the fraction data.

Similarity Metric

The monomodal case of KinematicDIR enables the use of an image-based similarity metric. For this, a binary target image is generated from the fraction CT by considering all voxel in the interval $[120, 2000]$ HU as bone tissue. In each step of the optimization, the overlap of the bones in the Puppetmaster and the binary target image is considered the objective value of the similarity metric. A representation of this binary target image is displayed in Figure 3.7. It indicates that a simple threshold does not extract the skeleton completely and also includes certain image artifacts (e.g. around the mandible and the catheter visible at the left chest). The underlying assumption of this approach is that the threshold image contains sufficient information on the actual bone location that the Puppetmaster model can identify the individual bones as extreme points in the overlap and hence optimize into this position.

Evaluation Data Set

To evaluate the performance of KinematicDIR regarding accuracy and robustness, the image data sets of three head and neck cancer patients were used. All patients underwent postoperative fractionated radiotherapy in 33 fractions (and up to 36 fraction CT scans). They were selected at random from a published data cohort including stereotactic fixation [130, 131]. Written informed consent to use their data was obtained from all patients. For each patient, there was a planning CT as well as daily kV-control CTs available. The authors of this study report

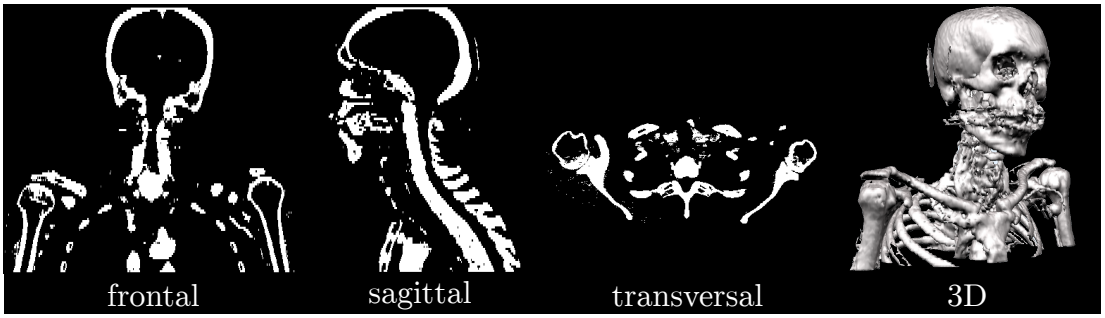

Figure 3.7: Binary target image for CT-CT registration. The fraction CT is threshold in the range $[120, 2000]$ HU to generate a binary target image. The 3D rendering shows the artifacts that are visible from the mandible and remaining artifacts from the table and positioning device.

inter-fractional anatomical deformations between 0 and 9 mm.

A total of 18 fractions (6 per patient) were used to evaluate KinematicDIR in the monomodal CT-CT case. They were distributed throughout the treatment course to cover potential anatomical changes during treatment. All images scans were acquired with a pixel size of $0.98 \times 0.98\,\mathrm{mm}^2$ and a slice thickness of 3 mm for patient 1_{CT} and 2_{CT} and 2 mm for patient 3_{CT}.[3] Planning CT scans were acquired by a Toshiba Aquilon scanner (Toshiba, Otawara, Japan) and the fraction CT scans by a Siemens Primatom in-room single-slice spiral scanner (Siemens OCS, Malvern, PA). In all fraction CT scans, a stereotactic frame registration was applied to establish their spatial alignment with the planning CT scans utilizing a stereotactic frame described earlier [130].

The quantification of the performance of KinematicDIR is achieved using a landmark-based evaluation. For this purpose, visibly identifiable points (landmarks) are localized on the planning CT and each of the considered fraction CTs. Patient 1_{CT} was used for the assessment of the absolute accuracy. Therefore, the large number of 161 corresponding landmark pairs is used. Outlier detection and a rigidity condition (< 3 mm violation) are applied. The inter-observer variability of the landmark identification was assessed on two out of six fractions by four independent observers and ranged from $0.1 - 2.9$ mm.

Patient 2_{CT} and 3_{CT} were used for the investigation of the robustness of KinematicDIR and in particular the question of whether different patient geometries can be registered. For those patients, $63 - 70$ landmarks were identified on 6 fractions for

[3] The notation of patient data sets is used to distinguish them from the subsequent multimodal data sets.

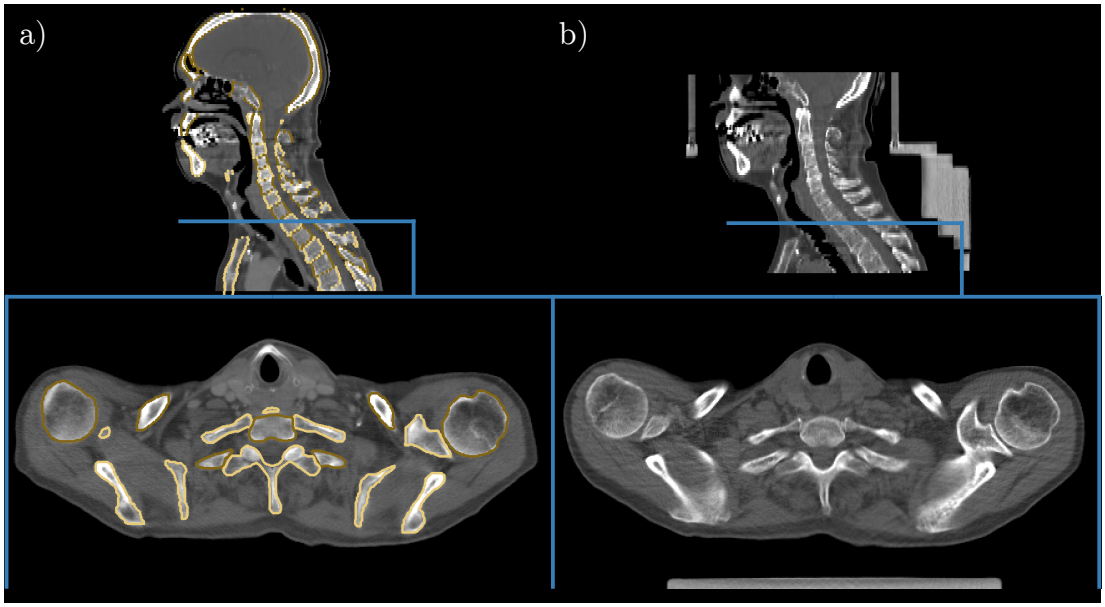

Figure 3.8: Image quality of CT slices for patient 1_{CT}. a) A sagittal and transversal slice of the planning CT indicating the measured field-of-view and the used manual bone segmentations. The blue line indicates the position of the transversal slice. b) The stereotactically aligned corresponding slice of a fraction CT (F01). The chosen fraction scan is representative of the imaging quality of all fraction scans, facilitating meaningful accuracy of manually identified correspondence points.

each of these two patients. Figure 3.9 shows an overlay of the three-dimensional bone rendering and the landmarks identified on the skeleton for patient 1_{CT}. The landmarks were distributed on the skeleton. There were no landmarks positioned at the upper part of the skull since it is not visible in any of the fraction images and therefore no counterpart would have been identifiable.

Evaluation Metrics

In the evaluation, the landmarks were used to calculate the target registration error (TRE) as the Euclidean distance of the transformed landmark on the planning CT L_{plan} and the landmark on the fraction CT L_{fraction}. Given the transformation $\mathscr{T}$ that connects the two images, the TRE is defined as

$$\mathrm{TRE}(L_{\mathrm{plan}}, L_{\mathrm{fraction}}, \mathscr{T}) = \left\| \mathscr{T} * L_{\mathrm{plan}} - L_{\mathrm{fraction}} \right\| . \tag{3.8}$$

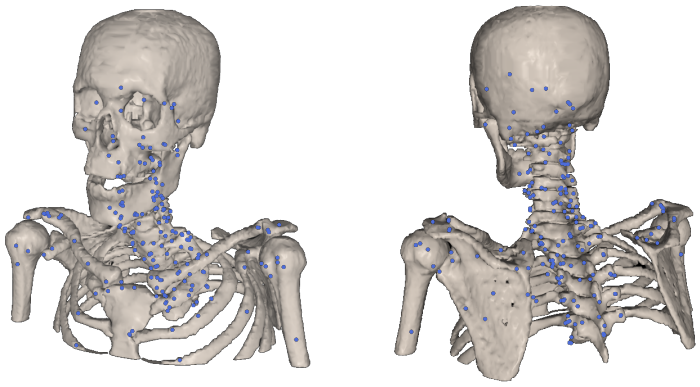

Figure 3.9: Overlay of the three-dimensional bone rendering and the 161 landmarks for patient 1_{CT}. Landmarks are positioned uniformly on the bones within the boundaries fo the field of view of the fraction image data. Landmarks are positioned one the bone surface and inside of bones.

A lower TRE means the alignment of both images can be considered to be superior locally at the landmark. Since there were a large number of landmarks considered, a statistical approach was used in the evaluation and visualization of the results in the TRE analysis. In this thesis, the distribution of TRE is shown in box plots using the median and a box size from the first quartile (Q_1) to the third quartile (Q_3). The distribution is then also classified using the *inter-quartile range* (IQR) defined as

$$\mathrm{IQR} = Q_3 - Q_1 \,. \tag{3.9}$$

Assuming a normal distribution of the data, the IQR would be equal to approximately 1.35σ. The IQR can hence be used as a measure of the width of the distribution even in skewed and asymmetric conditions. As a second indicator of the width of the distributions, whiskers are used. Due to the relatively broad distributions observed in this thesis, the whisker length of the box plot was chosen to be $3 \cdot \mathrm{IQR}$ meaning all values outside the range were considered outliers and marked as individual data points. Figure 3.10 displays a schematic box plot to illustrate the visualization that is most prominently used in this thesis.
To analyze the robustness of the registration without having landmarks available for all fractions, a second analysis based on the Dice similarity coefficient (DSC)

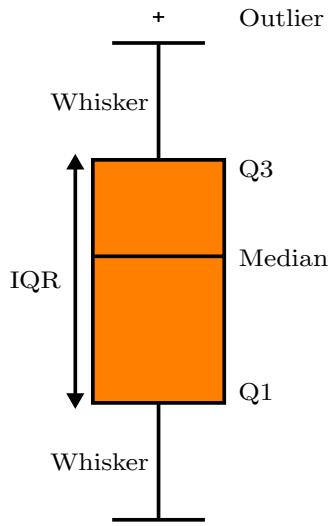

Figure 3.10: Illustration of a box plot. The colored box ranges from Q_1 to Q_3 with an explicit line marking the median. The length of the box is called the inter quartile range (IQR) The whiskers cover all data points within $3 \cdot \text{IQR}$. Any point outside the whiskers is considered an outlier and marked with a black cross.

was used [132]. For two sets A and B, the DSC is defined as

$$DSC(A, B) = \frac{2|A \cap B|}{|A| + |B|}, \tag{3.10}$$

where $|\cdot|$ denotes the cardinality of a set.[4] The DSC ranging from 0 to 1 indicates the relative overlap of the two sets.

For the analysis of KinematicDIR with the DSC, a segmentation of the skeleton was required on the planning as well as the fraction CT. In the presented experiment, a fast segmentation was generated by considering all voxels above 120 HU as skeleton, and correcting for the most prominent metal implant artifacts manually. After the application of KinematicDIR, the contours of the skeleton segmentation was then transformed with the resulting DVF and compared with the skeleton segmentation on the fraction CT using the DSC.

[4]Regarding a segmented image, the cardinality would be the number of voxels within the segmentations

Comparison with Intensity-Based Registration

While a fair comparison of two registration algorithms can be challenging, a first experiment to compare KinematicDIR with intensity-based registration approaches was performed in this work. For this purpose, the *Plastimatch* algorithm within the *SlicerRT* [133] framework embedded in *3DSlicer* [134] was used as it provided an out-of-the-box registration and is particularly developed for the application in (adaptive) radiation therapy workflows. A multi-stage B-spline-based registration is used in this framework. In each stage, the image is sub-sampled to reduce the resolution. Then control points are distributed in the image space following the grid size parameter. To provide the best possible chance for the intensity-based algorithm, a 3-stage registration was performed with subsequently reducing voxel size and grid spacing. Since the default parameter choice within 3DSlicer appeared optimized for computational speed, this 3-stage registration was performed once with the default parameters and once with a much finer grid of control points. The parameter choice is shown in Table 3.2.
It can be expected that the finer grid provides a more accurate registration of the landmarks. However, a stronger deformation of the image and in particular the bones can be the result. Additionally, the fine grid parameters increased the computation time to minutes even with GPU acceleration.

Table 3.2: Plastimatch parameters used for intensity-based DIR. The default settings yield a fast registration. To achieve the best possible accuracy a second parameter set of finer grid size is used.

Parameters	default			fine grid		
	Stage 1	Stage 2	Stage 3	Stage 1	Stage 2	Stage 3
sub-sampling	4,4,2	2,2,1	2,2,1	4,4,2	2,2,1	1,1,1
grid size [mm]	100	50	50	50	25	5
max iterations	50	50	50	50	50	150

To compare the performance of both algorithms, the data set of patient 1_{CT} was used in alignment with the KinematicDIR accuracy evaluation. For all fractions with available landmark data, KinematicDIR, as well as Plastimatch registration, were performed (once with default parameters, once with the fine grid). The TRE was calculated and used as a quantification of the registration accuracy.

3.3.2 Multimodal KinematicDIR with Cone Beam CT

In the clinical context, one of the most common imaging modalities available for image-guided or adaptive radiation therapy is cone beam CT (CBCT), since it offers a fast 3D image acquisition (see Section 2.2.3). To facilitate the use of KinematicDIR in a clinical scenario, it is important that the registration pipeline as introduced in Section 3.3.1 can be adapted and used to perform the registration of a planning CT to CBCTs acquired at each session of fractionated radiation therapy.
The principle KinematicDIR pipeline remained unchanged for this experiment. This included the transformation model, optimizer, and parameters introduced earlier as well as the used evaluation metrics. A different approach, however, was required regarding the similarity metric.

Evaluation Data Set for CT-CBCT Registration

To evaluate the performance of KinematicDIR in the context of CT-CBCT registration, one patient data set was used consisting of a planning CT and three fraction CBCTs. It was selected from a previously published study [135, 136]. The planning CT was acquired with an in-plane resolution of $0.98 \times 0.98\,\mathrm{mm}^2$ and a slice thickness of 3 mm. The CBCT was acquired with an isotropic resolution of $1.0 \times 1.0 \times 1.0\,\mathrm{mm}^3$. The CBCT image was interpolated and re-sampled to the resolution of the planning CT and a rigid manual alignment by visual assessment was used to achieve an approximate alignment as well as a transfer into the same image coordinate system. Only one data set was used since the scapula – an essential bone for the determination of the posture – was not visible for any other data set within this cohort. On the planning CT, manual bone segmentations were created by a human observer facilitating the build-up of the Puppetmaster. For the calculation of the TRE, 61 landmarks on the skeleton were positioned on the planning CT and all fraction CBCTs.

Similarity Metric for CT-CBCT registration

A typical cone beam CT has no reliable Hounsfield unit contrast and therefore the detection of bone tissue can not be performed using a threshold. In particular, insufficiently corrected cupping artifacts contribute to this effect in the center

of the image (e.g. at the vertebral bodies) [137]. To provide a proof-of-principle for the CT-CBCT registration, a semi-automatic segmentation of bone tissue on CBCTs was used. After initial thresholding to find most of the bone tissue, a human observer adds the rough shape of the bone tissue that was not recognized and cuts the most obvious artifacts using a scissor tool. This segmentation took minutes and should not be seen as a fully manual segmentation of individual bones as is required for the model build-up. Figure 3.11 shows the quality of the evaluation data set including the semi-automatic segmentation of the complete skeleton (orange) for the cone beam CT. The inaccuracy of the segmentation and the missing separation of individual bones was the result of the fast semi-automatic approach.

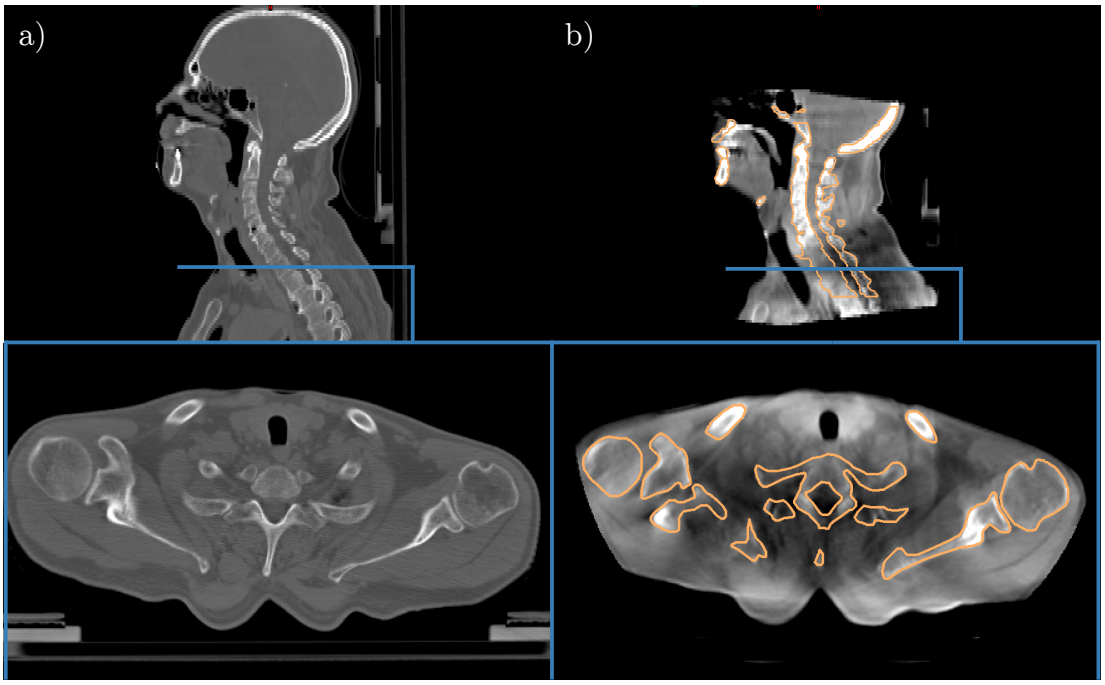

Figure 3.11: Image quality of CT and MRI slices for patient 1_{CBCT}. a) Sagittal and transversal slice for the CT indicating the measured field-of-view. The blue line indicates the position of the transversal slice. b) Cone beam CT with a limited field of view and semi-automatic skeleton segmentation (orange). The transversal slice shows that the semi-automatic segmentation does not reach the quality of a manual segmentation.

3.3.3 Multimodal KinematicDIR with MR Images

Registration of CT and MR images is a true multimodal image registration task since CT and MRI do not share a common contrast. As an example, bone and air are at the two ends of the HU spectrum on CT but both yield a similarly low MR signal. To enable the registration of CT to MRI in this thesis, an object-based approach was chosen. Instead of registering two images to each other, the whole skeleton was chosen as the object for which the overlap should be maximized. While this is the same goal as for the monomodal CT-CT registration, the process and implementation differ.

Evaluation Data Set for CT-MRI Registration

For the multimodal image registration task, three patient data sets were used. For each patient, a planning CT (Siemens Somatom Confidence) was acquired with a pixel size of $0.98 \times 0.98\,\mathrm{mm}^2$ and a slice thickness of 3 mm. On the same day, each patient received an MRI scan (Siemens Magnetom Sola: Vibe Dixon) with a pixel size of $1.30 \times 1.30\,\mathrm{mm}^2$ and a slice thickness of 3 mm. The in-phase image (see Section 2.2.4) was used since the characteristic dark band around the bones could provide a contrast to delineate bone tissue. The MRI data was interpolated and re-sampled to the resolution of the planning CT and a rigid manual alignment by visual assessment was used to achieve an approximate alignment as well as a transfer into the same image coordinate system.
To enable the quantification of the registration quality of KinematicDIR, 50 landmark pairs on the skeleton were positioned in the CT and the MRI.

Similarity metric and Bone Identification on the MRI

To identify the skeleton, segmentations of individual bones were created for both the CT and the MRI by manual delineation. Then, the unity of all individual bones of the MRI segmentations was considered as the skeleton in the MRI. From this segmentation, a binary *target cube* was created with all voxels inside the segmentation set to 1 and all other voxels set to 0. In the KinematicDIR pipeline, the similarity metric was calculated as the overlap of the skeleton model in the Puppetmaster and the binary target cube, comparable to the monomodal

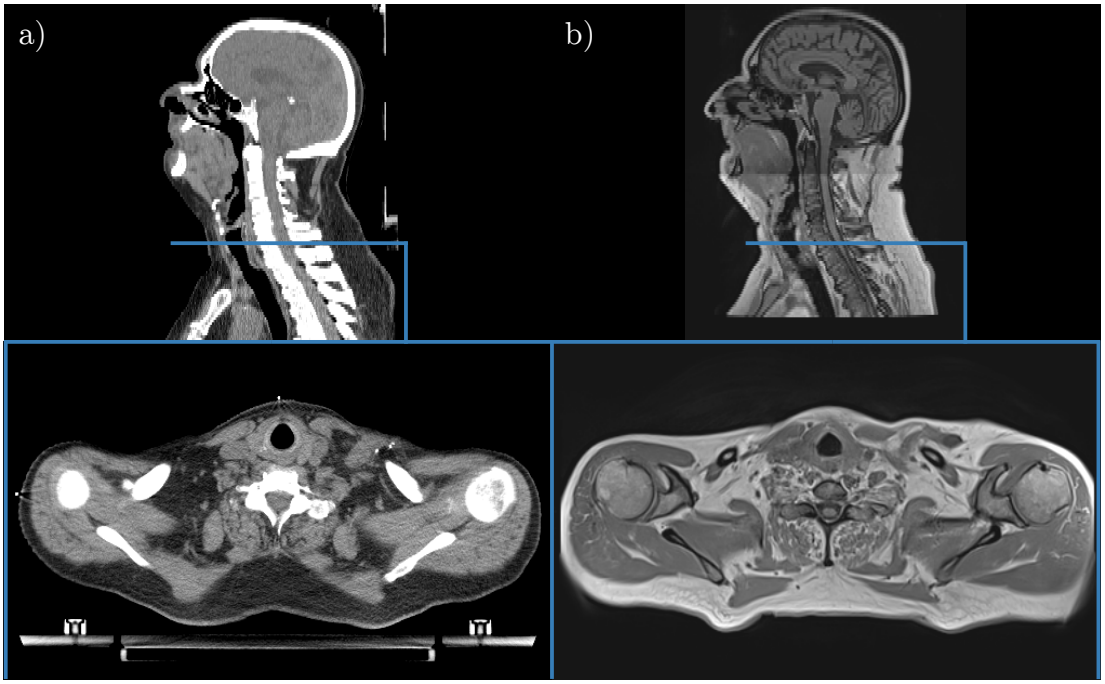

Figure 3.12: Image quality of CT and MRI slices for patient 1_{MRI}. a) Sagittal and transversal slice for the CT indicating the measured field-of-view. The blue line indicates the position of the transversal slice. b) Manually aligned and resampled corresponding slices of the MRI.

case. This approach separates the two problems of bone identification and object registration. The experiment investigated whether KinematicDIR is capable of accurately registering CT-MRI data if bone tissue is identified.

Investigation of the MRI segmentations

The identification of the skeleton on the MRI utilizing manual segmentations can be limited by the capability of the human observer to discern the bones given the limited bone contrast. To investigate the effect of the MRI image on the segmentations, for each patient data set, the volumetric ratio is calculated for each bone between the CT and the MR image. This gives a first approximation of how comparable the segmentations are. Ideally, the segmented volume would be equal regardless of the modality. To include the geometrical shape in the analysis, the standard deviation of the segmented bone in each of the image coordinates is considered for both images. This is a first-order measure for the geometrical shape of the segmented object. For each patient data set, the standard deviation for each bone was compared to yield an estimate of the similarity of the segmentations.

3.4 Bionic Augmentation

A second major application for the Puppetmaster model is based on the potential of biomechanical models to perform so-called forward calculations. This means that given an input motion or posture, the Puppetmaster can calculate the position for each bone and joint following the underlying kinematics and the Simbody optimizer. Then the Chainmail propagates the motion into the remaining image space to generate a DVF that represents the input motion. Finally, by transforming the respective planning CT and labels, a new fully segmented data set is created. The synthetic data can then be used as additional training data for a deep learning-based segmentation approach. Increasing the training data in such a scenario is called data augmentation. Since the augmentation data arises from the biomechanical model and is therefore inspired by human motion, the presented and investigated approach is called *Bionic Augmentation*.

3.4.1 Generation of Synthetic Image Data

To generate new data with the Puppetmaster and Chainmail pipeline, the model needs to be built-up from a segmented CT as described in Section 3.1. The forward calculation is performed, when a new posture is given as an input to the model. In the current implementation used in this thesis, this can be done in two ways.

As a first option, by defining new positions for all bones using the coordinate marker points for each bone, the Puppetmaster tries to solve the inverse kinematic problem to generate the posture that is as close as possible to the input without violating any kinematic constraints (see equation (3.7)). The second option is an interactive input. In the graphical user interface provided by the Puppetmaster, a pre-defined or individually placed point on the skeleton can be dragged to a new location in space. This yields an observation where the marker points on this bone have to move. The Puppetmaster can solve the inverse kinematics in real-time to adapt to the new posture. A combination of several manipulated points can be used to create arbitrary new postures.

In the prototype application of Bionic Augmentation, the second approach was used. While it is more labor intensive, the generation of postures via interactive human input has several advantages:

- The human input guarantees a certain degree of randomness without generating unrealistic postures.
- The interactive approach enables the user to visually access the generated posture during the process and ensure the resulting data will be consistent with the human motion range.
- Using the knowledge of CT systems and positioning devices, the human can choose postures that can realistically be found in CT scans. This means the underlying distribution that can be expected in any prediction data can be covered more closely.

3.4.2 Training of U-net-based Automatic Segmentation

To evaluate the feasibility and quantify the benefit of Bionic Augmentation, two experiments were considered. For both experiments, an artificial neural network for the semantic segmentation of individual bones was trained using 10 manually segmented patient data sets of the head and neck region. For each patient, the individual bones were segmented from the skull to the fifth thoracic vertebra (T5). For five of these patients, the Puppetmaster model was created. In the prototype for the proof-of-principle of Bionic Augmentation, an interactive human input was used to generate two realistic yet random postures for each of these patients. Each of these postures was checked to guarantee a realistic, overlap-free result. In total, 10 additional data sets were created in this way.
For the first experiment, a U-net [116] without any additional adaptations is trained once with the 20 data sets resulting from Bionic Augmentation (10 real data sets + 10 augmented). As a comparison, the same network is trained using standard augmentation based on image rotation. The 10 real data sets are augmented with 10 rotated image data sets with an angle of $\pm 2°$.
In the second experiment, the nnU-net was used as a state-of-the-art segmentation framework [138, 139]. The in-built augmentations that enhance the training data set with a large amount of artificial data sets was used for both the Bionic Augmentation training and the standard augmented training.

3.4.3 Analysis of Bionic Augmentation

To analyze the performance of the Bionic Augmentation approach, a DSC-based evaluation of the prediction quality is used. After training the network, the prediction on previously unknown patient data sets is used to calculate the DSC between the prediction and the manually segmented ground truth for each individual bone. For the U-net, one test data set is used, for the nnU-net, two test data sets.
To compare the performance of standard augmentation and bionic augmentation, the DSC is compared per individual bone. An equal DSC means both approaches have the same performance and a higher DSC coefficient indicates better quality of the prediction. Figure 3.13 illustrated the evaluation pipeline for Bionic Augmentation in the case of the U-net. The original training data is augmented with the same number of Bionic Augmentation data sets and standard augmentation data set. Two separate instances of the U-net are trained and the DSC between the ground truth and the prediction is calculated for the test data. A higher DSC means better prediction. To compare the performance, the corresponding DSC values are compared in a scatter plot.

3.4.4 Generalization of Postures

The approach to generate synthetic data in the prototype application of Bionic Augmentation relies on interactive human input and visual assessment of the resulting posture. To enable the generation of far more artificial data sets – in the order of hundreds to thousands – is essential to scale the process of posture generation and to be able to augment large data sets.
One approach investigated in this thesis is the generalization of postures in the Bionic Augmentation approach. This means a posture is created and curated on one individual patient's anatomy. It is then transferred and applied to all other patients in the training cohort. This theoretically yields a factor of N in the total augmented data given the number of patients N that are used. In a more general sense, this approach would also allow the generation of a library of large numbers of pre-computed postures that can be employed for all further augmentations regardless of the available training data set or the specific application.
The application of generalized postures can lead to an overlap of bones since the current version of Puppetmaster does not include collision detection and the resampling of the transformed contours is not ideal on a relatively large

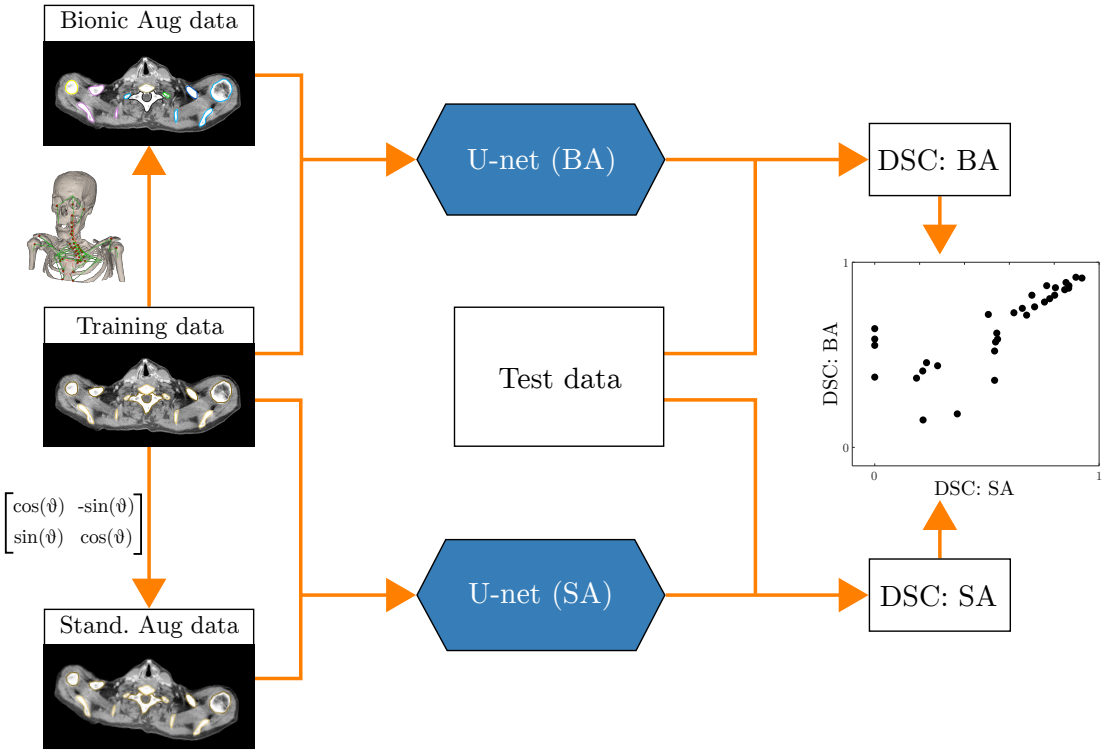

Figure 3.13: Evaluation Pipeline for Bionic Augmentation (BA) for the U-net. The original training data is augmented with the same number of Bionic Augmentation data sets and standard augmentation (SA) data sets. Then, two separate U-nets are trained. For each U-net the DSC between prediction and ground truth is calculated. A scatter plot shows the comparison of the performance.

slice thickness in most medical images. This overlap introduces a biomechanical inconsistency since the bones in the human body do not overlap in reality. To investigate how large this overlap effect is, the previously used 2 postures on 5 patients at applied to all other patients resulting in 50 data sets in total. Then the overlap of any two neighboring bones is quantified using the DSC. This yields the inconsistency metric for the generalized posture.

3.5 Model Build-up from Automatic Segmentations

One important aspect regarding the application of the Biomechanical model composed of the Puppetmaster and Chainmail in an automated image registration approach, is the model build-up. In the prototype version currently used for KinematicDIR, this is performed using labor-intensive manual segmentations. This cannot be expected in a clinical routine.
Therefore, the potential to use automatic image segmentations of the individual bones using an artificial neural network for semantic segmentation is examined. For this, a two-step investigation is performed.

3.5.1 Baseline: Segmentations by Human Observers

First, the baseline of human observers is analyzed to estimate the quality that is required from any automatic approach. For this purpose, the data set of patient 1_{CT} as used in Section 3.3.1 is segmented by a second independent observer using the same rules (e.g. inclusion of the costal cartilage, same CT windows, and same software for segmentation [140, 141]). Consequently, both sets of segmentations are compared in a geometrical analysis: For each individual bone, the DSC is calculated. This is the most common evaluation done for the deep learning-based approaches and allows a comparison of how well human segmentations align.
For a better understanding of the geometrical differences, the Hausdorff Distance is calculated [142]. For two sets X and Y in a metric space with distance metric d, the Hausdorff distance d_H is defined as

$$d_H = max \left\{ \sup_{x \in X} d(x, Y), \sup_{y \in Y} d(X, y) \right\} , \tag{3.11}$$

where $d(x, Y) = \inf_{y \in Y} d(x, y)$ denotes the distance of a point x to the set Y and vice versa [143]. In addition, the mean distance d_{mean} is calculated to mitigate the effect of single points in the segmentations.
Within the build-up of the Puppetmaster, the segmentations of individual bones are essential for the positioning of joints since they are either determined using a nearest-neighbor approach or a geometrical rule-based system, depending on the joint type. To investigate the effect of the different human observers on the joint

positioning, the distance between the joints with the different segmentations is used. Since the resolution of the image data set is not isotropic, the joint distance is considered in each of the image coordinates (denoted x and y in plane and z along the longitudinal axis).
Finally, to investigate the effect these different segmentations have for the application of the Puppetmaster as a transformation model in the image registration task, the registration of planning CT and fraction CT was performed for all six fractions with the KinematicDIR pipeline and a landmark-based TRE analysis (as described in Section 3.3.1) was performed for both human observers. This gave a baseline on how robust the model will behave for different segmentations and which quality is required for the automatic segmentations. Figure 3.14 illustrates the evaluation pipeline of the three-stage evaluation of how different segmentations affect the Puppetmaster.

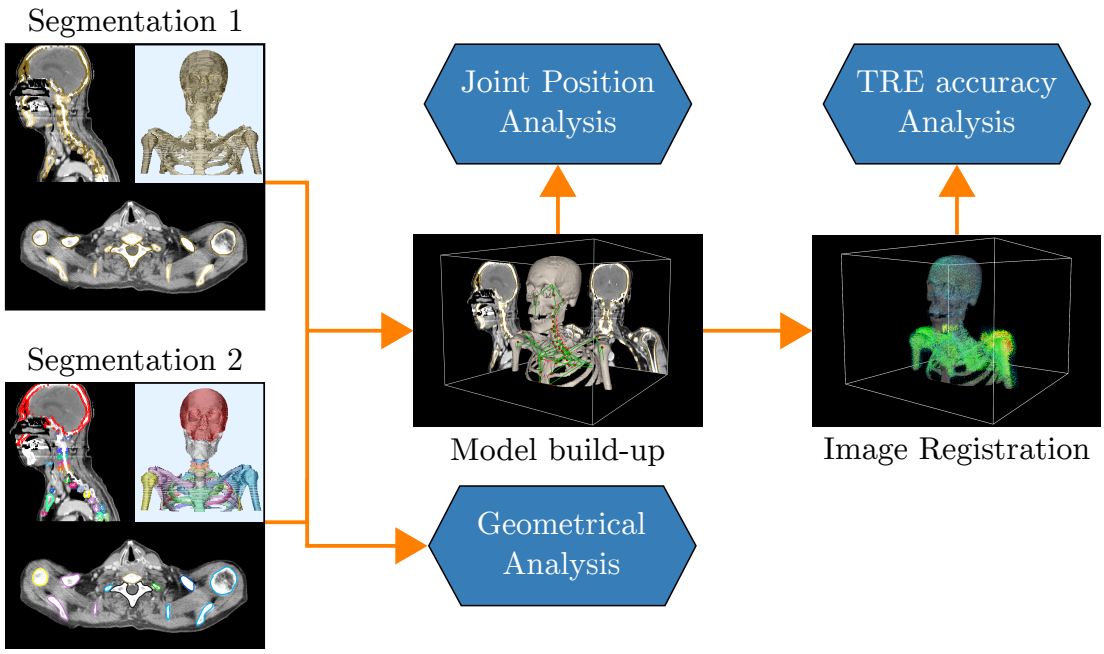

Figure 3.14: Evaluation Pipeline for different image segmentations. Both segmentations are compared using a geometrical analysis including DSC and Hausdorff distance. From both segmentations, the Puppetmaster is built-up to compare the joint position between the segmentations. Finally, the image registration is performed and evaluated using the landmark-based TRE accuracy analysis.

3.5.2 Automatic Segmentations by a Generalized U-net

For the segmentations of individual bones in a CT for the build-up of the Puppetmaster, it is essential to distinguish individual bones of the whole head and neck region without overlap. The *TotalSegmentator* toolkit [144] as available in Slicer3D [134] offers an automatic segmentation of up to 104 anatomical structures on CT data sets. This set includes most bones necessary for the Puppetmaster but lacks the skull, mandible, sternum, and hyoid. It is based on the nnU-net [138,139] and is trained on more than 1200 annotated patient data sets.
To investigate the potential of such an automatic segmentation for the model build-up of the Puppetmaster, the planning CT of patient 1_{CT} was used in analogy to the comparison of human observers in Section 3.5.1. The bones not available in the TotalSegmentator were left unchanged and for all other bones, the auto-segmentations provided by this toolkit are used. These automatic segmentations were compared to the manual segmentations of the first human observer regarding the geometrical properties and the joint positioning in the Puppetmaster in analogy to Section 3.5.1.
To investigate if the automatic segmentations by the TotalSegmentator are of sufficient quality for the KinematicDIR pipeline, the registration quality was assessed on the 6 fractions CTs with available landmarks for patient 1_{CT}. Overall, the analysis followed Figure 3.14.

4 Results

The results of this thesis can be separated into three major aspects. First, the performance of KinematicDIR is evaluated for monomodal as well as multimodal image registration. Second, the proof-of-principle for Bionic Augmentation is shown and a method to scale the augmentation approach is examined. Finally, the potential of automatic segmentations for the model build-up is investigated by systematic comparison of different human and automatic segmentations.

4.1 Monomodal Image Registration on Computed Tomography Images

The evaluation of the performance of the KinematicDIR pipeline composed of the Puppetmaster & Chainmail in monomodal image registration was performed on three patients with one planning CT and six fraction CTs distributed along the treatment (see Section 3.3.1 *Evaluation Metrics*). Parts of this study have been submitted for publication.

4.1.1 Visual Evaluation

For visual evaluation of the registration results, a color fusion with complementary colors (see Section 2.2.1) of the planning CT and the fraction CT F01 (i.e. the CT scan on the day of the first fraction) before and after registration is shown in Figure 4.1. As the first observation, the color fusion indicates that with stereotactic alignment, there are still large deformations remaining between the images. While the skull region is aligned well due to the stereotactic frame, there is apparent motion in the spine and shoulder region. This motion can be seen in all slice orientations.
Second, after the application of KinematicDIR, excellent alignment of the bone tissue can be seen (green arrows). There are residual deformations in the soft tissue, particularly in regions far away from any bone that do not align well (red arrows). This indicates that the optimization within the kinematic tree was successful,

while the soft tissue propagation as a simplified approach cannot achieve perfect alignment of the images when it comes to distant soft tissue.
The blue and orange areas in the frontal and sagittal slice arise from the limited field of view of the fraction CT, and the treatment couch and frame, respectively. They are not related to the registration accuracy of KinematicDIR.

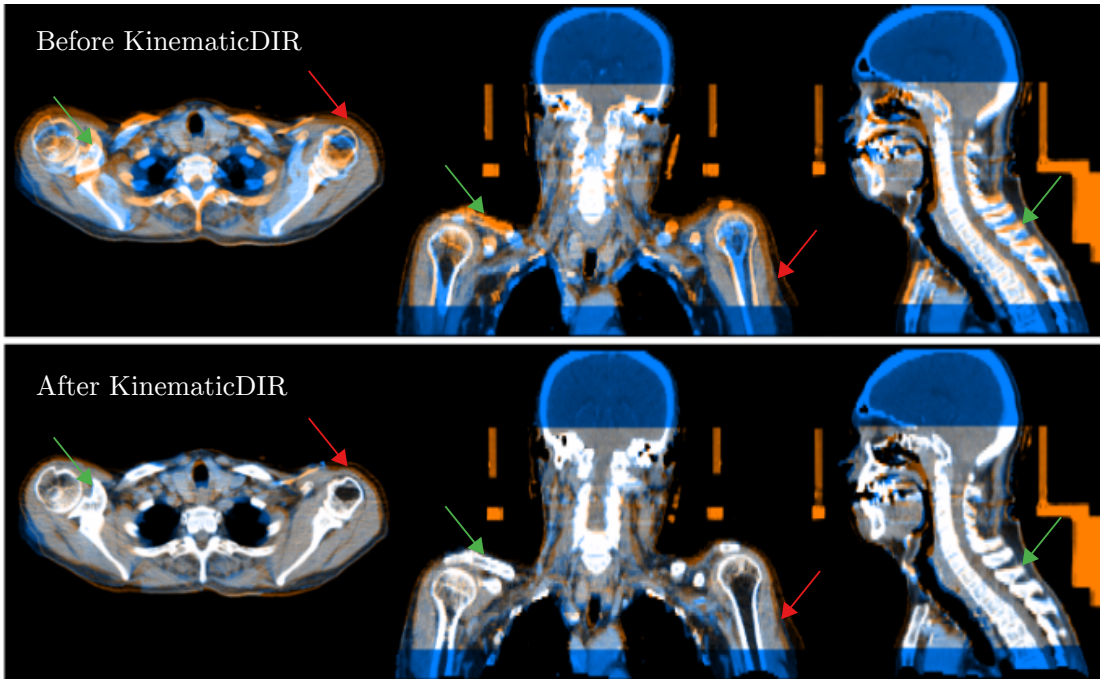

Figure 4.1: Visual evaluation of monomodal image registration with KinematicDIR. The upper row shows the transversal, frontal and sagittal slice of a color fusion before deformable image registration. planning CT (blue) and fraction CT (orange). The lower row displays the fusion after KinematicDIR. Green arrows indicate locations of excellent registration of bone tissue. Red arrows show residual soft tissue deformations.

4.1.2 Accuracy Evaluation

To quantify the accuracy of the registration provided by KinematicDIR, the TRE (see Section 3.3.1 *Evaluation Metrics*) is calculated for all landmarks of patient 1_{CT}. The distribution of TRE for all considered fractions before and after the application of KinematicDIR is shown in Figure 4.2. Before the registration, when the patient is fixated in a stereotactic frame, there is still large-scale motion in the image as was observed in the visual assessment. This translates to a median TRE between 4 mm and 10 mm and a peak TRE of 24 mm. These outliers are

omitted from the figure. After registration with KinematicDIR, the median TRE for patient 1_{CT} is (1.2 ± 0.1) mm with an IQR of (0.9 ± 0.2) mm. This improvement is substantial and statistically significant for all fractions. Registration with KinematicDIR can be considered accurate for the purpose of radiation therapy. In addition, there is no observable trend in registration accuracy during the course of the treatment. This is the first indicator regarding the robustness of the approach, which will be investigated in the following.

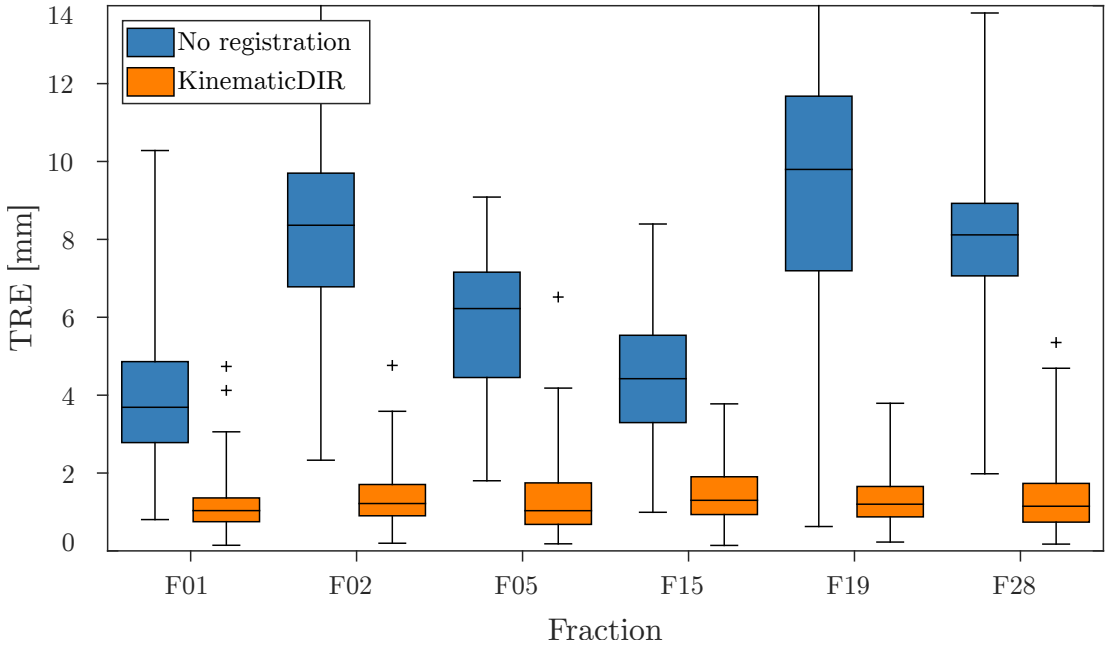

Figure 4.2: Accuracy evaluation for KinematicDIR for CT-CT registration. distribution of the TRE before (blue) and after registration with KinematicDIR (orange). The median TRE is reduced from 4 mm−10 mm to (1.2 ± 0.1) mm. The IQR after registration is (0.9 ± 0.2) mm indicating most landmarks could be registered to within an accuracy of 2 mm. There is no observable trend or change in the TRE distributions after KinematicDIR throughout the treatment.

4.1.3 Robustness Evaluation

To investigate the robustness of KinematicDIR, two aspects need to be shown. First, it is important whether the Puppetmaster model can be generalized to diverse patient geometries and set-ups. In particular, the effect of bones missing from the field of view can become relevant. Second, it is important to provide accurate registration through the treatment process for all patients. A physiological change during the treatment should ideally not affect the performance of KinematicDIR. In the presented experiment, three patient data sets with landmarks on 6 fraction CT scans per patient are used (see Section 3.3.1). After the registration with KinematicDIR, the TRE is calculated for all landmarks and all fractions. Figure 4.3 shows the TRE distribution for the three patient data sets. For patient 1_{CT}, this data is equal to Figure 4.2. The distributions are characterized by their median of (1.2 ± 0.1) mm. For patient 2_{CT} and patient 3_{CT} the median TRE is (1.6 ± 0.2) mm and (1.5 ± 0.1) mm, respectively. The slightly larger TRE coincides with the fact that the sternum was not visible for both patient 2_{CT} and 3_{CT}. The sternum typically fixates the rib cage as a highly correlated set of bones. Without this pre-positioning of the ribs, there is more opportunity for local errors. Importantly, however, the overall accuracy of less than 2 mm is consistent for all patients and remains valid for all considered fractions along the treatment course.

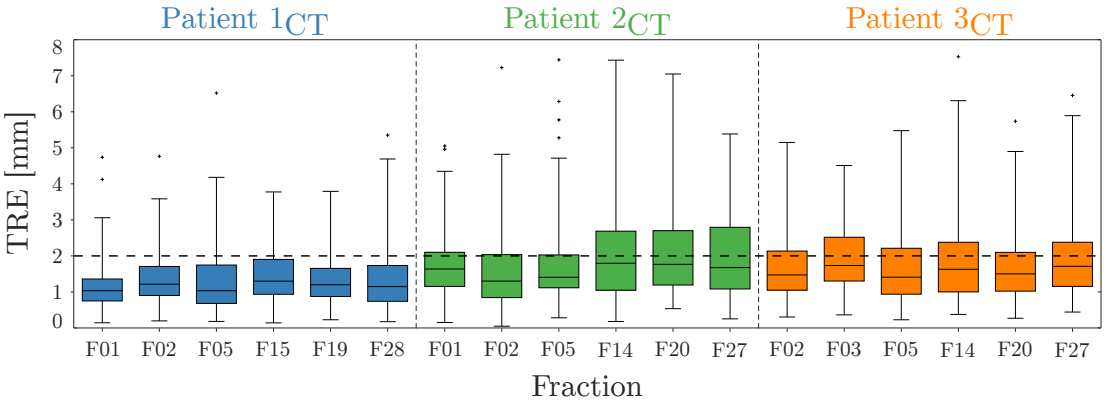

Figure 4.3: Robustness of KinematicDIR over three patients. For each patient, the TRE distributions after application of KinematicDIR are shown in box plots. The distributions remain consistent throughout the treatment course and for all fractions have a median TRE of less than 2 mm (dashed line).

An explicit change in the distribution is observed for patient 2_{CT} for the later three fractions. Here, the distribution of the TRE broadens slightly, meaning

the number of landmarks with a larger TRE is increased. This effect can be correlated with a plan adaptation after fraction F13. The unavoidable change in the patient positioning consequently led to relatively large shifts between the original planning CT and the fraction CT. While this does not cause any issue for the KinematicDIR pipeline, it has a detrimental effect on landmark identification since the landmark is quite often hidden within the relatively thick slice of 3 mm. The inferior TRE can hence be associated with a shortcoming of the evaluation metric rather than then KinematicDIR method.
In a second analysis of the robustness of KinematicDIR, the DSC is used to quantify the overlap of the skeleton tissue before and after the registration with KinematicDIR. An important limit for this evaluation is the overlap of the manual segmentations used for the model build-up and the semi-automatic segmentations used for the evaluation, which is just 0.89 when applied to the planning CT. This is a reasonable limit to how large the DSC can be after KinematicDIR. Figure 4.4 shows the DSC before and after registration for all available 36 fraction CT scans and the limiting dashed line of 0.89. The DSC overlap varies between 0.52 and 0.73 indicating the significant motion present in the skeleton between planning CT and fraction CT. After the application of KinematicDIR, the DSC is consistently above 0.84 with an average of 0.86 ± 0.01. This indicates that KinematicDIR performs equally well for all fractions of this patient.
Overall, the robustness analysis shows that KinematicDIR can be applied to a general patient geometry and that the registration quality remains consistent throughout the treatment course for all patients. The benefit of a visible sternum can be seen in the data.

4.1.4 Comparison with Intensity-Based Registration

To compare the performance of KinematicDIR with a state-of-the-art intensity-based deformable image registration algorithm, a direct comparison is performed for the data set of patient 1_{CT}. For this purpose, the registration between planning and each of the six fractions CTs is done using B-Spline-based deformable image registration within Plastimatch (performed in 3Dslicer, see Section 3.3.1). The registration is done once with default parameters and once with fine grid parameters. For each registered data set, the TRE is calculated using the available landmarks.
The distribution of TRE is shown in Figure 4.5 for the three-stage Plastimatch

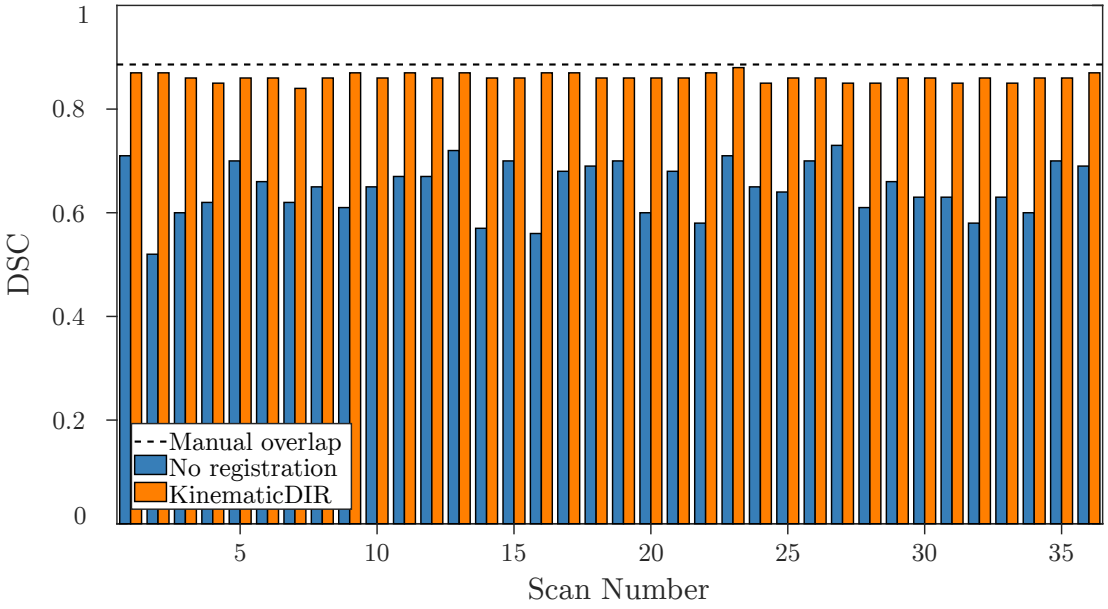

Figure 4.4: DSC-based robustness analysis of KinematicDIR. DSC indicating the overlap of the skeleton tissue before (no registration) and after application of KinematicDIR. The DSC overlap is consistent for all fractions and is close to the evaluation limit. KinematicDIR yields a strong improvement of the DSC compared to no registration.

with default parameters (blue), fine grid parameters (green), and KinematicDIR (orange). Using the default parameters, the three-stage Plastimatch approach performs worse for all considered fractions and shows a median TRE of (2.0 ± 0.3) mm while the median TRE for KinematicDIR is (1.2 ± 0.1) mm as described in Section 4.1.2. With the fine grid parameters, the registration quality can be considered equal, as Plastimatch achieves a TRE of (1.2 ± 0.2) mm.

It is noteworthy that the registration with default Plastimatch is substantially worse for the later fractions. The last two investigated fractions show a median TRE of (2.4 ± 0.1) mm while the first four fractions have a median TRE of (1.7 ± 0.1) mm. A possible explanation for this could be anatomical changes that lead to an incorrect bone deformation in the intensity-based registration with global support points. KinematicDIR, however, is unaffected by these soft tissue changes.

The registration with finer grid points achieves the same registration accuracy as KinematicDIR. Here, there is no trend to be observed during the treatment course. However, with the fine selection of grid points, local warping can be expected which does not conserve the rigidity of bones. Overall, the experiments show that KinematicDIR performs on par with a tuned intensity-based DIR algorithm while

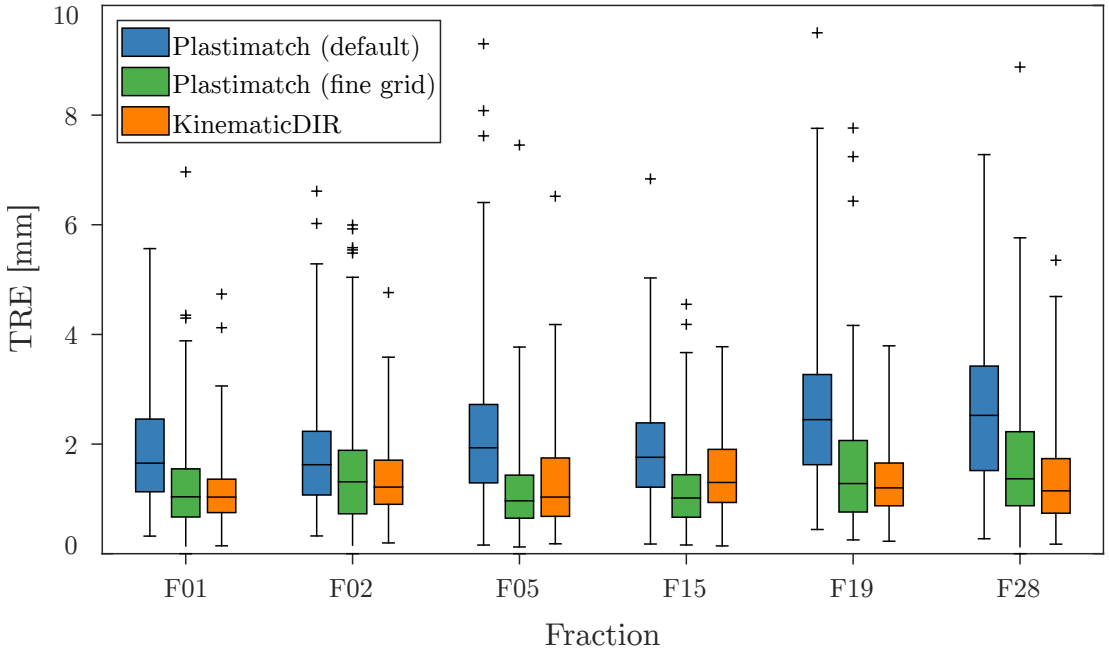

Figure 4.5: Comparison of KinematicDIR and intensity-based image registration. For each of the six fractions of patient 2_{CT}, a registration with Plastimatch is performed with default (blue) and with fine grid parameters (green). The TRE distribution is shown in box plots comparing the performance of Plastimatch and KinematicDIR (orange). The intensity-based registration achieves a median TRE of (2.0 ± 0.3) mm with default setting and (1.2 ± 0.2) mm over all fractions compared to (1.2 ± 0.1) mm for KinematicDIR.

guaranteeing the desired properties of the transformation. When no parameter tuning is applied, KinematicDIR yields a superior registration.

4.2 Multimodal Image Registration on Cone Beam CT Images

For the registration of CT and CBCT, the data set of one patient (patient 1_{CBCT}) was considered that showed a full field of view in the region of the shoulders without any truncation. This was necessary for the positioning of the scapulae in the registration process.

To evaluate the accuracy of the registration, the original planning CT was registered to three fractions CBCTs using the KinematicDIR approach. The considered fractions were chosen at the beginning, middle, and end of the fractionated

radiation therapy to investigate the performance through the treatment. For each

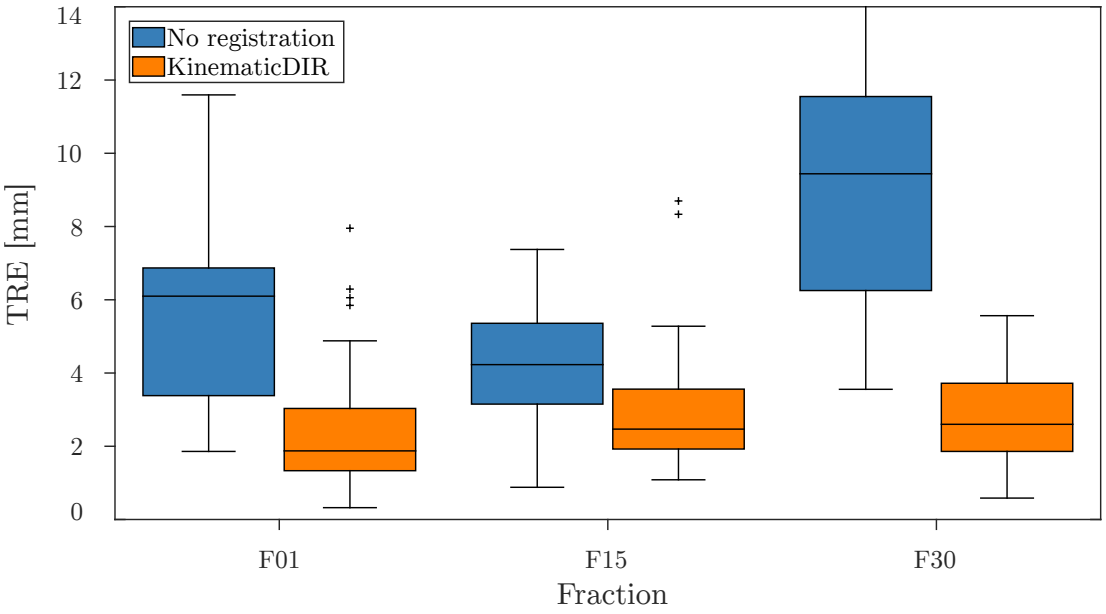

Figure 4.6: Accuracy evaluation for KinematicDIR for CT-CBCT registration. Distribution of the TRE before (blue) and after registration with KinematicDIR (orange). Before registration, the typical median TRE of 4 mm−10 mm can be seen. After KinematicDIR, the TRE is reduced to (2.3 ± 0.3) mm. There is no observable trend or change in the TRE distributions after KinematicDIR throughout the treatment.

fraction, the TRE is calculated before and after the registration. Figure 4.6 shows the distribution of the TRE in box plots. Before the application of KinematicDIR, typical deformations in the range of 4 mm−10 mm median TRE are observed. It should be noted that the distribution is very broad and includes maximum deviations of up to 18 mm that are omitted from the figure.
After the registration using KinematicDIR, the median TRE is reduced to (2.3 ± 0.3) mm. In addition, the TRE shows only minimal deterioration throughout the treatment fraction. The IQR for the distribution is (1.7 ± 0.1) mm indicating a broader distribution compared to the monomodal registration (see Section 4.1.2). It is noteworthy that the combination of higher median TRE and IQR are two indicators for an overall worse registration for the CT-CBCT case as compared to the monomodal registration.

4.3 Multimodal Image Registration on Magnetic Resonance Images

For the application of KinematicDIR in the context of CT-MRI registration, a major challenge could be bone identification on the MRI. In this work, this was solved using manual segmentation of the skeleton on the MRI. In this section, these manual segmentations are analyzed regarding their usability for bone detection and image registration. Then, the KinematicDIR pipeline is tested using these manual segmentations.

4.3.1 Skeleton Detection on MR Images

To identify bone tissue on the MR images, a manual, binary segmentation approach (bone-background) is used. However, it is known that geometric distortion can arise in MR images. To investigate the effect of the different contrast and these distortions on the delineation of bones, the segmentations of individual bones are used.
For each bone, the volume is calculated based on the contours and the slice thickness. Figure 4.7 a) shows the distribution of the volumetric ratios $\frac{V_{\mathrm{MRI}}}{V_{\mathrm{CT}}}$ per bone for all three CT-MRI data sets in box plots including individual data points. For this observer, the segmentations for each bone and patient are significantly smaller on the MRI than on the CT image data. The distributions have a median of (0.87 ± 0.01) and range from a minimum of 0.70 to a maximum of 1.08. Overall, the segmentations on the MR image data are substantially smaller, indicating that the human observer tends to under-segment bones on the MR image.
To provide sufficiently accurate bone detection for the optimization of the Puppetmaster in the KinematicDIR pipeline, the similarity regarding the shape of the bone segmentations can play a more important role. With sufficient geometrical similarity, the difference in size might not cause an incorrect local optimum. To investigate the geometrical similarity between the two segmentations, a shape analysis is performed on the voxel set of each bone. The standard deviation in each of the coordinate components is a first-order measure for the shape of the object. Figure 4.7 b) shows a scatter plot of the corresponding standard deviation in all three coordinate components for each bone and patient, as well as the identity line in black. Most of the data points are near the identity line, indicating

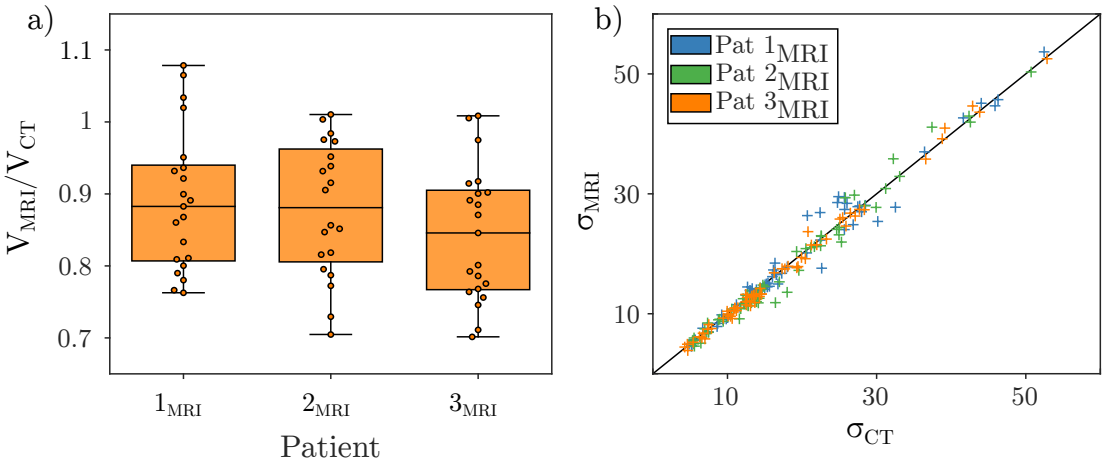

Figure 4.7: Geometric analysis of bone segmentations on MRI as compared to CT. a) Volumetric ratio of individual bone delineations. The volumes as segmented on the MRI are substantially smaller for all patients and nearly all bones. b) Analysis of the shape of the segmentations. The corresponding standard deviations (σ) of the MRI and CT segmentations are shown in a scatter plot. Most points are near the identity line (black) indicating geometric similarity.

that the shape in the cardinal directions is similar. This is an indicator that the segmentations vary in the volume that is covered but are very similar in their shape. This geometrical similarity can be essential for proper bone detection facilitating the use in KinematicDIR.

4.3.2 Accuracy of Image Registration

For the application test of KinematicDIR in the CT-MRI registration case, the three CT-MR data sets are registered. On the landmarks positioned on each of the image scans, the TRE is calculated before and after the registration as a quantification of the registration accuracy. Figure 4.8 shows the distribution of the TRE.

Before registration, the TRE ranges from 5 mm to 11 mm in the median. These distributions reflect the fact that patient 1_{MRI} did not wear a fixating face mask during the MR image acquisition while patient 2_{MRI} and 3_{MRI} wore such fixations. This results in strongly reduced motion between the planning CT and the MRI and hence lower median TRE. After registration with KinematicDIR, the median TRE is reduced to (2.0 ± 0.1) mm with an IQR of (1.5 ± 0.5) mm. These values are comparable to the TRE found in the CT-CBCT case. Since only one MRI is

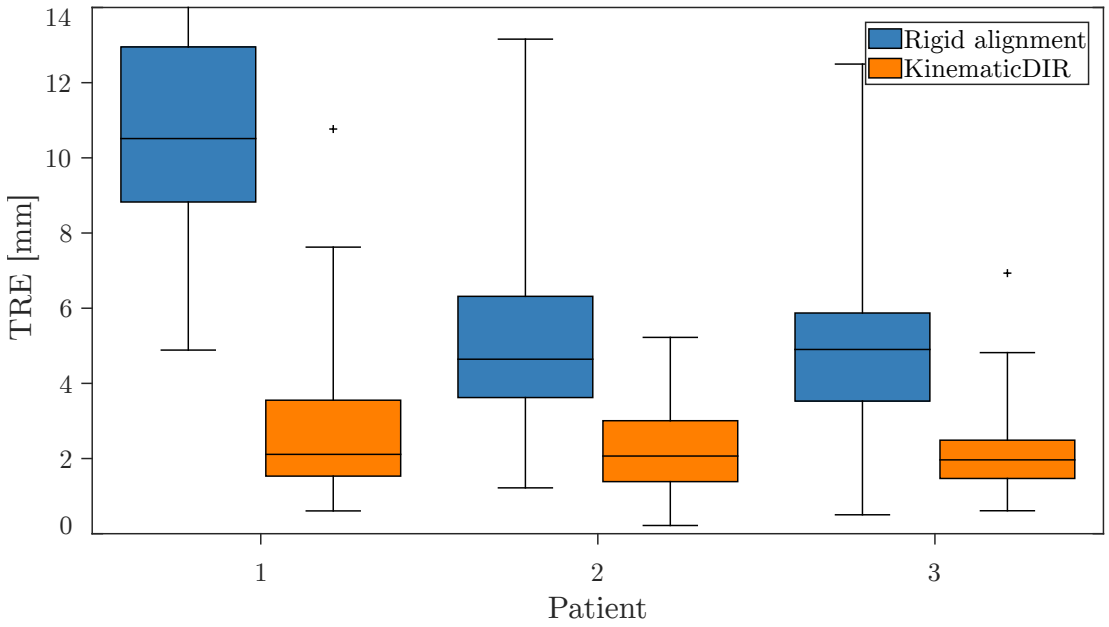

Figure 4.8: Accuracy evaluation for KinematicDIR for CT-MRI registration. Distribution of the TRE before (blue) and after registration with KinematicDIR (orange). Before registration median TRE is 5 mm−11 mm depending on the fixation of the patient during the MR acquisition. After KinematicDIR, the TRE is reduced to (2.0 ± 0.1) mm with an IQR of (1.5 ± 0.5) mm. For all three patients, the distributions are comparable. However, for patient 1_{MRI}, there is an outlier of more than 10 mm TRE.

available for each patient, it is not possible to state how the accuracy develops throughout the treatment course. However, all three patients show a comparable accuracy indicating the approach of KinematicDIR can be used for CT-MRI data sets in general. The accuracy of 2 mm can be considered adequate for radiation therapy applications. In this calculation, it should be noted that the localization of landmarks on MR images is challenging and in itself an error-prone process. This results in several outliers of the TRE after registration. For patient 1_{MRI} a single landmark showed a TRE of more than 10 mm, which might be due to incorrect positioning of the said landmark.

4.4 Bionic Augmentation

4.4.1 Proof-of-Principle: Bionic Augmentation of Bone Segmentation

The investigation of the potential for Bionic Augmentation is performed on the task of bone segmentation. For this purpose, two architectures are considered. First, a U-net is trained for the task of bone segmentation. The training collective of 10 manually segmented patient data sets is enriched with 10 standard augmented data sets in the first experiment. Then the same architecture is trained with the same 10 manually segmented data sets but this time enhanced with 10 Bionic Augmentation data sets. For both architectures, the predictions on one test patient data set are quantified using the DSC between the predicted segmentations and the ground truth segmentations provided by a human observer.
Figure 4.9 a) shows the evaluation for the U-net in a scatter plot of the DSC for Bionic Augmentation (BA) against standard argumentation (SA) including the identity line in black. The DSC with BA is higher for most bones that are considered. In particular, three ribs and the hyoid bone could not be detected with SA (DSC = 0) and could only be segmented with BA. For the vertebral bodies, a total of 8 showed improved DSC while four were detected worse with decreased DSC. Overall, the DSC increased by an average of 15% for all detected bones.
The state-of-the-art U-net-based segmentation is the nnU-net framework that automatically trains with a high number of standard augmentations. In a second experiment, the nnU-net is trained once with the same 10 manually segmented patient data sets and once with the additional 10 BA data sets. Figure 4.9 b) shows the evaluation of the nnU-net experiment. The DSC for BA and SA is shown in a scatter plot for two test data sets including the identity line in black. The first observation is that the highly augmenting U-net achieves higher DSC overall. In the region of DSC > 0.8, BA yields no further improvement as the prediction is already good. For the region of worse detection, however, there is a slight improvement for ribs, vertebrae, and the hyoid in one test data set.
These two experiments indicate that Bionic Augmentation can improve the training of a U-net-based automatic segmentation of individual bones when the number of training data sets is small and the architecture is the default U-net. Advanced

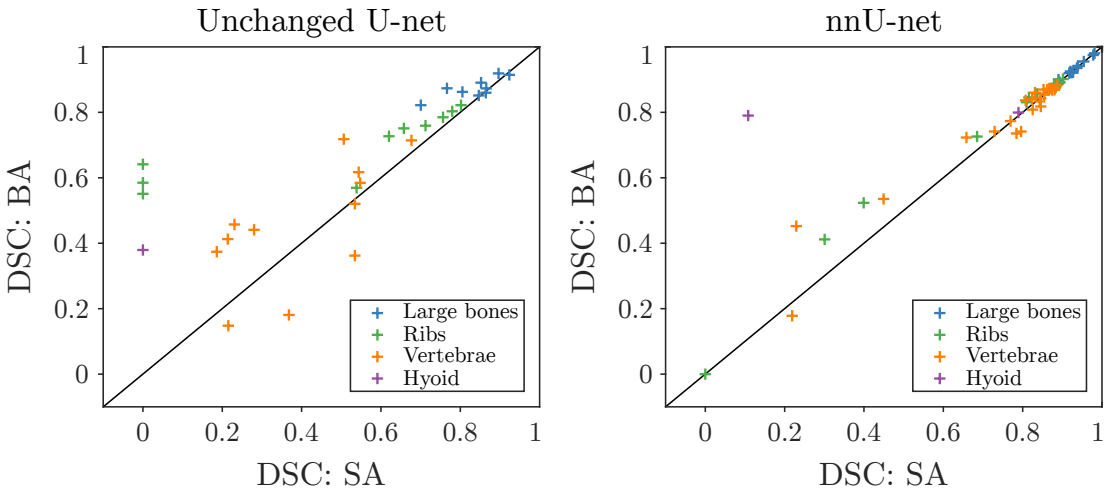

Figure 4.9: Evaluation of Bionic Augmentation (BA) for the improvement of bone segmentation U-nets. a) DSC for the U-net with one test data set. There are several bones (ribs and hyoid) only classified with BA. The average DSC is increased by 15% for the remaining bones. The vertebra segmentations are inconclusive with 8 improved but 4 inferior segmentations with BA as compared to standard augmentation (SA). b) DSC for the highly augmenting nnU-net with 2 test data sets. Overall increased DSC for both approaches as compared to the U-net. DSC for BA and SA are very similar for well-detected bones (DSC > 0.8) and a slight improvement for bad detected bones.

approaches like the highly augmenting nnU-net are capable of providing high-quality segmentations even with limited training data, since they use a wide range of standard augmentations. This limits the effect that a small number of Bionic Augmentation data sets can have.

4.4.2 Generalization of Postures

One bottleneck of the Bionic Augmentation pipeline as described in Section 3.4.1 is the manual creation of postures that only create a single additional data set. The proposed solution for this are generalized postures (Section 3.4.4) that lead to inconsistencies, by forcing the Puppetmaster into overlapping bones. Figure 4.10 displays the overlap quantified as the DSC between neighboring bones for the original postures, as well as the generalized postures for each of the five Bionic Augmentation patient data sets.

For 750 out of a total of 1745 bone pairs, the overlap is exactly zero. In addition, is was observed that there is a non-negligible inconsistency overlap even for the

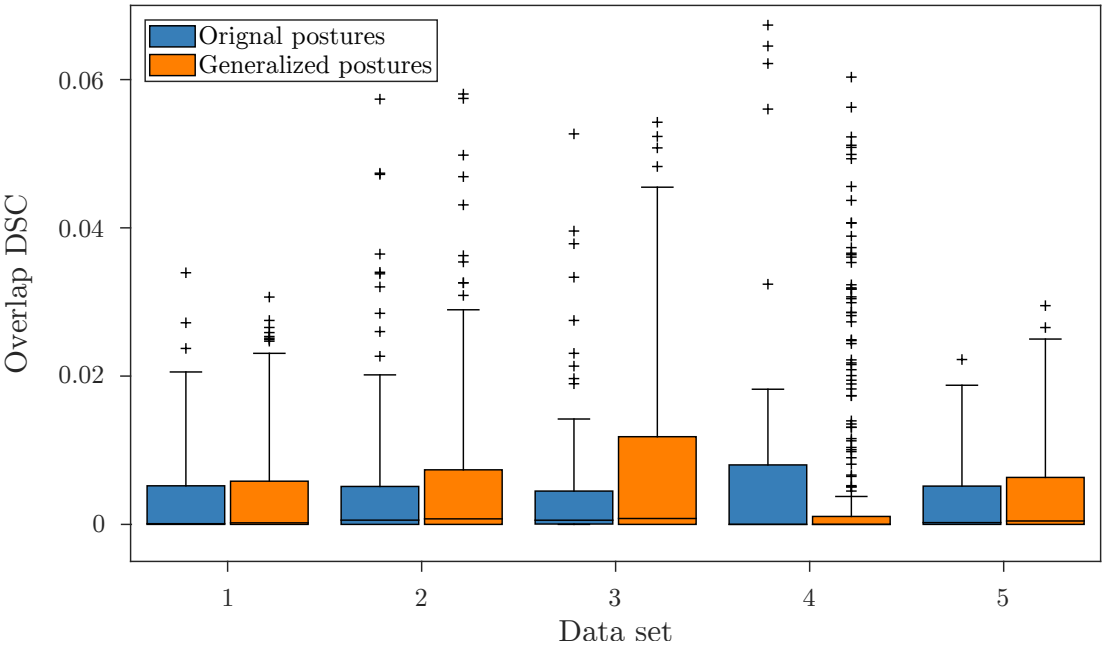

Figure 4.10: Overlap DSC for generalized postures for all Bionic Augmentation data sets. For 43% of all joints, there is no bone overlap (DSC = 0). For the bones with overlap, there is close to no change in the median inconsistency. It can be seen that the distribution becomes broader for data sets 2 and 3. For data set 4, there are many more outliers in the overlap DSC region of 0.02 to 0.06.

hand-curated original postures (blue). For some of the outliers, this is above 0.06. This distribution changes only slightly for the generalized postures (orange). The median remains unchanged for all patients and does not become higher than 0.001. The distribution for data sets 2 and 3 become slightly broader indicating that the overlap DSC has slightly increased. For data set 4, there are many bones with close to zero overlap (small box) but there are many outliers outside the 3· IQR whiskers. It is noteworthy that this data set has the highest outliers for the original postured, as well as the generalized postures. A closer observation reveals that data set 4 is the infant anatomy that behaves differently from the adult anatomy. This is a potential cause for this larger overlap for some bones.

4.5 Model Build-up from Automatic Segmentations

To translate the Puppetmaster model into a more realistic, clinical set-up, full automation is required. In particular, the model build-up in the current prototype relies on manual segmentations. In this section, the impact of the segmentations is evaluated. Then the potential of automatic segmentation strategies based on convolutional neural networks is investigated. The important metric in this part is the quality of the image registration the model can provide given a certain quality of segmentations. For this, the landmark-based evaluation is used.

4.5.1 Baseline: Segmentations by Human Observers

To investigate the effect of different segmentations in the build-up and performance of the Puppetmaster, the planning CT of patient 1_{CT} is segmented by two independent observers following the same set of rules (e.g. inclusion of costal cartilage).
A first geometrical analysis is performed, in order to evaluate how similar the segmentations of individual bones are when performed by a human observer. For this purpose, the DSC between each bone per observer, the Hausdorff distance, and the mean distance are evaluated for the 35 individual bones considered in this data set. Figure 4.11 a) shows the DSC between the two segmentations of individual bones in a box plot including individual data points. The distribution is centered at the median of 0.88 and ranges from 0.83 to 0.96.
In a second step, the geometric distance between the two contour sets is considered. Figure 4.11 b) shows the distributions of the Hausdorff distance and the mean distance. Per definition, the Hausdorff distance is always larger than the mean distance. In the considered data set, the median Hausdorff distance is 4.5 mm. It can be noted that the skull shows a very large Hausdorff distance of 19.2 mm that is found in the region where the mandible and skull meet. Here, the data set showed typical metal artifacts and both observers used a different approach to segmenting the frontal teeth.
In the mean distance, the large number of considered points leads to a very consistent mean distance of 0.8 mm. This indicates that the differences in the delineation are mostly local and are not caused by a global difference in the

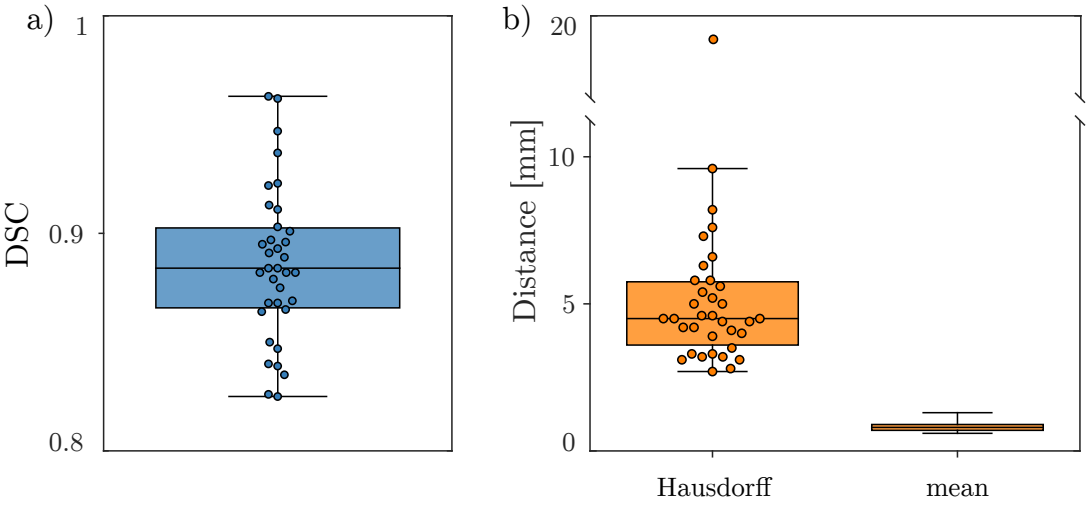

Figure 4.11: Geometric comparison between two manual segmentations of the same data set: The DSC a) shows a median of 0.88 and ranges from 0.82 to 0.96. The distance between the two segmentations b) is classified via the Hausdorff Distance (median 4.5 mm) and the mean distance between the contours (median 0.8 mm). It should be noted that the Hausdorff distance has one outlier(the skull).

segmentations.

Regarding the application of the manual bone segmentations for the model build-up of the Puppetmaster, it was analyzed how far the joint positions in the model deviate when using the different segmentations.

The distribution of the distances between individual joint positioning between the two different segmentations is shown in Figure 4.12 for the absolute distance and separated into the different coordinate contributions. The absolute distance has a median of 2.0 mm and includes three outliers. For the two skull-mandible joints and the connection of the first right rib and the sternum, a total distance of more than 14 mm is observed. For the skull-mandible region, this can be correlated to the large difference in the segmentations. The large difference in the connection of the rib and sternum can be explained as a result of the nearest neighbor approach to this joint positioning. In particular, for the first rib, the area where the sternum and rib are very close ranges over several slices and covers a relatively large region in each slice. Furthermore, the distinction of costal cartilage between the first and second rib is not always possible given the image information. Both effects can explain the large difference in joint positioning.

Regarding the contribution to the distance from the three image coordinates, it

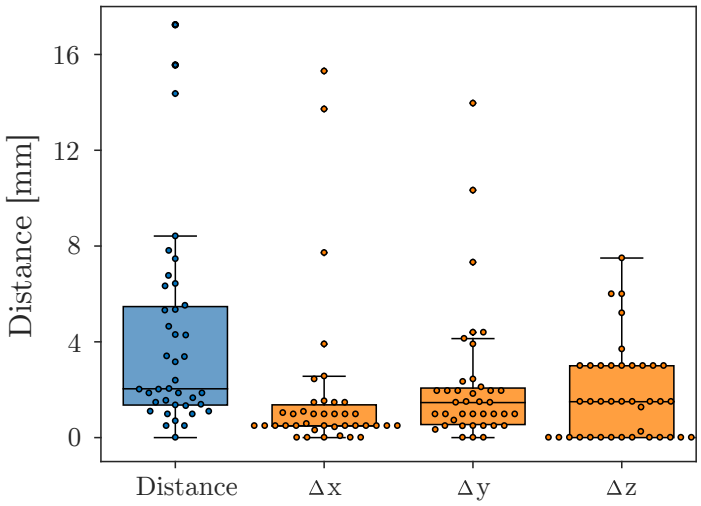

Figure 4.12: Joint distance from two manual segmentations created by human observers. The distribution of these distances is shown in box plots regarding the absolute distance as well as the distance in the image coordinates. The median distance is 2.0 mm and shows three outliers of more than 14 mm. These outliers are the two skull-mandible joints and the joint connecting the first right rib to the sternum. The coordinate distance is similar for the x and y coordinates. In the z-coordinate, a quantization of 1.5 mm is seen. This is half the slice distance in the data set.

can be seen that each contribution appears to be relevant. It is noteworthy that most joint positions in the z-direction are either in one image plane or between two image planes due to the nature of image processing in the Puppetmaster. With a slice distance of 3 mm, this explains why most values of Δz are given as multiples of 1.5 mm, i.e., half the slice distance. Finally, the question remains how different human segmentations affect the performance of the Puppetmaster as a biomechanical motion model. To evaluate this, the previously described registration scheme is performed for the same data set using both segmentations. Figure 4.13 shows the distributions of TRE after registration for both segmentations in box plots. There is no relevant difference between any of the distributions. Additionally, there is no indication of which segmentations would lead to better segmentation results. This all hints to the conclusion that both human segmentations – while varying geometrically and regarding the joint positioning – are sufficiently representing the human anatomy to allow the optimization process to find the proper posture and register the two images. This provides the baseline

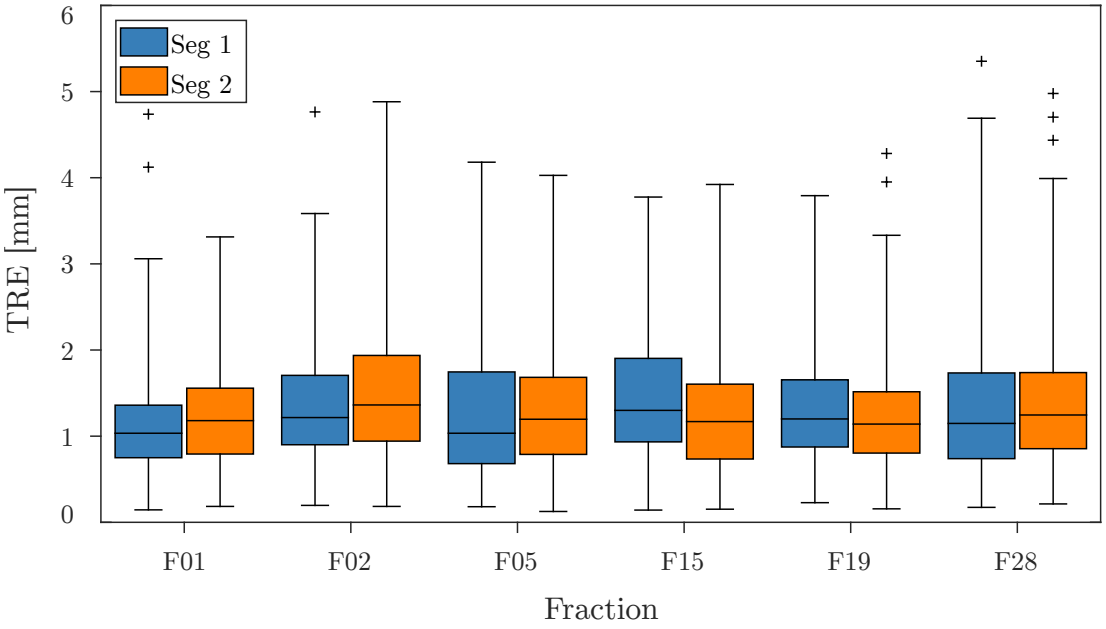

Figure 4.13: Registration quality for different segmentations. For each segmentation, the registration for all 6 fractions is performed and the distribution of the TRE is shown. There is no relevant difference between the distributions and no clear trend which segmentations would provide a better registration.

where automatic segmentations have to be equally well suited for the utilization of the Puppetmaster in the image registration pipeline. This is investigated in the next section.

4.5.2 Automatic Segmentations by a Generalized U-net

Using the TotalSegmentator framework, automatic bone segmentations of patient 1_{CT} are generated. The skull, mandible, hyoid, and sternum are not available in this automatic segmentation framework.
A visual comparison of the manual and automatic segmentations in the 3D rendering (Figure 4.14) reveals significant differences in the segmentations. The automatic segmentations do not include costal cartilage and misses the costovertebral connection between ribs and vertebral bodies. In addition, a difference in the joint positioning is visible (black arrows).
This is quantified in analogy to the human baseline, by performing a geometric analysis of the two segmentations. For the automatically segmented 31 individual bones, the DSC, Hausdorff distance, and mean distance are calculated.

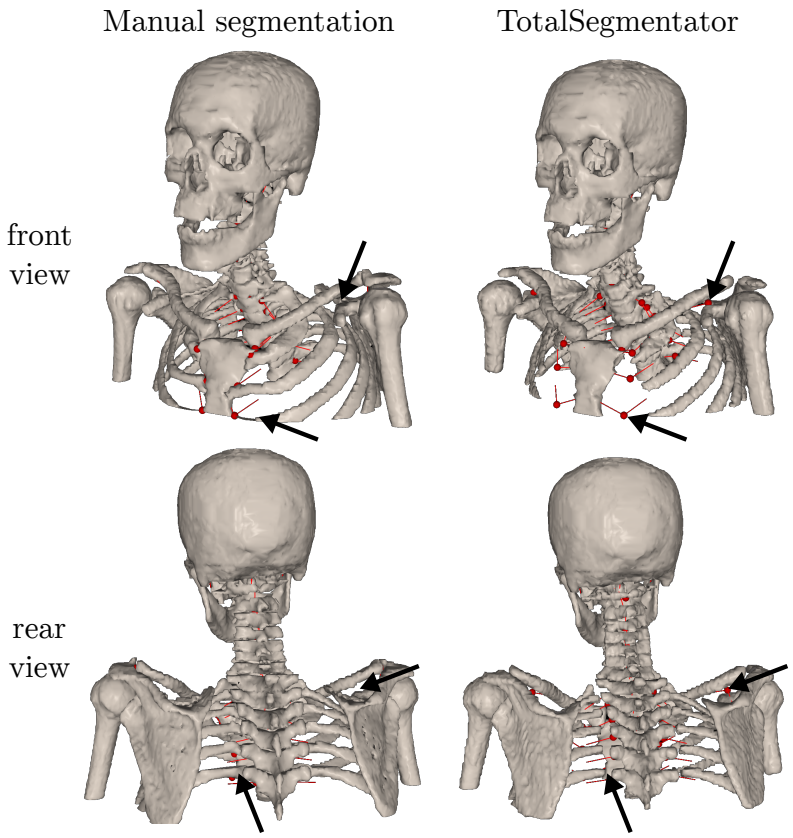

Figure 4.14: Comparison of manual and automatic segmentations. In the manual segmentations, the costal cartilage is included. In the TotalSegmentator-based segmentations, the attachments of ribs to the vertebral bodies are missing and there are generally slight differences in the segmentations. The effect this has on the position of joint anchor points is highlighted by orange arrows.

Figure 4.15 a) shows the distribution of the DSC in a box plot including individual data points. The distribution is described by a median of 0.81 ranging from 0.59 to 0.95. The low DSC scores arise from the ribs that are segmented without the costal cartilage in the TotalSegmentator framework. In addition, the ribs are not segmented all the way to the spine and therefore only a part of the rib is segmented. No ribs have a DSC of more than 0.76, confirming that the rib segmentations vary between the human observer and the automatic segmentations. It is noteworthy, however, that the vertebral bodies are segmented in high quality and are distinguished well by TotalSegmentator. This fact in particular is essential for the model build-up since any overlap or wrong identification of a vertebral body can prevent the correct build-up of the kinematic tree.
Figure 4.15 b) shows the distributions of the Hausdorff distance and the mean dis-

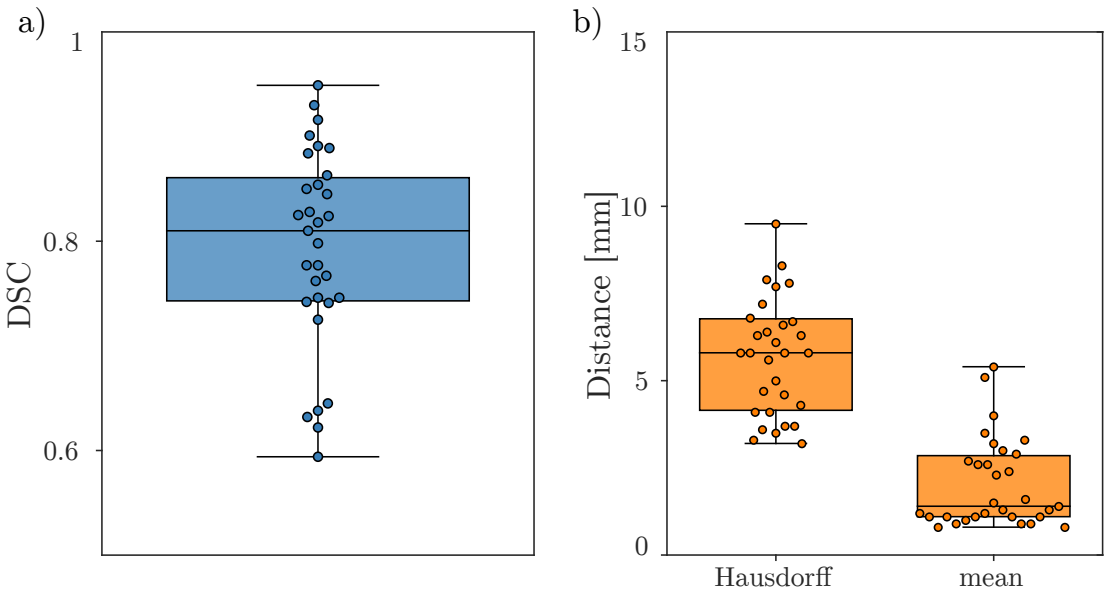

Figure 4.15: Geometric comparison between automatic and manual segmentations of the same data set. a) The DSC shows a median of 0.81 and ranges from 0.59 to 0.95. b) The distance between the two segmentations is classified via the Hausdorff Distance (median 5.8 mm) and the mean distance between the contours (median 1.4 mm).

tance between the manual segmentations and the automatic segmentations. The Hausdorff distance is substantially larger with a median of 5.8 mm, in particular, considering that the skull (the bone showing the largest Hausdorff distance for the manual segmentations) is not included in this analysis, because it is not segmented in the TotalSegmentator framework. The mean distance shows a median of more than 1.4 mm. This distribution is once again dominated by the ribs, since they are structurally different in the segmentations and hence have a large distance between their contours.

Overall, the automatic segmentations deviate in many aspects from the manual segmentations. The deviation arose from the different approaches to rib segmentations.

Nonetheless, the build-up of the Puppetmaster was successful with the automatic segmentations and the analysis of joint positions could be done. The Euclidean distance between each joint was calculated and is shown in Figure 4.16. To investigate the impact in the different image coordinate, the distance is also displayed split into its coordinate components. The median absolute distance is 11.3 mm and therefore significantly larger than the baseline value of 2.0 mm between human observers. This distribution is uneven. There are many joints positioned

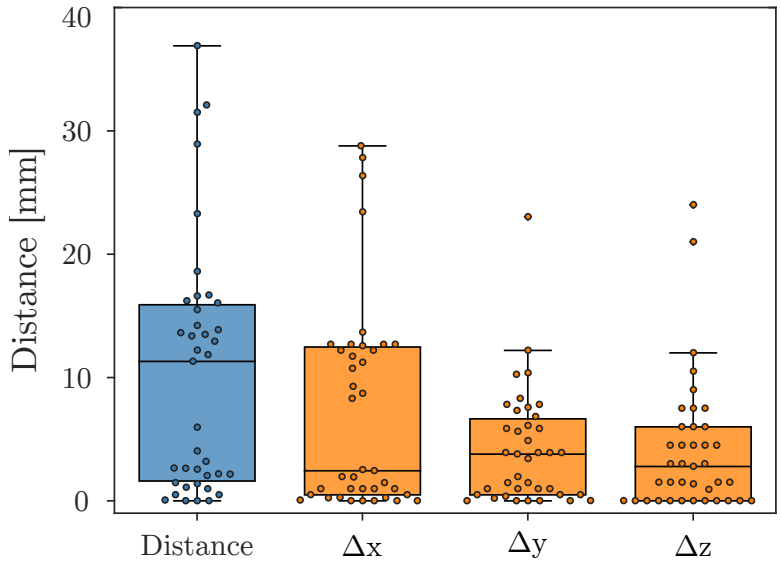

Figure 4.16: Joint distance between automatic and manual segmentations. The distribution of these distances is shown in box plots regarding the absolute distance as well as the distance in the image coordinates. The median distance is 11.2 mm and shows that several joints are placed at distances of up to 40 mm from their position following the manual segmentations. The distance in the rib joints is mostly in the medial-lateral coordinate that is denoted as x.

close to their position given the manual segmentations. These are exclusively the inter-vertebral joints. For the joints connecting the ribs to vertebrae and sternum a larger deviation is observed, as the joint is located on the middle point of the shortest connecting line between the rib and the sternum/vertebral body. This leads to a systematic shift that can be well seen in the x-coordinate of the distance (the x-coordinate is the medial-lateral dimension in this analysis).

It is noteworthy that the largest distances are the two joints between the scapulae and the clavicles. In the automatic segmentations, they are positioned further anterior as the automatic segmentations miss parts of the upper scapulae where the joint is positioned following the manual segmentations. This indicated how sensitive a nearest-neighbor positioning can be in this setting.

As before, the final test regarding the use of automatic segmentations for the model build-up of the Puppetmaster is once again the quality of the registration in KinematicDIR. In analogy to the baseline of human observers, the model is built-up from the automatic segmentations. The image registration is then performed using KinematicDIR for the planning CT and the six fractions CTs with available landmarks.

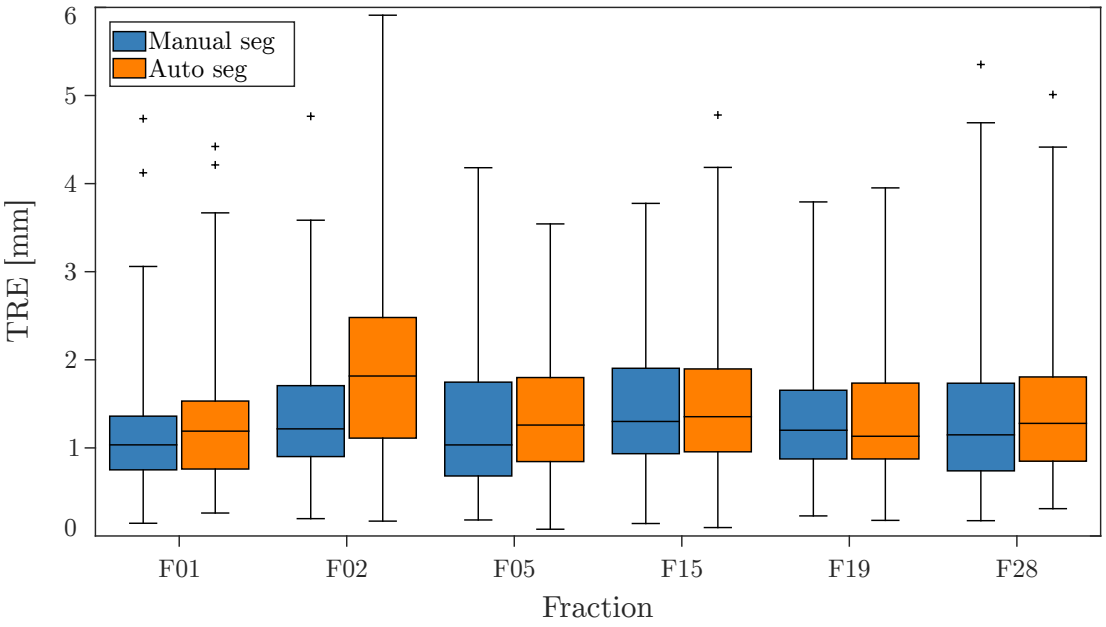

Figure 4.17: Registration quality for automatic and manual segmentations. For each set of segmentations, the result of the KinematicDIR registration is quantified in the TRE for all six fractions. The box plots indicate the equal performance of the automatic and the manual segmentations for all fractions but fraction F02. Here, the TRE is 1.8 mm and therefore significantly higher than for the manual segmentations.

Figure 4.17 shows the distributions of TRE after registration of patient 1_{CT} in box plots. In the same way as for different human segmentations, there is no relevant change in the registration quality for 5 out of 6 fractions. For fraction 2, the median TRE is 1.8 mm, the largest among all registered fractions and significantly larger than the median TRE of 1.2 mm observed with the manual segmentations. Over all fractions, the median TRE from the automatic segmentations is (1.3 ± 0.2) mm and therefore insignificantly worse than the manual segmentations (with a TRE of (1.1 ± 0.1) mm).

The noteworthy result is that the quality of the registration appears to be very robust regarding the segmentations of individual bones in the CT image. In the baseline, slight geometric differences were leading to slight deviations for most joints. This had no impact on registration accuracy. With the automatic segmentations as generated by the TotalSegmentator framework, the geometric differences were large and lead to drastic differences in the joint positions. Given the small range of motion in typically fixated radiation therapy of less than 10 mm, the KinematicDIR pipeline still provides an equally accurate registration as compared

to the human segmentations. In the performed experiment, even the worst fraction registered with the model built-up from the automatic segmentations still achieves a TRE of less than 2 mm.
The relevance of this experiment and an interpretation of why the registration accuracy remains unaffected is given in detail in Section 5.5.

5 Discussion

5.1 The Biomechanical Model

The selected biomechanical model is composed of a two-step patient motion model. At the center, the patient-tailored articulated skeleton multi-body model is used as a kinematic motion model. Using inverse kinematics, it can be employed to register the bony structures of two images to each other (in the KinematicDIR approach) or to use a given input of bone positions and orientations and find the closest kinematically feasible posture for the input motion (in the Bionic Augmentation approach).
A modified chainmail-based algorithm provides a second stage motion propagation throughout the adjacent soft tissue and is applied after the kinematic solution of the Puppetmaster is found. Taking the motion of individual bones as initiators, it calculates the transformation of all other voxels within the image.
The result of this composite model is a displacement vector field (DVF) that either maps between two images or provides the transformation to synthesize a new image from a reference.

5.1.1 Kinematic Model: The Puppetmaster

The Puppetmaster as a kinematic motion model is based on the general principles of how an articulated human skeleton can move. This general model is mapped to the individual patient geometry using the image information of a planning CT that is available for all current radiation therapy applications.
In the current implementation, the Puppetmaster is built-up using contours of individual bones on the planning CT of a radiation therapy patient. As reference, a manual, refined segmentation is performed to generate these contours. This guarantees sufficient accuracy for the joint positioning and a unique label for each voxel. At the build-up, this results in a 3D representation of the skeleton without any overlap. The manual segmentation does require labor-intensive delineation of each individual bone in all necessary slices. Depending on the level of experience, this process can take several hours. The task of automatically

segmenting bones in CT images has been approached in various ways in the literature. Statistical shape models [145] and atlas-based segmentation algorithms showed promising results but are typically not widely available or they are only intended to segment the skeleton as a whole [8, 146]. Recently, deep learning-based methods have become popular for segmentation tasks and prove capable of bone segmentation [147–149]. With the publication of the nnU-net and the consequent TotalSegmentator [139, 144] an automatic bone segmentation using deep learning has become widely available. Section 5.5 covers the effect manual or automatic segmentations can have on the build-up and application of the Puppetmaster in the context of radiation therapy.

Regarding the motion model, the joints in the Puppetmaster work as a regularization of which motion is possible within the skeleton. In the Puppetmaster, most joints are modeled as 3 DoF ball-and-socket joints. This grants mobility that is larger than typical in reality for many of the human joints, which can lead to unrealistic or unlikely motion if the input to the model is inconsistent or if the Simbody optimizer solving the inverse kinematics gets stuck in a local optimum. It should be noted, that this flexibility in the motion could also compensate for other inconsistencies within the model, like uncertainties in joint positioning. In particular, for the KinematicDIR approach, it can be assumed that the target location for each bone was realistic since it was directly derived from an imaged anatomy. Here, the high flexibility in the motion can compensate for potential inaccuracies in the joint positioning. This is a current hypothesis why the vastly different joint positions observed in the model build-up from automatic segmentations did not affect the registration accuracy in the KinematicDIR approach (see Section 4.5.2).

The joint parametrization is also relevant to the computational speed of the model. From the perspective of inverse kinematics in Simbody, more constraints result in higher computational demand to optimize the posture. Therefore, using ball-and-socket joints might increase the computation time of the kinematic model. For KinematicDIR, however, the time spent on the calculation of the posture (in the order of milliseconds) is small. The optimization of the complete skeleton in the hierarchical scheme takes about 20 minutes in the prototype since the down-hill simplex approach needs to compute a large number of postures to converge. A lower number of DoF could reduce the complexity of the optimization in the down-hill simplex approach and in the end result in faster optimization even if the time for each posture would increase.

Overall, the choice of the joint type remains a trade-off between the bio-fidelity of the model and computational speed regarding the specific application of the Puppetmaster.
A second aspect that currently limits the bio-fidelity of the Puppetmaster is the omitted collision detection. In the consideration of the kinematic tree, centroids of bones, and joints with their anchor points are not related to the 3D skeleton and can move independently of potential collisions. The option for bones to overlap in the model can be one source of inconsistencies in the DVF generated in KinematicDIR. This overlap becomes particularly important in the generation of synthetic data and is discussed further in Section 5.4. At the same time, the addition of collision detection would require many additional constraints, which consequently would increase the complexity and computational time required for the Puppetmaster.

5.1.2 Motion Propagation: The Chainmail

The Chainmail algorithm as the second stage of the composite biomechanical model in this work uses the results of the Puppetmaster as the initiators for each bone and propagates the motion into the surrounding soft tissue to create a DVF for the whole image space. While the DVF for the bone tissue can be considered accurate either as the results from the registration in KinematicDIR or the motion input from the Puppetmaster, the soft tissue DVF is calculated based on the result of motion propagation and is therefore extrapolated. For the head and neck region, this can be considered a valid approximation as the overall deformation in this region is dominated by skeletal motion. When considering the effect for particle therapy, the position of a bone can have a high impact on the range of the particles, while soft tissue can mostly be treated as water equivalent. Therefore, the approximation of soft tissue deformation can be adequate.
To allow for motion propagation, in heterogeneous tissues without explicit object segmentation, the Chainmail needs to assign elastic properties to each voxel. In the current implementation of the Chainmail, this is done using the self-parametrization from the HU values of the planning CT image. Assigning bones as rigid and air as fully elastic leaves the soft tissue that can be linearly interpolated. This interpolation does not incorporate elasticity effects that are not related to HU and should be seen as an estimate of the tissue parameters. Additionally, the geometrical constraint regarding maximum strain leads to a non-linear strain-

stress relation for large deformations. This limits the application to small soft tissue deformations [82]. In radiation therapy of the head and neck, this can be regarded as valid for most immobilization techniques. However, this deteriorate the potential of the Chainmail when considering large deformations in the abdominal or thoracic region. In the context of this thesis, relatively small deformations in the head and neck data sets in the order of 1 cm were considered and therefore the Chainmail algorithm could provide sufficiently accurate DVF for the whole image space.

5.2 Image Registration and Model Performance

The main approach investigated in this thesis was the biomechanically motivated deformable image registration KinematicDIR. The novelty of this approach is two-fold. Firstly, the inclusion of the Puppetmaster as the transformation model in the image registration process provides a kinematically feasible transformation that is rigid for each individual bone and respects the physiological joint positions. Secondly, the similarity metric for KinematicDIR is the overlap of bone tissue, which can be seen as an object-based similarity metric. This prevents adverse effects of image artifacts and enables the straightforward adaptation from monomodal CT-CT registration to multimodal cases including CBCT and MRI data sets.

For the monomodal registration, the identification of bone tissue was realized using thresholding based on the HU scale of the fraction CT without any corrections. This was sufficient to identify most of the bone tissue. The landmarks were well identifiable on the CT images with comparable contrast and hence provided a reliable evaluation. The good contrast facilitated consistently positioned landmarks in large numbers that could be distributed evenly within the skeleton.

For three patients and a total of 18 fractions, KinematicDIR registration was performed and reduced the median TRE to less than (1.6 ± 0.2) mm. This is significantly more accurate than reported results for an intensity-based approach [150] or a finite element-based approach [6]. In addition, 2 mm is the typical threshold that is deemed acceptable in photon radiation therapy [151–155]. For KinematicDIR, this is achieved for all patients and fractions. To compare KinematicDIR to a state-of-the-art DIR algorithm that might be used in radiation therapy, the SlicerRT algorithm Plastimatch was used. With the default parameter set,

Plastimatch resulted in a TRE of (2.0 ± 0.3 mm), which os similar to the published literature and worse than KinematicDIR. With a customized set of fine grid parameters, the accuracy of Plastimatch and KinematicDIR became comparable. However, it remains an issue of these fine grid DIR approaches that they can cause unrealistic local deformations of bone tissue, which would not be desirable for the quantification of electron densities in the context of radiation therapy [156].
The robustness of KinematicDIR was shown for several patients and fractions. The quality of registration did not deteriorate over the course of the radiation therapy treatment for any of the considered monomodal data sets. For one patient, a DSC-based analysis of the skeleton overlap was used to emphasize the robustness for all fractions. Overall, it could be concluded that the KinematicDIR approach can be used for diverse anatomies and fields of view without relevant loss in the registration quality.
For multimodal image registration, bone identification was the main challenge for object-based registration in KinematicDIR. To provide a proof-of-principle, semi-automatic (in the case of CBCT) or manual segmentations (in the case of MRI) were utilized. These were not a refined segmentations of bones but provided a rough estimate of which voxels can be considered bone tissue. For KinematicDIR, this provided sufficient information on bone tissue to perform image registration with a TRE in the order of 2 mm. For CBCT, the results in line with intensity-based DIR algorithms [157]. For CT-MRI registration, a comparison can be difficult since few works concentrate on bone tissue. In general, the accuracy of KinematicDIR was superior to what can be found in the literature for CT-MR registration with intensity-based DIR algorithms. [158]
An issue in the evaluation of the multimodal registration case was the positioning of the landmarks. For the observer, identifying the same points on the planning CT and the CBCT or MRI required experience and did benefit from additional anatomical knowledge. Furthermore, the positioning of landmarks in low-contrast regions can be infeasible leading to a lower number of landmarks and a non-uniform distribution. While all this limits the evaluation, KinematicDIR can be considered an accurate registration of multimodal images when evaluating the skeleton. In the multimodal case, the strength of the object-based registration becomes most apparent, since artifacts in the intensity distribution do not relevantly deteriorate the registration quality.
In the current version, KinematicDIR has certain drawbacks that remain to be solved. Most prominently, only bone tissue is registered. All soft tissue in the

image undergoes the soft tissue motion propagation of the Chainmail. This process does not incorporate image features representing specific individual motion. While this has limited effect in the head and neck region, where mostly muscles, fascias, and fatty tissue is present, KinematicDIR would require a modified soft tissue model for deformations of large deformable organs in the thoracic or abdominal region. From a computational point of view, the current prototype of KinematicDIR requires additional acceleration to be utilized in clinical ART workflows. An average computation time of 20 minutes with single-core optimization using the downhill simplex has the potential to be improved upon. There are several realistic approaches to accelerate KinematicDIR by parallelization and implementation on a GPU, the use of a faster optimizer, and pre-positioning of the bones. If this could bring the computation time to below one minute the application in adaptive workflows would be more realistic.

5.3 Methods for the Evaluation and Validation

Evaluating the quality of a DVF and the quality of an image registration is essential to validate the capability of a (biomechanical) model to represent the true anatomical motion [159]. In this thesis, a variety of concepts and metrics was utilized to evaluate the Puppetmaster and Chainmail model.
For the monomodal CT-CT registration with KinematicDIR, a visual evaluation was performed using a color fusion approach with complementary colors. Here, the same image intensity leads to a perceived grayscale value while any discrepancy yields a residual color. Perception of this residual color, however, depends on the intensity. This was acceptable in this evaluation since the primary observation was supposed to be the alignment of the skeleton, which typically has a high image intensity in CT images. Therefore, any discrepancy in the skeleton between the two images was well identified. The residual soft tissue discrepancies were still visible and can be assessed using image fusion. One drawback of the image fusion approach is the reliance on the same contrast in both images limiting the use of color fusion for multimodal image comparison. Additionally, in the CT-CT data sets, there was a contrast agent administered for the planning CT, which was not present during the acquisition of the fraction CTs. This led to a visible offset in the thoracic region that was no indicator of misregistration and could potentially be detrimental as it might mask real discrepancies in the local environment. Overall,

visual evaluation depends on many external factors, including the windows, the display, and the considered contrast. In addition, visual evaluation can depend on the perception of the observer. Therefore, it can serve as a first step and needs to be followed by a quantitative evaluation.
For this purpose, a landmark-based evaluation was employed. After a human observer positioned the same visually identifiable point on both images, the distance between these two points could be considered a quantification of the local difference between both images. In the context of image registration, this is called the target registration error (TRE). The landmark-based analysis relied on the accurate and precise positioning of the landmarks on the images. It can be expected that the inconsistency of a landmark is at least on the same order of magnitude as the voxel size. A perfectly accurate image registration can hence be expected to achieve a TRE in the order of the voxel size [160].
Additionally, landmarks are typically identified in contrast-rich sites [158, 161] or in the case of the presented work on the skeleton. Therefore, this evaluation is limited to the skeleton and requires a uniform distribution of these landmarks. It should be noted that landmark-based evaluations are typically not capable of determining the accuracy between landmarks.
The evaluation using landmarks is connected with a large amount of manual labor, as observers need to position the landmarks on each considered image scan. For an evaluation as performed for patient 1_{CT}, more than 160 landmarks were positioned. In a robustness analysis regarding all fractions, this would yield nearly 6000 landmarks. This was deemed an infeasible approach. As a substitute, the overlap of the skeleton was quantified using the Dice similarity coefficient (DSC). For this purpose, the skeleton needed to be segmented as one structure. A semi-automatic segmentation approach was chosen including thresholding above 120 HU and manual artifact removal. This method is fast but cannot be compared to the manual segmentations that provide accurate information about which voxel is part of the skeleton. Furthermore, such an analysis has to be critically appreciated regarding the sensitivity of the DSC. Since the skeleton is an overall large object, the DSC is not sensitive to errors of small region [162]. A completely wrong vertebra or ribs, for example, would lead to a small decrease of the DSC by 0.01 – 0.02. Overall, the DSC analysis can be seen as an additional test of the robustness of the KinematicDIR registration. In combination with the landmark-based TRE analysis, it can be an indicator of accuracy and robustness for all fractions.

5.4 Forward Generation of Postures and Bionic Augmentation

The generation of synthetic but realistic image data and corresponding labels of the segmentations is a novel and unique application of biomechanical models. In the presented work, the Puppetmaster was used to generate a new posture of manually segmented patient data. The Puppetmaster provides a graphical user interface that allows the interactive generation of new postures. In addition, an external input of position and orientation for each bone can be given and the Simbody optimizer for inverse kinematics finds the closest kinematic realization of this posture.

In a proof-of-principle experiment, 10 synthetic image data sets were created and combined with 10 manually segmented data sets. In the U-net, this Bionic Augmentation (BA) approach yielded a significant improvement over the use of standard augmentation (SA) based on image rotation for bone prediction. This improvement was particularly large for the ribs and hyoid bone. A possible explanation for this can be the large variety of the bone positions and the limited number of training data sets as compared to what is typically expected for medical image segmentation [163]. The inclusion of the BA data sets improved the overall coverage of the distribution in the training data set, which allowed for a better prediction. For the prediction of vertebral bodies, the use of BA had inconclusive effects. Here, the effect of overlap in the bones might have added confusion in the training data set. The overlap occurred due to the missing collision detection in the Puppetmaster but was also increased by interpolation artifacts after the transformation of the delineation with the DVF. The slice thickness of 3 mm in the considered image data did not allow for an ideal segmentation of the vertebral bodies, which additionally contributed to the overlap after the transformation.

In a second experiment, the state-of-the-art nnU-net was trained to investigate whether BA also provides improved predictions for a training routine that includes a large number of standard augmentations. The improvement of the nnU-net predictions was smaller than for the U-net. However, it should be noted that the overall prediction quality of the nnU-net was much higher. For most bones, the DSC was above 0.8, which already can be considered a good bone segmentation. In this case, an improvement was not realized. For bones with lower accuracy, the BA data sets did improve the prediction. Here the partial visibility of the ribs

and the lowest vertebral body might have contributed. BA can provide additional training data for these regions. Overall, this led to the hypothesis that BA can improve the prediction accuracy if the given training data insufficiently covers the underlying distribution of the label data.
An important aspect of the Puppetmaster is the potential to generate large skeletal motion. Explicitly, it could be shown in previous publications that the transfer of a CT in the arms-up position can be transformed into the arms-down position [122]. This can allow training data to be used in previously impossible scenarios. In this example, using a training data set of arms-down CT data could be augmented with an equal number of arms-up CT data from the same patient geometries to allow a more generalized model to be trained. This augmented model could provide a prediction for both arms-up and arms-down CT data, which would not be feasible without the augmentation and is not covered by standard augmentation techniques.
To facilitate the large-scale use of BA, the current prototype workflow of manually generating postures needs to be adapted to automate the process. The proposed method from this work is a library of generalized postures that could be applied to a large cohort of patient data sets, especially when the automated model build-up is possible (see Section 5.5). The current implementation of the Puppetmaster could have the issue of providing postures with an overlap in the bones, which would cause conflicting BA data. It could be shown, however, that the inconsistency in bone overlap does not significantly change when considering generalized postures applied to a similar patient geometry. When considering data sets of children this might not hold up. The child anatomy (data set 4 in Section 4.4.2) showed a different behavior when investigating the inconsistency overlap. This indicates that the inclusion of child data sets should be avoided or be done in sufficiently large numbers to facilitate proper generalization of the trained network. Regarding the posture generation, a separation of adult and child postures might be possible
Overall, the effect of interpolation after the transformation of the delineation appears to be the bigger challenge that needs to be solved for the use of bionic augmentation.

5.5 Model Build-up from Automatic Segmentations

An important aspect of the application of the Puppetmaster model in any clinical context would be the full automation of the model build-up. For the evaluations of the model performance in this thesis, manual segmentations of individual bones were used. For this, strict guidelines were given to the observers as to which parts of bones or cartilage were included in the delineation. Generating this manual delineations can take several hours or more, depending on the experience of the observer, the size of the image, and the slice thickness.
To enable the utilization of automatic segmentations, two investigations were performed in this work. First, the variability of segmentations between two experienced observers was quantified. A geometric analysis was used to examine the reproducibility of segmentations. The investigation showed a noticeable deviation between two segmentations of human observers under the same conditions. The DSC overlap as well as the Hausdorff distance indicated that both observers delineate following human intuition that is not represented in the guidelines given to them. The median DSC of 0.88 is in the same range as typical segmentation algorithms achieve in segmentation tasks that are considered successful [164].
Regarding the model build-up, the different segmentations had a limited impact. The analysis of the joint positions revealed that most joints were offset by less than 2 mm which would relate to less than two voxels in the investigated data set. In the final application test of how different segmentations affect the performance of the Puppetmaster, the image registration task with KinematicDIR for patient 1_{CT} was performed for all fractions with available landmarks. The TRE analysis showed no relevant difference between both registrations' qualities.
Overall, the baseline of human observers showed that two segmentations can have relatively large geometrical differences and still lead to an equal performance in registration accuracy. This is an additional aspect of why the object-based registration approach introduced in KinematicDIR is robust.
The equivalent analysis was performed for the automatic segmentations provided by the TotalSegmentator framework [144]. Since the TotalSegmentator is focused on the human torso, there were no segmentations available for the skull and mandible. In addition, neither the sternum nor the hyoid bone were part of the segmentations. In a fully automated Puppetmaster build-up, this would have to

be solved by using a hybrid model enhancing the TotalSegmentator by a custom network segmenting these four remaining bones. For the presented analysis, manual segmentations were used for the missing bones.
In the geometrical analysis, vastly larger discrepancies were found as compared to the two human segmentations. This was in part due to the decision to include costal cartilage in the rib segmentations, as it was found that the TotalSegmentator overall has a structurally different approach to the segmentation task. The costal cartilage is not included in the automatic segmentations and the connection of ribs to vertebral bodies is left out. This different approach led to a larger mean distance of the contours indicating that the differences are not locally limited but affect the whole delineation.
A beneficial aspect of the TotalSegmentator is the consistently high quality in differentiating adjacent bones. This enabled the build-up of the Puppetmaster model without any errors in the pipeline but revealed drastically different joint positions compared to the manual segmentations. While this was expected for the ribs given the different segmentation approach between the manual segmentations and the TotalSegmentator, it was also found that the scapula-humerus joint was more than 30 mm offset. This could be explained by the nearest-neighbor approach when positioning this particular joint. The TotalSegmentator had the tendency to under-segment the scapula, particularly in the cranial region. Since there are two regions where the scapula and humerus are close, this tendency resulted in the anterior region becoming the closest and hence the joint was positioned there. This emphasizes the limitations of nearest neighbor-based joint positioning when the fully automated Puppetmaster pipeline is considered. For the robustness of the model build-up, it would be beneficial to find simplified anatomical rules in analogy to the already introduced ones (e.g. following Veeger [165]).
In the application of the Puppetmaster in KinematicDIR, the quality of registration remained unchanged for most of the considered fractions. Only for F02, the median TRE of 1.8 mm was significantly larger for the automatic segmentations than for the manual segmentations. However, it was still within the 2 mm limit given for radiation therapy applications and hence deemed acceptable. There are two hypotheses as to why largely different segmentations led to nearly the same performance in the registration. Firstly, the expected motion in stereotactic radiation therapy can be considered rather small compared to the general human range of motion. The typical TRE of 10 mm is very relevant for radiation therapy, but can be considered small motion when compared to the overall range of skeletal

motion that can be assumed an order of magnitude larger in a free motion regime. The second aspect contributing to the good registration quality can be the choice of 3 DoF ball-and-socket joints in the Puppetmaster model. This additional flexibility might enable a correct positioning and orientation of the bones even with the incorrect joint positioning. Regarding the sternocostal joints connecting the ribs to the sternum, a relatively rigid connection can also be assumed. This means the precise location of this particular joint has less impact on the KinematicDIR pipeline.
In conclusion, different segmentations from human observers or automatic segmentations algorithms can vary in their geometry but lead to an accurate registration of nearly equal quality. This allows for the creation of a fully automated registration pipeline of KinematicDIR.

5.6 Outlook

Considering the Puppetmaster and Chainmail as a biomechanical model with the potential to be applied in various areas of radiation therapy, there are some further implementations and improvements that should be investigated.
Regarding the Puppetmaster model, the current version is implemented as a prototype that can be improved regarding realistic kinematics. In particular, adding more rule-based joints can become crucial when using automatic segmentations. In addition, replacing some of the ball-and-socket joints with more adequate joint models with less DoF could improve the bio-fidelity while also improving the computational speed. The costovertebral joint could be a prime candidate for this change of joint type. In general, the rib cage can be considered a highly correlated structure. This could be used to handle partially visible ribs and proper pre-positioning for faster optimization.
An essential enhancement of the kinematic model would be collision detection. This would greatly benefit the ability to produce realistic postures when it comes to the Bionic Augmentation approach. Additionally, this would provide a further constraint for the KinematicDIR optimization increasing the bio-fidelity.
Finally, the current version of the biomechanical model explicitly models bones and then applies forward propagation of the motion into surrounding soft tissue using the Chainmail. This approach can be useful for soft tissue without any increased risk of adverse side effects. However, more accurate modeling would

be beneficial for sensitive organs at risk or the tumor region. The tool of choice could be the finite element method that is already incorporated in some DIR algorithms [6]. Choosing specific organs and incorporating the initial kinematic input given by the Puppetmaster could provide a hybrid biomechanical model that realizes accurate quantification of anatomical changes within the human anatomy in reasonable time..

To achieve clinical relevance of the biomechanical model, it will also be crucial to increase the computational speed of KinematicDIR. The approach at hand would be the utilization of a faster optimizer that makes use of parallel computing. Other methods like the mentioned joint parametrization and pre positioning should be further investigated.

Considering advances in artificial intelligence methods, a second option for the clinical use of the Puppetmaster and Chainmail model could be as a quality and bio-fidelity check for the prediction of an artificial neural network. As an example, the predicted DVF of a machine learning-based DIR algorithm could be the input to the biomechanical model. This would enable the fast assessment of the bio-fidelity of the predicted DVF.

For these ideas, a fully automated model-build-up is required. It could be shown in this thesis that this automation is in principle possible using modern deep learning-based segmentation approaches. As the next step, the segmentations of all necessary bones and the joint positioning using automatic segmentations needs to be solved to facilitate the fully automatic build-up pipeline for the Puppetmaster.

6 Conclusion

In this work, the potential applications of a biomechanical patient model were investigated. The model consists of a kinematics-based articulated skeleton model and a soft tissue motion propagation to appropriately handle the deformation of the soft tissue and create the DVF for image transformation.
The first application was a model-based deformable image registration scheme called KinematicDIR for the head and neck region. It provided accurate and robust registration of the human skeleton in the monomodal CT-CT case. It was then adapted for the multimodal cases of CT-CBCT and CT-MRI by the addition of an object-based similarity metric. By using this object-based registration approach, Kinematic DIR was able to perform accurate registration on the medically relevant CBCT and MRI data sets.
In the second application, the biomechanical model was employed in the generation of synthetic yet realistic image data. Bionic Augmentation was used to enhance deep learning-based segmentation algorithms with additional, realistic image and label data in the training data set. This improved the performance of a U-net-based automatic bone segmentation model and achieved a slight improvement for the highly augmenting nnU-net framework. The biomechanically motivated augmentation approach can scale to large augmentation numbers by using generalized postures to synthesize large numbers of data sets.
Finally, the potential to build up the model from automatic segmentations was investigated. Geometrically, these automatic segmentations differ significantly from manual human segmentations. The model build-up was possible, but the positioning of joints varied between the two segmentations. KinematicDIR performed equally accurately and robustly with the automatic segmentation model. This emphasizes the robustness of the kinematic approach.

Glossary

This thesis introduces and used specific names for several applications. They are collected in this glossary for reference.

Bionic Augmentation	Generation of synthetic image and label data for augmented training of deep learning applications
Chainmail	Enhanced chainmail-based soft tissue propagation with self-parametriztation
KinematicDIR	Kinematics-based deformable image registration comprised of the Puppetmaster and the Chainmail
Puppetmaster	Kinematic articulated skeleton model

List of Acronyms

ANN	Artificial neural network
ART	Adaptive radiation therapy
BA	Bionic augmentation
CBCT	Cone beam computed (X-ray) tomography
CT	Computed (X-Ray) tomography
CTV	Clinical target volume
DoF	Degree of freedom
DSC	Dice similarity coefficient
DVF	Displacement vector field
FEM	Finite element method
FFD	Free form deformation
HU	Houndsfield unit
IGRT	Image guided radiotherapy
IQR	Inter quartile range
KinematicDIR	Kinematics-based deformable image registration
MI	Mutual information
MRI	Magnetic resonance image/imaging
MSE	Mean squared error
OAR	Organ at risk
PTV	Planning target volume
RBF	Radial basis function
RT	Radiation therapy
SA	Standard augmentation
SSD	Sum of squared differences
TPS	Thin-plate spline
TRE	Target registration error

List of Figures

List of Tables

Publications

Partial results of this dissertation were published in advance:

Journal articles

Bauer, Cornelius J.; Teske, Hendrik; Walter, Alexandra; Hoegen, Philipp; Adeberg, Sebastian; Debus, Jürgen; Jäkel, Oliver; Giske, Kristina:
Biofidelic Image Registration for Head and Neck Region utilizing an in-silico Articulated Skeleton as a Transformation Model,
in Review

Conference Contributions

Bauer, Cornelius J.; Jarutatsanangkoon, Pasit; Walter, Alexandra; Welzel, Thomas; Koerber, Stefan A.; Klüter, Sebastian; Jäkel, Oliver; Giske, Kristina:
CT-MR-Deformable Image Registration for MR-Guided Radiotherapy using a Biomechanical Skeleton Model 9$^{\text{th}}$ MR in RT Symposium, Los Angeles, United Stated of America, 2023; Oral presentation

Walter, Alexandra; **Bauer, Cornelius J.**; Dias Rodrigues, Joao; Hoegen, Philip; Adeberg, Sebastian; Welzel, Thomas; Koerber, Stefan A.; Paul, Katharina M.; Klüter, Sebastian; Jäkel, Oliver;Frank, Martin; Giske, Kristina:
The impact of additional MRI scans on training of supervised deep learning methods for automatic CTV delineation in head and neck cancer 9$^{\text{th}}$ MR in RT Symposium, Los Angeles, United Stated of America, 2023; Poster presentatio

Bauer, Cornelius J.; Yawson, Ama; Giske, Kristina:
Synthetische Bilddaten Augmentierung für DL-basierte Segmentierung mit einem biomechanischen, kinematischen Patientenmodell Proceedings of the 53rd Annual Scientific Meeting DGMP, Aachen 2022; Oral presentation

Yawson, Ama; **Bauer, Cornelius J.**; Giske, Kristina:
U-net-Based Segmentation for Individual Bones in radiation therapy planning CT scans for Head & Neck region Proceedings of the 26$^{\text{th}}$ Annual Conference, MIUA 2022, Cambridge, UK, 2022; Poster presentation

Bauer, Cornelius J.; Tekse, Hendrik; Bartelheimer, Katrin; Bendl, Rolf; Giske, Kristina:
Biomechanics in non-rigid image registration for adaptive radiotherapy of head and neck cancer: Articulated skeleton as an accurate transformation model 9$^{\text{th}}$ World Congress of Biomechanics, Taipeh, Taiwan, 2022; Oral presentation

Bibliography

[1] Rajamanickam Baskar, Kuo Ann Lee, Richard Yeo, and Kheng-Wei Yeoh. Cancer and radiation therapy: current advances and future directions. *International journal of medical sciences*, 9(3):193, 2012.

[2] C Nutting, DP Dearnaley, and S Webb. Intensity modulated radiation therapy: a clinical review. *The British journal of radiology*, 73(869):459–469, 2000.

[3] Byungchul Cho. Intensity-modulated radiation therapy: a review with a physics perspective. *Radiation oncology journal*, 36(1):1, 2018.

[4] O Jäkel, M Krämer, D Schulz-Ertner, P Heeg, CP Karger, B Didinger, A Nikoghosyan, and J Debus. Treatment planning for carbon ion radiotherapy in germany: review of clinical trials and treatment planning studies. *Radiotherapy and Oncology*, 73:S86–S91, 2004.

[5] Neil Kirby, Cynthia Chuang, and Jean Pouliot. A two-dimensional deformable phantom for quantitatively verifying deformation algorithms. *Medical physics*, 38(8):4583–4586, 2011.

[6] KK Brock, MB Sharpe, LA Dawson, SM Kim, and DA Jaffray. Accuracy of finite element model-based multi-organ deformable image registration. *Medical physics*, 32(6Part1):1647–1659, 2005.

[7] Hualiang Zhong, Jinkoo Kim, and Indrin J Chetty. Analysis of deformable image registration accuracy using computational modeling. *Medical physics*, 37(3):970–979, 2010.

[8] Stephen Yip, Timothy Perk, and Robert Jeraj. Development and evaluation of an articulated registration algorithm for human skeleton registration. *Physics in Medicine & Biology*, 59(6):1485, 2014.

[9] Wolfgang Schlegel, Christian P. Karger, and Oliver Jäkel, editors. *Medizinische Physik*. Springer Berlin Heidelberg, 2018.

[10] DE Velkley, DJ Manson, JA Purdy, and GD Oliver Jr. Build-up region of megavoltage photon radiation sources. *Medical physics*, 2(1):14–19, 1975.

[11] Hans A Bethe and Julius Ashkin. Experimental nuclear physics. *Wiley, New York*, 1953.

[12] Joseph Perl, Jungwook Shin, Jan Schümann, Bruce Faddegon, and Harald Paganetti. Topas: an innovative proton monte carlo platform for research and clinical applications. *Medical physics*, 39(11):6818–6837, 2012.

[13] Bruce Faddegon, José Ramos-Méndez, Jan Schuemann, Aimee McNamara, Jungwook Shin, Joseph Perl, and Harald Paganetti. The topas tool for

particle simulation, a monte carlo simulation tool for physics, biology and clinical research. *Physica Medica*, 72:114–121, 2020.

[14] Sea Agostinelli, John Allison, K al Amako, John Apostolakis, H Araujo, Pedro Arce, Makoto Asai, D Axen, Swagato Banerjee, GJNI Barrand, et al. Geant4—a simulation toolkit. *Nuclear instruments and methods in physics research section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.

[15] D Schardt, Heavy-Ion Therapy Collaboration, et al. Tumor therapy with high-energy carbon ion beams. *Nuclear Physics A*, 787(1-4):633–641, 2007.

[16] Silvia Scoccianti, Beatrice Detti, Davide Gadda, Daniela Greto, Ilaria Furfaro, Fiammetta Meacci, Gabriele Simontacchi, Lucia Di Brina, Pierluigi Bonomo, Irene Giacomelli, et al. Organs at risk in the brain and their dose-constraints in adults and in children: a radiation oncologist's guide for delineation in everyday practice. *Radiotherapy and Oncology*, 114(2):230–238, 2015.

[17] Steve Webb. The physical basis of IMRT and inverse planning. *The British journal of radiology*, 76(910):678–689, 2003.

[18] Jay S Cooper, Karen Fu, James Marks, and Sol Silverman. Late effects of radiation therapy in the head and neck region. *International Journal of Radiation Oncology* Biology* Physics*, 31(5):1141–1164, 1995.

[19] Anders Brahme. Development of radiation therapy optimization. *Acta oncologica*, 39(5):579–595, 2000.

[20] Lynn J Verhey. Immobilizing and positioning patients for radiotherapy. In *Seminars in Radiation Oncology*, volume 5-2, pages 100–114. Elsevier, 1995.

[21] Nicolas Depauw, Estelle Batin, Julianne Daartz, Anatoly Rosenfeld, Judith Adams, Hanne Kooy, Shannon MacDonald, and Hsiao-Ming Lu. A novel approach to postmastectomy radiation therapy using scanned proton beams. *International Journal of Radiation Oncology* Biology* Physics*, 91(2):427–434, 2015.

[22] Frank Lohr, Jürgen Debus, Claudia Frank, Klaus Herfarth, Otto Pastyr, Bernhard Rhein, Malte L Bahner, Wolfgang Schlegel, and Michael Wannenmacher. Noninvasive patient fixation for extracranial stereotactic radiotherapy. *International Journal of Radiation Oncology* Biology* Physics*, 45(2):521–527, 1999.

[23] Mark Ostyn, Thomas Dwyer, Matthew Miller, Paden King, Rachel Sacks, Ross Cruikshank, Melvin Rosario, Daniel Martinez, Siyong Kim, and Woon-Hong Yeo. An electromechanical, patient positioning system for head and neck radiotherapy. *Physics in Medicine & Biology*, 62(18):7520, 2017.

[24] Dante Amelio, Marcus Winter, Daniel Habermehl, Oliver Jäkel, Jurgen Debus, and Stephanie E Combs. Analysis of inter-and intrafraction accuracy

of a commercial thermoplastic mask system used for image-guided particle radiation therapy. *Journal of radiation research*, 54(suppl_1):i69–i76, 2013.

[25] Alexandra D Jensen, Marcus Winter, Sabine P Kuhn, Jürgen Debus, Olaf Nairz, and Marc W Münter. Robotic-based carbon ion therapy and patient positioning in 6 degrees of freedom: setup accuracy of two standard immobilization devices used in carbon ion therapy and IMRT. *Radiation Oncology*, 7(1):1–9, 2012.

[26] Meinhard Nevinny-Stickel, Reinhart A Sweeney, Reto J Bale, Andrea Posch, Thomas Auberger, and Peter Lukas. Reproducibility of patient positioning for fractionated extracranial stereotactic radiotherapy using a double-vacuum technique. *Strahlentherapie und Onkologie*, 180(2):117–122, 2004.

[27] Kimberley West, Majella Russo, Elizabeth Brown, Tamara Barry, Cathy Hargrave, and David Pryor. Evaluation of kidney motion with and without a pneumatic abdominal compression belt: considerations for stereotactic radiotherapy. *Journal of Medical Imaging and Radiation Oncology*, 62(1):128–132, 2018.

[28] Yong Hu, Yong-Kang Zhou, Yi-Xing Chen, and Zhao-Chong Zeng. Magnitude and influencing factors of respiration-induced liver motion during abdominal compression in patients with intrahepatic tumors. *Radiation Oncology*, 12(1):1–8, 2017.

[29] Gauthier Bouilhol, Myriam Ayadi, Simon Rit, Sheeba Thengumpallil, Joël Schaerer, Jef Vandemeulebroucke, Line Claude, and David Sarrut. Is abdominal compression useful in lung stereotactic body radiation therapy? a 4DCT and dosimetric lobe-dependent study. *Physica Medica*, 29(4):333–340, 2013.

[30] Marnix G Witte, Jan-Jakob Sonke, Jeffrey Siebers, Joseph O Deasy, and Marcel Van Herk. Beyond the margin recipe: the probability of correct target dosage and tumor control in the presence of a dose limiting structure. *Physics in Medicine & Biology*, 62(19):7874, 2017.

[31] Simon van Kranen, Olga Hamming-Vrieze, Annelisa Wolf, Eugène Damen, Marcel van Herk, and Jan-Jakob Sonke. Head and neck margin reduction with adaptive radiation therapy: robustness of treatment plans against anatomy changes. *International Journal of Radiation Oncology* Biology* Physics*, 96(3):653–660, 2016.

[32] Allen M Chen, Yao Yu, Megan E Daly, D Gregory Farwell, Stanley H. Benedict, and James A Purdy. Long-term experience with reduced planning target volume margins and intensity-modulated radiotherapy with daily image-guidance for head and neck cancer. *Head & neck*, 36(12):1766–1772, 2014.

[33] International Commission on Radiation Units and Measurements. *Prescrib-*

ing, recording, and reporting photon beam therapy. International Commission on Radiation Units and Measurements, 1993.

[34] C Thieke, U Malsch, W Schlegel, J Debus, P Huber, R Bendl, and C Thilmann. Kilovoltage CT using a linac-CT scanner combination. *The British journal of radiology*, 79(special_issue_1):S79–S86, 2006.

[35] Uwe Oelfke, Thomas Tücking, Simeon Nill, Annete Seeber, Bernd Hesse, Peter Huber, and Christoph Thilmann. Linac-integrated kv-cone beam CT: technical features and first applications. *Medical Dosimetry*, 31(1):62–70, 2006.

[36] Stephen Chin, Cynthia L Eccles, Alan McWilliam, Robert Chuter, Emma Walker, Philip Whitehurst, Joseph Berresford, Marcel Van Herk, Peter J Hoskin, and Ananya Choudhury. Magnetic resonance-guided radiation therapy: a review. *Journal of medical imaging and radiation oncology*, 64(1):163–177, 2020.

[37] Aswin Hoffmann, Bradley Oborn, Maryam Moteabbed, Susu Yan, Thomas Bortfeld, Antje Knopf, Herman Fuchs, Dietmar Georg, Joao Seco, Maria Francesca Spadea, et al. MR-guided proton therapy: a review and a preview. *Radiation Oncology*, 15(1):1–13, 2020.

[38] Wenjing Chen, Alexander Gemmel, and Eike Rietzel. A patient-specific planning target volume used in 'plan of the day'adaptation for interfractional motion mitigation. *Journal of radiation research*, 54(suppl_1):i82–i90, 2013.

[39] Sabrina T Heijkoop, Thomas R Langerak, Sandra Quint, Luiza Bondar, Jan Willem M Mens, Ben JM Heijmen, and Mischa S Hoogeman. Clinical implementation of an online adaptive plan-of-the-day protocol for nonrigid motion management in locally advanced cervical cancer IMRT. *International Journal of Radiation Oncology* Biology* Physics*, 90(3):673–679, 2014.

[40] Abdul Wahab M Sharfo, Sebastiaan Breedveld, Peter WJ Voet, Sabrina T Heijkoop, Jan-Willem M Mens, Mischa S Hoogeman, and Ben JM Heijmen. Validation of fully automated VMAT plan generation for library-based plan-of-the-day cervical cancer radiotherapy. *PloS one*, 11(12):e0169202, 2016.

[41] Qiuwen Wu, Yuwei Chi, Peter Y Chen, Daniel J Krauss, Di Yan, and Alvaro Martinez. Adaptive replanning strategies accounting for shrinkage in head and neck IMRT. *International Journal of Radiation Oncology* Biology* Physics*, 75(3):924–932, 2009.

[42] Tejinder Kataria, Deepak Gupta, Shikha Goyal, Shyam S Bisht, Trinanjan Basu, Ashu Abhishek, Kushal Narang, Susovan Banerjee, Shahida Nasreen, Sasikumar Sambasivam, et al. Clinical outcomes of adaptive radiotherapy in head and neck cancers. *The British Journal of Radiology*, 89(1062):20160085, 2016.

[43] Allen M Chen, Megan E Daly, Jing Cui, Mathew Mathai, Stanley Benedict, and James A Purdy. Clinical outcomes among patients with head and neck cancer treated by intensity-modulated radiotherapy with and without adaptive replanning. *Head & neck*, 36(11):1541–1546, 2014.

[44] Omar Bohoudi, Anna ME Bruynzeel, Suresh Senan, Johan P Cuijpers, Ben J Slotman, Frank J Lagerwaard, and Miguel A Palacios. Fast and robust online adaptive planning in stereotactic mr-guided adaptive radiation therapy (SMART) for pancreatic cancer. *Radiotherapy and Oncology*, 125(3):439–444, 2017.

[45] Patrick Kupelian and Jan-Jakob Sonke. Magnetic resonance–guided adaptive radiotherapy: a solution to the future. In *Seminars in radiation oncology*, volume 24-3, pages 227–232. Elsevier, 2014.

[46] A Webster, AL Appelt, and G Eminowicz. Image-guided radiotherapy for pelvic cancers: a review of current evidence and clinical utilisation. *Clinical Oncology*, 32(12):805–816, 2020.

[47] Dirk Verellen, Mark De Ridder, Nadine Linthout, Koen Tournel, Guy Soete, and Guy Storme. Innovations in image-guided radiotherapy. *Nature Reviews Cancer*, 7(12):949–960, 2007.

[48] Simon Jonathan Thomas. Relative electron density calibration of CT scanners for radiotherapy treatment planning. *The British journal of radiology*, 72(860):781–786, 1999.

[49] Christian Möhler, Tom Russ, Patrick Wohlfahrt, Alina Elter, Armin Runz, Christian Richter, and Steffen Greilich. Experimental verification of stopping-power prediction from single-and dual-energy computed tomography in biological tissues. *Physics in Medicine & Biology*, 63(2):025001, 2018.

[50] Lauren E Henke, JA Contreras, OL Green, B Cai, H Kim, MC Roach, JR Olsen, B Fischer-Valuck, DF Mullen, R Kashani, et al. Magnetic resonance image-guided radiotherapy (MRIgRT): a 4.5-year clinical experience. *Clinical Oncology*, 30(11):720–727, 2018.

[51] Sebastian Klüter. Technical design and concept of a 0.35 T MR-linac. *Clinical and Translational Radiation Oncology*, 18:98–101, 2019.

[52] Troels Bjerre, Sjoerd Crijns, Per Munck af Rosenschöld, Marianne Aznar, Lena Specht, Rasmus Larsen, and Paul Keall. Three-dimensional MRI-linac intra-fraction guidance using multiple orthogonal cine-MRI planes. *Physics in Medicine & Biology*, 58(14):4943, 2013.

[53] Sebastian Klüter, Sonja Katayama, C Katharina Spindeldreier, Stefan A Koerber, Gerald Major, Markus Alber, Sati Akbaba, Jürgen Debus, and Juliane Hörner-Rieber. First prospective clinical evaluation of feasibility and patient acceptance of magnetic resonance-guided radiotherapy in germany. *Strahlentherapie und Onkologie*, 196(8):691–698, 2020.

[54] Donald F Swinehart. The beer-lambert law. *Journal of chemical education*, 39(7):333, 1962.

[55] U Fano. Gamma-ray attenuation. part i. basic processes. *Nucleonics (US) Ceased publication*, 11, 1953.

[56] H Stanjek and WJHI Häusler. Basics of X-ray diffraction. *Hyperfine interactions*, 154(1):107–119, 2004.

[57] Godfrey N Hounsfield. Computerized transverse axial scanning (tomography): Part 1. description of system. *The British journal of radiology*, 46(552):1016–1022, 1973.

[58] James Ambrose. Computerized transverse axial scanning (tomography): Part 2. clinical application. *The British journal of radiology*, 46(552):1023–1047, 1973.

[59] Johann Radon. 1.1 über die bestimmung von funktionen durch ihre integralwerte längs gewisser mannigfaltigkeiten. *Classic papers in modern diagnostic radiology*, 5:21, 2005.

[60] Gabor T Herman. *Fundamentals of computerized tomography: image reconstruction from projections*. Springer Science & Business Media, 2009.

[61] Lee W Goldman. Principles of CT and CT technology. *Journal of nuclear medicine technology*, 35(3):115–128, 2007.

[62] AC Miracle and SK Mukherji. Conebeam CT of the head and neck, part 1: physical principles. *American Journal of Neuroradiology*, 30(6):1088–1095, 2009.

[63] AC Miracle and SK Mukherji. Conebeam CT of the head and neck, part 2: clinical applications. *American journal of neuroradiology*, 30(7):1285–1292, 2009.

[64] Anil Kumar Nagarajappa, Neha Dwivedi, and Rana Tiwari. Artifacts: The downturn of CBCT image. *Journal of International Society of Preventive & Community Dentistry*, 5(6):440, 2015.

[65] Charles Kittel, Paul McEuen, and Paul McEuen. *Introduction to solid state physics*, volume 8. Wiley New York, 1996.

[66] Felix Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946.

[67] Robert W Brown, Y-C Norman Cheng, E Mark Haacke, Michael R Thompson, and Ramesh Venkatesan. *Magnetic resonance imaging: physical principles and sequence design*. John Wiley & Sons, 2014.

[68] Robin A De Graaf. *In vivo NMR spectroscopy: principles and techniques*. John Wiley & Sons, 2019.

[69] Amy Walker, Gary Liney, Peter Metcalfe, and Lois Holloway. MRI distortion: considerations for MRI based radiotherapy treatment planning. *Australasian*

physical & engineering sciences in medicine, 37(1):103–113, 2014.

[70] Calvin R Maurer Jr, Georges B Aboutanos, Benoit M Dawant, Srikanth Gadamsetty, Richard A Margolin, Robert J Maciunas, and J Michael Fitzpatrick. Effect of geometrical distortion correction in MR on image registration accuracy. *Journal of computer assisted tomography*, 20(4):666–679, 1996.

[71] John P Mugler III. Overview of MR imaging pulse sequences. *Magnetic resonance imaging clinics of North America*, 7(4):661–697, 1999.

[72] W Thomas Dixon. Simple proton spectroscopic imaging. *Radiology*, 153(1):189–194, 1984.

[73] Jingfei Ma. Dixon techniques for water and fat imaging. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 28(3):543–558, 2008.

[74] Arthur Ardeshir Goshtasby. *2-D and 3-D image registration: for medical, remote sensing, and industrial applications*. John Wiley & Sons, 2005.

[75] Peter J Bickel and Kjell A Doksum. *Mathematical statistics: basic ideas and selected topics, volumes I-II package*. Chapman and Hall/CRC, 2015.

[76] Francisco PM Oliveira and Joao Manuel RS Tavares. Medical image registration: a review. *Computer methods in biomechanics and biomedical engineering*, 17(2):73–93, 2014.

[77] Roberto Manduchi and Gian Antonio Mian. Accuracy analysis for correlation-based image registration algorithms. In *1993 IEEE International Symposium on Circuits and Systems*, pages 834–837. IEEE, 1993.

[78] Frederik Maes, Andre Collignon, Dirk Vandermeulen, Guy Marchal, and Paul Suetens. Multimodality image registration by maximization of mutual information. *IEEE transactions on Medical Imaging*, 16(2):187–198, 1997.

[79] Adrian Andronache, Philippe Cattin, and Gábor Székely. Local intensity mapping for hierarchical non-rigid registration of multi-modal images using the cross-correlation coefficient. In *International Workshop on Biomedical Image Registration*, pages 26–33. Springer, 2006.

[80] Fred L. Bookstein. Principal warps: Thin-plate splines and the decomposition of deformations. *IEEE Transactions on pattern analysis and machine intelligence*, 11(6):567–585, 1989.

[81] Richard Szeliski and James Coughlan. Spline-based image registration. *International Journal of Computer Vision*, 22(3):199–218, 1997.

[82] Kathrin Bartelheimer. *A Heterogeneous and Multi-Range Soft-Tissue Deformation Model for Applications in Adaptive Radiotherapy*. PhD thesis, Heidelberg University, 2020.

[83] Hartmut Prautzsch, Wolfgang Boehm, and Marco Paluszny. *Bézier and B-spline techniques*, volume 6. Springer, 2002.

[84] Daniel Rueckert, Luke I Sonoda, Carmel Hayes, Derek LG Hill, Martin O Leach, and David J Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE transactions on medical imaging*, 18(8):712–721, 1999.

[85] Mike Fornefett, Karl Rohr, and H Siegfried Stiehl. Radial basis functions with compact support for elastic registration of medical images. *Image and vision computing*, 19(1-2):87–96, 2001.

[86] Michael Velec and Kristy K Brock. Biomechanical modeling applications in image-guided radiotherapy. In *Biomechanics of Soft Tissues*, pages 117–144. CRC Press, 2018.

[87] Jian-Kun Shen, Bogdan J Matuszewski, Lik-Kwan Shark, and Christopher J Moore. Deformable image registration using spring mass system. In *BMVC*, pages 1199–1208, 2006.

[88] Alexander Hagemann, Karl Rohr, H Siegfried Stiehl, Uwe Spetzger, and Joachim M Gilsbach. Biomechanical modeling of the human head for physically based, nonrigid image registration. *IEEE transactions on medical imaging*, 18(10):875–884, 1999.

[89] J Neylon, X Qi, K Sheng, R Staton, J Pukala, R Manon, DA Low, P Kupelian, and A Santhanam. A GPU based high-resolution multilevel biomechanical head and neck model for validating deformable image registration. *Medical physics*, 42(1):232–243, 2015.

[90] Andrew Nealen, Matthias Müller, Richard Keiser, Eddy Boxerman, and Mark Carlson. Physically based deformable models in computer graphics. In *Computer graphics forum*, volume 25-4, pages 809–836. Wiley Online Library, 2006.

[91] Olek C Zienkiewicz and Robert Leroy Taylor. *The finite element method for solid and structural mechanics*. Elsevier, 2005.

[92] Nitin S Gokhale. *Practical finite element analysis*. Finite to infinite, 2008.

[93] Yi-Je Lim, Dhanannjay Deo, Tejinder P Singh, Daniel B Jones, and Suvranu De. In situ measurement and modeling of biomechanical response of human cadaveric soft tissues for physics-based surgical simulation. *Surgical endoscopy*, 23(6):1298–1307, 2009.

[94] Gurpreet Singh and Arnab Chanda. Mechanical properties of whole-body soft human tissues: a review. *Biomedical Materials*, 2021.

[95] Kevin J Glaser, Armando Manduca, and Richard L Ehman. Review of MR elastography applications and recent developments. *Journal of Magnetic Resonance Imaging*, 36(4):757–774, 2012.

[96] Lei Shi, Wang Yao, Yu Gan, Lily Y Zhao, W Eugene McKee, Joy Vink, Ronald J Wapner, Christine P Hendon, and Kristin Myers. Anisotropic material characterization of human cervix tissue based on indentation and inverse finite element analysis. *Journal of biomechanical engineering*, 141(9), 2019.

[97] Sarah F Gibson. 3D chainmail: a fast algorithm for deforming volumetric objects. In *Proceedings of the 1997 symposium on Interactive 3D graphics*, pages 149–ff, 1997.

[98] Markus A Schill, Sarah FF Gibson, H-J Bender, and Reinhard Männer. Biomechanical simulation of the vitreous humor in the eye using an enhanced chainmail algorithm. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 679–687. Springer, 1998.

[99] Hendrik Teske, Kathrin Bartelheimer, Jan Meis, Rolf Bendl, Eva M Stoiber, and Kristina Giske. Construction of a biomechanical head and neck motion model as a guide to evaluation of deformable image registration. *Physics in Medicine & Biology*, 62(12):N271, 2017.

[100] Kathrin Bartelheimer, Hendrik Teske, Rolf Bendl, and Kristina Giske. Tissue-specific transformation model for CT-images. *Current Directions in Biomedical Engineering*, 3(2):525–528, 2017.

[101] Hendrik Tim Teske. *Articulated patient model in high-precision radiation therapy*. PhD thesis, Heidelberg University, 2019.

[102] Andrei Nikolaevich Tikhonov, AV Goncharsky, VV Stepanov, and Anatoly G Yagola. *Numerical methods for the solution of ill-posed problems*, volume 328. Springer Science & Business Media, 1995.

[103] Daniel E Whitney. Resolved motion rate control of manipulators and human prostheses. *IEEE Transactions on man-machine systems*, 10(2):47–53, 1969.

[104] William A Wolovich and Howard Elliott. A computational technique for inverse kinematics. In *The 23rd IEEE Conference on Decision and Control*, pages 1359–1363. IEEE, 1984.

[105] Charles W Wampler. Manipulator inverse kinematic solutions based on vector formulations and damped least-squares methods. *IEEE Transactions on Systems, Man, and Cybernetics*, 16(1):93–101, 1986.

[106] L-CT Wang and Chih-Cheng Chen. A combined optimization method for solving the inverse kinematics problems of mechanical manipulators. *IEEE Transactions on Robotics and Automation*, 7(4):489–499, 1991.

[107] Jianmin Zhao and Norman I Badler. Inverse kinematics positioning using nonlinear programming for highly articulated figures. *ACM Transactions on Graphics (TOG)*, 13(4):313–336, 1994.

[108] Samuel R Buss. Introduction to inverse kinematics with jacobian transpose, pseudoinverse and damped least squares methods. *IEEE Journal of Robotics and Automation*, 17(1-19):16, 2004.

[109] Michael Meredith and Steve Maddock. Real-time inverse kinematics: The return of the jacobian. 2004.

[110] Eliakim H Moore. On the reciprocal of the general algebraic matrix. *Bull. Am. Math. Soc.*, 26:394–395, 1920.

[111] Roger Penrose. A generalized inverse for matrices. In *Mathematical proceedings of the Cambridge philosophical society*, volume 51, pages 406–413. Cambridge University Press, 1955.

[112] Adi Ben-Israel and Thomas NE Greville. *Generalized inverses: theory and applications*, volume 15. Springer Science & Business Media, 2003.

[113] Michael A Sherman, Ajay Seth, and Scott L Delp. Simbody: multibody dynamics for biomedical research. *Procedia Iutam*, 2:241–261, 2011.

[114] Fabian Isensee. *From manual to automated design of biomedical semantic segmentation methods*. PhD thesis, Heidelberg University, Heidelberg, 2020.

[115] Shijie Hao, Yuan Zhou, and Yanrong Guo. A brief survey on semantic segmentation with deep learning. *Neurocomputing*, 406:302–321, 2020.

[116] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[117] Marc Peter Deisenroth, A Aldo Faisal, and Cheng Soon Ong. *Mathematics for machine learning*. Cambridge University Press, 2020.

[118] Gilbert Strang. *Linear algebra and learning from data*, volume 4. Wellesley-Cambridge Press Cambridge, 2019.

[119] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning representations by back-propagating errors. *nature*, 323(6088):533–536, 1986.

[120] William E Lorensen and Harvey E Cline. Marching cubes: A high resolution 3D surface construction algorithm. *ACM siggraph computer graphics*, 21(4):163–169, 1987.

[121] Timothy S Newman and Hong Yi. A survey of the marching cubes algorithm. *Computers & Graphics*, 30(5):854–879, 2006.

[122] Hendrik Teske, Kathrin Bartelheimer, Rolf Bendl, Eva M Stoiber, and Kristina Giske. Handling images of patient postures in arms up and arms down position using a biomechanical skeleton model. *Current Directions in Biomedical Engineering*, 3(2):469–472, 2017.

[123] Ajay Seth, Michael Sherman, Peter Eastman, and Scott Delp. Minimal formulation of joint motion for biomechanisms. *Nonlinear dynamics*, 62(1):291–303, 2010.

[124] Edda Eich. Convergence results for a coordinate projection method applied to mechanical systems with algebraic constraints. *SIAM Journal on Numerical Analysis*, 30(5):1467–1482, 1993.

[125] Roy Featherstone. *Rigid body dynamics algorithms*. Springer, 2014.

[126] Ajay Seth, Ricardo Matias, António P Veloso, and Scott L Delp. A biomechanical model of the scapulothoracic joint to accurately capture scapular kinematics during shoulder movements. *PloS one*, 11(1):e0141028, 2016.

[127] Richard Pio. Euler angle transformations. *IEEE Transactions on automatic control*, 11(4):707–715, 1966.

[128] Dong C Liu and Jorge Nocedal. On the limited memory BFGS method for large scale optimization. *Mathematical programming*, 45(1):503–528, 1989.

[129] John A Nelder and Roger Mead. A simplex method for function minimization. *The computer journal*, 7(4):308–313, 1965.

[130] Kristina Giske, Eva M Stoiber, Michael Schwarz, Armin Stoll, Marc W Muenter, Carmen Timke, Falk Roeder, Juergen Debus, Peter E Huber, Christian Thieke, et al. Local setup errors in image-guided radiotherapy for head and neck cancer patients immobilized with a custom-made device. *International Journal of Radiation Oncology* Biology* Physics*, 80(2):582–589, 2011.

[131] Michael Schwarz, Kristina Giske, Armin Stoll, Simeon Nill, Peter E Huber, Jürgen Debus, Rolf Bendl, and Eva M Stoiber. IGRT versus non-IGRT for postoperative head-and-neck IMRT patients: dosimetric consequences arising from a PTV margin reduction. *Radiation Oncology*, 7(1):1–7, 2012.

[132] Lee R Dice. Measures of the amount of ecologic association between species. *Ecology*, 26(3):297–302, 1945.

[133] Csaba Pinter, Andras Lasso, An Wang, David Jaffray, and Gabor Fichtinger. Slicerrt: radiation therapy research toolkit for 3D slicer. *Medical physics*, 39(10):6332–6338, 2012.

[134] Andriy Fedorov, Reinhard Beichel, Jayashree Kalpathy-Cramer, Julien Finet, Jean-Christophe Fillion-Robin, Sonia Pujol, Christian Bauer, Dominique Jennings, Fiona Fennessy, Milan Sonka, et al. 3D slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging*, 30(9):1323–1341, 2012.

[135] Eva Maria Stoiber, Nina Bougatf, Hendrik Teske, Christian Bierstedt, Dieter Oetzel, Jürgen Debus, Rolf Bendl, and Kristina Giske. Analyzing human decisions in IGRT of head-and-neck cancer patients to teach image registration

algorithms what experts know. *Radiation Oncology*, 12(1):1–7, 2017.

[136] Kerstin A Kessel, Christian Bohn, Uwe Engelmann, Dieter Oetzel, Nina Bougatf, Rolf Bendl, Jürgen Debus, and Stephanie E Combs. Five-year experience with setup and implementation of an integrated database system for clinical documentation and research. *Computer methods and programs in biomedicine*, 114(2):206–217, 2014.

[137] Allison K Hunter and WD McDavid. Characterization and correction of cupping effect artefacts in cone beam CT. *Dentomaxillofacial Radiology*, 41(3):217–223, 2012.

[138] Fabian Isensee, Jens Petersen, Andre Klein, David Zimmerer, Paul F Jaeger, Simon Kohl, Jakob Wasserthal, Gregor Koehler, Tobias Norajitra, Sebastian Wirkert, et al. nnu-net: Self-adapting framework for u-net-based medical image segmentation. *arXiv preprint arXiv:1809.10486*, 2018.

[139] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2):203–211, 2021.

[140] Rolf Bendl, Jürgen Pross, Mark Keller, Josef Bürkelbach, and Wolfgang Schlegel. VIRTUOS—a program for VIRTUal radiotherapy simulation. In *Computer Assisted Radiology/Computergestützte Radiologie*, pages 676–682. Springer, 1993.

[141] R Bendl, J Pross, A Hoess, M Keller, K Preiser, and W Schlegel. VIRTUOS-a program for virtual radiotherapy simulation and verification. In *Proc of 11th Int. Conf. on The Use of Computers in Radiation Therapy, AR Hounsell ua Manchester: North Western Med. Physics Dept*, pages 226–227, 1994.

[142] Felix Hausdorff. *Grundzüge der mengenlehre*, volume 7. von Veit, 1914.

[143] B. Sendov. *Hausdorff Distance*, pages 23–48. Springer Netherlands, Dordrecht, 1990.

[144] Jakob Wasserthal, Manfred Meyer, Hanns-Christian Breit, Joshy Cyriac, Shan Yang, and Martin Segeroth. Totalsegmentator: robust segmentation of 104 anatomical structures in CT images. *arXiv preprint arXiv:2208.05868*, 2022.

[145] Nazli Sarkalkan, Harrie Weinans, and Amir A Zadpoor. Statistical shape and appearance models of bones. *Bone*, 60:129–140, 2014.

[146] Yabo Fu, Shi Liu, H Harold Li, and Deshan Yang. Automatic and hierarchical segmentation of the human skeleton in CT images. *Physics in Medicine & Biology*, 62(7):2812, 2017.

[147] André Klein, Jan Warszawski, Jens Hillengaß, and Klaus H Maier-Hein. Automatic bone segmentation in whole-body CT images. *International*

journal of computer assisted radiology and surgery, 14(1):21–29, 2019.

[148] Noémie Moreau, Caroline Rousseau, Constance Fourcade, Gianmarco Santini, Ludovic Ferrer, Marie Lacombe, Camille Guillerminet, Pascal Jezequel, Mario Campone, Nicolas Normand, et al. Comparison between threshold-based and deep learning-based bone segmentation on whole-body CT images. In *Medical Imaging 2021: Computer-Aided Diagnosis*, volume 11597, pages 661–667. SPIE, 2021.

[149] Pengbo Liu, Hu Han, Yuanqi Du, Heqin Zhu, Yinhao Li, Feng Gu, Honghu Xiao, Jun Li, Chunpeng Zhao, Li Xiao, et al. Deep learning to segment pelvic bones: large-scale CT datasets and baseline models. *International Journal of Computer Assisted Radiology and Surgery*, 16(5):749–756, 2021.

[150] Christoph Hoffmann, Sonja Krause, Eva Maria Stoiber, Angela Mohr, Stefan Rieken, Oliver Schramm, Jürgen Debus, Florian Sterzing, Rolf Bendl, and Kristina Giske. Accuracy quantification of a deformable image registration tool applied in a clinical setting. *Journal of applied clinical medical physics*, 15(1):237–245, 2014.

[151] Carri K Glide-Hurst, Percy Lee, Adam D Yock, Jeffrey R Olsen, Minsong Cao, Farzan Siddiqui, William Parker, Anthony Doemer, Yi Rong, Amar U Kishan, et al. Adaptive radiation therapy (ART) strategies and technical considerations: a state of the art review from nrg oncology. *International Journal of Radiation Oncology* Biology* Physics*, 109(4):1054–1075, 2021.

[152] Hania A Al-Hallaq, Laura Cerviño, Alonso N Gutierrez, Amanda Havnen-Smith, Susan A Higgins, Malin Kügele, Laura Padilla, Todd Pawlicki, Nicholas Remmes, Koren Smith, et al. AAPM task group report 302: Surface-guided radiotherapy. *Medical physics*, 49(4):e82–e112, 2022.

[153] Eric E Klein, Joseph Hanley, John Bayouth, Fang-Fang Yin, William Simon, Sean Dresser, Christopher Serago, Francisco Aguirre, Lijun Ma, Bijan Arjomandy, et al. Task group 142 report: Quality assurance of medical accelerators a. *Medical physics*, 36(9Part1):4197–4212, 2009.

[154] Nahla K Saleh-Sayah, Elisabeth Weiss, Francisco J Salguero, and Jeffrey V Siebers. A distance to dose difference tool for estimating the required spatial accuracy of a displacement vector field. *Medical Physics*, 38(5):2318–2323, 2011.

[155] David Tilly, Nina Tilly, and Anders Ahnesjö. Dose mapping sensitivity to deformable registration uncertainties in fractionated radiotherapy–applied to prostate proton treatments. *BMC medical physics*, 13(1):1–12, 2013.

[156] Kana Motegi, Hidenobu Tachibana, Atsushi Motegi, Kenji Hotta, Hiromi Baba, and Tetsuo Akimoto. Usefulness of hybrid deformable image registration algorithms in prostate radiation therapy. *Journal of applied clinical medical physics*, 20(1):229–236, 2019.

[157] Angelo Mencarelli, Simon Robert van Kranen, Olga Hamming-Vrieze, Suzanne van Beek, Coenraad Robert Nico Rasch, Marcel van Herk, and Jan-Jakob Sonke. Deformable image registration for adaptive radiation therapy of head and neck cancer: accuracy and precision in the presence of tumor changes. *International Journal of Radiation Oncology* Biology* Physics*, 90(3):680–687, 2014.

[158] Kristy K Brock, Deformable Registration Accuracy Consortium, et al. Results of a multi-institution deformable registration accuracy study (MIDRAS). *International Journal of Radiation Oncology* Biology* Physics*, 76(2):583–596, 2010.

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

[160] Kristy K Brock, Sasa Mutic, Todd R McNutt, Hua Li, and Marc L Kessler. Use of image registration and fusion algorithms and techniques in radiotherapy: Report of the AAPM radiation therapy committee task group no. 132. *Medical physics*, 44(7):e43–e76, 2017.

[161] Richard Castillo, Edward Castillo, Rudy Guerra, Valen E Johnson, Travis McPhail, Amit K Garg, and Thomas Guerrero. A framework for evaluation of deformable image registration spatial accuracy using large landmark point sets. *Physics in Medicine & Biology*, 54(7):1849, 2009.

[162] Florian Kofler, Ivan Ezhov, Fabian Isensee, Fabian Balsiger, Christoph Berger, Maximilian Koerner, Johannes Paetzold, Hongwei Li, Suprosanna Shit, Richard McKinley, et al. Are we using appropriate segmentation metrics? identifying correlates of human expert perception for CNN training beyond rolling the DICE coefficient. *arXiv preprint arXiv:2103.06205*, 2021.

[163] Junghwan Cho, Kyewook Lee, Ellie Shin, Garry Choy, and Synho Do. How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *arXiv preprint arXiv:1511.06348*, 2015.

[164] Anjany Sekuboyina, Malek E Husseini, Amirhossein Bayat, Maximilian Löffler, Hans Liebl, Hongwei Li, Giles Tetteh, Jan Kukačka, Christian Payer, Darko Štern, et al. VerSe: A vertebrae labelling and segmentation benchmark for multi-detector CT images. *Medical image analysis*, 73:102166, 2021.

[165] HEJ Veeger. The position of the rotation center of the glenohumeral joint. *Journal of biomechanics*, 33(12):1711–1715, 2000.

Acknowledgements

- I would like to thank Prof. Oliver Jäkel for the opportunity to perform the research of this thesis in his department and for his guidance when it comes to medical physics.
- Prof. Seco agreed to be my first examiner for the Faculty of Physics and Astronomy. Thank you for your support and fruitful discussions.
- Prof. Rolf Bendl as the final member of my advisory committee was a fascinating help in all aspects of computational modeling. Thank you for your specific and general scientific input.
- Kristina Giske taught me all I needed during the last three years: code development, the Puppetmaster, and ever so often Virtuos. But more importantly, your ideas and strategies in sciences were valuable for my development as an independent scientist.
- I want to thank our group Computational Patients Models. Alex, Ama, Fabian, Goran, Jakob, Katrin, Kristina, Mark, Miki, Nora, Pasit, Pedro, Stephen, and Tom. I am glad I met you all.
- A special thank you to my HiWis Romy and Melanie for providing segmentations and landmarks for all the data in this work. Also thank you to Fabian for the TOPAS data creation and Ama and Alex for the training of the ANNs.
- Proofreading my work is not always an easy task. Thank you to all who helped me with their comments and suggestions.
- I have to thank Alex for all sorts of things. I will express my gratitude and appreciation in person, though.
- Der letzte Dank gilt meiner Familie. Ihr wart immer für mich da und habt mich auf diesem Weg begleitet. Danke für all das.

Declaration

I hereby affirm that I have written this work independently and without the use of resources and media other than those quoted.

Ich versichere hiermit, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe.

Heidelberg, den _____ _____

Cornelius J. Bauer