

Aus dem Deutschen Krebsforschungszentrum  
(Vorstand: Prof. Dr. Michael Baumann, Prof. Dr. Josef Puchta)  
Abteilung für Medizinische Bildverarbeitung  
(Leiter: PD Dr. rer. nat. Klaus Hermann Maier-Hein)

# AUTOMATIC IMAGE ANALYSIS OF C-ARM COMPUTED TOMOGRAPHY IMAGES FOR ANKLE JOINT SURGERIES

Inauguraldissertation  
zur Erlangung des Doctor scientiarum humanarum (Dr. sc. hum.)  
an der  
Medizinischen Fakultät Heidelberg  
der  
Ruprecht-Karls-Universität

vorgelegt von  
SARINA THOMAS  
aus  
Herborn  
2019

Prodekan: Prof. Dr. med. Hans-Georg Kräusslich

Doktorvater: Priv.-Doz. Dr. rer. nat. Klaus Hermann Maier-Hein

**CONTENTS**

**List of Figures** **v**

**List of Tables** **viii**

**List of Acronyms** **xi**

**1 Introduction** **1**

    1.1 Motivation . . . . . 1

    1.2 Objectives . . . . . 3

    1.3 Approach . . . . . 4

    1.4 Outline . . . . . 5

**2 Fundamentals** **7**

    2.1 Medical background . . . . . 7

        2.1.1 Anatomy of the upper ankle joint . . . . . 7

        2.1.2 Radiographic anatomical parameters . . . . . 10

        2.1.3 Syndesmotic injuries and diagnosis . . . . . 13

        2.1.4 Surgical treatment . . . . . 16

        2.1.5 Conclusion . . . . . 19

    2.2 Technical background . . . . . 19

        2.2.1 2D Radiography . . . . . 19

        2.2.2 Computed Tomography . . . . . 21

        2.2.3 Intra-operative C-arm imaging . . . . . 23

        2.2.4 Conclusion . . . . . 24

**3 State of the art** **25**

    3.1 Assistance systems in orthopedic and trauma surgery . . . . . 25

    3.2 3D anatomical shape representation . . . . . 27

        3.2.1 Deep learning approaches . . . . . 28

        3.2.2 Statistical shape modeling . . . . . 32

3.2.3	Combining shape model segmentation and deep learning . . . . .	38
3.3	Image understanding of radiological summation images . . . . .	39
3.4	2D-3D Reconstruction . . . . .	40
3.4.1	3D Model type . . . . .	41
3.4.2	Reconstruction strategy and definition of cost function . . . . .	42
3.4.3	Optimization strategies . . . . .	45
3.4.4	Setup and image data . . . . .	46
3.5	Discussion . . . . .	48
<b>4</b>	<b>Methods</b>	<b>53</b>
4.1	Approach . . . . .	53
4.1.1	Clinical workflow . . . . .	53
4.1.2	System design choices . . . . .	55
4.1.3	Sources of error . . . . .	59
4.2	3D bone segmentation . . . . .	62
4.2.1	Statistical shape and appearance model . . . . .	62
4.2.2	3D-CNN-based segmentation . . . . .	68
4.2.3	Combined approach: Deep learning-driven 3D-SSM segmentation . . . . .	74
4.3	Fluoroscopic image understanding . . . . .	77
4.3.1	2D image segmentation . . . . .	77
4.3.2	Landmark extraction . . . . .	81
4.4	2D-3D multiple bone reconstruction . . . . .	83
4.4.1	Coarse model initialization . . . . .	83
4.4.2	Optimization . . . . .	85
4.5	Visualization and clinical workflow integration . . . . .	91
4.5.1	3D-3D model-based registration . . . . .	91
4.5.2	Plane adjustment . . . . .	91
4.5.3	Visualization . . . . .	93
<b>5</b>	<b>Experiments &amp; Results</b>	<b>95</b>
5.1	Data acquisition . . . . .	95
5.2	Overview and purpose of the experiments . . . . .	98
5.3	Evaluation metrics . . . . .	100
5.4	3D bone segmentation . . . . .	103
5.4.1	Training data and 3D-SSM creation . . . . .	103
5.4.2	Evaluation method . . . . .	104
5.4.3	Evaluation 3D-CNN segmentation . . . . .	104
5.4.4	Evaluation of the 3D-SSM segmentation . . . . .	106
5.4.5	Evaluation of the new combined 3D-SSM segmentation approach . . . . .	106
5.4.6	Results . . . . .	106
5.5	Fluoroscopic image understanding . . . . .	109
5.5.1	Training data and network parameters . . . . .	109

5.5.2	2D segmentation accuracy . . . . .	110
5.5.3	Viewport restriction . . . . .	110
5.5.4	Landmark detection . . . . .	111
5.5.5	Results . . . . .	111
5.6	2D-3D bone reconstruction . . . . .	112
5.6.1	Training data and 3D-SSM parameters . . . . .	113
5.6.2	Performance evaluation under ideal conditions . . . . .	114
5.6.3	Influence of the 3D-SSM . . . . .	114
5.6.4	Influence of initialization . . . . .	114
5.6.5	Influence of the 2D segmentations . . . . .	115
5.6.6	Number of projection images . . . . .	115
5.6.7	Results . . . . .	115
5.7	System evaluation . . . . .	117
5.7.1	Adjustment of standard planes . . . . .	117
5.7.2	Overall error estimation . . . . .	118
5.7.3	Clinical evaluation . . . . .	119
5.7.4	Results . . . . .	119
<b>6</b>	<b>Discussion</b>	<b>125</b>
6.1	Intra-operative assistance system for ankle surgery . . . . .	125
6.2	3D bone segmentation . . . . .	126
6.3	2D fluoroscopic image understanding . . . . .	128
6.4	2D-3D reconstruction . . . . .	130
6.5	Clinical integration and visualization . . . . .	131
6.6	Conclusion . . . . .	133
6.7	Summary of contributions . . . . .	135
6.8	Future work . . . . .	136
<b>7</b>	<b>Summary</b>	<b>139</b>
<b>8</b>	<b>Zusammenfassung</b>	<b>141</b>
	<b>Bibliography</b>	<b>144</b>
	<b>Publications</b>	<b>162</b>
	<b>Curriculum Vitæ</b>	<b>165</b>
	<b>Acknowledgments</b>	<b>167</b>



## LIST OF FIGURES

1	Clinical workflow . . . . .	1
2	Appearance of the ankle joint in C-arm images . . . . .	2
3	Technical overview . . . . .	5
4	The upper ankle joint with syndesmotic ligaments . . . . .	8
5	Tubercles of the ankle joint at the level of the incisura . . . . .	9
6	Radiographic 2D measurements of the ankle joint on fluoroscopic images . . . . .	11
7	Radiographic measurements of the ankle joint on 3D C-arm images . . . . .	12
8	Weber fracture classification . . . . .	14
9	OR setup during ankle surgery . . . . .	16
10	Computed Tomography systems . . . . .	23
11	Network architectures . . . . .	30
12	Different model types . . . . .	42
13	Non-convex objective function . . . . .	43
14	Reconstruction strategies . . . . .	44
15	Medical workflow . . . . .	54
16	Technical overview . . . . .	56
17	Hierarchical representation of different sources of error . . . . .	60
18	3D C-arm images of the ankle joint after fibular reduction . . . . .	62
19	3D U-Net architecture . . . . .	68
20	Data augmentation strategies on 3D C-arm images . . . . .	72
21	Cropping and image translation for data augmentation . . . . .	73
22	The segmentation algorithm as a whole . . . . .	76
23	Fluoroscopic projection images of the ankle joint . . . . .	77

24	2D U-Net architecture . . . . .	78
25	2D segmentation of fluoroscopic images of the upper ankle joint . . . . .	80
26	Data augmentation methods on fluoroscopic images . . . . .	80
27	Landmarks of fluoroscopic images . . . . .	82
28	Acquisition of the 3D image $V$ and projection images $B$ . . . . .	84
29	2D-3D initialization procedure . . . . .	85
30	2D-3D reconstruction overview . . . . .	86
31	3D-SSM projection and silhouette generation . . . . .	87
32	Sampling of profile points . . . . .	88
33	Automatic standard plane adjustment . . . . .	92
34	Visualization of the reconstruction results . . . . .	94
35	C-arm volumes of injured ankle . . . . .	96
36	Standard projections of the ankle joint . . . . .	98
37	Overview of conducted experiments . . . . .	99
38	Voxel-/pixel-based overlap metrics for shape comparison . . . . .	101
39	2D slices relevant to the clinical evaluation . . . . .	102
40	SSM initialization scheme . . . . .	105
41	3D segmentation error evaluation . . . . .	107
42	Segmentation quality of the 3D-CNN . . . . .	108
43	Visual comparison of different 3D-CNN predictions . . . . .	108
44	2D-CNN segmentation evaluation . . . . .	112
45	Comparison of different viewport restrictions . . . . .	113
46	2D-3D reconstruction accuracy of tibia and fibula . . . . .	116
47	Visual inspection of the 2D-3D reconstruction using 2D-CNN predictions . . . . .	116
48	Influence of the number of images . . . . .	117
49	Two configurations for the overall evaluation . . . . .	118
50	Visual comparison of the overall pipeline . . . . .	122
51	Mean landmark variation of tibia and fibula . . . . .	126
52	Accurate 3D-SSM segmentation versus the outlier with worst performance . . . . .	128
53	Visual inspection of the 2D-CNN segmentation . . . . .	129
54	Reconstruction results with a varying number of projection images . . . . .	131



LIST OF TABLES

1	Standard plane definition for the ankle joint . . . . .	18
2	Overview on different approaches for evaluation of the reduction result . . .	55
3	SSM parameter selection . . . . .	104
4	Network hyperparameter selection for the 3D U-Net . . . . .	105
5	Comparison of different metrics for different bones and image segments . .	106
6	Network parameter selection for the 2D U-Net . . . . .	109
7	2D-CNN segmentation evaluation . . . . .	111
8	2D landmark extraction . . . . .	112
9	Reconstruction parameter selection . . . . .	114
10	Time consumption for 2D-3D reconstruction . . . . .	117
11	Standard plane deviation . . . . .	119
12	Quantitative evaluation of the whole pipeline for configuration 1 . . . . .	120
13	Quantitative evaluation of the whole pipeline for configuration 2 . . . . .	121
14	Mean time consumption: Pipeline methods . . . . .	121
15	Clinical evaluation performed by two orthopedic surgeons . . . . .	123

---

## LIST OF ACRONYMS

<b>ADAM</b>	<i>Adaptive Moment Estimation Algorithm</i> .....	71
<b>ALARA</b>	<i>As Low As Reasonably Attainable</i> .....	26
<b>AO</b>	<i>Arbeitsgemeinschaft für Osteosynthesefragen</i> .....	16
<b>AP</b>	<i>Anterior-posterior</i> .....	10
<b>ASD</b>	<i>Average symmetric surface distance</i> .....	100
<b>CAOS</b>	<i>Computer-Assisted Orthopedic Surgery</i> .....	26
<b>CBCT</b>	<i>Cone-beam Computed Tomography</i> .....	23
<b>CMAES</b>	<i>Covariance Matrix Adaptive Evolutionary Strategy</i> .....	45
<b>CNN</b>	<i>Convolutional Neural Network</i> .....	28
<b>CPD</b>	<i>Coherent Point Drift</i> .....	34
<b>CT</b>	<i>Computed Tomography</i> .....	11
<b>DRR</b>	<i>Digital Reconstructed Radiograph</i> .....	44
<b>ELU</b>	<i>Exponential Linear Unit</i> .....	69
<b>FCN</b>	<i>Fully Convolutional Network</i> .....	29
<b>FOV</b>	<i>Field of View</i> .....	21
<b>GAN</b>	<i>Generative Adversarial Network</i> .....	137
<b>GPA</b>	<i>General Procrustes Analysis</i> .....	33
<b>HU</b>	<i>Hounsfield Unit</i> .....	22
<b>ICP</b>	<i>Iterative Closest Point</i> .....	34
<b>IOU</b>	<i>Intersection Over Unit</i> .....	100

---

<b>MDL</b>	<i>Minimum description length</i> .....	34
<b>MPR</b>	<i>Multi-planar reconstruction</i> .....	17
<b>MRI</b>	<i>Magnetic Resonance Imaging</i> .....	11
<b>MITK</b>	<i>Medical Imaging and Interaction Toolkit</i> .....	13
<b>OR</b>	<i>Operating room</i> .....	20
<b>ORIF</b>	<i>Open reduction and internal fixation</i> .....	16
<b>PCA</b>	<i>Principal Component Analysis</i> .....	34
<b>PDM</b>	<i>Point Distribution Model</i> .....	32
<b>ReLU</b>	<i>Rectified Linear Unit</i> .....	69
<b>RMS</b>	<i>Root mean square</i> .....	71
<b>ROI</b>	<i>Region of interest</i> .....	22
<b>SNR</b>	<i>Signal-to-Noise ratio</i> .....	20
<b>SSM</b>	<i>Statistical Shape Model</i> .....	32
<b>SSIM</b>	<i>Statistical Shape and Intensity Model</i> .....	32
<b>SVD</b>	<i>Singular Value Decomposition</i> .....	35
<b>THA</b>	<i>Total hip arthroplasty</i> .....	27
<b>VO</b>	<i>Virtual Object</i> .....	26

## 1.1 Motivation

Fractures of the ankle joint rank among the most common fractures of the lower extremities and can be associated with injury of the syndesmotic complex (WARNER et al. 2015).

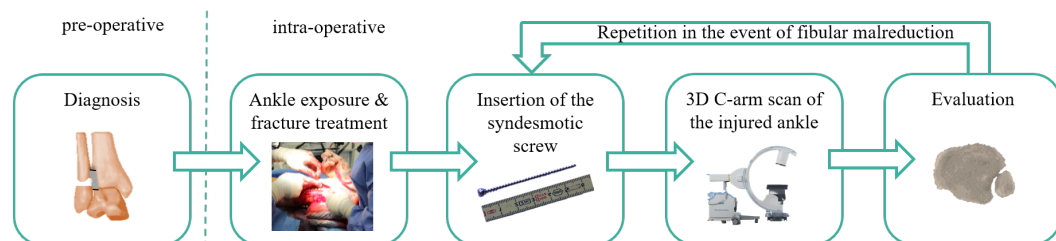


Figure 1: Clinical workflow: Following pre-operative diagnosis, the ankle is exposed to examine the fracture and the syndesmotic complex. The fracture is re-positioned and fixated with metal implants e.g. screws and a plate. A syndesmotic screw is inserted to connect the tibia and fibula so that the syndesmotic complex can heal. Afterward, a C-arm scan is usually acquired to evaluate the reduction result. If a malreduction is identified, the screw position is modified and the scan is repeated.

Usually, patients experienced accidental forced external rotation of the ankle causing a fracture of the fibula bone (SAXENA 2012). Diagnosed with a pre-operative X-ray (as illustrated in the clinical workflow in Fig. 1), the fractured fibula is carefully restored and relocated. Hence, this process is referred to as reduction. It is widely assumed that incorrect re-positioning, termed malreduction, of the fibula in the incisural notch of the tibia is a major reason for revision surgery. An undetected malreduction can lead to early chondral degeneration with osteoarthritis (VETTER et al. 2019a).

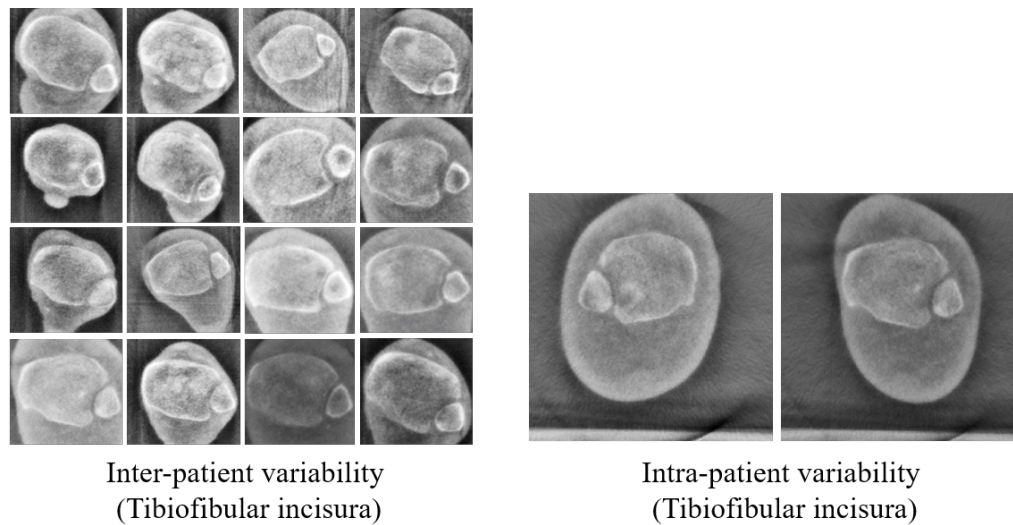


Figure 2: Appearance of the ankle joint in C-arm images: Inter-individual variation for 16 patients (left) and the intra-individual variation for one patient (right) shown on the axial slice at the level of the incisura.

Recent advances in 3D imaging technology allow novel insights into the bone anatomy. As shown in Fig. 1, an intra-operative C-arm device can be used to acquire a 3D image of the ankle immediately after screw insertion and carefully evaluate the reduction result afterward. FRANKE et al. 2012 showed that fibular malreduction was detected more often when 3D imaging was used for acute unstable syndesmotic injuries but remained undetected with 2D standard fluoroscopy. This is accompanied by a significantly increased need for intra-operative revisions. If a malreduction is identified, the reduction of the fibula in the incisural notch is modified before image acquisition is repeated. Thus, malreduction and implant malposition can be detected more easily and post-operative *Computed Tomography* (CT) becomes obsolete.

Intra-operative evaluation of the reduced fibula is a time-consuming process that is prone to errors since bone reduction is patient-specific and no reliable gold standard exists for reducing the fibula to achieve an optimal result. In a worst case scenario, intra-operative revision is repeated several times and significantly increases the intervention time as well as the cumulative radiation dose. Even with 3D imaging, the reduction process remains very challenging as the high inter-individual variability of the position and shape of the of the tibiofibular joint impedes verification of a correct reduction result. However, the intra-individual variability between symmetrical left and right ankles is significantly lower than the inter-individual variability (ELGAFY et al. 2010; VETTER et al. 2019a) as shown in Fig. 2. A study by MUKHOPADHYAY et al. 2011 suggests that si-

multaneous visualization of both ankles, e.g. by acquiring a pre-operative CT, provides the most reliable verification method, especially for difficult cases or inexperienced surgeons. This elicits safer and more assured treatment for the patient and also gives the reassurance of standardized quality. Despite this, standard clinical protocols typically suggest that only two pre-operative X-ray radiographs of the injured ankle be used for diagnosis and a pre-operative 3D *Magnetic Resonance Imaging* (MRI) or CT scan is usually not available. An additional intra-operative 3D C-arm scan of the contralateral side is not routinely acquired for the intervention due to concerns regarding radiation exposure for the surgical team and the patient and in order to avoid time-consuming and costly procedures.

Obviating the need for a second 3D scan and providing the option of using the information of the contralateral side could improve the surgical outcome, decrease the number of intra- and post-operative revisions and the patient's time under anesthetic while simultaneously reduce the cumulative radiation dose. Surgeons wish to have more useful information for assessing the syndesmotic region since the quantitative ankle measurements have a high inter-individual variability and are not standardized for the procedure.

To date, there is no intra-operative assistance available to guide surgeons during the reduction process. This thesis aims to develop methods to provide the surgeon with additional visual aid gained from the uninjured ankle of the patient without the need for an additional 3D scan while carefully considering intra-operative requirements.

## 1.2 Objectives

The main focus of this thesis is computer-assisted contralateral side comparison verifying reduction results during ankle surgeries. With the elaboration of software-based methods, an assistance system for surgeons, which can also be integrated into mobile C-arm systems, should be designed. The main objective of this thesis was the development of a system that should meet the following requirements:

- **Requirements regarding the benefits for the patient:** The assistance system should be tailored to help the surgeon in his decision process and improve the quality of the procedure in order to avoid post-operative complications or revision surgery which would potentially be harmful to the patient. The software should work without requiring an additional 3D scan of the contralateral side which prevents the patient from being exposed to additional radiation. The methods should be as fast as possible in order to decrease the time the patient would potentially spend under

anesthetic.

- **Requirements regarding the benefits for the surgeon:** The developed methods should offer a high level of automation and obviate any additional interaction steps for the surgeon in order to reduce the risk of contamination. The proposed workflow should limit the number of additional image acquisitions so that the surgeon is not exposed to additional radiation. The developed software should be compatible with the clinical workflow and meet the requirements of the current clinical protocol. The assistance system should visualize the extracted information to the surgeon in a concise way and easy to interpret.
- **Further technical requirements regarding the usability and accuracy:** The computed overall result should be as precise as possible. An error of less than 1 mm is regarded for achieving the required clinically relevant accuracy. All methods applied to C-arm images of the injured ankle should be robust for different fracture types and metal artifacts, which are often visible in intra-operative images. The whole pipeline should be implemented as a modular system so that the individual methods can be interchangeable and evaluated independently of each other.

### 1.3 Approach

The stated clinical need for a contralateral side comparison of the ankle joint was approached by the conceptual design and algorithmic development of an image-based assistance system. Fig. 3 shows an overview of the technical components and their integration in the clinical workflow. In order to develop a prototype capable of contralateral side comparison and to achieve the aforementioned objectives, the following steps were conducted which are visualized in Fig. 3:

- (A) A precise 3D model representation of the injured ankle was created by applying *Statistical Shape Model* (SSM) segmentation on the 3D C-arm image (Section 4.2).
- (B) Relevant image information was extracted from fluoroscopic projection images of the contralateral uninjured ankle using an approach based on deep learning (Section 4.3).
- (C) An automatic coarse initialization was performed to align the 3D model with the projection images (Section 4.4).
- (D) The 3D model and the 2D image information were combined for a 2D-3D multi-bone reconstruction of the uninjured ankle (Section 4.4).



- (E) The contralateral reconstruction result was matched with the injured ankle model using 3D-3D registration to provide an interpretable visualization of both 3D surfaces (Section 4.5).

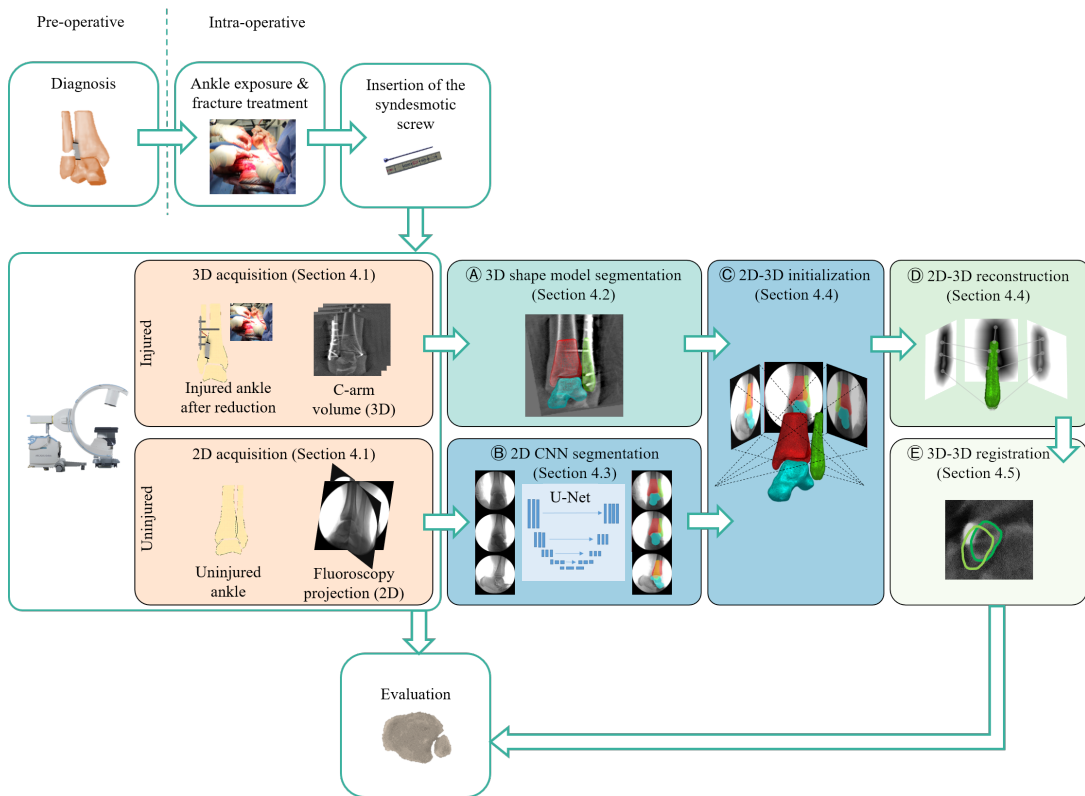


Figure 3: Technical overview: Components of the proposed automatic contralateral side comparison embedded in the clinical workflow

## 1.4 Outline

The remainder of this thesis is structured as follows: Chapter 2 provides the relevant information on the medical and technical background. Thereafter, the state of the art of assistance systems for orthopedic surgery, medical image segmentation and 2D-3D reconstruction is summarized in chapter 3.

All methods developed in the scope of this thesis are presented in chapter 4. An overview of the chosen approach and details on design choices and error propagation is given in section 4.1. Following on from the approach, section 4.2 introduces model-based 3D segmentation of the injured ankle using statistical shape models and convolutional neural networks. In section 4.3, the semantic interpretation of fluoroscopic 2D images using convolutional neural networks and basic image processing techniques are

investigated. Section 4.4 presents methods for 2D-3D multiple bone reconstruction of the ankle joint. Section 4.5 combines all of the developed methods and provides an insight into the overall integration of all components along with methods for the final visualization.

In chapter 5, a systematic evaluation of the different experiments conducted for this thesis is elaborated. Chapter 6 discusses the findings of the thesis, draws final conclusions and gives directions for future work. The thesis is concluded with a summary of the work presented herein (Chapter 7).

The following chapter is intended to describe the required background knowledge of both, the medical and the technical field. In the first section (2.1), background information on the anatomy of interest and the surgical intervention is presented. In the second section (2.2), insights are given on relevant radiological imaging modalities helpful for understanding the content of this thesis.

## **2.1 Medical background**

The human body has a variety of bilateral anatomically symmetric regions regarding bones and joints. Among others, the wrist, proximal humerus, proximal femur, knee, ankle and foot would all be suitable for contralateral side comparison. In this work, the ankle is the chosen joint of interest, since related interventions are highly challenging and frequent. In this chapter, the anatomy of the ankle joint (2.1.1) and the standard radiographic measurements (2.1.2) are described. Furthermore, an overview is given of frequent injuries (2.1.3) and corresponding interventions (2.1.4).

### **2.1.1 Anatomy of the upper ankle joint**

The ankle joint is defined as the joint connecting the foot with the lower limb. It involves three osseous structures, namely the distal part of the tibia, the distal part of the fibula and the talus. As shown in Fig. 4, the medial malleolus, which is a part of the tibia, and the lateral malleolus, which is composed of the fibula, together form a deep recess for the talus (STANDRING et al. 2015). Tibia and fibula are two adjacent bones that are connected by multiple tibiofibular ligaments. This complex is referred to as the syndesmosis.

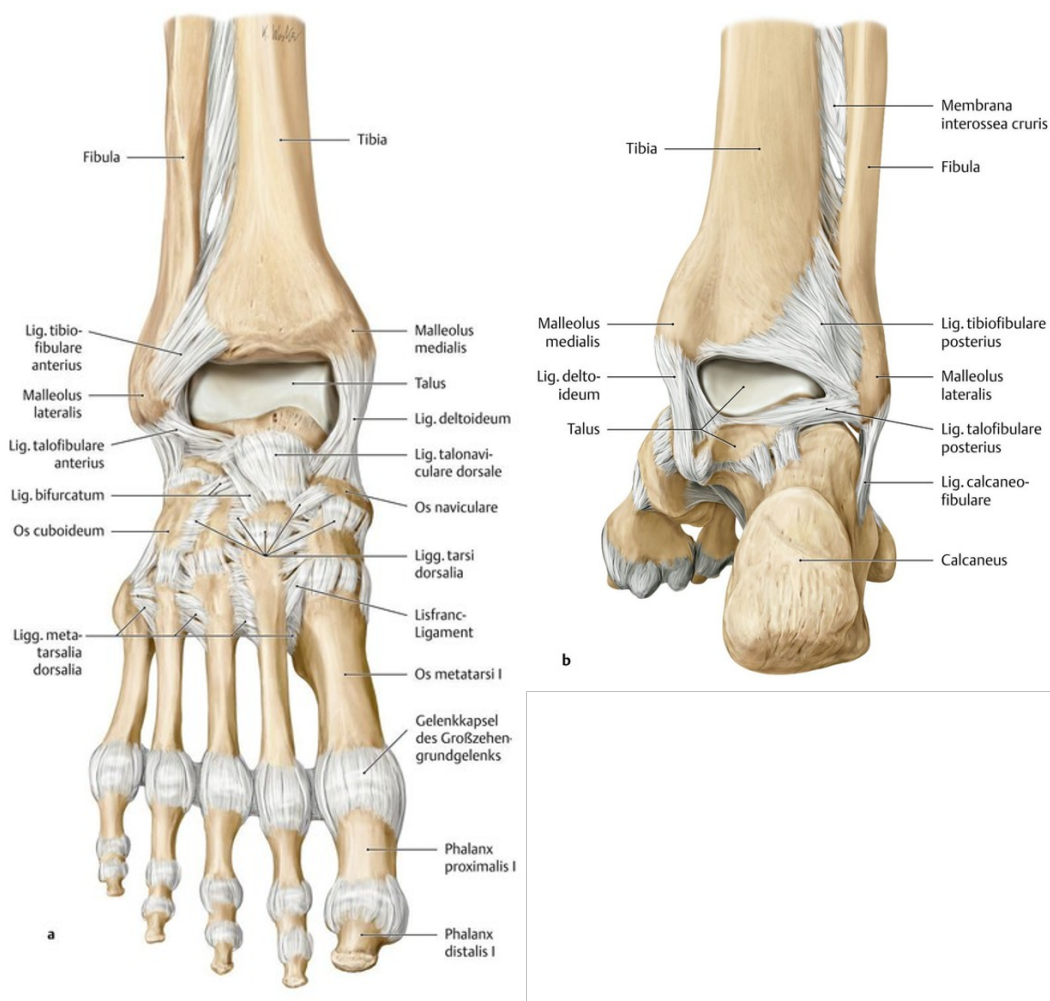


Figure 4: The upper ankle joint with syndesmosic ligaments: a) shows the ventral (front to back) and b) shows the dorsal (back to front) view. Reprinted from SCHÜNKE et al. 2018 with kind permission of Thieme.

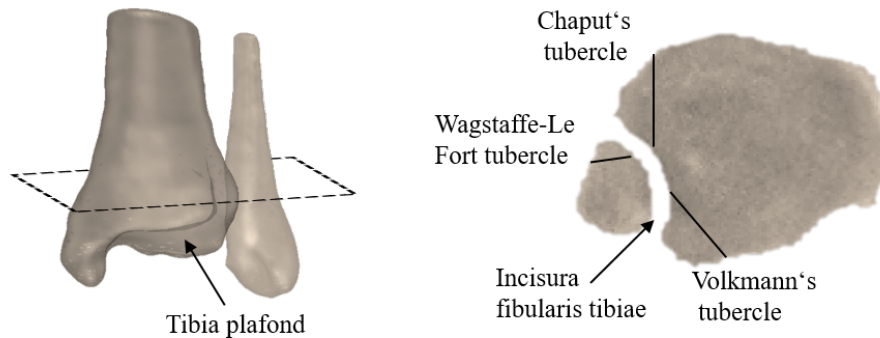


Figure 5: Tubercles of the ankle joint at the level of the incisura

The tibia is the second largest bone of the body. It bears the weight transmitted through the femur and knee joint at the proximal end and together with the fibula comprises the ankle joint at the distal end (STANDRING et al. 2015). The tibia builds a concave triangle at the level of the syndesmosis (cf. Fig. 5) referred to as the incisura fibularis tibiae, henceforth abbreviated with incisura. Described by HERMANS et al. 2010 for adult patients, the lateral ridge of the tibia bifurcates caudally to the incisura approx. 6 to 8 cm above the lowermost level of the talocrural joint. One of the margins becomes the anterior tubercle (Chaput) whereas the other margin builds the posterior tubercle (Volkman). The Chaput's tubercle is more distinctive and prevents the fibula from forward slipping. The posterior tubercle forms the pivot point during external rotation, in which the fibula rotates along its main axis.

The fibula is the much thinner but likewise tubular counterpart of the tibia (STANDRING et al. 2015). The ridge on the medial aspect of the distal fibula also bifurcates into two margins and forms a respective convex triangle located right above the articular facet of the lateral malleolus (HERMANS et al. 2010). The anterior tubercle (Wagstaffe-Le Fort) and the, rather negligible, posterior tubercle form the base of the fibular triangle (illustrated in Fig. 5). The apices of the fibula and the tibia triangle are situated at the same level. The shape of the fibula smoothly matches the shape of the incisural notch of the tibia. ELGAFY et al. 2010 found that the incisura can be divided into two types, which is either deep as in 67% of the cases or more shallow as in all other cases. Different studies showed that a rather shallow incisura might increase the risk of recurrent ankle sprains (YILDIRIM et al. 2003) or fracture dislocation (EBRAHEIM et al. 1998).

The syndesmosis is a ligamentous complex consisting of the following components (labeled in Fig. 4): The anterior inferior tibiofibular ligament (AITFL) ranges from the

Chaput's tubercle to the anterior distal fibula. The posterior inferior tibiofibular ligament (PITFL) is located on the Volkmann's tubercle, and connects at the posterior part of the distal fibula (SAXENA 2012). Between fibula and tibia, 0.5 to 2 mm above the tibio-talar joint, the interosseous ligament blends with the interosseous membrane. All these ligaments together provide stability to the ankle joint.

The ankle joint not only bears most of the bodyweight but is also essential for human gait. The human skeleton provides different sorts of joint types differentiated by their size, location and their degrees of freedom. The ankle joint is a hinge joint which implies that the joint movement is restricted to one plane. The rotation axis is dynamic with main movements of plantar flexion (about  $10^\circ$ ) and dorsiflexion (about  $30^\circ$ ) (STANDRING et al. 2015). Other movements like eversion and inversion of the whole extremity are caused by e.g. the subtalar joint, which is formed by talus and calcaneus. The ankle allows a small degree of fibula excursion in healthy subjects. The syndesmotic ligaments maintain the integrity between tibia and fibula at the distal end by sustaining translational, axial and rotational forces. However, if two or more of those ligaments are compromised, the ankle joint may face significant mechanical laxity.

### 2.1.2 Radiographic anatomical parameters

The ankle joint can be described by different anatomical measurements although their significance and correct definition remains controversial among different studies (NIELSON et al. 2005; BEUMER et al. 2004; EBRAHEIM and ELGAFY 2003).

2D measurements of anatomical parameters can be either performed on plain radiographs or by using intra-operative standard fluoroscopy (FRANKE et al. 2012). Fluoroscopy is an imaging modality that is often used in surgery since it allows to show continuous low-dose X-ray images. There are specific view directions defined for 2D image acquisition in clinical routine. In general, *Anterior-posterior* (AP) describes the view from front to back and *lateral* describes the view from left to right of the patient. SCHNETZKE et al. 2016 describe three 2D standard projections for diagnostics of the ankle joint: *mortise*, AP and *lateral* (shown in Fig. 6). AP is obtained by placing the foot with toes pointing upward and by directing the central beam of the C-arm or X-Ray source towards the joint. In this view, the anterior tibial tubercle overlaps with the fibula and the talus is slightly externally rotated. HAHN and C. L. COLTON 2018 describe the so-called ankle mortise as the articulation of three bones, in particular, the distal aspect of the tibia, the distal aspect of the fibula and the talus. Hence, to obtain the *mortise* view, the foot is internally rotated by approximately  $20^\circ$ . With the talus in its true AP position, the talar dome is fully visible and the clear spaces of the lateral and medial malleoli are clearly exposed.

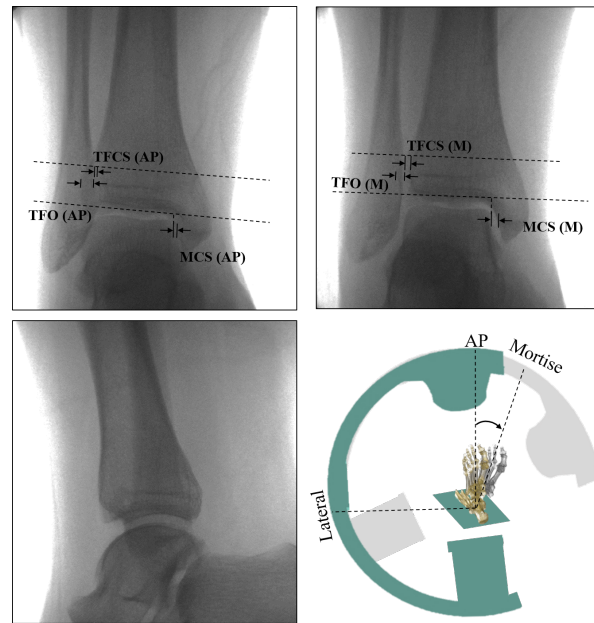


Figure 6: Radiographic 2D measurements of the ankle joint on fluoroscopic images: The top row shows the AP and mortise (M) view with medial clear space (MCS), tibiofibular clear space (TFCS) and tibiofibular overlap (TFO). The bottom row shows the lateral view and an illustration of different C-arm acquisition angles. For mortise view, the ankle is tilted by approximately 20°.

The *lateral* view is obtained sideways by superposition of both malleoli.

Different quantitative measurements can be found in literature to assess the ankle joint in 2D images. E.g. HARPER et al. 1989 defined the following for *mortise* and AP view (indicated in Fig. 6):

- Medial clear space (MCS)
- Tibiofibular clear space (TFCS)
- Tibiofibular overlap (TFO)

HARPER et al. 1989 concluded that the tibiofibular clear space is a more reliable measurement for diagnosis compared to other measurements since it is independent of talar rotation.

Measurements in 3D are obtained most commonly on *Computed Tomography* (CT) or *Magnetic Resonance Imaging* (MRI) scans (NIELSON et al. 2005). Tomographic imaging has the advantage that interpretation is not biased by the superposition of bones. Different methods exist to estimate quantitative measurements of the syndesmosis. Most of the parameters are measured on a selected image plane, approximately 10 mm proximal to the tibial plafond. E.g. MUKHOPADHYAY et al. 2011 measure the anterior and posterior

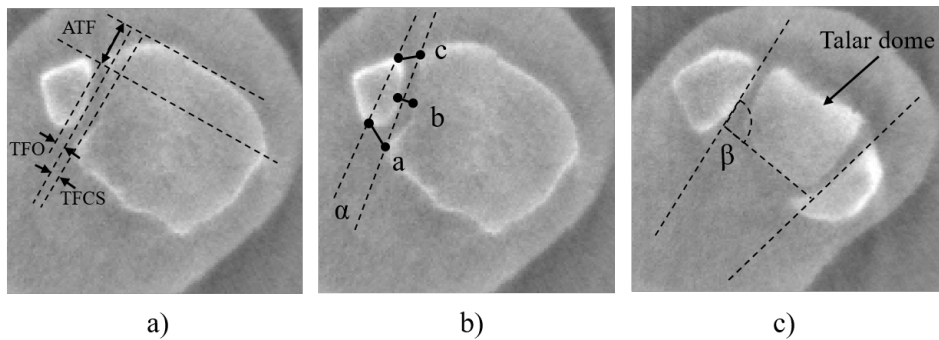


Figure 7: Radiographic measurements of the ankle joint on 3D C-arm images: Approximately 10 mm proximal the tibial plafond a) shows the TFO (tibiofibular overlap), ATF (anterior tibiofibular interval) and TFCS (tibiofibular clear space). b) shows distance measurements a, b and c and the angle between tibia and fibula. c) Approximately 6 mm distal the tibial plafond, at the talar dome level, the angle between tibia and fibula is measured to assess the rotation of the fibula. Measurements in a) are defined after DIKOS et al. 2012 and b) and c) are defined after NAULT et al. 2013.

clear space of the tibia and fibula on that particular plane and compare the deviation. On the same plane, DIKOS et al. 2012 measure the TFO (tibiofibular overlap), ATF (anterior tibiofibular interval) and TFCS (tibiofibular clear space) as shown in Fig. 7a. In addition, they assess the fibular rotation by measuring the angle between two lines, in particular, the tangent between both tibia tubercles and a line connecting both fibula tubercles (cf.  $\alpha$  in Fig. 7b). ELGAFY et al. 2010 conducted a study based on 100 subjects and suggested to use four standardized points at the level of the distal tibiofibular syndesmosis, 9-12 mm above the tibia plafond, for both, the shallow and deep incisura type. As shown in Fig. 7b, the posterior measurement is defined by two points, more specifically, one point is located at the medial edge of the fibula and the second represents the closest point at the lateral edge of the posterior tibial tubercle (annotated with a). The anterior measurement is defined by a point located at the tip of the anterior tibial tubercle and the nearest point of the fibula (annotated with c).

NAULT et al. 2013 adapted and extended the existing measurement techniques to an overall of eight measurements on the level 9.45 mm proximal to the tibial plafond. Similar to ELGAFY et al. 2010, they annotated the outermost anterior point of the incisura and estimated the length to the closest point of the fibula as well as connected the most posterior point of the incisura with the outermost posterior point of the fibula. Additionally, they measured the distance of both bones at the center of the incisura (annotated as  $b$  in Fig. 7b). They connected both fibula and both tibia points to measure the length of the fibula and the incisura as well as the angle between both lines (cf. Fig. 7b). Different



from other measurement techniques, NAULT et al. 2013 also measured the angle of the two malleoli on the level of the talar dome (annotated as  $\beta$  in Fig. 7c) to assess the fibular rotation. VETTER et al. 2019b demonstrated in their study that the most convenient and reproducible level to measure the fibular rotation is 4 mm or 6 mm below the talar joint line.

All of the conducted studies in literature that included ankle pairs have in common that a high inter-individual but a significantly lower intra-individual anatomical variability of the syndesmotic measurements was observed among different patients (VETTER et al. 2019a). DIKOS et al. 2012 and MUKHOPADHYAY et al. 2011 investigated the variation between contralateral ankles in different patients and observed no significant differences for most of the parameters. Dikos et al. observed a range of  $0.5^\circ$  to  $29.5^\circ$  for the internal rotation (annotated as  $\alpha$  in Fig. 7) of the fibula with a mean of  $12.7^\circ \pm 6.7^\circ$  for different healthy subjects. Comparing the fibular rotation of the same patient, a mean deviation of  $2.9^\circ \pm 1.8^\circ$  (max  $6.5^\circ$ ) was found between the left and right ankle.

In summary, different measurements are defined to assess the ankle joint in 2D and 3D images. But due to the high anatomical variation, some of them are considered unreliable for a precise reduction control in ankle fractures. All of the studies agreed on the hypothesis that intra-individual variability is most often significantly lower than inter-individual variability.

### 2.1.3 Syndesmotic injuries and diagnosis

One of the most common causes of syndesmotic injuries is an extreme forced rotational dislocation of the foot. Sport activities are considered a risk factor since those injuries are often reported in sports like skiing, ice hockey, football or basketball (SAXENA 2012). Sometimes, the injury is pure ligamentous but more frequently associated with fractures of the malleoli.

To decide on an adequate treatment, most frequently, the Danis-Weber classification after WEBER and C. COLTON 1991 is used to categorize different syndesmotic injuries related to the fibula fracture level (indicated in Fig. 8):

- *Weber A*: A fracture of the lateral malleolus occurs distal to the syndesmosis' ligaments. The medial malleolus is usually intact or sheared. In most cases, the tibiofibular syndesmosis remains intact and the joint stable.
- *Weber B*: A fracture of the lateral malleolus occurs right at the level of the syndesmosis complex. The fracture might extend upwards or backwards. The syn-

---

<sup>1</sup>Medical Imaging and Interaction Toolkit (MITK) workbench, [www.mitk.org/wiki/MITK](http://www.mitk.org/wiki/MITK)

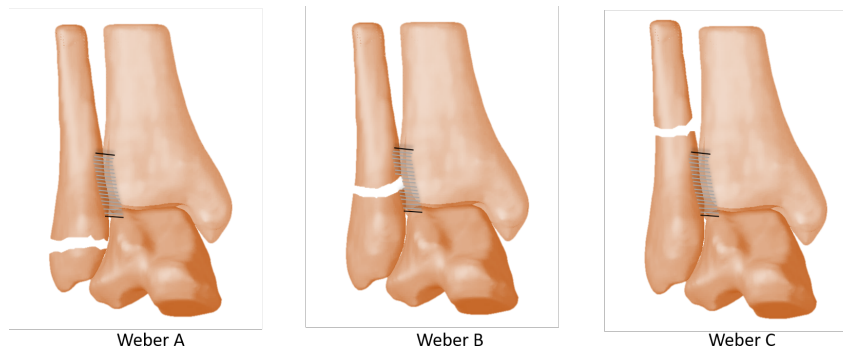


Figure 8: Weber fracture classification: Classifications are based on the location of the fibular fracture in relation to the syndesmosis. Created using MITK<sup>1</sup>.

desmosis complex can be either intact but also partially torn or detached without widening of the distal tibiofibular articulation. A fracture of the medial malleolus might be possible and involved ligaments might be torn. This might further reduce stability.

- *Weber C*: A fracture occurs proximal to the syndesmosis above the ankle joint between the syndesmosis and the head of the fibula. The syndesmosis is most likely disrupted with widening of the distal tibiofibular articulation. Smaller to fairly large fragments can be present. Furthermore, a medial malleolus fracture or an injury of other ligaments is present.

Additionally, two other types of injuries are often associated with unstable syndesmosis:

- *Maisonneuve*: Maisonneuve refers to a fracture of the upper third part of the proximal fibula accompanied with a disruption of the distal tibiofibular syndesmosis (KALYANI et al. 2010). The fracture is mainly caused by a pronation combined with external rotation mechanism and should be treated operatively.
- *Ankle sprain*: A distal tibiofibular syndesmostic rupture occurs without the presence of a bone fracture (HERMANS et al. 2010). Joint instability is believed to originate in a widening of the ankle mortise.

FRANKE et al. 2012 reported that 11% of all investigated 2286 ankle fractures had an unstable syndesmosis. 60.2% were fractures of type Weber C, 20.7% Maisonneuve fractures, 14.7% of type Weber B and the remaining 4.4% were isolated ankle sprains without a fracture. HERMANS et al. 2010 observed that a syndesmostic injury occurred in about 50% of type Weber B and in all type Weber C fractures. An injury of the distal tibiofibular syndesmosis occurs without any fractures in approximately 1-11% of all ankle sprains.

According to NIELSON et al. 2005, a syndesmotic injury is the single most predictive factor for long-term disability and chronic ankle pain if mis- or undiagnosed regardless of the grade of injury.

The level of the fibula fracture can be identified quite easily using standard radiographs. More difficult is the identification of a syndesmotic instability. In general, SAXENA 2012 suggest an intervention if an indication is given for two or more syndesmotic ligaments being compromised regardless of the fibula position. Furthermore, an indication is given if the syndesmotic widening is greater than 2 mm and the medial space widening of the ankles' mortise is greater than 4 mm. On the contrary, DIKOS et al. 2012 concluded that a variation of 2 mm can be found among healthy subjects and thus does not serve as a reliable threshold for identifying abnormalities.

Different researchers claimed that 2D measurements on plain radiographs are not sufficient for assessing the syndesmosis (DIKOS et al. 2012; ELGAFY et al. 2010; MUKHOPADHYAY et al. 2011; HARPER et al. 1989) and that a CT scan is more sensitive in detecting smaller syndesmotic widening. EBRAHEIM et al. 1997 showed that a widening of up to 3 mm can only be detected by a CT scan. MUKHOPADHYAY et al. 2011 claimed that even a minor imperfect placement of the fibula, referred to as malreduction may result in large changes to the tibiotalar joint. Thus, diagnostic imaging needs to be able to detect small differences reliably. BUCKLEY et al. 2011 further showed that a fibular rotation is even more difficult to detect using standard fluoroscopy and observed that external rotations of up 30° can remain undetected. Different authors conclude that the acquisition of the contralateral side for comparison is the most reliable source of information to assess the syndesmosis (MUKHOPADHYAY et al. 2011; DIKOS et al. 2012; HARPER et al. 1989; EBRAHEIM et al. 1997) since the intra-individual variability is significantly smaller than inter-individual variability as described in section 2.1.2. All of the studies comparing both ankles acquired post-operative CT scans of the ankle. SZEIMIES et al. 2012 recommend contralateral side comparison of X-ray images in evidence of chronic instabilities. SCHREIBER et al. 2013 conducted a study on the feasibility of contralateral side comparison with 2D standard fluoroscopy. In their study, they only measured translational variation of the fibula compared to the uninjured counterpart and showed the benefits of the comparison.

In conclusion, the significance of pre- or intra-operative measurements to assess a potential injury of the syndesmosis and give a reliable diagnosis remains controversial due to the high anatomical variation. To this date, a volumetric comparison to the contralateral side of the patient is believed to give the best indication. However, a routine contralateral side comparison is still not implemented as standard procedure.

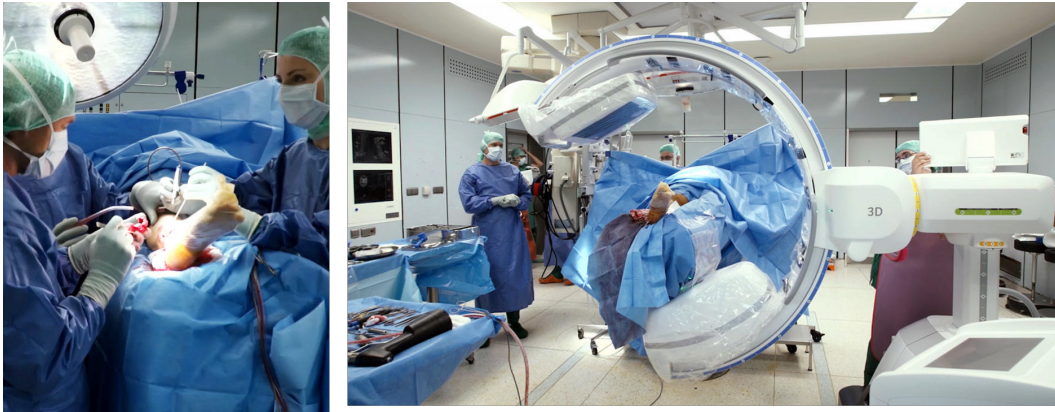


Figure 9: OR setup during ankle surgery: The OR is equipped with a Cios Spin<sup>®</sup> C-arm scanner for intra-operative imaging (BG Trauma center Ludwigshafen). Image courtesy of Siemens Healthcare GmbH.

#### 2.1.4 Surgical treatment

Pre-operatively, besides different manual rotation and stabilization tests performed by the trauma surgeon, diagnostic imaging is performed in clinical routine. As stated before, radiographs can help to diagnose fractures but often fail to identify occult syndesmosis injuries. Most commonly, the height of the fracture and the position of the involved bones are examined to decide whether a potential indication for syndesmotic instability is given.

**Procedure:** In case of positive indication, the standard procedure in syndesmotic reduction is called *Open reduction and internal fixation* (ORIF), which implies that the procedure is not minimal invasive but open and fracture fixation is done by inserting interior plates and screws. Fractures classified as Weber C always require a ORIF intervention. Intra-operatively, a small incision is performed after positioning and anesthetics of the patient to expose the fracture. Instability of the syndesmosis can be diagnosed during the intervention e.g. by using a bone hook to pull the distal fragment of the fibula under continuous fluoroscopy of the ankle joint in *mortise* view (*Hook test*) as described by FRANKE et al. 2012. Depending on fracture type and occurrence, different stabilization methods and metal implants are chosen for the treatment. The treatment of Weber B and Weber C fractures commonly involves the placement of a one-third tubular plate for osteosynthesis. Smaller fractures (e.g. if a larger Volkmann fragment is identified) are stabilized with a lag screw or less frequent using a band wire. For a more detailed description of different fracture treatments, the reader is referred to the *Arbeitsgemeinschaft für Osteosynthesefragen* (AO) manual of HAHN and C. L. COLTON 2018.

FRANKE et al. 2012 describe one possible approach for syndesmotic reduction after osteosynthesis as follows: A reduction clamp can be placed across the medial and lateral malleoli. After tightening the clamp, the reduction is verified by using fluoroscopy. Afterward, mortise reduction is performed by inserting one or two 3.5 mm small-fragment positioning screws from lateral to medial as well as from posterior to anterior at the angle of 30° applying standard AO techniques.

SORA et al. 2004 stated that the greatest problem in this kind of injury, even when treated surgically, is a correct alignment of the fibula, namely, the reduction of the fibula in the incisural notch. According to FRANKE et al. 2012, a so-called malreduction of the fibula and as a consequence thereof an intra-operative revision rate was observed in 32.7% of all cases although no malpositioning was visible using standard fluoroscopy. GARDNER et al. 2006 measured the distance between the fibula and the anterior and posterior facet of the incisura and regarded a discrepancy of 2 mm between the anterior and posterior measurement as malreduction. HERMANS et al. 2010 state that a widening of the ankle mortise by 1 mm already decreases the tibiotalar contact area by 42%, which may lead to joint instability and may be associated with early osteoarthritis.

**Intra-operative imaging:** FRANKE et al. 2012 recommend to use intra-operative 3D imaging in clinical routine and stated that a post-operative discovery of a malreduction may require an undesired revision surgery. Instead, they showed that a malreduction was revealed using an intra-operative 3D C-arm imaging device but remained overseen in standard fluoroscopy in 32.7% of the patients treated with an unstable syndesmosis. After fibular reduction, a mobile 3D C-arm can be used to acquire a 3D image of the reduction result to verify screw placement and bone alignment (the setup is shown in Fig. 9).

The resulting 3D image is displayed to the surgeon after applying a *Multi-planar reconstruction* (MPR) reconstruction. The method allows the generation of slices through the 3D image from arbitrary viewpoints and orientations (HANDELS 2009). With MPR, the volume is usually approached by three standard viewing planes. The *axial* plane allows the surgeon to view the volume from proximal to distal, the *sagittal* plane conventionally passes from lateral to medial and the *coronal* plane from ventral to dorsal.

The surgeon adjusts those MPR viewing planes of the reconstructed C-arm 3D volume to alternative viewing planes, hereinafter referred to as standard planes, suitable for an accurate evaluation. An overview of the standard plane adjustment for the ankle joint is given in Tab. 1. The axial plane is oriented orthogonal to the tibial shaft and centered between both malleoli at the level of the tibial plafond. The coronal standard plane is

Plane	Definition
Axial	Orthogonal to the tibia shaft at the level of the tibia plafond with the center in the middle between the medial and lateral malleoli.
Coronal	Orthogonal to the axial plane and building the angle bisector of tibia and fibula at 6 mm below the tibial plafond.
Sagittal	Orthogonal to the axial plane oriented 90° towards the sagittal plane.

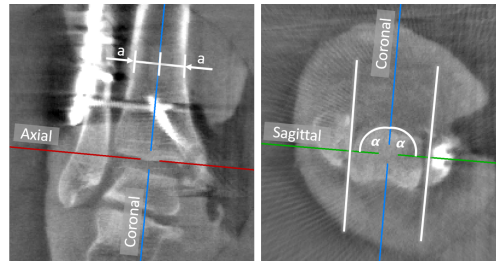


Table 1: Standard plane definition for the ankle joint

orthogonal to the axial plane and builds the angle bisector between the medial contour of the fibula and the tibia at the level of the talar dome approximately 6 mm below the tibial plafond. All planes intersect in the middle of the articular surface of tibial plafond. By scrolling along the axial plane the surgeon can analyze the reduction result at two levels - approximately 10 mm above the tibial plafond to examine the fibula in the incisural notch and 6 mm below to examine tibia and fibula at the talar dome level (the most proximal surface of the talus). If a malreduction is identified, the fibula position is modified and the screws are adjusted accordingly before the image acquisition is repeated until the surgeon is satisfied with the reduction result.

**Evaluation:** FRANKE et al. 2012 suggested to evaluate a potential malreduction on the 3D image based on following criteria:

1. In the ideal case, the anterior border of the fibula forms a harmonical elliptical line with the tibial plafond.
2. The fibula is positioned adequately in the incisura with respect to the syndesmosis topography and width.
3. The joint space width between the talus both malleoli is equal.
4. The correct rotation of the lateral malleolus (fibula) is reflected by the congruent positioning of the malleoli with respect to the talus.
5. In coronal direction, the width of the joint space between the talus and the malleoli is identical to the syndesmosis.

Recommendations for the intra-operative imaging of the contralateral side are given by FRANKE et al. 2012 although they state that an additional scan increases the cumulative radiation dose. HAHN and C. L. COLTON 2018 advise using contralateral side com-

parison at least for the pre-operative planning of complex cases, e.g. in case the fibula is multi-fragmentary.

### 2.1.5 Conclusion

The ankle joint comprises of tibia, fibula and talus as well as several ligaments and forms an important part of the human musculoskeletal system. Often, ankle joint fractures are accompanied by a compromised syndesmotic complex that requires an surgical intervention. The fracture reduction process benefits from 3D intra-operative imaging, however, syndesmotic reduction remains a challenging procedure. A widening of the ankle mortise by more than 1 mm is regarded as malreduction. An overseen malreduction can lead to poor clinical outcome. The high inter-individual variance of the bone shape and pose impedes the correct evaluation of the reduction result since there is no reliable gold standard available. Contralateral side comparison is recommended in the literature to provide a more reliable reference but has not been integrated into the clinical routine as it requires the acquisition of an additional 3D image and increases the cumulative radiation dose and intervention time.

## 2.2 Technical background

The use and analysis of radiological images is an essential part of this thesis. Without technical advances of X-ray imaging in the last decades, neither pre-, intra- nor post-operative observations would be possible. This chapter gives a brief overview and compares different imaging devices utilized in the diagnosis and treatment of syndesmotic injuries.

### 2.2.1 2D Radiography

Radiography enables visualization of the interior parts of the body and can be used to diagnose major pathologies like trauma, joint disease and structural changes of osseous tissue (PETTERSSON 1998). Acquiring 2D radiographic images is the standard procedure in clinical routine for the detection of any kind of fractures in the human body. A radiograph is an X-ray imaging device, which emits radiation through the body without incision. It consists of a detector and an X-ray tube, which are placed around the patient to allow the rays to travel right through the region of interest. The tube, consisting of a cathode and anode, creates photons with high energy in the kilovoltage (kV) range, which are collimated by a beam-limiting device (AICHINGER et al. 2012) and send in direction of the detector. The X-ray spectra vary and the penetration ability increases with

the potential or applied voltage between cathode and anode (BERGER et al. 2018). When interacting with material, X-rays can either be scattered, absorbed or remain unaffected.

The underlying physical principle of X-ray imaging is the varying attenuation of radiation when interacting with tissue. The intensity of the attenuated rays captured by the detector varies depending on the density and structural composition of the intersected tissue (HANDELS 2009). The intensity  $I$  depends on the base intensity  $I_0$ , the linear attenuation  $\mu$  as well as the thickness  $d$  of the intersected material which can be simplified to

$$I = I_0 \cdot e^{-\mu \cdot d} \quad (1)$$

for one energy and more general for all energies and different tissue  $z$

$$I = \int_E I_0 \cdot e^{-\int_z \mu(z,E) dz} dE \quad (2)$$

Osseous tissue has a high electron density accompanied by a high absorption and is surrounded by tissue with less density, which makes it highly visible on radiographs. Soft tissue with a weaker density is often not well distinguishable.

A differentiation is made between conventional and interventional radiographs. Conventional radiographs are static systems used for pre-operative diagnosis. Interventional radiographs, also referred to as interventional fluoroscopic systems or C-arms, can be static or mobile and are used in the *Operating room* (OR) during interventional procedures. Regarding trauma surgery, the use of mobile C-arms serves as the gold standard for planning, navigation and validation of surgical procedures (FRANKE and GRUTZNER 2013). Mobile C-arms comprise an X-ray source and a detector, which are connected through a C-shape linkage explaining its name. The C-arm base is capable of horizontal and vertical movements and allows to position the C-arm in an arbitrary position and orientation towards the OR table.

The contrast of fluoroscopic images is inherently weak compared to images of conventional radiographs. Lower energies are applied to reduce radiation exposure during continuous radiation. In the case of a constant tube exposure, this results in a lower *Signal-to-Noise ratio* (SNR). Fluoroscopic systems provide static images and real-time image sequences, but on the other hand, suffer from motion-related artifacts.

In modern fluoroscopic systems, either image intensifier or flat panel detectors are deployed. Although still many image intensifier systems can be found in the clinics, flat panel detector systems are the state of the art technology for interventional imaging (BERGER et al. 2018). Flat panel detectors are designed to deliver better spatial resolution while reducing noise. Furthermore, they do not suffer from distortion effects compared



to image intensifiers and the spatial resolution depends on the dimension of the detector, not the *Field of View* (FOV).

Since it is only possible to examine the patient from a limited number of directions, the view direction must be chosen considerately. The most suitable direction varies with the anatomical region and the indication of the examination. Most commonly *lateral* and AP projections are acquired (SZEIMIES et al. 2012). In general, it is recommended to intersect the examined bone areas perpendicular to the surface and - if not impeded by pathologies - parallel to the plane of the detector medium.

Plain 2D radiographs are useful to show internal anatomical structures but lack any depth information since intersected tissue absorption is summed along each X-ray. Depending on the indication, the use of a complementary examination with a 3D imaging device like MRI or CT is recommended to reveal findings often overseen in planar radiographs (PETTERSSON 1998) as already addressed in section 2.1.3.

### 2.2.2 Computed Tomography

The mathematical principle of CT (greek *tomos* = slice, *graphein* = to write) goes back to 1917, where Radon proved that it was possible to reconstruct a 3D image from summation images (CASSAR-PULLICINO 1998). The key idea was to acquire various summation images from different directions and to perform a back-projection of the attenuation signals. To acquire a 3D image, which allows separation of superimposed tissue and provides information on the internal structure of the body, Hounsfield applied the reconstruction principles to the medical domain and introduced the first commercial CT scanner in 1971 (HANDELS 2009). Nowadays, different methods for CT reconstruction exist, the most widely known being filtered back-projection and the algebraic reconstruction technique (HSIEH 2003).

Conventional CT scanners are mostly used for complementary pre- and post-operative imaging and still show superiority in the depiction of osseous tissue. However, due to the ionized radiation and potential harm to the patient, benefits must be weighed against costs. The static systems consist of an X-ray source and a detector component hidden behind a rotating gantry and a patient table (Fig. 10). Different generations of scanner systems have been developed with either rotating both, source and detector (third generation) or only the source (fourth generation). The latest generation of scanners are spiral multi-slice CTs, which rotate the source continuously, translate the patient table simultaneously and have a multi-beam collimator. As an advantage, the patient can be scanned in one single breath on a spiral path.

The detector converts incoming X-ray photons into visible light, which in turn is con-

verted into electrical signals that can be transformed into digital information by AD conversion (CASSAR-PULLICINO 1998). The resulting gray values are called *Hounsfield Unit* (HU) and cover more than 4000 values. HUs are standard values used in CT image interpretation and are normalized to the reference of attenuated water being 0 HU, bone roughly 1000 (depending on the density) and air -1000 HU.

$$HU_{\text{tissue}} = 1000 \cdot \left( \frac{\mu_{\text{tissue}} - \mu_{\text{water}}}{\mu_{\text{water}}} \right) \quad (3)$$

In the equation of Hounsfield,  $\mu_{\text{tissue}}$  describes the linear attenuation coefficient for a specific tissue and  $\mu_{\text{water}}$  describes the linear attenuation coefficient for water. This results in unimpeded rays travelling only through air represented as parts of the image with lower HU whereas osseous tissue usually has much higher HU. The human eye is capable of distinguishing between 700-900 shades of gray (KIMPE and TUYTSCHAEVER 2007). To benefit from the large HU range, an intensity windowing can be applied to narrow down the entire range to the range of interest. The window center is set to the average value of the *Region of interest* (ROI) and the width of the window can limit the range. CT values below or above the window range are set to the minimum and maximum, respectively.

The image quality of CT images is influenced by different factors. Scanner-related factors include the voltage of the detector tube (kV), tube current, pitch, scan rotation time, flying focal spot as well as beam filtration and collimation among others. Detector-related factors are the pixel size and electron noise. Reconstruction-related factors comprise slice thickness, slice increment and the chosen reconstruction algorithm. Besides, the scanned anatomy influences the image quality by its size, the organs' density and patient motion.

Furthermore, the image quality can be affected by image artifacts, which cause errors in the CT image that are unrelated to the actual information of the studied subject and might impede diagnostics (HSIEH 2003). Most common artifacts are volume averaging, in which two tissue types with a different intensity are averaged in the same voxel. This so-called partial volume effect occurs when the diameter of structures is smaller than the pixel size or slice thickness. Blooming occurs if the boundary shifts with the change of window center and width and might result in inaccurate tissue measurements e.g. for estimation of cortical thickness. Beam hardening artifacts occur due to energy shifts in the spectrum along the intersected object. The energy change is influenced by the length of the traveled trajectory as well as the varying attenuation of different materials. Since high energy photons are less likely to get absorbed, the average energy changes and leads to artifacts in the image (BERGER et al. 2018).

The reconstruction is affected in areas with varying beam intensities e.g. if a long-

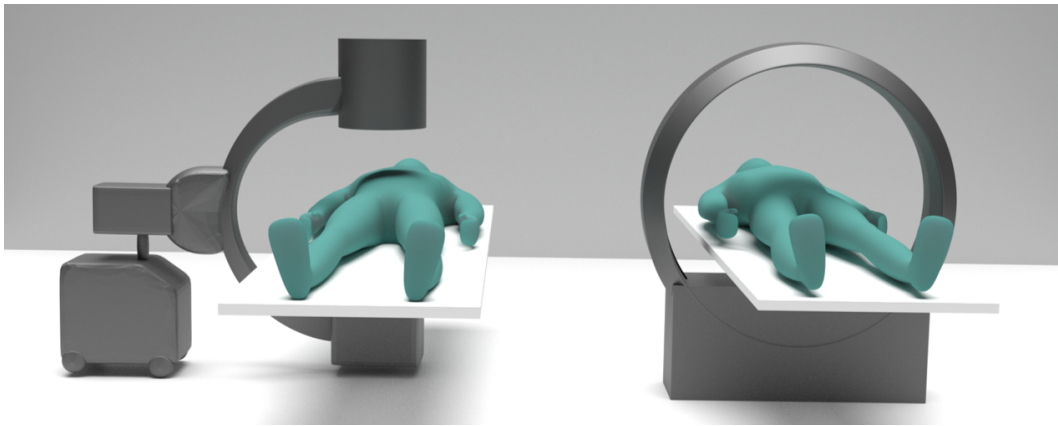


Figure 10: Computed Tomography systems: C-arm scanner (left), conventional spiral CT (right), created using Blender<sup>2</sup>.

shaft bone is cut from the FOV. Streak artifacts are present when external objects like metal implants or instruments of high density are present in the FOV and cause streaks in which the radiation is absorbed resulting in a star-shaped superimposition. Depending on the studied anatomy, motion artifacts from breathing or long scanning times can occur and cause streaks of blurred duplicated objects.

### 2.2.3 Intra-operative C-arm imaging

Besides its application in conventional static systems, the idea of using radiodensity for information gain has been also transferred to the OR. Intra-operative imaging has been used in a variety of different procedures and different surgical fields i.e. cardiology, orthopedics, vascular surgery, trauma surgery and radiotherapy. 3D fluoroscopy enables more rapid and accurate placement of needles, instruments, prosthetics and interventional devices (A. M. DAVIES et al. 2012). The 3D C-arm belongs to the so-called *Cone-beam Computed Tomography* (CBCT) systems (illustrated in Fig. 10). The main difference to a conventional CT is the acquisition of X-ray images to be used in the reconstruction process. A CBCT system uses a cone-shaped beam radiating from the X-ray source covering a large volume with one single motorized rotation about the patient instead of a spiral path as performed by conventional CTs. After the acquisition, the projection images are reconstructed in the same manner as static CT images usually using filtered back-projection.

Besides its great mobility, a CBCT has quicker motion in comparison to the spiral motion of a traditional CT scanner thus making it applicable in the OR. Furthermore, the radiation dose is less compared to a conventional CT, which is important because

<sup>2</sup>Blender version 2.8, Blender foundation, [www.blender.org](http://www.blender.org)

not only the patient but also the staff is exposed to radiation. One huge advantage of the iso-centric design is that the central beam always remains in the iso-center, which eliminates the need for re-positioning while reducing time and dose consumption (SIEMENS HEALTHCARE 2016). Mobile C-arm systems have a much smaller size than CT scanner, which allows integration into limited OR space. The availability of recently developed intra-operative CT scanner systems remains very limited in clinics.

On the contrary, the mobility of the C-arm not only provides advantages but at the same time leads to the disadvantage that no patient-to-device orientation is established. Therefore, the resulting images can vary significantly from patient to patient w.r.t the orientation and position of the present anatomical region. Furthermore, a low radiation dose implies a lower SNR and weak contrast. Compared to systems with flat panel technology, C-arms with image intensifiers have lower image quality and no absolute radiodensity, which means that the resulting values depend on all objects in the FOV and their relative positions. As a consequence, reconstructed volumes cannot be normalized based on HUs which hampers image interpretation. Furthermore, the image quality of image intensifier systems is affected by geometric distortion. Additionally, all CBCT systems are more prone to image artifacts like beam hardening and streak artifacts.

#### 2.2.4 Conclusion

In summary, tomographic imaging modalities like the conventional CT or CBCT can be used for diagnostics, treatment and evaluation in trauma surgery. Since mobile systems managed their way into the operation room due to their compact dimensions and lower radiation exposure, the surgeon can visually inspect the operative treatment. Mobile systems allow an arbitrary placement of the device towards the patient but at the same time do not provide knowledge on how the acquired images are oriented. Image understanding for diagnostics can be impeded, especially by image artifacts e.g. caused by metal implants often present in images that are acquired during trauma surgery. Compared to the acquisition with flat panel detectors, images acquired with image intensifiers are more affected by geometrical distortion, artifacts and noise. Another notable difference is the lack of normalized HU that in the case of image intensifier systems result in different appearances of the anatomy depending on the relative position and the surrounding tissue of the object of interest. This circumstance further complicates image analysis. All of the described methods have in common that they produce ionization radiation potentially harmful to the patient and in the case of mobile systems also to the surgeon. Therefore, one aim for the development of intra-operative assistance systems should be to reduce the cumulative radiation dose and always weigh potential benefits against harms.

Many systems for intra-operative assistance in trauma surgery have been developed in the last decades with arising imaging technologies and software development. To this date, however, no systems for intra-operative analysis of the ankle joint have been introduced neither in the market nor in the literature. This chapter reviews the development of intra-operative assistance systems for orthopedic surgery and gives an overview of different aspects required for the application (section 3.1). The developed system is a composition of different image processing methods covering the fields of image segmentation in 3D (section 3.2) and 2D (section 3.3) as well as 2D-3D bone reconstruction (section 3.4).

### 3.1 Assistance systems in orthopedic and trauma surgery

Assistance systems are designed specifically to help the surgeons during all phases of the surgery, either in planning before the procedure (pre-operative), navigation, monitoring and inspection during the procedure (intra-operative) or verification of the outcome after the procedure (post-operative).

Most of the systems proposed for orthopedic or trauma surgery focus on pre-operative planning of fracture assessment or prosthesis selection. E.g. FÜRSTHAHL et al. 2012 developed a pre-operative assistance system based on the comparison of anatomies that have a symmetric counterpart in the body. They performed the reconstruction of complex proximal humerus fractures by mirroring and registration of the uninjured and the injured humerus using CT images with both humeri present in the FOV. EHLKE et al. 2015 use 3D models of pelvis and femur to estimate anatomical measurements from 2D radiographs for the correct positioning of an osteosynthesis plate. The software iLeg (ZHENG

2013b) and iJoint (ZHENG 2013a) developed at the University of Bern offer a pre-operative planning to model different parts of the lower extremity and allow patient-specific 3D measurements based on 2D radiographs.

Intra-operative assistance systems are specifically designed for the application during surgery and can help the surgeons with navigation, augmented visualization and workflow monitoring. All systems have in common that they underlie strict requirements concerning time consumption (ZHENG and NOLTE 2018), limited OR-space and patient safety restrictions like the avoidance of additional radiation exposure following the *As Low As Reasonably Attainable* (ALARA) principle (STIEHL et al. 2007). Often, a registration step is required to either align the involved intra-operative components or to apply a pre-operative plan. The most widely known group of intra-operative assistance systems belong to the so-called *Computer-Assisted Orthopedic Surgery* (CAOS) systems. They consist of a *Virtual Object* (VO), e.g. a pre-operative CT, that serves as a virtual reference. This reference can be used for an enhanced visualization as well as for planning a drilling path, a screw location or a suitable prosthesis size. Another important component is the intra-operative data that can either originate from an intra-operative imaging device, the end-effector of a robotic system or any other external tool with a known position within an established coordinate system in the operation room. The most important step is to establish the relation between the VO and intra-operative data so that both share the same coordinate systems. Once the relation is established, the system can visualize any tracked instruments or manipulators in combination with the VO. In the same manner, a tracked robotic system can execute predefined plans based on the reference data. In the last decades, commercial CAOS systems have been developed for various types of surgeries. Currently available systems are mainly specialized in pedicle screw placement, using either tracked handheld instruments (e.g. Brainlab<sup>1</sup>) or robotic systems (e.g. Mazor Robotics<sup>2</sup>), and partial or total knee and hip replacement (e.g. Stryker<sup>3</sup>).

For enhanced visualization, augmented reality has been deployed in intra-operative assistance systems. NAVAB et al. 2010 supplemented a C-arm with a video camera and a double mirror system to overlay camera and radiographic images and investigated performance and knowledge gain for pedicle screw placement, vertebroplasty, and intramedullary nail locking procedure. Recently, FOTOUHI et al. 2019 presented a geometry-aware visualization of virtual 2D C-arm images augmented on the real patient anatomy to enable intuitive surgical guidance.

---

<sup>1</sup>Brainlab, Muinch, Germany, [www.brainlab.com/surgery-products/](http://www.brainlab.com/surgery-products/)

<sup>2</sup>Mazor Robotics, Caesarea Maritima, Israel, [www.mazorrobotics.com](http://www.mazorrobotics.com)

<sup>3</sup>Stryker, Kalamazoo, USA, [www.stryker.de/](http://www.stryker.de/)

Acquiring the VO pre-operatively is always dependent on the clinical indication and the procedure. Hence, other intra-operative assistance systems exist that do not require any pre-operative reference data. Instead, they rely on prior knowledge about the anatomy of interest. Related previous work on foot and ankle surgery concentrate on intra-operative analysis and improved visualization of calcaneus fractures. BREHLER et al. 2015 developed an assistance system for the automatic adjustment of MPR viewing planes using feature-based atlas registration and the analysis of fracture reduction quality along the computed planes. GÖRRES et al. 2016 proposed 3D segmentation method to be applied on the reduced calcaneus, cuboid and talus using a multi-bone statistical shape model. Furthermore, they developed a method for the detection of intra-articular screws since the metal implants are present in the 3D image after reduction. The methods were customized for 3D C-arm images acquired during calcaneal fracture surgery without any patient-specific pre-operative data.

Post-operatively, assistance systems are most commonly used to assess the quality of the outcome. E.g. ZHENG et al. 2012 employed a model-based 2D-3D reconstruction method to validate the correct cup orientation of the pelvic implant after performing a *Total hip arthroplasty* (THA).

In summary, many well-established assistance systems exist for pre-, intra- and post-operative applications. In comparison, intra-operative systems have the most restrictive requirements concerning execution time and data acquisition, especially if no pre-operative reference data is available. Intra-operative assistance systems usually comprise a number of processing steps that include image segmentation and analysis as well as the registration of images with different modalities and dimensions. In the following sections, an overview is given of the state of the art of individual components and technical approaches relevant for this thesis that as well build the foundation of many assistance systems for orthopedic surgery.

## 3.2 3D anatomical shape representation

The shapes of anatomical organs show a large variation (KOBATAKE and MASUTANI 2017), just like the human exterior appearance, and can be influenced by different factors such as age, size, sex, ethnicity, physical activity or pathologies. The extraction of 3D shapes from medical images to analyze anatomical variation has been extensively studied in the literature. A common shape representation is a voxel grid classifying each voxel as either belonging to a specific structure of interest or the background. This process is referred to as voxel-wise segmentation that partitions an image volume  $\mathbf{I} \subset \mathbb{R}^3$  into pairwise disjoint

sub-regions  $\mathbf{I}_k$ .

$$\mathbf{I} = \bigcup_{k=1}^K \mathbf{I}_k \quad \text{with} \quad \mathbf{I}_i \cap \mathbf{I}_j = \emptyset \quad \forall i \neq j \in \{1, \dots, K\} \quad (4)$$

Segmentation results can hold binary labels (single class with  $K = 2$ ) or a label for each object class (multi-class with  $K > 2$ ). All voxels belonging to a sub-region  $I_k$  share e.g. the same spatial or textural properties. The classification criteria to create sub-regions ranges from manual selection based on visual inspection to automatic determination.

Although many automatized segmentation techniques have been developed to this date, segmentation in the medical domain is still most commonly performed manually through labeling pixels in a slice-by-slice manner. If applicable, semi-automatic methods like region growing or thresholding approaches can be applied e.g. a histogram-based threshold selection using Otsu's method (OTSU 1979). However, manual segmentation is error-prone for difficult cases and can be dependent on the experience level of the operator resulting in disambiguity (KOHL et al. 2018) and a high inter-rater or even intra-rater variance (WARFIELD et al. 2006). Most importantly, manual labeling is tremendously time-consuming (GREENSPAN et al. 2016), especially in case of large or many data sets, which often makes it inapplicable for the clinicians to integrate into their daily workflow. Thus, the development of automatic segmentation methods remains an important field of research. In recent years, the field has undergone great advances by incorporating supervised machine learning techniques. Segmentation methods include the extraction and learning of task-specific computed features (NORAJITRA and MAIER-HEIN 2017), the registration and warping to templates or atlas images (SOTIRAS et al. 2013) and the use of active, deformable shape models (HEIMANN and MEINZER 2009) adapting their shape to the target organ. The choice of methods is always task-dependent. The current state of the art for voxel-wise segmentation are *Convolutional Neural Network* (CNN)-based methods outperforming almost all former segmentation techniques (LITJENS et al. 2017). Section 3.2.1 describes recent developments in this field. Nonetheless, in some specific cases, i.e. shape analysis, shape registration and anatomical reconstruction, a voxel grid provided by the segmentation is not sufficient. Instead, a more high-level representation with shape descriptors on the anatomical regions' boundary is required. Therefore, in section 3.2.2 an overview is given on model-based segmentation.

### 3.2.1 Deep learning approaches

A deep learning approach is different from other machine learning approaches, in which hand-crafted feature representations are learned. Instead, a deep neural network can



learn the feature representation implicitly by updating and optimizing its network parameters. By learning an intensity representation of the underlying image data distribution, deep learning methods are capable of applying the learned task on unseen images.

Originating from the recent success of image classification in computer vision, deep learning approaches have evolved that not only provide one label per entire image but instead create a pixel- or voxel-wise classification referred to as segmentation. Recently, CNN based approaches have received a lot of attention due to their ability to solve various segmentation problems surpassing previous methods (GREENSPAN et al. 2016). A typical CNN architecture consists of a stack of so-called layers, each performing a specific mathematical operation like convolution or pooling. Convolutions are performed on the input of the layer with different convolutional matrices, also referred to as kernels. Pooling is defined as the aggregation of pixel neighborhoods by computing the maximum or mean value (LITJENS et al. 2017). Each layer receives the input from the preceding layer and outputs a different representation followed by a non-linear activation function. This way, the input is projected onto different feature spaces with decreasing dimensions for each layer to build an abstract and more meaningful representation. The first layers rather emphasize low-level features like edges, blobs and corners (GREENSPAN et al. 2016). CNNs used for classification are often supplemented with a fully connected layer at the end of the convolution layer assembly without shared weights. This layer has a fixed size and therefore limits the spatial resolution of the output since the spatial information is lost at this position. Segmentation tasks in the medical domain, however, need to be solved on a pixel- or voxel-based level. This requires the CNN to be applied in a computationally expensive 'sliding window' approach by processing the image patch-wise.

One solution to the problem was provided by the *Fully Convolutional Network* (FCN) introduced by LONG et al. 2015 for image segmentation. In their work, they replaced all fully connected layers of the CNN with fully convolutional layers and supplemented the network by upsampling operations and skip connections to combine the semantic information of the final prediction layers with appearance information of the lower layers. With a fully convolutional architecture, the FCN can take an image with arbitrary size and produces an output with the same resolution (cf. Fig. 11). Building upon the idea of FCNs, the most prominent approach in the medical domain is the U-Net architecture introduced by RONNEBERGER et al. 2015. The original architecture is a symmetric FCN.

It consists of a downsampling path, namely the encoder, and an upsampling path, namely the decoder path. The symmetry creates a much higher number of feature maps in the upsampling path compared to a classical FCN. The main difference towards other

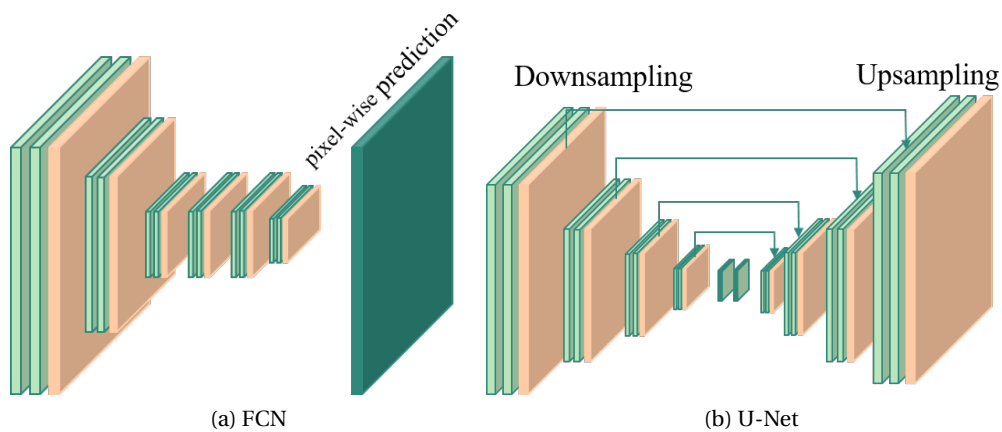


Figure 11: Network architectures

architectures are concatenating skip connections linking each layer of the two paths to provide additional spatial information during upsampling. In the original publication, the downsampling path is composed of blocks as illustrated in Fig. 11b. Each of the blocks consists of two convolution layers with an activation function and a normalization layer at the end substituted with a maximum pooling layer. Therewith, the contextual information of the image can be captured. The upsampling path comprises the same number of blocks with a deconvolutional layer each. The upsampled feature maps are concatenated with the corresponding feature maps from the downsampling layer to combine spatial with contextual information. The input size is variable since no dense layers are involved and the output is a segmentation map with one output channel for each class. The U-Net can be trained by comparing the output of the network to the ground truth class and calculating an appropriate loss metric. This loss is back-propagated through the network by an optimizer to update the weights in the individual layers.

Many different variations of the U-Net architecture have been proposed in the literature. Particularly interesting for the medical domain, ÇIÇEK et al. 2016 and MILLETARI 2016 were among the first to extend the 2D U-Net to 3D by replacing all 2D convolution and pooling operation by their 3D counterparts. The V-net by Milletari et al. not only extends the vanilla U-Net to allow the input of volumetric data but also introduces residual blocks as well as a different loss function based on the Dice score (DICE 1945). Although many modifications to the architecture have been proposed in the literature, their significance remains controversial. ISENSEE et al. 2019 recently showed the superiority of the vanilla U-Net architecture in an international challenge on a variety of different anatomies, suggesting that changes in the network architecture do not significantly

improve the performance. They recommended to rather focus on assemblies and adaptive pre-processing as well as a profound selection of hyperparameters.

The amount of training data needed for learning the underlying distribution is one major drawback of learning-based methods (GREENSPAN et al. 2016). In other research areas of computer vision, deep neural networks are trained using millions of data samples. In medical image analysis, however, often only a small data set is available due to ethics and legal regulations, limited possibilities of data transfer or lack of annotated and meaningful data (UNBERATH et al. 2018). Researchers have developed different strategies to overcome this limitation. Besides techniques like pre-training on large-scale data sets, domain adaptation and transfer learning, data augmentation is one of the methods most often applied when dealing with limited training data. Data augmentation aims at increasing the number of training images while avoiding the need for annotation and, as a result, helps the network to generalize to a larger set of image characteristics. It is usually performed by applying basic image operations on both, the images and the annotations. Images can be modified and data set variation extended either by applying a spatial transformation, such as translation or rotation and pixel-wise enhancement such as contrast augmentation.

CNN segmentation of the musculoskeletal regions has been employed in a variety of projects. Some authors like KLEIN et al. 2019 concentrate on the general segmentation of bone areas. Other groups rather focus on specific bones like the spine. X. LI et al. 2018 proposed a method to segment and localize intervertebral discs in MR images. PRA-SOON et al. 2013 used a 2.5D CNN for segmentation of knee cartilage. 2.5D denotes the segmentation of a 3D image by applying a 2D-CNN on slices in all three orthogonal 3D image planes and aggregating the results. ZENG et al. 2017 employed deep supervision in a 3D U-Net like architecture to segment femoral bone tissue in MR images.

Besides the need for a huge amount of training data, one other drawback of CNNs is that the result is a voxel-based classification of the image. Some medical properties like volume or size can be directly computed and used for medical analysis. However, a voxel-wise segmentation as such does not hold any low-dimensional feature representation nor any semantic correspondences necessary for more complex shape analysis or comparison. Up to this date, learning a statistical shape model representation is the most prominent approach for shape correspondence estimation. Since the establishment of corresponding landmarks plays an important role in the remainder of this thesis, the following sections provide a detailed description of the state of the art in statistical shape modeling.

### 3.2.2 Statistical shape modeling

As opposed to pure voxel-based segmentation methods, model-based methods integrate semantic a-priori knowledge of certain anatomy-specific properties and statistics (HANDELS 2009). Those models can help to gain insights on similarities and dissimilarities among a population or be used for segmentation, registration and reconstruction of new unknown images. Statistical models capture the variance and prior knowledge of the shape and appearance of an anatomical structure given annotated training data. Following notations are used for the remainder of this thesis: If the model is comprising the statistics of an objects' shape, for example by describing the surface as a set of anatomical landmarks and their distribution, instances are named *Statistical Shape Model* (SSM). If also image-related properties like the intensity are modeled, instances are referred to as *Statistical Shape and Intensity Model* (SSIM).

A broad overview of fundamentals and modifications of shape and appearance models can be found in the reviews by HEIMANN and MEINZER 2009 for medical image segmentation and SARKALKAN et al. 2014 for bone modeling in particular. In the following, the major aspects of statistical shape models and some variations beneficial for the project are described in detail.

#### Model representation

Analyzing the underlying training data distribution is the key step towards building a model representation. The morphological variance is most commonly modeled by a distribution of landmarks on the structures' boundary, the so-called *Point Distribution Model* (PDM). The PDM is a versatile, simple method which can be used to model objects of arbitrary topology. Each training sample can be described by a landmark vector  $\mathbf{X}_{\text{training}} = \{\mathbf{x}_k | k = 1, \dots, N\}$  with  $N$  corresponding Euclidean landmarks  $\mathbf{x}_k = (x_k, y_k, z_k)^T \in \mathbb{R}^3$ . Those landmarks are either selected based on unique anatomical landmarks or by voxel-wise segmentation of the medical image and applying a surface extraction algorithm like Marching Cube developed by LORENSEN and CLINE 1987. This iso-surfacing algorithm results in a polygonal mesh with a varying number of landmarks. Consequently, the latter approach requires an additional correspondence building step, since a PDM requires an equal number of distinct landmarks across all training shapes.

Besides the PDM, other shape feature representations have been investigated in the literature. STAIB and DUNCAN 1992 transfer the principle of Fourier transformation to the surface domain to model shape descriptors. SZEKELY et al. 1996 build a parametric model for segmentation of spherical objects by mapping the surface on a sphere and parametrizing spherical harmonics (SPHARMs). A different approach to model shape

variation is called shape-DNA introduced by REUTER et al. 2006, in which the shape representation is transferred to the Laplace-Beltrami space. Alternatively, M-reps (Medial representation) introduced by PIZER et al. 2003 model the object as a compact mesh or chain of so-called medial atoms. Medial atoms are defined as a discretized collection of landmarks on the medial axis of an organ with vectors directing at the outer surface. M-reps were successfully applied in bone segmentation e.g. by ANAS et al. 2016 to extract the wrist joint in ultra-sound images. However, applications are limited by high computation times and therefore not widely used. Another approach found in the literature is based on level-sets (MALLADI et al. 1995). Level-sets can be used to model the object's boundary with implicit functions by embedding it in a higher dimensional surface representation. Due to their simplicity and computational efficiency, PDMs are still widely used in shape analysis and build the basis of many segmentation algorithms in the medical domain. All developments in this thesis are based on the PDM, since the representation meets the requirements for this project in terms of high computational efficiency and adaption to bone segmentation and reconstruction. Therefore, the further sections emphasize on different shape model related aspects all based on PDMs.

Several steps are necessary for the construction of a PDM. The training shapes are aligned and transferred into a common coordinate system. Afterward, corresponding landmarks are established. Finally, the dimensionality of the landmark distribution is reduced to derive a more compact representation of the models' statistical distribution. The following paragraphs describe the steps in more detail.

### **Shape alignment and correspondence optimization**

One major drawback of PDMs is the required registration of landmarks to establish a meaningful distribution. In the medical domain, training images are not necessarily aligned towards each other, can contain different anatomical sizes and do not share the same coordinate system. As a consequence, the landmarks in the PDM vector do not yet correspond to the same anatomical landmarks. A common approach to align the individual shapes is *General Procrustes Analysis* (GPA), firstly introduced by GOWER 1975. Procrustes analysis as such offers a solution in closed form that aligns two shapes by translation, rotation and uniform scaling to minimize the Euclidean distance of the landmarks. The GPA is an extension of the Procrustes analysis that allows the alignment of multiple shapes. One essential pre-requisite of the GPA is that the number of landmarks does not vary among the shapes. Therefore, correspondences must be established and may also be optimized and changed during optimization.

The definition of well-defined corresponding landmarks among the training shapes is

the most challenging, nonetheless crucial, part of the shape model construction according to HEIMANN and MEINZER 2009. The statistical variation of the anatomical shape can only be analyzed if corresponding points represent the same anatomical landmarks on the surface for all shapes. In simple point-based correspondence building methods like the *Iterative Closest Point* (ICP) by PAUL BESL 1992, the Euclidean distance between two shapes is minimized iteratively. Applying a similarity transform based on the minimum Euclidean distance may cause inaccuracies in correspondence selection, especially in case of large shape variation. To overcome that limitation, more flexible, deformable methods have been introduced like e.g. the *Coherent Point Drift* (CPD) algorithm introduced by SONG and XUBO 2010. Instead of using the minimum Euclidean distance, an optimization criterion can be defined with respect to the determinant of the covariance matrix to find correspondence points among the dense meshes of the training shapes. R. H. DAVIES et al. 2002 were the first to apply a now widely adapted approach, in which optimization was performed based on the *Minimum description length* (MDL). The MDL is inspired by the Occam's Razor principle which claims that the simpler the model is, the better it performs, which in terms of the SSM implies that it can be described by fewer variational modes. To reduce the computation time and make it more applicable in the model construction, THODBERG 2003 presented a simplified version of the MDL, which is based on a gradient descent optimization. After successful correspondence building, the resulting landmark distribution captures the shape variation.

### Model building

Under the assumption that the corresponding landmarks of  $\mathbf{X}_{\text{training}}$  are normally distributed, the model can be described by the mean  $\bar{\mathbf{X}}$

$$\bar{\mathbf{X}} = \frac{1}{M} \sum_{m=1}^M \mathbf{X}_{\text{training},m} \quad (5)$$

built over landmarks of all M training shapes and the corresponding covariance matrix  $\mathbf{C}$ :

$$\mathbf{C} = \frac{1}{M-1} \sum_{m=1}^M (\mathbf{X}_{\text{training},m} - \bar{\mathbf{X}})(\mathbf{X}_{\text{training},m} - \bar{\mathbf{X}})^T \quad (6)$$

The mean model describes the extent of a principal mode along the training samples. To reduce the dimensionality of the data and find the set of modes that best describes the underlying shape variation, dimensionality reduction is performed. The most prominent approach is using a *Principal Component Analysis* (PCA) on the resulting landmarks and performing an eigendecomposition on the covariance matrix to compute eigenvectors and their corresponding eigenvalues. The eigenvectors are sorted by their corre-

sponding eigenvalues, which resemble the shape variance along each vector. The eigenvectors with the highest eigenvalues reflect the largest shape variations and are therefore most relevant for statistical shape analysis. Eigenvectors with lower eigenvalues, on the other hand, reflect small shape variations which may be mostly caused by noise (HEIMANN and MEINZER 2009). After *Singular Value Decomposition* (SVD) each training sample  $\mathbf{x}_m$  can be approximated using

$$\mathbf{x}_m \approx \bar{\mathbf{X}} + \mathbf{P}\mathbf{b}_m \quad \text{with} \quad \mathbf{P} = \{p_i | 1, \dots, t\} \quad (7)$$

$\mathbf{P}$  contains  $t$  eigenvectors,  $\mathbf{b}$  is a  $t$  dimensional vector of the eigenvalues and  $\bar{\mathbf{X}}$  denotes the mean landmark vector. By altering eigenvalues  $\mathbf{b}$  of the SSM, new shape representations can be created.

Besides the standard PCA, different variations of dimensionality reduction have been used for shape modeling in recent years. LUTHI et al. 2018 propose a reformulation of the PCA by applying Gaussian morphable models (GPMM) to model the shape variation with a Gaussian process. Recently, MA et al. 2017 proposed a more robust method, the so-called kernel-regularized robust PCA (KRPCA) to model the anatomy of the ankle in healthy subjects of MRI. They claim that the model is more robust towards out-of-distribution samples and showed promising results in the presence of pathologies. However, the method was not evaluated with regard to the presence of metal artifacts or time consumption.

### Appearance Modeling

To allow the SSM to be used in image segmentation and create correspondences between the model and new unseen images, not only the shape characteristics but also the appearance of an anatomical structure is modeled. The first and simplest approach proposed in the literature by COOTES et al. 1995 was the adaption of the model to the strongest gradient or distinct edge features. Those features, however, might be influenced by noise and artifacts or do not necessarily resemble the true boundaries of the anatomy. Therefore, more complex and specialized appearance models have been developed (HEIMANN and MEINZER 2009).

Appearance models can be categorized in profile-based and region-based models. Profile-based models were originally proposed by COOTES et al. 1995 who modeled a Gaussian distribution of profiles sampled at landmarks of the training data. During segmentation, the Mahalanobis distance between a current landmark profile in the unseen image and the learned distribution is compared. It was shown that normalized profiles or the normalized derivative profiles of the samples served best for comparison as op-

posed to comparing plain intensity values. The main pre-condition of the training data is the assumption that the intensity value distribution is Gaussian, which is not necessarily the case according to BRUIJNE and NIELSEN 2004. To overcome this limitation, they introduced a kNN-classifier sampling both, true boundary landmarks and non-boundary landmarks, to build profiles that afterward classify the current landmark and provide a goodness of fit.

Due to its simplicity and efficiency, profile-based methods modeled solely along the shape boundary are most commonly found in SSM segmentation applications. Region-based models capture a wider range of intensity values by building a feature vector, in which all intensity or gradient values are stored. This can be either done on the whole region of the shape model by transforming the sample images to the mean shape and then perform a PCA on the textures or by sampling only a region of interest to cope with general higher computational costs.

HEIMANN and MEINZER 2009 state that clustering of appearance model landmarks improves the generalizability of the model. BREJL and SONKA 2000 were the first to introduce a c-means clustering for reducing the complexity and combing appearance profiles with similar properties.

### 3D model-based segmentation

Besides shape analysis of anatomical structures, a main application of SSMs is model-based segmentation. Given an initial estimate of the shape parameters  $\mathbf{b}$  (often the mean shape  $\mathbf{b}=0$ ) and an initial placement of the SSM described by transformation  $\mathbf{T}$ , parameters of the SSM are iteratively updated to fit the SSM to the image and obtain a model presentation  $\mathbf{X}_{\text{final}}$  of the anatomy.

$$\arg \min_{\mathbf{T}, \mathbf{b}} \left| \mathbf{X}_{\text{final}} - \mathbf{T} \underbrace{\left( \bar{\mathbf{X}} + \sum_{p=1}^t \mathbf{P}_p \mathbf{b}_p \right)}_{\mathbf{X}} \right| \quad (8)$$

$\mathbf{T}$  represents a similarity transformation aligning the mean model to the target structure. The modifiable range of shape parameters  $\mathbf{b}$  is usually constrained to create more plausible shape variations by allowing the different modes to deviate by  $|\mathbf{b}_p| \leq 3 \sigma_p$ . The optimization is performed by iteratively estimating the best local displacement of each landmark  $\mathbf{x} \in \mathbf{X}$ . For this purpose, the local image intensity information at the landmark positions is analyzed. Intensity profiles are evaluated at different probe positions perpendicular to the surface boundary with varying probe spacing to estimate their good-



ness of fit. The iterative optimization process is twofold: First, all optimal local landmark displacements are computed and the GPA algorithm is applied on the resulting displacements to determine an update of  $\mathbf{T}$ . Secondly, the displacements are projected in shape space to obtain shape deviation  $d\mathbf{b}$ . Thus,  $\mathbf{b}$  of the SSM is globally altered to resemble the prior learned representation based on the estimated landmark adjustments. The steps are repeated until the model converges at the local optimum. Most works found in the literature employ a multi-resolution strategy with appearance models trained for each resolution (LINDNER 2017). This approach is said to be more robust towards initialization and can speed up the adaption process.

### Optimal surface integration

One major disadvantage of SSMs is the sensitivity towards unexpected appearance values as they occur in the presence of image artifacts, especially whenever metal implants are involved. A step to overcome this limitation was proposed by HEIMANN et al. 2007. They integrated an optimal surface approach, firstly introduced by K. LI et al. 2006, to improve the quality of the model adaption, especially in areas affected by noise and artifacts. The approach complements a global shape prior during the evaluation of local appearances. This way, the overall landmark displacements lead to smoother surface deformation. The optimal surface algorithm solves the shape adaption efficiently by transforming the task into a graph problem. A directed graph  $G = (N, D)$  consists of  $N$  nodes and  $D$  connecting edges with a source  $n_s \in N$  and a sink  $n_t \in N$ . The graph is constructed comprising a node  $n \in N$  for each landmark profile position of the mesh. Each node is assigned with the fitting cost of the respective appearance model and a sink and source is supplemented and connected to each node. Additionally, the nodes of the neighboring profile positions are connected. By applying the graph cut algorithm in a maximum flow and minimum cut manner as described by BOYKOV and KOLMOGOROV 2004, the optimal landmark displacements can be computed while ensuring a certain smoothness of the overall shape. KAINMUELLER et al. 2008 were among the first to extend the optimal surface approach for the use of multi-object segmentation by introducing inter-mesh intensity profiles. GÖRRES et al. 2016 extended the graph optimization approach of HEIMANN et al. 2007 to multi-object SSMs by connecting the individual graphs with interconnecting edges.

### Model initialization and robustness

Another main drawback of SSM segmentation is the potential convergence into local minima. Model-based segmentation methods often require at least an approximate prior

knowledge about the position of the anatomical structure. FRIPP et al. 2006 applied an affine intensity-based pre-registration before the actual correspondence optimization. ECABERT and THIRAN 2004 employed a generalized Hough transformation to estimate the coarse transformation between the model and the data set. Recently, NORAJITRA and MAIER-HEIN 2017 presented an SSM approach with appearance modeling based on random forests and majority voting that alleviates the need for initialization and covers large translation offsets and rotations up to  $30^\circ$ . One often mentioned disadvantage of shape model approaches is their lack of generalizability towards pathologies and image artifacts. MA et al. 2017 showed that the kernel-regularized robust PCA can cope with pathologies but did not compare to classical PCA. In their study, no image artifacts were present in the data. GÖRRES et al. 2016 developed several methods to improve multi-object segmentation of the calcaneotalar joint in C-arm images. The work addressed the presence of metal implants and was designed in particular to cope with the challenges of image artifacts in C-arm data as well as intra-operative requirements and therefore built the basis of this work.

### 3.2.3 Combining shape model segmentation and deep learning

In the last few years, authors have combined the high accuracy that CNN-based approaches can offer with the capability of SSMs to incorporate semantic relations. A junction of both approaches can be manifold and integrated on different levels. The SSM can be used as a shape prior in the network training to create meaningful representations in a semantic manner. The predictions of a CNN can be used for initialization and detection of the anatomical structure to be utilized as a starting point for the local search of the SSM. AVENDI et al. 2016 presented a method in which CNN-based methods are applied to localize and initialize a deformable model for ventricle segmentation in MRI images. The predictions can be further used, not for mere pre-alignment of the model, but also during the actual segmentation process. MA et al. 2018 proposed a Bayesian model combining shape model segmentation with CNN-based predictions for pancreas segmentation. KOREZ et al. 2016 use a 3D CNN to localize vertebral bodies and use the networks' outcome during the segmentation process. Very recently, few researchers, as e.g. BHALODIA et al. 2018 with DeepSSM, have worked on alleviating the need for local model adaption by directly inferring shape parameters from images using a neural network.

### 3.3 Image understanding of radiological summation images

Conventional X-ray or fluoroscopy images differ from images of other medical image modalities in their nature of being summation images as already described in detail in section 2.2.1. Independent from the used technology, challenges include overlapping anatomical structures and external objects superimposing the region of interest. The literature on analysis of summation images mainly includes segmentation and feature extraction. The segmentation task of summation images differs to the classical 3D segmentation in the way that one pixel can hold more than one class since two or more anatomical structures can be superpositioned at a pixel location. Thus, this particular pixel-wise segmentation partitions the image  $\mathbf{I} \subset \mathbb{R}^2$  in a union of non-disjoint subsets  $\mathbf{I}_k$ .

$$\mathbf{I} = \bigcup_{k=1}^K \mathbf{I}_k \quad (9)$$

Traditional image processing methods applied on X-rays include simple semi-automatic thresholding and region-based approaches like region growing (ADAMS and BISCHOF 1994). Class affiliation is determined by analyzing neighboring pixels with respect to a uniformity or connectivity criteria. Those methods often require a manual seed or threshold selection and cannot fully adapt to varying intensities in case of superposition of structures or noise. Edge-based methods extract distinct edges rather than regions by applying convolution operators on the image. Besides simple operators like Prewitt, Sobel, Roberts and Laplacian of Gaussian (BOZEK et al. 2009), the most prominent edge detector is the so-called the Canny edge detector (CANNY 1986). Successfully applied in many segmentation and registration applications (BAKA et al. 2012; BENAMEUR et al. 2003; ZHENG et al. 2009), the Canny edge detector computes intensity gradients on the smoothed image and applies non-maximum suppression and hysteresis to extract strong edges. All of the mentioned traditional contour and region extraction methods can solve simple segmentation task efficiently but are often not capable to adapt to image noise or artifacts, especially, when being applied on fluoroscopic images instead of standard radiographs.

For complex segmentation problems and inhomogeneous medical images, more advanced methods like deformable-, atlas- or learning-based methods seem to be more appropriate. The use of SSM is not limited to 3D surfaces but can just as well be employed to learn the shape characteristics of contours. XIE et al. 2014 applied a 2D-SSM to segment the proximal femur in AP projection images. In the last decade, the quality of segmentation or feature extraction tasks on summations images has been improved by incorporating machine learning approaches. One popular approach is the use of ran-

dom forests for landmark classification. Random forests consist of several decision trees which are trained on subsets of features and data. During training time, the algorithm splits branches according to a split rule separating positive and negative samples. LINDNER et al. 2013 apply random forest regression voting in a sliding window approach to extract potential landmark positions in an AP radiograph. Those landmark candidates are used to initialize the local search of a 2D-SSM. CHEN and ZHENG 2013 also employed random forest regression to directly derive potential 2D-SSM landmark positions for the 2D segmentation of the pelvis and femur.

Earlier work on neural networks for segmentation of X-rays by CERNAZANU-GLAVAN and HOLBAN 2013 already used back-propagation and feed-forward neural networks with rather mixed, cluttered results on ribcage segmentation. Recently, CNN-based 2D segmentation of summation images has been introduced e.g. by BOHM et al. 2018. An U-Net was implemented to extract cell boundaries in microscopic images. CNN-based 2D bone segmentation of radiological images, in particular, has been presented in the literature. Among others, NOVIKOV et al. 2018 compare different network architectures and achieved best results by using an U-Net inspired CNN for segmentation of clavicles in chest radiographs. KORDON et al. 2019 showed a comparison of different augmentation methods for the segmentation of the knee joint in lateral X-rays and fluoroscopic images. AL ARIF et al. 2018 showed that an U-Net inspired CNN with a specific shape loss function outperforms traditional SSM based approaches when being applied on vertebral bodies present in lateral X-rays.

In conclusion, learning-based methods have shown their superiority in fluoroscopic image segmentation tasks. However, all of the mentioned learning-based methods restrict the learned representation to one particular viewport and do not generalize to arbitrary views.

### 3.4 2D-3D Reconstruction

Three-dimensional reconstruction from medical images has been proven to be very useful for pre-operative planning and computer-aided orthopedic surgeries. However, conventional techniques for image reconstruction of a 3D model from a modality like CT have the drawback that they entail additional costs and time consumption during imaging (REYNEKE et al. 2018). Furthermore, such imaging devices often cause high radiation or pre-operative or intra-operative data is simply not available. Especially during musculoskeletal surgery, only 2D images are acquired routinely for navigation, guidance and verification. Therefore, researchers have started to investigate how to obtain a 3D model from a limited number of 2D images. This approach, known as 2D-3D recon-

struction, also often referred to as deformable, model-to-modality or atlas-to-modality registration, has become a quite large but still highly challenging research field due to its ill-posed nature (MARKELJ et al. 2012).

The problem can be formulated as follows: Let  $A$  be 3D data defined over the domain  $\Omega_A \subset \mathbb{R}^3$  and  $B_j$  2D intra-operative data defined over  $\Omega_B \subset \mathbb{R}^2$  with  $j = 1, \dots, N$  images. Then, 2D-3D registration can be formulated as the mapping  $f(\theta) : \Omega_A \mapsto \Omega_B$  that aligns data points  $x \in \Omega_A$  to  $y \in \Omega_B$  such that a certain comparison criterion  $\epsilon$  is minimized.

$$\theta_{opt} = \arg \min_{\theta} \epsilon(A, B) \quad (10)$$

In classical rigid 2D-3D registration, the 3D data  $A$  is patient-specific and corresponds to the object present in  $B$ . The mapping  $f$  can be a rigid transformation parametrized by  $\mathbf{T}$  comprising three translational parameters  $t_x, t_y, t_z$  and three rotational parameters  $\alpha, \beta$  and  $\gamma$ . During registration, the parameter vector  $\mathbf{T}$  for the rigid transformation, necessary to align the 3D image to the 2D images, is established by transforming the data set  $A$  until it is best aligned with  $B_j$ . As opposed to that, 2D-3D reconstruction lacks a corresponding 3D reference image of the same patient, but instead uses a 3D model or morphable atlas/template image of the anatomy. In consequence, not only the spatial relationship defined by parameters  $\mathbf{T}$  but also shape deformation parameters  $\mathbf{b}$  are optimized to fit  $A(\mathbf{x}^{3D})$  to  $B_j(\mathbf{y}^{2D})$ . The combined set of parameters is denoted by  $\theta := [\mathbf{T}, \mathbf{b}]$ .

Different reconstruction methods found in the literature can be characterized by their 3D data representation, their reconstruction strategy, the type of optimizer and the nature of the input. An overview of the different aspects is given in the following.

### 3.4.1 3D Model type

The 3D reference can be a deformable model encoding prior learned knowledge and assumptions about the anatomical variation, either based on a representative atlas, a statistical representation or parametric relations (examples are shown in Fig. 12).

- A template or atlas image can be e.g. a mean intensity representation of an anatomical structure or a 3D image of one sample patient. These images can be deformed during optimization to the effect that the corresponding simulated projection images fit to the reference projections. YU and ZHENG 2018 used a template image of the distal femur and tibia as well as three thigh muscles. The template is warped by computing a local deformation field and applying a thin-plate b-spline transformation.

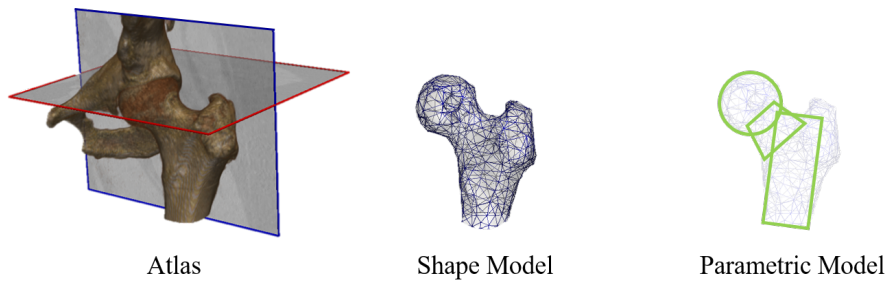


Figure 12: Different model types

- SSM is a pure shape-based statistical model as described in section 3.2.2 encoding transformation and shape parameters of the bone anatomy. The model is either deformed by modifying the shape  $\mathbf{b}$  and transformation parameters  $\mathbf{T}$  or by free-form deformation.
- SSIM is an extended version of the SSM complementing the shape with intensity-related information. For a comprehensive overview of SSIM based bone reconstruction, in particular, the reader is referred to the review of REYNEKE et al. 2018. EHLKE et al. 2013 created a tetrahedral SSM by supplementing the boundary surface model with landmarks within the anatomical object. By incorporating the entire volume density distribution, they aim to create more realistic simulated 2D images. BONARETTI et al. 2014 applied an alternative approach by warping all training data to a reference image and directly perform the PCA on the voxel grid to create an image-based SSIM.
- Parametric models are based on mathematical relationships approximating the anatomy by primitive geometrical parameters. E.g. DONG and ZHENG 2008 defined the femoral head as a circle and the corresponding shaft as a cylinder shaft in a Bayesian inference network to define the initial position of the femur.

### 3.4.2 Reconstruction strategy and definition of cost function

2D-3D reconstruction can be approached with two different strategies: Intensity-based and feature-based reconstruction, which differ in the definition of the cost function  $c$ . Intensity-based approaches apply a forward projection on the 3D model or template to create simulated reference 2D images. The real and simulated 2D images are compared to each other to maximize a similarity-based cost function  $c_{sim}$ . Feature-based methods most commonly establish correspondences between 3D model landmarks and computed features of the 2D images and minimize a distance-based cost function  $c_{dist}$ .

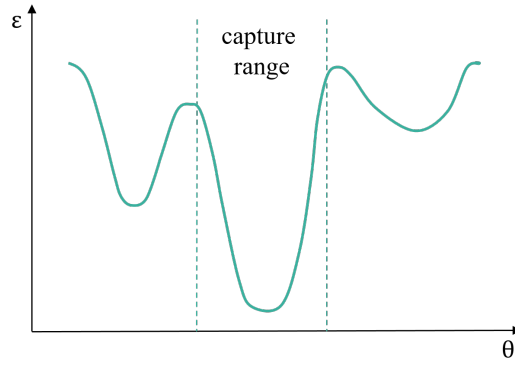


Figure 13: Non-convex objective function: The capture range defines an interval, in which the objective function is considered to show a convex behavior.

In all optimization problems, a cost function  $\epsilon$  with a collection of variables  $\theta$  is formulated and the optimization aims at minimizing the function value by altering  $\theta$ . Cost functions with a convex nature have the advantage that the local optimum also resembles the global optimum. Thus, even local optimizers are guaranteed to converge into the global optimum. However, especially intensity-based cost functions tend to be highly non-convex and most likely possess local minima and a small capture range (REYNEKE et al. 2018; SOTIRAS et al. 2013). The capture range describes the range, in which the function has a quasi-convex shape and can be formulated by constraining the value range of  $\theta$  (shown in Fig. 13). The choice of cost function depends on the type of reconstruction and a distinction is made between feature-based and intensity-based reconstruction.

### Feature-based reconstruction

Feature-based reconstruction finds correspondence pairs between the landmarks of the 3D model  $\mathbf{x} \in \Omega_A$  and features  $\mathbf{y} \in \Omega_{B_j}$  computed on the 2D projection images. Those 2D features can be distinct anatomical landmarks, contours, edges, points or the result of a neighborhood operator. Fig. 14a depicts an example of feature-based reconstruction, in which the segmentation is used for feature creation and the deformable model is a 3D-SSM learned from training data. Feature-based reconstruction seeks to iteratively minimize the distance between the features in the respective spatial space. The required correspondence building can be either performed by transforming the 3D model and projecting it onto 2D space using a perspective projection matrix  $P$  and calculating the distances between the geometrical features (MARKELJ et al. 2012).

$$\mathbf{x}^{2D} = P(\mathbf{T}(\mathbf{y}^{3D})) \quad (11)$$

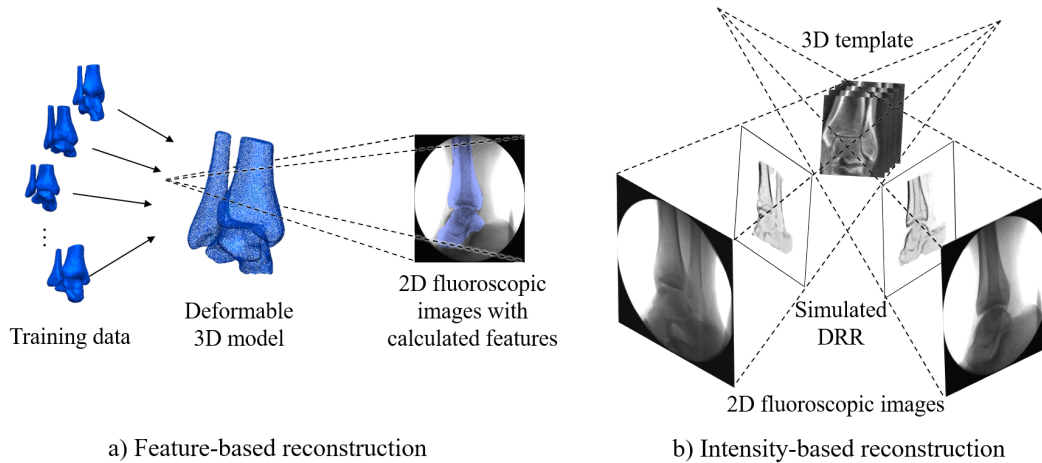


Figure 14: Reconstruction strategies: a) Feature-based reconstruction with a 3D deformable model and features computed from the fluoroscopic image. b) Intensity-based reconstruction with the generation of simulated DRRs.

An alternative approach is the back-projection and triangulation of 2D features in 3D space and calculating the 3D distance to the 3D model landmarks (ZHENG and SCHUMANN 2009; BAKA et al. 2012).

$$\mathbf{x}^{3D} = \mathbf{T}(P^{-1}(\mathbf{y}^{2D})) \quad (12)$$

Correspondence building always depends on the quality of the extracted features (MARKELJ et al. 2012) and is the most important part of the reconstruction. Most commonly, the cost function  $\epsilon_{\text{dist}}$  is the Euclidean distance between model landmarks and the assigned corresponding points or lines in the least square manner. Instead of the Euclidean distance, BENAMEUR et al. 2003 employed the Mahalanobis distance that allows accounting for the correlation between variables.

### Intensity-based reconstruction

As opposed to feature-based approaches, intensity-based reconstruction methods directly employ the pure pixel information of the 2D projection images and the correspondence building step can be omitted. In general, those approaches have higher accuracy, since they are not biased by feature extraction inaccuracies and exploit all information of the projection images (REYNEKE et al. 2018). On the contrary, without any pre-processing or feature extraction, the reconstruction is more sensitive to noise and external structures that are not present in both compared images. A widely used approach includes the generation of simulated 2D fluoroscopic images, called *Digital Reconstructed Radiograph* (DRR), from the 3D model or atlas and the comparison of the simulated with the real fluoroscopic images (cf. Fig. 14b). Correspondences, as they are established in



feature-based reconstruction, are assumed to be given implicitly at every pixel location in which the respective images overlap. This approach seeks to minimize the difference between the simulated and the real image based on pixel-wise similarity measures. Research has mainly focused on how to improve the computationally expensive step of the DRR generation in terms of efficiency by e.g. integrating GPU computing (EHLKE et al. 2013) and quality by selecting adequate similarity measures. Different measures, specifically for intensity comparison, are employed in 2D-3D reconstruction like e.g. mutual information (SADOWSKY et al. 2006; YU et al. 2016), gradient correlation and difference (PENNEY et al. 1998; OTAKE et al. 2015), normalized cross-correlation (KLIMA et al. 2016) and gradient direction (DE SILVA et al. 2016). A brief description of the different metrics can be found in the comparative study of PENNEY et al. 1998. They concluded that the performance of metrics incorporating spatial information, like gradient difference, were superior to metrics purely based on intensity values.

### 3.4.3 Optimization strategies

Optimization aims at minimizing or maximizing a cost function by selecting the best set of parameters  $\theta_{opt}$ . An optimizer can be defined as a certain strategy to select the parameters for the next iteration to lead to the optimum of the cost function. Optimizers used in 2D-3D bone reconstruction can be characterized in two categories: derivative-based and derivative-free optimizers.

Gradient descent is a first-order iterative algorithm that takes steps proportional to the negative gradient of the current function value (BENAMEUR et al. 2003). According to KLIMA et al. 2016, the classical gradient descent is outperformed by the Levenberg-Marquardt, which extends the former by a Gauss-Newton behavior in areas close to the optimum. Other derivative-based optimizers found in the literature are the non-linear conjugate gradients and an extension of the original gradient descent called stochastic gradient descent used by YU et al. 2016.

The derivative-free Down-Hill Simplex optimizer, used by SADOWSKY et al. 2007, BOUSSAID et al. 2011 and FLEUTE and LAVALLÉE 1999, iteratively refines sampled function values without differentiation. Another derivative-free algorithm is the Powell optimizer, successfully applied by KADOURY et al. 2009, based on a bi-directional search and iterative refinement of a search vector by a linear combination of former search vectors. More recently, HANSEN 2006 introduced the *Covariance Matrix Adaptive Evolutionary Strategy* (CMAES), a stochastic derivative-free optimizer, which is based on the biological principle of evolution. In each iteration, new candidate solutions are sampled based on the parent solution and selected as new parents based on their fitness. The algorithm

is said to be highly suitable for non-convex optimization problems with a medium large dimensionality and was applied for 2D-3D reconstruction by CERVERI et al. 2017 and GONG 2011. FLEUTE and LAVALLÉE 1999 presented an alternative approach similar to the widely known ICP (PAUL BESL 1992) adapted to 2D-3D registration. The iterative procedure alternates between correspondence building and transformation estimation. In the corresponding step, point correspondence updates are determined, while in the transformation step the transformation parameters of the model are updated by performing a closed-form least-square registration of the corresponding point pairs obtained in the first step.

#### 3.4.4 Setup and image data

Besides the chosen algorithm and model type, the input data highly affects the performance of the application. The size, number and the anatomical object of interest can be different for each task. Besides, settings like the number of images or the initial alignment of the model may influence accuracy, computation time and convergence. For the development of a 2D-3D reconstruction application, different design choices should be considered, which are illuminated in the following.

**Number of 2D projection images:** Most of the work found in the literature works with two or more calibrated 2D images taken from quasi-orthogonal directions. TOMAŽEVI 2008 investigated the impact on the number of X-ray views and concluded that the accuracy gain flattens with more than two images. LANGTON et al. 2009 were among the first to show a successful femur reconstruction from a single conventional X-ray. However, according to REYNEKE et al. 2018, 2D-3D reconstruction of both shape and intensity information from a single image remains an unsolved problem.

**Degree of interaction:** In all reconstruction applications, the 3D model has to be placed within the coordinate system of the X-ray images. Especially in intensity-based reconstruction, the impact on the convergence is highly dependent on an adequate initialization. Different strategies range from manual landmark selection to fully automatic approaches. For intra-operative applications, effort is made to make the method as automatized as possible to reduce human-computer interaction, potentially accompanied by a harmful effect on the surgical workflow. Earlier methods for initialization rely on the manual annotation of geometrical parameters on the anatomy (CHAIBI et al. 2012), by placing distinct stereo-corresponding landmarks (AUBERT et al. 2016) or by manual contouring the outer boundary (ZHENG and SCHUMANN 2009). For a fully automated feature

extraction and initialization, YU et al. 2016 used random forest regression and achieved accurate results but only tested their approach on simulated X-ray images. ZHENG et al. 2007 proposed a representation of the proximal femur based on geometrical primitives and employed a particle filter for fully automatic initialization. KARADE and RAVI 2015 initialize the template by performing a 2D ICP algorithm on the extracted contour and the projected template.

**Anatomy:** The musculoskeletal system of an adult human comprises of 206 bones according to HOMBACH-KLONISCH et al. 2019. However, 2D-3D reconstruction applications have been proposed for a limited number of bones only. A large number of orthopedic applications for various bones have been developed and evaluated under simulated or clinical conditions. But some applications were not investigated further because they failed in terms of accuracy or robustness or transpired to be too complex to be used within intra-operative setups (ZHENG and NOLTE 2018). Most of the methods are successfully applied to the proximal or distal femur (BAKA et al. 2012; SADOWSKY et al. 2007; KLIMA et al. 2016; STEININGER et al. 2008; EHLKE et al. 2013; ZHENG and YU 2017) or the pelvis (JIANHUA YAO and TAYLOR 2003; SADOWSKY et al. 2007). Other work focus on the spine and ribcage (BENAMEUR et al. 2003; BOISVERT et al. 2008; AUBERT et al. 2016) and the humerus (MUTSVANGWA et al. 2017). The most related anatomy to the ankle joint, the wrist was investigated by GONG 2011. Multi-bone reconstruction has been applied mostly on adjacent bones with small overlap. EHLKE et al. 2013 investigated the performance of articulated SSIMs, in which the femur is connected to the pelvis by a rotational and translational constraint inspired by their biomechanical relation. ZHENG and YU 2017 applied their reconstruction method simultaneously on pelvis and femur. The only work on parallel adjacent bone compounds was conducted by GONG 2011. REYNEKE et al. 2018 claimed that the success of a reconstruction algorithm on one specific bone does not guarantee the translation to other bones.

**View port dependency:** JIANHUA YAO and TAYLOR 2003 concluded that the best registration accuracy is obtained with an orthogonal angle between two perspectives and least correlated with a view angle between 75 and 105. However, they neither investigated the performance on different anatomical structures nor on cases of partial occlusion. Furthermore, they showed that a calibration error of more than  $5^\circ$  can significantly increase the reconstruction error. REYNEKE et al. 2018 mentioned that the correlation between accuracy and orthogonality is not given in case of symmetric shapes like the pelvis with a large superposition and concluded that the best imaging direction remains anatomy

specific.

**Computation time:** A comparison of run times for bone reconstruction is difficult since existing methods are neither implemented within the same software framework, executed with the same hardware nor performed on the same data. In general, intensity-based optimization is more computationally expensive (MARKELJ et al. 2012) due to the time-consuming DRR calculation step, which has to be repeated numberless times. GPU-based DRR calculation helps to significantly speed up the optimization process as proposed by EHLKE et al. 2013. Most of the authors do not mention any specific time measurements for the reconstruction. Computation times found in the literature range from less than 1 minute to 5 minutes (YU et al. 2016; KARADE and RAVI 2015; ZHENG and SCHUMANN 2009). CERVERI et al. 2017 reported a computation time of 45 minutes for the 2D-3D reconstruction of the distal femur when using the C-MAES and elastic 2D-2D registration for correspondence building.

In conclusion, many different approaches for the 2D-3D reconstruction of bones have been introduced in the literature and applied to various bone compounds. The main difference is made between feature-based and intensity-based reconstruction. The intensity-based reconstruction is assumed to be more accurate but also more time-consuming. There is no approach that provides an optimal solution for all problems, hence, the selection of an appropriate algorithm remains task-specific.

### 3.5 Discussion

Assistance systems in orthopedic surgery are all targeted at improving the clinical workflow by providing the surgeon with useful additional information on the anatomy. Intra-operative assistance systems have more strict requirements since their application is dependent on the OR environment including the often very limited time frame of the surgery, the changing appearance of the anatomy and the limited OR space. To date, no assistance systems for ankle surgery are available. Although many other applications can be found in literature, there are only a few that managed their way into clinical routine. This might be also explained by the high accuracy demands of the clinical application and the complexity of the developed systems. Assistance systems often comprise a conjunction of several complex components. The sections in this chapter reviewed the state of the art with respect to the different components essential for the development of the proposed application. Transferring the findings in the literature to the specific application of contralateral side comparison, different conclusions could be drawn.

**3D anatomical shape representation:** Different methods for shape analysis and segmentation have been introduced in the literature. The main criteria for selecting an appropriate approach to meet intra-operative restrictions are robustness towards arbitrary initialization and metal-induced artifacts, compatibility with time requirements and a sufficient accuracy. The PDM model has been successfully applied in many SSM based segmentation tasks since construction and application is computationally efficient and allows for correspondence comparison and shape analysis. Other shape representations like m-reps were not further investigated since the time consumption does not meet the time requirements of intra-operative applications. GÖRRES et al. 2016 presented an approach that yielded a high accuracy for the segmentation task of the calcaneus in C-arm images, which is closely related to the task addressed in this thesis. They investigated the impact of metal-induced artifacts and arbitrary SSM placement. It has to be noted that their work did not address the effect of varying metal implant positions on the SSM performance. In all images, the metal plate was roughly placed at the same position relative to the calcaneus.

CNN-based methods have set the state of the art in image segmentation outperforming SSM-based approaches. However, CNNs jointly learn the appearance and shape of the anatomy that not necessarily leads to anatomical plausible shapes. Consequently, the segmentation can contain holes or mislabeled isolated regions. Furthermore, a pure voxel-based segmentation lacks unique corresponding landmarks to derive shape descriptors. The SSM is learned by decoupling shape and appearance. This leads to anatomical plausible shapes with corresponding points that can be used for the subsequent 2D-3D reconstruction. On the contrary, the SSM is more sensitive towards changes in appearance and requires an initialization. Therefore, a combination of the shape prior of an SSM with the highly accurate CNN-based segmentation might be beneficial to overcome limitations of both approaches.

**Image understanding of radiological summation images:** CNN-based methods, especially the U-Net architecture, are well suited for both, 2D and 3D segmentation tasks. One of the main requirements of those learning-based methods is the availability of an acceptable number of training data to cover a large variation of the input. Especially fluoroscopic images show a high degree of freedom but suffer from a sparse sampling limited by the C-arm rotation angle.

**2D-3D bone reconstruction:** Different possible 3D model types can be found in the liter-

ature, namely SSM, SSIM, templates and parametric models. All approaches developed for 2D-3D bone reconstruction establish the 3D model based on CT volumes. According to MARKELJ et al. 2012 other modalities like MRI are not suitable for DRR generation from the SSIM or template due to the lack of normalization. The same problems arise, if the training data is built based on C-arm images acquired with an image intensifier as described in section 2.2.2. The use of C-arm data for SSIM generation has not yet been investigated in the literature. Also, parametric models, e.g. applied on the femur by DONG and ZHENG 2008 can not be transferred to the ankle joint since there are no unique landmarks or mathematical relations to describe the bones.

Fluoroscopic images, especially when acquired with an image intensifier, entail image noise, device-specific geometrical distortion, intensity inhomogeneities and external objects. Obstacles like the operating room table, additional screws or tubes present in the field of view may impede the performance of the reconstruction.

Possible reasons for the lack of any bone reconstruction methods applied on the ankle joint could originate in the bones completely superpositioned on all 2D images. Recent approaches for multi-bone reconstruction were applied on adjacent bones with little to no contact area with respect to the bone size. If the bone overlap is high, intensity-based methods might not be applicable since the appearance on the DRRs are dependent on both models which may not necessarily correlate in their variation. Similar to the application concerning the femur bone, varying shaft length and out-of-field areas have to be taken into account. Another challenge is the rotation-symmetric nature of the distal fibula. Even minor deviations during feature calculation can cause the model to rotate along its shaft axis. Other anatomies provide more distinct anatomical landmarks and may be less likely to fall into local minima.

Similar to the highly varying computation time, there is a controversy on the validation of bone reconstruction algorithms. Hardly any work is reported on real fluoroscopic data although most of the researchers propose that their approach could also be used during surgery, in which calibrated conventional X-ray images are often not available. This might be explained by the lack of reference data to evaluate the algorithm quantitatively. A comparison of different proposed algorithms is tedious because the used settings and data sets differ and usually test data sets are of limited size that does not guarantee generalizability to other anatomies or image modalities (REYNEKE et al. 2018).

Summarizing, although many different approaches have been developed, there is no approach, which can serve all applications and must be adapted to the specific anatomy, the nature of the input data and intra-operative restrictions. The proposed work aims to

create an intra-operative assistance system to be used in ankle joint surgery by carefully considering all aforementioned challenges. Those challenges include the strict time requirements that do not allow to use computational expensive methods, the limitations given by the image intensifier C-arm device such as inhomogeneities and metal artifacts, and the complexity of the ankle joint.





## 4.1 Approach

In this work, a prototype for a contralateral side comparison for the ankle joint was developed. This chapter gives an overview of the medical workflow implied by the proposed prototype (section 4.1.1) and presents design choices based on the state of the art (section 4.1.2). The chapter is concluded by a description of potential sources of error and their relations within the system (section 4.1.3).

### 4.1.1 Clinical workflow

As already described in section 2.1.4, ORIF is one of the state of the art approaches in upper ankle surgery. The fractured fibula is relocated and reduced using e.g. a one-third tubular plate and screws during open surgery. Hereinafter, an overview of a commonly used clinical workflow is provided (see Fig. 15, colored in blue):

1. **Diagnosis:** A pre-operative X-ray image is acquired and an indication for the surgery is given depending on the height of the fracture (cf. section 2.1.3).
2. **ORIF procedure:** The ankle is exposed and the dislocated fragments are relocated and stabilized with metal implants using standard AO techniques.
3. **Screw placement:** A syndesmotic screw is inserted to connect fibula and tibia for the syndesmotic compound to heal properly.
4. **3D image acquisition:** An intra-operative scan of the injured ankle is acquired using a mobile 3D C-arm device.

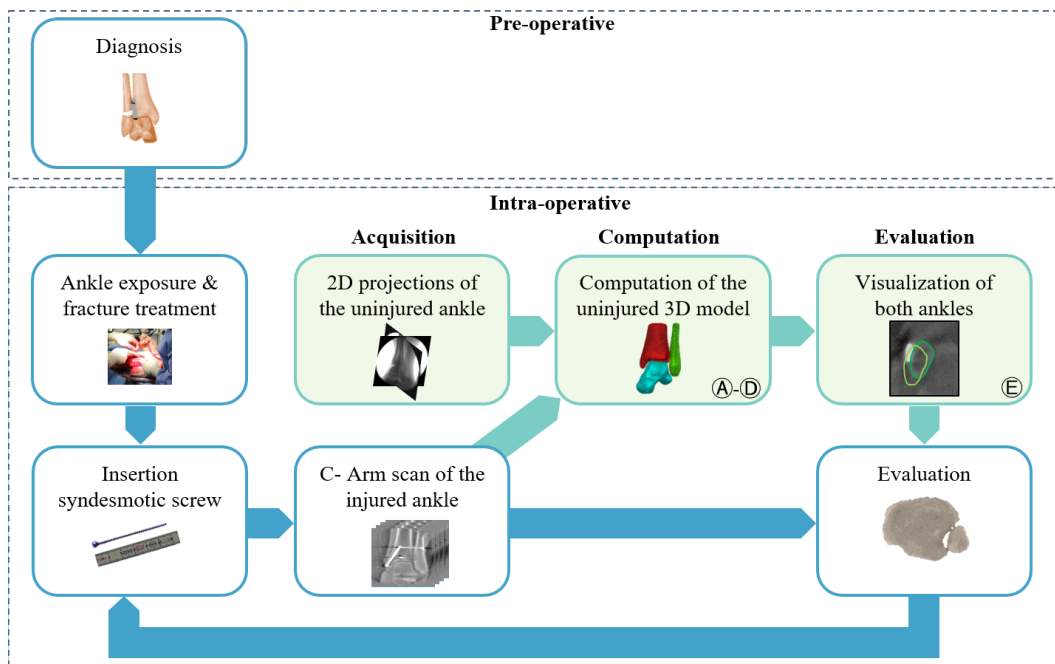


Figure 15: Medical workflow: Different steps of the procedure with (green) and without (blue) the proposed automatic contralateral side comparison

5. **Evaluation:** The 3D image is examined at relevant regions to decide whether the fibular reduction result is sufficient.

In the event of malreduction, the fibula position is modified, and steps 4 and 5 are repeated until the surgeon is satisfied with the result. This may lead to repeated intra-operative revision scans and high cumulative radiation dose. To date, manual contralateral side comparison is only performed in rare cases, in which there is an indication given that the anatomy of the ankle appears to be suspicious or the reduction process is complicated impeding an appropriate evaluation. Even if the surgeon is certain about the reduction result, there is still a chance of overlooking natural anatomical pose variation. The main research question of this project was to answer whether and how it is possible to perform an automatized contralateral side comparison. As the main prerequisite, the volumetric data of both ankle joints need to be available for a successful profound 3D comparison. But up to date, C-arm scanners in clinical routine are mostly limited by their detector size and only capable of fitting one ankle joint in the FOV.

Summarized in Tab. 2, there are three approaches for evaluating the reduction result: No comparison, 3D-3D comparison and 2D-3D comparison. In this thesis, 2D-3D comparison was elaborated after considering the advantages and disadvantages of each approach. Compared to the standard procedure without any comparison, a full scan of the contralateral side and a direct 3D-3D comparison can show intra-individual differences

Evaluation Method	Acquisition Injured ankle	Acquisition Uninjured ankle	Remarks
No comparison	3D C-arm scan	None	+ Less images - Decision based on surgical skills - No inter-individual variability - Revision-related scans with additional radiation
3D-3D comparison	3D C-arm scan	3D C-arm scan	+ Inter-individual variability + Less revision-related scans - More radiation exposure - Increased OR time
2D-3D comparison	3D C-arm scan	2D C-arm projections	+ Inter-individual variability + Less revision-related scans - Lack of 3D information

Table 2: Overview on different approaches for evaluation of the reduction result

and may avoid further revision-related scans. On the contrary, a full scan of the contralateral side not only causes additional radiation but also requires a time-consuming resetting of the scanner and an increase of OR duration due to additional C-arm acquisition. Instead of acquiring a full 3D scan, one possible solution is to acquire low-dose 2D projection images and compute a 3D model to perform a 2D-3D comparison. The proposed workflow requires additional 2D fluoroscopic projections from the contralateral side of the patient, which can either be acquired before starting the fibular reduction or after acquiring the 3D image of the reduced fibula. Providing the surgeon with an overlay of the reconstructed surface on top of the fractured volume allows an instant comparison between both ankles (as illustrated in Fig. 15, colored in green). The main problem to consider was that reconstructing 3D information from 2D is always an ill-posed problem so that the algorithm needs to perform the reconstruction as precise as possible to meet the accuracy requirements.

#### 4.1.2 System design choices

The main objective of this project was to design and implement a clinically compatible concept for automatized contralateral side comparison, which can be integrated into the current workflow for syndesmotic interventions.

Most relevant in translating the prototype to the operating room was the consideration of intra-operative requirements. All developed methods should only add reasonable time to the intervention time to avoid costs and risks for the patient. Furthermore, the cumulative radiation exposure to the patient and staff should not significantly exceed the currently used.

With the intra-operative C-arm as designated imaging modality, a 2D-3D reconstruction method was developed that avoids a full 3D scan of the uninjured ankle and aims at reducing the number of revision scans. Regarding time requirements, all proposed methods needed to be designed to be computationally efficient.

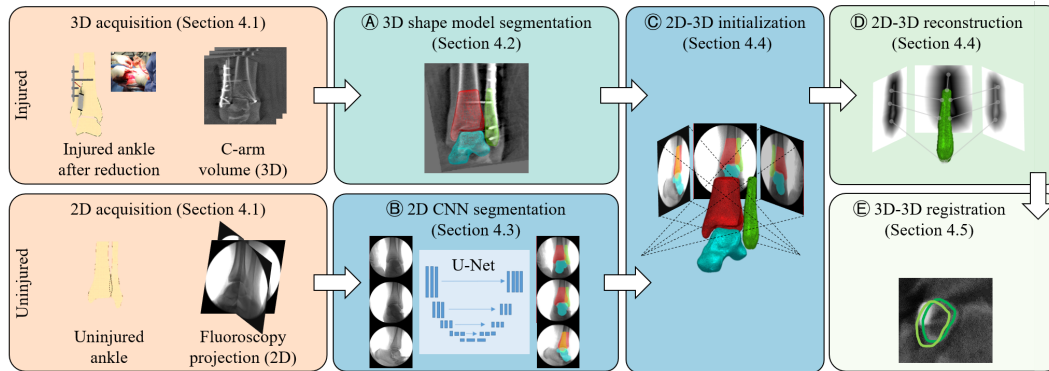


Figure 16: Technical overview: Different components of the proposed automatic contralateral side comparison

2D-3D reconstruction is not an end-to-end method but a complex pipeline involving a number of pre-processing and intermediate steps. Design choices of the different components were made based on the findings of the state of the art in 2D-3D reconstruction in section 3.5. Fig. 16 shows the technical workflow that comprises of five major steps: Segmentation in 3D using a statistical shape model (SSM), 2D segmentation of the fluoroscopic projections, coarse alignment of both data sets leading to a fine 2D-3D reconstruction and finalized by a model-to-model back-registration. The embedding of the components in the clinical workflow is labeled in Fig. 15 (corresponding alphanumeric symbols). Hereinafter, design choices for the single components are described in detail:

#### (A) 3D shape model segmentation of the injured ankle:

The injured ankle is scanned after performing ORIF using an intra-operative C-arm. The acquired 3D images have the following common properties:

- All bone fragments are re-positioned to resemble the ankle before the fracture. In most cases, thin fracture lines indicate the fracture but no extremely dislocated fragments are present.
- Metal implants are visible on the 3D image and the osseous tissue is superimposed by metal-related artifacts. Especially, the fibula shaft is affected since it is fixated with the plate covering a large section of the bone.

- The visible area of the tibia and fibula shaft varies for each scan depending on the position of the ankle relative to the scanner's iso-center and ankle size.
- The intensity value range of the reconstructed volume varies with each image and 3D images are not normalized with HU.

To analyze the shape and pose of the injured ankle after fibular reduction, the bones need to be extracted from the 3D image first. A simple segmentation allows to extract the bones, but to compare both ankles, corresponding landmarks are desired. With those landmarks, shape variation can be modeled by a more compact representation. For the proposed prototype, a 3D-SSM segmentation approach is employed. The 3D-SSM must be implemented and trained so that it can handle varying shaft lengths and is robust towards metal-induced artifacts. Furthermore, the 3D-SSM segmentation must be executed in a reasonable time frame to meet intra-operative requirements and create accurate surfaces of the bones. One important advantage of a model representation of the injured ankle is that the obtained shape parameters closely resemble shape parameters of the final uninjured model. Thus, the reference model of the uninjured model for the 2D-3D reconstruction can be initialized with those parameters accordingly.

#### **(B) 2D-CNN segmentation and feature extraction of the uninjured ankle:**

For the proposed prototype, additional 2D projection images of the contralateral ankle are acquired. C-arm projection images of the uninjured ankle have the following common properties:

- Projection images only show the uninjured ankle. Metal implants, bone fragments or fractures are not present in the image under the assumption that the uninjured ankle is not affected by previous injuries or pathologies.
- In some viewports, the ankle can be superimposed by external objects like the OR table or the second foot.
- The ankle is acquired from different rotational angles in a range of  $190^\circ$  around the scanner's iso-center which is not necessarily the center of the ankle.
- The different projection images are calibrated towards each other by their respective calibration matrices.
- Each projection image is taken with a different kV resulting in different contrasts and different value ranges. Projection images suffer from noise and inhomogeneities towards the projection border and are not normalized.

In all projection images, multiple bones are overlapping and impeding the reconstruction quality. The first step is to extract only valuable and bone-specific information distinguishing between the relevant bones. Since the ankle is shown from different viewports, a method is required that is generalizable and not limited to specific anatomical landmarks. Therefore, the prototype uses a CNN-based segmentation approach, because it is fast and can create an accurate segmentation of each single bone. When trained on a reasonable amount of data, the method should be neither dependent on the viewport nor the number of bones.

### **(C) Initialization of model with projection images**

2D-3D reconstruction is an ill-posed optimization problem, especially, if the SSM is placed far from the global optimum. Therefore, after pre-processing of the 3D image and the projections, one component of the pipeline must be an initialization step, in which the model is coarsely pre-aligned with all the projection images. The projection images are acquired from different viewports that intersect at the iso-center of the C-arm. The proposed method needs to be fully automatic, reasonably fast and independent from the viewport (e.g. not dependent on viewport-specific features). The method should provide a coarse but sufficient initialization for the preceding fine reconstruction step. Since 2D segmentations are computed in the preceding step, the method can use the prior information as features and encode landmarks in the mean model representation of the 3D-SSM. This way no extensive computation is required.

### **(D) 2D-3D Reconstruction**

In this step, the shape and transformation parameters of the 3D-SSMs must be altered to reconstruct the uninjured ankle. As already discussed in section 3.5, intensity-based reconstruction is difficult to apply to image-intensifier-based 3D images. Thus, the proposed method is based on extracted features. This has the advantage that a multi-bone reconstruction can be performed that is not biased by the bone overlap. The method is variable in terms of the number of viewports and bones and can be improved steadily by using more training data. The number of images is set to the number of standard projections taken in the clinical routine. Preliminary experiments showed that two images might not allow revealing enough spatial information to fit the model to the contours. Since no anatomical landmarks are available, the reconstruction must rely on automatic assignment of correspondence points. The choice of a suitable optimizer is influenced by the number of the parameters, the nature of the objective function and the efficiency of the calculation. The number of parameters is reasonably low if the number of shape

parameters is restricted. The objective function is not convex and might contain many local minima depending on the start position. Instead of a global optimizer that would not take into account local shape deviation of the bones, a multi-resolution ICP-inspired approach with an adaptive update strategy is used to solve the problem.

### **(E) 3D-3D Registration and Visualization**

The last step comprises an adequate visualization of the reconstruction result. The reconstructed 3D-SSMs of the uninjured ankle must be aligned with the injured ankle and displayed as an overlay. The aim of the overall pipeline is the identification of the fibula deviation. Thus, the alignment must be performed while remaining the spatial relation between the bones. Furthermore, the surgeon must be provided with utility methods to ease the reduction evaluation process. This includes an automatic standard plane adjustment to the standard planes described in section 2.1.4. In addition, the workflow can be improved by providing methods to switch between the two levels relevant for the evaluation.

The following section presents an overview of possible sources of error and influencing factors of all components used in the proposed workflow.

#### **4.1.3 Sources of error**

The proposed system is comprised of several individual components and all of them introduce an uncertainty contributing to the overall error. To allow for a systematic evaluation, this section presents a hierarchical representation of the different sources of error. As shown in Fig. 17, their influencing factors are either based on the input data, the implemented algorithm, the quality of the learning-based method or on the output(s) of preceding steps.

At the bottom of the hierarchy, components are situated that only depend on the implementation of the component itself and the input images. At the top, components are affected by all preceding steps. The 3D-SSM segmentation creates the 3D model of the injured ankle and is influenced by the output of the CNN that in turn highly depends on the training data. The acquisition of the 3D image of the injured ankle is influenced by user interaction and image quality. The operator chooses the anatomy of interest, the laterality and the C-arm configuration. Errors in the settings are severe and lead to a failure of the software. Furthermore, the operator has to position the ankle in the iso-center with the shaft lines of tibia and fibula orthogonal to the C-arm. Any misalignment can influence the result of the system. Concerning the anatomy and image quality, the

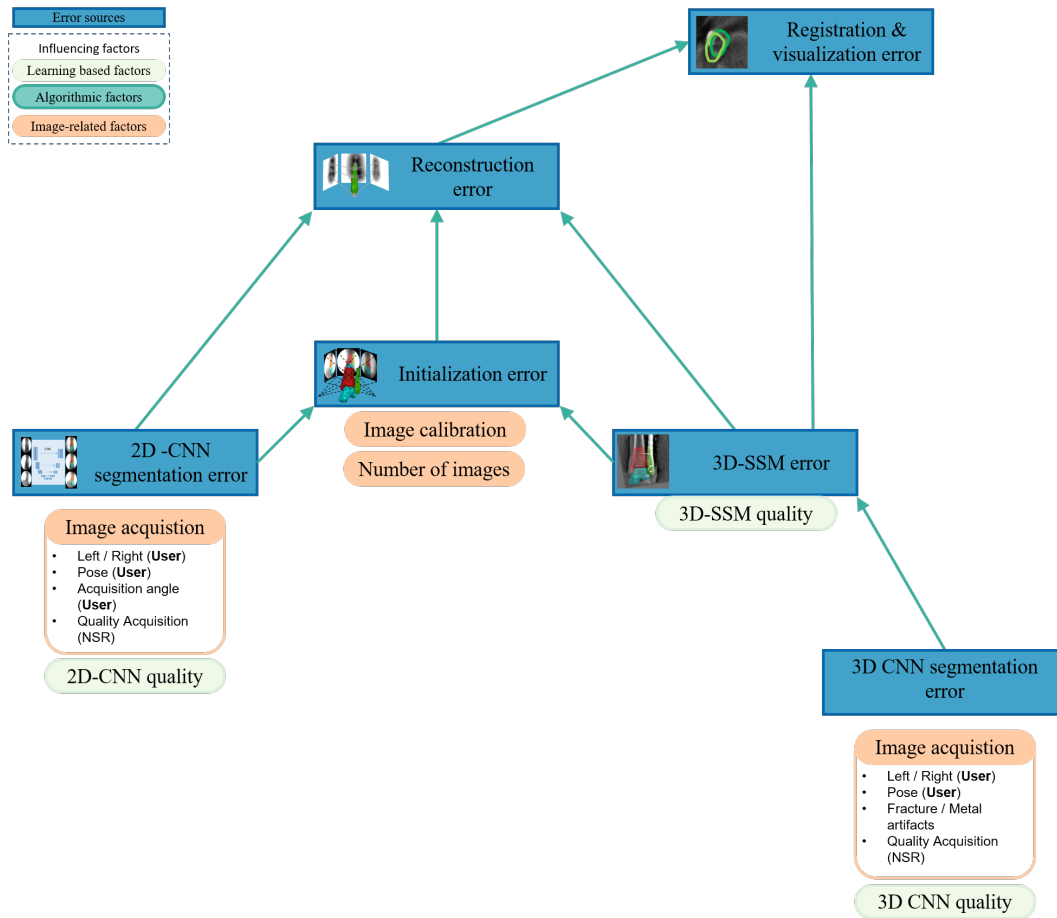


Figure 17: Hierarchical representation of different sources of error and their influencing factors. The arrows indicate the dependency between different components.

main influencing factors are misaligned and isolated bone fragments as well as artifacts from metal implants impeding the 3D segmentation result. The SSM error reflects the discrepancy between the actual shape of the injured ankle and the 3D model. The output of the 2D-CNN segmentation step depends on the input data and the quality of the trained CNN network. The image acquisition of the contralateral ankle is affected by the patient positioning towards the C-arm scanner and influences the performance of the 2D segmentation result. The surgeon acquires all projection images from certain viewports that may also affect the result. The projection images are related to each other by perspective projection matrices provided by device calibration. Thus, not only the selection of the viewports but also calibration inaccuracies influence the result as shown by JIANHUA YAO and TAYLOR 2003.

The 2D-3D initialization is influenced by the 2D segmentation and the 3D-SSM accuracy



and in turn, its outcome affects the 2D-3D reconstruction, which is dependent on all of the three components. Consequently, uncertainties of all components are propagated. Finally, the last component, namely the 3D-3D registration between the injured and the reconstructed uninjured model, depends on the shape deviation of the tibia models. Registration errors during rigid registration are induced by a deviation between the injured and uninjured tibia models. This discrepancy can be caused either during 3D-SSM segmentation and reconstruction or, on the other hand, originate from a natural intra-individual variance between the left and right ankle.

The experiments presented in chapter 5 were specifically designed to assess the combined overall error as well as different sources of error independently.

## 4.2 3D bone segmentation

Segmentation of multiple bones with metal artifacts in C-arm images has already been tackled in previous research projects. However, problems arise when transferring the methods to the anatomy of the ankle joint given the limitations of intra-operative imaging (examples are depicted in Fig. 18). One challenge of the segmentation task is the varying bone shaft length. As described in section 4.1.2, fibula and tibia are tubular bones that are not fully visible in the 3D image due to their length. The visible shaft proportion highly varies and depends on the ankle position relative to the C-arm iso-center during acquisition. Furthermore, metal implants cause image artifacts, often in large areas along the fibula shaft. Since different fracture types lead to a variety of treatments in which the placement of metal implants can highly differ, statistical appearance modeling is challenging. The following sections describe different methods for bone segmentation used in this project as well as extensions to previous methods to cope with arising challenges. The statistical shape model has been complemented with a deep learning-based approach that is presented in section 4.2.2.

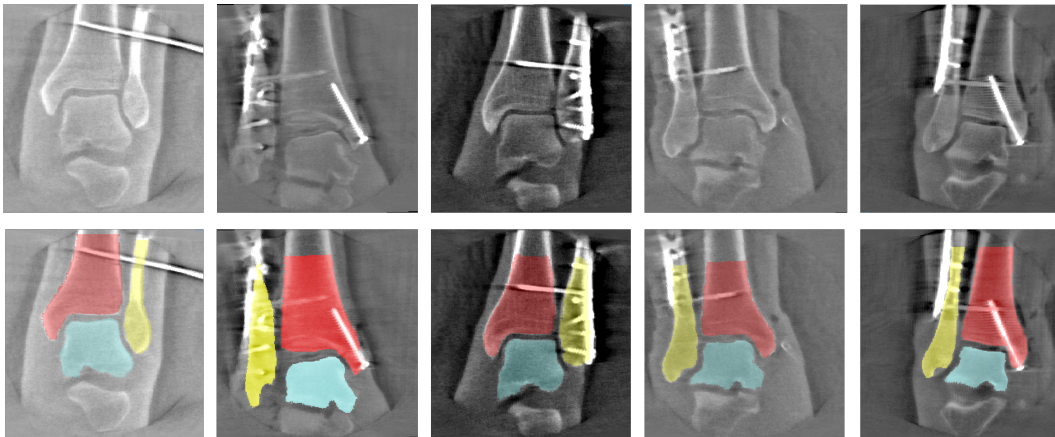


Figure 18: 3D C-arm images of the ankle joint after fibular reduction: Examples show different fracture treatments of the ankle and 3D images with different contrasts and metal implants. The second row shows the corresponding segmentations of tibia (red), fibula (yellow) and talus (cyan).

### 4.2.1 Statistical shape and appearance model

Statistical shape models (SSM) incorporate prior learned knowledge and provide a compact representation of the anatomy of interest. An anatomy can be described in terms of its shape characteristics and in terms of its appearance. Shape information can be derived from the training data represented by polygonal meshes of the structure's boundary. In addition, the pure image intensity of the training data can be analyzed to derive

the appearance of the anatomy represented by image characteristics. In this project, intensity-based modeling is used in combination with shape modeling, such that image intensities are evaluated at each position of the polygonal mesh. The mesh is updated based on a new expectation of point locations. All point expectations are then evaluated as a whole to determine a new plausible shape for the mesh. This is solved by maximizing the fit of the SSM to the expected point locations. The following sections provide the reader with an overview of how to get from annotated data of the ankle joint towards a statistical shape and appearance model resembling the anatomical variation by solving task-specific challenges.

### SSM creation

The underlying model representation forms the basis of the 3D-SSM. The PDM has been proven to be very successful in many segmentation applications due to its simplicity and efficient computation (HEIMANN and MEINZER 2009). A PDM consists of a distribution of landmarks and represents the mean shape geometry as well as descriptive modes of the shape variation. For model creation, each annotated patient data is described as a collection of  $k$  Euclidean landmarks for each bone.

$$\mathbf{X}_{\text{bone,patient}} = \{\mathbf{x}_i \mid i \in 1, \dots, k\} \text{ with } \mathbf{x}_i = [x_i, y_i, z_i] \quad (13)$$

The first step towards the PDM is the conversion from a voxel-based representation to a valid surface mesh with respective landmarks.

### Mesh creation

Manual segmentation of patient data results in a binary voxel grid for each bone. Segmentation inaccuracies might provoke holes or isolated voxels in the binary image, which are not related to the actual shape. In a first step, small holes are removed by applying basic image processing operations like binary closing or hole filling (SOILLE 2004). To eliminate isolated voxels, only the largest connected component is considered for mesh generation by analyzing the neighborhood of each voxel. The Marching Cube algorithm is applied to all remaining voxels to create a polygonal mesh from the iso-surface (LORENSEN and CLINE 1987). For each voxel, the algorithm builds a polygon based on the configuration of adjacent voxel positions. Manual segmentations may appear chiseled due to the limited spatial resolution of the image. Edges at image boundaries may impede the creation of natural shapes. Hence, the mesh generation step is finalized by a Gaussian smoothing after surface generation. In addition, a surface reduction can help

to accelerate the consecutive steps of the SSM generation by reducing the number of mesh vertices. The resulting triangle mesh is a closed, non-manifold surface.

### PDM creation

All polygonal meshes are created independently and not based on distinct anatomical landmarks. The PDM, however, is based on the assumption that all landmarks correspond to each other. Hence, a crucial step of the PDM creation is an accurate correspondence building. Successful in many segmentation applications, the widely used GPA (GOWER 1975) is employed for the efficient alignment of the training shapes. GPA computes the pairwise alignment of shapes iteratively to optimize a mean shape that minimizes the distances to all other samples. To ease the convergence of the GPA, the scale of the mean shape is normalized and all other shapes are scaled to the tangent space of the mean shape for convergence. One pre-requisite of the GPA is that an equal number of corresponding landmarks are provided that are distributed well over each training surface. This can be ensured by a simultaneous correspondence optimization step. The optimization of landmark correspondences is conducted following the approach of HEIMANN et al. 2005, who used the MDL as optimization criteria. In their work, they chose Gradient descent as optimization method since it allows the derivation of the objective function and can be computed time efficiently and automatically. Prior to optimization, a spherical parametrization is employed by mapping the meshes to a unity space for initial correspondence placement to ease the distribution of surface landmarks. This parametrization presumes closed, non-manifold surfaces that need to be guaranteed during foregone pre-processing steps.

Given a successful optimization, the PDM comprises an equal number of corresponding landmarks. As described in the state of the art (section 3.2) in detail, the final SSM is a concise representation of the PDM with reduced dimensionality. For that particular purpose, a PCA is applied to the landmark distribution  $\mathbf{X}$ . The resulting SSM can be described by

$$\mathbf{X}_{\text{bone}} = \bar{\mathbf{X}}_{\text{bone}} + \mathbf{P}\mathbf{b} \quad (14)$$

where  $\bar{\mathbf{X}}_{\text{bone}}$  denotes the mean landmark vector. Instead of using all eigenvectors and eigenvalues retrieved by the PCA, it is sufficient to rely on the largest eigenvalues to cover the majority of the shape variation. Consequently,  $\mathbf{P} = (p_1, \dots, p_t)$  contains the respective  $t$  eigenvectors associated with the largest  $t$  eigenvalues  $\mathbf{b}$ .

For the particular case of ankle segmentation, one major challenge is the cut contact zone of the fibula and tibia shaft with the image border. Tibia and fibula exceed the FOV of the scanner. As a consequence, a manual segmentation can only be performed up to

the area where the bones blur with the border of the image. At segmentation borders, surfaces suffer from sharp edges with high curvature. To make the model independent from highly varying shaft lengths, all tibia and fibula segmentations are cut along the shaft to a common length before 3D-SSM creation. It has to be noted, however, that this does not alleviate the problem completely since the respective shafts are cut absolute and scale-based differences are not taken into consideration. Another anatomy-specific characteristic is the quasi-rotation symmetric shape of the fibula. Since 3D-SSM generation and correspondence building is optimized for sphere-shaped objects with rather distinct morphology, the 3D-SSM generation of the fibula is prone to misalignment along the rotational axis. Therefore, one pre-processing step is the pre-registration of the segmentation volumes. This ensures the correspondence optimization process to converge into the local optimum.

After PCA, the variation of the bone morphology is incorporated in  $\mathbf{b}$  and can be used for statistical analysis by investigating the impact of  $\mathbf{b}$  among the population as well as for segmentation by finding a set of  $\mathbf{b}$  that fits the 3D-SSM to an unseen target image.

### Appearance model generation

One possibility to train an appearance model follows the approach proposed by HEIMANN 2009b. During training phase, intensity profiles along the surface normal of each landmark  $\mathbf{x} \in \mathbf{X}_{\text{training}}$  are sampled and an appearance model with mean and principal modes is created for each landmark. By altering the profile sample number and the spacing between two consecutive samples, the capture range can be increased. This allows training the model on different resolutions by creating models for different spacings. The definition of the appearance model can be manifold e.g. based on the sampling of image intensities or gradient values. In the thesis of HEIMANN 2009b, an extensive comparison of different sampling methods was conducted for a variety of applications and best results were achieved by using non-linear profiles. For this project, the gradient magnitude of the directional derivative is sampled which yielded good results in other related applications on C-arm images (GÖRRES et al. 2016). Furthermore, the approach of BRUIJNE and NIELSEN 2004 is employed in order to train a k-NN classifier from positive samples on the actual boundary and negative samples on distant profile locations. Learning the appearance for single landmarks leads to a rather sparse distribution since training data is usually limited. In medical images, neighboring landmarks are likely to show a profile similarity, thus, a landmark clustering can be performed on the resulting profiles. Several landmarks with similar appearances are combined to train a more compact shared appearance model with more samples that may improve generalization (HEIMANN 2009a).

### Shape Analysis

Since a 3D-SSM describes the statistical distribution of a shape, it can be used for comparison and variation analysis. The PCA creates a compact representation, in which the shape is defined only by the shape parameters  $\mathbf{b}$  modifying the mean model  $\bar{\mathbf{X}}$ . Two shapes are considered to be equal if they can be transformed into each other only by applying rotation, translation and scaling operations without modifying  $\mathbf{b}$ . For a pure shape analysis, different  $\mathbf{b}$  of all  $N$  shapes in a population can be compared directly.

$$\delta \mathbf{b}_p = \sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{|\mathbf{b}_{p,j} - \mathbf{b}_{p,i}|}{N} \quad (15)$$

One disadvantage of this approach is that the parameters are altered globally which implies that the comparison is not capable of a spatial local encoding of the global deviation. The alternative approach is based on the assumption that dependencies related to translation, rotation and scaling are invariant for the used 3D-SSM representation. Hence, the spatial landmark distribution can be employed directly to evaluate the local deviation at each landmark  $\mathbf{x}_n$  of all  $N$  shapes.

$$\delta \mathbf{x}_n = \sum_{i=1}^N \sum_{j=1, j \neq i}^N \frac{|\mathbf{x}_{n,j} - \mathbf{x}_{n,i}|}{N} \quad (16)$$

This evaluation allows to identify regions of high anatomical variance within each model and to compare intra- and inter-individual differences.

### 3D-SSM segmentation

As the main purpose besides shape analysis, the generated 3D-SSM can also be used for image segmentation on unseen patients. A new shape not contained in the training data can be approximated by:

$$\mathbf{X}_{\text{new}} \approx \bar{\mathbf{X}} + \mathbf{P}\mathbf{b}_{\text{new}} \quad (17)$$

During image segmentation, the 3D-SSM is fitted to the target anatomy by estimating shape parameters  $\mathbf{b}$  as well as a similarity transformation  $\mathbf{T}$  aligning the shape with the target anatomy. The task can be formulated as an optimization problem

$$\underset{\mathbf{T}, \mathbf{b}}{\operatorname{argmin}} |\mathbf{X}_{\text{new}} - \mathbf{T}(\bar{\mathbf{X}} + \sum_{p=1}^t \mathbf{P}_p \mathbf{b}_p)| \quad (18)$$

The first step is to place the initial shape instance e.g. the mean shape of the created 3D-SSM at an initial position in the image desirably close to the target. At starting position

$\mathbf{T}_{i=0}$ , the intensity is evaluated at each landmark position  $\mathbf{x} \in \mathbf{X}$ . A profile perpendicular to the surface is sampled from  $\mathbf{x}$  with profile spacing  $s$ . The aforementioned optimization problem is solved by minimizing the fitting costs for each profile and compute optimal landmark displacements. The fitting costs are based on the output of the trained appearance model for each landmark.

After determining local landmark displacements, a graph-cut algorithm is used to estimate final displacements following the approach of HEIMANN et al. 2007. The graph comprises nodes that resemble all profile positions as well as a sink and a source. Fitting costs are translated to edges and the graph is optimized in a maximum flow and minimum cut manner. It outputs an optimal solution of landmark updates based on global shape priors. A transformation update of  $\mathbf{T}_{i+1}$  can be computed by applying a GPA on the updated landmark positions. Alternating with the transformation update, estimated landmark adjustments are then projected onto shape space. A shape update  $\mathbf{b}_{i+1}$  is computed by altering eigenvalues  $\mathbf{b}_i$  of the 3D-SSM globally. This process is performed iteratively until an optimal solution is found.

Different approaches exist to increase the robustness of the 3D-SSM segmentation such as multi-resolution and multi-object optimization. The segmentation process is implemented as a multi-stage process, in which the profile spacing and the resolution of the image is decreased in each stage starting with a coarse resolution to roughly fit the model. To further increase robustness and to prevent the 3D-SSM from drifting towards false contours, a multi-object GPA is applied as suggested by GÖRRES et al. 2016. Instead of updating  $\mathbf{T}$  for each bone, a transformation  $\mathbf{T}_{\text{composite}}$  is defined, which simultaneously updates the transformation of all 3D-SSMs and preserves the relative position among the models. The single GPA is usually performed on the displacement vector of each model, while for the multi-object GPA all displacement vectors are concatenated first and Procrustes analysis is performed on the entire displacement vector. In a finer resolution stage, the 3D-SSMs are optimized separately with a finer-grained profile spacing.

Transferring the approach to the specific application of ankle surgery, some remarks have to be taken into consideration. During surgery and before image acquisition, the fibula is surgically restored in most cases. It can be assumed that the C-arm image does not contain any large bone fragments. The occurrence of smaller isolated bone fragments does not influence the 3D-SSM segmentation significantly. However, the sheer amount and the varying position of metal implants induced in surgery can lead to strong differences between expected image appearance and actual appearance, thus challenging the 3D-SSM segmentation. To overcome limitations caused by the classical appear-

ance modeling, the 3D-SSM is complemented with convolutional neural networks which are introduced in the following sections.

#### 4.2.2 3D-CNN-based segmentation

Recent developments in medical image processing showed that convolutional neural networks (CNN) can be successfully employed for the task of image segmentation and outperform previous methods (LITJENS et al. 2017). This section gives an overview of the specific task of 3D multiple bone segmentation in C-arm images.

##### Network architecture

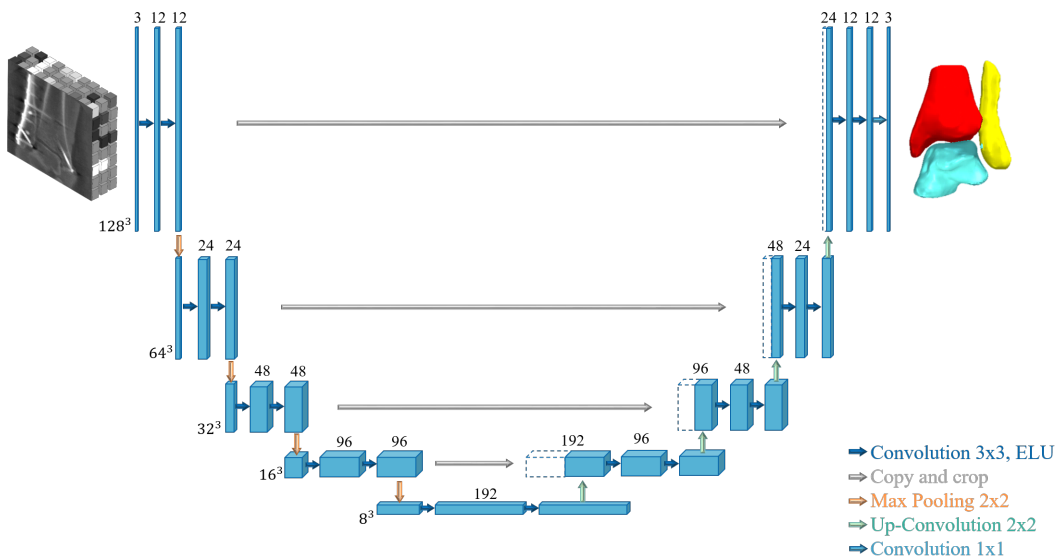


Figure 19: 3D U-Net architecture: Symmetric CNN for multi-class bone segmentation.

The original 2D U-Net architecture by RONNEBERGER et al. 2015 has already been introduced in section 3.2.1. It consists of a symmetric down- and upsampling path. Fig. 19 shows the architecture for the task of 3D segmentation which is similar to the 2D U-Net w.r.t the general structure. In the following, the different components of both paths are described in detail.

**Downsampling path:** As depicted in Fig. 19, each level of the downsampling path consists of two convolution layers with a kernel size of 3x3 and a stride of 1. A stride can be defined as the number of pixels, by which the convolutional filter is shifted over the input. Zero padding is applied to the convoluted image. Adding zeros around the border of



the input image ensures that the convolutional filter is performed on all pixels so that the resolution of in- and outputs remain the same (GERON 2017). In the original architecture, the aggregation of convolutional layers is supplemented by a batch normalization layer that normalizes the input across the entire batch and spatial locations. Here, a batch describes the number of input images that can be processed at a time. In the proposed architecture, this particular layer is replaced by an instance normalization layer, which normalizes each image independently and only across the spatial locations. Each level of the downsampling path is completed by an activation function, which introduces non-linearity to the network and decides whether an output feature is activated. Different non-linear activation functions can be employed in those so-called hidden layers of the network. In the proposed architecture, the *Rectified Linear Unit* (ReLU) of the original U-Net is replaced by *Exponential Linear Unit* (ELU) that in comparison to the former can produce negative outputs by adding a constant  $\alpha$  and have a smoother transition whereas ReLUs sharply smooth with input values  $y < 0$ . As stated by CLEVERT et al. 2015, ELUs accelerate network training while maintaining a more robust and stable representation due to their clear saturation plateau.

$$f_{\text{ReLU}}(y) = \begin{cases} y & \text{for } y > 0 \\ 0 & \text{for } y \leq 0 \end{cases}, \quad f_{\text{ELU}}(y) = \begin{cases} y & \text{for } y > 0 \\ \alpha \cdot (e^y - 1) & \text{for } y \leq 0 \end{cases} \quad (19)$$

The different levels of the downsampling paths are connected with a maximum pooling operation, which downsamples the input to half of the size for each dimension with a stride of 2.

**Upsampling path:** The bottommost layer with two more 3x3 convolution layers mediates between the downsampling and the so-called upsampling path. This path consists of an identical number of levels to build a fully symmetric composition. It consists of three convolutional layers, in particular, two strided convolutions with kernel size 3x3 and one transposed convolution with kernel size 2x2 to upsample and double the spatial dimension of the current feature map. Most importantly, the input of each layer is concatenated by the corresponding feature maps of the downsampling path through so-called skip-connections (see gray connections in Fig. 19) and further propagated through the subsequent convolutions. This ensures that the network retrieves the spatial information that is lost during maximum pooling layers otherwise. On the highest level, the transposed convolution layer is replaced by a convolution with kernel size 1x1. This final layer outputs a prediction map with the size of the input image times the number of the target classes.

The architecture is modified to allow a third dimension in the input path. This way, the whole volume can be processed at a time and volumes of arbitrary sizes can be inserted so that the processing is only limited by the GPU memory. There are several ways to deal with limited GPU memory capacity, which is a common issue in the field of 3D segmentation. The volume can be downsampled, cropped to the region of interest or processed through overlapping patches and aggregated afterward.

In summary, the proposed architecture is designed so that it can forward an unseen 3D image through the network and output a segmentation mask for each class once it is trained accordingly.

### Optimization

The resulting network architecture has got an enormous number of weight parameters which influence the network output. An essential part of training the CNN is the adjustment and optimization of those parameters. The image segmentation task can be formulated as an optimization problem that requires the definition of an objective function. Similar to the original U-Net paper, the widely applied cross-entropy loss is used as an objective function to measure the discrepancy between the output  $y$  and the ground truth reference  $\hat{y}$ . In each iteration, the network is predicting a probability distribution over all classes  $C$  for each input. To interpret the outcome  $y$  as probabilities, a softmax classifier is used for normalization prior to loss calculation.

$$\sigma_{\text{softmax}}(y_c) = \frac{\exp(y_c)}{\sum_{c'=1}^C \exp(y_{c'})} \quad (20)$$

It has to be noted that the softmax is not applied independently for each class but accounts for all classes. The cross-entropy loss can be defined as

$$\text{Loss}(f) = - \sum_{c=1}^C \hat{y}_c \log(y_c) \quad (21)$$

In multi-class segmentation, the ground truth labels are usually one-hot-encoded implying that only one class element  $c_p$  of the ground truth  $\hat{y}$  is not zero. Disregarding all other elements, softmax activation and cross-entropy can be combined and the loss can be reformulated for the given task to

$$\text{Loss}(f) = - \frac{1}{\text{BXYZ}} \sum_{b,i,j,k}^{\text{BXYZ}} \log(\sigma_{\text{softmax}}(y_{c=c_p, bijk})) \quad (22)$$

The loss function returns a scalar loss value by averaging over all dimensions includ-

ing the batch  $b = \{1, \dots, B\}$  and all three spatial dimensions  $i \in \Omega_X$ ,  $j \in \Omega_Y$  and  $k \in \Omega_Z$ . During optimization, the network parameters are altered with the aim of minimizing the loss function value. In each iteration, the input data is forward-propagated through the network. The output of each consecutive layer is computed and until the final output layer is reached to compute the loss. The algorithm also computes to which extent each neuron in the previous layer has contributed to that error. The computation is performed in a reversed manner until the input layer is reached. This so-called back-propagation step measures the error gradient across all network weights and allows to update the weights in a gradient descent manner (GERON 2017). For weight optimization, the *Adaptive Moment Estimation Algorithm* (ADAM) optimizer (KINGMA and BA 2015) is widely used in image segmentation tasks due to its minimal memory consumption and computational efficiency. ADAM is a combination of the gradient descent with momentum and the *Root mean square* (RMS) that optimizes the moments of the given objective function. The algorithm calculates an exponential weighted moving average and then squares the computed weight gradients. It has got two decay parameters that control the decay rates of these calculated moving averages and can be modified for the network training.

### Network training

The training process is not only influenced by architecture design choices, like the layer composition or the optimizer with the respective learning rate, but also by so-called hyperparameters set prior to the learning process. During training, the network is shown different input data to optimize network weights. Per iteration, the network is fed with a batch of images simultaneously. The training is conducted for a certain number of epochs. An epoch denotes the number of training iterations to optimize the network. The exact definition of an epoch varies among different projects. The epoch should be designed so that each training image is shown to the network once in each epoch. If the training data collection is augmented, often a pre-defined number of images are inserted randomly instead to statistically cover the entire data. For training purpose, the available data sets are usually split into training, validation and test sets. Training sets are used to modify the networks' floating weights. The validation set is used to estimate the quality of the learned network weights and to adjust the hyperparameters. The test set is a separate set, which must not be used during any of the training steps but solely for the final evaluation. After each epoch, the performance of the network with the current set of weight parameters is tested on the validation set to compute an additional loss. For detecting occurrences like overfitting this validation loss can be observed and compared

to the training loss. Overfitting describes the behavior that occurs whenever the validation loss starts to increase while the training data loss continues to decrease. It happens when the network is too complex in relation to the variation of the training data and not capable of generalization (GERON 2017).

The choice of hyperparameters is important but the key factor for the success of a neuronal network remains the reasonable choice of input data. Since the amount of annotated C-arm images is sparse, it is essential to enlarge the training data set by data augmentation before training the network.

### Input data augmentation

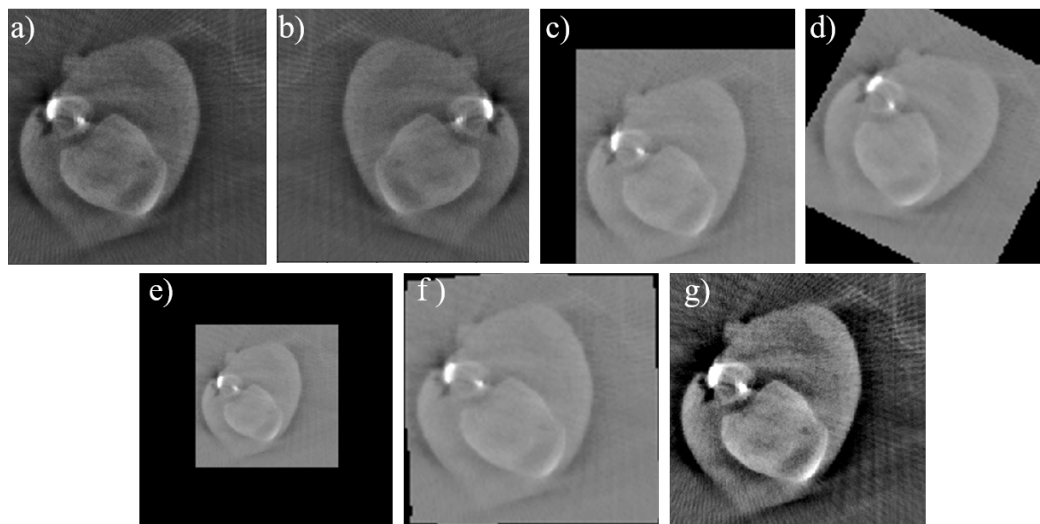


Figure 20: Data augmentation strategies on 3D C-arm images: (a) Original image, (b) reflection, (c) translation, (d) rotation, (e) scaling, (f) deformation and (g) contrast enhancement.

To derive and learn a meaningful representation of the underlying image intensity distribution, a reasonable amount of annotated training data is required, which remains one of the major challenges in the medical domain (UNBERATH et al. 2018). For the given task, only a limited number of training data is available and must be enlarged by applying data augmentation. As already mentioned in section 3.2.1, the methods used for data augmentation remain application-specific and based on the nature of the input. In the case of C-arm volumes of the ankle joint, useful augmentation methods mainly aim at an invariance of the training data towards spatial transformation and changes in intensity. The following data augmentation operations are applied:

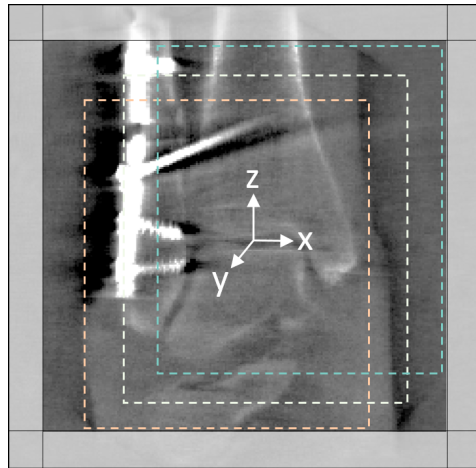


Figure 21: Cropping and image translation for data augmentation: The image is cropped randomly (colored rectangles) by sparing a border margin (transparent rectangle). Only two axes are displayed for better illustration although the augmentation is performed on all three axes.

**Reflection:** The input data contains left and right ankles. To attain invariance towards laterality, all input volumes are reflected along all spatial axes  $x$ ,  $y$  and  $z$ .

**Cropping and translation:** Each image is cropped to ensure that the segmentation of the bones does not reach the outer boundaries of the volume. In the ideal case, the ankle would be placed in the iso-center of the C-arm, but real clinical data shows deviations from the ideal placement. Therefore, a random crop with a margin of 15 voxels from the center is allowed which omits the need of any other translational transformation (cf. Fig. 21).

**Rotation:** The starting angle of C-arm 3D acquisition is not restricted. This flexibility introduces rotational variance, which can be further increased by the tilting angle of the foot relative to the scanner. To be invariant against different rotation angles, a rotation transformation is applied to the image. The rotation can be described by  $\alpha$ ,  $\beta$  and  $\gamma$  around each image axis  $x$ ,  $y$  and  $z$  in the interval of  $[-180^\circ, 180^\circ]$ .

**Scaling:** The input consists of ankles of different sizes and proportions depending e.g. on age and sex of the patient. Therefore, an isotropic scaling is applied to the data by applying a scaling factor  $s \in [0.9, 1.1]$  to obtain independence from varying sizes.

**Deformation:** A slight deformation was also applied but reduced to a minimum since the bones are rigid objects and only have minor local deformations due to the anatomi-

cal shape variance.

**Normalization and contrast variation:** As pre-normalization, a min-max normalization is applied to each 3D image.

$$\mathbf{y}_{\text{normalized}} = \mathbf{y} \cdot \frac{\max(\mathbf{y}) - \min(\mathbf{y})}{\max(\mathbf{y})} \quad (23)$$

The normalization linearly transforms the input values to the range of 0 to 1 based on the minimum and maximum intensity value. In literature, often z-scoring can be found as an alternative normalization method, which transforms the input values to a zero mean and unit variance distribution. However, this approach is not feasible whenever the intensity values of the input do not follow a Gaussian distribution. The contrast of each image varies depending on the intersected material density. To account for varying contrasts and brightness, contrast augmentation was applied to the image by modifying each input value of  $\mathbf{y}$  according to  $\mathbf{y}' = (\mathbf{y} - \mu) \cdot c + \mu$  with factor  $c \in [0.8, 1.2]$ , where  $\mu$  denotes the mean of the entire 3D image.

### 4.2.3 Combined approach: Deep learning-driven 3D-SSM segmentation

Although SSMs have been proven to be very useful, pure 3D-SSM-based segmentation has two disadvantages, namely its sensitivity towards variation from the learned intensity appearance as well as its dependence on a well-chosen starting position. CNN-based segmentation, on the other hand, can be more accurate even in the presence of metal implants but do not force a closed surface boundary. Hence, mislabeled voxels can lead to holes or non-realistic boundaries. Furthermore, a segmentation alone cannot replace the model representation since landmark correspondences are required for the successive 2D-3D reconstruction step.

In this project, the advantages of both approaches are combined and used in a reciprocal manner. The semantic knowledge, derived from CNN predictions, is first used for initialization of the 3D-SSM. Then, the 3D-SSM uses the predictions as underlying appearance. This way the 3D-SSM accuracy can be improved by providing a sufficient starting position and distinct binary edges. The following sections give a detailed description of modifications regarding 3D-SSM initialization and segmentation.

#### Model initialization

In general, conventional SSMs are sensitive towards initialization and depend on a sufficiently good starting point to find the global optima during the local search. The ini-

tialization of the 3D-SSM can be improved by incorporating the predictions from the U-Net based bone segmentation. For each predicted binary volume, the largest connected component is extracted and the center of mass is calculated. Simultaneously, the center of mass of all single 3D-SSMs is calculated. All resulting centers are used as landmarks for a landmark-based registration. The resulting transform  $\mathbf{T}_{\text{initial}}$  resembles the best mapping between the two landmarks sets in a least-square sense. To further refine the initial starting position, a surface is created from all voxels belonging to the tibia. A fine transformation is established by applying an ICP (PAUL BESL 1992) on the predicted tibia surface and the tibia 3D-SSM. The resulting transformation matrix is used to transform all 3D-SSMs. This way the 3D-SSM is positioned close to the target position and independent towards patient positioning.

### SSM segmentation using CNN predictions

Similar to the original 3D-SSM workflow, a global graph-cut based optimization is used to determine the final local landmark displacements. Instead of using the pure intensity values, the proposed method uses the CNN predictions for the model fitting. At first, an unsigned distance transform  $f_{\text{dist}}(x)$  is applied to the predicted segmentations  $\mathbf{S}$ .

$$f_{\text{dist}}(x) = d(x, \delta\Omega_{\mathbf{S}}) \text{ if } x \in \Omega_{\mathbf{S}} \quad (24)$$

$$d(x, \delta\Omega_{\mathbf{S}}) := \min_{y \in \delta\Omega_{\mathbf{S}}} d(x, y) \quad (25)$$

where  $\Omega_{\mathbf{S}}$  denotes the boundary of the segmentation  $\mathbf{S}$ . This way, a 3D image is created, in which each voxel is mapped to its Euclidean distance towards the closest boundary with zero resembling the border itself. A local landmark shift can be calculated for each landmark position by identifying the minimum value of the distance image along a landmark profile. This shift can be directly used as fitting costs for the optimization. The proposed method creates intact shape representations and reliable landmark correspondences combined with high segmentation quality as produced by the CNN method.

### The segmentation algorithm as a whole

The modifications described in the previous chapters can be included in the training and segmentation process. Fig. 22 illustrates the whole training and segmentation pipeline. 3D images and corresponding annotated manual segmentations are used to train not only an SSM for each bone but also an U-Net. In the online segmentation phase, the U-Net creates segmentation masks for each bone. With those, the SSM can be pre-aligned

to a position close to the optimal position. A distance transform on the segmentation masks serves as input for the final SSM adaption.

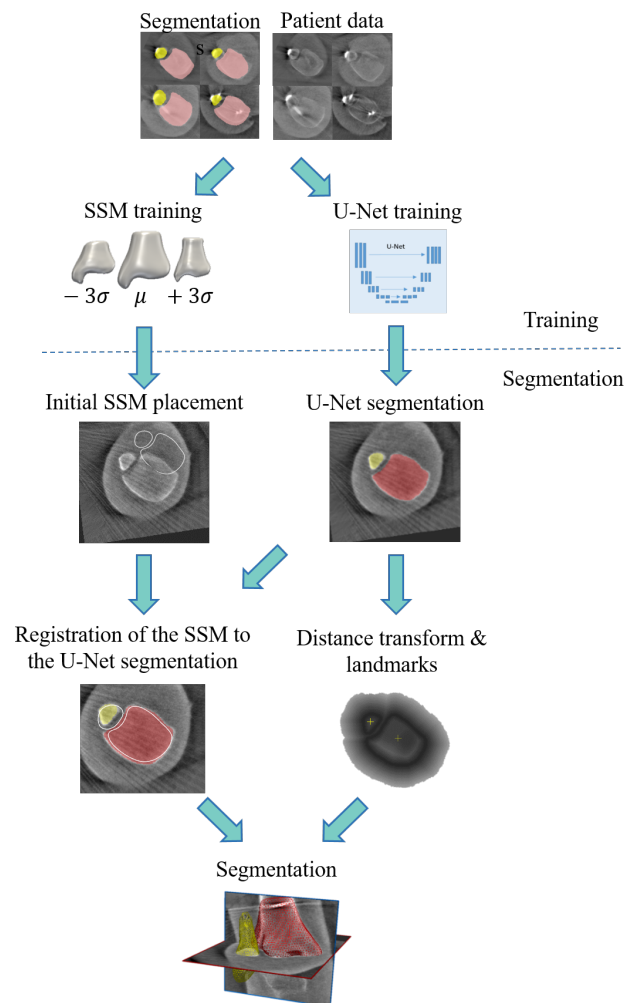


Figure 22: The segmentation algorithm as a whole: 3D-SSM segmentation incorporating CNN prediction during initialization and optimization. The figure was adapted from THOMAS et al. 2019.



### 4.3 Fluoroscopic image understanding

In this project, a semantic analysis of the fluoroscopic projection images is an essential part of the proposed workflow. Unique features, derived from the images, serve as anchor points for the subsequent steps. The following chapter provides technical details on image segmentation and contour extraction (section 4.3.1). The resulting segmentations can be utilized to compute anatomical landmarks that are necessary for an automatic and reliable coarse 2D-3D registration. Section 4.3.2 provides details on the landmark extraction in 2D C-arm images of the ankle joint acquired from varying viewports.

#### 4.3.1 2D image segmentation

In the specific task of bone segmentation applied to 2D C-arm images, major challenges include noise and inhomogeneities as well as artifacts from superpositioned tissue and external objects (cf. Fig. 23).

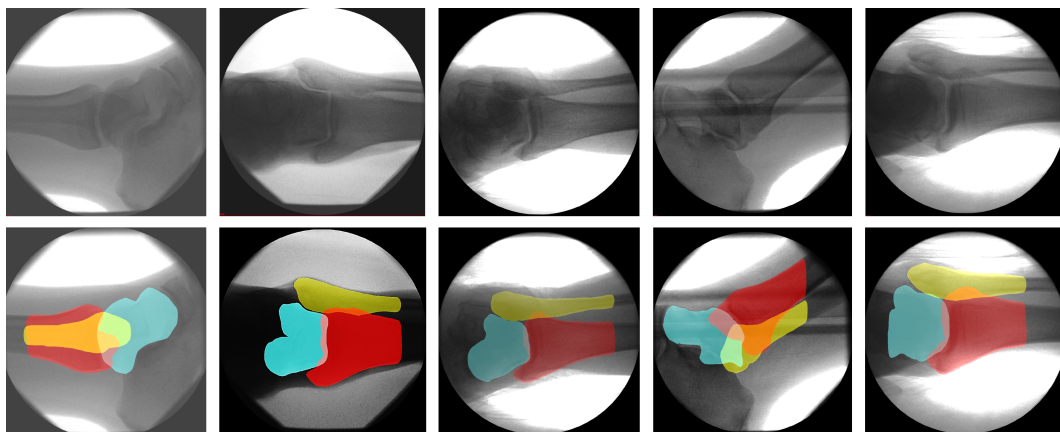


Figure 23: Fluoroscopic projection images of the ankle joint: Different examples show varying viewports on the ankle that are affected by different contrasts, noise or external objects. The second row shows the corresponding segmentations of tibia (red), fibula (yellow) and talus (cyan).

The aforementioned can often not be handled by classical image processing algorithms, which only provides features that are not distinguishable with regards to single bones. In the case of the ankle joint, each fluoroscopic projection image shows three or more bones with highly overlapping regions. Those particular multiple bone areas and the respective overlap proportion varies with the pose of the single bones and can be different for each patient. Superimposing areas usually contain diffuse image boundaries that further impede feature extraction. This results in ambiguous and miss-classified features with undesired background noise (THOMAS et al. 2017a). Furthermore, the viewport is

not limited with respect to the C-arm pose towards the ankle during the orbital rotation. This circumstance causes a highly varying appearance of shape and intensity.

To prevent the 3D-SSM from adapting its boundaries to false contours, unique features need to be extracted for each of the bones. A CNN-based segmentation method is applied to label the three partly overlapping bones tibia, fibula and talus. The segmentation masks can be used in turn to compute edge features for the next algorithmic steps. This comes with the advantage that the approach can be performed on each bone separately without any noise or false features resulting from overlap or external objects.

**Network architecture**

Segmentation of superpositioned bones can only be performed under the assumption that each pixel may potentially belong to more than one class as described in section 3.3. A modified version of the original U-Net developed by RONNEBERGER et al. 2015 is employed to address this multi-label, multi-class problem. Details of the network architecture can be taken from section 4.2.2. The main difference to the aforementioned architecture is the dimensionality of the input. While the proposed 3D U-Net is designed to segment the 3D images, the U-Net for fluoroscopic image segmentation is provided with a 2D input like proposed in the original architecture. Similar to the proposed modifications made to the 3D U-Net, instance-wise normalization and ELU activation layers are employed as opposed to the original implementation.

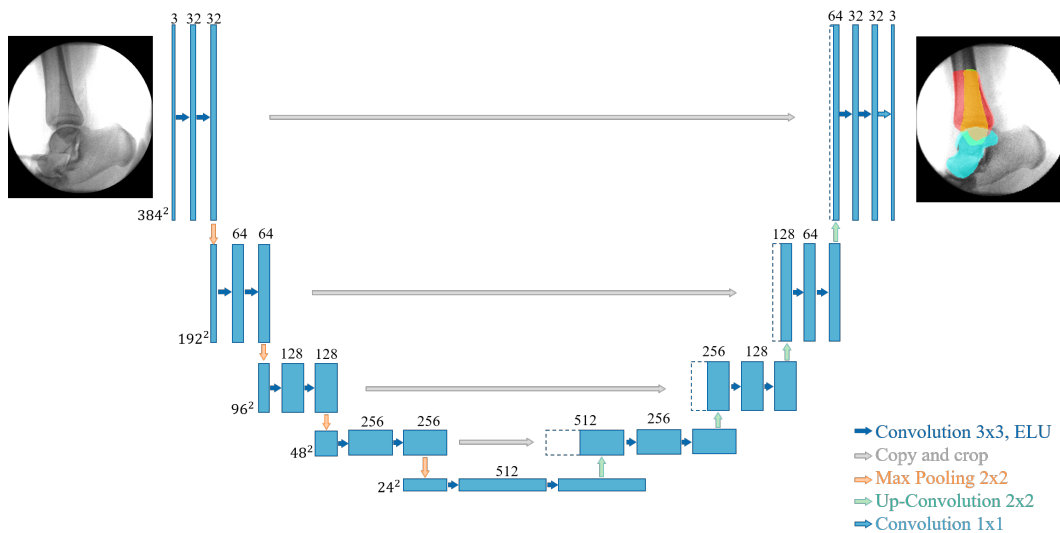


Figure 24: 2D U-Net architecture

### Optimizer and loss function

The weight optimization of the 2D segmentation task is conducted in the same fashion as the 3D segmentation task. Network parameters are optimized by comparing the networks' output to the 2D ground truth segmentation for each bone separately. The output feature maps with one layer per class contain probabilities for each class and each pixel. The final layer comprises of a sigmoid layer as opposed to a softmax layer in the original architecture to account for overlapping labels.

$$\sigma_{\text{sig}}(y) = \frac{\exp(y)}{\exp(y) + 1} \quad (26)$$

The sigmoid function produces a "S-like" shape and maps the logits of the output layer to  $[0,1]$ . The softmax function returns a probability per class and the target class should have the highest value since all probabilities cumulate to 1. As opposed to the softmax output, resulting probabilities of the sigmoid function do not necessarily cumulate to 1. Hence, the target class is not determined by identifying the maximum probability but by applying a threshold of  $\sigma_{\text{sigmoid}}(y) \geq 0.5$ . The pixel-wise difference between the binary ground truth segmentation and the resulting logits used as loss metric.

To allow an assignment to multiple classes, a binary cross-entropy loss is applied as opposed to the original U-Net implementation. The computed loss between prediction  $y$  and ground truth  $\hat{y}$  should not be influenced by the output of other classes. The name of the loss originates in its reformulation of the classical cross-entropy loss as a binary classification problem for each class with  $C = 2$ . If the number of classes  $C$  is greater than 2, the sum over all classes is built and can be applied in multi-class and multi-label segmentation tasks.

$$\text{Loss}(f) = \frac{1}{\text{BXY}} \sum_{b,i,j} \sum_{c=1}^C -\hat{y}_{c,bij} \cdot \log(\sigma_{\text{sig}}(y_{c,bij})) - (1 - \hat{y}_{c,bij}) \cdot \log(1 - \sigma_{\text{sig}}(y_{c,bij})) \quad (27)$$

Again, the loss is averaged over all dimensions including the batch  $b = \{1, \dots, B\}$  and both spatial dimensions  $i \in \Omega_X$  and  $j \in \Omega_Y$ .

In literature, e.g. elaborated in the original U-Net paper (RONNEBERGER et al. 2015), overlapping regions were handled differently by the loss function. A weighting scheme can be added to penalize under-represented labels like overlapping segmentations or border regions. This way, under-represented labels can be given more impact by increasing the weight and penalizing the error in those regions. However, in the case of images of the ankle joint, all classes are represented almost equally (examples are given in Fig. 25). The potential bone overlap varies with each image and can not be handled as

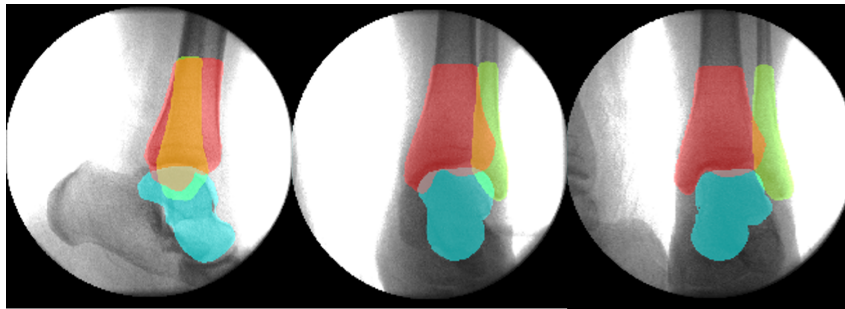


Figure 25: 2D segmentation of fluoroscopic images of the upper ankle joint: Overlapping regions vary among the different viewports.

an under-represented label. Thus, a weighting scheme to emphasize overlapping structures is rather unfeasible and not further elaborated.

The ADAM optimizer is selected for weight optimization since it shows good performance for the proposed 3D segmentation and has been applied successfully in many recent CNN optimization tasks. For details on optimization and network training, the reader is referred to section 4.2.2. After weight optimization through back-propagation of the loss, the network can be used to perform a segmentation on a fluoroscopic projection image and create a segmentation mask for each of the bones.

#### Data augmentation on 2D fluoroscopic projections

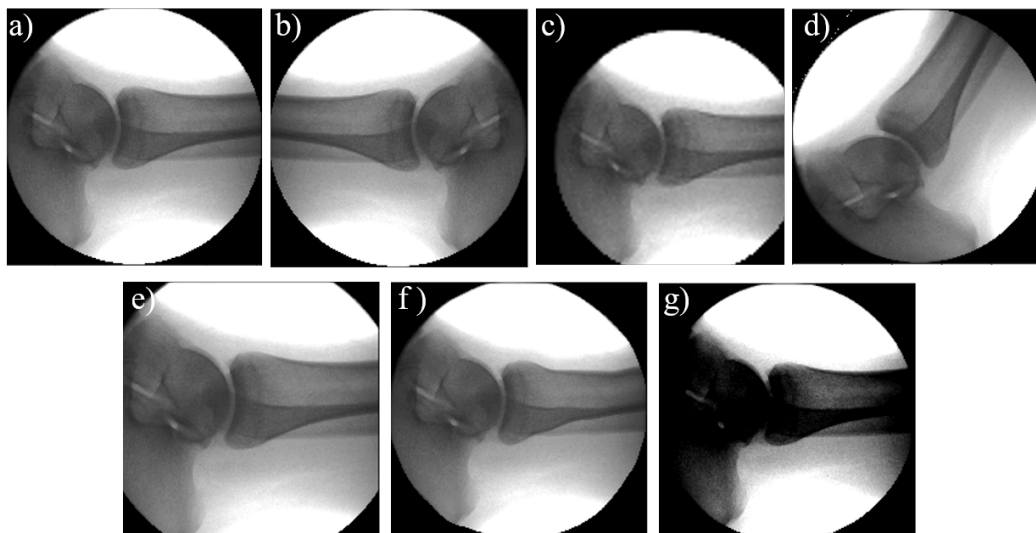


Figure 26: Data augmentation methods on fluoroscopic images: (a) Original image, (b) reflection, (c) translation, (d) rotation, (e) scaling, (f) deformation and (g) contrast enhancement.

Since one limitation of all machine learning-based approaches is the need for enough

training data, data augmentation is essential for the diverse fluoroscopic images especially considering the degree of freedom and the variability of the presentation. Different strategies have been already elaborated in section 4.2.2 for enlargement of the 3D training data set and are modified for the 2D case. Analogously to 3D data augmentation, operations like reflection, rotation and scaling operations are applied to all image axes  $i \in [0, 1]$ . The size of the input image is reduced by a random crop with a deviation of 50 pixels from the image center in both dimensions. Slight deformation, contrast enhancement and min-max normalization are applied similarly to the 3D case. Fig. 26 shows examples for all different augmentation strategies.

### 4.3.2 Landmark extraction

A sufficiently well-chosen starting position is essential for the success of a 2D-3D reconstruction. The preceding segmentation already distinguishes between different bones and reveals their relative spatial location. To benefit from this prior semantic knowledge, landmarks are extracted from the segmentation masks directly instead of the intensity image.

In a first step, the tilting angle of the ankle joint is extracted taking advantage of the assumption that bone shafts can be approximated by straight line segments in 2D. Since the shaft lines highly vary with the viewport, the amount of data in this project is rather insufficient for applying machine learning approaches for shaft line detection. Instead, classical image processing methods are applied to the segmentation images  $\mathbf{S}$ . First, a Sobel operator extracts binary edges on  $\mathbf{S}_{\text{fibula}}$  and  $\mathbf{S}_{\text{tibia}}$ . A Hough transformation, introduced by DUDA and HART 1972, is applied to each edge image to extract all straight lines. Each pixel  $\mathbf{p} = [x_1, x_2]$  can be represented by a curve in a space spanned by angle  $\alpha$  and radius  $r$  following the parametric equation:

$$x_2 = -\frac{\cos \alpha}{\sin \alpha} \cdot x_1 - \frac{r}{\sin \alpha} \quad (28)$$

Every line in the edge image corresponds to a point cluster in the  $\alpha$ - $r$  space given that  $\alpha$  is restricted to  $[0, \pi]$ . Thus, point clusters allow extracting even disconnected line segments with the same slope. Long line segments have a high probability of belonging to a bone shaft. By computing the mean angle of all long line segments, the tilting angle of the ankle can be estimated.

In the next step, all tibia and fibula points are projected 1) along and 2) orthogonal to the estimated tilting angle. With this, the width and the length of the ankle can be estimated (cf. Fig. 27). One landmark, denoted as  $\mathbf{p}_{\text{intersect}}^{2D}$  resembles the lowermost point of the tibia. It is defined as the first intersection point of the tibia with the talus

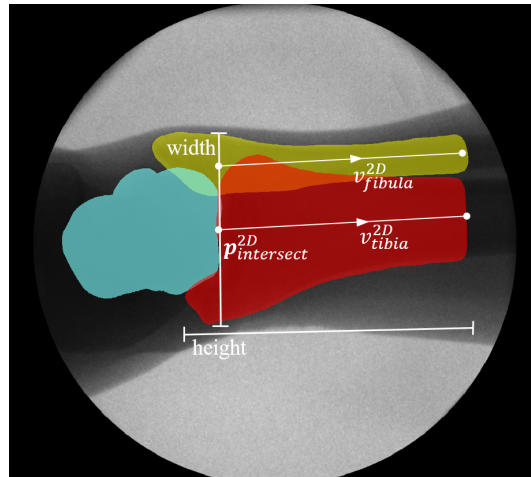


Figure 27: Landmarks of fluoroscopic images: Shaft line vectors for tibia (red) and fibula (yellow) and intersection point of tibia and talus (cyan).

segmentation. To derive the shaft line vectors of tibia and fibula, a 2D PCA is applied to the pixel grid of each  $\mathbf{S}$ . Since tibia and fibula are tubular bones, the first eigenvector of the PCA is most likely orientated along the shaft line and can be directly used as a shaft vector. Thus, the landmarks can be defined by a center point  $\mathbf{p}_{\text{intersect}}^{2D}$  and a shaft line vector for fibula denoted as  $\mathbf{v}_{\text{fibula}}^{2D}$  and tibia denoted as  $\mathbf{v}_{\text{tibia}}^{2D}$  (as illustrated in Fig. 27).

## 4.4 2D-3D multiple bone reconstruction

2D-3D multiple bone reconstruction forms the basis of the proposed pipeline and aims at the creation of a 3D anatomical surface model by fitting a 3D-SSM as precisely as possible to a set of 2D images. In the proposed work the 3D-SSM of the injured ankle  $\mathbf{X}_{\text{injured,bone}}$  serves as a reference and is mapped to 2D segmentations extracted from fluoroscopic images of the uninjured ankle. This way, an anatomical surface model of the uninjured ankle can be created without the need of acquiring an additional 3D image. The previous chapters provide the reader with the required pre-processing steps like the creation of the 3D-SSM and the semantic segmentation of fluoroscopic images. For a successful feature-based 2D-3D reconstruction, two core requirements need to be considered, namely, a well-chosen initial placement of the 3D-SSM and an accurate, robust optimization of the 3D-SSM parameters  $\theta := [\mathbf{T}, \mathbf{b}]$ . In the initialization step, described in section 4.4.1, the 3D-SSM is aligned to the coordinate system spanned by the virtual intersection of the projection images in 3D space to obtain an initial starting position. After initialization, 3D-SSM optimization is performed iteratively by altering the transformation parameters  $\mathbf{T}_{\text{uninjured}} \in \theta$  and the shape parameters  $\mathbf{b}_{\text{uninjured}} \in \theta$  to fit the model to the contours of the projection images, which is elaborated in section 4.4.2.  $\mathbf{X}_{\text{injured,bone}}$  denotes the 3D-SSM of the injured ankle and  $\mathbf{X}_{\text{uninjured,bone}}$  denotes the 3D-SSM of the uninjured ankle. In the following chapter, the subscript 'bone' is left out for simplicity unless the method is explicitly applied to a specific bone instead of all bones.

### 4.4.1 Coarse model initialization

A well chosen set of initial starting values  $\theta_0$  is essential for the success of a non-convex optimization. The preceding 3D segmentation of the injured ankle, described in chapter 4.2, results in a 3D-SSM  $\mathbf{X}_{\text{injured}}$  with parameters  $\hat{\theta}_{\text{injured}}$ . The 3D-SSM of the uninjured ankle  $\mathbf{X}_{\text{uninjured}}$  can be initialized with that  $\hat{\theta}_{\text{injured}}$ , in particular with the transformation  $\hat{\mathbf{T}}_{\text{injured}}$  and the shape parameters  $\hat{\mathbf{b}}_{\text{injured}}$ . Under the assumption that the intra-individual variation of the patients' left and right ankle is considerably low, shape parameters  $\hat{\mathbf{b}}_{\text{injured}}$  should not deviate significantly from the desired  $\mathbf{b}_{\text{uninjured}}$ . However, the transformation parameters  $\hat{\mathbf{T}}_{\text{injured}}$  and  $\mathbf{T}_{\text{uninjured}}$  are likely to show a substantial discrepancy. This circumstance is based on the independent acquisition of the acquired 3D image of the injured ankle  $\mathbf{A}$  and the projection images  $\mathbf{B}_j$  with  $j \in \{1, \dots, n\}$  of the uninjured ankle. For both acquisitions, the mobile C-arm system offers a large number of possible device-to-patient orientations. Thus, both modalities do not share the same coordinate system. The goal of a coarse initialization established by transformation  $\mathbf{T}_{\text{initialization}}$  is to provide a good estimate for  $\mathbf{T}_{0,\text{uninjured}}$  by aligning both coordinate

systems.

$$\mathbf{T}_{0,\text{uninjured}} = \mathbf{T}_{\text{initialization}} \cdot \hat{\mathbf{T}}_{\text{injured}} \quad (29)$$

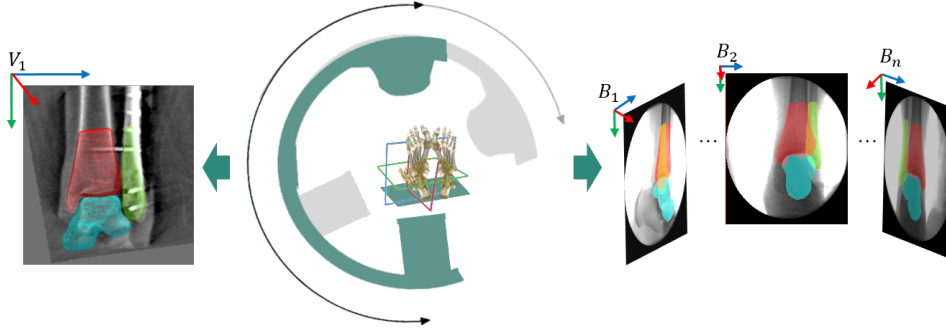


Figure 28: Acquisition of the 3D image  $V$  and projection images  $B$ : Both image modalities do not share the same coordinate system since the C-arm is moved during acquisition.

The injured and uninjured ankle of one patient is always assumed to be symmetric, which implies that in order to compare or align them, one ankle needs to be mirrored. This problem can be solved by applying a 3D reflection transformation  $\mathbf{T}_{\text{reflection}}$  on  $\mathbf{X}_{\text{uninjured}}$  prior to initialization.

$$\mathbf{T}_{0,\text{uninjured}} = \mathbf{T}_{\text{initialization}} \cdot \mathbf{T}_{\text{reflection}} \cdot \hat{\mathbf{T}}_{\text{injured}} \quad (30)$$

In the previous chapter 4.3.2, a method is described to calculate landmarks in the projection images, in particular the lower intersection point of the tibia ( $\mathbf{p}_{\text{intersect},j}^{2D}$ ), the tibia shaft direction vector ( $\mathbf{v}_{\text{tibia},j}^{2D}$ ) and the fibula shaft direction vector ( $\mathbf{v}_{\text{fibula},j}^{2D}$ ). Assuming that at least two projection images are provided, the landmarks can be triangulated by back-projection of the points with the inverse projection matrix  $\mathbf{P}^{-1}$  to obtain  $\mathbf{p}_{\text{intersect}}^{3D}$ ,  $\mathbf{v}_{\text{tibia}}^{3D}$  and  $\mathbf{v}_{\text{fibula}}^{3D}$ . Given  $j > 2$ , the points can be pairwise triangulated and averaged afterward to refine the result.

Prior to reconstruction, the upper- and lowermost points of the shafts in the SSM mean representation of tibia  $\bar{\mathbf{X}}_{\text{tibia}}$  and fibula  $\bar{\mathbf{X}}_{\text{fibula}}$  are annotated as landmarks to build two respective SSM shaft vectors. The initialization is conducted by the following steps:

- The lowermost point of  $\mathbf{X}_{\text{tibia},\text{uninjured}}$  is translated to the 3D intersection point  $\mathbf{p}_{\text{intersect},j}^{3D}$  and a rotation matrix  $\mathbf{R}$  is computed to rotate the SSM tibia shaft vec-



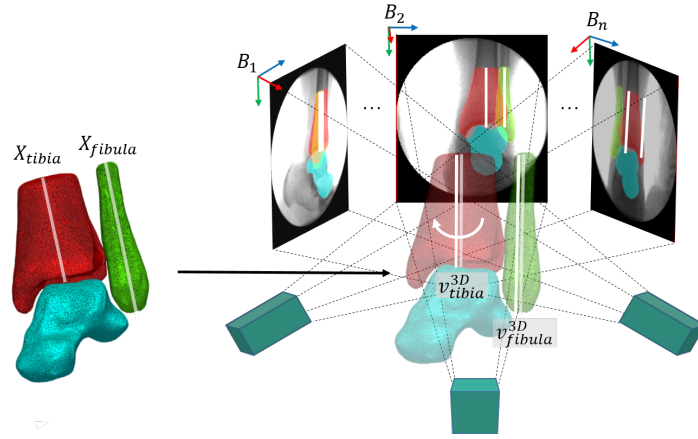


Figure 29: 2D-3D initialization procedure: Triangulation of 2D landmarks and mapping to 3D-SSM landmarks

tor to the triangulated 2D shaft line.

$$\mathbf{R} = \mathbf{I} + [v]_x + [v]_x^2 \frac{1}{1 + \cos(\angle(s_1, s_2))} \quad \text{with} \quad v = s_1 \times s_2 \quad , \quad [v]_x = \begin{vmatrix} 0 & -v_3 & v_2 \\ v_3 & 0 & -v_1 \\ -v_2 & v_1 & 0 \end{vmatrix} \quad (31)$$

where  $s_1$  and  $s_2$  denote the shaft line vectors. As a result, both, the triangulated 2D and the 3D vectors related to the tibia are aligned according to their position. The only remaining degree of freedom is given by the angle of rotation  $\alpha$  along the triangulated 2D shaft axis.

- The fibula vector is rotated along the computed tibia shaft axis and for each angle  $\alpha$  the distance to the 3D fibula shaft axis is determined. The 3D-SSM is transformed according to the rotation angle  $\alpha_{\min}$  with the shortest distance to obtain the coarse alignment  $\mathbf{T}_{\text{initialization}}$ .

After initialization, the 3D-SSMs are located closely within the boundaries of the corresponding segmented bone contours and the actual reconstruction process can be started.

#### 4.4.2 Optimization

The optimization aims to find transformation parameters  $\mathbf{T}$  and shape parameters  $\mathbf{b}$  to fit the initialized SSM to the 2D images as precisely as possible. In order to solve the optimization problem, 2D-3D correspondence pairs between the 3D model defined over  $\Omega_A \in \mathbb{R}^3$  and a set of 2D images  $B = \{B_j \mid j \in 1, \dots, N_B\}$  need to be established.

The optimization problem can be solved by minimizing the following distance-based cost function  $\epsilon_{\text{dist}}$ :

$$\arg \min_{\theta} \sum_{j=1}^{N_B} \epsilon_{\text{dist}} \left( A^T(\mathbf{x}), \mathbf{P}_j^{-1}(B_j(\mathbf{y}_j)) \right) \quad (32)$$

where  $\mathbf{x} \in \Omega_A$  denote landmarks of the 3D-SSM  $\mathbf{X}_{\text{injured}} = \{\mathbf{x}_i \in \mathbb{R}^3 \mid i \in 1, \dots, k\}$ . Feature points  $\mathbf{y}_j \in \Omega_{B_j}$  are extracted from 2D images  $B_j$  and are back-projected in 3D space by the inverse perspective projection matrix  $\mathbf{P}^{-1}$ .

In feature-based reconstruction, unique correspondence pairs are usually not available (MARKELJ et al. 2012). Instead, corresponding feature subsets  $\mathbf{x}_j^* \subseteq \mathbf{x}$  and  $\mathbf{y}_j^* \subseteq \mathbf{y}_j$  have to be searched iteratively for each projection by minimizing the distance between both subsets. Shown in Fig. 30, the correspondence building step is performed by

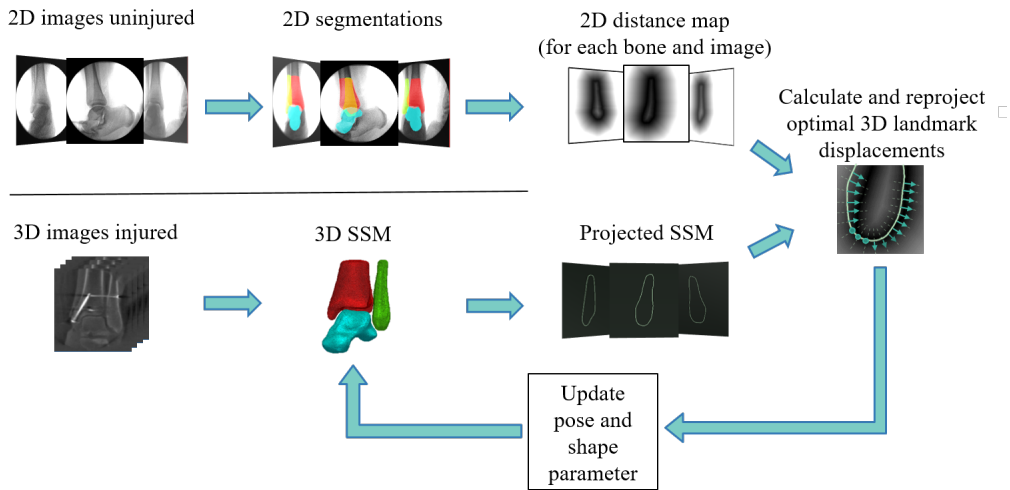


Figure 30: 2D-3D reconstruction overview: From all 2D images, segmentation masks and distance maps are extracted (top). The 3D-SSM is initialized using the 3D image of the injured ankle and projected onto the respective projection planes (bottom). Optimal landmark displacements of the 3D-SSM are computed and used to update shape and pose parameter in each iteration (right). The figure was adapted from THOMAS et al. 2019.

projecting  $\mathbf{X}_{\text{injured}}$  onto each projection plane of  $\mathbf{B}_j$  first so that both modalities share the same dimension. The iterative optimization alternates between correspondence building in 2D, back-projection of the established corresponding points  $\mathbf{y}_j^*$  in 3D space and updating of pose and shape parameters. Hereinafter, details are given on the correspondence estimation for both modalities as well as optimization strategies.

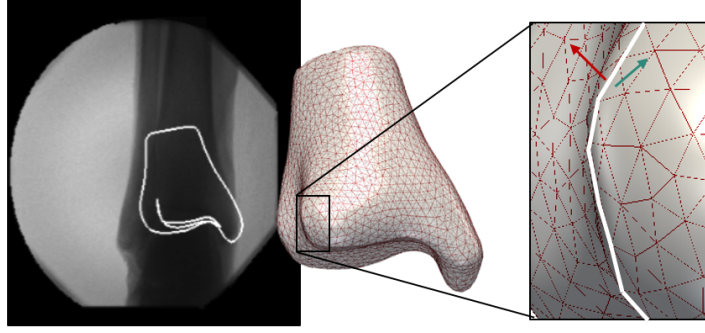


Figure 31: 3D-SSM projection and silhouette generation

### Correspondence building and optimization metric

The feature set  $\mathbf{x}_j^*$  holds a landmark subset of the 3D-SSM  $\mathbf{X}_{\text{uninjured}}$  and will be defined for each projection  $j$ . Shown in Fig. 31, the subset resembles the 2D silhouette of the SSM once it is projected onto the projection plane of  $\mathbf{B}_j$  using the corresponding projection matrix  $\mathbf{P}_j$ . To extract the subset, a silhouette can be computed based on the assumption that silhouette edges have the following property: Both adjacent face normals of the edge have a differently signed direction towards the normal  $\mathbf{n}_{B_j}$  of  $\mathbf{B}_j$  (illustrated by the red and green arrow in Fig. 31). A disadvantage of this approach is the creation of inner contours, which hamper the comparison to the closed segmentation mask. To eliminate all inner contours, the entire triangle mesh of the 3D-SSM is projected to create an image stencil. Projected landmark candidates are omitted if they are distant to the outer boundary of the stencil. Each candidate in  $\mathbf{x}_j^*$  of the resulting silhouette has a 2D pixel location on the image plane  $\mathbf{B}_j$ .

The reference features  $\mathbf{y}_j^*$  are derived from the CNN segmentations  $\mathbf{S}_j$  of image  $\mathbf{B}_j$  (described in section 4.3.1) for each bone and projection  $j$ . First, segmentation image  $\mathbf{S}_j$  is pre-processed by applying a Sobel edge filter to extract contours  $\mathbf{C}_j$  from the segmentations. A subsequent dilation operation ensures that no small holes occur by morphological increasing the width of the contour. On the resulting binary contour images an unsigned distance function  $f_{\text{dist}}(\mathbf{x})$  is applied.

$$f_{\text{dist}}(\mathbf{x}) = d(\mathbf{x}, \delta\mathbf{C}_j) \quad \text{if } \mathbf{x} \in \mathbf{C}_j \quad \forall \mathbf{x} \in \Omega_{C_j} \quad (33)$$

$$d(\mathbf{x}, \delta\mathbf{C}_j) := \min_{\mathbf{y} \in \delta\mathbf{C}_j} d(\mathbf{x}, \mathbf{y}) \quad (34)$$

The output is a distance map, also referred to as Chamfer distance (BARROW et al. 1977), in which each pixel corresponds to the distance  $d(\mathbf{x}, \delta\mathbf{C}_j)$  to the closest border of con-

tour  $\mathbf{C}_j$  with zero resembling the border itself. Hence, the distance of each pixel to the contour is directly mapped to the image which allows to guide the 3D-SSM towards the bone boundary efficiently. For each projection image  $\mathbf{B}_j$  all  $N_x$  silhouette points in  $\mathbf{x}_j^*$  are translated to pixel positions and sampled on the distance map. The resulting sampled distances can be used as  $\mathbf{y}_j^*$  in the cost function  $\epsilon_{\text{dist}}$  (cf. Eq. 32).

### Multi-resolution pose optimization

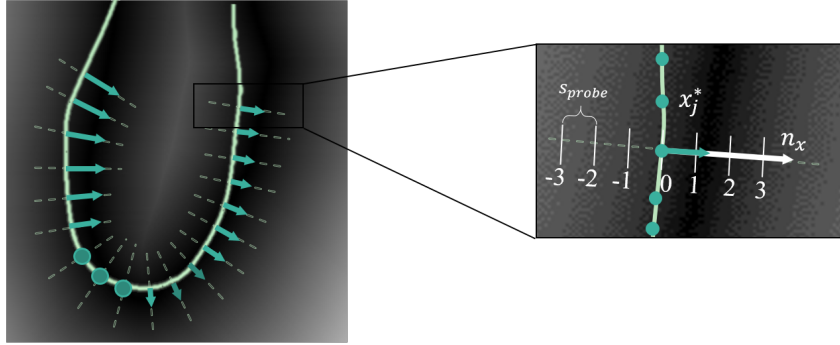


Figure 32: Sampling of profile points: Points are sampled along the contour normals with the green arrow indicating the displacement to the minimal distance of the distance map.

A multi-resolution approach akin to the segmentation phase described in chapter 4.2.1 is implemented and applied to each bone. Inspired by the profile sampling procedure, not only the actual contour points  $\mathbf{x}_j^* = \{\mathbf{x}_{j,k}^* \mid k \in 1, \dots, N_x\}$  are sampled but also  $N_{\text{profile}}$  points along the contour point normal  $\mathbf{n}_{\mathbf{x}_{j,k}^*}$  in positive and negative direction with a probe spacing  $s_{\text{probe}}$  (cf. Fig. 32). By also comparing the neighborhood, optimal local adjustments can be calculated in each iteration. Along the contour point normal  $\mathbf{n}_{\mathbf{x}_{j,k}^*}$ , the distance map is re-sampled and the minimal distance is estimated for each profile. This way, local optimal corresponding points  $\mathbf{y}_j^* = \{\mathbf{y}_{j,k}^* \mid k \in 1, \dots, N_x\}$  can be computed by:

$$\mathbf{y}_{j,k}^* = \min \left( f_{\text{dist}}(\mathbf{x}_{j,k}^* + i \cdot s_{\text{probe}} \cdot \mathbf{n}_{\mathbf{x}_{j,k}^*}) \mid i = -N_{\text{profile}}, \dots, N_{\text{profile}} \right) \quad (35)$$

$$\forall j \in \{1, \dots, N_B\} \quad \forall k \in \{1, \dots, N_x\}$$

Since  $\mathbf{x}_j^*$  only comprises points perpendicular to the 3D-SSM boundary, the sampled points  $\mathbf{y}_j^*$  along the contour normal correspond to 3D points along the surface normal and can be back-projected easily. By changing  $s_{\text{probe}}$ , the distance between sampled positions can be varied to realize a multi-resolution approach. The optimization can be

started with a high value of  $s_{\text{probe}}$  to reach a fast convergence and can be decreased towards a smaller value to ensure an accurate adaption during fine adjustment.

With this approach, optimal local landmark adjustments can be computed based on the correspondence pairs. Subsequently, a GPA is applied exclusively on the feature sets  $\mathbf{x}_j^*$  and back-projected adjustments  $\mathbf{y}_j^*$  for all  $B_j$ . This way, an iterative update of transformation  $\mathbf{T}_{\text{uninjured}}$  can be obtained.

### Multi-object procrustes pose optimization

To increase the robustness and accuracy of the proposed approach and to cope with inaccuracies of the preceding initialization step, the multi-object GPA described in 4.2.1 is integrated into the reconstruction step. In coarse reconstruction steps, where larger profile distances are sampled and analyzed, a composite transformation  $\mathbf{T}_{\text{composite}}$  helps to ensure that the 3D-SSMs migrate to a common location. The multi-object GPA uses the correspondence pairs of all bones simultaneously and establishes a common transformation update, which is applied to each bone. This is useful, particularly to prevent the quasi-rotation symmetric fibula and tibia 3D-SSM from an early ill-posed rotation along the corresponding shaft axis. Since the pose of the single bones varies for each patient, the multi-object GPA can only be applied during an early, coarser stage. It is necessary to decouple the pose adaption in finer stages and optimize each bone separately.

### Landmark-specific adaption

One drawback of the 3D-SSM trained on cut bone shafts is that the varying length is not fully covered in the model representation.  $\mathbf{X}_{\text{uninjured}}$  is initialized with  $\mathbf{b}_{\text{injured}}$  and the assigned shaft length may be different from the visible shaft length in the fluoroscopic projection images  $\mathbf{B}$ . During reconstruction, landmark subset  $\mathbf{x}_{\text{shaft}} \subset \mathbf{x} \subset \mathbf{X}_{\text{uninjured}}$  close to the upper bound might let  $\mathbf{X}_{\text{uninjured}}$  drift towards the proximal end of the bone instead of focusing on the more relevant distal part. Therefore, it can be helpful to reduce the influence of the shaft related landmarks. For that purpose, the upper 30% of the shaft landmarks are identified and tagged in the mean representation of the 3D-SSM by measuring the distance from each landmark to the tibial plafond, in other words, the most distal point of the tibia shaft. For that purpose, the axial standard plane (definition given in Tab. 1) can be used as a reference plane. All landmarks of the subset  $\mathbf{x}_{\text{shaft}}$  are ignored during GPA.

### Shape optimization

Since the reference 3D-SSM with  $\hat{\mathbf{b}}_{\text{injured}}$  is assumed to already closely resemble the desired uninjured ankle, shape parameters  $\mathbf{b}_{\text{uninjured}}$  and  $\hat{\mathbf{b}}_{\text{injured}}$  should only deviate slightly. Thus, in a coarse resolution stage, only transformation  $\mathbf{T}_{\text{uninjured}}$  is optimized to avoid local minima resulting from early shape adaption. Shape parameters  $\mathbf{b}_{\text{uninjured}}$  are globally altered in a second finer resolution stage. The local adjustments, denoted as  $\delta \mathbf{x}^*$  are used to project the shape model onto shape space. In shape space, an update of the shape parameters  $\delta b$  can be calculated by rearranging the SSM definition taken from Eq. 14.

$$\delta b = \mathbf{P}^T \tilde{\mathbf{T}}^{-1} (\delta \mathbf{x}^*) \quad (36)$$

where  $\tilde{\mathbf{T}}$  denotes the transformation  $\mathbf{T}$  without the translational part and  $\mathbf{P}$  denote eigenvectors of the SSM. Since the shape deviation between  $\mathbf{b}_{\text{uninjured}}$  and  $\mathbf{b}_{\text{injured}}$  should not differ significantly, the shape space is constrained to a range of  $\mathbf{b} \pm 2 \cdot \sigma_b$ .

## 4.5 Visualization and clinical workflow integration

The focus of this work is the realization of an assistance system that enables an intra-operative contralateral side comparison for ankle surgery based on C-arm imaging. A description of the proposed workflow can be found in chapter 4.1 and the individual components are introduced in chapter 4.2, 4.3 and 4.4. This chapter describes the integration and visualization of the computed 3D surface models into the workflow. Section 4.5.1 depicts the alignment of the uninjured and injured surface models. The chapter also gives insights on the automatic adjustment of the viewing planes (section 4.5.2) and methods to provide a sound visualization of the results (section 4.5.3). 2D-3D reconstruction leads to the creation of a 3D model representation of the uninjured ankle. In order to assist the surgeon during the fibular reduction process, the surface models must be visualized together with the injured ankle. To bridge the gap between the computation and visualization and to enable a direct comparison, the computed surface models need to be transferred into the coordinate system of the 3D image of the injured ankle. Furthermore, the developed system must allow the surgeon to display specific areas of the ankle decisive for the evaluation.

### 4.5.1 3D-3D model-based registration

After obtaining two 3D-SSMs,  $X_{\text{bone,injured}}$  and  $X_{\text{bone,uninjured}}$ , a contralateral side comparison can be performed in 3D space. To enable the comparison, the two SSMs need to be aligned towards each other, which requires two steps: Reflection and 3D-3D registration. After reconstruction,  $X_{\text{bone,uninjured}}$  does not share the same laterality (left or right) with  $X_{\text{bone,injured}}$  and needs to be back-reflected first. Afterward, a rigid registration can be performed. One advantage of the 3D-SSMs is their ability to capture meaningful correspondence points. This allows an instant assignment of correspondence pairs in both 3D-SSMs to simplify the registration. The contralateral side comparison is targeted at revealing the pose deviation of the fibula between the two ankles. Hence, a composite transformation of all 3D-SSMs is established to retain the relative positions of the individual bones. The composite transformation can be computed by a landmark-based least-square optimization of  $X_{\text{tibia,injured}}$  and  $X_{\text{tibia,uninjured}}$ . This way only the tibia bones are optimally aligned and the relative spatial arrangement of all other bones is preserved.

### 4.5.2 Automatic standard plane adjustment

Intra-operative imaging allows the surgeon to visually verify the surgical outcome during all phases of an intervention. The surgeon is usually only interested in a specific

anatomical ROI and a profound visualization of those relevant structures is most essential for a comprehensive evaluation. One possibility to achieve a suitable visualization is the anatomy-specific adjustment of the MPR viewing planes of the 3D C-arm image. A detailed description of the so-called standard planes for the ankle joint is given in section 2.1.4. The adjustment of the standard planes must be repeated for each scan and increases the intervention time. The work of BREHLER et al. 2015 allows to calculate the standard planes automatically for calcaneus fractures based on a pre-defined atlas image. This approach relies on two assumptions: The spatial inter-relations between the respective bones do not significantly vary and standard planes are defined in relation to one specific bone. Furthermore, the approach is limited to rigid registrations between features derived from the 3D image and the atlas. Transferring the approach to the ankle joint is not readily applicable since the positional and rotational relation between fibula and tibia varies, especially in the event of fibular fractures, and the definition of the standard planes depends on both bones. Therefore, an alternative approach is proposed, in which the standard planes are derived from the 3D-SSM for each bone separately.

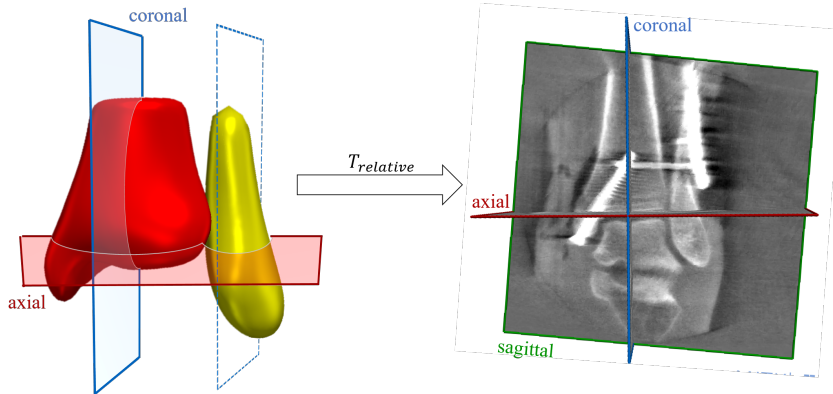


Figure 33: Automatic standard plane adjustment: Standard planes are annotated in the mean SSM and transferred to the image after SSM segmentation.

In the mean 3D-SSM representation  $\bar{\mathbf{X}}$ , standard planes are adjusted to resemble the ideal plane adjustment (cf. Fig. 33). The affine transformation parameters of all 3D-SSMs are extracted twice, before 3D segmentation  $\mathbf{T}_{0,\text{bone}}$  of the ankle and after segmentation  $\hat{\mathbf{T}}_{\text{final},\text{bone}}$ . Afterward, the relative transformation can be established.

$$\mathbf{T}_{\text{relative},\text{bone}} = \mathbf{T}_{0,\text{bone}} \cdot \hat{\mathbf{T}}_{\text{final},\text{bone}}^{-1} \quad (37)$$

Using the relative transformation, the standard planes can be transformed independently. The axial standard plane is transformed using the resulting transformation  $\mathbf{T}_{\text{relative},\text{tibia}}$  of

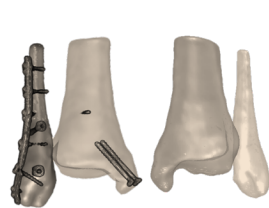
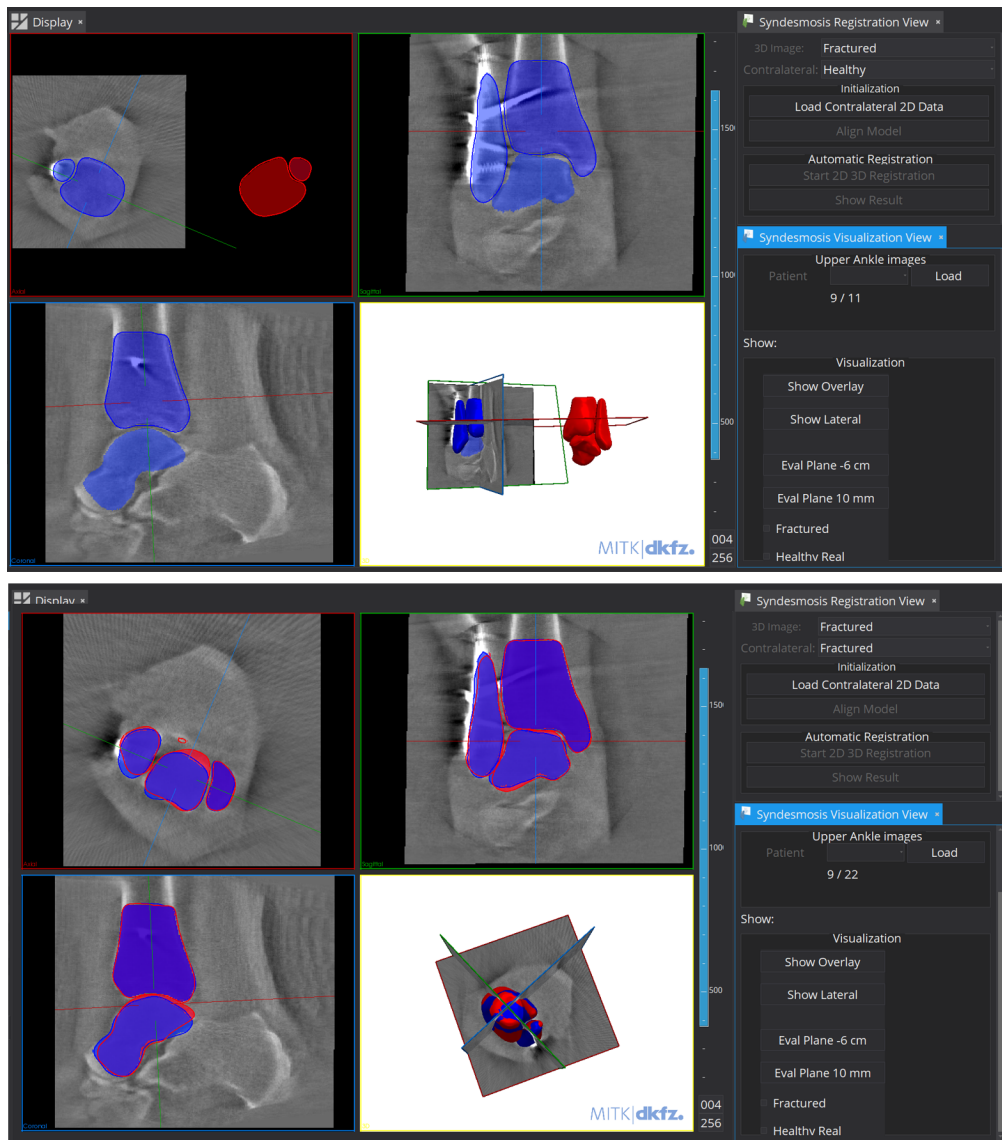


the tibia and the sagittal standard plane is transformed by the resulting transformation  $\mathbf{T}_{\text{relative, fibula}}$  of the fibula. The coronal standard plane is orthogonal to the axial and sagittal standard plane and thus can be derived automatically.

As a result, the standard planes of the 3D image can be calculated considering not only the pose of the tibia but also the relative pose between tibia and fibula.

### 4.5.3 Visualization

An exact adjustment of the standard planes is necessary for an adequate visualization and evaluation of the reduction result. However, the axial plane located at the level of the tibial plafond does not reveal any clinical relevant information. It provides the correct orientation of the planes but is not set to the correct image slice. Hence, the manual clinical workflow requires the surgeon not only to adjust the standard planes. Instead, the surgeon uses the adjusted axial standard plane to scroll along the image stack and to approach the two relevant evaluation planes. This process can be automatized by computing the standard planes and then translating the axial plane along its normal. Either the plane is translated  $\delta s = 10\text{mm}$  (superior) for a clear view on the incisura or  $\delta s = -6\text{mm}$  (inferior) for evaluating the internal or external rotation of the fibula at the talar dome level. The values are recommended for clinical evaluation as already described in detail in section 2.1.2. The software allows to alternate between both views. Furthermore, the application provides an overlaid and a side-by-side view. In the overlaid view, both surfaces are displayed together with the 3D input image. The side-by-side view shows the reconstructed surface reflected along a utility plane, which is built by intersecting the image border of the 3D image with the axial plane. Fig. 34 shows the two visualization modes implemented as a plugin for the MITK workbench (NOLDEN et al. 2013).



Side-by-side view



Overlay view

Figure 34: Visualization of the reconstruction results: The prototype allows to visualize the computed surfaces in a side-by-side view (top, 10 mm above the tibial plafond) and as overlays (bottom, 6 mm below the tibia plafond). The blue surfaces resemble the injured ankle whereas the red surfaces resemble the reconstructed contralateral uninjured ankle.

The proposed workflow was carefully validated for a potential successful integration into the clinical environment. The quality and performance of the prototype were investigated on different levels following the error hierarchy introduced in section 4.1.3. The individual components, such as e.g. the 3D-SSM segmentation, were tested with different setups to examine the accuracy and robustness of each component. Once the individual components were evaluated, the overall pipeline was tested with clinical data of both ankles to assess the overall uncertainty. The following sections provide the reader with details of data material used in the conducted experiments (section 5.1) as well as an overview of all different experiments and their motivation (section 5.2). For a meaningful and interpretable evaluation of the results, different image-based and point-based metrics are introduced in section 5.3. Afterward, a detailed description is given on the different experiments assessing the quality of 3D bone segmentation (section 5.4), 2D fluoroscopic segmentation (section 5.5), 2D-3D reconstruction (section 5.6) and the overall system (section 5.7). Results are presented directly after the respective experiment description.

## 5.1 Data acquisition

All data sets used in this project were acquired in the BG Trauma Center Ludwigshafen between 2011 and 2018 and anonymized for the retrospective evaluation. The device used in this project was an Arcadis<sup>®</sup> Orbic 3D (Siemens Healthcare GmbH), a mobile 3D C-arm system with an image intensifier, an isocentric design and 190° orbital movement. During intra-operative acquisition, the operator uses the interface of the device to select the respective anatomical region, in case of bilateral osseous regions the laterality, the

pose of the C-arm with respect to the OR table as well as whether the scan is taken with reduced (50) or full amount (100) of images. The images were acquired with a pulse rate of 66 at a kV range of 59-65.

Each data set consisted of one 3D image, 100 fluoroscopic 2D images and their respective projection matrices. The volume was reconstructed from all 100 fluoroscopic images generated by an orbital rotation of  $190^\circ$  around the device's iso-center. All volumes had an isotropic voxel spacing ranging from 0.475 to 0.5 mm and a size of  $256^3$  voxels. The fluoroscopic images had  $1024^2$  pixels and covered a detector FOV of 23 cm. In most of the images, the tibiofibular joint is located in the image center and always contains tibia, fibula and talus. In all cases, tibia and fibula were only partially visible limited by the FOV. Some images showed only a proportion of the talus in case the ankle was not placed in the iso-center.

Hereinafter, an overview of the different data sets is given. The data sets can be categorized in three groups:  $\mathbf{S}_{\text{clinical,injured}}$  including clinical images of the injured ankle,  $\mathbf{S}_{\text{clinical,pairs}}$  comprises clinical images of both ankles and  $\mathbf{S}_{\text{cadaver,uninjured}}$  includes images of the uninjured ankle acquired within the context of a cadaver study. Hereinafter, a detailed description is provided.

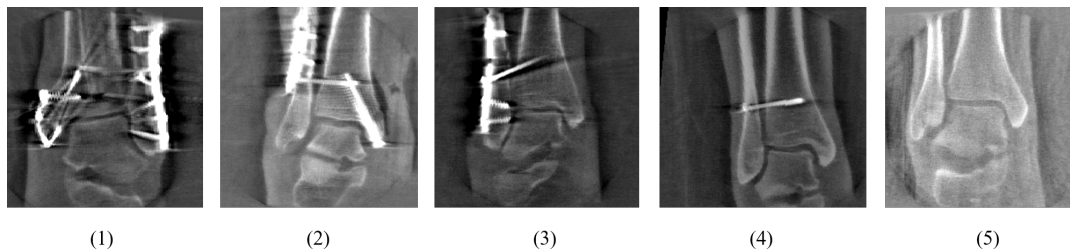


Figure 35: C-arm volumes of injured ankle: Weber B and C fractures with an one-third tubular plate (1-3) plus the medial malleolus stabilized with a band wire (1) or with a lag screw (2), Maisonneuve fracture with syndesmotic screw (4) and uninjured ankle without metal implants (5).

1.  $\mathbf{S}_{\text{clinical,injured}}$  comprised 42 ankles which are clinical cases after ORIF and with metal implants scanned during intervention for the purpose of visual evaluation of the reduction result. The acquired 3D images were acquired right after reduction and relocation of the bone fragments so that no isolated fragments were present. The data set collection contained different fracture types and treatments ranging from the insertion of a plate and screws, additional screws for fixating small bone fragments to the mere insertion of a syndesmotic screw. Weber B, Weber C and Maisonneuve fractures were included. Fig. 35 shows representative examples of the different treatment types.

2.  $S_{\text{cadaver, uninjured}}$  consisted of 30 ankles without isolated bone fragments or any metal implants. The subjects in this study were received from the Institute for Anatomy and Cell Biology, University of Heidelberg, Germany. A written informed consent was provided by the deceased to allow the use of their bodies for research purpose. The cadaver study was carried out at the BG Trauma Center Ludwigshafen, Germany. It got approved by the ethical review committee in Heidelberg, Germany with the registration number S-013/2014. For the conducted experiments in this thesis, 30 ankles were selected. Besides 12 bilateral ankle pairs, the cadaver study was conducted with unilateral ankles.
3.  $S_{\text{clinical, pairs}}$  comprises a collection of 11 ankle pairs acquired during intervention for the purpose of manual contralateral side comparison. One pair includes one ankle with metal implants after reduction and one corresponding uninjured ankle without metal implants.

**3D ground truth segmentations:** Manual segmentations of all 3D images were obtained by a medical expert. All segmentations were validated by an orthopedic surgeon. The MITK workbench (NOLDEN et al. 2013) was used to segment the images by manually contouring the bone boundary in a slice-by-slice manner and converting the binary segmentations to surface representations. MITK was further used to annotate an evaluation plane for each volume. The axial plane was defined as orthogonal to the shaft of the tibia with the origin at the tibial plafond. This plane is later used to cut the shafts to avoid uncertainties resulting from varying shaft lengths.

**3D standard planes:** A manual adjustment of standard planes was performed by a surgeon using the MITK workbench. Axial, sagittal and coronal planes were adjusted following the definition given in Tab. 1. These adjustments were used for defining the reference plane for cutting the tibia and fibula shaft as well as served as a reference when evaluating the automatic standard plane adjustment.

**2D ground truth segmentations:** Manual segmentations are only available for 3D images since the process of 2D fluoroscopic image segmentation is tremendously time-consuming and at the same time highly challenging due to unreliable, noisy contours at overlapping bones. To create 2D ground truth segmentations, it was taken advantage of the relation between the 3D image and the respective 2D images. The 3D ground truth segmentations and the respective projection matrices were used to project all binary voxels onto each of the 2D image planes. This way, 2D segmentation masks could be generated for all single fluoroscopic images. It must be noted, however, that those

segmentations did not cover the entire visible shaft of tibia and fibula since manual 3D segmentations were likewise constrained by the FOV. Furthermore, the resulting 2D segmentations merely approximated manual 2D segmentations. Uncertainties in the computation of projection matrices, which are generated during device calibration, can translate to segmentation uncertainties.

**Selection of standard projections:** For all experiments, three projection images were selected by a trauma surgeon from all 100 projection images for each patient. The selection approximated the radiological standard projections of the ankle joint: Lateral, mortise and anterior-posterior view (FRANKE et al. 2012) as depicted in Fig. 36. For  $S_{\text{clinical,pairs}}$  the selection of the real standard projections was not always possible due to the superposition of the contralateral ankle. In those cases, the nearest projection image with one visible ankle was selected instead.

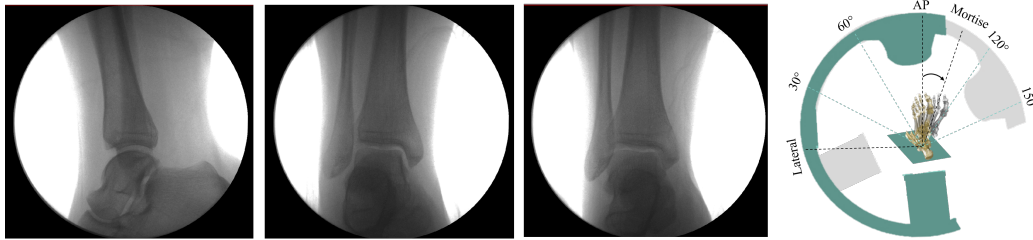


Figure 36: Standard projections of the ankle joint: lateral view (L), mortise view (M) and anterior-posterior view (AP).

## 5.2 Overview and purpose of the experiments

The proposed workflow consists of a variety of individual components and all induce a certain uncertainty to the system. A variety of different experiments were conducted to assess certain aspects and error sources. This section provides a brief description of the experiments and their incorporation in the overall context. Fig. 37 represents an overview of all components, the conducted experiments and indicates the dependencies on each other.

A systematic evaluation was performed in two stages: First, the uncertainties of all components, which are solely dependent on the image data were assessed under different aspects:

- 3D-SSM segmentation: The method was evaluated under different setups comparing the pure intensity based SSM method, the CNN segmentation and the CNN-complemented SSM segmentation. The segmentation quality is influenced by the

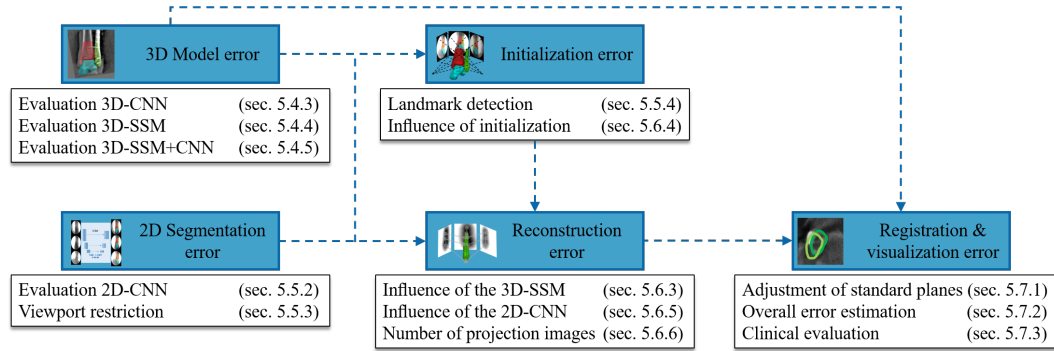


Figure 37: Overview of conducted experiments: Experiments were designed to cover error sources resulting from 2D segmentation, 3D segmentation and 2D-3D initialization as well as the overall reconstruction error. The arrow indicate that the respective component serves as input for another component.

performance of the trained 3D-SSM and 3D-CNN as well as the input 3D image. The SSM as such is also dependent on the initial placement. Therefore, the robustness and accuracy of the different methods were investigated.

- 2D-CNN segmentation: The accuracy of the method was assessed by training on all data sets as opposed to training on a limited viewport since the CNN is influenced by the training and the input data.
- 2D landmark extraction: The quality of the landmarks was computed on the CNN predictions and compared to landmarks extracted from ground truth segmentations.

In the second stage of the evaluation, components were evaluated that are affected by preceded pipeline components. The results of all former methods function as inputs for the 2D-3D reconstruction so that their uncertainties get accumulated. The influence of each uncertainty was estimated by replacing the output of the component by the perfect result, e.g. using the ground truth instead of the computed 2D-CNN segmentation.

- 2D-3D reconstruction: The reconstruction was compared in different stages by systematically replacing inputs by their respective ground truth to eliminate different error sources.
- Overall evaluation: The accuracy of the overall pipeline was assessed before and after 3D-3D registration to investigate the performance on real clinical data. The overall evaluation is not only affected by the 2D-3D reconstruction but also by the 3D-3D registration of  $\mathbf{X}_{\text{tibia,injured}}$  and  $\mathbf{X}_{\text{tibia,uninjured}}$ .

- Further experiments: The quality of the standard plane adjustment was evaluated on the clinical pairs.

### 5.3 Evaluation metrics

The conducted experiments resulted either in 2D or 3D segmentation results or respective surfaces. Evaluation of the results can either be performed by mathematically quantifying the discrepancy between the result and the reference. Alternatively, the results can be examined in a qualitative manner by visually inspecting challenging regions and main problems of the different approaches relevant to the overall pipeline.

Segmentation results can be compared to the ground truth segmentation of each bone using different metrics for shape comparison. Evaluation can be either based on the outer surface of the SSM or the segmentation. For the former, the final landmark positions of the SSM can be compared to a surface generated from the ground truth segmentation. In case of 2D segmentations, not the surfaces but the contours are used and the metrics are based on a pixel-wise comparison.

A widely used metric is the *Average symmetric surface distance* (ASD) that measures the reciprocal euclidean distance between two surfaces or contours in mm (HEIMANN et al. 2009).

$$\text{ASD}(X_{\text{Result}}, X_{\text{GT}}) = \frac{1}{|X_{\text{Result}}| + |X_{\text{GT}}|} \int_{x \in X_{\text{GT}}} \delta(x, X_{\text{Result}}) dx + \int_{x \in X_{\text{Result}}} \delta(x, X_{\text{GT}}) dx \quad (38)$$

where  $|X_{\text{Result}}|$  and  $|X_{\text{GT}}|$  denote the area of the surfaces.

A popular volume-based metric to measure the segmentation quality is the Sørensen-Dice coefficient by JACCARD 1912. The overlap of the segmentation result  $X_{\text{Result}}$  and the ground truth  $X_{\text{GT}}$  segmentation can be measured by:

$$\text{Dice}(X_{\text{Result}}, X_{\text{GT}}) = \frac{2 \cdot |X_{\text{Result}} \cap X_{\text{GT}}|}{|X_{\text{Result}}| + |X_{\text{GT}}|} \quad (39)$$

where  $|X_{\text{Result}}|$  and  $|X_{\text{GT}}|$  are the measured volumes of the binary segmentations.

Another commonly applied metric is the Jaccard index also referred to as *Intersection Over Unit* (IOU) introduced by JACCARD 1912.

$$\text{Jaccard}(X_{\text{Result}}, X_{\text{GT}}) = \frac{|X_{\text{Result}} \cap X_{\text{GT}}|}{|X_{\text{Result}}| + |X_{\text{GT}}| - |X_{\text{Result}} \cap X_{\text{GT}}|} \quad (40)$$

Dice and Jaccard have in common that their value is influenced by the overlapping region



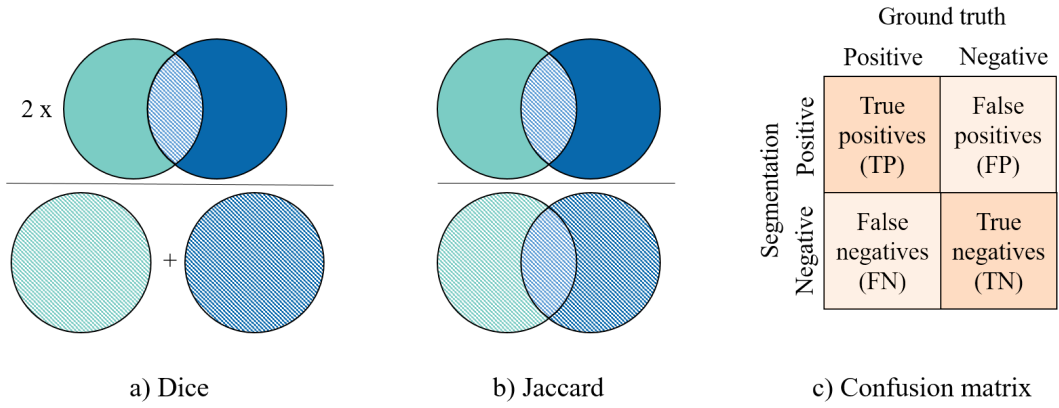


Figure 38: Voxel-/pixel-based overlap metrics for shape comparison: a) Dice score and b) the Jaccard are illustrated by two circular sets, the result (green) and the ground truth (blue). c) shows the confusion matrix for the metrics sensitivity, specificity, precision and accuracy.

and approaches 1 in case of perfect overlap. An illustration of both metrics is provided in Fig. 38 a) and b).

In machine learning applications, other metrics are commonly used to assess the performance based on the assumption that the segmentation is a pixel-/voxel-wise classification task. Fig. 38 c) displays a confusion matrix introduced by FAWCETT 2006 that distinguishes between positive and negative labels being classified as either positive or negative. The following metrics are all based on the decisions drawn from the confusion matrix:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \quad \text{Specificity} = \frac{TN}{TN + FP} \quad (41)$$

$$\text{Precision} = \frac{TP}{TP + FP} \quad \text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN}$$

where TP denotes true positives, FP false positives, TP true negatives and FN false negatives.

The voxel-wise comparison is affected by the divergent shaft lengths of computed and manual segmentation. As pre-condition for a reliable evaluation independent from the shaft lengths, all segmentations are cropped w.r.t. the shortest shaft first. The cut is performed using the axial standard plane perpendicular to the shaft which is translated

towards the shortest of the shafts  $l_{\text{shaft1}}$  and  $l_{\text{shaft2}}$ .

$$\mathbf{p}_{\text{cut}} = \mathbf{p}_{\text{axial}} + \mathbf{n}_{\text{axial}} \cdot \min \{l_{\text{shaft1}}, l_{\text{shaft2}}\} \quad \mathbf{n}_{\text{cut}} = \mathbf{n}_{\text{axial}} \quad (42)$$

There are areas of the ankle, like e.g. the incisura, that are more important than others for the clinical application. The previous mentioned metrics are global measures that do not emphasize or distinguish specific areas. Besides 3D measurements like ASD, Jaccard and Dice score, slice-wise 2D measurements on specific relevant slices are introduced to quantify the quality of the result. The two 2D planes are depicted in Fig. 39. The Hausdorff distance is a common quantitative measure in shape comparison to estimate the distance between two point sets A and B (HUTTENLOCHER et al. 1993).

$$d_H(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} d(a, b), \sup_{b \in B} \inf_{a \in A} d(a, b) \right\} \quad (43)$$

The function  $d_H(A, B)$  is referred to as the symmetric Hausdorff distance of A and B. It evaluates the distance of each point A to the nearest point of B as well as the distance of each point B to the nearest point of A. The largest of these distances determines the value of  $d_H(A, B)$ .

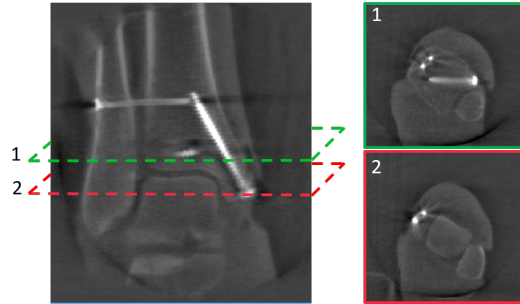


Figure 39: 2D slices relevant to the clinical evaluation: Position 1 shows the fibula in the incisura approximately 10 mm above the tibia plafond. Position 2 shows the tibia and fibula approximately 6 mm below the tibia plafond at the talar dome level.

Experiments in section 5.5.4 require a quantitative evaluation of landmarks. The landmarks were computed for ground truth and predicted segmentations. The 2D landmark positions can be compared by computing the pixel distance between both computed landmarks  $p_1$  and  $p_2$  and the rotational deviation can be estimated by the angle between normal landmark vectors  $v_1$  and  $v_2$ .

$$\delta(p_1, p_2) = \|p_1 - p_2\|_2 \quad \angle(v_1, v_2) = \arccos(v_1 \cdot v_2) \quad (44)$$

## 5.4 3D bone segmentation

All methods described in chapter 4.2 are tailored to perform an accurate bone segmentation of C-arm images under the presence of metal implants. The 3D segmentation is influenced by the trained 3D-SSM, its initial placement in the image and the input image itself. The described task requires the 3D-SSM to cope with arbitrary initialization since the ankle position and orientation varies for each patient and 3D-SSM is known to be sensitive towards initialization. Thus, it was essential to investigate not only the accuracy but also the robustness of the methods if the 3D-SSM is not placed close to the final position. Segmentation performance of all three approaches, namely, the pure intensity-based 3D-SSM segmentation, the CNN-based segmentation and the 3D-SSM segmentation incorporating the CNN complement were compared.

### 5.4.1 Training data and 3D-SSM creation

Experiments were conducted with a six-fold cross-validation on the data sets of  $S_{\text{clinical, injured}}$ . The data set collection was split into training and test subsets with 35 and 7 data sets, respectively. All images in the training set were reflected to correspond to the same laterality as the respective test set. A 3D-SSM was created using the methods described in section 4.2.1. The tibia and talus 3D-SSM consist of 2000 landmarks and the fibula 3D-SSM of 1500 landmarks. Prior to training, the segmentations were pre-aligned in a pre-processing step using the center of mass of all segmentations as landmarks and tibia and fibula shafts were cut to roughly match the same bone length.

The chosen parameters for the 3D-SSM segmentation were inspired by the selection of GÖRRES et al. 2016 for the calcaneus and experimentally modified for the ankle joint. The probe spacing was set to  $s = 0.5$ , which equals the largest possible voxel size of the input images. The number of sampled probes for each landmark profile was set to 3 probes per side and not changed during the procedure. The segmentation procedure was executed in 5 iterations with varying parameters (cf. Tab. 3). In the first iteration, only the first two shape parameters were altered whereas in the consecutive iterations the number of shape parameters was increased. In addition, the resolution was increased during iterations and implied that e.g. resolution 2 refers to the image being downsampled with factor 2 and the probe spacing was multiplied by 2. The maximum point movement of landmarks between two consecutive steps was chosen as convergence criterion and also altered among iterations. The SSM segmentation was stopped when the maximum point movement of all meshes was less than the predefined threshold. The multi-object GPA was only applied in the first two iterations before the SSMs were optimized independently.

Iteration	1	2	3	4	5
Num. of shape parameters	2	2	20	20	20
Resolution	2	2	1	0	0
Max. move threshold	0.8	0.3	0.3	0.5	0.2
Multi-object GPA	yes	yes	no	no	no

Table 3: SSM parameter selection

### 5.4.2 Evaluation method

Prior to the 3D-SSM segmentation, the 3D-SSM had to be placed at a certain position in the 3D image. The ideal position could be derived from the ground truth segmentation by initializing the 3D-SSM with  $\mathbf{T}_{\text{train}}$ . For a simultaneous evaluation of accuracy and robustness, the position and orientation of the SSM were varied following the systematic approach of GÖRRES et al. 2016. A rigid transformation  $\mathbf{T}_{\text{offset}}$  was applied to the 3D-SSM to create a translational or rotational displacement as illustrated in Fig. 40.

$$\mathbf{T}_{\text{start}} = \mathbf{T}_{\text{train}} \cdot \mathbf{T}_{\text{offset}} \quad (45)$$

The translational offset was increased by 2 mm intervals to a maximum of 12 mm in six perpendicular directions. The rotational angle was increased by  $5^\circ$  intervals to a maximum of  $25^\circ$  for all rotation axes. The shape parameters were set to  $\mathbf{b} = 0$  to resemble the mean model representation  $\bar{\mathbf{X}}_{\text{bone}}$ . Afterward, the actual 3D-SSM segmentation was executed at the current location and the ASD was measured. The evaluation scheme was performed on the pure intensity based and the CNN complemented SSM method. On the contrary, the evaluation of the CNN network was conducted by directly comparing the output segmentation to the ground truth segmentations since the network is fed by the entire volume independent from pre-initialization.

### 5.4.3 Evaluation 3D-CNN segmentation

The aim of the experiment was to evaluate whether an accurate segmentation of tibia, fibula and talus can be created using the U-Net approach described in section 4.2.2. A summary of network hyperparameters can be taken from Tab. 4. In a pre-processing step, the volumes were downsampled to  $180^3$  voxels and cropped to an input size of 128. Data augmentation strategies included cropping, rotation, mirroring, deformation and scaling. A detailed description of the different strategies can be found in section 4.2.2.

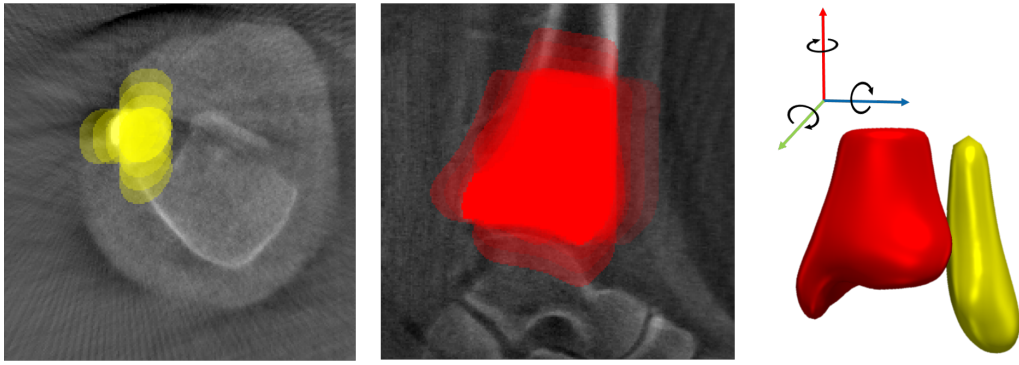


Figure 40: SSM initialization scheme: The 3D-SSM is placed at different positions in the image by applying a translation or rotation.

For augmentation, a publicly available in-house framework<sup>1</sup> was used. The U-Net was trained for approx. 300 epochs for each fold with a learning rate of  $10^{-3}$ , a learning decay of  $10^{-4}$  and a batch size of 8.

Data augmentation was not only applied to enlarge the training data during training but also during test time. This approach referred to as test time augmentation, performs modifications to a test image so that the network is shown the image several times. Since the modifications can be reversed, prediction results can be aggregated and averaged with to increase the accuracy. In the conducted experiments, each image was reflected along all axes similar to the conventional training data augmentation.

Parameter	Value
Downsampled image size	180
Cropped image size	128
Batch size	8
Optimizer	ADAM
Loss	3D cross entropy
Activation function	ELU
Normalization	Instance
Learning rate	$10^{-3}$
Learning decay	$10^{-4}$
Epochs	300

Table 4: Network hyperparameter selection for the 3D U-Net

Since the images were center cropped before used as input for the network, the evaluation was performed on the full image as well as a cropped ROI. The cropping was

<sup>1</sup>Batchgenerators <https://github.com/MIC-DKFZ/batchgenerators>

performed solely along the z-axis by 70 pixels in each direction to account for the influence of the center crop in the training as well as the influence of the varying shaft lengths and the partial visible talus.

#### 5.4.4 Evaluation of the 3D-SSM segmentation

With this experiment, the approach by HEIMANN 2009b described in section 4.2.1 was evaluated using the parameters in Tab. 3 and served as a baseline for the subsequent experiments. The main question to answer was whether a pure intensity-based 3D-SSM can create an accurate segmentation robust to metal-related artifacts and independent from an initial model placement.

#### 5.4.5 Evaluation of the new combined 3D-SSM segmentation approach

In the last experiment, the proposed CNN complement was combined with the 3D-SSM segmentation to examine whether the resulting SSM representation achieves a similar segmentation accuracy compared to the CNN-based segmentation. After initial placement and misalignment of the SSM, the landmark-based pre-initialization, described in 4.2.3, was performed using the CNN-based predictions. This pre-processing step should ideally assure the segmentation result to be independent to initialization. Network parameters remained constant and can be taken from the previous experiment.

#### 5.4.6 Results

	Tibia		Fibula		Talus	
	Full	ROI	Full	ROI	Full	ROI
Dice	0.939± 0.025	0.959± 0.021	0.883± 0.048	0.905± 0.044	0.910± 0.028	0.958± 0.019
Jaccard	0.886± 0.432	0.922± 0.036	0.794± 0.072	0.830± 0.067	0.836± 0.046	0.919± 0.033
Precision	0.938± 0.428	0.956± 0.034	0.889± 0.069	0.916± 0.054	0.940± 0.043	0.958± 0.024
Sensitivity	0.942± 0.038	0.963± 0.023	0.887± 0.080	0.902± 0.077	0.885± 0.052	0.959± 0.030
Specificity	0.999± 0.001	0.992± 0.007	0.999± 0.000	0.993± 0.005	0.999± 0.000	0.995± 0.002
Accuracy	0.997± 0.001	0.987± 0.007	0.999± 0.000	0.985± 0.006	0.997± 0.001	0.992± 0.004

Table 5: Comparison of different metrics for different bones and image segments

**3D-CNN segmentation:** The results of the 3D-CNN segmentation method on the six folds of  $S_{\text{clinical, injured}}$  are displayed in Fig. 42. The boxplots in light grey show the evaluation of the entire 3D image whereas the dark blue boxplots show the evaluation performed on a cropped ROI. The cropped region yielded better results since it was not

affected by segmentation errors at image borders. In Tab. 5 the mean and standard deviation of all metrics are displayed. The best performance was achieved by the CNN applied to the tibia. Visual inspection of the segmentation results for different cases (cf. Fig. 43) confirms that the errors were mainly attributed to inaccuracies along the fibula shaft. Fig. 43c depicts the example with the worst performance caused by the high number of metal implants present in the image.

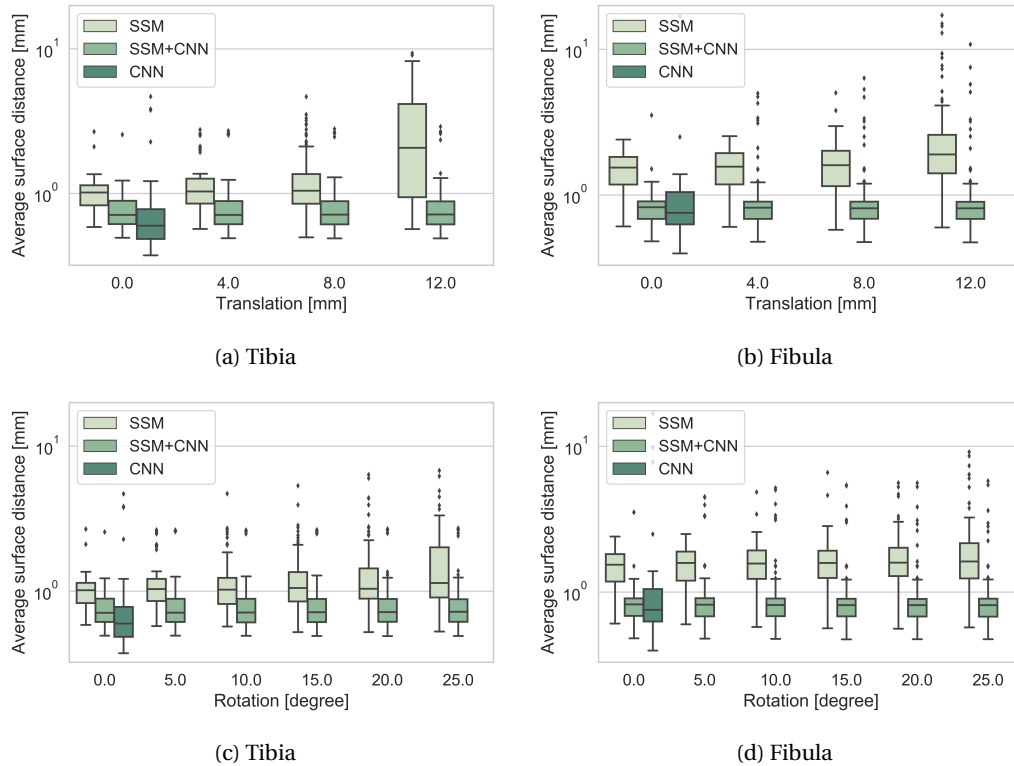


Figure 41: 3D segmentation error evaluation: Segmentation error (log. scale) with translational and rotational offset without (light green) and with using CNNs (middle green). The segmentation error of the 3D-CNN (dark green) is shown for comparison. Mean surface distance (ASD) is measured in mm.

**Comparison of all three segmentation approaches:** Fig. 41 compares the performance of the pure intensity based 3D-SSM with the CNN complemented 3D-SSM segmentation for different starting positions. Fig. 41a) and Fig. 41b) show the results for the translational offset whereas Fig. 41c) and Fig. 41d) show the rotational offset. The CNN (indicated in dark green) yielded the best results. The 3D-SSM segmentation (light green) provided the required corresponding landmarks but, on the other hand, resulted in a higher ASD that increased with a larger displacement. Combining the 3D-SSM and the CNN complement achieved nearly as good results as the CNN segmentation and at the same time was independent of any rotational or translational displacement while pro-

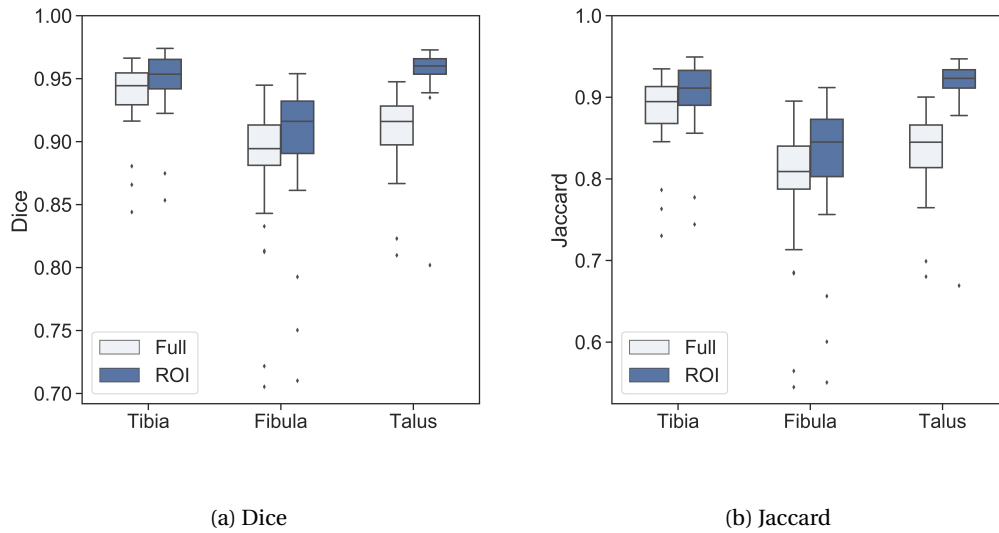


Figure 42: Segmentation quality of the 3D-CNN: Comparison of Dice score and Jaccard for the whole segmentation and the cropped region of interest to make the result independent from the partly segmented bone shafts.

viding the benefits of the SSM. In general, a lower ASD was observed for the tibia in comparison to the fibula for all methods. The talus was left out for the evaluation since it was only partially visible on a number of images.

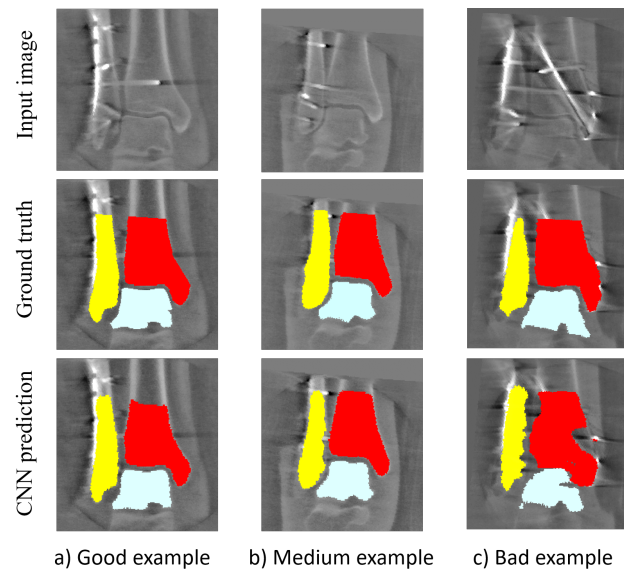


Figure 43: Visual comparison of different 3D-CNN predictions: Three segmentation examples of tibia (red), fibula (yellow) and talus (cyan) are presented (third row) in comparison to the ground truth segmentation (second row). Examples a) and b) mainly show a discrepancy along the fibula shaft. c) shows an input image containing many metal implants at different positions for which the CNN yielded the worst performance.



## 5.5 Fluoroscopic image understanding

2D bone segmentation of the fluoroscopic images is a crucial part of the proposed workflow since segmentation inaccuracies may directly affect the reconstruction error adversely. This section gives an overview of conducted experiments to assess the accuracy of the CNN-based segmentation and landmark calculation under different settings.

### 5.5.1 Training data and network parameters

Experiments were conducted by performing a four-fold cross-validation on all 30 data sets of  $\mathbf{S}_{\text{cadaver, uninjured}}$ . The resulting data set collection was split into 19 training, 3 validation and 6 to 8 test subsets respectively. One data set comprised 100 2D images each. The projected 3D ground truth segmentations were used for training the network and as a reference for comparison. Hyperparameters of the 2D U-Net are shown in Tab. 6. In a pre-processing step, the images were downsampled from  $1024^2$  to  $512^2$  pixels and normalized with min-max normalization. All images were cropped to an input size of 384. The U-Net was trained for 150 epochs for each fold with a learning rate of  $10^{-3}$ , a learning decay of  $10^{-4}$  and a batch size of 12. Different types of data augmentation were applied based on the variance of the data. A detailed description of the different strategies can be found in section 4.3.1.

Parameter	Value
Downsampled image size	512
Cropped image size	384
Batch size	12
Optimizer	ADAM
Loss	Binary cross entropy
Activation function	ELU
Normalization	Instance
Learning rate	$10^{-3}$
Weight decay	$10^{-4}$
Epochs	150

Table 6: Network parameter selection for the 2D U-Net

Test time augmentation was applied by reflecting each image along all image axes similar to the 3D-CNN segmentation described in section 5.4.

### 5.5.2 2D segmentation accuracy

The accuracy of the 2D-CNN directly affects the performance of the pipeline since the outcome is used as features for the reconstruction. Hence, it was important to investigate the accuracy of the resulting segmentations. The method, assuming that the architecture design is established and the hyperparameters are chosen, is dependent on the training data used to optimize the network weights as well as the input images. For each fold, the network was trained on the training set, validated on the validation set and accuracy was measured on the test set. The accuracy was quantified by measuring the overlap metrics described in section 5.3. Since the test data was center cropped with a margin of 128 pixels, the comparison was only performed on the cropped image.

### 5.5.3 Viewport restriction

One challenge in segmentation of fluoroscopic images is that the super-positioned appearance of the bones is highly dependent on the C-arm viewport. This offers infinite possibilities that can barely be covered in the training data unless it is enlarged by simulated training data. However, a simulation of training data with C-arm volumes could not be realized in a preliminary experiment due to artifacts and inhomogeneities attributed to the image intensifier technology. As a consequence, the acquisition of training data was constrained to a single orbital rotation of the C-arm. Rotation angles were sampled in  $1.9^\circ$  steps and volume acquisition diverged for each patient in terms of the starting position of the C-arm and the patient positioning. Therefore, the training data contained a sparse representation of the anatomical variance. One possible solution to reduce the variability was to train the network only on specific viewports and restrict the variability. In this experiment, the accuracy and sensitivity of three different configurations were compared:

- a) Training is performed on all subjects of  $\mathbf{S}_{\text{cadaver, uninjured}}$ .
- b) Training data is limited to three specific viewports  $\mathbf{S}_{\text{SP}} \in \mathbf{S}_{\text{cadaver, uninjured}}$  with  $\text{SP} \in [\text{AP}, \text{Mortise}, \text{Lateral}]$  corresponding to the standard projections AP, mortise and lateral.
- c) Training and testing is performed for each  $\mathbf{S}_{\text{SP}} \in \mathbf{S}_{\text{cadaver, uninjured}}$  with  $\text{SP} \in [\text{AP}, \text{Mortise}, \text{Lateral}]$  individually.

Ideal standard projections were not available and had to be approximated by the most similar projection chosen by an orthopedic surgeon from all 100 fluoroscopic images. To enlarge the training data set, 8 images were sampled in a range of approx.  $\pm 8^\circ$  around the

selected standard projection. Analogous to the previous experiment, the overlap metrics were used to assess the accuracy of all images in each fold.

#### 5.5.4 Landmark detection

A sufficiently well-chosen initialization plays an important role in bone reconstruction since the method is often prone to get stuck in local minima. One essential part of the proposed initialization is the accurate detection of corresponding 2D landmarks in the fluoroscopic images. In this experiment, all landmarks, described in section 4.3.2, were computed on the CNN-based segmentations and compared to the landmarks derived from the corresponding ground truth segmentations. The translational and rotational deviation between the landmarks was measured to investigate the influence of segmentation inaccuracies on landmark detection. For landmark detection, the Hough space is spanned with a resolution of 5 pixels and 1 degree. To extract meaningful contour lines with a higher probability of belonging to the shaft, the minimum line length is set to 30 pixels and the minimum line gap to 20 pixels.

#### 5.5.5 Results

**2D-CNN segmentation:** The best performance could be achieved when applying the 2D-CNN on the tibia with a Dice score of  $0.965 \pm 0.016$  averaged over all 100 images and all data sets (see Tab. 7). Slightly worse results were yielded for the fibula ( $0.937 \pm 0.041$ ) and the talus ( $0.937 \pm 0.059$ ). The performance evaluated only on the selected standard projections yielded better results for the fibula and reduced the number of outliers (see Fig. 44). The mean run time of the CNN prediction for a single fluoroscopic projection was 50 ms.

	Tibia		Fibula		Talus	
	All	SP	All	SP	All	SP
Dice	$0.965 \pm 0.016$	$0.965 \pm 0.013$	$0.937 \pm 0.041$	$0.946 \pm 0.021$	$0.937 \pm 0.059$	$0.937 \pm 0.550$
Jaccard	$0.932 \pm 0.029$	$0.933 \pm 0.024$	$0.883 \pm 0.062$	$0.898 \pm 0.037$	$0.887 \pm 0.090$	$0.886 \pm 0.087$
Precision	$0.969 \pm 0.023$	$0.970 \pm 0.019$	$0.950 \pm 0.045$	$0.956 \pm 0.029$	$0.940 \pm 0.093$	$0.941 \pm 0.090$
Sensitivity	$0.962 \pm 0.029$	$0.961 \pm 0.027$	$0.926 \pm 0.056$	$0.938 \pm 0.035$	$0.941 \pm 0.039$	$0.939 \pm 0.038$
Specificity	$0.992 \pm 0.007$	$0.992 \pm 0.005$	$0.993 \pm 0.985$	$0.994 \pm 0.004$	$0.990 \pm 0.016$	$0.990 \pm 0.014$
Accuracy	$0.985 \pm 0.008$	$0.985 \pm 0.006$	$0.985 \pm 0.009$	$0.987 \pm 0.006$	$0.982 \pm 0.014$	$0.983 \pm 0.012$

Table 7: Comparison of different metrics for different bones and image segments computed for all images (All) and three standard projections (SP)

**Viewport restriction:** Fig. 45 depicts the results when the network training was restricted to the specific viewports AP, lateral and mortise. The Dice score indicates that the performance of the CNN could not be improved by limiting the viewport compared to the

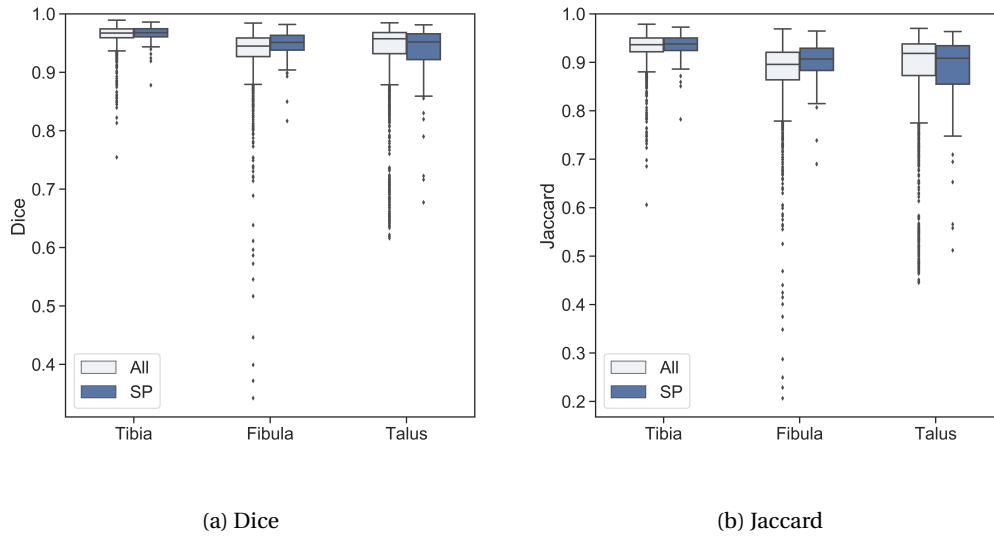


Figure 44: 2D-CNN segmentation evaluation: Comparison of Dice score and Jaccard for tibia, fibula and talus. The evaluation was performed on all 100 fluoroscopic images (All, light grey) and on three selected standard projections (SP, dark blue).

training on all training images (All).

**Landmark extraction:** Evaluating the accuracy of the extracted landmarks from the fluoroscopic projections using the ground truth segmentations as opposed to the 2D-CNN predictions shows that the mean intersection point deviation is 39 pixels. 39 pixels roughly translate to 5 mm.

Intersection point error [px]	Angular deviation fibula [°]	Angular deviation tibia [°]
$38.74 \pm 31.88$	$0.94 \pm 0.87$	$4.23 \pm 5.07$

Table 8: 2D landmark extraction

## 5.6 2D-3D bone reconstruction

The experiments conducted in this section all focused on the main research question, whether an accurate 3D multiple bone model can be reconstructed from a limited number of 2D images. 2D-3D reconstruction is dependent on a variety of aspects either regarding the quality of the input or individual components of the algorithm. The series of experiments described in this section investigated how influencing factors affect the accuracy and robustness of the approach. After performing the 2D-3D reconstruction

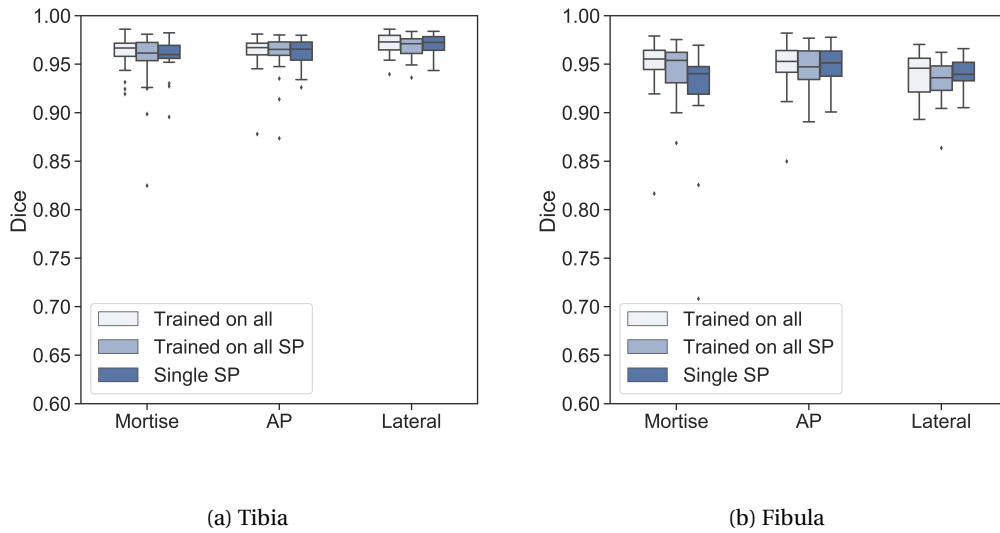


Figure 45: Comparison of different viewport restrictions: Dice scores for the three standard projection images mortise, AP and lateral. Training was performed on all images (All, white), on all three standard projections (SP, light blue) and on each standard projections separately (Single SP, dark blue).

under different configurations, the resulting 3D surface models were validated against the 3D ground truth segmentation. The accuracy was measured using the ASD.

### 5.6.1 Training data and 3D-SSM parameters

A four-fold cross-validation on  $S_{\text{cadaver, uninjured}}$  was conducted for this experiment. A 3D-SSM was created using the methods described in section 4.2.1. The tibia and talus 3D-SSM comprised of 2000 landmarks and the fibula 3D-SSM of 1500 landmarks. Segmentations were pre-aligned and tibia and fibula shafts were cut as already described in section 5.4. All data sets were reflected to correspond to right ankles.

The reconstruction was performed in 4 iterations with varying parameters. The number of sampled probes per side was reduced from 6 to 3 probes after iteration 2 to increase the robustness of the coarse alignment step. The multi-object GPA was applied in the first two iterations to further assist the SSM towards the global optimum solution. As convergence criterion, a maximum movement threshold of 0.1 was applied together with a maximum number of iterations. If one of the criteria was reached, the reconstruction was stopped. The spacing between two sampled probes was set to  $s = 0.5$  for all iterations and the maximum number of altered shape parameters  $\mathbf{b}$  was set to the first 20 parameters. The shape parameters were only modified in the last iteration since they were only expected to change slightly.

Besides the input image, the method was provided with the laterality of the presented

ankle. Since the SSM was only trained on right ankles, it was reflected in case the image showed a different laterality.

Iteration	1	2	3	4
Num. of probes per side	6	6	3	3
Resolution	2	1	0	0
Multi-object GPA	yes	yes	no	no
Shape adaption	no	no	no	yes
Max. number of iterations	50	50	10	30

Table 9: Reconstruction parameter selection

### 5.6.2 Performance evaluation under ideal conditions

In the first experiment, the 2D-3D reconstruction was performed under ideal conditions. For that purpose, shape parameters  $\mathbf{b}$  of  $\mathbf{X}_{\text{bone, uninjured}}$  were initialized with the shape parameters of the corresponding training image  $\mathbf{b}_{\text{train}}$ . Additionally, the transformation  $\mathbf{T}_{\text{train}}$  of the training shape was used to place the 3D-SSM at the ground truth location. The ground truth 2D segmentations were used in the actual reconstruction step to eliminate any errors resulting from segmentation inaccuracies. This way, the reconstruction result was not biased by any input errors and served as a baseline for comparison.

### 5.6.3 Influence of the 3D-SSM

In the proposed workflow, the initial 3D-SSMs should closely resemble the reconstructed 3D-SSMs given that the prior 3D-SSM segmentation worked well and left and right ankle did not show any intra-individual shape deviations. Under realistic conditions, the proposed 2D-3D reconstruction method should cope with shape deviations. Therefore, the 3D-SSM was first initialized with  $\mathbf{b}_{\text{train}}$  of the respective training image and then perturbed by adding random Gaussian noise with  $\sigma = \sigma_b$  to the first six shape parameters.

### 5.6.4 Influence of initialization

Depending on the chosen optimization method, 2D-3D reconstruction can be sensitive towards an erroneous initial model placement. This experiment was designed to elaborate whether the proposed initialization method affects the overall reconstruction accuracy. For that purpose, the proposed initialization method was applied to pre-initialize the 3D-SSM with the 2D ground truth segmentation before executing the 2D-3D reconstruction. The result was compared to the baseline, in which the 3D-SSM was placed at the ground truth location of the test image.

### 5.6.5 Influence of the 2D segmentations

Feature-based reconstruction methods are known for the dependence on the quality of the features. In this work, contours extracted from CNN-based 2D segmentations of the bones serve as features for the 2D-3D reconstruction. Falsely predicted labels lead to inaccuracies and disturb the reconstruction process. The accuracy of the reconstruction was examined using predicted CNN segmentations and compared to the baseline using ground truth 2D segmentations to investigate the influence of the features.

### 5.6.6 Number of projection images

As elaborated in section 3.4.4, the number of fluoroscopic images varies in literature ranging from one single image to several images. Most commonly, bones are reconstructed from two well-defined viewports. In this experiment, the number of fluoroscopic images was varied to assess how many images are sufficient for the application. Since the initialization and the reconstruction method depend on triangulation, the minimal number of images is set to two.

### 5.6.7 Results

**Comparison of different influencing factors:** As depicted in Fig. 46, the influence of the different components was compared by measuring the ASD. The boxplot in the first column serves as baseline given optimal conditions and yielded the best result by using an SSM initialized with  $\mathbf{T}_{\text{train}}$  and  $\mathbf{b}_{\text{train}}$  and ground truth 2D segmentations. The proposed initialization (second column) did not increase the ASD for the tibia and only slightly increased the ASD of the fibula in comparison to the ground truth position. The third column simulated the impact of shape perturbation on the performance and resulted in a slight increase of the ASD for both bones. The fourth column shows how the ASD changed when the predicted CNN segmentations were used instead of the 2D ground truth segmentations. The ASD of the resulting 3D model shown in Fig. 46 for the tibia and the fibula is higher when being compared to the former experiments due to 2D-CNN segmentation errors influencing the result in a negative way. The CNN segmentations seem to have the greatest impact on the accuracy. The fifth experiment simulated the error of a defective shape model segmentation and actual 2D deep learning predictions after initialization (Fig. 46, fifth column). The results show that the proposed 2D-3D reconstruction method was capable of adapting well to the shape variation for the fibula and slightly increased the ASD for the tibia. As depicted in Fig. 47, inaccuracies were mainly attributed to a local shape deviation at the incisura.

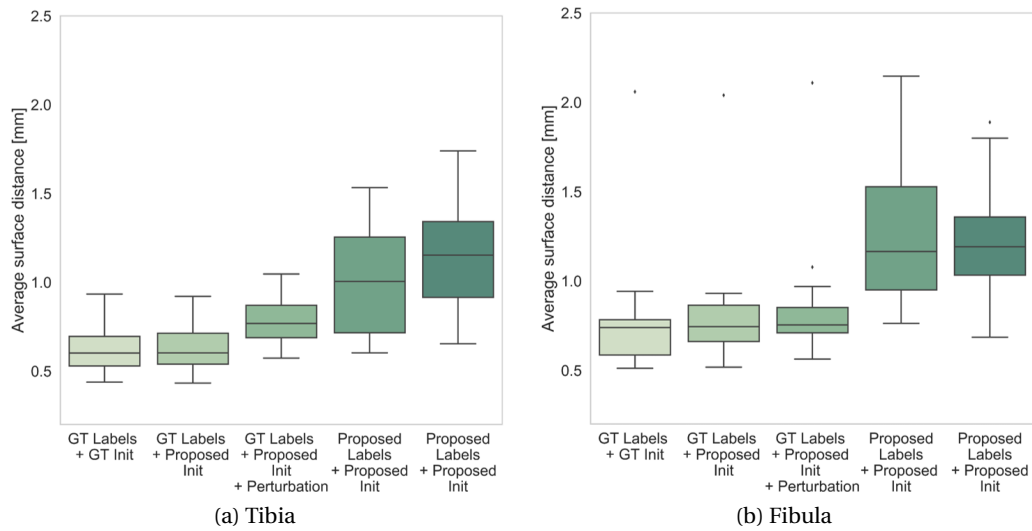


Figure 46: 2D-3D reconstruction accuracy of tibia and fibula: Errors were systematically added to assess the influence of the respective error source: Ground truth 2D segmentation with ground truth initialization (baseline, first column), ground truth 2D segmentation with proposed initialization (second column), ground truth 2D segmentation with proposed initialization and shape perturbation (third column), 2D-CNN segmentation with proposed initialization (fourth column) 2D-CNN segmentation with proposed initialization and shape perturbation (fifth column). Results show that the quality of the 2D-CNN segmentations influenced the outcome most.

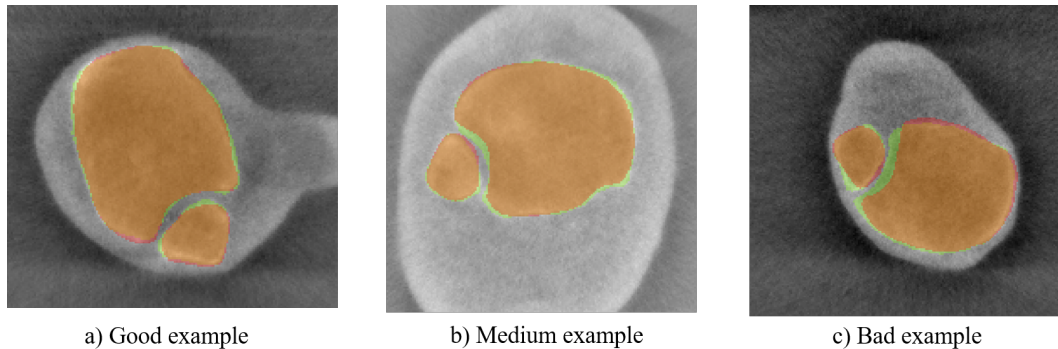


Figure 47: Visual inspection of the 2D-3D reconstruction using 2D-CNN predictions: Visual results of different quality are presented. Local inaccuracies in b) and c) are most prominent at the convex incisura.

**Number of projection images:** By increasing the number of projection images, the accuracy was improved for the fibula bone (Fig. 48). The accuracy of the tibia even slightly decreased with the number of 2D images.



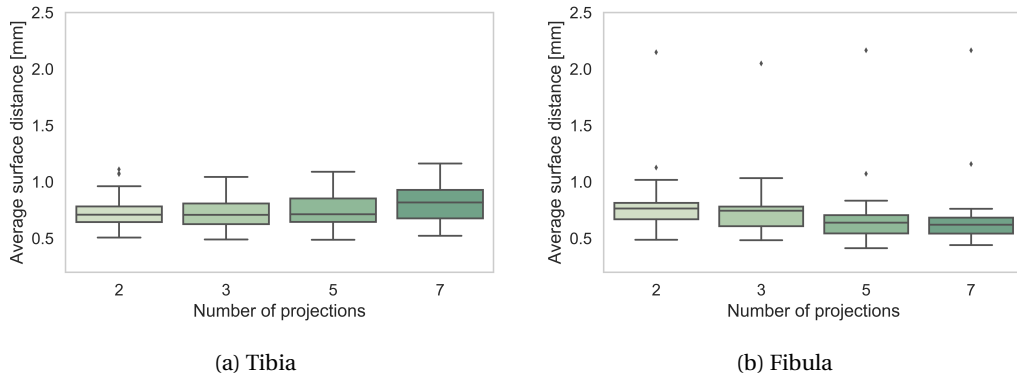


Figure 48: Influence of the number of images: 2D-3D reconstruction accuracy of tibia and fibula with a varying number of 2D images.

The time consumption depicted in Tab. 10 increases almost linearly with the number of 2D images by 10 seconds per additional image. The time measurements were averaged over all 30 executions of the reconstruction method. The C++ based MITK implementation was executed using an Intel® Core™i7-7820HQ CPU at 2.90GHz with 32 GB of RAM.

Number of 2D images	2	3	5	7
Time [s]	$28.74 \pm 1.61$	$38.77 \pm 1.90$	$61.92 \pm 2.92$	$81.23 \pm 4.84$

Table 10: Time consumption for 2D-3D reconstruction

## 5.7 System evaluation

The following experiments were designed to evaluate the entire pipeline on a limited number of realistic clinical cases for which 3D and 2D images were available for the injured and uninjured ankle. The overall performance was evaluated in terms of 3D accuracy, 2D accuracy for specific slices of relevant regions and computation time.

### 5.7.1 Adjustment of standard planes

With the proposed standard plane adjustment method, the three standard viewing planes are adjusted automatically. The proposed method was applied to all injured ankles of  $S_{\text{clinical,pairs}}$ . A trauma surgeon manually adjusted the standard planes according to the definition described in Tab. 1 to create a reference adjustment. After applying the 3D-SSM segmentation, the angle deviation of the axial and sagittal plane and the positional

deviation of the crosshair position between the computed and the reference planes were measured. The term crosshair position refers to the 3D intersection point of all three planes.

### 5.7.2 Overall error estimation

The purpose of this experiment was to test the whole pipeline under realistic settings on the eleven clinical pairs of  $\mathbf{S}_{\text{clinical,pairs}}$  in a retrospective study. The 3D U-Net and SSMs were trained using all ankles in  $\mathbf{S}_{\text{clinical,injured}}$ . The 2D U-Net was trained on all data sets in  $\mathbf{S}_{\text{cadaver,uninjured}}$ . To enlarge the number of training data, the 2D-CNN was trained in a leave-one-out cross-validation. For each patient, the 2D-CNN was trained on all other data sets in  $\mathbf{S}_{\text{clinical,pairs}}$  and  $\mathbf{S}_{\text{cadaver,uninjured}}$ . 3D-SSM, CNN and reconstruction parameters remained constant and can be taken from Tab. 3, 4 and 9. The proposed workflow, shown in 4.1.2, was conducted on the entire data set corresponding to each ankle pair. First, the 3D-SSM with the CNN-complement was used to perform a segmentation on the 3D image of the injured ankle. Next, the 2D-3D reconstruction was performed using the 3D-SSM  $\mathbf{X}_{\text{bone,injured}}$  and the 2D projection images of the respective uninjured ankle. The resulting 3D-SSM  $\mathbf{X}_{\text{bone,uninjured}}$  was compared to the 3D ground truth segmentation of the uninjured ankle. Afterward, both,  $\mathbf{X}_{\text{bone,uninjured}}$  and the reference ground truth were registered to  $\mathbf{X}_{\text{bone,injured}}$  and comparison was repeated.

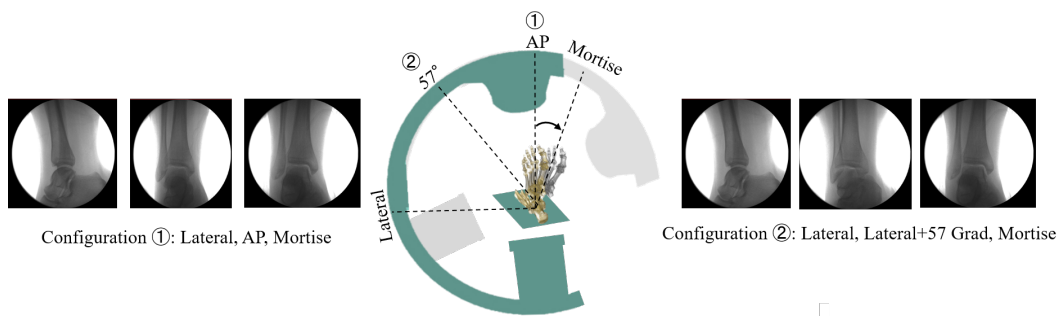


Figure 49: Two configurations used for the overall evaluation: Configuration 1 consists of projection images lateral, AP and mortise (left). Configuration 2 consists of projection images lateral, lateral+57° and mortise.

Preliminary experiments already indicated that using the three standard projections might not be the optimal selection for the 2D-3D reconstruction. To examine whether a different selection is beneficial, two configurations were chosen (cf. Fig. 49):

- 1 The overall pipeline was conducted using three standard projections, namely AP, mortise and lateral view.

- 2 Instead of using the AP view as third standard projection, the overall pipeline was conducted using mortise, lateral and a third projection image between mortise and lateral with  $B_3 = \text{lateral} + 57^\circ$ .

With this experiment, not only the overall error for two configurations but also the influence of the 3D-3D registration could be assessed. In clinical routine, the surgeon examines the result of the reduction at two levels, in particular, approximately 10 mm above the tibia plafond to evaluate the fibula in the incisural notch and 6 mm below to evaluate the fibular rotation (cf. Fig. 39). Besides conducted 3D measurements before and after the 3D-3D registration, these two slices were considered for a more specific comparison most relevant to the surgeon. The Hausdorff distance was used to measure the largest deviation between the two shapes.

### 5.7.3 Clinical evaluation

The computation of shape metrics is a common and adequate way to quantify errors of the pipeline. However, those metrics are not easy to interpret in the clinical context. Therefore, another experiment was conducted to assess the qualitative and quantitative error from a surgeons' perspective. The surgeon was asked to examine the reconstruction result  $\mathbf{X}_{\text{bone, uninjured}}$  overlaid on the 3D image of the injured ankle and to decide whether and how the reduction result must be modified. Then, the surgeon was shown the real 3D intensity image of the uninjured ankle as an overlay and asked, whether the same decision would have been made given the entire information of both ankles. Furthermore, the surgeon was asked to report, whenever the shape of the reconstructed  $\mathbf{X}_{\text{bone, uninjured}}$  appeared suspicious or irregular.

### 5.7.4 Results

**Standard plane adjustment:** Tab. 11 presents the mean deviation of the computed standard planes compared to the ground truth standard planes adjusted by an orthopedic surgeon. Results were averaged over all 11 patients. The mean angle deviation of the axial standard plane is smaller than the mean sagittal standard plane deviation.

Positional deviation [mm]	Deviation axial plane [°]	Deviation sagittal plane [°]
$2.34 \pm 0.75$	$2.47 \pm 1.74$	$7.44 \pm 2.71$

Table 11: Standard plane deviation

## Overall evaluation

Tab. 12 and 13 show the 3D and the 2D results for eleven clinical cases evaluated on the two evaluation planes (cf. section 5.2 for a definition). The results in Tab. 13 refer to configuration 2 in which a third projection image halfway between mortise and lateral view was chosen. All measures show an improvement when compared to the results achieved with the conventional selection (mortise, lateral, AP).

	3D-SSM segmentation				2D-3D reconstruction			
	$\mu$	$\sigma$	min	max	$\mu$	$\sigma$	min	max
3D surface evaluation tibia								
ASD [mm]	0.653	0.151	0.446	0.942	0.890	0.113	0.731	1.053
3D surface evaluation fibula								
ASD [mm]	0.879	0.534	0.498	2.417	1.014	0.345	0.682	1.691
	2D-3D reconstruction- induced error				Overall error after 3D-3D registration			
	$\mu$	$\sigma$	min	max	$\mu$	$\sigma$	min	max
2D Position 1 (10 mm above tibial plafond)								
Dice	0.913	0.027	0.857	0.956	0.882	0.060	0.748	0.945
Jaccard	0.842	0.045	0.750	0.915	0.794	0.060	0.598	0.895
Hausdorff [mm]	1.70	0.39	1.08	2.50	2.09	0.52	1.11	2.97
2D Position 2 (6 mm below tibial plafond)								
Dice	0.918	0.033	0.846	0.954	0.892	0.047	0.784	0.954
Jaccard	0.849	0.056	0.733	0.913	0.807	0.074	0.645	0.912
Hausdorff [mm]	2.45	0.87	1.49	4.14	2.76	0.93	1.49	4.78

Table 12: Quantitative evaluation of the whole pipeline for configuration 1

The first column compares the reconstruction result directly to the ground truth segmentation. The second column depicts the result after the reconstruction result and ground truth segmentation are mapped to the injured ankle. The mean Hausdorff distance of configuration 2 is  $1.65 \pm 0.48$  mm before registration and slightly increases to  $1.72 \pm 0.57$  mm after 3D-3D surface registration at 10 mm above the tibial plafond (position 1). In addition, the mean Dice score decreases after registration. At 6 mm below the tibial plafond (position 2), the mean Hausdorff distance is higher before and after registration. This might be attributed to the smaller contour size at that position that is more sensitive to shape deviations. The Dice score decreases similar to the score at position 1. The visual results depicted in Fig. 50 (third column) show that in most cases ground truth segmentation (green) and reconstruction of the uninjured side (red) overlap reasonably. Some cases e.g. patient 2 show a shape deviation due to errors of the workflow mainly attributed to inaccuracies in the 2D deep learning segmentation.

	3D-SSM segmentation				2D-3D reconstruction			
	$\mu$	$\sigma$	min	max	$\mu$	$\sigma$	min	max
3D surface evaluation tibia								
ASD [mm]	0.653	0.151	0.446	0.942	0.873	0.150	0.704	1.269
3D surface evaluation fibula								
ASD [mm]	0.879	0.534	0.498	2.417	0.935	0.301	0.578	1.536
2D-3D reconstruction-induced error								
	2D-3D reconstruction-induced error				Overall error after 3D-3D registration			
	$\mu$	$\sigma$	min	max	$\mu$	$\sigma$	min	max
2D Position 1 (10 mm above tibial plafond)								
Dice	0.917	0.018	0.8777	0.941	0.898	0.045	0.834	0.957
Jaccard	0.848	0.031	0.781	0.888	0.753	0.110	0.495	0.863
Hausdorff [mm]	1.65	0.48	1.08	2.61	1.72	0.57	0.99	3.11
2D Position 2 (6 mm below tibial plafond)								
Dice	0.919	0.032	0.855	0.960	0.910	0.064	0.731	0.934
Jaccard	0.852	0.053	0.747	0.923	0.836	0.064	0.731	0.934
Hausdorff [mm]	2.32	0.63	1.49	3.43	2.35	0.64	1.49	3.43

Table 13: Quantitative evaluation of the whole pipeline for configuration 2

Method	Time [s]
3D-CNN segmentation	$1.65 \pm 0.48$
3D-SSM segmentation	$22.60 \pm 0.78$
2D-CNN segmentation	$<0.06$
2D-3D initialization	$1.24 \pm 0.06$
2D-3D reconstruction	$38.27 \pm 2.89$
3D-3D reconstruction	$5.25 \pm 0.49$

Table 14: Mean time consumption: Pipeline methods

Tab. 14 shows the mean time consumption for the different components measured for all 11 cases. The C++ based MITK implementation was developed and executed using an Intel® Core™i7-7820HQ CPU at 2.90GHz with 32 GB of RAM. All CNN based methods were implemented with the python based framework PyTorch and trained as well as executed using an Intel® Core i7-3930K™CPU at 3.2 GHz with 32 GB of RAM. The results show that the 2D-3D reconstruction and the 3D-SSM segmentation are the most time-consuming methods of the pipeline. All individual components were executed in less than a minute.

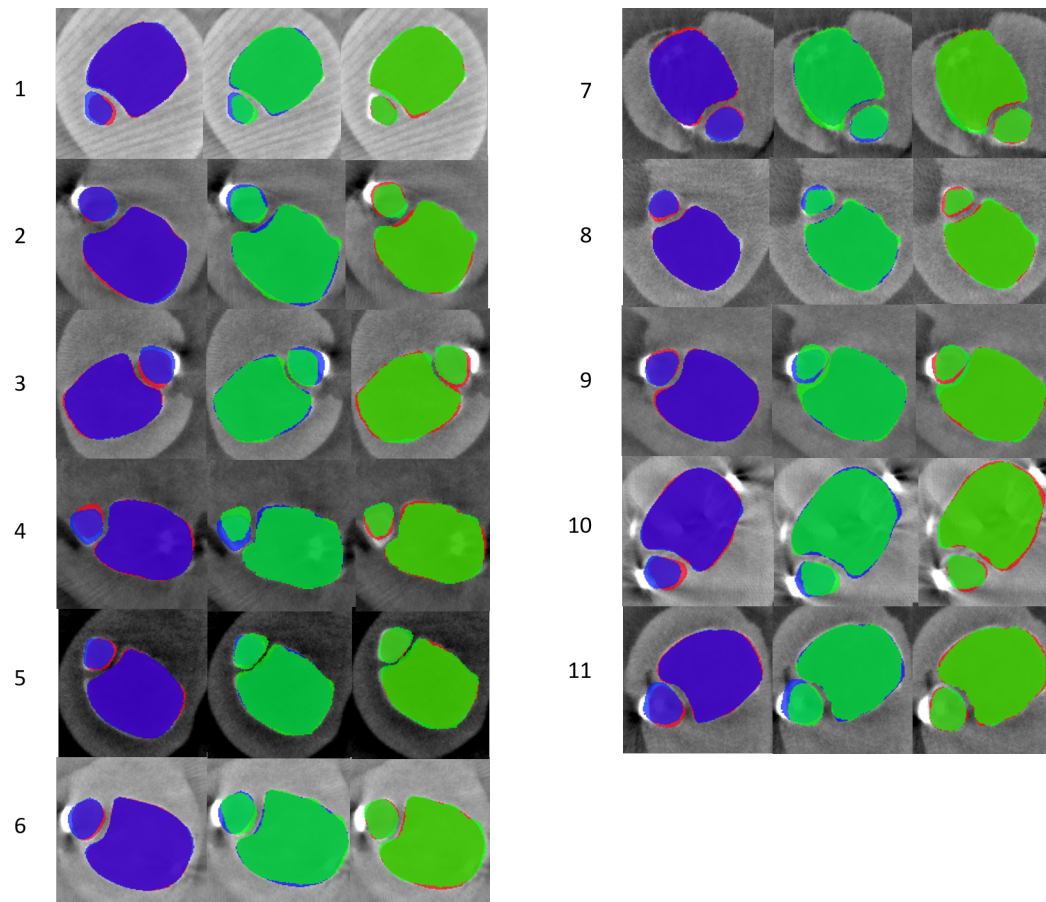


Figure 50: Visual comparison of the overall pipeline: Results are depicted for all eleven clinical cases after executing the whole pipeline including the 3D-3D registration. In the background, the 3D image of the injured ankle is displayed at evaluation position 1 (10 mm above the tibia plafond). The blue surface refers to the segmentation of the injured ankle. The red surface resembles the computed reconstruction result of the uninjured ankle mapped to the segmentation of the injured ankle. For comparison, also the ground truth segmentation of the uninjured ankle is mapped to the injured ankle (green surface). Column 3 shows the discrepancy between the computed and the ground truth segmentation.

### Clinical evaluation

Tab. 15 depicts the results of the visual inspection performed by two trauma surgeons. Both were asked to decide whether they would perform a correction of the reduction result given the information of the computed reconstruction result (left) and the real intensity image of the contralateral side (right). In case of an optional correction (Tab. 15,(O)), the deviation between the injured and uninjured fibula was almost small enough so that the surgeon would not necessarily have performed an intra-operative revision of the reduction result. Surgeon 1 made the same decision in ten out of eleven cases.

Patient	Surgeon	Correction based on Reconstruction result						Correction based on Contralateral image					
		LT	MT	AT	PT	IR	ER	LT	MT	AT	PT	IR	ER
1	1				X						O		
	2												
2	1												
	2		X										
3	1		X						X				
	2												
4	1	X			X			X			X		
	2				X						X		
5	1		(O)						(O)				
	2												
6	1		O				O		O				O
	2												O
7	1												X
	2								X				X
8	1		(O)					(O)					
	2												
9	1	O						O					
	2												
10	1				X						X		
	2				(X)						X		
11	1	X			X			X			X		
	2	O			X			O			X		

LT = Lateral translation    MT = Medial translation    AT = Anterior translation  
 PT = Posterior translation    IR = Internal rotation    ER = External rotation  
 X = Correction    O = Slight correction    (O) = Optional correction

Table 15: Clinical evaluation performed by two orthopedic surgeons

Surgeon 2 agreed in seven out of eleven cases with the reconstruction result. In one case (patient 10, (X)), he graded the reconstruction result as not valid, although he stated that - if he had trusted the result - both would have led to the same decision. In half of the cases, both surgeons disagree in their decision whether and how the reduction result needed to be corrected.





This chapter discusses general remarks (section 6.1) as well as the different methods presented in this thesis with respect to the findings of the conducted experiments and results (section 6.2-6.5). Section 6.6 gives an overview of conclusions drawn from the conducted work, section 6.7 summarizes the main contributions and section 6.8 provides future directives in the field of computer-assisted ankle surgery.

## 6.1 Intra-operative assistance system for ankle surgery

In this work, a novel approach for intra-operative assistance in ankle surgery was proposed. To enable a translation into the surgical workflow, the methods must follow strict time constraints. Often, advanced methods found in the literature (e.g. M-reps mentioned in section 3.2) can not be transferred into real clinical scenarios since their execution is too time-consuming. Thus, in this work time-efficient methods were developed and employed. Comparable systems found in the literature often are trained on normalized pre-operative CT images that allow to employ intensity-based reconstruction approaches (BONARETTI et al. 2014; EHLKE et al. 2013; ZHENG 2011). The conducted work was limited to feature-based reconstruction since C-arm images contain inhomogeneities and artifacts. A simulation of the required realistic DRRs is not feasible. Furthermore, all of the experiments conducted in the scope of this thesis were trained on a limited amount of data that especially in the case of learning-based approaches can highly affect the performance.

The proposed workflow relies on the assumption that the intra-individual variance of ankle pairs is significantly lower than the inter-individual variance among the population. An evaluation was conducted to verify the hypothesis by examining the shape

deviation of pairs as opposed to the whole population. For this particular purpose, all ankle pairs from data set collection  $S_{\text{cadaver, uninjured}}$  were selected. Those data sets were not affected by fractures or pathologies and serve the purpose of a direct shape comparison. A 3D-SSM was trained on all 24 subjects of the respective 12 pairs. Afterward, the variance of each landmark of the 3D-SSM was calculated using Eq. 16. The results depicted in Fig. 51 could substantiate the clinical hypothesis that intra-individual variation is smaller than inter-individual variation.

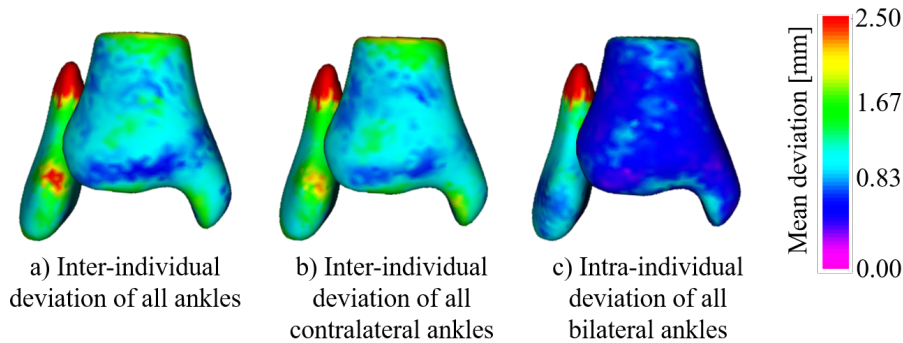


Figure 51: Mean landmark variation of tibia and fibula: Inter-individual variation between each ankle and a) all other ankles b) all other contralateral ankles from the population. c) shows the intra-individual variation between all twelve pairs. The intra-individual variation is lower compared to the inter-individual variation. The deviation at the proximal end of the fibula is caused by cutting the shafts. The figure was adapted from THOMAS et al. 2019.

## 6.2 3D bone segmentation

The results presented in section 5.4 showed that two main limitations of the state of the art methods when being applied to the ankle joint could be alleviated to enable an accurate and robust model-based segmentation of the ankle joint in C-arm images.

Errors of the SSM were mainly attributed to its sensitivity in terms of initialization or significant changes in appearance. Those changes, e.g. caused by metal artifacts at different locations can only be incorporated during SSM training when the occurrence of metal implants can be modeled by a statistical distribution. The training and testing data included different fracture types.  $S_{\text{injured, clinical}}$  contained 3D images of the injured ankle after fibula reduction and showed a large variety of different metal implants that not only varied in terms of location (different positions of the plate at the fibula shaft) but also implant type (single syndesmotoc screw or plate and multiple screws). The most notable variance appeared to be the height of the plate fixated to the fibula.

With the limited number of training data it was challenging for the SSM to generalize. A visual inspection of the results revealed that errors were mainly caused by an initialization outside of the capture range of the SSM so that it failed to converge into local minima. The fibula is a tubular slim bone that was prone to failures whenever the SSM and the real fibula in the image did not overlap. A common observation was that the fibula SSM did not reach to the distal end of the real fibula and was rather guided by the location of the metal implant.

The 3D-CNN segmentation yielded accurate results. Occasionally surface bleeding, dislocated voxels or holes were observed. The approach is further limited by the varying shaft. To reduce the impact of the cut shaft, a comparison to the ground truth segmentations was performed on the cropped image. This induces a certain bias, however, it only affects areas that are less relevant for the subsequent pipeline. The performance of the CNN network might be further improved in the future by performing a more advanced hyperparameter search since the parameters used in the conducted experiments were selected by manual search starting from recommendations found in the literature.

Results in section 5.4.6 clearly demonstrate the superiority of the proposed method over the pure intensity-based SSM. The accuracy of the segmentation could be sustained for arbitrary initial placements of the model. The segmentation quality is independent of the initial position. This robustness is attributed to the coarse registration of the model to the semantic segmentations. The results further demonstrated that the proposed method could even outperform the pure intensity-based segmentation in case of an ideal SSM placement.

The proposed hybrid segmentation comes with the advantage of providing corresponding landmarks at the cost of only a slight reduction of accuracy compared to the 3D-CNN segmentation. The CNN-complemented SSM method was capable of handling the varying positions of the one-third tubular plate. One patient of the data set collection was treated using a one-third tubular plate and additionally multiple, large screws (cf. Fig. 52). This sample yielded one of the highest errors in the SSM evaluation which can be explained by its out-of-distribution appearance that has not been learned. In that particular case, the subsequent SSM segmentation is dependent on the CNN and failed to perform a valid segmentation. A solution would be to re-train the SSM and CNN with more training data to cover a larger variety of treatment types.

Another advantage of the proposed hybrid SSM method is the possibility to transfer to different modalities. Since shape and appearance are completely decoupled, the SSM is independent from the modality and can be applied to e.g. MRI data by training an

additional neural network.

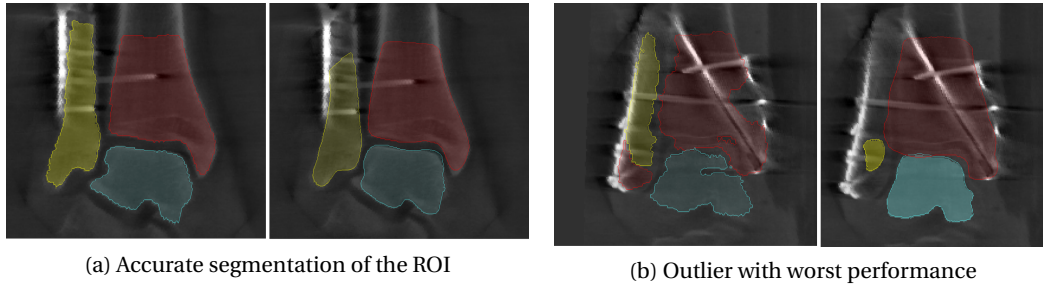


Figure 52: Accurate 3D-SSM segmentation of the ROI versus the outlier with worst performance: The 3D-CNN is shown on the left and 3D-SSM with CNN complement on the right side. The worse example shows a 3D image with larger and more metal implants that were not fully represented in the training.

The talus was only considered for the CNN experiment since its SSM segmentation was often hampered by being only partially visible on the C-arm image. The pure-intensity based method completely fails if a large proportion of the mesh is outside of the image. The CNN-complemented method only relies on the segmented sections that allows the model to remain at the desired position but fails to establish the correct shape appearance. Still, it can be useful for the initialization of the multi-object model. By training and applying the SSM only on images with the talus fully present this limitation can be eliminated. However, to transfer the method into the clinical workflow, one pre-requisite would be to ensure that the ankle is placed with the tibia plafond in the iso-center.

### 6.3 2D fluoroscopic image understanding

2D-CNN segmentation plays an important role in the pipeline. Experiments were conducted on the data sets of  $S_{\text{clinical,pairs}}$  since the task requires the segmentation of uninjured ankle joints in C-arm images.

The conducted cross-validation results showed mixed results for the different bones (examples are provided in Fig. 53). The accuracy was impeded by obstacles like the table or other external objects that could not be sufficiently represented in the training data and visual inspection showed that the table and the fibula were very similar in intensity and shape which further hampered especially fibula segmentation.

One other explanation for the segmentation performance might be based on the creation of the ground truth data. Since there was no ground truth data for the fluoroscopic images, it was derived by projecting the 3D ground truth segmentations onto the projection planes using the projection matrices. Hence, the created segmentations were influenced

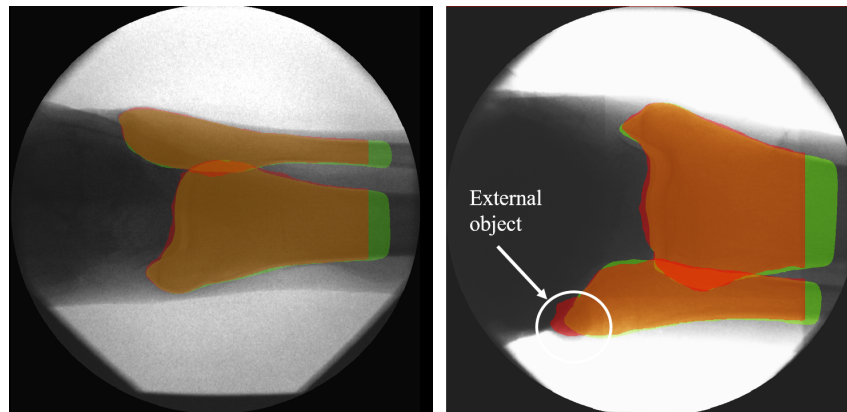


Figure 53: Visual inspection of the 2D-CNN segmentation applied to data of the cadaver study: The left image shows an accurate 2D segmentation comparing the result (red) with the ground truth segmentation (green). The right image shows an example that contains an external fixation plate used for the study in the FOV. Since plate and bone had similar intensity values, the plate was mislabeled as fibula.

by potential calibration errors.

A sampling from only 100 images that are acquired by an orbital rotation around the ankle and different for each subject results in a sparse sampling of the overall distribution and naturally did not cover all degrees of freedom. It is therefore expected to contain out-of-distribution data in the test set.

The experiment conducted in section 5.5.3 was designed to investigate whether the segmentation accuracy could be increased by limiting the viewport. Viewport restriction should naturally decrease the intensity variation while increasing the prediction accuracy. However, results indicate the opposite. The reason for this rather contradictory result might be based on the circumstance that a viewport restriction also highly reduced the number of training data. The CNN seems to be not unable to generalize well towards unseen data when trained on less training data even if the appearance of the data in terms of the shape of the bone is more similar.

The data sets of the cadaver study have the additional limitation that they do not completely resemble the variation of the clinical data. One reason is the positioning of the limbs, which were separated from the body of the donor. This allowed to place the foot in different positions from a realistic configuration. Both cadaver studies entail different limb positions and different C-arm configurations. Furthermore, 18 of the limbs were fixated by a plastic plate to ensure an even positioning. This plate is visible in the 2D images and might have added additional errors to the CNN training. A C-arm of identical

construction and type was used for the clinical experiment and the cadaver study. However, both devices were calibrated independently and might have different calibration errors. This has to be taken into account for the ground truth 2D segmentation creation.

It must be noted that all images were center cropped before CNN segmentation. Thus, the image borders were not taken into account for training and evaluation. However, the relevant part of the ankle joint is most likely located in the image center and therefore a sparing of image borders should not affect the pipeline adversely. Similar to the 3D-CNN training, hyperparameter search was performed manually and could be further improved by more advanced search methods.

The landmark extraction showed reasonably good results. Even the maximal observed deviation of 110 pixels (approximately 14 mm) placed the model close enough to its final position as shown in the subsequent experiments. However, the performance is again dependent on the CNN segmentation and therefore can be improved accordingly.

As observed in the results of the 2D-3D reconstruction-related experiments, the accuracy of the 2D segmentation plays the most important role in the pipeline. 2D segmentation inaccuracies have the greatest impact on the reconstruction result. Thus, it is essential to further improve performance. In conclusion, the limitations of the proposed methods are explainable and can probably be addressed using a larger amount of training data with accurate ground truth segmentations.

## 6.4 2D-3D reconstruction

Different assumptions about the 2D-3D reconstruction can be derived from the conducted systematic experiments. The proposed initialization scheme provides a sufficient pre-alignment of the 3D-SSM. No significant differences could be observed for the reconstruction accuracy compared to the perfect initialization. Perturbing the shape parameters with the variance of the SSM slightly increased the ASD. It has to be noted that the evaluation is limited because the noise added to the 3D-SSM changed the shape globally and probably with a higher variance than the error induced by the 3D-SSM segmentation of the injured ankle. To evaluate realistic model deviations, the 3D-SSM would have to be altered locally. The results indicate that the quality of the 2D segmentation has the greatest impact on the performance of the 2D-3D reconstruction. Hence, to improve the overall performance, attention has to be given to the improvement of the 2D-CNN segmentation that, as already stated, can be achieved by increasing the amount of training data.

Increasing the number of fluoroscopic images could only improve the accuracy of the fibula but slightly decreased the accuracy of the tibia. This rather contradictory result

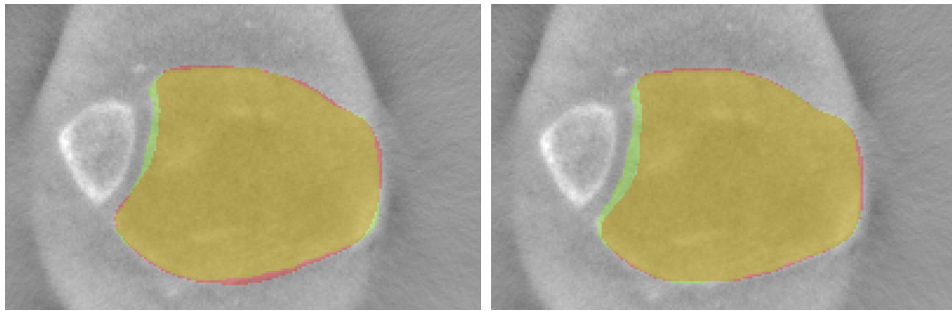


Figure 54: Reconstruction results with a varying number of projection images: Visual inspection of the tibia when reconstruction is performed using 2 projection images (left) and 7 projection images (right). When using 7 projection images, the outer contour is reconstructed more accurately but adversely affects concave areas of the incisura.

can be explained by visual inspection (cf. Fig. 54). The fibula 3D-SSM benefits from the increase of contours points on the outer surface. The tibia 3D-SSM is also guided to the contour points and, in case of 7 projections, closely approaches the bone surface (Fig. 54, right). On the contrary, the 3D-SSM alters its global shape parameters to reach the surface and deforms in areas in which the model lacks any contour points. The reconstruction method only employs the outer surface which means that concave areas like the tibia plafond or the incisura will never add information to the reconstruction. This explains that the error increases in concave areas of the tibia.

This unwanted behavior could be reduced by tagging areas of the incisura and the tibia plafond and make them independent of any shape deformation in the last step of the reconstruction. A similar approach has been shown by GÖRRES et al. 2016 for the 3D-SSM segmentation. Another solution would be to integrate inner contours or intensity values in the reconstruction process. The fibula is a convex bone that is not affected by this behavior and improves with more standard projections. Using more standard projections naturally increases the run time of the reconstruction and would increase the cumulative radiation exposure.

## 6.5 Clinical integration and visualization

The overall pipeline can be executed by providing 2D and 3D image data as well as the laterality and the corresponding projection matrices. No further human-machine interaction is necessary.

The current state of the art in ankle surgery is the manual adjustment of the single standard planes. The presented approach automatically derives standard planes from 3D images of the upper ankle joint after 3D-SSM segmentation. It is an easy and fast

method and does not require any user-interaction. The method is adaptable to different anatomies like lower ankle joint, wrist or shoulders and allows to compute standard planes defined with inter-bone relations. The method is limited by the prior 3D-SSM segmentation as an additional step but is applied as part of the overall pipeline in any case.

The overall pipeline evaluation revealed that the reconstruction as such provides accurate surface models in most cases. It could be shown that the error increases with the 3D-3D registration even though the reconstruction worked well. The 3D-3D registration is influenced not only by the reconstruction result but also by shape deviations for the previous step and the natural anatomical variance of both ankles. One possible solution would be to provide the surgeon only with the side-by-side view of both results to eliminate the bias by the registration. It could be shown that the viewport of the third standard projections influences the performance of the reconstruction. Since the standard projections AP and mortise view are relatively close to each other the 3D-SSM does not gain distinct spatial information.

The results presented in THOMAS et al. 2019 for the overall evaluation could be further improved as depicted in section 5.7. Three main differences explain the discrepancy between both evaluations. The 2D-CNN was trained in a leave-one-out cross-validation using all data sets except the test set. This increased the amount of training data and supplemented the training data with clinical cases that were more similar in appearance than the data of the cadaver study. The increased robustness and accuracy of the 2D-CNN substantiate the assumption that the results can be strongly improved by more training data.

In addition, a multi-object GPA was used to globally optimize the pose of the bones and make the model more robust towards initialization. The results could be further improved by using a third standard projection different from the AP view.

Comparing the results for both configurations lead to the conclusion that selecting a third standard projection with a high angular margin between the other two yields better results. This can be naturally explained by a better spatial distribution of anchor landmarks for the reconstruction. However, all conclusions drawn from the experiment must be interpreted with caution, since the examined clinical cases were very limited and may not reflect the entire distribution.

The entire pipeline can be executed in an average time frame of 69 s. The computation of all components except the 3D-CNN and 2D-CNN predictions are implemented for execution on the CPU only. Leveraging GPU computation power for more steps of the pipeline would make it even more efficient and most probably result in a sufficient



performance increase for typical clinical usage scenarios.

The clinical evaluation showed that the opinion of both surgeons disagreed not only in terms of the comparison between the computed reconstruction result and the real intensity image. Their decision also varied in terms of their decision whether and how they would correct the reduction result given the intensity image. Surgeon 1 agreed with the reconstruction result in all cases except in one case for which it was not possible to measure the fibula rotation. Surgeon 1 reported that the reason for that rather severe misinterpretation was attributed to the shape of the distal part of the computed fibula that was more and unnaturally convex compared to the real fibula. A convex shape in this particular region made it difficult to measure a correct rotation. In addition, Surgeon 2 reported that the talus bone looked suspicious in two cases. As stated in the previous section, the reconstruction of the talus was often hampered when it was only partially visible on the 2D images. Surgeon 2 further mistrusted the computed reconstruction result whenever the reconstructed tibia bone appeared to be shifted compared to the injured tibia. This behavior might be explained by inaccuracies in the 3D-3D registration or intra-individual variance of the left and right ankle.

## 6.6 Conclusion

With recent advances in technology and data science, many intra-operative assistance systems have been developed but not many of them are integrated into the clinical routine as yet. One reason is that the feasibility of intra-operative assistance systems is often constrained by high demands concerning the accuracy, patient safety, execution time and robustness. Furthermore, an assistance system often comprises a complex conjunction of several methods. This work concentrated on the development of robust and fast methods to enable direct clinical translation.

Assistance systems for orthopedic surgery can be found for many surgical applications, mainly concentrating on the femur, pelvis and knee. The current state of the art in ankle surgery, however, is the manual visual inspection of the reduction result using an intra-operative C-arm device. A standardized verification of the result is challenging due to inter-individual differences of the fibula shape and pose and rather based on the experience of the surgeon. Only if a strong indication is given, a 3D scan of the contralateral side is performed for comparison. In this thesis, a novel approach is presented that provides an automatized objective assistance while avoiding additional intervention time and radiation dose caused by a full 3D scan. The prototype implementation of this approach showed good results on the available data in reconstructing the contralateral side of the patient using only three projection images.

The predefined objectives of this thesis concerned time consumption, radiation dose and accuracy relevant to successful clinical integration. Most of the predefined objectives were carefully revised and fulfilled although a prospective clinical evaluation is necessary to investigate the potential integration into the clinical workflow, the usability and the robustness of the approach.

An assistance system was designed to help the surgeon during ankle surgery. The presented approach does not require any additional 3D C-arm images and thus avoids further harm to the patient. Instead, the system works with low-dose projection images of the uninjured ankle joint which is beneficial to patient and surgeon.

All methods were automatized and can be executed within one step so that the surgeon does not risk any contamination or induce time-consuming re-sterilization resulting from manual user-interaction. The technical workflow is compatible with the clinical workflow and can be integrated without additional hardware. The proposed software provides a visualization of the results with colored contour overlays that were designed in collaboration with trauma surgeons to be intuitive and easy to interpret.

The proposed system consists of time-efficient and interchangeable methods. A 3D-SSM approach was used to create a reference surface model from C-arm images of the injured ankle after fixation and reduction from 3D C-arm images. Those images were challenging in terms of the arbitrary position and orientation of the ankle in the 3D image and the sheer amount and varying position of metal implants leading to a high variance in appearance. By incorporating CNN based predictions in the initialization and SSM adaption process, the accuracy and robustness of the method could be improved. The 2D-3D multi-bone reconstruction method was targeted at providing a fast and robust estimate of the contralateral uninjured ankle by global optimization of all SSMs and an efficient sampling of distance maps.

The mean execution time of 69 s does not increase the intervention time by a large margin. The execution time can be even further reduced by the employment of GPUs to facilitate the 2D-3D reconstruction. The methods are trained on data comprising different fracture types and metal implants and thus are robust towards implant placement.

A number of potential limitations need to be considered. First, the prototype trained by using the available training data could not as yet fulfill the high accuracy standards needed for ankle surgery. In literature, clinical evidence for a malreduction is given by a discrepancy ranging from 1 mm according to HERMANS et al. 2010 to 2 mm according to GARDNER et al. 2006. The proposed system achieves a mean Hausdorff distance of 1.74 mm for the reconstructed contour after 3D-3D reconstruction. To meet both clinical assumptions, the accuracy would have to be further improved. In the future, special

attention has to be given to the local shape adaption. Concerning the tibia, the convex shape of the incisura should be reconstructed more accurately since it plays an important part in the evaluation.

The proposed pipeline consists of several components inducing certain inaccuracies to the system. All individual inaccuracies are propagated and accumulate to the overall error. This thesis provided a systematic evaluation of all different components and correlation of associated sources of error to identify possible bottlenecks of the approach and give further directions. By eliminating the influence of the 2D-CNN segmentation, it could be shown that the pipeline has a high probability of computing an accurate estimate of the uninjured ankle. There is evidence, that the inaccuracies of the CNN-driven methods can be significantly reduced with more training data. Besides algorithmic limitations, the quality of the images acquired with the image intensifier is influenced by geometric distortion, artifacts and calibration errors that need to be taken into consideration to gain further improvement.

Nevertheless, the first prototype showed that the results already provide a good estimate of the injured ankle, even with the limitations described in this section. The proposed approach can help to provide a non-biased view on the ankle that would be beneficial especially for inexperienced surgeons. However, the pipeline should be evaluated with more clinical cases to validate its benefits for surgeons and patients.

In conclusion, the novel concept for automatized contralateral side comparison has a high potential for clinical transfer in the future, because it can assist the orthopedic surgeon during the reduction process and enables an objective visual verification of the surgical outcome without increasing radiation dose or time consumption.

## 6.7 Summary of contributions

In this thesis, a novel assistance system has been designed, implemented and validated, which comprises different components. All of the components were either customized for the application or newly engineered. The main contribution of this thesis was to compute a 3D surface of the contralateral uninjured ankle from 2D fluoroscopic images to enable an intra-operative assistance method for ankle surgeries. The following paragraphs give a brief summary of all contributions:

**Concept design for contralateral side comparison:** In this thesis, the first automatized assistance in ankle surgery was proposed by designing a workflow that is capable of a contralateral side comparison. Current pitfalls like ethical concerns and economical costs were alleviated by developing a workflow based on 2D

fluoroscopic projections instead of a full 3D scan.

**3D segmentation on C-arm images of the injured ankle:** A method was developed that combines the shape prior of a 3D-SSM with the accurate segmentation results of a 3D-CNN. By bridging the gap between statistical shape models and convolutional neural networks, the 3D-SSM could be made more robust in terms of initialization and more accurate in the presence of metal artifacts.

**2D segmentation on C-arm fluoroscopic projections of the uninjured ankle:** A 2D-CNN was employed to create accurate 2D bone segmentations from fluoroscopic images addressing problems like the highly varying viewport, superpositioned bones and the sparse training data.

**Multi-bone 2D-3D reconstruction:** This thesis presents the first 2D-3D reconstruction applied to the ankle joint. A feature-based reconstruction method was introduced based on CNN segmentations of fluoroscopic images. The method optimizes the shape and pose of multiple 3D-SSMs jointly to increase efficiency and robustness. Furthermore, an initialization method was proposed to automatically place the 3D-SSMs close to the final position.

**Visualization:** A new approach was proposed to give a visual representation and guideline for evaluation and correction of fracture reduction. An automatic visual inspection tool was developed that automatically adjusts the viewing planes of the 3D image to clinically important standard planes. In addition, the surgeon is directed towards the region of interest by automatically computing and switching between two evaluation planes. The prototype provides the surgeon with a simultaneous side-by-side and overlay view on the ankle to ease the clinical evaluation.

## 6.8 Future work

The assistance system is a conjunction of different methods and components. Some components may be further evaluated to benefit from possible optimizations:

**Hybrid appearance learning:** Complementing the 3D-SSM with CNN predictions already improved the segmentation performance. However, it would be interesting to combine the information gained from both learned representations. In presence of metal-induced artifacts, the SSM should rather trust the CNN prediction. The intensity variation in these areas is extremely high due to training samples without metal implants. By evaluating the classifier output of the appearance model with the predicted label e.g.

using cross correlation, suspicious areas might be identified during the segmentation. Furthermore, a second CNN targeted at segmenting metal implants in images could be used to directly classify suspicious regions. Alternatively, landmarks with high variance could be tagged and handled separately inspired by the approach of GÖRRES et al. 2016.

**Hybrid intensity and feature-based 2D-3D reconstruction:** For this work, the 2D-3D reconstruction is purely based on the SSM shape and corresponding features extracted from the fluoroscopic images. This has the advantage that the SSM is independent of the modality once the CNN is trained and could be transferred to other anatomies easily. But since intensity-based approaches are assumed to be more accurate than feature-based approaches, the incorporation of intensity values would be beneficial for the fine adaptation of the SSM. One possible hybrid approach would employ intensity-based reconstruction only in the final stage to improve the results in a reasonable time. The idea of assigning different landmarks with weights or tags presented in section 4.4.2 could be also transferred to the reconstruction step. This would allow emphasizing certain regions while reducing the impact of others.

**CNN-based 2D-3D reconstruction:** Recent results from TOTH et al. 2018 showed that 2D-3D registration could be solved using CNNs to regress the transformation parameters. But so far no work has been published on CNN-based 2D-3D reconstruction. Looking into the field of computer vision, KUNDU et al. 2018 showed that a 3D voxel grid could be created using a *Generative Adversarial Network* (GAN) to compare rendered artificial from real 2D images. They claim that even eigenvalues of a PCA on aligned voxel grids could be derived by only analyzing 2D images. An interesting future research question would be if those aforementioned principles could be transferred to the medical domain and applied to projection images to recover the shape parameters of a mesh representation. Compared to photographic images, 3D information cannot be inferred by analyzing the perspective. On the other hand, there is no occlusion but only superposition of pixels that still contain all the information.

**Flat panel detector technology:** 3D C-arm systems used in clinical routine are either based on image intensifiers (e.g. Arcadis<sup>®</sup> Orbic 3D, Siemens Healthcare GmbH) or flat panel detectors (e.g. Cios<sup>®</sup> Spin, Siemens Healthcare GmbH). The latter offers several advantages w.r.t. the system used in this work. Recent improvements in C-arm image quality might further improve the reconstruction quality. Furthermore, the new system offers a motorized adjustment of the C-arm that could allow a standardized acquisition

of standard projections to ensure quality and reproducibility. In section 2.2, differences of both technologies have been briefly introduced. Notable are the elimination of geometric distortion, less noise and inhomogeneities and the possibility to normalize the gray values to resemble HU similar to a CT. This offers a variety of new directions such as the simulation of more 2D training data would be possible to improve the quality of the 2D-CNN. The creation of simulated DRRs from a volume acquired by a C-arm with an image intensifier was not feasible due to image artifacts and blurring. With the next generation of flat panel detectors, a DRR simulation from volume data could create more realistic training images. The 2D-CNN segmentation could greatly benefit from simulated training data that could be further enhanced by employing the DeepDRR generation proposed by UNBERATH et al. 2018. Additionally, the flat panel system would allow developing intensity-based or hybrid reconstruction methods that cannot be implemented yet due to the inhomogeneities and noise of the 3D images.

Also worth mentioning is the FOV of the Cios<sup>®</sup> Spin that is much larger than of the previously used system. If both ankles are placed within the FOV, simultaneous comparison of both ankles could be realized by merely reflection and registration of both images. A direct comparison between both intensity images would be the most reliable method to compare the injured and uninjured ankle. Preliminary experiments on body donors did already indicate that the acquisition of both ankles at the same time is possible with the new system. Since the image intensifier system is still widely installed, the proposed approach remains applicable and might even benefit from the new system by using evaluation data with simultaneous acquisition of both ankles.

The proposed concept has only been applied to images of the ankle joint. However, the human body has a variety of different bilateral bone compounds, for which the concept would only need to be slightly modified in terms of the landmark calculation during 2D-3D initialization. All other methods are not anatomy-specific and allow to generalize. One potential application would be fracture reduction of the wrist and the distal humerus.

Open reduction and internal fixation is a standard procedure in ankle surgery for treating a fractured fibula. Since fibula fractures are often accompanied by an injury of the syndesmosis complex, it is essential to restore the correct relative pose of the fibula relative to the adjoining tibia for the ligaments to heal. Otherwise, the patient might experience instability of the ankle leading to arthritis and ankle pain and ultimately revision surgery. Incorrect positioning referred to as malreduction of the fibula is assumed to be one of the major causes of unsuccessful ankle surgery. 3D C-arm imaging is the current standard procedure for revealing malreduction of fractures in the operating room. However, intra-operative visual inspection of the reduction result is complicated due to high inter-individual variation of the ankle anatomy and rather based on the subjective experience of the surgeon. A contralateral side comparison with the patient's uninjured ankle is recommended but has not been integrated into clinical routine due to the high level of radiation exposure it incurs.

This thesis presents the first approach towards a computer-assisted intra-operative contralateral side comparison of the ankle joint. The focus of this thesis was the design, development and validation of a software-based prototype for a fully automatic intra-operative assistance system for orthopedic surgeons. The implementation does not require an additional 3D C-arm scan of the uninjured ankle, thus reducing time consumption and cumulative radiation dose.

A 3D statistical shape model (SSM) is used to reconstruct a 3D surface model from three 2D fluoroscopic projections representing the uninjured ankle. To this end, a 3D-SSM segmentation is performed on the 3D image of the injured ankle to gain prior knowledge of the ankle. A 3D convolutional neural network (CNN) based initialization method was developed and its outcome was incorporated into the SSM adaption step. Segmen-

tation quality was shown to be improved in terms of accuracy and robustness compared to the pure intensity-based SSM. This allows us to overcome the limitations of the previously proposed methods, namely inaccuracy due to metal artifacts and the lack of device-to-patient orientation of the C-arm.

A 2D-CNN is employed to extract semantic knowledge from all fluoroscopic projection images. This step of the pipeline both creates features for the subsequent reconstruction and also helps to pre-initialize the 3D-SSM without user interaction. A 2D-3D multi-bone reconstruction method has been developed which uses distance maps of the 2D features for fast and accurate correspondence optimization and SSM adaptation. This is the central and most crucial component of the workflow. This is the first time that a bone reconstruction method has been applied to the complex ankle joint and the first reconstruction method using CNN based segmentations as features. The reconstructed 3D-SSM of the uninjured ankle can be back-projected and visualized in a workflow-oriented manner to procure clear visualization of the region of interest, which is essential for the evaluation of the reduction result. The surgeon can thus directly compare an overlay of the contralateral ankle with the injured ankle.

The developed methods were evaluated individually using data sets acquired during a cadaver study and representative clinical data acquired during fibular reduction. A hierarchical evaluation was designed to assess the inaccuracies of the system on different levels and to identify major sources of error. The overall evaluation performed on eleven challenging clinical data sets acquired for manual contralateral side comparison showed that the system is capable of accurately reconstructing 3D surface models of the uninjured ankle solely using three projection images. A mean Hausdorff distance of 1.72 mm was measured when comparing the reconstruction result to the ground truth segmentation and almost achieved the high required clinical accuracy of 1-2 mm. The overall error of the pipeline was mainly attributed to inaccuracies in the 2D-CNN segmentation. The consistency of these results requires further validation on a larger data set.

The workflow proposed in this thesis establishes the first approach to enable automatic computer-assisted contralateral side comparison in ankle surgery. The feasibility of the proposed approach was proven on a limited amount of clinical cases and has already yielded good results. The next important step is to alleviate the identified bottlenecks in the approach by providing more training data in order to further improve the accuracy. In conclusion, the new approach presented gives the chance to guide the surgeon during the reduction process, improve the surgical outcome while avoiding additional radiation exposure and reduce the number of revision surgeries in the long term.



Die offene Reposition und die anschließende osteosynthetische Versorgung gelten als Standardverfahren zur operativen Behandlung von Sprunggelenksfrakturen. Da diese Frakturen oft mit einer Verletzung der Syndesmose einhergehen, ist es unerlässlich, die korrekte Lage zwischen Fibula und Tibia operativ wiederherzustellen. Eine Fehlstellung der Fibula kann zu einer Instabilität der Syndesmose führen und damit einhergehend eine frühzeitige Arthritis und Schmerzen am Sprunggelenk begünstigen oder gar eine Revision der Operation erfordern.

Der Einsatz intra-operativer 3D C-Bögen ermöglicht eine räumliche Interpretation und umfassende Analyse des Repositionsergebnisses. Die objektive Beurteilung wird jedoch durch die hohe inter-individuelle Varianz der Form und Lage der am Sprunggelenk beteiligten Knochen erschwert. Ein kontralateraler Seitenvergleich mit dem unverletzten Sprunggelenk des Patienten wird zwar empfohlen, aber aufgrund der zusätzlichen Strahlenbelastung und dem erhöhten Zeitaufwand bisher nicht in der klinischen Routine eingesetzt. Ziel dieser Arbeit war es, Methoden für ein Computer-gestütztes intra-operatives Assistenzsystem zum kontralateralen Seitenvergleich des Sprunggelenks zu entwickeln. Die vorliegende Arbeit umfasst das Design, die Entwicklung und die Validierung eines softwarebasierten automatisierten Prototyps. Der entwickelte Ansatz erfordert keinen zusätzlichen 3D C-Bogen Scan des unverletzten Sprunggelenks, um sowohl die Zeit des Eingriffs als auch die Strahlenbelastung zu reduzieren.

Um ein 3D Oberflächenmodell des gesunden Sprunggelenks aus drei 2D Projektionsbildern zu rekonstruieren, wird ein 3D Statistisches Formmodell (SSM) verwendet. Zu diesem Zweck wird zunächst eine 3D-SSM Segmentierung des 3D Volumens des verletzten Sprunggelenks durchgeführt, um eine gute erste Annäherung des zu rekonstruierenden Modells zu erhalten. Die 3D-SSM Segmentierung wird mit einem 3D Convo-

lutional Neural Network (CNN) kombiniert, welches zum einen zur Initialisierung des Formmodells und zum anderen zur Anpassung während der Segmentierung eingesetzt wird. Somit können zwei wesentliche Herausforderungen des konventionellen 3D-SSMs, die Anfälligkeit gegenüber Metallartefakten und Ungenauigkeiten aufgrund einer ungünstigen Initialisierung, gelöst werden. Die Qualität der Segmentierung kann so im Vergleich zu einer rein auf Bildintensitäten basierenden Segmentierung verbessert werden.

Ein 2D-CNN wird verwendet, um Konturen und Landmarken aus allen 2D Projektionsbildern zu extrahieren. Dieser Schritt ermöglicht eine automatische Initialisierung des Formmodells und generiert gleichzeitig Landmarken für den nachfolgenden Rekonstruktionsschritt. Es wurde eine 2D-3D Rekonstruktionsmethode entwickelt, welche aus den 2D Landmarken generierte Distanzkarten verwendet, um eine simultane, schnelle und genaue Korrespondenzoptimierung und Formanpassung aller beteiligten Knochen zu ermöglichen. Die Rekonstruktion ist der zentralste und wichtigste Schritt des vorgestellten Ansatzes. Dies ist das erste Mal, dass eine Rekonstruktionsmethode auf das komplexe Sprunggelenk angewendet wird. Die Arbeit stellt gleichzeitig die erste Rekonstruktionsmethode vor, die CNN-basierte Segmentierungen in den Rekonstruktionsschritt integriert. Das rekonstruierte Formmodell des unverletzten Sprunggelenks kann nach der Rekonstruktion rückprojiziert und mit dem frakturierten Sprunggelenk überlagert dargestellt werden. Dieser Schritt ermöglicht einen direkten Vergleich des kontralateralen und des frakturierten Sprunggelenks, um eine mögliche Fehlstellung der Fibula zu identifizieren.

Die entwickelten Methoden wurden mit Hilfe von Datensätzen aus einer Kadaverstudie und repräsentativen klinischen Daten ausgewertet. Eine systematische Evaluation wurde durchgeführt, um die Ungenauigkeiten des Systems und seiner Teilkomponenten zu bewerten und wesentliche Fehlerquellen zu identifizieren. Eine Auswertung des Gesamtsystems wurde auf elf Datensätzen durchgeführt, die im Rahmen eines manuellen kontralateralen Seitenvergleichs in der klinischen Routine akquiriert wurden. Die Auswertung zeigte, dass das System in der Lage ist, 3D-Oberflächenmodelle des unverletzten Sprunggelenks aus drei Projektionsbildern zu rekonstruieren. Das System zeigte vielversprechende Ergebnisse auf den verfügbaren Daten. Beim Vergleich des Rekonstruktionsergebnisses mit der Referenzsegmentierung wurde ein mittlerer Hausdorffabstand von 1,72 mm gemessen, welcher die geforderte hohe klinische Genauigkeit, die zwischen 1 - 2 mm angegeben wird, bereits fast erreicht. Der Gesamtfehler konnte hauptsächlich auf Ungenauigkeiten in der 2D-CNN-Segmentierung zurückgeführt werden.

Die in dieser Arbeit entwickelten Methoden stellen den ersten Ansatz für den automa-

tischen Computer-gestützten kontralateralen Seitenvergleich in der Sprunggelenkschirurgie dar. Es konnte gezeigt werden, dass eine gute Annäherung der Oberfläche mit nur drei kontralateralen Projektionsbildern erreicht werden kann. Die Realisierbarkeit des vorgeschlagenen Ansatzes wurde an einer begrenzten Anzahl von klinischen Fällen nachgewiesen und hat bereits zu vielversprechenden Ergebnissen geführt. Der nächste wichtige Schritt besteht darin, die identifizierten Fehlerquellen zu reduzieren. Durch die Erweiterung der Trainingsdatenmenge und der Auswertung auf einer größeren Datenmenge soll das Verfahren in Zukunft weiter verbessert und validiert werden.

Zusammenfassend lässt sich sagen, dass der vorgestellte neue Ansatz die Möglichkeit bieten kann, den Chirurgen während des Reduktionsprozesses zu unterstützen und gleichzeitig eine zusätzliche Strahlenbelastung zu vermeiden mit dem Ziel die Qualität der Operation zu verbessern und die Anzahl der Revisionseingriffe langfristig zu reduzieren.



- ADAMS, R. and L. BISCHOF (1994). “Seeded region growing”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 16.6, pp. 641–647. DOI: [10.1109/34.295913](https://doi.org/10.1109/34.295913).
- AICHINGER, H., S. JOITE-BARFUSS, J. DIERKER, and M. SAEBEL (2012). “Principles of X-Ray Imaging”. In: *Radiation Exposure and Image Quality in X-Ray Diagnostic Radiology: Physical Principles and Clinical Applications*. Ed. by H. SAEBEL, J. AICHINGER, S. DIERKER, and M. JOITE-BARFUSS. 1st ed. Springer-Verlag Berlin Heidelberg, pp. 3–7. DOI: [10.1007/978-3-642-11241-6](https://doi.org/10.1007/978-3-642-11241-6).
- AL ARIF, S. M. R., K. KNAPP, and G. SLABAUGH (2018). “SPNet: Shape prediction using a fully convolutional neural network”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11070 LNCS, pp. 430–439. DOI: [10.1007/978-3-030-00928-1\\_49](https://doi.org/10.1007/978-3-030-00928-1_49).
- ANAS, E. M. A., A. RASOULIAN, A. SEITEL, K. DARRAS, D. WILSON, P. S. JOHN, D. PICHORA, P. MOUSAVI, R. ROHLING, and P. ABOLMAESUMI (2016). “Automatic Segmentation of Wrist Bones in CT Using a Statistical Wrist Shape + Pose Model”. In: *IEEE Transactions on Medical Imaging* 35.8, pp. 1789–1801. DOI: [10.1109/TMI.2016.2529500](https://doi.org/10.1109/TMI.2016.2529500).
- AUBERT, B., C. VERGARI, B. ILHARREBORDE, A. COURVOISIER, and W. SKALLI (2016). “3D reconstruction of rib cage geometry from biplanar radiographs using a statistical parametric model approach”. In: *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization* 4.5, pp. 281–295. DOI: [10.1080/21681163.2014.913990](https://doi.org/10.1080/21681163.2014.913990).
- AVENDI, M. R., A. KHERADVAR, and H. JAFARKHANI (2016). “A combined deep-learning and deformable-model approach to fully automatic segmentation of the left ventricle in cardiac MRI”. In: *IEEE Transactions on Medical Image Analysis* 30, pp. 108–119. DOI: [10.1016/j.media.2016.01.005](https://doi.org/10.1016/j.media.2016.01.005).

- BAKA, N., M. DE BRUIJNE, T. VAN WALSUM, B. L. KAPTEIN, J. E. GIPHART, M. SCHAAP, W. J. NIESSEN, and B. P. LELIEVELDT (2012). “Statistical shape model-based femur kinematics from biplane fluoroscopy”. In: *IEEE Transactions on Medical Imaging* 31.8, pp. 1573–1583. DOI: [10.1109/TMI.2012.2195783](https://doi.org/10.1109/TMI.2012.2195783).
- BARROW, H. G., J. M. TENENBAUM, R. C. BOLLES, and H. C. WOLF (1977). “Parametric Correspondence and Chamfer Matching: Two New Techniques for Image Matching”. In: *Proceedings of the 5th International Joint Conference on Artificial Intelligence - Volume 2*. IJCAI’77. San Francisco, CA, USA: Morgan Kaufmann Publishers Inc., pp. 659–663.
- BENAMEUR, S., M. MIGNOTTE, S. PARENT, H. LABELLE, W. SKALLI, and J. DE GUISE (2003). “3D/2D registration and segmentation of scoliotic vertebrae using statistical models”. In: *Computerized Medical Imaging and Graphics* 27.5, pp. 321–337. DOI: [10.1016/S0895-6111\(03\)00019-3](https://doi.org/10.1016/S0895-6111(03)00019-3).
- BERGER, M., Q. YANG, and A. MAIER (2018). “X-Ray Imaging”. In: *Medical imaging systems - An introductory guide*. Ed. by A. MAIER, S. STEIDL, V. CHRISTLEIN, and J. HORNEGGER. 1st ed. Springer International Publishing, pp. 119–146. DOI: [10.1063/1.2914691](https://doi.org/10.1063/1.2914691).
- BEUMER, A., W. L. W. V. HEMERT, R. NIESING, C. A. C. ENTIUS, A. Z. GINAI, P. G. H. MULDER, and B. A. SWIERSTRA (2004). “Radiographic Measurement of the Distal Tibiofibular Syndesmosis Has Limited Use”. In: *Clinical Orthopaedics and Related Research* 423, pp. 227–234. DOI: [10.1097/01.blo.0000129152.81015.ad](https://doi.org/10.1097/01.blo.0000129152.81015.ad).
- BHALODIA, R., S. Y. ELHABIAN, L. KAVAN, and R. T. WHITAKER (2018). “DeepSSM: A Deep Learning Framework for Statistical Shape Modeling from Raw Images”. In: DOI: [10.1007/978-3-030-04747-4\\_23](https://doi.org/10.1007/978-3-030-04747-4_23).
- BOHM, A., A. UCKER, T. JAGER, O. RONNEBERGER, and T. FALK (2018). “ISOODL: Instance segmentation of overlapping biological objects using deep learning”. In: *Proceedings - International Symposium on Biomedical Imaging*, pp. 1225–1229. DOI: [10.1109/ISBI.2018.8363792](https://doi.org/10.1109/ISBI.2018.8363792).
- BOISVERT, J., F. CHERIET, X. PENNEC, H. LABELLE, and N. AYACHE (2008). “Articulated spine models for 3-D reconstruction from partial radiographic data”. In: *IEEE Transactions on Biomedical Engineering* 55.11, pp. 2565–2574. DOI: [10.1109/TBME.2008.2001125](https://doi.org/10.1109/TBME.2008.2001125).
- BONARETTI, S., C. SEILER, C. BOICHON, M. REYES, and P. BÜCHLER (2014). “Medical Engineering & Physics Image-based vs . mesh-based statistical appearance models of the human femur : Implications for finite element simulations”. In: *Medical Engineering and Physics* 36.12, pp. 1626–1635. DOI: [10.1016/j.medengphy.2014.09.006](https://doi.org/10.1016/j.medengphy.2014.09.006).
- BOUSSAID, H., S. KADOURY, I. KOKKINOS, J.-y. LAZENNEC, G. ZHENG, and N. PARAGIOS (2011). “3D Model-based Reconstruction of the Proximal Femur from Low-dose Bi-

- planar X-Ray Images”. In: *Proceedings of the British Machine Vision Conference 2011*, pp. 1–10. DOI: [10.5244/C.25.35](https://doi.org/10.5244/C.25.35).
- BOYKOV, Y. and V. KOLMOGOROV (2004). “An Experimental Comparison of Min-Cut / Max-Flow Algorithms for Energy Minimization in Vision”. In: *Transactions on Pattern Analysis and Machine Intelligence* 26.9, pp. 1124–1137.
- BOZEK, J., M. MUSTRA, K. DELAC, and M. GRGIC (2009). “A survey of image processing algorithms in digital mammography”. In: *Studies in Computational Intelligence* 231, pp. 631–657. DOI: [10.1007/978-3-642-02900-4\\_24](https://doi.org/10.1007/978-3-642-02900-4_24).
- BREHLER, M., J. GÖRRES, S. Y. VETTER, J. FRANKE, P. A. GRÜTZNER, and H.-p. M. IVO (2015). “Intra-operative assessment of fractured articular surfaces in cone beam CT image data”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.4, pp. 603–612. DOI: [10.1007/s11548-015-1304-0](https://doi.org/10.1007/s11548-015-1304-0).
- BREJL, M. and M. SONKA (2000). “Object localization and border detection criteria design in edge-based image segmentation: Automated learning from examples”. In: *IEEE Transactions on Medical Imaging* 19.10, pp. 973–985. DOI: [10.1109/42.887613](https://doi.org/10.1109/42.887613).
- BRUIJNE, M. D. and M. NIELSEN (2004). “Shape Particle Filtering for Image Segmentation”. In: *Proceedings - International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 168–175.
- BUCKLEY, J., H. K. HAN, M. MARMOR, E. HANSEN, and A. MATITYAHU (2011). “Limitations of Standard Fluoroscopy in Detecting Rotational Malreduction of the Syndesmosis in an Ankle Fracture Model”. In: *Foot & Ankle International* 32.6, pp. 616–622. DOI: [10.3113/fai.2011.0616](https://doi.org/10.3113/fai.2011.0616).
- CANNY, J. (1986). “A Computational Approach To Edge Detection”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 8.6, pp. 679–714.
- CASSAR-PULLICINO, V.N. (1998). “Computer Tomography”. In: *Orthopedic Imaging - Medical Radiology*. Ed. by A. M. DAVIES and H. PETERSSON. Berlin, Heidelberg: Springer-Verlag. DOI: [10.1007/978-3-642-60295-5](https://doi.org/10.1007/978-3-642-60295-5).
- CERNAZANU-GLAVAN, C. and S. HOLBAN (2013). “Segmentation of bone structure in X-ray images using convolutional neural network”. In: *Advances in Electrical and Computer Engineering* 13.1, pp. 87–94. DOI: [10.4316/AECE.2013.01015](https://doi.org/10.4316/AECE.2013.01015).
- CERVERI, P., C. SACCO, G. OLGATI, A. MANZOTTI, and G. BARONI (2017). “2D/3D reconstruction of the distal femur using statistical shape models addressing personalized surgical instruments in knee arthroplasty: A feasibility analysis”. In: *International Journal of Medical Robotics and Computer Assisted Surgery* 13.4, pp. 1–13. DOI: [10.1002/rcs.1823](https://doi.org/10.1002/rcs.1823).

- CHAIBI, Y., T. CRESSON, L. HUMBERT, J. A. DE GUISE, and W. SKALLI (2012). “Fast 3D reconstruction of the lower limb using a parametric model and statistical inferences and clinical measurements calculation from biplanar X-rays”. In: *Computer Methods in Biomechanics and Biomedical Engineering* 15.5, pp. 457–466. DOI: [10.1080/10255842.2010.540758](https://doi.org/10.1080/10255842.2010.540758).
- CHEN, C. and G. ZHENG (2013). “Fully Automatic Segmentation of AP Pelvis X-rays via Random Forest Regression”. In: *International Conference on Computer Analysis of Images and Patterns*, pp. 335–343. DOI: [10.1007/978-3-642-40261-6\\_40](https://doi.org/10.1007/978-3-642-40261-6_40).
- ÇIÇEK, A., A. ABDULKADIR, S. S. LIENKAMP, T. BROX, and O. RONNEBERGER (2016). “3D U-Net : Learning Dense Volumetric Segmentation from Sparse Annotation”. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 424–432. DOI: [10.1007/978-3-319-46723-8\\_49](https://doi.org/10.1007/978-3-319-46723-8_49).
- CLEVERT, D.-A., T. UNTERTHINER, and S. HOCHREITER (2015). “Fast and Accurate Deep Network Learning by Exponential Linear Units (ELUs)”. In: *Arxiv*, pp. 1–14. URL: <http://arxiv.org/abs/1511.07289>.
- COOTES, T. F., C. J. TAYLOR, D. H. COOPER, and J. GRAHAM (1995). “Active Shape Models - Their Training and Application”. In: *Computer Vision and Image Understanding* 61.1, pp. 38–59. DOI: [10.1006/cviu.1995.1004](https://doi.org/10.1006/cviu.1995.1004).
- DAVIES, A. M., R. W. WHITEHOUSE, and J. P. R. JENKINS (2012). *Imaging of the Foot & Ankle: Techniques and Applications*. 1st ed. Medical Radiology. Springer Berlin Heidelberg.
- DAVIES, R. H., C. J. TWINING, T. F. COOTES, J. C. WATERTON, and C. J. TAYLOR (2002). “A Minimum Description Length Approach to Statistical Shape Modeling”. In: *IEEE Transactions on Medical Imaging* 21.5, pp. 525–537.
- DE SILVA, T., A. UNERI, M. KETCHA, S. REAUNGAMORNARAT, G. KLEINSZIG, S. VOGT, N. AYGUN, S.-F. LO, and J.-P. WOLINSKY (2016). “3D-2D image registration for target localization in spine surgery: investigation of similarity metrics providing robustness to content mismatch”. In: *Physics in medicine and biology* 61.8, pp. 3009–3025. DOI: [10.1088/0031-9155/61/8/3009](https://doi.org/10.1088/0031-9155/61/8/3009).
- DICE, L. R. (1945). “Measures of the Amount of Ecologic Association Between Species”. In: *Ecology* 26.3, pp. 297–302. DOI: [10.2307/1932409](https://doi.org/10.2307/1932409).
- DIKOS, G. D., J. HEISLER, R. H. CHOPLIN, and T. G. WEBER (2012). “Normal tibiofibular relationships at the syndesmosis on axial CT imaging”. In: *Journal of Orthopaedic Trauma* 26.7, pp. 433–438. DOI: [10.1097/BOT.0b013e3182535f30](https://doi.org/10.1097/BOT.0b013e3182535f30).
- DONG, X. and G. ZHENG (2008). “Automatic extraction of femur contours from calibrated x-ray images: A Bayesian inference approach”. In: *Proceedings - 5th IEEE International*



- Symposium on Biomedical Imaging: From Nano to Macro*, pp. 57–60. DOI: [10.1109/ISBI.2008.4540931](https://doi.org/10.1109/ISBI.2008.4540931).
- DUDA, R. O. and P. E. HART (1972). “Use of the Hough Transformation to Detect Lines and Curves in Pictures”. In: *Communications of the ACM* 15.1, pp. 11–15. DOI: [10.1145/361237.361242](https://doi.org/10.1145/361237.361242).
- EBRAHEIM, N. A. and H. ELGAFY (2003). “Syndesmotoc Disruption in Low Fibular Fractures Associated With Deltoid”. In: *Clinical Orthopaedics and Related Research* 409, pp. 260–267. DOI: [10.1097/01.blo.0000052935.71325.30](https://doi.org/10.1097/01.blo.0000052935.71325.30).
- EBRAHEIM, N. A., J. LU, H. YANG, A. O. MEKHAIL, and R. A. YEASTING (1997). “Radio-graphic and CT Evaluation of Tibiofibular Syndesmotoc Diastasis : A Cadaver Study”. In: *Foot & Ankle International* 18.11, pp. 693–698.
- EBRAHEIM, N. A., J. LU, H. YANG, and J. ROLLINS (1998). “The Fibular Incisure of the Tibia on CT Scan : A Cadaver Study”. In: *Foot & Ankle International* 19.5, pp. 318–321.
- ECABERT, O. and J. P. THIRAN (2004). “Adaptive Hough transform for the detection of natural shapes under weak affine transformations”. In: *Pattern Recognition Letters* 25.12, pp. 1411–1419. DOI: [10.1016/j.patrec.2004.05.009](https://doi.org/10.1016/j.patrec.2004.05.009).
- EHLKE, M., M. HEYLAND, S. MÄRDIAN, G. N. DUDA, and S. ZACHOW (2015). “Assessing the relative positioning of an osteosynthesis plate to the patient-specific femoral shape from plain 2D radiographs”. In: *Proceedings of the 15th Annual Meeting of CAOS-International (CAOS)*.
- EHLKE, M., H. RAMM, H. LAMECKER, H.-C. HEGE, and S. ZACHOW (2013). “Fast Generation of Virtual X-ray Images from Deformable Tetrahedral Meshes”. In: *IEEE Transactions on Visualization and Computer Graphics* 19.12, pp. 2673–2682.
- ELGAFY, H., H. B. SEMAAN, B. BLESSINGER, A. WASSEF, and N. A. EBRAHEIM (2010). “Computed tomography of normal distal tibiofibular syndesmosis”. In: *Skeletal Radiology* 39.6, pp. 559–564. DOI: [10.1007/s00256-009-0809-4](https://doi.org/10.1007/s00256-009-0809-4).
- FAWCETT, T. (2006). “An introduction to ROC analysis”. In: *Pattern Recognition Letters* 27.8, pp. 861–874. DOI: [10.1016/j.patrec.2005.10.010](https://doi.org/10.1016/j.patrec.2005.10.010).
- FLEUTE, M. and S. LAVALLÉE (1999). “Nonrigid 3-D/2-D Registration of Images Using Statistical Models”. In: *Proceedings - International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 138–147. DOI: [10.1007/10704282\\_15](https://doi.org/10.1007/10704282_15).
- FOTOUHI, J., M. UNBERATH, T. SONG, W. GU, A. JOHNSON, G. OSGOOD, M. ARMAND, and N. NAVAB (2019). “Interactive Flying Frustums ( IFFs ): spatially aware surgical data visualization”. In: *International Journal of Computer Assisted Radiology and Surgery* 14.6, pp. 913–922. DOI: [10.1007/s11548-019-01943-z](https://doi.org/10.1007/s11548-019-01943-z).

- FRANKE, J. and P. A. GRÜTZNER (2013). “3D Navigation with a Mobile C-Arm”. In: *Computer and Template Assisted Orthopedic Surgery*. Ed. by R. HAAKER and W. KONERMANN, pp. 153–160.
- FRANKE, J., J. V. RECUM, A. J. SUDA, P. A. GR, and K. WENDL (2012). “Intraoperative three-dimensional imaging in the treatment of acute unstable syndesmotic injuries”. In: *The Journal of bone and joint surgery. American volume* 94.15, pp. 1386–1390. DOI: [10.1029/2008JB006151](https://doi.org/10.1029/2008JB006151).
- FRIPP, J., S. WARFIELD, S. CROZIER, and S. OURSELIN (2006). “Automatic Segmentation of the Knee Bones using 3D Active Shape Models”. In: *Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, pp. 167–170. DOI: [10.1109/icpr.2006.306](https://doi.org/10.1109/icpr.2006.306).
- FÜRNSTAHL, P., G. SZÉKELY, C. GERBER, J. HODLER, J. G. SNEDEKER, and M. HARDERS (2012). “Computer assisted reconstruction of complex proximal humerus fractures for preoperative planning”. In: *IEEE Transactions on Medical Image Analysis* 16.3, pp. 704–720. DOI: [10.1016/j.media.2010.07.012](https://doi.org/10.1016/j.media.2010.07.012).
- GARDNER, M. J., D. DEMETRAKOPOULOS, S. M. BRIGGS, D. L. HELFET, and D. G. LORICH (2006). “Malreduction of the tibiofibular syndesmosis in ankle fractures”. In: *Foot and Ankle International* 27.10, pp. 788–792. DOI: [10.1177/107110070602701005](https://doi.org/10.1177/107110070602701005).
- GERON, A. (2017). *Hands-On Machine Learning with Scikit-Learn and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. Ed. by N. TACHE, N. ADAMS, and R. MONAGHAN. 1st. O’Reilly Media, Inc., p. 261.
- GONG, R. H. (2011). “Two-Dimensional-Three-Dimensional Registration Methods for Computer-Assisted Orthopaedic Surgery”. PhD thesis. Queen’s University Kingston Ontario.
- GÖRRES, J., M. BREHLER, J. FRANKE, S. Y. VETTER, P. A. GRÜTZNER, H. P. MEINZER, and I. WOLF (2016). “Articular surface segmentation using active shape models for intraoperative implant assessment”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.9, pp. 1661–1672. DOI: [10.1007/s11548-015-1316-9](https://doi.org/10.1007/s11548-015-1316-9).
- GOWER, J. C. (1975). “Generalized procrustes analysis”. In: *Psychometrika* 40.1, pp. 33–51. DOI: [10.1007/BF02291478](https://doi.org/10.1007/BF02291478).
- GREENSPAN, H., B. van GINNEKEN, and R. M. SUMMERS (2016). “Guest Editorial Deep Learning in Medical Imaging : Overview and Future Promise of an Exciting New Technique”. In: *IEEE Transactions on Medical Imaging* 35.5, pp. 1153–1159. DOI: [10.1109/TMI.2016.2553401](https://doi.org/10.1109/TMI.2016.2553401).

- HAHN, D. M. and C. L. COLTON (2018). “Malleoli”. In: *AO Principles of Fracture Management - Vol. 2: Specific fractures*. Ed. by R. E. BUCKLEY, C. G. MORAN, and T. APIVATTHAKAKUL. 3rd ed. Thieme Verlagsgruppe, Stuttgart, New York, Delhi, Rio. Chap. 6.9.
- HANDELS, H. (2009). *Medizinische Bildverarbeitung: Bildanalyse, Mustererkennung und Visualisierung für computergestützte ärztliche Diagnostik und Therapie*. 2nd ed. Vieweg+Teubner Verlag. DOI: [10.1007/978-3-8348-9571-4](https://doi.org/10.1007/978-3-8348-9571-4).
- HANSEN, N. (2006). “The CMA evolution strategy: a comparing review”. In: *Towards a New Evolutionary Computation. Studies in Fuzziness and Soft Computing*. Ed. by J. LOZANO, P. LARRAÑAGA, I. INZA, and E. BENGOTXEA. 192nd ed. Springer Berlin, Heidelberg, pp. 75–102. DOI: [10.1007/11007937\\_4](https://doi.org/10.1007/11007937_4).
- HARPER, M. C., T. S. KELLER, and D. PH (1989). “A Radiographic Evaluation of the Tibiofibular Syndesmosis”. In: *Foot & Ankle* 10.3, pp. 156–160.
- HEIMANN, T. (2009a). *Statistical Shape Models for 3D Medical Image Segmentation*. Saarbruecken, Germany, Germany: VDM Verlag.
- (2009b). “Statistical shape models for 3d medical image segmentation”. PhD thesis. Universität Heidelberg.
- HEIMANN, T. and H. P. MEINZER (2009). “Statistical shape models for 3D medical image segmentation: A review”. In: *IEEE Transactions on Medical Image Analysis* 13.4, pp. 543–563. DOI: [10.1016/j.media.2009.05.004](https://doi.org/10.1016/j.media.2009.05.004).
- HEIMANN, T., M. SASCHA, H.-p. MEINZER, and I. WOLF (2007). “A Shape-Guided Deformable Model with Evolutionary Algorithm Initialization for 3D Soft Tissue A Shape-Guided Deformable Model with Evolutionary Algorithm Initialization for 3D Soft Tissue Segmentation”. In: *Information Processing in Medical Imaging*, pp. 1–12. DOI: [10.1007/978-3-540-73273-0](https://doi.org/10.1007/978-3-540-73273-0).
- HEIMANN, T., B. VAN GINNEKEN, M. A. STYNER, Y. ARZHAeva, V. AURICH, C. BAUER, A. BECK, C. BECKER, R. BEICHEL, G. BEKES, F. BELLO, G. BINNIG, H. BISCHOF, A. BORNIK, P. M. CASHMAN, Y. CHI, A. CÓRDOVA, B. M. DAWANT, M. FIDRICH, J. D. FURST, D. FURUKAWA, L. GRENACHER, J. HORNEGGER, D. KAINMÜLLER, R. I. KITNEY, H. KOBATAKE, H. LAMECKER, T. LANGE, J. LEE, B. LENNON, R. LI, S. LI, H. P. MEINZER, G. NÉMETH, D. S. RAICU, A. M. RAU, E. M. VAN RIKXOORT, M. ROUSSON, L. RUSKÓ, K. A. SADDI, G. SCHMIDT, D. SEGHERS, A. SHIMIZU, P. SLAGMOLEN, E. SORANTIN, G. SOZA, R. SUSOMBOON, J. M. WAITE, A. WIMMER, and I. WOLF (2009). “Comparison and evaluation of methods for liver segmentation from CT datasets”. In: *IEEE Transactions on Medical Imaging* 28.8, pp. 1251–1265. DOI: [10.1109/TMI.2009.2013851](https://doi.org/10.1109/TMI.2009.2013851).

- HEIMANN, T., I. WOLF, T. WILLIAMS, and H.-P. MEINZER (2005). “3D Active Shape Models using gradient descent optimization of description length”. In: *Lecture Notes in Computer Science* 3565, pp. 566–577.
- HERMANS, J. J., A. BEUMER, T. A. W. DE JONG, and G. J. KLEINRENSINK (2010). “Anatomy of the distal tibiofibular syndesmosis in adults: A pictorial essay with a multimodality approach”. In: *Journal of Anatomy* 217.6, pp. 633–645. DOI: [10.1111/j.1469-7580.2010.01302.x](https://doi.org/10.1111/j.1469-7580.2010.01302.x).
- HOMBACH-KLONISCH, S., T. KLONISCH, and J. PEELER (2019). *Sobotta Clinical Atlas of Human Anatomy, one volume, English*. Elsevier Health Sciences.
- HSIEH, J. (2003). *Computed Tomography: Principles, Design, Artifacts, and Recent Advances*. 1st ed. SPIE Press monograph. SPIE Press.
- HUTTENLOCHER, D. P., G. A. KLANDERMAN, and W. J. RUCKLIDGE (1993). “Comparing Images Using the Hausdorff Distance”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.9, pp. 850–863. DOI: [10.1109/34.232073](https://doi.org/10.1109/34.232073).
- ISENSEE, F., J. PETERSEN, A. KLEIN, D. ZIMMERER, P. F. JAEGER, S. KOHL, J. WASSERTHAL, K. GREGOR, T. NORAJITRA, and C. V. SEP (2019). “nnU-Net: Self-adapting Framework for U-Net-Based Medical Image Segmentation”. In: URL: <https://arxiv.org/abs/1809.10486>.
- JACCARD, P. (1912). “The distribution of the flora in the alpine zone.1”. In: *New Phytologist* 11.2, pp. 37–50. DOI: [10.1111/j.1469-8137.1912.tb05611.x](https://doi.org/10.1111/j.1469-8137.1912.tb05611.x).
- JIANHUA YAO and TAYLOR (2003). “Assessing accuracy factors in deformable 2D/3D medical image registration using a statistical pelvis model”. In: *Proceedings Ninth IEEE International Conference on Computer Vision*. Vol. 2, pp. 1329–1334. DOI: [10.1109/ICCV.2003.1238644](https://doi.org/10.1109/ICCV.2003.1238644).
- KADOURY, S., F. CHERIET, S. KADOURY, F. CHERIET, and H. LABELLE (2009). “Personalized X-Ray 3-D Reconstruction of the Scoliotic Spine From Hybrid Statistical and Image-Based Models”. In: *IEEE Transactions on Medical Imaging* 28.9, pp. 1422–1435. DOI: [10.1109/TMI.2009.2016756](https://doi.org/10.1109/TMI.2009.2016756).
- KAINMUELLER, D., H. LAMECKER, S. ZACHOW, M. HELLER, and H.-c. HEGE (2008). “Multi-object Segmentation with Coupled Deformable Models”. In: *Proceedings of Medical Image Understanding and Analysis*, pp. 34–38.
- KALYANI, B. S., C. S. ROBERTS, and P. V. GIANNOUDIS (2010). “The Maisonneuve Injury: A Comprehensive Review”. In: *Orthopedics* 33.3, pp. 190–195. DOI: [10.3928/01477447-20100129-25](https://doi.org/10.3928/01477447-20100129-25).
- KARADE, V. and B. RAVI (2015). “3D femur model reconstruction from biplane X-ray images: a novel method based on Laplacian surface deformation”. In: *International Jour-*

- nal of Computer Assisted Radiology and Surgery* 10.4, pp. 473–485. DOI: [10.1007/s11548-014-1097-6](https://doi.org/10.1007/s11548-014-1097-6).
- KIMPE, T. and T. TUYTSCHAEVER (2007). “Increasing the number of gray shades in medical display systems - How much is enough?” In: *Journal of Digital Imaging* 20.4, pp. 422–432. DOI: [10.1007/s10278-006-1052-3](https://doi.org/10.1007/s10278-006-1052-3).
- KINGMA, D. P. and J. L. BA (2015). “Adam: A Method for Stochastic Optimization”. In: *3rd International Conference for Learning Representations*. San Diego, pp. 1–15.
- KLEIN, A., J. WARSZAWSKI, J. HILLEGASS, and K. H. MAIER-HEIN (2019). “Automatic bone segmentation in whole-body CT images”. In: *International Journal of Computer Assisted Radiology and Surgery* 14.1, pp. 21–29. DOI: [10.1007/s11548-018-1883-7](https://doi.org/10.1007/s11548-018-1883-7).
- KLIMA, O., P. KLEPARNIK, M. SPANEL, and P. ZEMCIK (2016). “Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation”. In: *Medical Imaging 2016: Biomedical Applications in Molecular, Structural, and Functional Imaging*. DOI: [10.1117/12.2216529](https://doi.org/10.1117/12.2216529).
- KOBATAKE, H. and Y. MASUTANI (2017). *Computational Anatomy Based on Whole Body Imaging: Basic Principles of Computer-Assisted Diagnosis and Therapy*. Springer Japan, p. 137.
- KOHL, S. A. A., B. ROMERA-PAREDES, C. MEYER, J. DE FAUW, J. R. LEDSAM, K. H. MAIER-HEIN, S. M. A. ESLAMI, D. J. REZENDE, and O. RONNEBERGER (2018). “A Probabilistic U-Net for Segmentation of Ambiguous Images”. In: *NeurIPS*, pp. 1–28. URL: <https://arxiv.org/abs/1806.05034>.
- KORDON, F., R. LASOWSKI, B. SWARTMAN, J. FRANKE, P. FISCHER, and H. KUNZE (2019). “Improved X-Ray Bone Segmentation by Normalization and Augmentation Strategies”. In: *Informatik aktuell - Bildverarbeitung für die Medizin*, pp. 104–109. DOI: [10.1007/978-3-658-25326-4\\_24](https://doi.org/10.1007/978-3-658-25326-4_24).
- KOREZ, R., B. LIKAR, F. PERNUS, and T. VRTOVEC (2016). “Model-Based Segmentation of Vertebral Bodies from MR Images with 3D CNNs”. In: *Proceedings - Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pp. 433–440. DOI: [10.1007/978-3-319-46723-8](https://doi.org/10.1007/978-3-319-46723-8).
- KUNDU, A., Y. LI, and J. M. REHG (2018). “3D-RCNN: Instance-level 3D Object Reconstruction via Render-and-Compare”. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3559–3568. DOI: [10.1109/CVPR.2018.00375](https://doi.org/10.1109/CVPR.2018.00375).
- LANGTON, C. M., S. PISHARODY, and J. H. KEYAK (2009). “Generation of a 3D proximal femur shape from a single projection 2D radiographic image”. In: *International Journal of Osteoporosis* 20, pp. 455–461. DOI: [10.1007/s00198-008-0665-4](https://doi.org/10.1007/s00198-008-0665-4).

- LI, K., S. MEMBER, X. WU, and S. MEMBER (2006). “Optimal Surface Segmentation in Volumetric Images - A Graph-Theoretic Approach”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 28.1, pp. 119–134.
- LI, X., Q. DOU, H. CHEN, C.-w. FU, X. QI, D. L. BELAVY, G. ARMBRECHT, D. FELSEBERG, G. ZHENG, and P.-a. HENG (2018). “3D multi-scale FCN with random modality voxel dropout learning for Intervertebral Disc Localization and Segmentation from Multi-modality MR Images”. In: *IEEE Transactions on Medical Medical Image Analysis* 45, pp. 41–54. DOI: [10.1016/j.media.2018.01.004](https://doi.org/10.1016/j.media.2018.01.004).
- LINDNER, C., S. THIAGARAJAH, J. M. WILKINSON, G. A. WALLIS, and T. F. COOTES (2013). “Fully automatic segmentation of the proximal femur using random forest regression voting”. In: *IEEE Transactions on Medical Imaging* 32.8, pp. 1462–1472. DOI: [10.1109/TMI.2013.2258030](https://doi.org/10.1109/TMI.2013.2258030).
- LINDNER, C. (2017). “Automated Image Interpretation Using Statistical Shape Models”. In: *Statistical Shape and Deformation Analysis: Methods, Implementation and Applications*. 1st ed. Elsevier Ltd, pp. 3–32. DOI: [10.1016/B978-0-12-810493-4.00002-X](https://doi.org/10.1016/B978-0-12-810493-4.00002-X).
- LITJENS, G., T. KOOI, B. E. BEJNORDI, A. A. A. SETIO, F. CIOMPI, M. GHAFORIAN, J. A. van der LAAK, B. van GINNEKEN, and C. I. SANCHEZ (2017). “A survey on deep learning in medical image analysis”. In: *IEEE Transactions on Medical Image Analysis* 42, pp. 60–88. DOI: [10.1016/j.media.2017.07.005](https://doi.org/10.1016/j.media.2017.07.005).
- LONG, J., E. SHELHAMER, and T. DARRELL (2015). “Fully Convolutional Adaptation Networks for Semantic Segmentation”. In: *Arxiv*. DOI: [10.1109/CVPR.2018.00712](https://doi.org/10.1109/CVPR.2018.00712).
- LORENSEN, W. E. and H. E. CLINE (1987). “Marching Cubes: A high resolution 3D surface construction algorithm”. In: *Computer Graphics* 21.4, pp. 163–169. DOI: [10.1088/2053-1591/aade2a](https://doi.org/10.1088/2053-1591/aade2a).
- LUTHI, M., T. GERIG, C. JUD, and T. VETTER (2018). “Gaussian Process Morphable Models”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40.8, pp. 1860–1873. DOI: [10.1109/TPAMI.2017.2739743](https://doi.org/10.1109/TPAMI.2017.2739743).
- MA, J., F. LIN, S. WESARG, and M. ERDT (2018). “A Novel Bayesian Model Incorporating Deep Neural Network and Statistical Shape Model for Pancreas Segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 11073 LNCS, pp. 480–487. DOI: [10.1007/978-3-030-00937-3\\_55](https://doi.org/10.1007/978-3-030-00937-3_55).
- MA, J., A. WANG, F. LIN, S. WESARG, and M. ERDT (2017). “Nonlinear Statistical Shape Modeling for Ankle Bone Segmentation Using a Novel Kernelized Robust PCA”. In:

- Proceedings - Medical Image Computing and Computer Assisted Intervention (MICCAI)* 10433, pp. 108–115. DOI: [10.1007/978-3-319-66182-7](https://doi.org/10.1007/978-3-319-66182-7).
- MALLADI, R., J. A. SETHIAN, and B. C. VEMURI (1995). “Shape Modeling with Front Propagation : A Level Set Approach”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17.2, pp. 158–175.
- MARKELJ, P., D. TOMAZEVIC, B. LIKAR, and F. PERNUS (2012). “A review of 3D/2D registration methods for image-guided interventions”. In: *IEEE Transactions on Medical Image Analysis* 16.3, pp. 642–661. DOI: [10.1016/j.media.2010.03.005](https://doi.org/10.1016/j.media.2010.03.005).
- MILLETARI, F. (2016). “V-Net : Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation”. In: *Fourth International Conference on 3D Vision 2016 Fourth International Conference on 3D Vision*, pp. 565–571. DOI: [10.1109/3DV.2016.79](https://doi.org/10.1109/3DV.2016.79).
- MUKHOPADHYAY, S., A. METCALFE, A. R. GUHA, K. MOHANTY, S. HEMMADI, K. LYONS, and D. O'DOHERTY (2011). “Malreduction of syndesmosis - Are we considering the anatomical variation?” In: *Injury* 42.10, pp. 1073–1076. DOI: [10.1016/j.injury.2011.03.019](https://doi.org/10.1016/j.injury.2011.03.019).
- MUTSVANGWA, T., W. WASSWA, V. BURDIN, B. BOROTIKAR, and T. S. DOUGLAS (2017). “Interactive patient-specific 3D approximation of scapula bone shape from 2D X-ray images using landmark-constrained statistical shape model fitting”. In: *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, pp. 1816–1819. DOI: [10.1109/EMBC.2017.8037198](https://doi.org/10.1109/EMBC.2017.8037198).
- NAULT, M. L., J. HÉBERT-DAVIES, G. Y. LAFLAMME, and S. LEDUC (2013). “CT scan assessment of the syndesmosis: A new reproducible method”. In: *Journal of Orthopaedic Trauma* 27.11, pp. 638–641. DOI: [10.1097/BOT.0b013e318284785a](https://doi.org/10.1097/BOT.0b013e318284785a).
- NAVAB, N., S.-M. HEINING, and J. TRAUB (2010). “Camera Augmented Mobile C-Arm (CAMC): Calibration, Accuracy Study, and Clinical Applications”. In: *IEEE Transactions on Medical Imaging* 29.7, pp. 1412–1423.
- NIELSON, J., M. GARDNER, M. PETERSON, J. SALLIS, H. POTTER, D. HELFET, and D. LORICH (2005). “Radiographic Measurements Do Not Predict Syndesmotic Injury in Ankle Fractures”. In: *Clinical Orthopaedics and Related Research* 436, pp. 216–221. DOI: [10.1097/01.blo.0000161090.86162.19](https://doi.org/10.1097/01.blo.0000161090.86162.19).
- NOLDEN, M., S. ZELZER, A. SEITEL, D. WALD, M. MÜLLER, A. M. FRANZ, D. MALEIKE, M. FANGERAU, M. BAUMHAUER, L. MAIER-HEIN, K. H. MAIER-HEIN, H. .-.P. MEINZER, and I. WOLF (2013). “The Medical Imaging Interaction Toolkit: challenges and advances”. In: *International Journal of Computer Assisted Radiology and Surgery* 8.4, pp. 607–620. DOI: [10.1007/s11548-013-0840-8](https://doi.org/10.1007/s11548-013-0840-8).

- NORAJITRA, T. and K. H. MAIER-HEIN (2017). “3D Statistical Shape Models Incorporating Landmark-Wise Random Regression Forests for Omni-Directional Landmark Detection”. In: *IEEE Transactions on Medical Imaging* 36.1, pp. 155–168. DOI: [10.1109/TMI.2016.2600502](https://doi.org/10.1109/TMI.2016.2600502).
- NOVIKOV, A. A., D. LENIS, D. MAJOR, J. HLADUVKA, M. WIMMER, and K. BUHLER (2018). “Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs”. In: *IEEE Transactions on Medical Imaging* 37.8, pp. 1865–1876. DOI: [10.1109/tmi.2018.2806086](https://doi.org/10.1109/tmi.2018.2806086).
- OTAKE, Y., A. S. WANG, A. UNERI, G. KLEINSZIG, S. VOGT, N. AYGUN, S. F. L. LO, J. P. WOLINSKY, Z. L. GOKASLAN, and J. H. SIEWERDSEN (2015). “3D-2D registration in mobile radiographs: Algorithm development and preliminary clinical evaluation”. In: *Physics in Medicine and Biology* 60.5, pp. 2075–2090. DOI: [10.1088/0031-9155/60/5/2075](https://doi.org/10.1088/0031-9155/60/5/2075).
- OTSU, N. (1979). “A threshold selection method from gray-level histograms”. In: *IEEE Transactions on systems, Man, and Cybernetics* 9.1, pp. 62–66. DOI: [10.1109/TSMC.1979.4310076](https://doi.org/10.1109/TSMC.1979.4310076).
- PAUL BESL, N. M. (1992). “A Method for Registration of 3-D Shapes”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 14.2, pp. 239–256.
- PENNEY, G. P., J. WEESE, J. A. LITTLE, P. DESMEDT, D. L. G. HILL, and D. J. HAWKES (1998). “A Comparison of similarity measures for use in 2D-3D medical image registration”. In: *IEEE Transactions on Medical Imaging* 17.4, pp. 586–595.
- PETTERSSON, H. (1998). “Radiography”. In: *Orthopedic Imaging - Medical Radiology*. Ed. by A. M. DAVIES and H. PETTERSSON. Berlin, Heidelberg: Springer-Verlag.
- PIZER, S. M., P. T. FLETCHER, S. JOSHI, A. THALL, J. Z. CHEN, Y. FRIDMAN, D. S. FRITSCH, G. GASH, J. M. GLOTZER, M. R. JIROUTEK, C. LU, K. E. MULLER, G. TRACTON, P. YUSHKEVICH, and E. L. CHANEY (2003). “Deformable M-Reps for 3D Medical Image Segmentation”. In: *International Journal of Computer Visualization* 55.2-3, pp. 85–106.
- PRASOON, A., K. PETERSEN, and C. IGEL (2013). “Deep Feature Learning for Knee Cartilage Segmentation Using a Triplanar Convolutional Neural Network”. In: *Proceedings - Medical Image Computing and Computer-Assisted Intervention (MICCAI)* 9, pp. 246–253.
- REUTER, M., F. E. WOLTER, and N. PEINECKE (2006). “Laplace-Beltrami spectra as ‘Shape-DNA’ of surfaces and solids”. In: *CAD Computer Aided Design* 38.4, pp. 342–366. DOI: [10.1016/j.cad.2005.10.011](https://doi.org/10.1016/j.cad.2005.10.011).
- REYNEKE, C. J., M. LUTHI, V. BURDIN, T. S. DOUGLAS, T. VETTER, and T. E. MUTSVANGWA (2018). “Review of 2D/3D reconstruction using statistical shape and intensity models



- and X-ray image synthesis: towards a unified framework”. In: *IEEE Reviews in Biomedical Engineering* 3333.c, pp. 1–17. DOI: [10.1109/RBME.2018.2876450](https://doi.org/10.1109/RBME.2018.2876450).
- RONNEBERGER, O., P. FISCHER, and T. BROX (2015). “U-net: Convolutional networks for biomedical image segmentation”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 9351, pp. 234–241. DOI: [10.1007/978-3-319-24574-4\\_28](https://doi.org/10.1007/978-3-319-24574-4_28).
- SADOWSKY, O., K. RAMAMURTHI, L. M. ELLINGSEN, G. CHINTALAPANI, J. L. PRINCE, and R. H. TAYLOR (2006). “Atlas-assisted tomography: registration of a deformable atlas to compensate for limited-angle cone-beam trajectory”. In: *3rd IEEE International Symposium on Biomedical Imaging: Nano to Macro, 2006*. Pp. 1244–1247. DOI: [10.1109/ISBI.2006.1625150](https://doi.org/10.1109/ISBI.2006.1625150).
- SADOWSKY, O., G. CHINTALAPANI, and R. H. TAYLOR (2007). “Deformable 2D-3D registration of the pelvis with a limited field of view, using shape statistics BT”. In: *Proceedings - Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. Vol. 4792 LNCS. PART 2, pp. 519–526.
- SARKALKAN, N., H. WEINANS, and A. A. ZADPOOR (2014). *Statistical shape and appearance models of bones*. DOI: [10.1016/j.bone.2013.12.006](https://doi.org/10.1016/j.bone.2013.12.006).
- SAXENA, A. (2012). *International Advances in Foot and Ankle Surgery*. Ed. by A. SAXENA. Springer-Verlag London.
- SCHNETZKE, M., S. Y. VETTER, N. BEISEMANN, B. SWARTMAN, P. A. GRÜTZNER, and J. FRANKE (2016). “Management of syndesmotic injuries: What is the evidence?” In: *World Journal of Orthopaedics* 7.11, pp. 718–725. DOI: [10.5312/wjo.v7.i11.718](https://doi.org/10.5312/wjo.v7.i11.718).
- SCHREIBER, J. J., A. S. MCLAWHORN, C. J. DY, and E. M. GOLDWYN (2013). “Intraoperative Contralateral View for Assessing Accurate Syndesmosis Reduction”. In: *Orthopedics* 36.5, pp. 360–361. DOI: [10.3928/01477447-20130426-03](https://doi.org/10.3928/01477447-20130426-03).
- SCHÜNKE, M., E. SCHULTE, U. SCHUMACHER, M. VOLL, and K. H. WESKER (2018). *Prometheus - Allgemeine Anatomie und Bewegungssystem - Lernatlas der Anatomie*. 5th ed. Stuttgart, New York: Georg Thieme Verlag. DOI: [10.1055/b-006-149643](https://doi.org/10.1055/b-006-149643).
- SIEMENS HEALTHCARE (2016). *ARCADIS Orbic /Orbic 3D Enhanced precision in the OR*. Tech. rep.
- SOILLE, P. (2004). *Morphological Image Analysis -Principles and Applications*. 2nd ed. Springer-Verlag Berlin Heidelberg.
- SONG, A. M. and XUBO (2010). “Point Set Registration: Coherent Point Drift”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 32.12, pp. 2263–2275. DOI: [10.1364/OE.25.013857](https://doi.org/10.1364/OE.25.013857).

- SORA, M. C., B. STROBL, D. STAYKOV, and S. FÖRSTER-STREFFLEUR (2004). "Evaluation of the ankle syndesmosis: A plastination slices study". In: *Clinical Anatomy* 17.6, pp. 513–517. DOI: [10.1002/ca.20019](https://doi.org/10.1002/ca.20019).
- SOTIRAS, A., C. DAVATZIKOS, S. MEMBER, and N. PARAGIOS (2013). "Deformable Medical Image Registration : A Survey". In: *IEEE Transactions on Medical Imaging* 32.7, pp. 1153–1190. DOI: [10.1109/TMI.2013.2265603](https://doi.org/10.1109/TMI.2013.2265603).
- STAIB, L. H. and J. S. DUNCAN (1992). "Boundary Finding with Parametrically Deformable Models". In: *IEEE Trans. on Pattern Analysis and Machine Intelligence* 14.11, pp. 1061–1075.
- STANDRING, S., H. ELLIS, and B. BERKOVITZ, eds. (2015). *Gray's Anatomy*. 41st ed. Elsevier Churchill Livingstone.
- STEININGER, P., K. D. FRITSCHER, G. KOFLER, B. SCHULER, M. HÄNNI, K. SCHWIEGER, and R. SCHUBERT (2008). "Comparison of different metrics for appearance-model-based 2D/3D-registration with X-ray images". In: *Proceedings- Bildverarbeitung für die Medizin*, pp. 122–126.
- STIEHL, J. B., W. H. KONERMANN, R. G. HAAKER, and A. M. DIGIOIA (2007). *Navigation and MIS in Orthopedic Surgery*. Springer Berlin Heidelberg.
- SZEIMIES, U., A. STÄBLER, and M. WALTHER, eds. (2012). *Bildgebende Diagnostik des Fußes*. Georg Thieme Verlag KG.
- SZEKELY, G., A. KELEMEN, C. BRECHBUEHLER, and G. GERIG (1996). "Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models". In: *IEEE Transactions on Medical Image Analysis* 1.1, pp. 19–34.
- THODBERG, H. H. (2003). "Minimum Description Length Shape and Appearance Models". In: *International conference on Information Processing in Medical Imaging*. Ed. by C. TAYLOR and J. NOBLE. Springer-Verlag Berlin Heidelberg, pp. 51–62. DOI: [10.1007/978-3-540-45087-0\\_5](https://doi.org/10.1007/978-3-540-45087-0_5).
- TOMAŽEVI, D. (2008). "3D/2D Registration of Medical Images". PhD thesis. University of Ljubljana, Slovenia.
- TOTH, D., S. MIAO, T. KURZENDORFER, C. A. RINALDI, R. LIAO, T. MANSI, K. RHODE, and P. MOUNTNEY (2018). "3D/2D model-to-image registration by imitation learning for cardiac procedures". In: *International Journal of Computer Assisted Radiology and Surgery* 13.8, pp. 1141–1149. DOI: [10.1007/s11548-018-1774-y](https://doi.org/10.1007/s11548-018-1774-y).
- UNBERATH, M., J. N. ZAECH, S. C. LEE, B. BIER, J. FOTOUHI, M. ARMAND, and N. NAVAB (2018). "DeepDRR - A Catalyst for Machine Learning in Fluoroscopy-Guided Procedures". In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Ar-*

- tificial Intelligence and Lecture Notes in Bioinformatics*) 11073 LNCS, pp. 98–106. DOI: [10.1007/978-3-030-00937-3\\_12](https://doi.org/10.1007/978-3-030-00937-3_12).
- VETTER, S. Y., M. PRIVALOV, N. BEISEMANN, B. SWARTMAN, H. KEIL, J. KIRSCH, P. A. GRÜTZNER, and J. FRANKE (2019a). “Influence of ankle joint position on angles and distances of the ankle mortise using intraoperative cone beam CT: A cadaveric study”. In: *PLoS ONE* 14.5, pp. 1–12. DOI: [10.1371/journal.pone.0217737](https://doi.org/10.1371/journal.pone.0217737).
- VETTER, S. Y., M. GASSAUER, L. UHLMANN, B. SWARTMAN, M. SCHNETZKE, H. KEIL, J. FRANKE, P. A. GRÜTZNER, and N. BEISEMANN (2019b). “A standardised computed tomography measurement method for distal fibular rotation”. In: *European Journal of Trauma and Emergency Surgery*, pp. 1–6. DOI: [10.1007/s00068-019-01120-6](https://doi.org/10.1007/s00068-019-01120-6).
- WARFIELD, S. K., K. H. ZOU, and W. M. WELLS (2006). *LNCS 4191 - Validation of Image Segmentation by Estimating Rater Bias and Variance*. Tech. rep.
- WARNER, S. J., P. D. FABRICANT, M. R. GARNER, P. C. SCHOTTEL, D. L. HELFET, and D. G. LORICH (2015). “The Measurement and Clinical Importance of of Rotational Ankle Fractures”. In: *Journal of Bone and Joint Surgery* 97-A.23, pp. 1935–1944.
- WEBER, B. G. and C. COLTON (1991). *Malleolar Fractures*. Ed. by M. MÜLLER, M. ALLGÖWER, R. SCHNEIDER, and H. WILLENEGGER. 3rd ed. Springer-Verlag Berlin Heidelberg. Chap. 14, pp. 595–612.
- XIE, W., J. FRANKE, C. CHEN, P. A. GRÜTZNER, S. SCHUMANN, L. P. NOLTE, and G. ZHENG (2014). “Statistical model-based segmentation of the proximal femur in digital antero-posterior (AP) pelvic radiographs”. In: *International Journal of Computer Assisted Radiology and Surgery* 9.2, pp. 165–176. DOI: [10.1007/s11548-013-0932-5](https://doi.org/10.1007/s11548-013-0932-5).
- YILDIRIM, H., A. MAVI, O. BÜYÜKBEBECI, and E. GÜMÜŞBURUN (2003). “Evaluation of the fibular incisura of the tibia with magnetic resonance imaging”. In: *Foot and Ankle International* 24.5, pp. 387–391. DOI: [10.1177/107110070302400502](https://doi.org/10.1177/107110070302400502).
- YU, W., C. CHU, M. TANNAST, and G. ZHENG (2016). “Fully automatic reconstruction of personalized 3D volumes of the proximal femur from 2D X-ray images”. In: *International Journal of Computer Assisted Radiology and Surgery* 11.9, pp. 1673–1685. DOI: [10.1007/s11548-016-1400-9](https://doi.org/10.1007/s11548-016-1400-9).
- YU, W. and G. ZHENG (2018). “Atlas-based 3D intensity volume reconstruction from 2D long leg standing X-rays: Application to hard and soft tissues in lower extremity”. In: *Advances in Experimental Medicine and Biology*. DOI: [10.1007/978-981-13-1396-7\\_9](https://doi.org/10.1007/978-981-13-1396-7_9).
- ZENG, G., X. YANG, J. LI, L. YU, and P.-a. HENG (2017). “3D U-net with Multi-level Deep Supervision : Fully Automatic Segmentation of Proximal Femur in 3D MR Images”. In: *Wang Q., Shi Y., Suk H.I., Suzuki K. (eds) Machine Learning in Medical Imaging. MLMI*

2017. *Lecture Notes in Computer Science*. Vol. 10541, pp. 274–282. DOI: [10.1007/978-3-319-67389-9](https://doi.org/10.1007/978-3-319-67389-9).
- ZHENG, G., J. VON RECUM, L. P. NOLTE, P. A. GRÜTZNER, S. D. STEPPACHER, and J. FRANKE (2012). “Validation of a statistical shape model-based 2D/3D reconstruction method for determination of cup orientation after THA”. In: *International Journal of Computer Assisted Radiology and Surgery* 7.2, pp. 225–231. DOI: [10.1007/s11548-011-0644-7](https://doi.org/10.1007/s11548-011-0644-7).
- ZHENG, G. (2011). “Personalized X-ray reconstruction of the proximal femur via intensity-based non-rigid 2D-3D registration”. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. DOI: [10.1007/978-3-642-23629-7\\_73](https://doi.org/10.1007/978-3-642-23629-7_73).
- (2013a). *iJoint - 2D/3D reconstruction of patient-specific hip joint from conventional X-ray radiographs*. URL: [https://www.istb.unibe.ch/research/information\\_processing\\_in\\_medical\\_interventions/ijoint/index\\_eng.html](https://www.istb.unibe.ch/research/information_processing_in_medical_interventions/ijoint/index_eng.html) (visited on 03/04/2019).
- (2013b). *iLeg. 2D/3D reconstruction of patient-specific hip joint from clinically available X-ray*. URL: [https://www.istb.unibe.ch/research/information\\_processing\\_in\\_medical\\_interventions/ileg/index\\_eng.html](https://www.istb.unibe.ch/research/information_processing_in_medical_interventions/ileg/index_eng.html) (visited on 03/04/2019).
- ZHENG, G., X. DONG, and M. A. BALLESTER (2007). “Unsupervised Reconstruction of a Patient-Specific Surface Model of a Proximal Femur from Calibrated Fluoroscopic Images”. In: *Proceedings - Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. February. DOI: [10.1007/b12345](https://doi.org/10.1007/b12345).
- ZHENG, G., S. GOLLMER, S. SCHUMANN, X. DONG, T. FEILKAS, and M. A. GONZÁLEZ BALLESTER (2009). “A 2D/3D correspondence building method for reconstruction of a patient-specific 3D bone surface model using point distribution models and calibrated X-ray images”. In: *IEEE Transactions on Medical Image Analysis* 13.6, pp. 883–899. DOI: [10.1016/j.media.2008.12.003](https://doi.org/10.1016/j.media.2008.12.003).
- ZHENG, G. and L.-P. NOLTE (2018). “Computer-Aided Orthopaedic Surgery: State-of-the-Art and Future Perspectives”. In: *Advances in Experimental Medicine and Biology 1093 - Intelligent Orthopaedics*. Ed. by G. ZHENG, W. TIAN, and X. ZHUANG. Springer Nature Singapore.
- ZHENG, G. and S. SCHUMANN (2009). “3D reconstruction of a patient-specific surface model of the proximal femur from calibrated x-ray radiographs: A validation study”. In: *Medical Physics* 36.4, pp. 1155–1166. DOI: [10.1118/1.3089423](https://doi.org/10.1118/1.3089423).
- ZHENG, G. and W. YU (2017). “Statistical Shape and Deformation Models Based 2D-3D Reconstruction”. In: *Statistical Shape and Deformation Analysis: Methods, Implemen-*

*tation and Applications*, pp. 329–349. DOI: [10.1016/B978-0-12-810493-4.00015-8](https://doi.org/10.1016/B978-0-12-810493-4.00015-8).



Some ideas, figures and tables shown in this thesis have appeared previously in the following publications:

### **Peer-reviewed international conference proceedings and journals**

THOMAS, S., F. ISENSEE, S. KOHL, M. PRIVALOV, N. BEISEMANN, B. SWARTMAN, H. KEIL, S. Y. VETTER, J. FRANKE, P. GRUETZNER, L. MAIER-HEIN, M. NOLDEN, and K. MAIER-HEIN (2019). “Computer-assisted intra-operative verification of surgical outcome for the treatment of syndesmotic injuries through contralateral side comparison”. In: *International Journal of Computer Assisted Radiology and Surgery*. DOI: [10 . 1007 / s11548-019-02043-8](https://doi.org/10.1007/s11548-019-02043-8).

THOMAS, S., M. SCHNETZKE, M. BREHLER, B. SWARTMAN, S. VETTER, J. FRANKE, P. A. GRÜTZNER, H.-P. MEINZER, and M. NOLDEN (2017a). “Upper ankle joint space detection on low contrast intraoperative fluoroscopic C-arm projections”. In: *Proc. SPIE 10135, Medical Imaging 2017: Image-Guided Procedures, Robotic Interventions, and Modeling*. Ed. by R. J. WEBSTER and B. FEI. DOI: [10 . 1117/12.2255633](https://doi.org/10.1117/12.2255633).

### **Peer-reviewed national conferences**

THOMAS, S., M. SCHNETZKE, J. FRANKE, S. VETTER, B. SWARTMAN, P. A. GRÜTZNER, H.-P. MEINZER, and M. NOLDEN (2017b). “Abstract: Detektion des tibiotalaren Gelenkspaltes in intraoperativen C-Bogen Projektionen”. In: *Bildverarbeitung für die Medizin 2017*. Ed. by K. H. MAIER-HEIN GEB. FRITZSCHE, T. M. DESERNO GEB. LEHMANN, H. HANDELS, and T. TOLXDORFF. Berlin, Heidelberg: Springer Berlin Heidelberg, p. 217.





### **Personalien**

Name: Sarina Thomas  
Geburtsdatum: 09.09.1987  
Geburtsort: Herborn  
Nationalität: Deutsch

### **Akademische Ausbildung**

Seit 07/2017 Doktorandin am  
Deutsches Krebsforschungszentrum (DKFZ),  
Abteilung für Medizinische Bildverarbeitung  
(PD Dr. Klaus Maier-Hein)

09/2015-07/2017 Doktorandin am  
Deutsches Krebsforschungszentrum (DKFZ),  
Abteilung für Medizinische und Biologische Informatik  
(Prof. Dr. Hans-Peter Meinzer)

10/2012-07/2015 Master of Science Visual Computing, Master of Science Autonome  
Systeme  
Technische Universität Darmstadt, Fakultät für Informatik  
Durchschnitt: sehr gut (1,3 Visual Computing,  
1,4 Autonome Systeme)

03/2009-09/2012 Bachelor of Science Medizintechnik  
Hochschule Mannheim, Fakultät für Informationstechnik  
Durchschnitt: sehr gut (1,3)

07/2007           Abitur  
                  Johanneum Gymnasium Herborn  
                  Durchschnitt: gut (2,1)

### **Berufliche Erfahrung**

03/2012-09/2012   Praktikum und Bachelorarbeit  
                  KUKA Laboratories GmbH, Augsburg

03/2010-06/2010   Tutorin für Digitaltechnik 1  
                  Hochschule Mannheim

02/2010-03/2010   Praktikum in der Strahlentherapie  
                  Universitätsklinikum Mannheim

09/2008-10/2008   Laborpraktikum  
                  Hessisches Landes Prüfungs- und Untersuchungsamt (im Gesund-  
                  heitswesen), Dillenburg

02/2010-03/2010   Technisches Grundpraktikum  
                  CLOOS Schweisstechnik, Haiger

### **Ausland**

08/2013-12/2013   Auslandssemester Computer Science Department  
                  Simon Fraser University, Vancouver, Kanada

08/2011-10/2011   Software Praktikum in der Robotics Group  
                  Accuray Inc., Sunnyvale, Silicon Valley, USA

08/2010-02/2011   Praxissemester Biomedical Engineering Department  
                  College of Engineering, Peking University, Peking, China

### **Stipendien**

2015            Helmholtz Stipendium

2013            PROMOS Reisetstipendium Canada

2009            GIZ Stipendium

15. Oktober 2019

## ACKNOWLEDGMENTS

There are many people I would like to acknowledge in this section. First of all, I thank PD Dr. Klaus Maier-Hein for all his useful comments, fruitful discussions and his patience with me. I further want to thank Dr. Marco Nolden for his supervision especially during the concept building phase of the project.

I enjoyed the inspiring research environment and I am thankful for the possibility to collaborate with partners from industry and hospitals, which was essential for this thesis. Therefore, I want to thank all partners within the context of this project, Dr. Holger Kunze, Dr. Alexander Gemmel and Gerhard Kleinszig from Siemens Healthineers for their technical instructions and for partly funding this project. Not to forget all highly motivated clinicians from the BG Ludwigshafen, especially Dr. Jochen Franke, for giving me insights to the OR and the clinical practice in trauma surgery.

I want to thank all my CAMIC co-workers and of course all my roommates, who showed great team spirit throughout the entire time in the department as well as a helping hand whenever or for whatever I needed it. I thank Lisa Kausch for all the knowledge exchange and many peculiar but mostly amusing excursions to clinical studies. Furthermore, a special thanks goes to the "Mädels", Esther and Jasmin for all their life saving support by any means. I would like to thank the members of my most favorite "Waschbären-party" running group, who not only discussed work-related topics during our practise but also kept me motivated to achieve my goals. I also want to thank all my amazing proof-reading angels, Andre, Ralf, Lisa, Christoph, Christian, Joseph, Maxim, Clemens, Michel and Matthias for their constructive feedback during the writing of this thesis.

At the end of this section a truly deep "Thank you" goes to my family and friends, who always believed in and supported me in good but -even more important- in rough times. Finally, I want to thank Leonard for his unconditional support, for always encouraging me and for sharing the best moments with me. This thesis is dedicated to all of you!