HABILITATION THESIS

A Decade in Learning 3D Shape for Medical Applications

written by

Ing. Michal Španěl, Ph.D.

Brno, November 2023

Abstract

This habilitation thesis explores the intersection of clinical medicine, computer graphics, and machine learning in computer-aided diagnosis and treatment planning. Over the past decade, the author has actively contributed to the field, particularly in 3D tissue shape reconstruction for pre-operative treatment planning, emphasizing practical applications in clinical settings. The author acted as a catalyst for research that translated theoretical concepts into clinical practice and mentorship of young researchers and PhD students. The thesis recounts experiences in both academic and industry settings, including leadership at a company driving medical product innovations. The thesis, built on published works, focuses on statistical shape models for 3D tissue modelling, with applications of 2D-to-3D registration for fractured bone reconstruction in osteosynthesis. The second part explores deep learning methods for automatic skull shape reconstruction for cranial implant design and tooth detection on digital dental cast models. Drawing upon personal experiences, the thesis offers insights into the past, present, and potential future research directions.

Keywords

Computer-aided diagnosis, computer-guided surgery, pre-operative treatment planning, statistical shape models, 2D-3D reconstruction from radiographs, digitally rendered radiographs, osteosynthesis, cranioplasty, skull shape reconstruction, cranial implant, deep learning, convolutional neural networks, dental cast, landmark detection, tooth detection on digital dental scan.

Acknowledgments

This habilitation thesis was born not only from my efforts but mainly from the efforts of my students and colleagues, who collaborated with me with outstanding commitment to our joint research projects. Most of all Oldřich Kodym, Ondřej Klíma, Petr Kleparník and Tibor Kubík. Thank you all very much! Many thanks also to my wife Markéta and both my children Nikol and Filip for their enormous tolerance to my long working days and all their smiles. I also thank all my colleagues and friends for our fruitful conversations and exceptionally inspiring working environment. Víťa Beran and Adam Herout, to you first. I would also like to mention and thank doc. Ing. Přemysl Kršek and doc. MUDr. Petr Krupa, CSc., who attracted me to the topic of medical image data analysis and were my long-term partners, first at the university and later in our joint medical startup.

Contents

Contents

List of Figures

List of Abbreviations

AP Anterior-Posterior

CAD Computer-Aided Design

CAD Computer-aided Diagnosis

CGS Computer-guided Surgery

CMA Covariance Matrix Adaptation

CMSA Covariance Matrix Self Adaptation

CNN Convolutional Neural Network

CT Computed Tomography

DRR Digitally Reconstructed Radiograph

DSC Dice Coefficient

FEM Finite Element Method

GAN Generative Adversarial Network

GNN Graph Neural Network

GPA Generalized Procrustes Analysis

HD Hausdorff Distance

LAT Lateral

LM Levenberg Marquardt

NMI Normalised Mutual Information

PCA Principal Component Analysis

List of Abbreviations

PDM Point Distribution Model

PPCA Probabilistic Principal Component Analysis

RANSAC RANdom SAmple Consensus

SAM Statistical Appearance Model

SSIM Statistical Shape and Intensity Model

SSM Statistical Shape Model

1 Introduction

With the progress of medical imaging methods and the rapid development of digital technologies, a new research field has been established - computer-aided diagnosis and treatment planning. Modern computer graphics, image processing and machine learning techniques began to affect medicine as a tool for scientific analysis and data modelling and to facilitate diagnosis or planning of surgical treatments. As a technically oriented person, I was fortunate to be involved in this process and to collaborate with clinicians on various exciting research projects.

This habilitation thesis maps the last ten years of my work in medical applications of image processing and machine learning, focusing on the reconstruction of the 3D shape of tissues and its modelling that would make treatment planning easier for clinicians.

During those years, I searched for research topics with the potential for practical application in clinical practice and helped connect enthusiastic young scientists, PhD students, and the medical world. I motivated young researchers and helped them direct their research. I feel more like a mentor who sometimes gives bits of research advice and draws on the experience I have gathered on research projects in a university environment and an industry with a strong focus on medical product innovations, especially SW solutions. I was lucky to supervise brilliant and hardworking students who were the main motor of the research, and I helped them define what "motor" for what "car" we needed.

An essential element of our research activities has always been applying scientific methods in clinical practice. An excellent experience for me was working as a head of software development in the company 3Dim Laboratory s.r.o., which has its roots in the university and develops innovative products in the healthcare sector. At the beginning of all the company's products was the applied research in 3D visualization, 3D modelling and geometric reconstruction of tissues. The company created products that are used in clinical practice today, both in the Czech Republic and abroad.

Currently, the original company has two followers. TESCAN 3DIM, s.r.o.* focuses

---

*https://www.tescan3dim.com/

on custom software development for 3D image data analysis, 3D visualization, image and geometry processing using conventional and deep learning methods. And the company develops dental treatment planning software for computer-guided surgery. TecuMed, s.r.o.[†] specializes in manufacturing and innovating medical devices for individual treatment in neurosurgery, maxillofacial surgery, oncology, traumatology and orthopaedics. It brings individual patient treatment solutions into clinical practice.

The habilitation thesis builds on published papers and articles to which I contributed. The main topics are applications of Statistical Shape Models (SSM) and newer approaches based on deep learning in modelling the 3D shape of tissues, especially bones, for planning complex procedures and designing implants tailored to the patient.

SSM is a traditional tool for modelling shape and its variability in a population [1, 2, 3]. One of the exciting tasks is the so-called 2D-to-3D registration, where a bone 3D shape is reconstructed from one or a few 2D radiographs [4, 5, 6, 7, 8]. Chapter 3 is devoted to this topic [9, 10, 11, 12].

With the evolution of deep learning methods, it was natural to start investigating the possibilities of their use in 3D shape analysis. The second part of the habilitation in Chapter 4 focuses on the use of deep learning for tasks such as automatic reconstruction of the skull shape for cranial implant design [13, 14, 15] or detection of landmark points on digital dental cast models [16].

A summary of my personal experiences in the context of the mentioned areas and the past years, including a look at the possible focus of future research directions, is in Section 3.4, Section 4.4 and Chapter 5.

[†] https://www.tecumed.com/

2 3D Modelling in Computer-assisted Treatment Planning

3D reconstruction of anatomy and exact shape of tissues has been the core of modern imaging methods such as Computed Tomography (CT) for many years. It plays a crucial role in the diagnosis of diseases. Compared to 2D radiographs (e.g. X-ray images), 3D reconstruction is more accurate and sensitive, which is why 3D reconstruction is used mainly to diagnose complicated fractures and deformities. Many interdisciplinary scientific fields combining clinical medicine and 3D computer analysis have developed dramatically in recent years.

- **Computer-aided Diagnosis (CAD)** – represents systems that assist clinicians in the interpretation of medical image data like X-rays or CT scans.

- **Computer-guided Surgery (CGS)** – software usually helps determine optimal location for implant placement and creates surgical guides to help the clinician place the implant.

Digital Orthopaedics is a great example. It helps clinicians with pre-operative selection of appropriate procedure and material but also during the surgical procedure itself. The software creates various types of surgical guides to help the clinician do corrective cuts, reposition bone fragments into anatomical position and finally place the implant (Figure 2.1). Digital Orthopaedics also deals with patient monitoring in the post-operative phase.

More widely, these technologies include not only pre-operative planning and virtual surgeriy, but also navigated surgery or education and training, which provides young medics with another tool to improve their skills. Computer-assisted medicine is an expansive area, and the mentioned technologies have been widely promoted in orthopaedics, dental surgery, traumatology, microsurgery, spine surgery, maxillofacial surgery and more in recent years.

2.1 Pre-operative Treatment Planning

This thesis addresses *Computer-assisted Pre-operative Treatment Planning* (Figure 2.2) in traumatology, dental implantology, orthodontics and cranioplasty. I prefer the term *computer-assisted* because the technology is meant to aid the physician, not guide. For example, a dentist plans the location and type of implant and models the shape of the crown with the help of computer software.

The basis for the preparation and planning of operations are typically X-ray images or CT scans, based on which the doctor chooses the appropriate treatment strategy, selects the most suitable material (implants, fixation screws, etc.) and plans the operation itself.

However, the human body has a very complex anatomical structure. In complicated cases, with large bone deformations, with atypical bone shape, after severe injuries or in cases of bone tumours, these standard procedures cannot ensure optimal planning and execution of surgery.

Osteosynthesis (Figure 2.3) is defined as fixation of a bone. It is a surgical procedure to treat bone fractures in which bone fragments are joined with screws, plates, nails or wires. The fractured bone is fixed with the aforementioned and can knit stably in the correct position.

Figure 2.1: Radius corrective osteotomy – is a bone-cutting procedure to realign forearm bones into anatomically correct position to guarantee proper functionality. The top row shows the planning of the osteotomy comparing both the hands. The bottom row shows planning of the plate and drilling and cutting guides to help cut the bone and rotate it into the correct position. Image courtesy of TecuMed, s.r.o.

In cases where it is not possible to use standard implants, it is necessary to plan and manufacture an implant tailored to the patient (Figure 2.4). In these cases, 3D tissue shape reconstruction methods come into play. By creating a computerized 3D model of the tissue (so-called digital anatomical models), it is possible to provide the clinician with a better spatial idea and guidelines for planning treatment, virtual simulation of surgery, but also to enable navigated surgery, to design and manufacture an individual implant using rapid prototyping methods.

Computer-assisted medicine also aims to minimize the risks, time and costs of the treatment, prepare in advance for possible complications, and ensure maximum accuracy, functionality and reliability of the resulting replacement. Implant stress simulation concerning its placement and fixation can become a helpful aid (Figure 2.5).
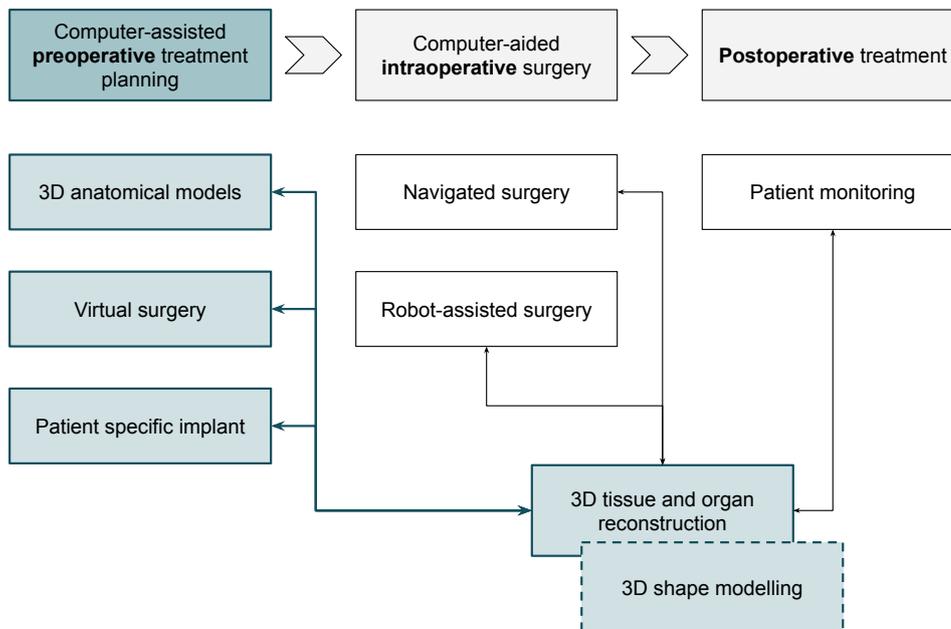
Figure 2.2: 3D shape modelling is essential for evolving all areas of computer-assisted planning of medical treatment. The necessary step is creating an anatomical 3D model, typically from a CT scan, which does not necessarily mean we need a computer with a deeper understanding of the tissue shape in the population. A technician can manually annotate (segment) the data and create the 3D model. However, a better understanding of the 3D shape of tissues and organs in the population provides a much more effective tool for faster and less tedious, more accurate model creation, automated repositioning of bone fragments into the anatomically correct position, design of implant shape, or navigated surgery.

2.2 Computer Planning in Clinical Practice

Research into using 3D modelling and machine learning in medicine and prototyping methods is fundamentally faster than the practical deployment of these methods in specific SW tools for clinicians or laboratory technicians. There is often a long way from an experimental method to practice, and preparing the researched methods for
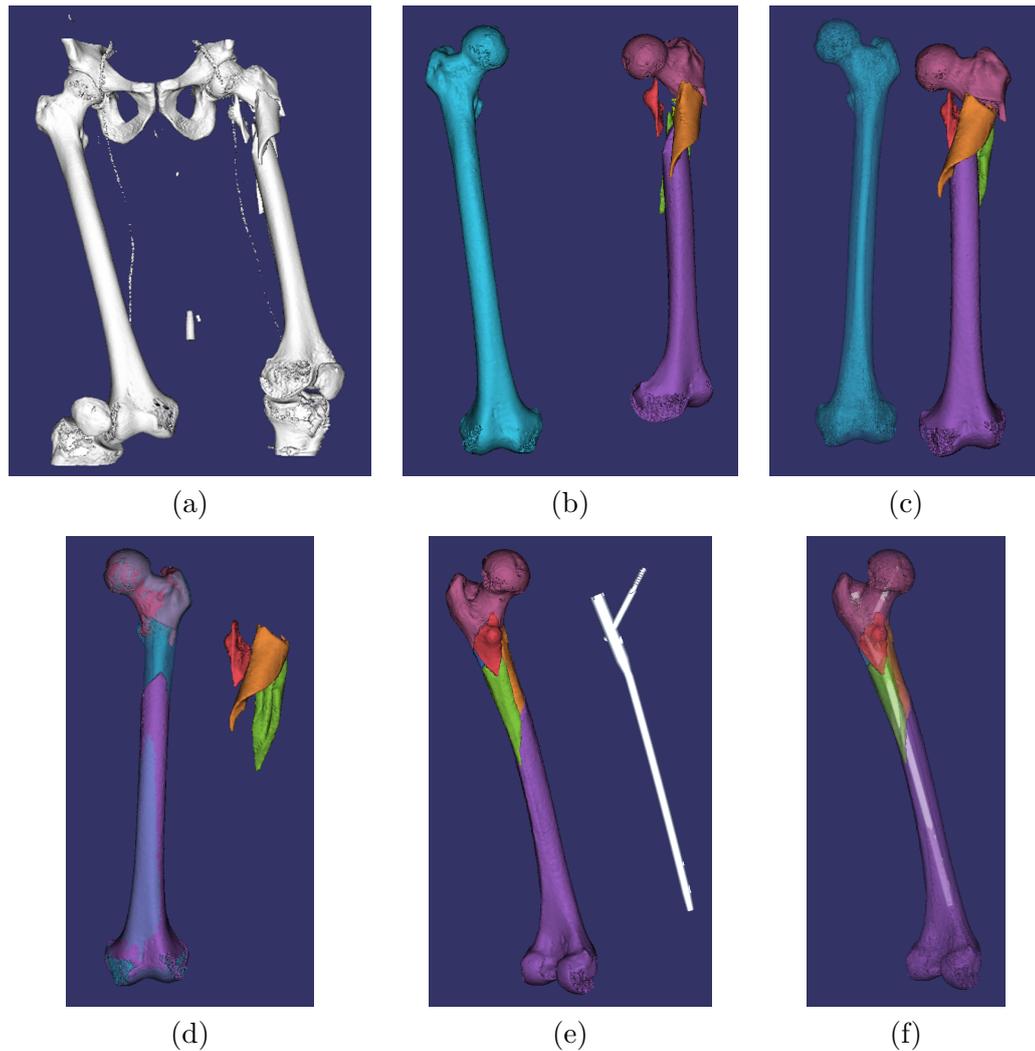
Figure 2.3: Osteosynthesis – virtual fixation of a bone before surgery. 3D surfaces extracted from a CT scan (a) must be separated into individual bone fragments (b). Then, the repositioning into the anatomically correct position takes place, which, in this case, uses a mirrored 3D model of the femur of the other limb. Finally, the appropriate length and diameter of the nail and fixation screws are chosen, preventing complications during the nail's insertion. Image courtesy of TESCAN 3DIM, s.r.o.

clinical practice is not trivial.

The healthcare domain has its specifics. Progress in it may have a considerable impact on people's health. However, it is bound by many standards and the understandable preference for years of practice and well-proven and established procedures. Hence, its inertia is long. Close cooperation with doctors, understanding of medical procedures and their goals, and precise identification of the suitable treatment planning steps for their simplification and automation should be considered absolutely fundamental. Moreover, automated processing should not take control over the treatment from clinicians' hands. Hence, I always thought it was essential to create assistant-like tools for clinicians to help them speed up the diagnosis and make the treatment more accurate.

Two Worlds, One Goal

Digital treatment planning systems connect two different worlds: that of clinicians and that of computer experts. Both worlds speak a different language, and their perspective on assistive computer systems differs significantly.

The barrier between clinical medicine and IT engineering is challenging, and insufficient communication between both worlds leads to the design of difficult-to-use SW solutions that require a lot of manual actions from the user during planning and unnecessary technical knowledge of the principles of how the planning software works. Mutual understanding and willingness to listen and look into the world of "the other"
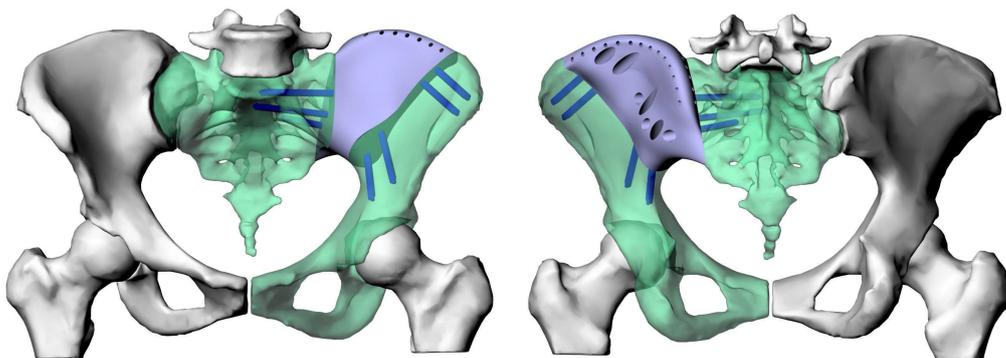
Figure 2.4: Patient specific implant designed on digital 3D anatomical models of the pelvis and hip joint from a CT scan. If the part of the anatomy replaced by the implant is more significant, and it is impossible to use mirroring of the healthy part of the anatomy, 3D shape modelling approaches learned on a large patient population greatly help in designing the suitable implant shape. Image courtesy of TecuMed, s.r.o.

is essential. This requires long-term collaboration and much time that clinicians do not have.

In some medical fields, such as dentistry, clinical laboratories exist that provide technical support and services to clinicians. Currently, it is the technicians of these
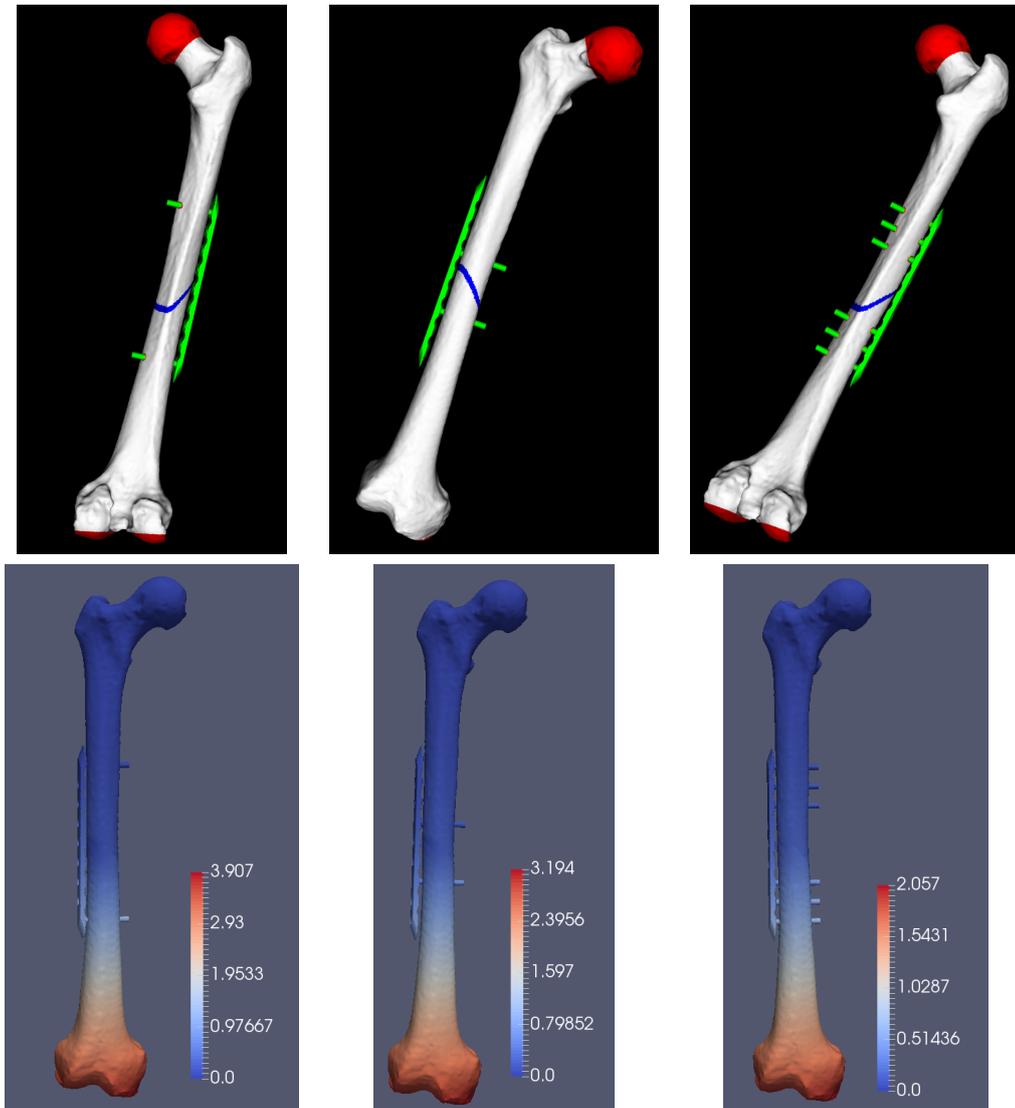
Figure 2.5: Stress calculation using FEM (Finite Element Methods) can be used to compare different variants of reconstruction plate and fixation nails placement what affects the stress on the plate and movement of the bone fragments (bottom line). For the FEM analysis, it was necessary to prepare 3D models of bone fragments from a CT scan and reposition them into the anatomically correct position and define external forces and boundary conditions (top line). Image courtesy of TESCAN 3DIM.

laboratories who are the users of software tools for treatment planning. Clinicians just use the software to assess the case and define the treatment goal and limitations. This trend will probably become more prevalent in other fields as well, especially with the development of tailor-made implants and replacements. Cooperation with technicians is essential for automating the planning process, and their world is even closer to the world of computer experts. However, when developing completely new digital treatment procedures, the knowledge and experience of clinicians cannot be replaced.

No Single Solution is Right

High-grade training data is essential for any statistical data analysis or machine learning method. However, these are challenging to obtain in the medical environment. Medical data are sensitive personal patient data and access to them on a larger scale is problematic. Public datasets are available for some typical tasks but only for some or have too small variability and do not sufficiently cover complex human anatomy and atypical cases. However, these complex and atypical cases are where the clinicians will appreciate the computer-assisted pre-operative planning.

In addition, data analysis is further complicated because there is no single correct solution. Clinicians work in different conditions, have different preferences regarding surgical procedures and materials used, and thus choose more or less different solutions. The complexity of real patient data and considerable variability within the population further intensifies this. A typical example is the task of tissue segmentation in an image, where the difference in manual annotation from individual specialists can be higher than the theoretical accuracy of an automatic method [17, 18], and this divergence has also been explored in the context of human anatomy shape generation [19, 20].

The long-term strategy is the collection of real cases where annotations are created by a semi-automatic method, and users manually correct the results. In practice, however, a revision of each individual sample is always necessary. Concerning saving their time and case-specific planning needs, experts often fix only the most relevant part of the data and leave the rest inaccurately annotated.

For the said reasons, in addition to the quality of the user interface itself, the usability of the software in clinical practice is also affected by the complexity, or rather the simplicity, of manual adjustments of the outputs from automatic methods. The software must give laboratory technicians and clinicians the ability to quickly adapt
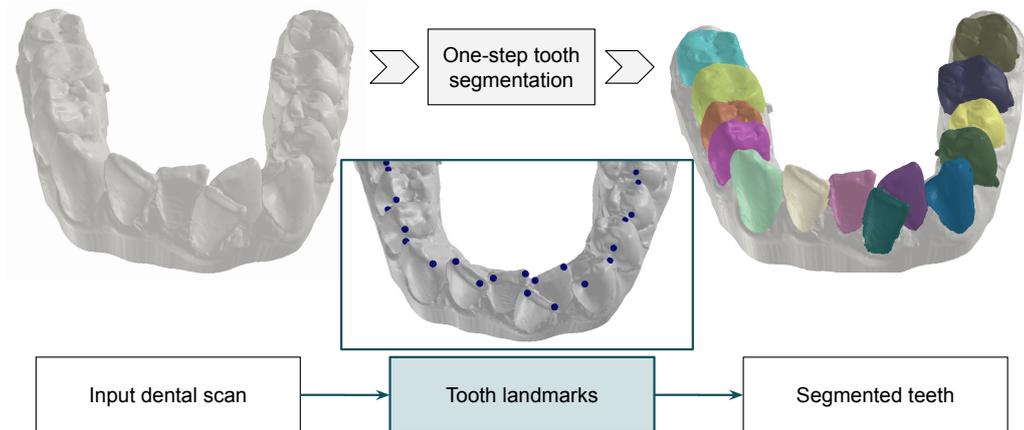
Figure 2.6: Tooth segmentation in orthodontics is the process of outlining and delineating the boundaries of individual teeth on surface dental scans so the isolated individual teeth can be virtually moved. This process usually includes teeth identification (incissors, molars, etc.) and numbering. If the automatic method performs both steps at the same time, the errors are of a more fundamental nature and tend to be much more time-consuming to correct manually (merged teeth, etc.). If we separate the steps, the user can first easily repair the detected teeth annotated with landmarks points and avoid subsequent segmentation failures and complicated fixes to the boundaries. Images are courtesy of Tibor Kubìk [21].

the shape of the implant or dental crown to their individual ideas. The computer assistant must allow easy changes and fine-tuning of details. It is often wise to design the machine data processing methods themselves with such a possibility in mind. It may be better to divide the treatment planning into more steps (see Figure 2.6) and allow the user to correct a partial sub-result and then proceed with the following automatic planning step.

3 Statistical Shape Models for Fractured Bones Reconstruction

This chapter summarizes our research results in using the Statistical Shape and Intensity Model (SSIM) to reconstruct a complete intact patient-specific 3D model of long bones from 2D radiographs (X-ray images) of a fracture, which is achieved by fitting a deformable 3D bone model onto the radiographs (Figure 3.3) of the fragments. The main researchers in this area were my colleagues Ondřej Klíma, who successfully defended his dissertation thesis [22] on this topic in 2022, and Petr Klepárník.

The publication results presented in this chapter resulted from the research project *TramaTech**. The project aimed to develop computer software for rapid pre-operative planning of diaphyseal fractures reduction – computer reconstruction of bone fragments into the correct anatomical position and identification of the optimal treatment and best fitting plate or intramedullary nail.

The role of FIT BUT in the project was the creation of statistical shape models of long bones (e.g. femur, humerus, tibia) for the needs of pre-operative planning, research and development of methods for fitting the shape models to a specific patient and speeding up the fitting using modern GPU architectures.

3.1 TraumaTech Project

The project was focused on creating software for rapid pre-operative digital planning of osteosynthesis (treatment of fractures) in traumatology. It focused on the pre-operative phase and development of a software system for:

- fracture identification and computer reconstruction of bone fragments into the correct anatomical position,

- planning corrective cuts,

---

*TraumaTech – Use of modern image data processing techniques and computer planning in traumatology, grant project supported by Technology Agency of the Czech Republic, TA04011606, 2014 – 2017

- choosing the optimal implant (nail or plate) or modifying standard bone plates by bending, including planning the necessary bending angle, shortening or other changes to the plate,

- and selection of osteosynthetic material and introduction of auxiliary fixation wires and screws.

The surgeon should have been able to plan the overall workflow already in the pre-operative phase to improve the accuracy of the osteosynthesis, prepare for possible complications, reduce the operating time and reduce the radiological burden on the patient and the operative team.

The project benefited from the collaboration of traumatology experts (University Hospital Ostrava), 3D software developers (3Dim Laboratory s.r.o.[†]), researchers in shape modelling (FIT BUT) and experts development and manufacturing of traumatological implants (Medin, a.s.).

At the time of the project, from 2014 to 2017, pre-operative planning was mainly based on Computed Tomography (CT) and used in orthopaedic and neurological surgery. The use of CT imaging in trauma procedures was marginal. The possibilities of computer-assisted pre-operative planning in traumatology were limited. The vast majority used simple 2D planning in one or two X-ray images because the surgeon lacks the time and space for a more complex examination. The complexity of existing systems based on the use of CT data and the need for time-consuming pre-operative planning discouraged surgeons from using these technologies, as time is a critical factor in traumatology.

So, traumatological treatment of fractures and deformities was usually planned and prepared on the basis of X-ray images, where the surgeon decides on the optimal approach based on his experience and intuition. Only in the case of more complex fractures an examination using computed tomography is first performed. For simple operations and polytrauma (simultaneous injury to at least two body systems), this approach is sufficient. However, this established diagnostic process does not allow fast and accurate planning concerning the trends are directed towards using instruments related to individual anatomical parts of the skeleton with an emphasis on optimal conditions for forming bone tissue and healing, minimal damage to soft tissues during the surgery, and reduction of infection risk in open operations.

The designed digital planning system used a combination of traditional diagnostic approaches to treating fractures and bone defects – X-ray examination and, in more

---

[†]The company's successor in the field of software development is TESCAN 3DIM, s.r.o.

complex cases, CT examination, with the possibility of virtual relocation of bone fragments to their correct position and planning of the osteosynthesis. Madeja et al. published a case study using the CT-based pre-operative planning of comminuted scapula fracture [23].

The fundamental innovation consisted of creating 3D statistical shape and intensity
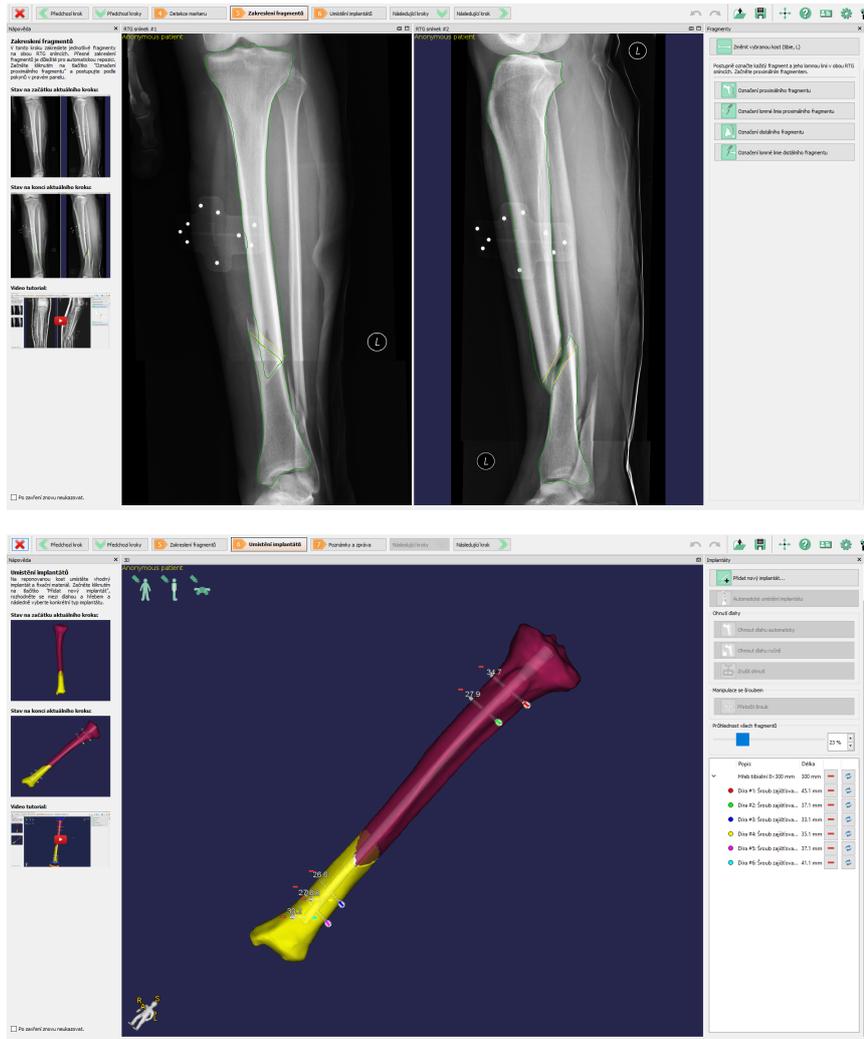
Figure 3.1: Demonstration of virtual osteosynthesis in TraumaTech software using AP and LAT radiographs. First, two X-ray images are taken with a calibration marker with contrast balls, which are easily automatically detected and serve to calibrate both images spatially. Individual fragments and main fracture lines are manually drawn in the X-ray images. The result of the 2D-3D registration of the SSIM model is a 3D model of the repositioned fragments in the anatomical position. Further treatment planning takes place in 3D; an intramedullary nail was used in this case.

models of long bones and developing methods for their adaptation to a specific patient using two calibrated X-ray images (Figure 3.1). The goal of this combination of accurate 3D modelling while acquiring 2D X-ray images was to increase the accuracy of pre-operative planning while maintaining the speed of traditional examinations using X-ray images. The planning in 3D would allow for a much more accurate assessment of the situation, repositioning of bone fragments, and comparing various implant placement options and implant fixation screws, where the goal was to prevent repeated surgical interventions and corrections.

3.2 Statistical Shape Intensity Models

Statistical Shape Models (SSM) allow a quantitative description of the shape of the studied anatomy within a given population. They are created based on a set of training data – real patients. Shape models are typically created by first obtaining training 3D models by segmenting CT scans and annotating them (manually or automatically) with a set of landmark points. Subsequently, the models are aligned using Generalized Procrustes Analysis (GPA) [24], while the models' size remains unchanged. GPA aligns all models to a chosen reference one, computes the mean shape, sets the reference to the mean shape, and repeats the alignment. A linear model describing the shape variability of femoral bones can be obtained by applying the Principal Component Analysis (PCA) [25]. The whole pipeline is shown in Figure 3.2.

Statistical models then make it possible to estimate the anatomically correct shape of the damaged part by optimising the shape parameters of the model and its pose. However, statistical shape models also allow estimation of the 3D shape of the tissue from a set of plain X-ray images using the so-called elastic 2D-3D registration [4, 8, 10]. The result is a lower radiation exposure for the patient, a reduction in costs and, last but not least, the possibility of 3D planning even in cases where only an X-ray examination is performed as a standard.

The 2D-3D reconstruction pipeline, schematically indicated in Figure 3.3, is built as a numerical optimisation. In each iteration, Digitally Reconstructed Radiographs (DRR) are rendered from the shape and intensity model. Differences between the original X-ray images and the corresponding DRRs are evaluated using an image similarity measure. The shape model's initial pose and shape parameters are then adjusted to minimise the dissimilarities between the original and the rendered images. The patient-specific bone model is reconstructed when the similarity between the DRR and X-ray images is maximal.
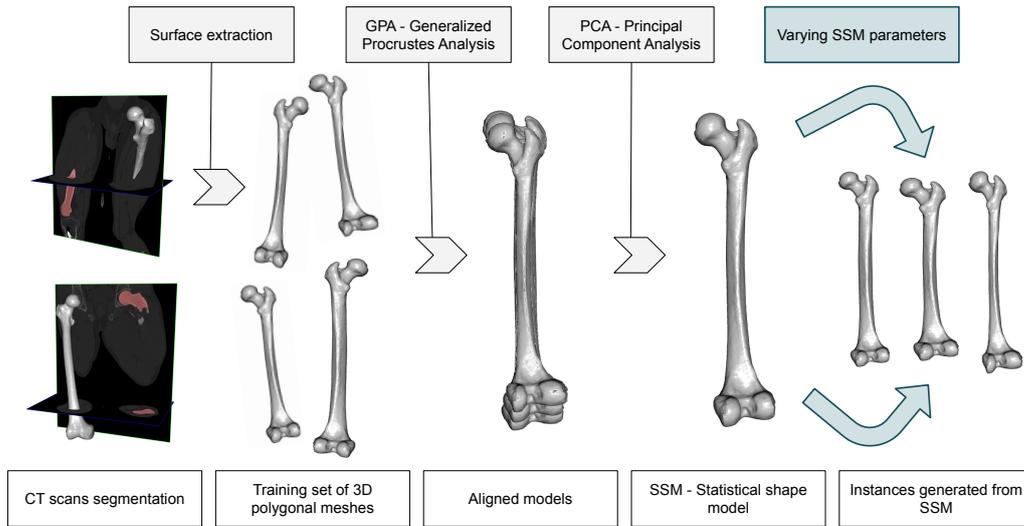
Figure 3.2: Construction of a femur statistical shape model. First, polygonal models are extracted from segmented CT scans. The models are brought into correspondence, aligned using Generalized Procrustes Analysis (GPA) and analysed using Principal Component Analysis (PCA). Model instances corresponding to different values of the shape modes are shown on the right.

There are two categories of 2D-3D reconstruction approaches. The first category comprises methods based on polygonal mesh SSM, which can reconstruct only the bone shape. These methods rely on bone silhouettes extracted from the original X-ray images and the SSM. Intensity-based methods represent the second category. These work directly with the original X-ray images' pixels and involve an appearance model; therefore, beyond the bone shape, they can reconstruct bone densities and anatomical features such as compact and spongy bone. On the other hand, constructing a plain shape model requires considerably less effort than training an appearance model.

Yao and Taylor proposed a novel intensity-based method for reconstructing pelvic and femoral bones named Statistical Shape and Intensity model (SSIM) [26, 27]. Their SSIM model is based on a 3D triangulation of the bone interior and describes bone densities inside its elements using analytical functions that allow efficient manipulation of the bone geometry compared to the voxel-based appearance models. Yao's work was continued by Sadowsky [28, 29], who focused on the effective rendering of a virtual X-ray image – Digitally Reconstructed Radiograph (DRR). Sadowsky replaced the numerical evaluation of integrals of the rays intersecting the SSIM by the closed formula solution and involved the GPU acceleration.
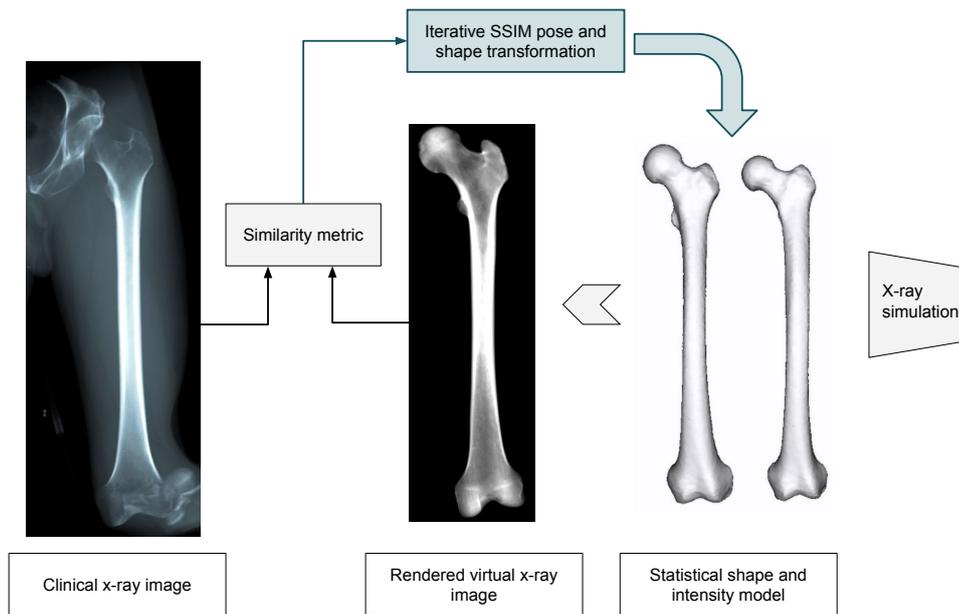
Figure 3.3: Reconstruction of a three-dimensional femur model from an X-ray image. 2D-3D registration is based on minimising the reprojection error between the original and the synthetic image (DRR), artificially rendered from a statistical shape and intensity model.

3.3 Summary of Included Papers

The following paper proposes a first version of a fast GPU accelerated femoral bone reconstruction method based on fitting its statistical shape and intensity model onto a pair of calibrated X-ray images usually taken from the anterior-posterior (AP) and lateral (LAT) views[‡]:

> Klima, O., Kleparnik, P., Spanel, M. & Zemcik, P. *Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation* in *Proceedings of the SPIE* (eds Gimi, B. & Krol, A.) **9788.** Cited by (Scopus): 8; (Mar. 2016), 97880F.

Yao's SSIM appearance model has been adopted, which describes the shape variability of the femur using a Point Distribution Model (PDM) [30] and a reference tetrahedral mesh of the bone (see Figure 3.4). Virtual X-ray images are rendered by a projection of the SSIM appearance model following Ehlke's GPU accelerated approach [7].

[‡]This and all following publications are presented as edited and shortened versions of the papers.
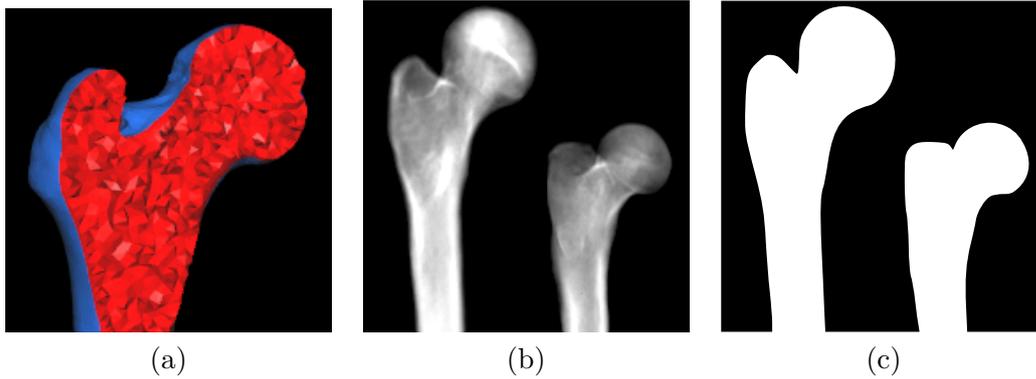
Figure 3.4: A cross-section of a tetrahedral model of the proximal femur (a). Digitally reconstructed radiographs rendered from instances of the shape and intensity model (b) and binary masks rendered from the same instances of the point distribution model (c).

The article proposes two principal variants of the optimisation:

- **The intensity-based registration** is performed by minimising differences between the original radiographs and the DRRs rendered from SSIM. As CT imaging has higher X-ray energy than plain radiography, the X-ray and DRR images may differ in contrast of the corresponding tissues. Therefore, we involve the Normalised Mutual Information (NMI) similarity measure, commonly used in inter-modal registration.

- **Black & White (BW) registration**, in contrast to the intensity-based registration, involves a plain shape model instead of the SSIM. The DRR images are replaced by binary masks, as shown in Figure 3.4. The binary masks must be extracted from the original radiographs as well.

The registration is formulated as a non-linear least squares problem, allowing for the involvement of Levenberg-Marquardt (LM) optimisation, which leads to a significantly faster and more reliable 2D/3D reconstruction method that requires a lower number of iterations (Figure 3.5) to converge when compared to Elke's gradient-descent based optimisation approach.

The reconstruction accuracy was evaluated using the symmetric Hausdorff distance between reconstructed and ground-truth bones using a synthetic dataset rendered from CT scans and real X-ray images of phantom bones whose relative pose was calibrated using a custom contrastive marker.
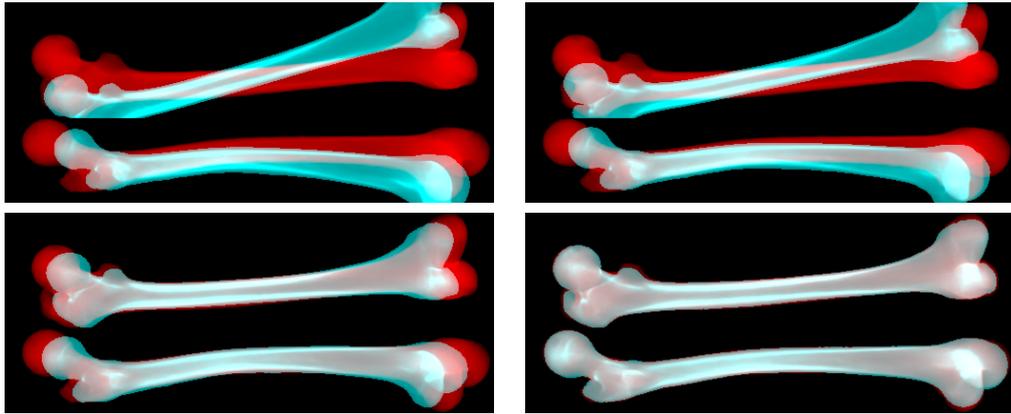
Figure 3.5: Illustration of the intensity-based model registration. The figure shows the first 50 iterations. An approximate initial SSIM pose is essential for proper convergence.

The best accuracy has been achieved by the B&W method, with a measure targeting the optimisation on similarities of local bone features. Here, the Levenberg-Marquardt method updates the optimised parameters with respect to derivatives of individual pixels of the X-ray and DRR binary masks instead of derivatives of the global similarity measures, leading to a very accurate registration method. However, the BW method requires precise X-ray image segmentation what requires either a manual annotation or an automatic method.

The accuracy of the intensity-based method was 1.18 ± 1.57 mm on average, and the registration took 8.76 seconds on average. The most significant advantage of the proposed least squares registration formulation is that it allows a straightforward extension to the multifragment 2D/3D reconstruction.

In the following study, performance of the previously proposed deformable 2D/3D registration approach based on the Levenberg-Marquardt optimisation with methods exploiting Covariance Matrix Adaptation (CMA) [31] and Covariance Matrix Self Adaptation (CMSA) [32] evolution strategies is compared:

> Klima, O., Chromy, A., Zemcik, P., Spanel, M. & Kleparnik, P. A Study on Performace of Levenberg-Marquardt and CMA-ES Optimization Methods for Atlas-based 2D/3D Reconstruction. *IFAC-PapersOnLine* **49.** 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016, 121–126. ISSN: 2405-8963 (2016).

Besides 2D-3D registration approaches based on numerical optimisation methods like the Ehlke's gradient-descent algorithm, strong attention has been paid to the gradient-free evolution strategies represented by the CMA-ES method proposed by Gong and Abolmaesumi [6, 5, 33]. The main benefit of the CMA-ES family of methods is that there is no need for the Jacobian matrix numerical approximation during the registration.

The comparison of different optimisation methods focused on both the robustness and the speed. The results were obtained using a large-scale data set of synthetic X-ray images. Certain parts of the registration pipeline (DRRs rendering, etc.) were accelerated using graphics hardware to speed up the reconstruction process.

The reconstruction based on the LM optimization is several times faster than CMA-ES and CMSA-ES based methods while reaching similar accuracy. The median time of
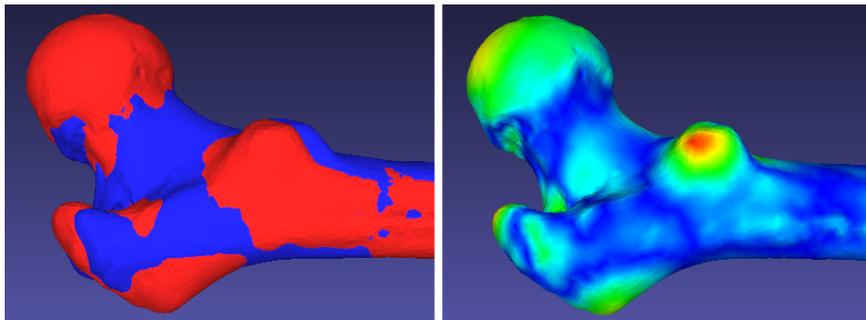
Figure 3.6: A proximal part of the femoral bone reconstructed with a mean error of 1.12 mm. The reconstructed bone, highlighted in red, is aligned with the ground-truth bone model obtained from the CT image (left). The reconstruction error is visualised by a heatmap; the red colour shows places with the highest error of 5.68 mm (right).

the registration using the LM method was more than 16-19 times lower in comparison to the CMSA-ES method and approximately 6-7 times lower in comparison to the CMA-ES method.

The median error of the proposed method was 1.12 mm, and the median reconstruction time was 7.2 s. The median time reached by the CMA-ES and CMSA-ES methods was 48.5 s and 138.5 s, respectively.

Most studies dealing with 3D pre-operative planning based on 2D stereo radiographs have only worked with uninjured bones, which limits their possible planning applications. In the subsequent paper, we proposed a method for reconstructing 3D polygonal models of the intact bones from the calibrated radiographs of diaphyseal fractures of long bones:

Klíma, O., Madeja, R., Španel, M., Čuta, M., Zemčík, P., Stoklásek, P. & Mizera, A. *Virtual 2D-3D Fracture Reduction with Bone Length Recovery Using Statistical Shape Models* in *Shape in Medical Imaging* (eds Reuter, M., Wachinger, C., Lombaert, H., Paniagua, B., Lüthi, M. & Egger, B.) (Springer International Publishing, Cham, 2018), 207–219. ISBN: 978-3-030-04747-4.

The first attempt to reconstruct injured bones using statistical atlases was proposed in a study focused on the reduction of multi-fragment fractures of the distal radius by Gong [5]. The reconstruction and the fracture reduction were achieved simultaneously by a 2D-3D registration of a single statistical appearance model of an intact distal radius into individual fracture segments. The method was evaluated in silico using simulated fractures, concluding that the atlas-based reconstruction may provide a more accurate distal radius template than the conventionally used mirrored model obtained from the contralateral limb.

A later study by Markelj et al., using a similar principle, focused on diaphyseal fractures of the long bones of the lower limbs [2]. It aimed to determine the rotation alignment between the proximal and distal fragments along the longitudinal axis. In addition to the rotation angles, the study considered the reconstruction of surface
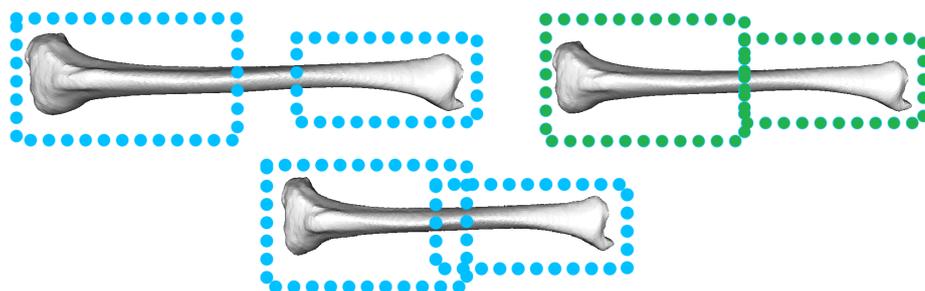
Figure 3.7: Relation between vertex assignment to fragment regions and the resulting bone length. If vertices in the middle of the shaft are not assigned to any of the two fragments, the bone is too long (left). If some vertices are assigned to both the fragments, the bone is too short (middle). The correct length is when each vertex gets assigned to exactly one bone fragment (green).
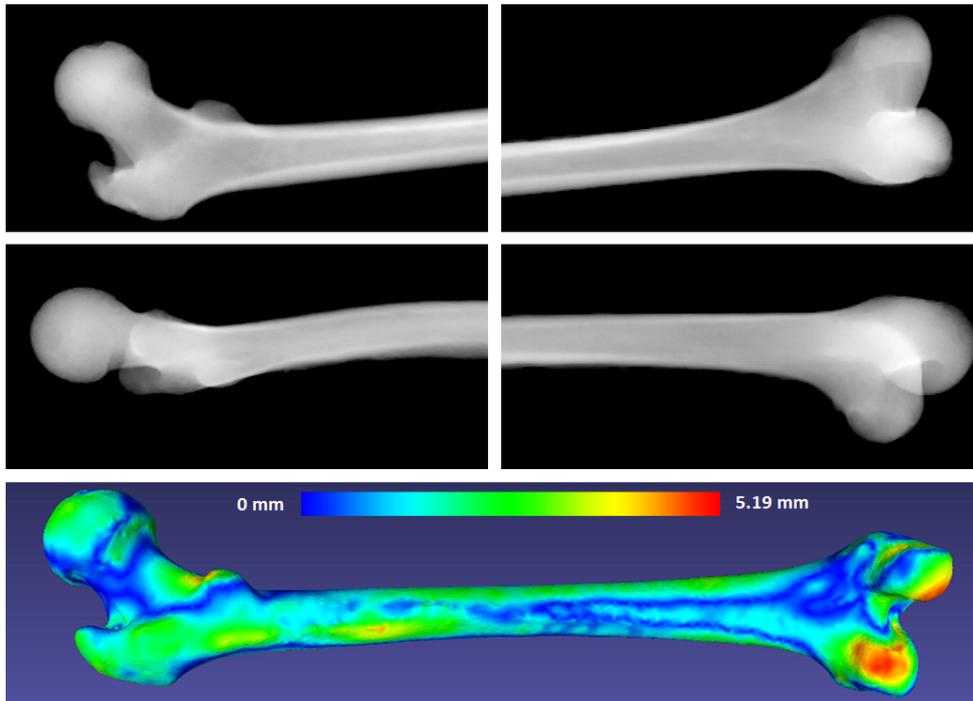
Figure 3.8: Accuracy of the sample case reconstruction. A sample test case of stereo radiographs with simulated transversal femoral shaft fracture. The heatmap shows the differences between reconstructed and ground-truth surfaces, evaluated using the symmetric Hausdorff distance. The areas with the highest deviation are coloured red and reach 5.19 mm.

models of the individual fracture segments. However, the approach could not perform virtual fracture reduction and provide a model of the intact bone, as the method could not determine the correct length of the target bone. Moreover, the shape model had to be divided into fragments in advance without further refinement during the registration process.

The work's main contribution is the ability to recover an accurate length of the intact bone. Unlike Markelj et al., the division of the statistical shape model into segments is performed automatically by the registration, enabling optimisation of the shape model length. The reconstruction is based on 2D-3D registration of a single statistical shape model onto the pairs of radiographs of each individual fragment, performed simultaneously with the virtual fracture reduction.

The method works with simple displaced transverse, oblique and wedge fractures when the injured bone is split into two main fragments. Each fragment is captured in its region of interest, forming a pair. The idea of the bone length recovery is to assign

Figure 3.9: Physically reduced dry cadaveric femur involved in the study (left) and sample AP and LAT radiographs of cadevers. Image courtesy of Ondřej Klíma[22].

each vertex of the shape model to only one of the main fragments. Consequently, each vertex should be re-projected in precisely one of the two regions (Figure 3.7).

The method was tested on a synthetic data set containing 96 virtual fractures (Figure 3.8) and on real radiographs of dry cadaveric bones (Figure 3.9) suffering perimortem injuries (bone damage occurred at or near the time of death without any evidence of healing). The accuracy was evaluated using the Hausdorff distance between the reconstructed and ground-truth bone models. On the synthetic data set, the average surface error reached 1.48 ± 1.16 mm. The method's accuracy is comparable with single-fragment reconstruction approaches, presented by Baka et al. in brief summary [1].

The main contributions of the presented series of papers can be summarised as follows:

- A new method for reconstructing 3D polygonal models of the intact bones from the calibrated radiographs of diaphyseal fractures of long bones based on the Levenberg-Marquardt optimisation method.

- GPU-accelerated rendering of digital radiographs from statistical shape and intensity models, which provides fast convergence and state-of-the-art reconstruction accuracy when combined with the LM optimisation.

- Extensive testing of reduction accuracy with both virtual fractures generated from CT scans and cadavers.

Figure 3.10: X-ray calibration marker and annotation of bone fragments in stereo radiographs. The user approximately annotates fragments' boundaries (green) and subsequently the fracture lines (yellow).

3.4 Insights and Remarks

The method has been integrated into pre-operative planning software. The mutual position of the stereo radiographs is determined using a calibration marker (Figure 3.10), usually attached to the lower limb plate. The shape model is initially aligned with the longitudinal axes of the fragments reconstructed in 3D from the user's bone fragment annotations on radiographs. The user must also annotate the fracture detachment sites as simple lines in the 2D images.

As part of the TraumaTech project, X-ray planning was tested with clinicians. The goal was to perform a virtual reduction of the bone, plan the osteosynthetic material, and then compare it with the material used in real life. The clinicians determined an acceptable deviation for the length of nails to be ±10 mm, for screws ±4 mm. The results showed that the deviation was minimal in approximately 1/2 of the cases, but it was unacceptable in 1/4 of the cases. The accuracy of the intact bone reduction method itself seems to be sufficient, but in practice, it is strongly influenced by the precision of the radiograph calibration when the marker must always be in exact place when taking the images – it is necessary to fix it as best as possible to the patient's limb. Further errors are then caused by the inaccurate drawing of fragments' boundaries and their fracture lines. It would be necessary to automate the annotation of X-ray images as much as possible using pre-trained detectors and fundamentally improve the calibration process.

Statistical shape models are a well-described mathematical tool for modelling shape and its variability in a population. Their clear definition and interpretability is their most significant advantage over newer deep learning methods discussed in Chapter 4. For some tasks like strength analysis using FEM, the precise structure of their shape representation by the point distribution model can also be an advantage.

On the other hand, because they work with the PDM, they require accurate sample alignment and matching, typically using anatomical landmark points. From a practical point of view, it is necessary to supplement the use of SSM with robust machine learning methods for accurate landmark point detection for the initialisation of the fitting process, which optimises the model parameters for a specific sample (patient).

This process of parameter optimisation turns out to be problematic in practice. Typically, this is a global optimisation over all model parameters, and it is challenging to adjust the minimised metric and set the boundary conditions to meet the specific needs of a particular task, often also considering the character of the dataset. It requires deep knowledge of the optimisation methods. Often, even a small change in the requirements for the resulting reconstruction leads to repeated tuning of the whole optimisation process. The practical fine-tuning of SSM optimization is very difficult and tends to be fragile.

4 Deep Learning Methods for 3D Shape Analysis

In the last decade, deep learning and convolutional neural networks have driven dramatic progress in various computer vision tasks. The combination of convolutional and deep architectures, efficient learning algorithms, a large amount of training data, enormous computing power, and the flexibility of deploying different model architectures make it possible to create high-quality predictive models for almost any task. Convolutional neural networks are the basis of state-of-the-art methods in the never-ending list of tasks such as image classification, semantic segmentation, object detection, person identification and pose estimation of non-rigid objects [34] through low-level information extraction tasks [35] like the estimation of depth, illumination, surface properties; to pure image processing tasks such as super resolution [36], noise removal [37, 38]. The current big bang in the field of generative models then moves the task domain much further than we could have imagined a few years ago. It was so natural that we started looking into using neural networks for tasks related to 3D shape analysis in medical data processing.

This chapter presents the results of our research in this area. The main topic is deep learning for automatic reconstruction of the defective part of the human skull and designing a patient-specific cranial implant. The main driver of this research was Oldřich Kodym, who defended his dissertation [39] based on this research in 2022. We collaborated with TESCAN 3DIM s.r.o. and TecuMed, s.r.o.* companies on the automatic design of cranial implants in the form of contract research because the automatic reconstruction of the skull shape represents a notable time saving for the technician who models the implant using 3D SW tools.

The second research area discussed in this chapter is the analysis of dental surface scans for tooth detection and tooth margin line segmentation. Tibor Kubík, who is now continuing his research as a doctoral student at FIT BUT, devoted himself to this area as part of his bachelor [40] and master [21] thesis.

---

*Formerly TESCAN Medical, s.r.o.

4.1 Skull Shape Reconstruction for Cranioplasty

Cranioplasty is the surgical repair of skull to correct a deformity or defect of the skull. These defects and deformities can happen as the result of a traumatic injury or neurosurgery, such as a craniectomy which is the most commonly performed surgery for brain tumour removal that requires access to the brain. During craniectomy, a specific part of the skull is resected and eventually replaced with a cranial implant. Examples of cranial implants are depicted in Figure 4.1.

For the cranioplasty, patient-specific implants are often used. When designing the implant, the correct skull shape reconstruction is critical to restore the protective and aesthetic function of the skull and also fit very precisely along the border [41, 42].

Current state-of-the-art methods usually involve using a patient CT scan to design the implant pre-operatively and then 3D printing of the result using bio-compatible materials such as Titanium, Porous Polyethylene or Polyether ether ketone (PEEK) [43, 44]. Alternatively, implants can be cast in a 3D-printed mould from bone cement, which can be loaded with antibiotics to decrease the risk of infection [45]. Such approaches lead to a reduction of operative time and improved patient results [46].

Provided that a precise enough anatomical 3D model is obtained from the CT data after its tissue segmentation, implant design is the most challenging and tedious task for a laboratory technician. In unilateral defects, techniques based on mirroring the healthy part of the skull are often used in combination with Computer-Aided Design software (CAD) [47]. However, the assumption of a perfectly symmetric skull does not hold in most cases, and manual corrections are often required. Moreover, estimating the correct shape is much more challenging in the case of bilateral defects. Recent methods addressing these challenges aim to be entirely or mostly automatic and to

Figure 4.1: Examples of cranial PEEK[†] implants designed by laboratory technicians. The implant can also be made from other materials, such as bone cement or 3D printed from titanium. Image courtesy of TecuMed, s.r.o.

reconstruct an arbitrary part of the skull. These methods include statistical shape models [48] and convolutional neural networks [49, 50].

Several conditions must be met to deploy reconstruction methods into the clinical practice. The symmetry of the skull should be preserved as well as possible, including in cases where the patient's skull itself is partly asymmetric and where the defect reaches partly into both sides of the skull. Second, the automatic reconstruction should fit the defect borders very precisely.

4.2 Deep Neural Networks for 3D Geometry Analysis

The use of deep neural networks for 3D geometry analysis is a research topic that has received more and more attention in recent years. The typical approach is to naturally extend Convolutional Neural Networks (CNNs) into the 3D space, represent the input shape as a regular binary voxel grid and train a 3D CNN with an auto-encoder architecture to output the completed binary shape [51, 52]. Unfortunately, such a natural extension of core deep neural network operations like the convolution from 2D to 3D domain introduces additional computational complexity. Having the volumetric grid data as input to deep neural networks has a considerable drawback in computational time and memory requirements. These approaches cannot be utilized for larger volumes due to GPU memory limitations.

The methods currently used by researchers and published in the literature can be divided into the following categories:

- **Coarse-to-fine volumetric strategies** – utilize the 3D CNN to estimate a coarse shape only, which is further refined in a post-processing step [53, 54] or with another 3D CNN working at higher resolution but with a smaller context [13] or more complex octree like structures [55].

- **Multi-view approaches** – render the 3D shape into several 2D images from different viewpoints, process the resulting 2D views with the standard image vision architectures, and finally combine the different viewpoints' results [56, 57]. Su et al. presented the multi-view CNN idea [56] obtaining state-of-the-art results on 3D classification task. And the multi-view approach was later on used to identify feature points on facial surfaces [58]. Their method reached state-of-the-art results in feature point detection on facial 3D scans while decreasing the GPU memory requirements needed for 3D processing. Additionally, they

also proposed a consensus method to find the final estimate, which combines the least-squares fit and RANdom SAmple Consensus (RANSAC) [59].

- **Point-based solutions** – represent the first category of methods that exploit different data representations [60, 61, 62] and define all the needed operations like the convolution and pooling on top of point clouds.

- **Mesh-based solutions** – work directly with the irregular triangular meshes redefining the convolution and pooling either on faces or edges as the main element [63, 64, 65, 66].

- **Graph networks** – relies on learning over graph structures. The convolution is defined spatially on the graph or in a spectral domain [67, 68].

Our papers described below mainly rely on the two first groups of methods: (i) coarse-to-fine volumetric grids and (ii) the multi-view approach. Although the mesh-based and graph networks are advancing significantly and are up-and-coming, their computational complexity still limits their ability to process input 3D meshes with a sufficiently large context and simultaneously generate shape details, e.g. at the cranial implant boundaries. We are investigating the possibilities of using these methods but do not have publications in this field yet.

Besides the challenge of efficient 3D shape representation suitable for deep learning, an exciting research direction is represented by Generative Adversarial Networks (GANs) or variational auto-encoders [54, 69], suggesting that the shape completition task has multiple correct solutions. Other authors have also raised this issue in the context of anatomical shape reconstruction [19, 20], where inter-expert variability of the resulting shape is also considered. However, the argument that the variability of the output should be enforced at the cost of precision measured against the original shape conflicts with the current literature, where the original shape is usually considered the ground-truth. Proving the benefit of this approach is not easy without a more extensive study directly with users.

4.3 Summary of Included Papers

The following paper was the first in which we entered the topic of skull shape reconstruction using deep learning techniques for cranial implant design. The work presented in the paper aimed to address two challenges: (i) design a fully automatic 3D shape reconstruction method that can address diverse shapes of real skull defects in various stages of healing and (ii) provide an open dataset for validation of anatomical reconstruction methods on a set of synthetically broken skull shapes[‡]:

> Kodym, O., Španěl, M. & Herout, A. Skull shape reconstruction using cascaded convolutional networks. *Computers in Biology and Medicine* **123.** Cited by (Scopus): 9; 103886. ISSN: 0010-4825 (2020).

Statistical shape models combined with geometric morphometrics have been studied extensively by Fuessinger et al. in the context of skull reconstruction [70, 71, 48], achieving an average surface error of 0.47 mm for defects of the parietal and temporal area and 0.75 mm for small mid-facial defects, as measured on simple synthetically created defects against the original bone shape. While providing a good reconstruction result, these methods rely on a clean, well-defined defect border, which is rarely the case in real patient cases with complex fractures in various stages of healing and bone resorption. In the first attempt to use deep learning-based shape completion for skull reconstruction, Morais et al. has presented an approach that uses convolutional denoising auto-encoders [50], although it only operates at a very coarse resolution.

In this work, we proposed a cascaded CNN architecture for the estimation of a high-resolution 3D anatomy shape conditioned on the input defective skull. The method can successfully reconstruct defects reaching into both sides of the skull as well as into more complex anatomical regions such as orbitals. To the authors' best knowledge, this was the first deep learning-based method of 3D shape reconstruction that reached a high enough resolution to be clinically viable for the skull reconstruction task. The architecture uses a combination of two CNN models with a 3D U-net [72] backbone. The individual models differ from the original 3D U-net. Figure 4.2 gives and overview of the architecture. Further details can be found in the full paper. Each of the models operates on a different resolution. During inference, the first model provides enough contextual information about the overall shape of the defective skull while the second model can ensure precise contact at the defect border.

[‡]This and all following publications are presented as edited and shortened versions of the papers.
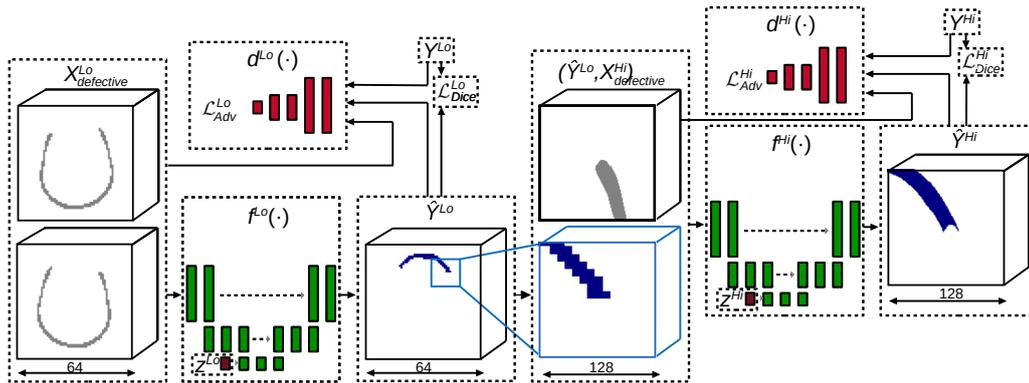
Figure 4.2: Overview of the 3D CNN cascade for skull reconstruction. The symmetrized low-resolution input X^{Lo} is fed into the first model f^{Lo} to produce the missing shape estimation which gets ubsampled and concatenated to the high-resolution input X^{Hi} and fed into the second model f^{Hi} to produce the final high-resolution missing shape. The discriminator CNNs d and latent vectors z are only used in the generative model.

As the ability to preserve the anatomical symmetry is a critical part of the skull defect reconstruction, we concatenated a sagitally flipped copy of the volume to the low-resolution CNN input. This makes it easier to propagate the symmetry information using convolutional kernels and skip connections of the U-net architecture.

To address the issue of multiple possibly correct solutions, we also experimented with a generative version of the proposed model. To make the described reconstruction model generative, the adversarial loss function in the form of a discriminator d has been added. The discriminator allows the model inject a random latent vector into the reconstruction CNNs. We concatenate the random latent vector with the bottleneck tensor of both CNNs as shown in Figure 4.2. The discriminator CNNs have the same architecture as the encoder part of the reconstruction CNNs with additional dense layers that output the discriminator scores. Given a combination of the defective skull shape and the missing anatomy shape, the discriminator is trained to assign a low score to the ground-truth missing shape and a high score to the reconstructed missing shape at both a low and high resolution.

In order to improve the reproducibility of research in the area of automatic skull reconstruction, we have introduced an open dataset of skull shapes with synthetic defects. The dataset mimics the variability in shape, position and bone resorption present among real patients. We used a public CQ500 dataset [73] as a source of head CT patient data. CT scans were rigidly aligned and the skulls were segmented and
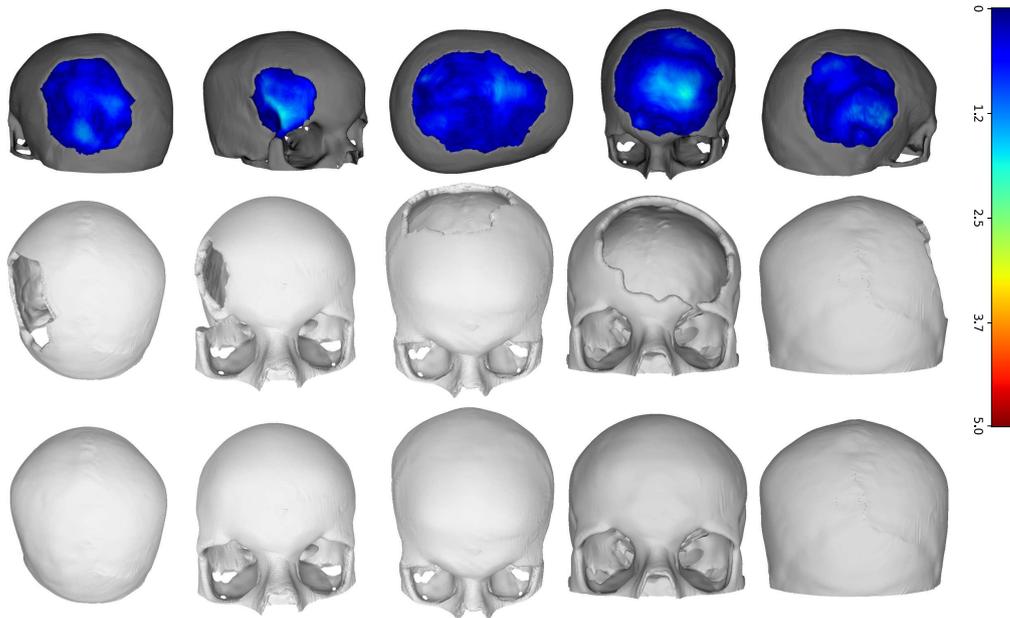
Figure 4.3: Results of the discriminative model reconstructions for parietal, frontal, bilateral and random defects. From top to bottom: Surface error maps, input synthetically broken skulls and reconstructed skull shapes. Most of the reconstructed surface reaches less than one millimeter error in all cases.

saved as 3D binary arrays. 5 different defects were created on each segmented skull with an emphasis on simulating the variability in real defective patient skulls. To allow for structured validation, the defects were categorized into unilateral parietal, unilateral frontal (the orbital area) and bilateral groups. The resulting dataset of 945 defective skulls with ground-truth original shapes along with further details is publicly available as the SkullBreak dataset[§].

The precision of each method was measured as the average symmetric distance between the surface voxels of the output reconstruction and the original anatomy shape, which we considered to be ground truth. We only measured the error on the skull's outer surface because the inner surface is irrelevant for cranial implant design in clinical practice. Although the models presented in this work occasionally produce slightly asymmetric results or fail to avoid some depressions around the defect border, our results show that the proposed method can achieve an overall satisfactory performance in this regard, as illustrated by example reconstructions in Figures 4.3, 4.4.

We first evaluated the discriminative model's performance and the symmetrised

---

[§]https://www.fit.vutbr.cz/~ikodym/skullbreak/
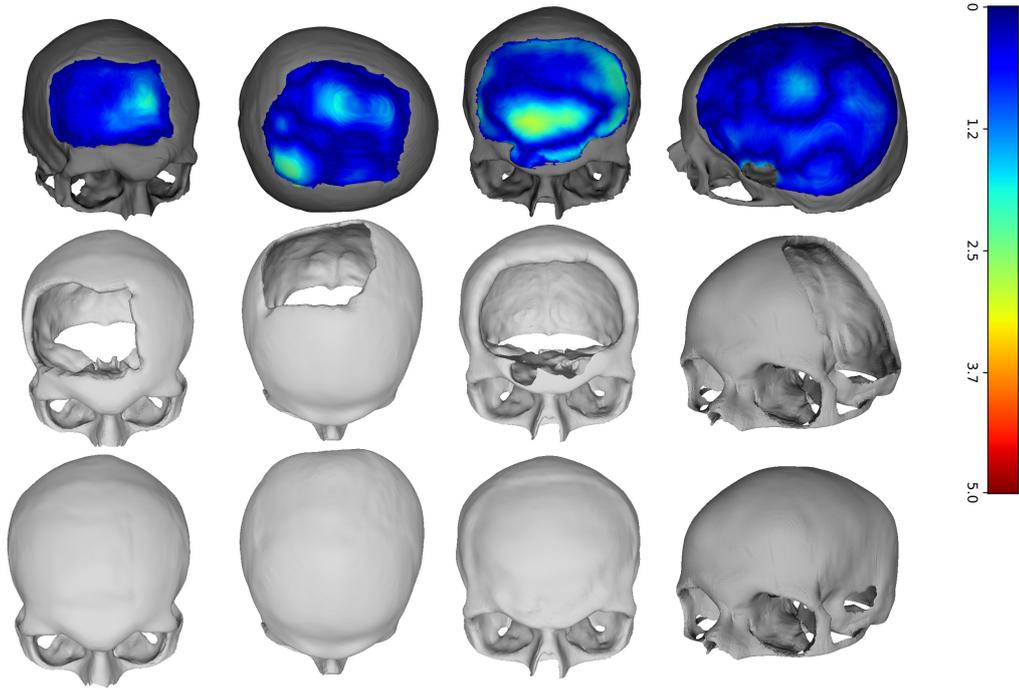
Figure 4.4: Results of the discriminative model for real patient data. Although some reconstruction faults can be seen in the last two cases, the reconstruction is usually correct. The surface distance to the original shape is below one millimeter on average.

input's effect. The model with a simple input could reconstruct each testing skull successfully. The effect of symmetrising input was that the average measured error of the reconstructed unilateral defects dropped from 0.69 mm to 0.48 mm for parietal and from 0.69 mm to 0.60 mm for frontal defects. As expected, the symmetrised input less affected the bilateral defects group. However, the error still slightly decreased since the symmetry partly constrains some bilateral defects. The overall average surface error of the discriminative model with a symmetrised input for the whole testing set was 0.59 ± 0.21 mm. Several examples of the discriminative model reconstructions are shown in Figure 4.3.

We also evaluated the performance of the generative model with symmetrised input and random input latent vectors. The overall average surface error was 0.68 ± 0.28 mm. For each defect group, the error of the generative model was higher than that of the discriminative model. However, since we consider multiple correct reconstructions for a single skull defect, the error measured against the ground-truth shape might not be a good indicator of the method's performance. The generative model allows us

to sample multiple different outputs for a single input defective skull by changing the input latent vectors. Therefore, we also experimented with generating multiple reconstructions and measuring the best-achieved result. The overall average surface error when measuring the best-of-five sampled reconstructions for each testing skull was 0.56 ± 0.21 mm. The results for individual defect groups were similar to the discriminative model. However, the error was reduced in the bilateral group from 0.73 mm to 0.65 mm compared to the discriminative model.

Finally, to evaluate our approach's ability to generalise, we tested the performance of the models trained exclusively on our synthetic dataset on an internal dataset of real defective patients without fine-tuning the model. Both models reconstructed the real defects mostly successfully. However, there was an expected increase in the surface error in both the discriminative model output and the best-of-five generative model output. In some cases, there were also visible faults, such as slight depressions or holes, as illustrated in Figure 4.4. This could be partly attributed to the fact that the real testing patients come from a different geographic location, in which the anatomical variability of the skull is different.

In the context of the state of the art, Fuessinger et al. [71] achieved an average surface error of 0.47 mm when reconstructing unilateral spherical defects of the cranial area with a radius of 5 cm. Compared to the performance of our discriminative and generative models reached 0.48 mm and 0.46 mm average surface error on the unilateral parietal defect group. In contrast, our method does not require any manual cleaning of the defect border, as the CNN model handles the seamless fit of the reconstructed part to the skull. It would be interesting to see the performance of the statistical shape model on more challenging parts of the introduced dataset.

Since the discriminative model outputs reach a lower average surface error than the randomly sampled outputs of the generative model, it can be concluded that it is more suitable for a completely automatic setting. The generative model could alternatively be used in a semi-automatic setting. If the initial reconstruction is not satisfactory for further processing, several subsequent samples from the generative model could be offered to the expert to increase the chance of avoiding falling back to a less efficient conventional workflow.

The following paper describes a method submitted to the MICCAI 2020 AutoImplant Challenge [74], which aims to provide a public platform for benchmarking skull reconstruction methods. The proposed method adapts the discrimative cascaded reconstruction CNN architecture described above [13]. Furthermore, the method is extended by an automatic landmark-based registration and a detail-preserving morphological post-processing step:

> Kodym, O., Španěl, M. & Herout, A. *Cranial Defect Reconstruction Using Cascaded CNN with Alignment* in *Towards the Automatization of Cranial Implant Design in Cranioplasty* (eds Li, J. & Egger, J.) Cited by (Scopus): 4; (Springer International Publishing, Cham, 2020), 56–64. ISBN: 978-3-030-64327-0.

The skull is first aligned using landmark detection, followed by a cascade of low-resolution and high-resolution reconstruction 3D CNN. As the defects in the AutoImplant dataset are generated on a static position inside the data volumes, and the variability in their shapes and positions comes from the variability of the positions of the skulls, reconstructing the shapes with a volumetric CNN model introduces some
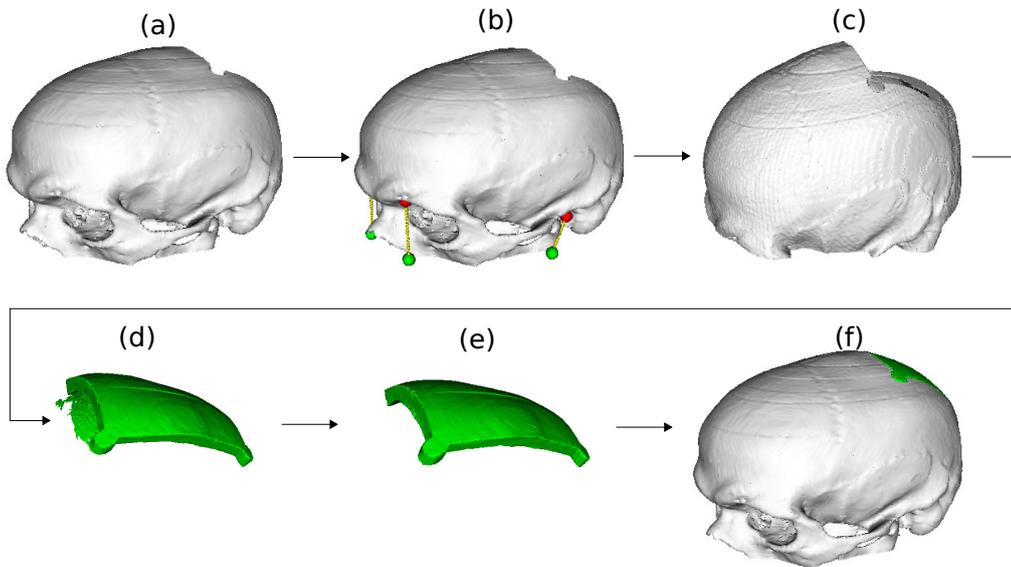
Figure 4.5: Overview of the MICCAI 2020 AutoImplant Challenge skull defect reconstruction method. In the input skull volume (a) 4 landmarks are detected (b). The skull pose and scale is normalized (c) so that the detected landmarks (red) are registered to the reference landmarks (green). Then, the skull is reconstructed by estimating the missing shape (d). Finally, the result is post-processed (e) and transformed back into the original skull coordinates (f).)

difficulties. The reconstruction model needs to implicitly learn rotational and translational invariance, making it cumbersome to exploit the symmetric properties of the skulls. To address this, rigid transformations are applied to normalise the scale and the position of the skulls. The positions of four anatomical landmarks, namely the left and right auditory meatus and left and right supraorbital notch (see Figure 4.5), are used to compute the transform. We trained a simple 3D CNN model for landmark detection with a U-net architecture using the heatmap regression approach [75].

The reconstruction model will occasionally produce outputs that contain noise, such as disjoint objects or protuberances covering the healthy part of the skull as shown in Figure 4.5(d). Assuming the missing shape should only consist of a single compact object, connected component analysis can discard all objects except the largest one. Next, morphological operations can remove any shape protuberances with less than the desired minimum shape thickness while preserving the fine details produced by the reconstruction model along the defect edges as shown in Figure 4.5(e).

The proposed method successfully reconstructs every skull in the standard test dataset, achieving an accuracy of 0.920 DSC (Dice Coefficient) and 4.137 mm in terms of HD (Hausdorff Distance). The proposed skull alignment and data augmentation techniques are general concepts that could be applied to any other reconstruction task. Although we only encountered one failure case in our additional test set, it hints that more defect shape augmentations should be used to increase the robustness of the reconstruction model.

The solution we submitted to the AutoImplant 2020 Challenge was finally ranked 3rd out of 10. The first three places were occupied by deep learning methods, followed by a method using SSM [76].

In the last paper, which is devoted to the topic of automatic cranial implant design, the method based on the cascade volumetric CNNs was further extended by adding multi-branches to allow simultaneous training on two different types of cranioplasty ground-truth data: (i) the skull patch, which represents the exact shape of the missing part of the original skull, and which can be easily created artificially from healthy skulls, (ii) and expert-designed cranial implant shapes that are much harder to acquire Figure 4.6. The idea was to train the overall shape of the implant on the large artificial dataset and to simultaneously train the model on a smaller real dataset to fine-tune the shape of the implant boundary resting on the skull according to the needs of the clinical use and implant manufacturing:

Kodym, O., Španěl, M. & Herout, A. Deep learning for cranioplasty in clinical practice: Going from synthetic to real patient data. *Computers in Biology and Medicine* **137.** Cited by (Scopus): 7; 104766. ISSN: 0010-4825 (2021).

The final shape of the cranial implant differs from the shape of the skull reconstruction patch. The implants have a constant thickness different from the original bone
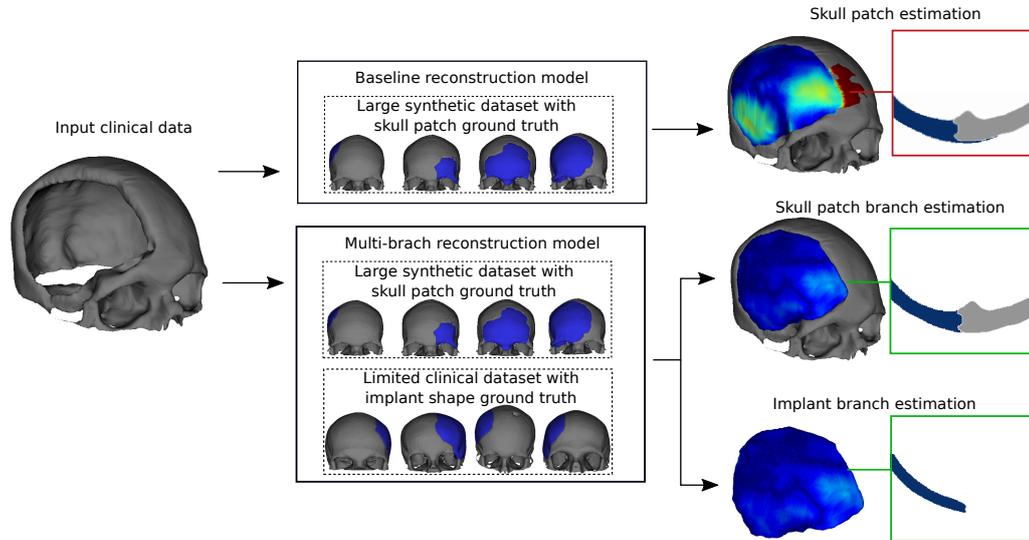
Figure 4.6: The multi-branch architecture uses multi-task learning on different skull reconstruction datasets. The multi-branch architecture uses multi-task learning on different skull reconstruction datasets. In addition to the higher overall accuracy and ability to directly output cranial implant shapes, the skull patch output of the multi-branch model also better fits the shape to complex defect borders in real clinical cases.
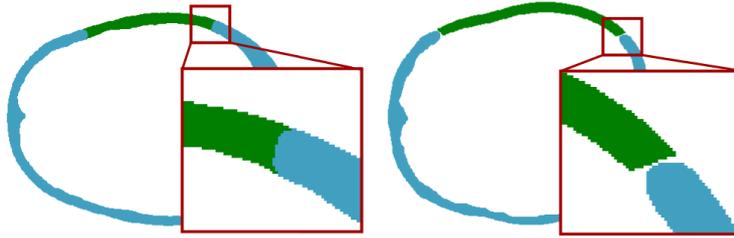
Figure 4.7: Axial slices through samples from the datasets used in this work.
Skull patch sample from a synthetic dataset (left), manually designed implant shape
sample from an in-house clinical dataset (right).

and have some spatial tolerance along the defect border to account for scar tissue and
continuing bone growth, ensuring implantability (see Figure Figure 4.7).

The shape of the implant can be estimated directly by a CNN model, provided that
sufficient training data is available for training. Synthetic datasets for the automatic
estimation of skull patches recently became available because they are easy to cre-
ate from public databases of healthy skulls. Unfortunately, they do not necessarily
fully cover the defective skull shape distribution of target clinical data (i. e. different
anatomical variability of the target population, defect shapes and sizes, complex mor-
phology of the defect border), which may affect the resulting reconstruction quality
in practice [13]. Real clinical data with expert-designed implant models are, on the
other hand, difficult to obtain. Furthermore, the more challenging bilateral and front-
orbital defects are less common. Still, in these challenging cases, correct automatic
skull patch reconstruction can have the most immense impact on clinical practice.

We use two different cranioplasty datasets in this work. Our publicly available
Skullbreak dataset [77] is a synthetic skull shape which contains 570 training and 100
testing pairs of defective skulls and corresponding skull patches. The second in-house
dataset was provided by the TecuMed company and contains 387 real patient cases
indicated for cranioplasty. Each patient case consists of CT data with manual skull
segmentation and a mesh model corresponding to an expert-designed cranial implant.
We split the in-house dataset into 312 training cases and 75 test cases.

We use the same baseline reconstruction method for both the skull patch and im-
plant estimation tasks, with the only difference being the data used for training. The
method is based on a cascade of two U-net-like volumetric CNNs proposed in our
previous work [13]. To simultaneously facilitate training of the CNN cascade using
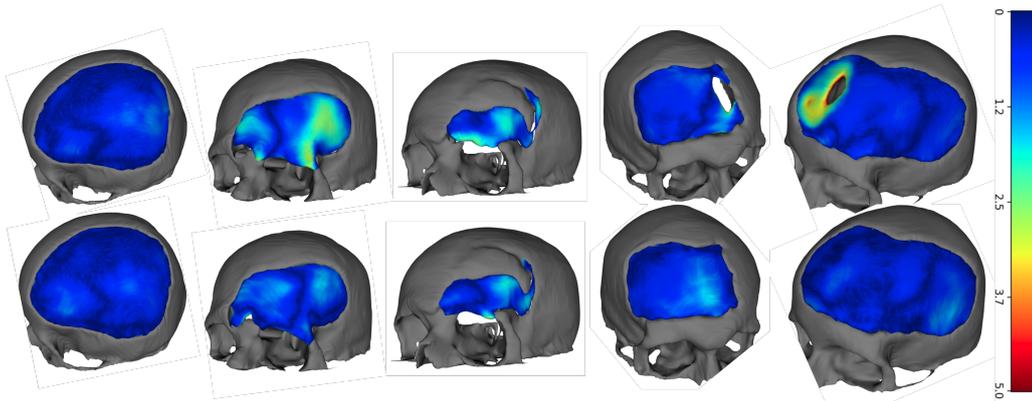
Figure 4.8: Implant estimates of the baseline implant model (top) and the multi-branch model (bottom).

both the synthetic skull patch dataset and the clinical implant dataset, we split the outputs of the model into a separate skull patch estimation branch and implant estimation branch at both coarse and high resolution. During the multi-branch CNN cascade training, we use mixed mini-batches containing samples from the Skullbreak and in-house datasets.

For the sake of the quantitative evaluation, we assume that the expert-designed shapes in the test set represent the only correct solution to the shape estimation tasks. So, the quality of the output can be quantified using segmentation metrics such as volumetric overlaps (i.e. Dice Coefficient) and surface distance [78]. However, the shape reconstruction task is specific in allowing some variability in the reconstructed shape in some cases, as long as there are no imperfections along the fit of the reconstructed shape to the input shape.

Because the thickness of the ground-truth patch is different from the thickness of the original bone in the Skullbreak dataset, we measure average surface error only at the outer surface voxels of the skull. We pay special attention to the fit quality along the defect border of the skull patches. Like other authors [49], we report the outer surface distance computed along the defect border. However, this metric may not precisely convey some common errors of skull reconstruction that impact the aesthetic outcome of cranioplasty, such as slight trenches or bumps on the surface along the defect border. To this end, a novel metric which compares approximate Gaussian curvatures of reconstructed skulls and reference skulls along the defect border as shown in Figure 4.9 has been proposed in the paper as shown in Figure 4.9.

The proposed method reaches an average surface distance of the reconstructed skull
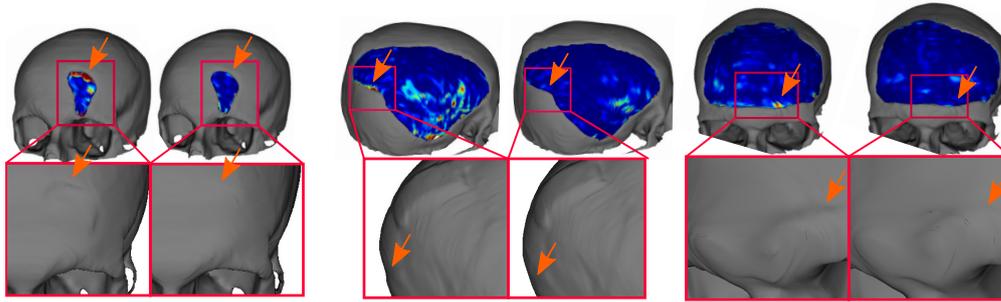
Figure 4.9: Examples of the baseline skull patch model outputs and the multi-branch model skull patch outputs with colour-coded Gaussian curvature error. The multi-branch model can produce smoother results with lower curvature error. Here, the entire Gaussian curvature error maps are shown for illustration, while only defect border voxels are considered when computing the mean errors.

patches of 0.67 mm on a clinical test set of 75 defective skulls. It also achieves a 12% reduction of the newly proposed defect border Gaussian curvature error metric compared to the baseline model trained on synthetic data only. Additionally, it produces 3D printable cranial implant shapes (see Figure 4.8) with a Dice Coefficient of 0.88 and a surface error of 0.65 mm. The proposed skull reconstruction method outputs reach good quality and can be considered for use in semi- or fully automatic clinical cranial implant design workflows.

The main contributions of the presented series of cranial implant design papers can be summarised as follows:

- The novel multi-head volumetric CNN model with a cascade of U-Net 3D CNNs allows simultaneous training on a large dataset of synthetically generated skull patches and a smaller expert-designed cranial implant shapes dataset so the model can generate the final cranial implant shape with the required implant boundary.

- The Skullbreak open dataset [77] for optimisation and validation of anatomical skull reconstruction methods on a set of more challenging synthetically broken skull shapes.

- Proposal of the generative model that could generate multiple possibly correct solutions, which can be further explored in a semi-automatic workflow when the initial reconstruction is not satisfactory so more solutions could be offered to the expert.

Landmark detection is frequently an intermediate step in medical data analysis. An example is the 3D intraoral scan of dentition used in orthodontics, where landmarking is notably challenging due to malocclusion, teeth shift, and frequent missing teeth. With the following paper, we entered a new research field – the automatic analysis of dental scans using deep neural networks:

Kubík, T. & Španěl, M. *Robust Teeth Detection in 3D Dental Scans by Automated Multi-view Landmarking* in *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 2: BIOIMAGING* Best Student Paper Award; (SciTePress, 2022), 24–34. ISBN: 978-989-758-552-4.

Dental casts used in digital orthodontics software are typically obtained from patients with various levels of malocclusion and numerous kinds of teeth shifting. Another challenging problem is the absence of teeth, a common phenomenon in terms of human dentition. For example, extracting third molars (also known as wisdom teeth) is one of the most frequent procedures in oral surgery as it eliminates future problems due to unfavourable orientation. Thus, the landmark method should be robust to such variations. A few examples of samples from our challenging dataset of 337 3D dental scans of human dentition represented as polygon meshes are illustrated in Figure 4.11. The dataset contains cases of both maxillary and mandibular dentition.

(a) Input polygonal mesh   (b) Detected landmarks with our method
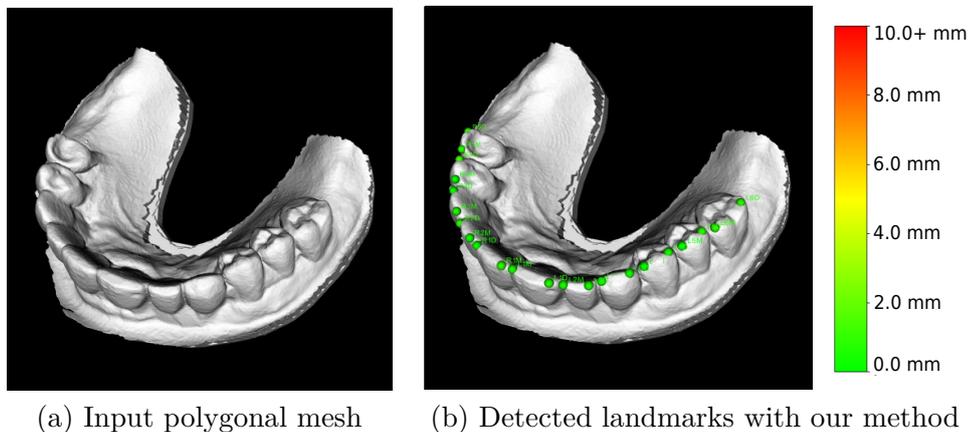
Figure 4.10: An example of a 3D scan of dentition (a) and appropriately detected landmarks (b). Our method automatically detects two landmarks on each tooth – mesial and distal. This type of landmark is necessary in orthodontics, as it defines the rotation of teeth from an anatomically perfect arrangement. Furthermore, the method correctly detects whether a tooth is missing or not.
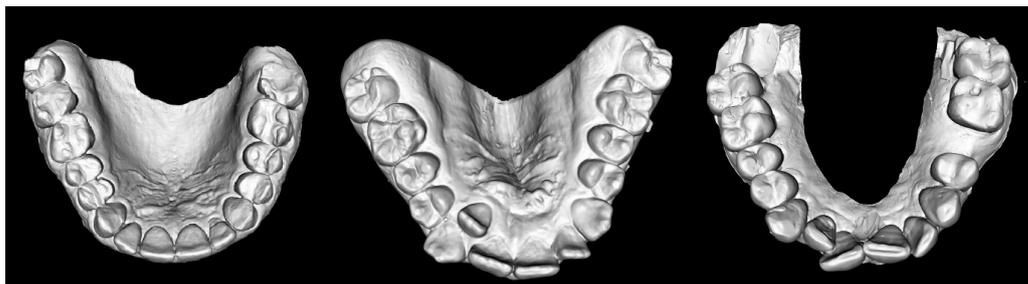
Figure 4.11: Examples of dental casts within the dataset were collected from orthodontics patients, so they usually suffer from different kinds of malocclusion.

As CNNs gained popularity, more scientific papers concerning their usage in landmark detection emerged. Some of these methods detected the landmark position directly by regressing its x and y coordinates [79, 80]. Over time, extensive literature has developed on landmarking by heatmap regression. Pfister et al. worked on a model that regresses human joint positions [81]. Instead of directly regressing the (x, y) joint position, they regressed a joint position's heatmap. During the training, the ground truth labels are transformed into heatmaps by placing a Gaussian with fixed variance at each joint coordinate. They concluded that the benefits of regressing a heatmap rather than (x, y) coordinates are: (i) the process of network training can be visualised in such a way that one can understand the network learning failures, and (ii) the network output can acquire confidence at multiple spatial locations. The incorrect ones are slowly suppressed later in the training process. In contrast, regressing the (x, y) coordinates directly, the network would have a lower loss only if it predicts the coordinate correctly, even if it was "growing confidence" in the correct position. Their approach outperformed direct coordinate regression and became a standard way of landmark detection in 2D images. Inspired by this method, Payer et al. presented multiple architectures that detect keypoints in X-ray images of hands and 3D hand MR scans [75].

A 3D segmentation for annotation of individual teeth and gingiva based on Graph Neural Networks (GNNs) was presented by Sun et al. [82]. Their network produces a dense correspondence that helps accurately locate individual orthodontics landmarks on teeth crowns. Another recent work was presented by Wu et al. [83]. They introduced a two-stage framework based on mesh deep learning for joint tooth labelling and landmark identification. To accurately detect tooth landmarks, they designed a modified PointNet [60] to learn the heatmaps encoding landmark locations.

In the paper, we present a method that considers the limitation of the dataset size,

the need for low computational time (as it should be part of interactive orthodontics treatment planning software), and the importance of robustness to missing and shifted teeth. The method is based on a multi-view approach, and it uses the heatmap regression [84] to predict landmarks in 2D and the RANSAC consensus method to robustly propagate the information back into 3D space (Figure 4.12). In order to address the problem of landmark presence on missing teeth, our method comprises post-processing based on a heatmap regression uncertainty analysis combined with the uncertainty of the multi-view approach. It ensures that our method correctly detects landmark presence without any additional computations. This is inevitable for orthodontic planning as it robustly detects the presence of teeth even in challenging cases.

The model is observed from various viewpoints. We used uniformly distributed camera positions with a maximal angle of ± 30 degrees from the initial aligned dental scan position. Images in the form of depth maps and direct geometry rendering are used as the inputs to the neural network. The predicted 2D heatmaps can be interpreted as the landmarks' screen coordinate positions. Each output channel contains a heatmap with a Gaussian representing the probability of a given landmark's screen coordinate in each pixel. Thus, the resulting screen coordinate must be extracted from the predicted heatmap by finding the coordinates of the peak value. The RANSAC consensus method combines these individual predictions from all viewpoints. Based on the maximum value in the activation map, only certain predictions are sent to the consensus method.

Our post-processing contains an analysis of the presence of each tooth. This is, in fact, a binary classification task whose result is based on two uncertainty hypotheses:

- Like Drevicky and Kodym [81, 84], the network is trained to regress heatmaps with the amplitude 1. Then, the fundamental assumption is that during the inference, the certainty is measured by the maximum value in the activation map, with a proportional increase to the network's confidence.

- The RANSAC consensus method robustly estimates the landmark position by eliminating outlier predictions. Thus, the proportion of inliers and outliers is another valuable output of this consensus method, assuming the number of inliers is proportional to the overall confidence.

Conducted experiments have shown that the multi-view approach combined with the RANSAC consensus method brings promising results in automating landmark detection (Figure 4.13). Evaluated on a dataset of real orthodontics dental casts with
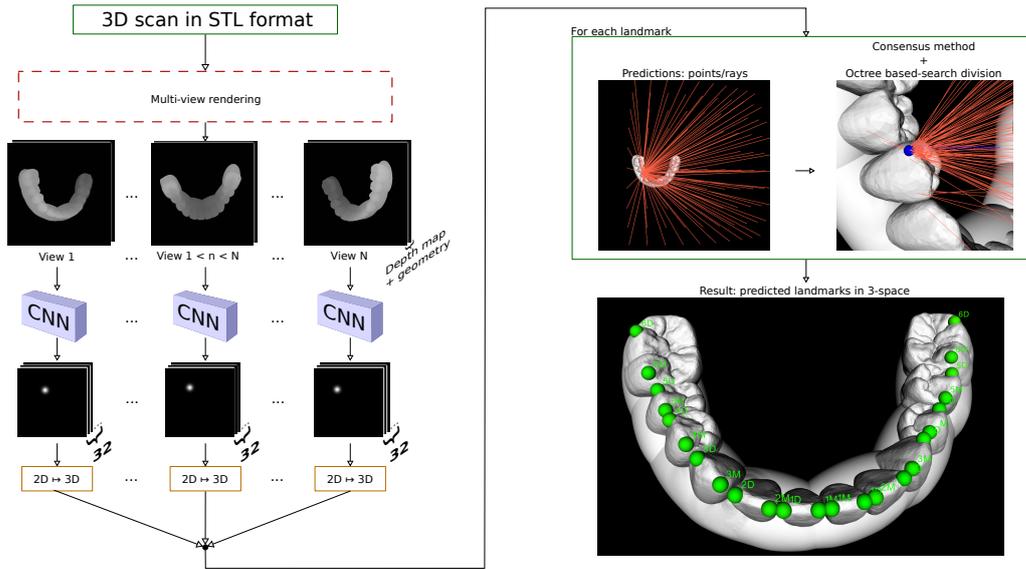
Figure 4.12: Outline of the proposed method for orthodontics landmark detection. Following the multi-view approach, the input 3D model is observed from various viewports and sent to the CNNs to predict landmark location heatmaps. Landmark screen coordinates are extracted from all the heatmaps, re-projected to the surface mesh, and processed by the consensus method, producing final estimates. Additionally, the maximum value in the activation heatmap and the consensus method output are used to detect tooth presence during post-processing.

significant diversity, the method performs the best with Attention U-Net architecture [85] as the backbone and with two-channeled input of depth maps and geometry renders. The proposed method can detect orthodontics landmarks on surface models with an error of 0.75 ± 0.96 mm.

When comparing our results to the framework from Wu et al. [83], they achieve a slightly better error of 0.623 ± 0.718 mm. However, it is necessary to recognise that their dataset consists of 36 samples. Such a relatively small number should be increased to ensure the method's robustness to the large variability of orthodontic cases. Our dataset is more challenging and consists of problematic cases with severe teeth shiftings and many cases with missing teeth. In addition, they detect landmarks only on 10 teeth, excluding, for example, very problematic third molars. For a fair comparison, it would be vital to benchmark methods on a public dataset, which was not available when writing the paper.

We have also shown that the uncertainty measures based on the analysis of the maximum values of regressed heatmap predictions in combination with multi-view uncertainty yield convenient information in the process of landmark presence detec-
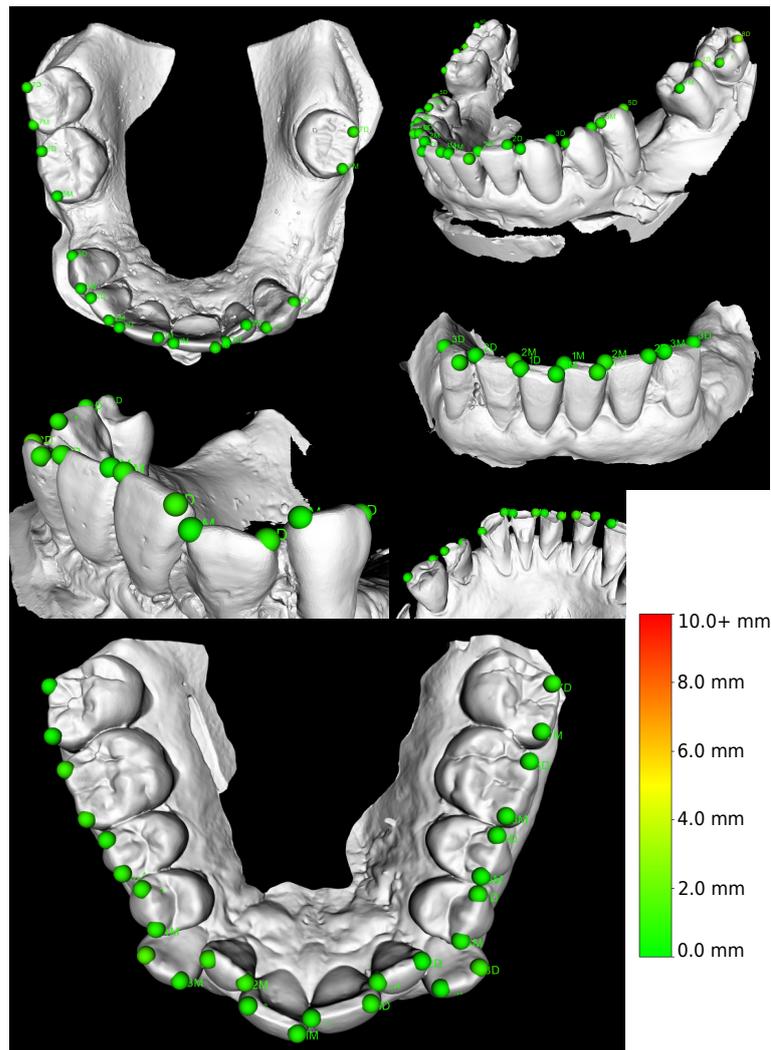
Figure 4.13: Examples of automatically detected landmarks with our method. The majority of predictions have a landmark localisation error below 2 mm. The method correctly detects if a tooth is missing and does not produce landmark predictions.

tion. Our method correctly detects missing teeth in 97.68% of cases. This means the method can be applied to data where landmarks' presence is not granted. In addition, the method meets the needs of clinical applications, as the inference at the user's side takes seconds to be calculated, even on less powerful CPUs.

4.4 Insights and Remarks

A more general comparison of shape modelling methods is currently limited by the need for standardized demanding datasets and methodology to evaluate the anatomical reconstruction methods. The distance from a ground-truth shape might estimate how well a method performs, but it is not the only criterion of correct reconstruction. The most relevant metric to measure the reconstruction method performance would be the time an expert requires to design a clinically acceptable implant from the initial automatic reconstruction.

Kodym presented the results of such a study in his habilitation thesis [39]. He created a dataset of automatically reconstructed defective skulls. He let an expert technician with years of experience in skull reconstruction and patient-specific implant design perform a subjective quality evaluation. Comparing the subjective expert scores with metrics of similarity between the reconstructed and the original shape can give an idea of what to look for when evaluating the reconstructions.

Routinely used metrics, including volumetric Dice Coefficient and average symmetric outer surface distance, correlate with the expert subjective score, confirming that they are appropriate for the comparison of different reconstruction methods. However, their correlation is weak in regions of higher subjective expert scores, making it impossible to use them for discrimination between good and perfect results. Also, several cases satisfied the quantitative metrics while being seen as low-quality by the expert and vice versa.

As the smoothness of the surface closest to the defect border significantly impacts the aesthetic outcome of cranioplasty, Kodym also studied the outer surface distance of the defect border and the mean square error of Gaussian curvature. These quantitative border metrics also do not always agree with the subjective quality score, but the correlation with the expert score was higher in the cases where the correlation of the global metrics was low. Although the study was performed using only one type of automatic reconstruction method and the results were evaluated by a single implant design expert, it can be suggested that to measure the quality of results of an automatic shape reconstruction, different types of quantitative metrics should be combined, and both global and border metrics should be considered.

The development of deep learning in various domains shows that a huge strength of these methods is the potential to apply similar techniques in different domains. Principles successful in language models, such as transformers, are deployed for image

analysis. Similarly, the lack of training data is not only specific to the medical field. The deep learning researchers explore many self-supervised techniques that work remarkably well, even after retraining on a small specific dataset. Due to the massive overlap of common principles in many target domains, deep learning is a fast-growing area with which classical methods such as SSMs will have difficulty keeping up, even though the statistical models also have their advantages.

My experience using deep learning shows that its practical deployment is more straightforward. Well, data is needed (a lot of data), annotations are needed (time-consuming annotations), and fine-tuning models for a specific task requires very expert knowledge. However, modifying a model concerning changes in the desired results does not mean such dramatic changes. It mainly involves collecting more data, reannotating data and retraining models, which is time-consuming but does not require fundamental changes in the approach. In addition, today, we already have the newly developing MLOps (Machine Learning Operations) specialisation as a tool for an end-to-end machine learning development process to design, build and manage reproducible, testable, and evolvable ML-powered software.

The interpretability of the *decisions* of deep learning models can be considered problematic not only in the medical field. However, this strongly depends on the specific task. In areas such as the design of implants or the automation of treatment planning software, where it is not an automatic diagnosis of diseases, it is essential that an expert evaluates the result and decides whether to use it or not.

5 Conclusion

Computer-assisted pre-operative planning is an area that has benefited heavily from the growth of machine and deep learning in the past years. Pre-operative treatment planning no longer has to be technically demanding and time-consuming. In addition to traditional image analysis, 3D modelling and 3D shape understanding allow digital planning to be more accurate and heavily automated. In this habilitation thesis, I summarized the past decade of our contribution to the field of 3D shape analysis and its applications in clinical practice.

The results of the young, enthusiastic researchers I have worked with and had the pleasure of helping guide their research are summarized in Chapter 3 and Chapter 4. The first part is about the use of statistical shape and intensity models, where the most significant contribution was the new method for reconstructing 3D polygonal models of the intact long bones (femur, tibia and radius) from the calibrated radiographs of fractures of the bones based on the Levenberg-Marquardt optimization method, and GPU-accelerated rendering of digital radiographs from statistical shape and intensity models, which provided fast convergence and state-of-the-art reconstruction accuracy when combined with the LM optimization.

The second part is about deep convolutional neural networks for 3D shape modelling. Here, the main topic of several papers was the reconstruction of the skull shape for the automatic design of cranial implants, where the main contribution is the publication of a novel multi-head and multi-scale volumetric CNN which can deal with actual patient data and generate the final cranial implant shape with required details of the implant boundary what considerably reduces further manual adjustments. Then, there is also the publication of The SkullBreak open dataset for validation of anatomical skull reconstruction methods and, subsequently, our participation in the organizing team of the MICCAI 2021 AutoImplant Challenge.

The last research topic in the thesis is the processing of 3D dental casts for orthodontic planning and 3D printing of invisible braces. Our paper on detecting landmark points annotating individual teeth shows that the multi-view approach, where the neural network processes 2D views of the 3D model, still has yet to say the last word

and achieves equal results to more recent architectures. In this area, we are currently finishing the second article based on Tibor Kubík master's thesis and show that an adequately designed multi-view approach is an accurate and more efficient solution compared to emerging architectures like MeshCNN or graph neural networks.

The contribution of deep learning to many domains is a huge step forward, but I also look at it critically. There are areas in which their deployment is questionable and controversial. However, in the area of pre-operative planning, where these methods assist the clinician or technician in planning the treatment, help them reduce the time of the surgery and prevent complications, their advantages outweigh the traditional approaches. The powerful benefits are easier adaptation to different conditions, the possibility of retraining on problematic cases, or even the discussed possibility of using generative networks to propose several alternative designs to the clinician and to make the entire planning process an interactive *discussion* between the computer-assistant and the clinician.

Another interesting topic for future research is the combination of deep learning with exact discrete mathematics used in the 3D geometry processing, which often fails if the input 3D meshes are incorrect and contain errors (unclosed surfaces, small holes, wrong orientation of normals, etc.). Deep learning methods pre-trained on extensive data can add an *element of experience* to failing exact mathematical algorithms. The element that is typically given by an experienced technician who repairs and corrects 3D models based on his experience.

6 References

1. Baka, N., Kaptein, B., de Bruijne, M., van Walsum, T., Giphart, J., Niessen, W. & Lelieveldt, B. 2D–3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models. *Medical Image Analysis* **15,** 840–850. ISSN: 1361-8415 (2011).

2. Markelj, P., Tomaževič, D., Likar, B. & Pernuš, F. A review of 3D/2D registration methods for image-guided interventions. *Medical Image Analysis* **16.** Computer Assisted Interventions, 642–661. ISSN: 1361-8415 (2012).

3. Lamecker, H. & Zachow, S. in *Computational Radiology for Orthopaedic Interventions* (eds Zheng, G. & Li, S.) 1–23 (Springer International Publishing, Cham, 2016). ISBN: 978-3-319-23482-3.

4. Zheng, G., Schumann, S., Balestra, S., Thelen, B. & Nolte, L.-P. in *Computational Radiology for Orthopaedic Interventions* (eds Zheng, G. & Li, S.) 197–215 (Springer International Publishing, Cham, 2016). ISBN: 978-3-319-23482-3.

5. Gong, R. H., Stewart, J. & Abolmaesumi, P. Reduction of multi-fragment fractures of the distal radius using atlas-based 2D/3D registration. *Proc SPIE* **7261** (Feb. 2009).

6. Gong, R. H. & Abolmaesumi, P. 2D/3D registration with the CMA-ES method - art. no. 69181M. *Proceedings of SPIE - The International Society for Optical Engineering* **6918** (Mar. 2008).

7. Ehlke, M., Ramm, H., Lamecker, H., Hege, H.-C. & Zachow, S. Fast Generation of Virtual X-ray Images for Reconstruction of 3D Anatomy. *IEEE transactions on visualization and computer graphics* **19,** 2673–82 (Dec. 2013).

8. Han, R., Uneri, A., Vijayan, R., Wu, P., Vagdargi, P., Sheth, N., Vogt, S., Kleinszig, G., Osgood, G. & Siewerdsen, J. Fracture reduction planning and guidance in orthopaedic trauma surgery via multi-body image registration. *Medical Image Analysis* **68,** 101917. ISSN: 1361-8415 (2021).

6 References

9. Klima, O., Kleparnik, P., Spanel, M. & Zemcik, P. *Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation* in *Proceedings of the SPIE* (eds Gimi, B. & Krol, A.) **9788.** Cited by (Scopus): 8; (Mar. 2016), 97880F.

10. Klima, O., Chromy, A., Zemcik, P., Spanel, M. & Kleparnik, P. A Study on Performace of Levenberg-Marquardt and CMA-ES Optimization Methods for Atlas-based 2D/3D Reconstruction. *IFAC-PapersOnLine* **49.** 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016, 121–126. ISSN: 2405-8963 (2016).

11. Klíma, O., Madeja, R., Španel, M., Čuta, M., Zemčík, P., Stoklásek, P. & Mizera, A. *Virtual 2D-3D Fracture Reduction with Bone Length Recovery Using Statistical Shape Models* in *Shape in Medical Imaging* (eds Reuter, M., Wachinger, C., Lombaert, H., Paniagua, B., Lüthi, M. & Egger, B.) (Springer International Publishing, Cham, 2018), 207–219. ISBN: 978-3-030-04747-4.

12. Klíma, O., Novobilský, P., Madeja, R., Bařina, D., Chromý, A., Španěl, M. & Zemcčík, P. Intensity-Based Nonoverlapping Area Registration Supporting "Drop-Outs" in Terms of Model-Based Radiostereometric Analysis. *Journal of Healthcare Engineering* **2018,** 1–10 (May 2018).

13. Kodym, O., Španěl, M. & Herout, A. Skull shape reconstruction using cascaded convolutional networks. *Computers in Biology and Medicine* **123.** Cited by (Scopus): 9; 103886. ISSN: 0010-4825 (2020).

14. Kodym, O., Španěl, M. & Herout, A. *Cranial Defect Reconstruction Using Cascaded CNN with Alignment* in *Towards the Automatization of Cranial Implant Design in Cranioplasty* (eds Li, J. & Egger, J.) Cited by (Scopus): 4; (Springer International Publishing, Cham, 2020), 56–64. ISBN: 978-3-030-64327-0.

15. Kodym, O., Španěl, M. & Herout, A. Deep learning for cranioplasty in clinical practice: Going from synthetic to real patient data. *Computers in Biology and Medicine* **137.** Cited by (Scopus): 7; 104766. ISSN: 0010-4825 (2021).

16. Kubík, T. & Španěl, M. *Robust Teeth Detection in 3D Dental Scans by Automated Multi-view Landmarking* in *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 2: BIOIMAGING* Best Student Paper Award; (SciTePress, 2022), 24–34. ISBN: 978-989-758-552-4.

17. Joskowicz, L., Cohen, D., Caplan, N. & Sosna, J. Inter-observer variability of manual contour delineation of structures in CT. *European Radiology* **29,** 1391–1399 (Sept. 2019).

18. Li, J., Udupa, J. K., Tong, Y., Wang, L. & Torigian, D. A. Segmentation evaluation with sparse ground truth data: Simulating true segmentations as perfect/imperfect as those generated by humans. *Medical Image Analysis* **69,** 101980. ISSN: 1361-8415. https://www.sciencedirect.com/science/article/pii/S1361841521000268 (2021).

19. Abdi, A., Borgard, H., Abolmaesumi, P. & Fels, S. *AnatomyGen: Deep Anatomy Generation From Dense Representation With Applications in Mandible Synthesis* in *Proceedings of The 2nd International Conference on Medical Imaging with Deep Learning* (eds Cardoso, M. J., Feragen, A., Glocker, B., Konukoglu, E., Oguz, I., Unal, G. & Vercauteren, T.) **102** (PMLR, Oct. 2019), 4–14.

20. Abdi, A. H., Pesteie, M., Prisman, E., Abolmaesumi, P. & Fels, S. *Variational Shape Completion for Virtual Planning of Jaw Reconstructive Surgery* in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019* (eds Shen, D., Liu, T., Peters, T. M., Staib, L. H., Essert, C., Zhou, S., Yap, P.-T. & Khan, A.) (Springer International Publishing, Cham, 2019), 227–235. ISBN: 978-3-030-32254-0.

21. Kubík, T. *Deep Learning for 3D Geometry Analysis in Medicine* Supervisor Španěl, Michal. Master's thesis (Brno University of Technology, Faculty of Information Technology, 2023).

22. Klíma, O. *Model-based 2D-3D Registration Methods for Analysis of Conventional Radiographs* Ph.D. thesis (Brno University of Technology, Faculty of Information Technology, Brno, CZ, 2022). https://www.fit.vut.cz/study/phd-thesis/648/.

23. Madeja, R., Bajor, G., Klima, O., Bialy, L. & Pometlova, J. Computer-assisted preoperative planning of reduction of and osteosynthesis of scapular fracture: A case report. *Open Medicine* **16,** 1597–1601 (2021).

24. Gower, J. C. Generalized procrustes analysis. *Psychometrika* **40,** 33–51 (1975).

25. Tipping, M. E. & Bishop, C. M. Probabilistic Principal Component Analysis. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)* **61,** 611–622. ISSN: 13697412, 14679868 (1999).

26. Yao, J. & Taylor, R. H. *Construction and simplification of bone density models* in *Medical Imaging 2001: Image Processing* (eds Sonka, M. & Hanson, K. M.) **4322** (SPIE, 2001), 814–823.

27. Yao, J. & Taylor, R. H. *Assessing Accuracy Factors in Deformable 2D/3D Medical Image Registration Using a Statistical Pelvis Model* in *Computer Vision, IEEE International Conference on* **3** (IEEE Computer Society, Los Alamitos, CA, USA, Oct. 2003), 1329.

28. Sadowsky, O., Cohen, J. D. & Taylor, R. H. Rendering tetrahedral meshes with higher-order attenuation functions for digital radiograph reconstruction. *VIS 05. IEEE Visualization, 2005.*, 303–310 (2005).

29. Sadowsky, O., Cohen, J. D. & Taylor, R. H. Projected Tetrahedra Revisited: A Barycentric Formulation Applied to Digital Radiograph Reconstruction Using Higher-Order Attenuation Functions. *IEEE Transactions on Visualization and Computer Graphics* **12,** 461–473. ISSN: 1077-2626 (June 2006).

30. Cootes, T., Cooper, D., Taylor, C. & Graham, J. Trainable method of parametric shape description. *Image and Vision Computing* **10.** BMVC 1991, 289–294. ISSN: 0262-8856 (1992).

31. Hansen, N. & Kern, S. *Evaluating the CMA Evolution Strategy on Multimodal Test Functions* in *Parallel Problem Solving from Nature - PPSN VIII* (eds Yao, X., Burke, E. K., Lozano, J. A., Smith, J., Merelo-Guervós, J. J., Bullinaria, J. A., Rowe, J. E., Tiňo, P., Kabán, A. & Schwefel, H.-P.) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2004), 282–291. ISBN: 978-3-540-30217-9.

32. Beyer, H.-G. & Sendhoff, B. *Covariance Matrix Adaptation Revisited – The CMSA Evolution Strategy* in *Parallel Problem Solving from Nature – PPSN X* (eds Rudolph, G., Jansen, T., Beume, N., Lucas, S. & Poloni, C.) (Springer Berlin Heidelberg, Berlin, Heidelberg, 2008), 123–132. ISBN: 978-3-540-87700-4.

33. Gong, R. H., Stewart, J. & Abolmaesumi, P. Multiple-Object 2-D–3-D Registration for Noninvasive Pose Identification of Fracture Fragments. *IEEE Transactions on Biomedical Engineering* **58,** 1592–1601 (2011).

34. Wei, S., Ramakrishna, V., Kanade, T. & Sheikh, Y. *Convolutional Pose Machines* in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, Los Alamitos, CA, USA, May 2016), 4724–4732.

35. Roy, A. & Todorovic, S. *Monocular Depth Estimation Using Neural Regression Forest* in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (2016), 5506–5514.

36. Wang, X., Xie, L., Dong, C. & Shan, Y. *Real-ESRGAN: Training Real-World Blind Super-Resolution with Pure Synthetic Data* in *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)* (2021), 1905–1914.

37. Krull, A., Buchholz, T.-O. & Jug, F. Noise2Void - Learning Denoising From Single Noisy Images. *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR),* 2124–2132 (2018).

38. Chaudhary, S., Moon, S. & Lu, H. Fast, efficient, and accurate neuro-imaging denoising via supervised deep learning. *Nature Communications* **13** (Sept. 2022).

39. Kodym, O. *Deep Learning for Virtual Patient-Specific Skull Modelling and Reconstruction* Ph.D. thesis (Brno University of Technology, Faculty of Information Technology, Brno, CZ, 2022). https://www.fit.vut.cz/study/phd-thesis/1225/.

40. Kubík, T. *Deep Neural Networks for Landmark Detection on 3D Models* Supervisor Španěl, Michal. Batchelor's thesis (Brno University of Technology, Faculty of Information Technology, 2021).

41. Lee, M.-Y., Chang, C.-C., Lin, C.-C., Lo, L.-J. & Chen, Y.-R. Custom implant design for patients with cranial defects. *IEEE Engineering in Medicine and Biology Magazine* **21,** 38–44 (2002).

42. Kurland, D., Khaladj-Ghom, A., Stokum, J., Carusillo, B., Karimy, J., Gerzanich, V., Sahuquillo, J. & Simard, J. M. Complications Associated with Decompressive Craniectomy: A Systematic Review. *Neurocritical Care* **23** (June 2015).

43. Jardini, A. L., Larosa, M. A., Filho, R. M., de Carvalho Zavaglia, C. A., Bernardes, L. F., Lambert, C. S., Calderoni, D. R. & Kharmandayan, P. Cranial reconstruction: 3D biomodel and custom-built implant created using additive manufacturing. *Journal of Cranio-Maxillofacial Surgery* **42,** 1877–1884. ISSN: 1010-5182 (2014).

44. Oh, J.-h. Recent advances in the reconstruction of cranio-maxillofacial defects using computer-aided design/computer-aided manufacturing. *Maxillofacial Plastic and Reconstructive Surgery* **40** (Dec. 2018).

45.   Belt, H., Neut, D., Schenk, W., Horn, J., van der Mei, H. & Busscher, H. Infection of orthopedic implants and the use of antibiotic-loaded bone cements: A review. *Acta orthopaedica Scandinavica* **72,** 557–71 (Jan. 2002).

46.   Rudman, K., Hoekzema, C. & Rhee, J. Computer-Assisted Innovations in Craniofacial Surgery. *Facial Plastic Surgery* **27,** 358–365. ISSN: 1098-8793 (04 July 2011).

47.   Chen, X., Xu, L., Li, X. & Egger, J. Computer-aided implant design for the restoration of cranial defects. *Scientific Reports* **7** (2017).

48.   Fuessinger, M. A., Schwarz, S., Neubauer, J., Cornelius, C.-P., Gass, M., Poxleitner, P., Zimmerer, R., Metzger, M. C. & Schlager, S. Virtual reconstruction of bilateral midfacial defects by using statistical shape modeling. *Journal of Cranio-Maxillofacial Surgery* **47,** 1054–1059. ISSN: 1010-5182 (2019).

49.   Matzkin, F., Newcombe, V., Stevenson, S., Khetani, A., Newman, T., Digby, R., Stevens, A., Glocker, B. & Ferrante, E. in. Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 390–399 (Sept. 2020). ISBN: 978-3-030-59712-2.

50.   Morais, A., Egger, J. & Alves, V. *Automated Computer-aided Design of Cranial Implants Using a Deep Volumetric Convolutional Denoising Autoencoder* in *New Knowledge in Information Systems and Technologies* (eds Rocha, Á., Adeli, H., Reis, L. P. & Costanzo, S.) (Springer International Publishing, Cham, 2019), 151–160. ISBN: 978-3-030-16187-3.

51.   Wu, Z., Song, S., Khosla, A., Yu, F., Zhang, L., Tang, X. & Xiao, J. *3D ShapeNets: A deep representation for volumetric shapes* in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, 2015), 1912–1920.

52.   Sharma, A., Grau, O. & Fritz, M. *VConv-DAE: Deep Volumetric Shape Learning Without Object Labels* in *Computer Vision – ECCV 2016 Workshops* (eds Hua, G. & Jégou, H.) (Springer International Publishing, Cham, 2016), 236–250. ISBN: 978-3-319-49409-8.

53.   Dai, A., Qi, C. & Nießner, M. Shape Completion Using 3D-Encoder-Predictor CNNs and Shape Synthesis. *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* 6545–6554 (2016).

54. Litany, O., Bronstein, A. M., Bronstein, M. M. & Makadia, A. Deformable Shape Completion with Graph Convolutional Autoencoders. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition,* 1886–1895 (2017).

55. Wang, P.-S., Liu, Y., Guo, Y.-X., Sun, C.-Y. & Tong, X. O-CNN: Octree-Based Convolutional Neural Networks for 3D Shape Analysis. *ACM Trans. Graph.* **36.** ISSN: 0730-0301. https://doi.org/10.1145/3072959.3073608 (July 2017).

56. Su, H., Maji, S., Kalogerakis, E. & Learned-Miller, E. G. *Multi-view Convolutional Neural Networks for 3D Shape Recognition* in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015* (2015), 945–953.

57. Le, T., Bui, G. & Duan, Y. A multi-view recurrent neural network for 3D mesh segmentation. *Computers & Graphics* **66.** Shape Modeling International 2017, 103–112. ISSN: 0097-8493. https://doi.org/10.1016/j.cag.2017.05.011 (2017).

58. Paulsen, R. R., Juhl, K. A., Haspang, T. M., Hansen, T., Ganz, M. & Einarsson, G. in *Lecture Notes in Computer Science* 706–719 (Springer International Publishing, 2019). ISBN: 9783030208875.

59. Fischler, M. A. & Bolles, R. C. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM* **24,** 381–395 (1981).

60. Qi, R. C., Su, H., Kaichun, M. & Guibas, L. J. *PointNet: Deep Learning on Point Sets for 3D Classification and Segmentation* in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE Computer Society, Los Alamitos, CA, USA, June 2017), 77–85.

61. Qi, C. R., Yi, L., Su, H. & Guibas, L. J. *PointNet++: Deep Hierarchical Feature Learning on Point Sets in a Metric Space* in *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA* (2017), 5099–5108.

62. Wu, W., Qi, Z. & Fuxin, L. *PointConv: Deep Convolutional Networks on 3D Point Clouds* in *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (2019), 9613–9622.

63. Hanocka, R., Hertz, A., Fish, N., Giryes, R., Fleishman, S. & Cohen-Or, D. MeshCNN: a network with an edge. *ACM Trans. Graph.* **38,** 90:1–90:12 (2019).

64. Hansen, B., Lowes, M., Ørkild, T., Dahl, A., Dahl, V., De Backer, O., Camara, O., Paulsen, R., Ingwersen, C. & Sørensen, K. SparseMeshCNN with Self-Attention for Segmentation of Large Meshes. *Proceedings of the Northern Lights Deep Learning Workshop* **3** (Apr. 2022).

65. Feng, Y., Feng, Y., You, H., Zhao, X. & Gao, Y. *MeshNet: Mesh Neural Network for 3D Shape Representation* in *The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019* (AAAI Press, 2019), 8279–8286.

66. Hu, S.-M., Liu, Z.-N., Guo, M.-H., Cai, J.-X., Huang, J., Mu, T.-J. & Martin, R. R. Subdivision-based Mesh Convolution Networks. *ACM Transactions on Graphics* **41,** 1–16 (Mar. 2022).

67. Dong, Q., Wang, Z., Li, M., Gao, J., Chen, S., Shu, Z., Xin, S., Tu, C. & Wang, W. Laplacian2Mesh: Laplacian-Based Mesh Understanding. *IEEE Transactions on Visualization and Computer Graphics,* 1–13 (2023).

68. Sharp, N., Attaiki, S., Crane, K. & Ovsjanikov, M. DiffusionNet: Discretization Agnostic Learning on Surfaces. *ACM Trans. Graph.* **41,** 27:1–27:16 (2022).

69. Wang, W., Huang, Q., You, S., Yang, C. & Neumann, U. *Shape Inpainting Using 3D Generative Adversarial Network and Recurrent Convolutional Networks* in *2017 IEEE International Conference on Computer Vision (ICCV)* (IEEE Computer Society, Los Alamitos, CA, USA, Oct. 2017), 2317–2325.

70. Fuessinger, M., Schlager, S., Schwarz, S., Ellis, E., Cornelius, C., Probst, F., Metzger, M. & Semper-Hogg, W. Virtual reconstruction of midface defects using statistical shape models. *Journal of Cranio-Maxillofacial Surgery* **45** (Dec. 2016).

71. Fuessinger, M., Schwarz, S., Cornelius, C., Metzger, M., Ellis, E., Probst, F., Semper-Hogg, W., Gass, M. & Schlager, S. Planning of skull reconstruction based on a statistical shape model combined with geometric morphometrics. *International Journal of Computer Assisted Radiology and Surgery* **13** (Oct. 2017).

72. Çiçek, Ö., Abdulkadir, A., Lienkamp, S. S., Brox, T. & Ronneberger, O. *3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016* (eds Ourselin, S., Joskowicz, L., Sabuncu, M. R., Unal, G. & Wells, W.) (Springer International Publishing, Cham, 2016), 424–432. ISBN: 978-3-319-46723-8.

73. Chilamkurthy, S., Ghosh, R., Tanamala, S., Biviji, M., Campeau, N., Venugopal, V., Mahajan, V., Rao, P. & Warier, P. Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study. *The Lancet* **392** (Oct. 2018).

74. Li, J. & Egger, J. *Towards the Automatization of Cranial Implant Design in Cranioplasty: First Challenge, AutoImplant 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings* ISBN: 9783030643263 (Springer International Publishing, 2020).

75. Payer, C., Štern, D., Bischof, H. & Urschler, M. *Regressing Heatmaps for Multiple Landmark Localization Using CNNs* in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016: 19th International Conference, Proceedings, Part II* **9901** (Springer-Verlag, Athens, Greece, 2016), 230–238. ISBN: 978-3-319-46723-8.

76. Li, J., Pimentel, P., Szengel, A., Ehlke, M., Lamecker, H., Zachow, S., Estacio, L., Doenitz, C., Ramm, H., Shi, H., Chen, X., Matzkin, F., Newcombe, V., Ferrante, E., Jin, Y., Ellis, D. G., Aizenberg, M. R., Kodym, O., Španěl, M., Herout, A., Mainprize, J. G., Fishman, Z., Hardisty, M. R., Bayat, A., Shit, S., Wang, B., Liu, Z., Eder, M., Pepe, A., Gsaxner, C., Alves, V., Zefferer, U., von Campe, G., Pistracher, K., Schäfer, U., Schmalstieg, D., Menze, B. H., Glocker, B. & Egger, J. AutoImplant 2020-First MICCAI Challenge on Automatic Cranial Implant Design. *IEEE Transactions on Medical Imaging* **40,** 2329–2342 (2021).

77. Kodym, O., Li, J., Pepe, A., Gsaxner, C., Chilamkurthy, S., Egger, J. & Španěl, M. SkullBreak / SkullFix – Dataset for automatic cranial implant design and a benchmark for volumetric shape learning tasks. *Data in Brief* **35,** 106902 (Apr. 2021).

78. Marreiros, F. M. M., Heuzé, Y., Verius, M., Unterhofer, C., Freysinger, W. & Recheis, W. Custom implant design for large cranial defects. *International Journal of Computer Assisted Radiology and Surgery* **11,** 2217–2230 (June 2016).

79. Sun, Y., Wang, X. & Tang, X. *Deep Convolutional Network Cascade for Facial Point Detection* in *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013* (IEEE Computer Society, 2013), 3476–3483. ISBN: 978-1-5386-5672-3. https://doi.org/10.1109/CVPR.2013.446.

6 References

80. Lv, J., Shao, X., Xing, J., Cheng, C. & Zhou, X. *A Deep Regression Architecture with Two-Stage Re-initialization for High Performance Facial Landmark Detection* in *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017* (IEEE Computer Society, 2017), 3691–3700. ISBN: 978-1-5386-0457-1.

81. Pfister, T., Charles, J. & Zisserman, A. *Flowing ConvNets for Human Pose Estimation in Videos* in *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015* (IEEE Computer Society, 2015), 1913–1921. ISBN: 978-1-4673-8391-2.

82. Sun, D., Pei, Y., Li, P., Song, G., Guo, Y., Zha, H. & Xu, T. *Automatic Tooth Segmentation and Dense Correspondence of 3D Dental Model* in *Medical Image Computing and Computer Assisted Intervention – MICCAI 2020* (eds Martel, A. L., Abolmaesumi, P., Stoyanov, D., Mateus, D., Zuluaga, M. A., Zhou, S. K., Racoceanu, D. & Joskowicz, L.) (Springer International Publishing, Cham, 2020), 703–712. ISBN: 978-3-030-59719-1.

83. Wu, T.-H., Lian, C., Lee, S., Pastewait, M. F., Piers, C., Liu, J., Wang, F., Wang, L., Chiu, C.-Y., Wang, W., Jackson, C. B., Chao, W.-L., Shen, D. & Ko, C.-C. Two-Stage Mesh Deep Learning for Automated Tooth Segmentation and Landmark Localization on 3D Intraoral Scans. *IEEE Transactions on Medical Imaging* **41,** 3158–3166 (2021).

84. Drevický, D. & Kodym, O. *Evaluating Deep Learning Uncertainty Measures in Cephalometric Landmark Localization* in *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOIMAGING,* (SciTePress, 2020), 213–220. ISBN: 978-989-758-398-8.

85. Oktay, O., Schlemper, J., Folgoc, L. L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B. & Rueckert, D. *Attention U-Net: Learning Where to Look for the Pancreas* in *Medical Imaging with Deep Learning* (2018).

7 Discussed Papers

7.1 Papers Discussed in the Thesis

The following papers are discussed in this postdoctoral thesis. The best three of these papers are marked with an asterisk (*).

1. Klima, O., Kleparnik, P., Spanel, M. & Zemcik, P. *Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation* in *Proceedings of the SPIE* (eds Gimi, B. & Krol, A.) **9788.** Cited by (Scopus): 8; (Mar. 2016), 97880F.

2. Klima, O., Chromy, A., Zemcik, P., Spanel, M. & Kleparnik, P. A Study on Performace of Levenberg-Marquardt and CMA-ES Optimization Methods for Atlas-based 2D/3D Reconstruction. *IFAC-PapersOnLine* **49.** 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016, 121–126. ISSN: 2405-8963 (2016).

3. *Klíma, O., Madeja, R., Španel, M., Čuta, M., Zemčík, P., Stoklásek, P. & Mizera, A. *Virtual 2D-3D Fracture Reduction with Bone Length Recovery Using Statistical Shape Models* in *Shape in Medical Imaging* (eds Reuter, M., Wachinger, C., Lombaert, H., Paniagua, B., Lüthi, M. & Egger, B.) (Springer International Publishing, Cham, 2018), 207–219. ISBN: 978-3-030-04747-4.

4. *Kubík, T. & Španěl, M. *Robust Teeth Detection in 3D Dental Scans by Automated Multi-view Landmarking* in *Proceedings of the 15th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2022) - Volume 2: BIOIMAGING* Best Student Paper Award; (SciTePress, 2022), 24–34. ISBN: 978-989-758-552-4.

5. Kodym, O., Španěl, M. & Herout, A. Deep learning for cranioplasty in clinical practice: Going from synthetic to real patient data. *Computers in Biology and Medicine* **137.** Cited by (Scopus): 7; 104766. ISSN: 0010-4825 (2021).

6. *Kodym, O., Španěl, M. & Herout, A. Skull shape reconstruction using cascaded convolutional networks. *Computers in Biology and Medicine* **123.** Cited by (Scopus): 9; 103886. ISSN: 0010-4825 (2020).

7. Kodym, O., Španěl, M. & Herout, A. *Cranial Defect Reconstruction Using Cascaded CNN with Alignment* in *Towards the Automatization of Cranial Implant Design in Cranioplasty* (eds Li, J. & Egger, J.) Cited by (Scopus): 4; (Springer International Publishing, Cham, 2020), 56–64. ISBN: 978-3-030-64327-0.

7.2 Papers of High Scientific Value

3D modeling in medicine was not my only topic of interest. An essential step for the analysis of 3D image data is, for example, image segmentation. So I couldn't avoid this topic as well.

Because computer vision in the context of 3D data interests me in a broader context, I was also working on more distant topics such as the reconstruction and mapping of 3D space using cameras and sensors like Velodyne LiDAR. The 3D reconstruction with Velodyne LiDARs was motivated by my collaboration with the Czech geodetic company GEODROM s.r.o. and its CEO Jirka Habrovec, and it resulted in succesfull Ph.D. defense of my colleague Martin Veľas who was the main driver of practical research here.

These activities leaded to other publications which are not discussed within this postdoctoral thesis. The following papers were categorized as being of high scientific value according to my personal, subjective rating.

1. Kodym, o., Spanel, M. & Herout, A. *Segmentation of Head and Neck Organs at Risk Using CNN with Batch Dice Loss* in *German Conference on Pattern Recognition* Cited by (Scopus): 18; (2018).

2. Kodym, O. & Španěl, M. *Semi-automatic CT Image Segmentation using Random Forests Learned from Partial Annotations* in *Proceedings of the 11th International Joint Conference on Biomedical Engineering Systems and Technologies (BIOSTEC 2018) - Volume 2: BIOIMAGING* Cited by (Scopus): 8; (SciTePress, 2018), 124–131. ISBN: 978-989-758-278-3.

3. Velas, M., Spanel, M., Sleziak, T., Habrovec, J. & Herout, A. Indoor and Outdoor Backpack Mapping with Calibrated Pair of Velodyne LiDARs. *Sensors* **19.** Cited by (Scopus): 18; ISSN: 1424-8220 (2019).

4.   Velas, M., Spanel, M., Hradis, M. & Herout, A. *CNN for IMU assisted odometry estimation using velodyne LiDAR* in *2018 IEEE International Conference on Autonomous Robot Systems and Competitions (ICARSC)* Cited by (Scopus): 49; (Apr. 2018), 71–77.

5.   Velas, M., Spanel, M. & Herout, A. *Collar Line Segments for fast odometry estimation from Velodyne point clouds* in. **2016-June.** Cited by (Scopus): 55; (Institute of Electrical and Electronics Engineers Inc., 2016), 4486–4495. ISBN: 978-146738026-3.

6.   Velas, M., Spanel, M., Hradis, M. & Herout, A. *CNN for very fast ground segmentation in velodyne LiDAR data* in (eds J., C., L.C., B., P., O., H., C. & N., L.) Cited by (Scopus): 45; (Institute of Electrical and Electronics Engineers Inc., 2018), 97–103. ISBN: 978-153865221-3.

8 Original Papers

The papers discussed in this habilitation thesis are reprinted on the following pages.

Robust Teeth Detection in 3D Dental Scans by Automated Multi-view Landmarking

Tibor Kubík[1] and Michal Španěl[1,2] [a]

[1]*Department of Computer Graphics and Multimedia, Faculty of Information Technology,*
Brno University of Technology, Brno, Czech Republic
[2]*TESCAN 3DIM, Brno, Czech Republic*

Keywords: Landmark Detection in 3D, Polygonal Meshes, Multi-view Deep Neural Networks, RANSAC, U-Net, Heatmap Regression, Teeth Detection, Dental Scans.

Abstract: Landmark detection is frequently an intermediate step in medical data analysis. More and more often, these data are represented in the form of 3D models. An example is a 3D intraoral scan of dentition used in orthodontics, where landmarking is notably challenging due to malocclusion, teeth shift, and frequent teeth missing. What's more, in terms of 3D data, the DNN processing comes with high memory and computational time requirements, which do not meet the needs of clinical applications. We present a robust method for tooth landmark detection based on a multi-view approach, which transforms the task into a 2D domain, where the suggested network detects landmarks by heatmap regression from several viewpoints. Additionally, we propose a post-processing based on Multi-view Confidence and Maximum Heatmap Activation Confidence, which can robustly determine whether a tooth is missing or not. Experiments have shown that the combination of Attention U-Net, 100 viewpoints, and RANSAC consensus method is able to detect landmarks with an error of 0.75 ± 0.96 mm. In addition to the promising accuracy, our method is robust to missing teeth, as it can correctly detect the presence of teeth in 97.68% cases.

1 INTRODUCTION

The localization of landmarks plays a crucial role in many tasks related to image analysis in medicine. Deep learning has demonstrated great success in this field, outperforming conventional machine learning methods. With the widespread availability of accurate 3D scanning devices, this task has moved into a 3D domain. This brings us the possibility of increased automation of clinical application tasks that operate on 3D models, such as in the case of digital orthodontics.

In terms of direct 3D data processing by neural networks, a noticeable challenge has emerged as the size of the input feature vector substantially increases. The time of computation of such deep neural networks is not suitable for clinical applications used during treatment planning in digital orthodontics. 3D medical data analysis reckons with another challenge – the limited amount of medical data, a common struggle in medical image processing.

Dentition casts used in digital orthodontics software are typically obtained from patients with various levels of malocclusion and numerous kinds of teeth shifting. Another challenging problem in this domain is the absence of teeth, a common phenomenon in terms of human dentition. The 3rd Molars (also known as *Wisdom teeth*) are worth taking a look at. Their extraction is one of the most frequent procedures in oral surgery as it eliminates future problems due to unfavorable orientation (Normando, 2015). Thus, the method should be robust to such variations.

In this paper, we present a method that considers the limitation of the dataset size, the need for low computational time, and the importance of robustness to missing and shifted teeth. It is based on a multi-view approach and it uses heatmap regression to predict landmarks in 2D and the RANSAC consensus method to robustly propagate the information back into 3D space. In order to address the problem of estimation of landmarks on missing teeth, our method comprises a post-processing based on a heatmap regression uncertainty analysis combined with the uncertainty of the multi-view approach.

[a] https://orcid.org/0000-0003-0193-684X

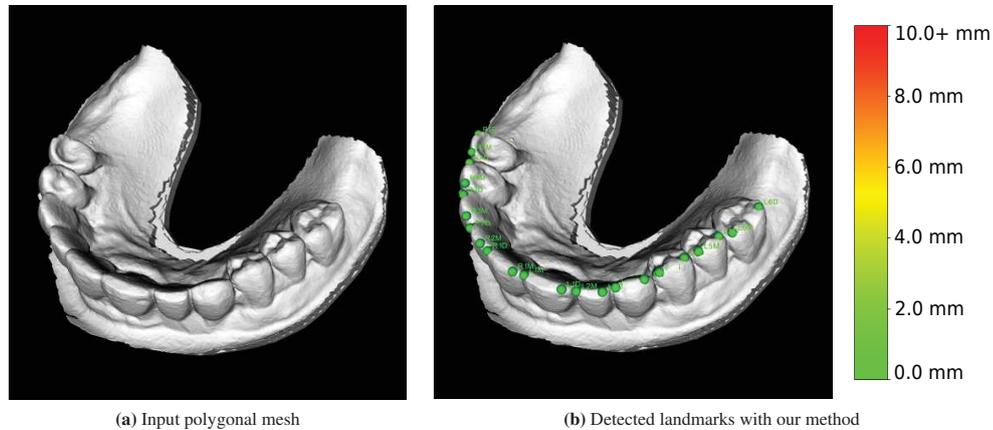(a) Input polygonal mesh   (b) Detected landmarks with our method

Figure 1: An example of a 3D scan of dentition (a) and appropriate detected landmarks (b). Our method automatically detects two landmarks on each tooth – mesial and distal. This type of landmarks is important in orthodontics, as it defines the rotation of teeth from anatomically perfect arrangement. Whatsmore, the method correctly detects whether a tooth is missing or not.

Conducted experiments have shown that the proposed method can detect orthodontics landmarks on surface models with an error of 0.75 ± 0.96 mm while 98.07% of detected landmarks achieve an error less than 2 mm. As for the robustness to missing teeth, our method's post-processing correctly detects missing teeth in 97.68% of cases.

2 CURRENT APPROACHES TO LANDMARKING

Early studies in this area relied on conventional machine learning approaches. Hough forests were used for landmark detection. Authors in (Donner et al., 2013) combined regression and classification, which brought better results comparing to both a single voxel's classification and classification of the volume of interest. As convolutional neural networks (CNNs) gained in popularity, an increasing number of scientific papers concerning their usage in landmark detection emerged. Some of these methods detected the landmark position directly by regressing its x and y coordinates. For example, in (Sun et al., 2013), the authors adopted cascaded convolutional neural networks for facial point detection. Another study (Lv et al., 2017) proposed a regression in a two-stage manner, still locating landmarks directly.

2.1 Heatmaps in Landmarking

Over time, extensive literature has developed on landmarking by heatmap regression. The authors in (Pfis-

ter et al., 2015) worked on a model that regresses human joint positions. Instead of directly regressing the (x, y) joint position, they regressed a joint position's heatmap. During the training, the ground truth labels are transformed into heatmaps by placing a Gaussian with fixed variance at each joint coordinate.

On top of the appliance of spatial fusion layers and optical flow, they discussed the benefits of regressing a heatmap rather than (x, y) coordinates directly. They concluded that the benefits are twofold: (i) the process of network training can be visualized in such a way that one can understand the network learning failures, and (ii) the network output can acquire confidence at multiple spatial locations. The incorrect ones are slowly suppressed later in the training process. In contrast, regressing the (x, y) coordinates directly, the network would have a lower loss only if it predicts the coordinate correctly, even if it was "growing confidence" in the correct position. Concerning these, such an approach outperformed direct coordinate regression and became a standard way of landmark detection in 2D images.

This approach seemed alluring for people in the medical image processing community. Inspired by this method, authors in (Payer et al., 2016) presented multiple architectures that detect keypoints in X-Ray images of hands and 3D hand MR scans. They affirmed that by regressing heatmaps, it is possible to achieve state-of-the-art localization performance in 2D and 3D domains while dealing with medical data shortage.

25

2.2 Processing of 3D Data by Neural Networks

Although the extension of deep neural network operations such as convolution from 2D to 3D domain seems natural, the additional computational complexity introduces notable challenges. Having volumetric data (for example, voxel models) or 3D surface data (for example, represented as polygon meshes) as an input to deep neural networks has a considerable drawback in computational time and memory requirements.

An alternative way of 3D data processing by neural networks is the *multi-view approach*. Obtaining state-of-the-art results on 3D classification, authors in (Su et al., 2015) presented the multi-view CNN idea. It is relatively straightforward and consists of three main steps:

1. Render a 3D shape into several images using varying camera extrinsics.

2. Extract features from each acquired view.

3. Process features from different viewpoints in a way suitable for a given task. In (Su et al., 2015), a pooling layer followed by fully connected layers was used to get class predictions.

The multi-view approach was later on used to identify feature points on facial surfaces (Paulsen et al., 2018). The authors discussed multiple geometry derivatives and experimented with their combinations to bring state-of-the-art results in feature point detection on facial 3D scans while decreasing the prohibitive GPU memory requirements needed for true 3D processing. Additionally, they proposed a consensus method to find the final estimate, which combines the *least-squares fit* and *RANdom SAmple Consensus* (RANSAC) (Fischler and Bolles, 1981). For each landmark, *N* rays in 3D space are the outputs of the proposed method.

Based on Graph Neural Networks (GNNs), authors in (Sun et al., 2020) presented coupled 3D segmentation for annotation of individual teeth and gingiva. Their network produces a dense correspondence that helps to accurately locate individual orthodontics landmarks on teeth crowns. Another recent work in landmark localization on dental mesh models was presented by authors in (Wu et al., 2021). They introduced a two-stage framework based on mesh deep learning (TS-MDL) for joint tooth labeling and landmark identification. To accurately detect tooth landmarks, they designed a modified PointNet (Qi et al., 2017) to learn the heatmaps encoding landmark locations.

We have developed a generic method based on the current approaches in landmarking to solve a variety of problems that arose from the medical character of the dataset:

- the method should be robust to missing teeth,

- tens of cases should be sufficient to train the network,

- and the speed of the inference should be fast enough to be used in a clinical application.

Especially valuable is the introduced post-processing based on heatmap regression uncertainty analysis and analysis of the uncertainty of the multi-view approach. It ensures that our method correctly detects landmark presence without any additional computations. This is inevitable for orthodontic flow as it robustly detects teeth presence even in challenging cases (e.g., already discussed 3rd molars). This aspect was not discussed in recent works that deal with orthodontics landmarks on teeth crowns.

In addition to the post-processing and the method itself, this paper presents valuable comparisons and experiments on various factors that impact the efficiency of alternative variations of the method:

- rendering type of the processed 3D object to be used as an input (depth map, geometry or combination of both),

- comparison of several network designs (U-Net, Attention U-Net, and Nested U-Net),

- the analysis of the results of two consensus methods: a method that calculates the centroid of multiple predictions and a geometric method based on the RANSAC algorithm and least-squares fit,

- and the analysis of the correlation between the number of viewpoints and the method accuracy.

3 DATASET OF 3D DENTAL SCANS AND LANDMARKS IN THIS STUDY

Our method was trained and evaluated on a dataset of 337 3D dental scans of human dentition represented as polygon meshes. The dataset contains cases of both maxillary and mandibular dentition. Since all dentition scans were anonymized, it is not possible to undertake complex analysis of patients' age or ethnicity. Therefore, the data analysis was empirical and focused on aspects such as the frequency of absence of teeth, the rate of healthy dentition, and dentition with malocclusion and shifted teeth. Concerning these aspects, our data reflect real orthodontics patients since
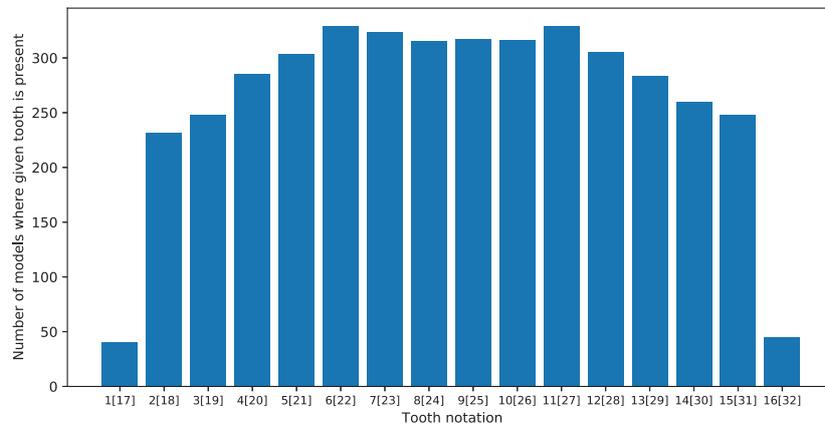
Figure 2: Distribution of casts where given tooth is present on the dentition. For example, out of 337 scanned dentition from the dataset, less than 50 cases contain either left or right 3rd molar. This distribution reflects the reality, as 3rd molars are often extracted (Normando, 2015). On the other hand, canines and incisors are present in the vast majority of models. Please note that the Universal Numbering System is used to refer teeth. Also note that teeth 1 and 17 are considered as the same category, likewise to the rest of the teeth.
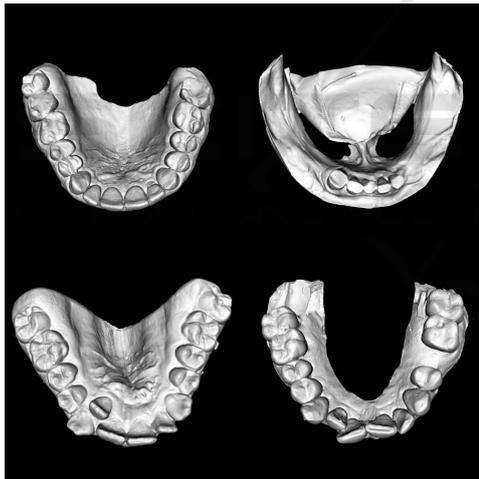

Figure 3: Examples of dental casts within the dataset. Data were collected from orthodontics patients, so patients usually suffer from different kinds of malocclusion, as depicted on the bottom examples.

the diversity of data is significant, which is essential for the algorithm's robustness. Figure 3 depicts the variety of dentitions in our dataset. The frequency of missing teeth confirms the diversity in orthodontics cases as well. Figure 2 shows the number of cases where individual teeth are not missing within the dataset. Landmarks used in this study address the digital orthodontics flow in the existing planning software. These landmarks define the mesial and distal

location of each tooth. They are placed on the occlusal surface of molars and premolars and the incisal surface on canines and incisors, as close to the cheek-facing surfaces as possible. In other words, 32 landmarks must be placed on one arch in case of full dentition, two for each tooth. Ground truth positions of landmarks were annotated by one person only.

4 PROPOSED SOLUTION FOR ORTHODONTICS LANDMARK DETECTION

An outline of our method can be found in Figure 4. Prior to each evaluation, there is a precondition to align the evaluated mesh so the occlusal surfaces face the camera. Afterward, following the multi-view approach, the model is observed from various camera extrinsics. We used uniformly distributed positions of the camera with a maximal angle of ± 30 degrees from the initial aligned position.

Network Inputs and Outputs

Images in the form of depth maps and direct rendering of the geometry are used as the inputs to the neural network.

From each acquired view, features are extracted and processed in the heatmap regression manner. In a similar way as in (Pfister et al., 2015), during training, the input example is denoted as a tuple (X, y),
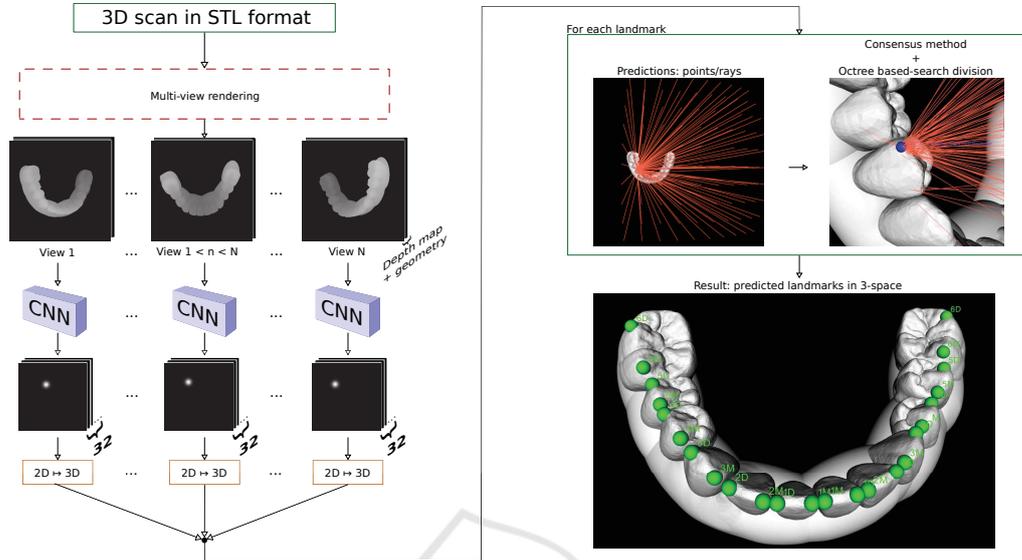
27

Figure 4: Outline of the proposed method for orthodontics landmark detection. Following the multi-view approach, input 3D model is observed from various viewports and sent to the CNNs to produce heatmaps. Landmark screen coordinates are extracted from obtained heatmaps and further processed by the consensus method, which produces final estimates. Additionally, the maximum value in the activation map, together with the output of the consensus method, are used to detect tooth presence during post-processing.

where X is the 2-channel input and y stands for the coordinates of 32 landmarks located in input X. Furthermore, the training data are denoted as $N = \{X, y\}$ and the network regressor as ϕ. Then, the training objective is the estimation of the network weights λ:

$$\arg\min_{\lambda} \sum_{(\mathbf{X},\mathbf{y}) \in N} \sum_{i,j,k} \| G_{i,j,k}(\mathbf{y}_k) - \phi_{i,j,k}(\mathbf{X}, \lambda) \|^2 \quad (1)$$

where $G_{i,j,k}(y_i) = \frac{1}{2\pi\sigma^2} e^{-[(y_k^1 - i)^2 + (y_k^2 - j)^2]/2\sigma^2}$ is a Gaussian centered at landmark y_k with fixed σ. Using this approach, the last convolutional layer's output is a heatmap represented as a fixed-size $i \times j \times 32$-dimensional matrix. This implies that the the predicted results are 32 channels (as we intend to predict 32 landmarks in our data).

Interpretation of Heatmap Regression Output in Terms of 3D Data

The predicted 2D heatmap can be interpreted as the landmark's screen coordinate (in $\mathbb{R}^2$) position (x, y). Each output channel contains a heatmap with a Gaussian representing the probability of a given landmark's screen coordinate in each pixel. Thus, the resulting screen coordinate must be extracted from the predicted heatmap by finding coordinates of the peak value. It is indispensable to propagate the coordinates into a world coordinate system $\mathbb{R}^3$ and find a final estimate by combining outputs from all camera views.

With the known position of the center of projection, the prediction for a single view of one landmark can be interpreted as (i) **a ray** defined by the origin in the corresponding center of projection and the point on the view plane at detected screen coordinates or (ii) simply **a point** in the 3D scene, i.e. the converted display coordinate into 3D space.

Consensus Methods

These individual predictions are combined in a consensus method, which is a standard post-processing step in the multi-view approach. Based on the maximum value in the activation map, only certain predictions above the experimentally determined threshold are sent to the consensus method. Certainty analysis will be discussed later in this work. If the predictions are interpreted as rays, the consensus method combines the RANSAC algorithm to eliminate partial predictions classified as outliers with the least-squares fit.

To achieve this, we defined each ray by its origin a_i and a unit direction vector n_i, similarly as (Paulsen et al., 2018). Then, the sum of squared distances from

a point p is calculated as follows:

$$\sum_i d_i^2 = \sum_i [(p-a_i)^T(p-a_i) - [(p-a_i)^T n_i]^2]. \quad (2)$$

It is necessary to differentiate this equation with respect to p. It brings the solution $p = S^+C$, where S^+ denotes the pseudo-inverse of S. In this case, $S = \sum_i(n_i n_i^T - I)$ and $C = \sum_i(n_i n_i^T - I)a_i$. The RANSAC procedure initially estimates the value of p by three randomly chosen rays. The residual is computed as the sum of squared distances (see Equation 2) from p to the included rays, and the iterative RANSAC algorithm then performs I iterations. In each of these iterations, the number of *inliers* and *outliers* is calculated, respecting a predefined threshold τ. This is a minimizing task that finds a point in $\mathbb{R}^3$ with the shortest distance to all remaining lines.

This method can be interchanged with a more statistical approach that is less computationally demanding, and it simply finds the mean position of the predicted points. Let's consider N as the number of views used in the multi-view approach. Let's also interpret the single-view evaluation output as a point on the target polygonal model. With N views, the final output P is a single point in $\mathbb{R}^3$ and is calculated from N points as a mean value of these points.

Finding Closest Point on Mesh Surface

These methods find the estimation among multiple predictions, but do not guarantee that the predicted landmark is placed on the surface of the evaluated polygonal model. Thus, the last necessary step is to find the closest point on the surface of the polygonal model. An octree data structure contains a recursively subdivided target polygonal model. The center of the closest face on the surface of the polygonal model to the consensus output is considered the final estimate.

4.1 Post-processing for Classification of Teeth Presence

As discussed in previous sections, assuming that the evaluated 3D scan represents full dentition would be loose. Therefore, apart from the accurate placement of the present landmarks, our post-processing contains an analysis of the presence of each tooth (i. e., of corresponding couple of landmarks). This is in fact a binary classification task, whose result is based on two uncertainty hypotheses:

- Like in (Drevický and Kodym, 2020), the network is trained to regress heatmaps with the amplitude of 1. Then, the fundamental assumption
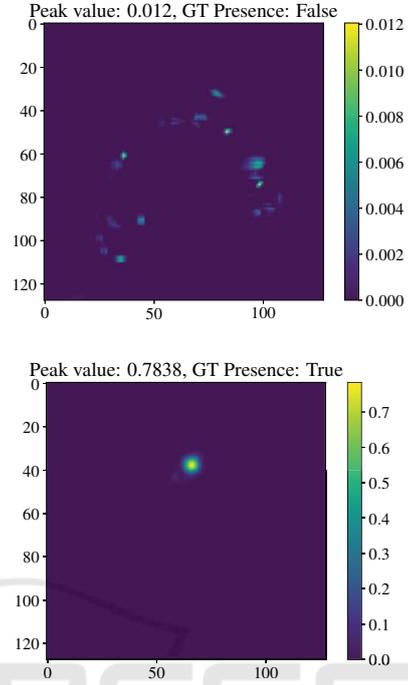

Figure 5: Examples of predicted heatmaps and analysis of the uncertainty. The top picture illustrates an example of a prediction with low peak value (0.012). Referencing to corresponding ground truth, this landmark is not present on the surface of the polygonal model. The bottom picture, on the other hand, shows the opposite situation. According to the ground truth, the peak value is relatively high, and this landmark is really present on evaluated polygon mesh. Note that the maximal amplitude value in a heatmap is 1.

is that during the inference, the certainty is measured by the maximum value in the activation map, with a proportional increase to the network's confidence (*Maximum Heatmap Activation Confidence*). See Figure 5 for an example.

- The RANSAC consensus method robustly estimates the landmark position by eliminating outlier predictions. Thus, the proportion of inliers and outliers is another valuable output of this consensus method, assuming the number of inliers is proportional to the overall confidence (*Multi-view Confidence*).

These assumptions result in a threshold value, which combines the Maximum Heatmap Activation Confidence and Multi-view Confidence, both in a unit range and equally weighed. The optimal threshold value can be determined by standard approaches for

29

a binary classifier, as an example by the ROC curve. This goes to show that such post-processing delivers vital data for classification of landmarks presence by *self-evaluation*, i. e., no additional computations or network evaluations are needed to obtain such information. Having the requirement of low computational time in mind, this is more than eligible.

5 EXPERIMENTS AND RESULTS

To find the best possible results, we experimentally investigated and compared several parts of the method:

- **Architecture Design:** we compared the U-Net architecture with two of its offshoots: the Attention U-Net and the Nested U-Net.

- **Consensus Methods:** a comparison of RANSAC consensus method with centroid calculation is presented.

- **Viewpoint Numbers for the Multi-view Approach:** we analysed whether the increase of viewpoint number has an impact on the method accuracy. We experimented with 1, 9, 25, and 100 views.

- **CNN Inputs:** depth map, direct geometry rendering and its combination (2-channel input) were compared.

All metrics are measured in physical units (mm) since the end clinical application is related to physical units.

5.1 Training Procedure

The input to the neural network is either a single-channel depth map, single-channel image of the rendered geometry, or two-channel combination of both, depending on experiment. In all cases, the size of input was set to 128×128. The training procedure ran on an NVIDIA GeForce RTX 2060 with 6 GB of memory.

The dataset of 337 dental scans was divided into a set of 247 cases used for training and a test set of 90 cases. Furthermore, the training set was split in the ratio of approximately 4:1 into a training and validation set, respectively.

Following augmentation techniques were applied to both, the 2D input(s) and the ground truth heatmaps:

- **Scale:** in the range $[0.90, 1.10]$,

- **Rotation:** in the range $[-30, 30]$ degrees,

- **Translation:** in the range $[-10\,\text{px}, 10\,\text{px}]$ and applied in both vertical and horizontal directions.

Training Parameters and Loss Function

Networks were trained using the Adam optimizer with the weight decay set to 10^{-3}. The learning rate was initially set to 10^{-3}. Its value was dynamically reduced using *learning rate scheduler*. The learning rate was reduced by a factor of 0.5 every time the value of validation loss has not improved for 5 consecutive epochs. The validation loss was monitored for the *early stopping*. If the validation loss value did not improve for more than 20 consecutive epochs, the training was stopped. To reduce the memory requirements during training, the *automatic mixed precision* was used. The batch size was set to 32. To train the models on a regression problem, the Root Mean Square Error (RMSE) loss was used.

5.2 Overall Results

The main focus of the experiments was to find the best setup of the method. Overall results are summarized in Table 1. Our results show that the acquired accuracy is mostly influenced by the consensus method, where RANSAC outperforms the Centroid by a large margin in all setups. As for the used architecture, the overall results show that the Attention U-Net performs slightly better than the rest. Combination of depth maps and geometry renders impacts the results in a positive way as well. See Figure 7 for box plots of radial errors of individual detected landmarks. The Attention U-Net has 526 534 trainable parameters and inference takes 4 seconds on average on Intel Core i7-8750H CPU @ 2.20 GHz with 6 cores (using 25 views).

When comparing our results to the framework from (Wu et al., 2021), specifically with their best-performing strategy, *2-stage iMeshSegNet+PointNet-Reg*. In terms of accuracy, they achieve a slightly better error of 0.623 ± 0.718 mm. Their approach slightly outperforms ours (in best-performing configuration, 0.75 ± 0.96 mm), but it is necessary to keep in mind several factors. As a matter of fact, their dataset consists of 36 samples. Such relatively small number should be increased to ensure the method's robustness to the large variability of orthodontic cases. Our dataset is more challenging and consists of problematic cases with severe teeth shiftings and of many cases with missing teeth. In addition, they detect landmarks only on 10 teeth, excluding, for example, very problematic 3rd molars. Thus, for a fair comparison, it would be vital to benchmark

Table 1: Overall results of the individual networks with different multi-view settings. Table compares the system performance with different combinations of architectures, network inputs, consensus methods, and number of viewpoints. A combination of the Attention U-Net architecture, the RANSAC consensus method, and 100 rendered views achieves the best performance. $\overline{R}$ stands for the mean radial error, and *SD* stands for standard deviation. Values are calculated from all predicted landmarks on dental scans from the test dataset and measured in millimeters (mm). All values are measured on networks with class-balanced loss. Please note that the alignment of evaluated 3D scans influence the measured values.

Architecture & consensus method		Single-view		Multi-view					
				$N = 9$		$N = 25$		$N = 100$	
		$\overline{R}$	SD	$\overline{R}$	SD	$\overline{R}$	SD	$\overline{R}$	SD
BN U-Net (Depth)	Centroid	2.24	4.02	2.00	2.37	1.74	2.33	1.80	1.96
	RANSAC			1.24	2.86	1.02	3.75	1.01	4.28
BN U-Net (Geom)	Centroid	2.13	4.41	2.03	3.14	1.69	2.21	1.67	2.41
	RANSAC			1.20	3.01	1.17	2.16	1.06	2.22
BN U-Net (Depth & Geom)	Centroid	2.02	4.10	1.90	2.12	1.82	2.48	1.85	3.23
	RANSAC			1.01	3.77	0.84	2.05	0.77	1.94
Att U-Net (Depth)	Centroid	1.73	3.48	2.37	3.37	2.02	2.87	2.01	1.99
	RANSAC			1.18	1.88	1.10	2.05	0.95	1.62
Att U-Net (Geom)	Centroid	1.72	3.62	2.31	2.68	1.98	2.09	1.96	2.38
	RANSAC			1.14	1.51	1.02	3.75	0.91	1.11
Att U-Net (Depth & Geom)	Centroid	1.67	3.06	2.00	2.37	1.74	2.33	1.80	1.96
	RANSAC			0.93	1.03	0.79	1.01	**0.75**	**0.96**
Nes U-Net (Depth)	Centroid	1.77	3.32	2.29	2.12	2.32	1.99	2.12	3.04
	RANSAC			1.09	2.60	1.00	1.85	0.95	2.82
Nes U-Net (Geom)	Centroid	1.77	3.00	2.44	1.98	2.30	3.01	2.23	2.58
	RANSAC			1.11	1.83	0.93	1.67	0.93	1.99
Nes U-Net (Depth & Geom)	Centroid	1.69	2.62	2.30	3.18	2.31	2.72	2.16	2.55
	RANSAC			0.98	2.09	0.83	2.12	0.80	1.45

our results on a public dataset, which is not currently available.

Impact of Viewpoint Number

As for the number of views used in the multi-view approach, a negligible increase in accuracy is achieved, comparing 25 and 100 views. This increase in viewpoint number, however, significantly raises the inference time, so it is necessary to cross-validate this number to obtain desirable accuracy as well as computational time. For example, an increase of 0.04 mm in accuracy as a trade-off for $4\times$ higher inference time is considerable. See Figure 6, which analyzes the Success Detection Rate (SDR) of various numbers of views.

Robustness to Model Rotations

Generally speaking, the multi-view approach is not invariant to rotation. The requirement of initial model alignment stems from this matter of fact. Therefore, we were interested in how the method performs with increasing alignment error. With an alignment error of less than 20 degrees, the method brings sufficiently accurate predictions. With higher alignment errors, especially above 30 degrees, the results should be visually checked and if needed, manually fixed. This correlation is depicted in Figure 8.
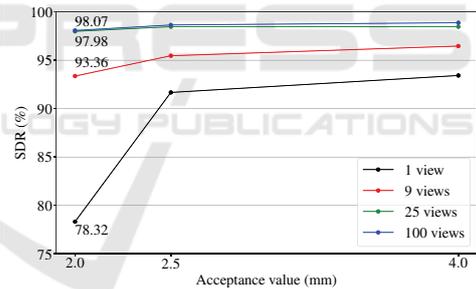

Figure 6: Success Detection Rates (SDRs) for Attention U-Net, 2-channel input and the RANSAC consensus method. Assuming the acceptable distance is 2 mm, setting the number of viewpoints higher than 25 does not bring any significant increase in performance.

5.3 Detection of Teeth Presence

The main focus of the experiments was to determine whether the method's *self-evaluation* can detect the presence of landmarks (and therefore, teeth). In line with previous studies in uncertainty measures, each prediction's peak value is considered one of the decision factors. Networks were trained by regressing heatmaps containing a Gaussian activation with the amplitude of 1. The predictions should follow the similar trend. There was no Gaussian in the ground
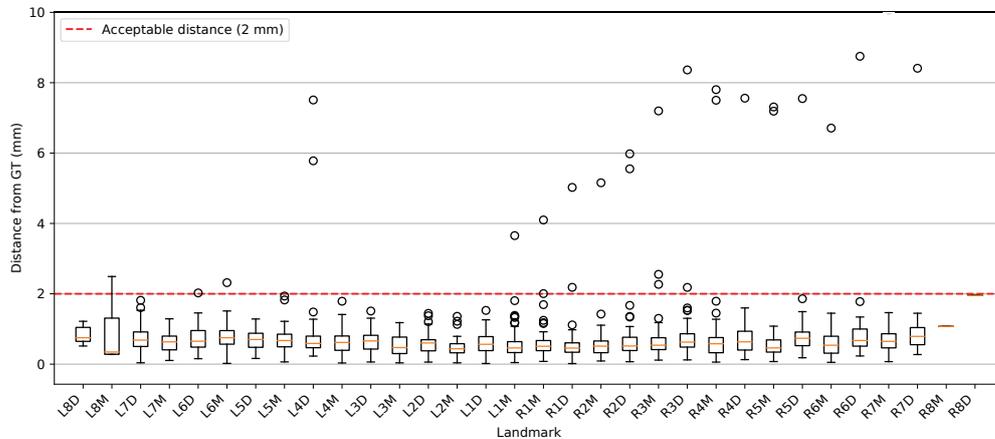
31

Figure 7: Box plots of the radial error values of individual landmarks. These values were measured with following method configuration: Attention U-Net, two-channel input, 25 views, and RANSAC consensus method. Additionally, the class-balanced loss was used for training. The landmark notation describes the type of landmark as follows: L stands for Left dentition part and R for Right, values 1 - 8 describe tooth in the quadrant (1 for centran incisor and 8 for 3rd molars) and letters M and D stand for mesial and distal landmark, respectively. Note that the outlier values in Right dentition part were caused by one problematic case, where all teeth in right part were shifted by one and our method misclassified each tooth with its adjacent tooth.
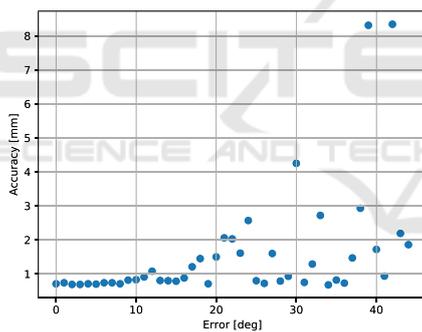

Figure 8: Correlation between error from required alignment and landmarking accuracy. As the 3D model is observed from different angles, the method robustly estimates landmarks even when the model is slightly rotated. Overall, the method becomes less stable with increasing error in alignment, especially above 30 degrees.

truth image if a landmark was missing on the polygonal model during training. This implies that the predictions should be either heatmaps with a peak value close to 1 or heatmaps with all values close to 0.

By plotting an ROC curve, it was found that the threshold value that brings off the best *sensitivity* and *specificity* values is 0.375. Please note that this value should be always cross-validated for each task. The accuracy of the detection was 96.36%. After empirical observations, there were situations where on a tooth, one landmark was classified as missing and the second one as present. This undesirable situation was eliminated by measuring the certainty in couples, averaging its confidences. It leads to better results, even if the improvement is negligible, achieving an accuracy of 96.69%. Another promising finding comes from the RANSAC consensus method output. The Multi-view Confidence, measured as the ratio between inliers and outliers, was again monitored by an ROC curve. The threshold was set to 0.85 and combined with the analysis of heatmap maximum value. Superior results are seen for this combination, as 97.68% of landmarks are correctly classified as missing or present.

Detecting Presence of 3rd Molars

A special category of detected teeth is 3rd molars. As discussed in Section 3, those teeth are represented in approximately 15% of the cases. The approach utilized for detection of teeth presence suffers from this imbalance, as the 3rd molars were always classified as missing. This was due to the training, where, in most cases, wisdom teeth were not present. To address this problem, the loss was balanced in class-wise manner (Cui et al., 2019). With this technique, 9 out of 12 wisdom teeth in the test set were correctly detected.
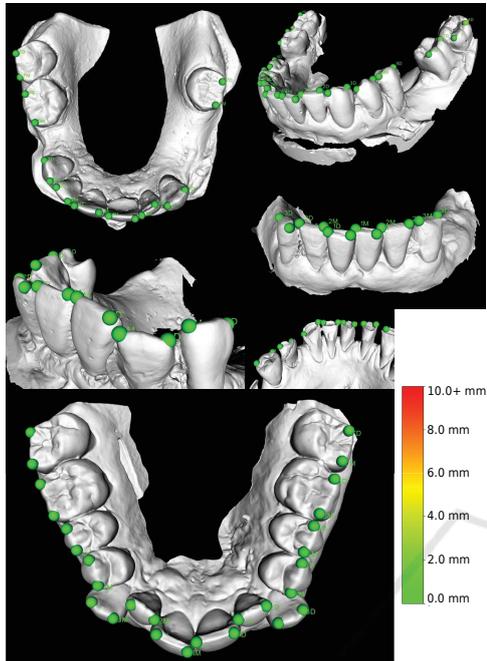
Figure 9: Examples of automatically detected landmarks with our method. Majority of predictions have the landmark localization error less than 2 mm. Our method correctly detects if a tooth is missing and does not produce predictions of corresponding landmarks.

6 CONCLUSIONS

The present findings confirm that the multi-view approach combined with the RANSAC consensus method brings promising results in the automation of landmark detection. Evaluated on a dataset of real orthodontics dental casts with significant diversity, the method performs the best with Attention U-Net architecture and with two-channeled input of depth maps and geometry renders. This method setup achieves a landmarking accuracy of 0.75 ± 0.96 mm.

Importantly, we have also shown that the uncertainty measures based on the analysis of the maximum values of regressed heatmap predictions in combination with multi-view uncertainty yield convenient information in the process of landmark presence detection. Combining these uncertainty measures, our method correctly detects landmark presence in 97.68% of cases. This means that the method is suitable to be applied to data where landmarks' presence is not granted. In addition, the method meets the needs of clinical applications, as the inference at

the user's side takes seconds to be calculated, even on less powerful CPUs.

Even though the accuracies are satisfying, the size of the dataset could not cover every bit of a malocclusion case and teeth shifting. Future research could examine the method on a larger dataset of dentition with even more complex cases. Furthermore, future studies should focus on the improvements in the invariance of rotation. The association between the rotation from the aligned position and the landmarking accuracy was investigated in this work, and it is the main shortcoming of the proposed method.

ACKNOWLEDGEMENTS

REFERENCES

Cui, Y., Jia, M., Lin, T.-Y., Song, Y., and Belongie, S. (2019). Class-balanced loss based on effective number of samples.

Donner, R., Menze, B. H., Bischof, H., and Langs, G. (2013). Global localization of 3d anatomical structures by pre-filtered hough forests and discrete optimization. *Medical Image Analysis*, 17(8):1304–1314.

Drevický, D. and Kodym, O. (2020). Evaluating deep learning uncertainty measures in cephalometric landmark localization. In *Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies - BIOIMAGING,*, pages 213–220. INSTICC, SciTePress.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6):381–395.

Lv, J., Shao, X., Xing, J., Cheng, C., and Zhou, X. (2017). A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pages 3691–3700. IEEE Computer Society.

Normando, D. (2015). Third molars: To extract or not to extract? *Dental press journal of orthodontics*, 20(4):17–18.

Paulsen, R. R., Juhl, K. A., Haspang, T. M., Hansen, T. F., Ganz, M., and Einarsson, G. (2018). Multi-view consensus CNN for 3d facial landmark placement. In Jawahar, C. V., Li, H., Mori, G., and Schindler, K., editors, *Computer Vision - ACCV 2018 - 14th Asian Conference on Computer Vision, Perth, Australia, December 2-6, 2018, Revised Selected Papers, Part I,*

33

volume 11361 of *Lecture Notes in Computer Science*, pages 706–719, Cham. Springer.

Payer, C., Stern, D., Bischof, H., and Urschler, M. (2016). Regressing heatmaps for multiple landmark localization using cnns. In Ourselin, S., Joskowicz, L., Sabuncu, M. R., Ünal, G. B., and Wells, W., editors, *Medical Image Computing and Computer-Assisted Intervention - MICCAI 2016 - 19th International Conference, Athens, Greece, October 17-21, 2016, Proceedings, Part II*, volume 9901 of *Lecture Notes in Computer Science*, pages 230–238. Springer.

Pfister, T., Charles, J., and Zisserman, A. (2015). Flowing convnets for human pose estimation in videos. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 1913–1921. IEEE Computer Society.

Qi, C. R., Su, H., Mo, K., and Guibas, L. J. (2017). Pointnet: Deep learning on point sets for 3d classification and segmentation.

Su, H., Maji, S., Kalogerakis, E., and Learned-Miller, E. G. (2015). Multi-view convolutional neural networks for 3d shape recognition. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 945–953. IEEE Computer Society.

Sun, D., Pei, Y., Li, P., Song, G., Guo, Y., Zha, H., and Xu, T. (2020). Automatic tooth segmentation and dense correspondence of 3d dental model. In *MICCAI*.

Sun, Y., Wang, X., and Tang, X. (2013). Deep convolutional network cascade for facial point detection. In *2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, June 23-28, 2013*, pages 3476–3483. IEEE Computer Society.

Wu, T.-H., Lian, C., Lee, S., Pastewait, M., Piers, C., Liu, J., Wang, F., Wang, L., Jackson, C., Chao, W.-L., Shen, D., and Ko, C.-C. (2021). Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3d intraoral scans.

34

Deep learning for cranioplasty in clinical practice: Going from synthetic to real patient data

Oldřich Kodym [*], Michal Španěl, Adam Herout

Department of Computer Graphics and Multimedia, Brno University of Technology, Božetěchova 2, 612 66, Brno, Czech Republic

ARTICLE INFO

ABSTRACT

Correct virtual reconstruction of a defective skull is a prerequisite for successful cranioplasty and its automatization has the potential for accelerating and standardizing the clinical workflow. This work provides a deep learning-based method for the reconstruction of a skull shape and cranial implant design on clinical data of patients indicated for cranioplasty. The method is based on a cascade of multi-branch volumetric CNNs that enables simultaneous training on two different types of cranioplasty ground-truth data: the skull patch, which represents the exact shape of the missing part of the original skull, and which can be easily created artificially from healthy skulls, and expert-designed cranial implant shapes that are much harder to acquire. The proposed method reaches an average surface distance of the reconstructed skull patches of 0.67 mm on a clinical test set of 75 defective skulls. It also achieves a 12% reduction of a newly proposed defect border Gaussian curvature error metric, compared to a baseline model trained on synthetic data only. Additionally, it produces directly 3D printable cranial implant shapes with a Dice coefficient 0.88 and a surface error of 0.65 mm. The outputs of the proposed skull reconstruction method reach good quality and can be considered for use in semi- or fully automatic clinical cranial implant design workflows.

1. Introduction

Cranioplasty is a procedure that restores the aesthetic, mechanical, and protective function of a defective skull by implanting material into the defect area. Although autologous bone or pre-formed titanium meshes can be used as implants, 3D printable implants have been shown to be more versatile and to have several other advantages, such as a lower risk of complications or lower chance of requiring secondary surgery [1,2]. Manufacturing these implants requires modeling their shape in computer-assisted design (CAD) software as the first step. This virtual reconstruction, however, requires the human operator to have sufficient knowledge of skull anatomy as well as skill in 3D modeling. Even if these requirements are met, correctly modeling the implant is time-consuming even for a skilled operator, especially in cases of defects reaching into both lateral sides of the skull [3]. Automatically producing fast and precise estimations of the implant shapes could therefore lead to increased standardization and efficiency of cranioplasty clinical workflow. In recent years, skull shape reconstruction methods based on volumetric convolutional neural networks (CNNs) have shown great promise in this regard [4,5,6], yet they remain mostly untested on real

patient data, which limits their potential of translation into clinical practice. This article deals with the issue of using these CNN-based models on real patient data and improving their performance with the use of multi-task learning, as illustrated in Fig. 1.

Most recent (semi-)automatic skull reconstruction methods aim to solve the task of finding the exact shape of the missing part of the skull. We refer to this type of reconstruction output as *a skull patch* in this article. The main criteria for a successful skull patch estimation is an anatomically plausible, symmetric shape with a smooth and seamless fit along the defect border. In clinical practice, this allows the operator to use the estimated skull patch as a template for the final cranial implant design in CAD software. Conventional skull reconstruction methods use mirroring of the healthy side of the skull onto the defective side [7], surface interpolations [8,9] or their combination [10] to estimate the skull patch. Statistical shape models [11] greatly expanded the range of skull defects that can be reconstructed automatically [12,13,14]. In recent years, the research focus shifted to volumetric convolutional neural networks (CNNs) which have shown great promise in fast and robust skull patch reconstruction [4,6,15] and became the method of choice in the 2020 AutoImplant challenge [5]. The CNN-based methods

---

* Corresponding author.
E-mail address: ikodym@fit.vutbr.cz (O. Kodym).

are usually trained and evaluated using synthetic defects created by removing some part of a healthy skull, resulting in virtually an infinite amount of different samples.

The final shape of the cranial implant (referred to simply as *implant* for the remainder of this article) differs from the shape of the skull patch in several ways (see Fig. 2). The implants have a constant thickness different from the original bone and have some spatial tolerance along the defect border to account for scar tissue and continuing bone growth, ensuring implantability. The shape of the implant can also be estimated directly by a CNN model, provided that sufficient training data is available for training. Although it is more difficult to edit this kind of shape in CAD software due to fine details along the defect border, it has the potential to be used in a fully automatic setting when no human operator, or not enough time for a manual design, is available, for example in intra-operative rapid manufacturing of cranial implants [16].

Synthetic datasets for automatic estimation of skull patches recently became available because they are easy to create from public databases of healthy skulls, such as CQ500 [17]. However, they do not necessarily fully cover the defective skull shape distribution of target clinical data (i. e. different anatomical variability of the target population, defect shapes and sizes, complex morphology of defect border), which may affect the resulting reconstruction quality in practice [4]. Real clinical data with expert-designed implant data are, on the other hand, difficult to obtain. Furthermore, in our experience, the distribution of available clinical data is often biased towards simple unilateral defect cases and not easily extendable by synthetic defect and implant shapes. The more challenging bilateral and fronto-orbital defects are less common, yet it is in these challenging cases where correct automatic skull patch reconstruction or implant design can have the largest impact on clinical practice. It is therefore desirable to design a method that will be able to leverage both types of cranioplasty data.

The main contributions of this article are the following:

● A multi-branch CNN architecture is proposed as an extension to the cascaded CNN used for skull reconstruction in our previous work.

The architecture allows for training on both synthetic and clinical data samples.
● The proposed CNN model is evaluated on a large dataset of real defective skulls with expert-designed implants for the first time. The positive effect of the proposed method on reconstruction performance is demonstrated.
● A novel metric based on Gaussian curvature is implemented to quantify surface imperfections along the defect border.

2. Materials and methods

2.1. Datasets

We use two different cranioplasty datasets in this work. The Skull-break dataset [18] is a synthetic skull shape reconstruction dataset adapted from the CQ500 public database of head CT scans [17]. The CT scans were rigidly aligned and segmented to provide normalized shapes of healthy skulls. Then, synthetic defects were created by subtracting random shapes from several regions in each skull. Morphological operations were additionally used to mimic some degree of bone healing processes along the defect borders. The dataset contains 570 training and 100 testing pairs of defective skulls and corresponding skull patches.

The second in-house dataset was provided by TESCAN Medical company. It contains a total of 387 real patient cases indicated for cranioplasty. Each patient case consists of CT data with manual skull segmentation and a mesh model corresponding to an expert-designed cranial implant. 75 of these cases additionally contain expert-designed mesh models of patches covering the full area of the defects that were used as an initial template for the final implant design by an expert. Although these expert-designed patches have a different thickness from the original bone, their outer surface can be used as a reference for the outer surface of automatically reconstructed patches. This naturally led us to split the in-house dataset correspondingly into 312 training cases and 75 test cases, ensuring that a real clinical test set of reasonable size is available for the evaluation of both the skull patch shape estimation and the final implant shape estimation tasks. All implant and patch mesh
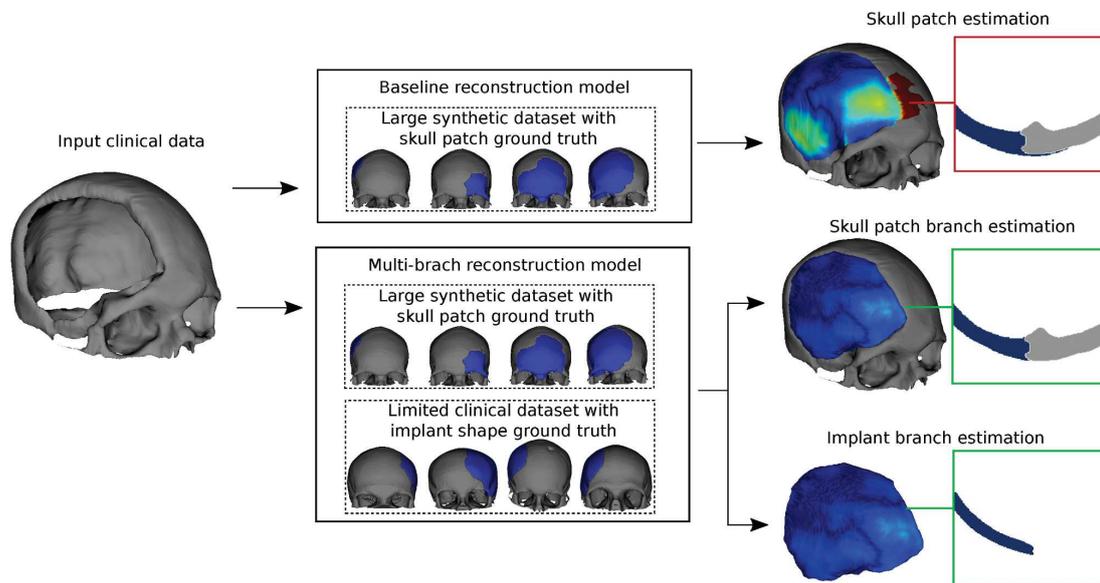

Fig. 1. The proposed multi-branch architecture makes use of multi-task learning on different skull reconstruction datasets. In addition to the higher overall accuracy and ability to directly output cranial implant shapes, the skull patch output of the multi-branch model also better fits the shape to complex defect borders in real clinical data.
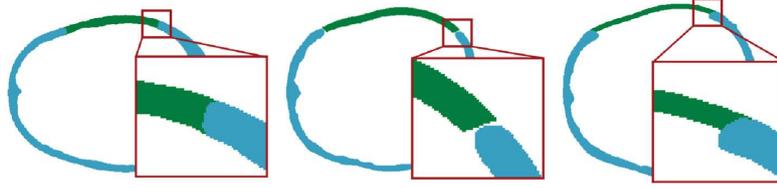
2

Fig. 2. Axial slices through samples from the datasets used in this work. From left to right: skull patch sample from a synthetic dataset, manually designed implant shape sample from an in-house clinical dataset, manually designed skull patch surface sample from an in-house clinical dataset.

models in the clinical in-house dataset were rasterized into voxel grids and the data were rigidly aligned to conform with the Skullbreak data. Several examples from all datasets can be seen in Fig. 2.

The two datasets also differ in several more aspects. Because they come from geographically distant sources, the average size and the anatomical variability of the skulls differ [19,20]. The scale and positional variability of the defects are also different. While the Skullbreak dataset was created specifically to contain a balanced amount of unilateral, bilateral, and fronto-orbital defects, the clinical in-house dataset contains a higher amount of unilateral defects with a larger size and reaching farther into lower parts of temporal and sphenoid bones. Although some of these differences could be addressed by tailoring the synthetic defects in the Skullbreak dataset to fit the distribution of clinical data more closely, some aspects such as skull shape variation and defect border complexity cannot be precisely emulated.

2.2. Baseline CNN models for shape estimation

We use the same baseline reconstruction method for both the skull patch estimation and the implant estimation tasks, with the only difference being the data used for training. The method is based on a cascade of two U-net-like volumetric CNNs proposed in our previous work [4]. The first, coarse CNN $g(\cdot)$ with weights θ_g takes a binary shape of defective skull in coarse resolution, denoted x_{coarse}, and produces an initial output shape estimate with the same resolution y_{coarse}:

$$y_{coarse} = g(x_{coarse}; \theta_g) \tag{1}$$

The second, high-resolution CNN $f(\cdot)$ with weights θ_f then takes a single crop of the upscaled coarse shape estimate y_{coarse} and a corresponding crop of the high-resolution defective skull $x_{high-res}$ and produces a high-resolution shape estimate $y_{high-res}$ of that crop, effectively performing super-resolution of the coarse shape estimate locally conditioned on the high-resolution defective skull:

$$y_{high-res} = f(y_{coarse}, x_{high-res}; \theta_f) \tag{2}$$

The coarse CNN model additionally uses a mirrored copy of the input volume, which has been shown to improve lateral symmetry of output shapes [4].

We use 12 initial feature channels and an input volume size of $128 \times 128 \times 128$ for both the coarse and high-resolution CNNs. The final output is created by first inferring the coarse shape estimate and then inferring the high-resolution CNN in a sliding window manner. Both the original input and the final high-resolution output volumes have size $512 \times 512 \times 512$ voxels with a resolution of 0.4 mm per voxel. We train the CNN cascade for 300 000 steps on mini-batches of size 4 using the soft Dice loss [21]. Each training step consists of updating the weights θ_g using the loss computed on coarse resolution and then updating both weights θ_g and θ_f using the loss computed on random high-resolution crops. More details about the CNN architecture and training procedure can be found in the original work [4].

2.3. Multi-branch CNN model for joint shapes estimation

To facilitate training of the CNN cascade using both the synthetic

skull patch dataset and the clinical implant dataset simultaneously, we split the outputs of the model into a separate skull patch estimation branch and implant estimation branch at both coarse and high resolution. The shape estimation branches are formed by a single conv-ReLU-conv-softmax block with the convolutional layers having the same number of features as the last layer of the U-net backbone. We denote the weights of the U-net backbone θ^B, the weights of the skull patch estimation branch θ^{SP}, and the weights of the implant estimation branch θ^I. The outputs of the coarse CNN in the multi-branch model are given as

$$y_{coarse}^{SP} = g(x_{coarse}; \theta_g^B, \theta_g^{SP}) \tag{3}$$

$$y_{coarse}^I = g(x_{coarse}; \theta_g^B, \theta_g^I), \tag{4}$$

where y_{coarse}^{SP} denotes the coarse shape estimate of the skull patch and y_{coarse}^I denotes the coarse shape estimate of the implant. These two coarse shape estimates are then both used as an input into the high-resolution CNN, along with the high-resolution shape of the input skull. For the high-resolution CNN, the outputs are given as

$$y_{high-res}^{SP} = f(x_{high-res}, y_{coarse}^{SP}, y_{coarse}^I; \theta_f^B, \theta_f^{SP}) \tag{5}$$

$$y_{high-res}^I = f(x_{high-res}, y_{coarse}^{SP}, y_{coarse}^I; \theta_f^B, \theta_f^I) \tag{6}$$

Such architecture ensures that although two slightly different types of shape outputs can be produced by the model at both resolutions, the shared U-net backbone is forced to learn to extract meaningful local features that are suitable for correct shape estimation on both datasets.

During the training of the multi-branch CNN cascade, we use mixed mini-batches containing two samples from the Skullbreak dataset and two samples from the in-house dataset. Accordingly, two loss components are computed at each step: one for the skull patch estimation branch output y^{SP} using the Skullbreak samples, and one for the implant estimation branch output y^I using the in-house dataset samples. These loss components are then added together before updating the respective CNN weights. The iterative training of coarse and high-resolution model weights is otherwise the same as in the baseline method described in Section 2 and the multi-branch model overview is shown in Fig. 3.

2.4. Metrics

For the sake of the quantitative evaluation, we assume that the expert-designed shapes in the test set represent the only correct solution to the shape estimation tasks. This means that the quality of the output can be quantified using segmentation metrics such as volumetric overlaps (i.e. Dice coefficient) and surface distance [22]. However, it should be noted that the shape reconstruction task is specific in allowing some variability in the reconstructed shape in some cases, as long as there are no imperfections along the fit of the reconstructed shape to the input shape. See Appendix A for an illustration of how different segmentation metrics correlate with a subjective quality score of an expert implant designer. For these reasons, we evaluate the automatic reconstruction outputs using multiple different metrics in this work.

In the case of implant shape evaluation, we use the Dice coefficient and average surface distance for quantification of the estimated implant
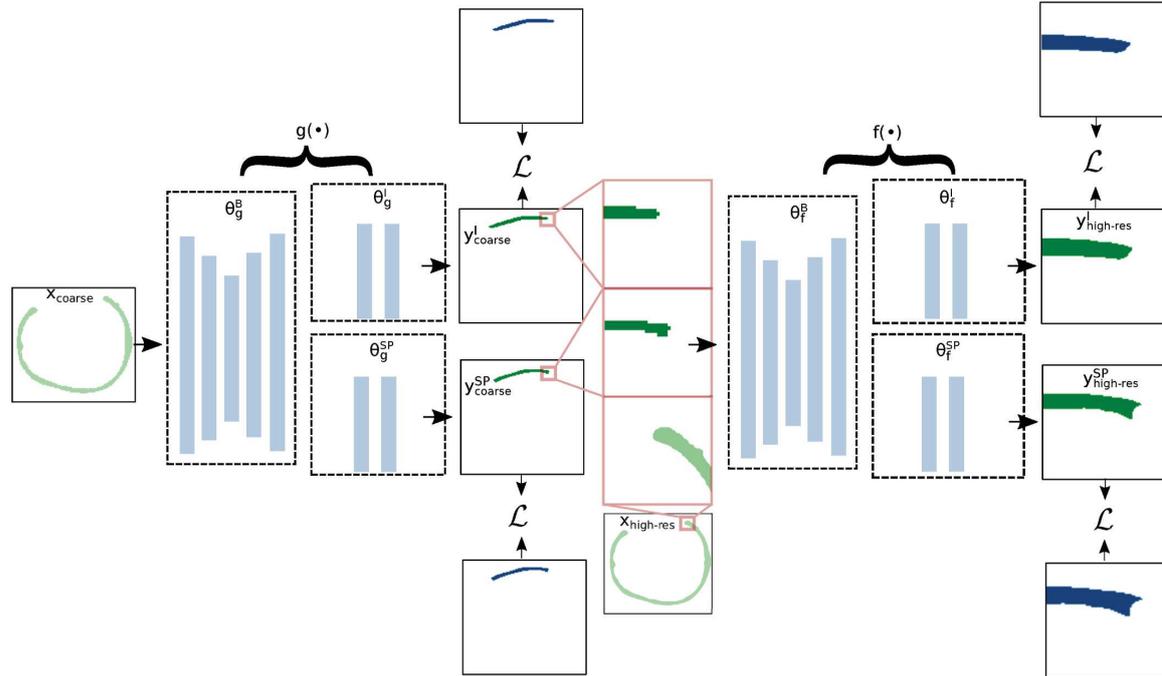
Fig. 3. Illustration of the multi-branch CNN cascade training process. Inputs and outputs of the network in light and dark green colors, respectively, and ground-truth shapes in blue. In each training step, the coarse network weights are first updated using the sum of the coarse losses, and then both coarse and high-resolution network weights are updated using the sum of the high-resolution losses.

shape quality, similarly to recent relevant works [12,23,6]. In the case of skull patch shape evaluation, however, the expert-designed ground truth patches and model outputs have different characteristics and this prevents us from using these metrics directly (see Fig. 2). Because the thickness of the ground-truth patch is different from the thickness of the original bone in the Skullbreak dataset, we measure average surface error only at the outer surface voxels of the skull.

We pay special attention to the quality of fit along the defect border of the skull patches. Similar to other authors [12], we report the outer surface distance computed along the defect border. However, this metric may not precisely convey some types of common errors of skull reconstruction which have an impact on the aesthetic outcome of cranioplasty, such as slight trenches or bumps on the surface along the defect border. To this end, we compare approximate Gaussian curvatures of reconstructed skulls and reference skulls along the defect border to supply this information.

Gaussian curvature is routinely used in 3D model surface analysis literature [24]. For simplification, we chose to approximate the Gaussian curvature error of the reconstructed skull shapes by first smoothing the binary images of skull shapes with a Gaussian blur with $\sigma = 5$, then normalizing back to a range between 0 and 1 and computing the Gaussian curvature K_i at each voxel i using the following equation:

$$K_i = -\frac{\left|\left| \begin{array}{cc} H(F_i) & \nabla F_i^T \\ \nabla F_i & 0 \end{array} \right|\right|}{|\nabla F_i|^4} \tag{7}$$

where F is the blurred volume containing the skull, ∇F_i is the vector of the first-order spatial differences in voxel i and $H(F_i)$ is the square matrix of the second-order spatial differences in voxel i [25]. The resulting Gaussian curvature volumes are then compared directly by computing voxel-wise squared error and we report the mean of this error computed

along the defect border voxels as

$$MSE_K = \frac{1}{N_B} \sum_{i \in B} (K_i^{ref} - K_i^{pred})^2 \tag{8}$$

where B is the set of outer border voxels of the predicted patch and N_B is their count. Although the exact result of this method is partially dependent on voxel resolution, value σ and on the absolute distance between the reconstructed and the reference skull surfaces, it eliminates the need for finding exact vertex correspondences and our experiments show that high resulting values correspond to dented or uneven parts of the surfaces.

3. Results

The baseline implant model was trained using the 312 training implant shapes from the in-house dataset and the baseline skull patch model was trained using the 570 Skullbreak training data samples. Because we noticed that the average size of the Skullbreak skulls differs from the average size of the in-house test skulls, we trained another baseline skull patch model on a modified version of the Skullbreak dataset that was rescaled to match the average height, length, and breadth of the in-house skulls. The multi-branch model was trained using a combination of the in-house and the rescaled Skullbreak dataset. Outputs of all models were morphologically denoised by removing smaller connected components and shape artifacts [26] before comparing them to the reference expert-designed shapes in the in-house test set. All models were implemented in Python programming language using the PyTorch[1] framework and the results were rendered using the

---

[1] https://pytorch.org.

Visualization Toolkit.[2]

3.1. Implant shape estimation performance

The implant shapes produced by the baseline implant model reached an average Dice coefficient of 0.85 ± 0.10 and average surface error of 0.77 ± 0.44 mm, confirming that it is possible to learn the direct mapping of defective skull shapes to the final cranial implant shapes using the CNN cascade. However, because central and fronto-orbital defects are not well represented in the in-house training dataset, the baseline implant model fails to correctly estimate implant shapes in these cases, as shown in Fig. 4. This issue may be amplified by the fact that the coarse CNN model learns to rely too much on the mirrored input to provide initial information about the missing shape, leading to overfitting and an inability to correctly deal with bilateral defects.

The implant estimates of the multi-branch model reached an average Dice coefficient of 0.88 ± 0.07 and an average surface error of 0.65 ± 0.33 mm, showing an increase in accuracy and decreased variability of output shape quality. Closer inspection of the outputs reveals an increased success rate of bilateral and fronto-orbital implant shape estimation. This can be attributed to better generalization of the U-net backbone which needs to account for more diverse defect positions in the Skullbreak dataset. Several example implant shape estimates from both the baseline implant model and the multi-head model implant estimation branch are shown in Fig. 4. The distribution of Dice coefficients and average surface distances achieved by both models can be found in Fig. 7 (top).

3.2. Skull patch estimation performance

The skull patches produced by the baseline skull patch model trained on the original Skullbreak data resulted in an average outer surface error of 0.98 ± 0.45 mm on the in-house test set. Rescaling the Skullbreak training skulls to match the average size of the in-house skulls decreased the error by 15% to 0.83 ± 0.38 mm, supporting the hypotheses that the model learns the average skull shape of the training data. However, the skull patch estimates still produced shapes with a high surface error and occasional artifacts such as holes and uneven surfaces, especially in cases of large defects. One of the causes may be the fact that the defects in the Skullbreak dataset do not fully cover the lower areas of the skull. This could be addressed by extending the dataset with additional synthetic defects, but Figs. 5 and 6 show that there are multiple different sources of error.

The skull patch estimates produced by the multi-branch CNN model further decreased the average surface error to 0.67 ± 0.37 mm. In addition to a lower amount of visible holes and artifacts in the estimated shapes, the multi-branch model also predicted the skull patches with an overall lower outer surface distance from the reference expert-designed patches, as shown in Fig. 5. The distributions of all error metrics for the three models are shown in Fig. 7 (bottom).

Interestingly, the multi-branch model output also reached a lower defect border surface error of 0.75 mm, compared to 0.96 mm and 0.94 mm for the baseline models trained on the original and the rescaled Skullbreak dataset, respectively. Similarly, the Gaussian curvature errors of the baseline skull patch model trained on the original Skullbreak and on the rescaled Skullbreak datasets also did not differ significantly, but the curvature error decreased by around 12% in the case of the multi-branch model skull patch branch outputs. This suggests that the multi-branch model learned to better fit the reconstructed skull patches to the more complex borders of the in-house defective skulls, despite only encountering the corresponding implant shapes with spatial tolerance along the border during training (see Fig. 2). Fig. 6 shows how the Gaussian curvature error reacts to different types of surface errors

compared to the distance-based metrics, helping to visually identify problematic regions of the skull patch shape reconstruction outputs.

3.3. Statistical analysis

We performed a statistical analysis to report the significance of the performance gain achieved by the multi-branch CNN. The statistical significance levels are shown in Fig. 7.

A one-sided paired t-test was used to test the hypothesis that the error measurements of the multi-branch CNN outputs were significantly lower (or higher in the case of Dice coefficient) than in the case of the baseline CNN outputs. For the global metrics (i. e. surface distance and Dice coefficient), p-value was below the level of 0.05 for both the reconstructed skull patches and implants, which led us to accept the hypothesis that combining the data using the multi-branch CNN provides better global results when compared to the baseline models which use only one type of training data.

In the case of the border error metrics (i. e. border distance and Gaussian curvature error), the hypotheses could not be accepted using t-test as the p-values were over 0.05. This is likely because the shape artifacts along the border were often concentrated into a relatively small area (see Fig. 6), which resulted in a smaller quantitative difference. Therefore, we used a non-parametric Wilcox sign test to test whether the proposed approach lowers the border error when compared to the baseline methods. The hypothesis was accepted, showing that albeit small, the border error reduction is consistent across the test cases.

4. Discussion and conclusions

CNN-based skull reconstruction methods are becoming a hot topic in medical imaging. One of the major drawbacks in the current research is that the reconstruction outputs are most often evaluated on a held-out synthetic dataset in which similar anatomical variability and defect shape and type distribution can be ensured. One of the goals of this study was to illustrate the behavior of CNN-based skull reconstruction models trained on an easily accessible synthetic dataset when evaluated on real patient data. Our experiments showed that the transfer of the trained CNN model to a different population can negatively affect the reconstruction quality. Furthermore, by looking at differences in Gaussian curvature, we found that the shape complexity of the defect border in real clinical data can cause faults in the smoothness of the resulting surface.

We showed that when training the model on real clinical patient data, synthetic data can be effectively leveraged using the proposed multi-branch CNN model to significantly improve the model performance and compensate for common issues of clinical patient datasets (i. e. data scarcity and imbalance). Although a similar effect could possibly be achieved by collecting a vast amount of well-balanced clinical data, or by perfectly matching their distribution by meticulously tailoring synthetic data, we believe that the proposed approach of combining a large amount of imperfect synthetic data and a limited set of target clinical data is generally simpler and easily extendable to different types of cranioplasty data, for example, different population, additional defect areas such as the orbital floor or zygomatic bone or even different preferences for the final implant shape. The error of the outer surface of reconstructed skulls achieved by the proposed method is higher than some other recent works evaluated on synthetic defects [12,4]. However, we believe that factors such as a higher average area of the defects in our test set may be the cause and that the results are overall very promising.

The synthetic and clinical datasets used in this work contained different types of ground truths: the original missing skull patch shape and final cranial implant shapes. This allowed us to automatically produce 3D printable and directly implantable shapes, although this use case will require further evaluation of the clinical applicability in cooperation with experienced implant designers. More importantly, the
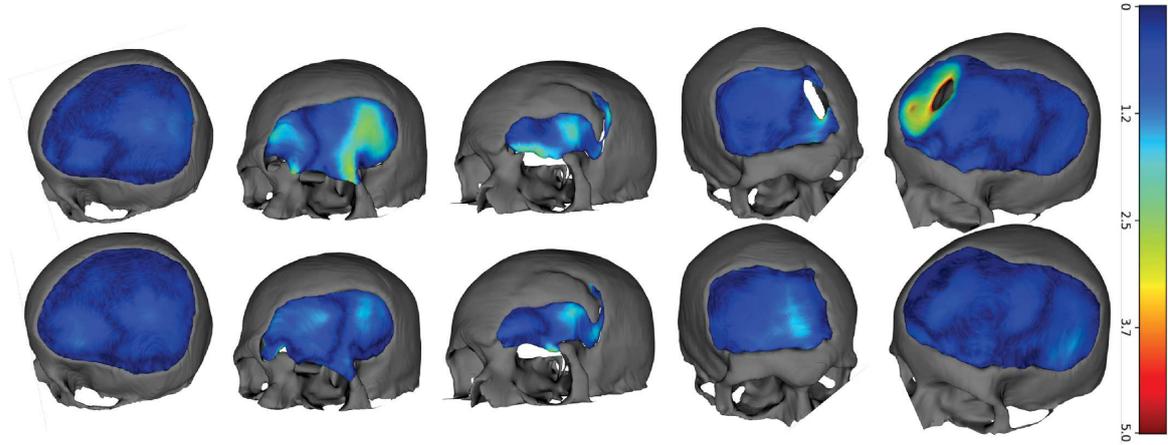
---

[2] https://vtk.org.

Fig. 4. Implant estimates of the baseline implant model (top) and the multi-branch model (bottom).
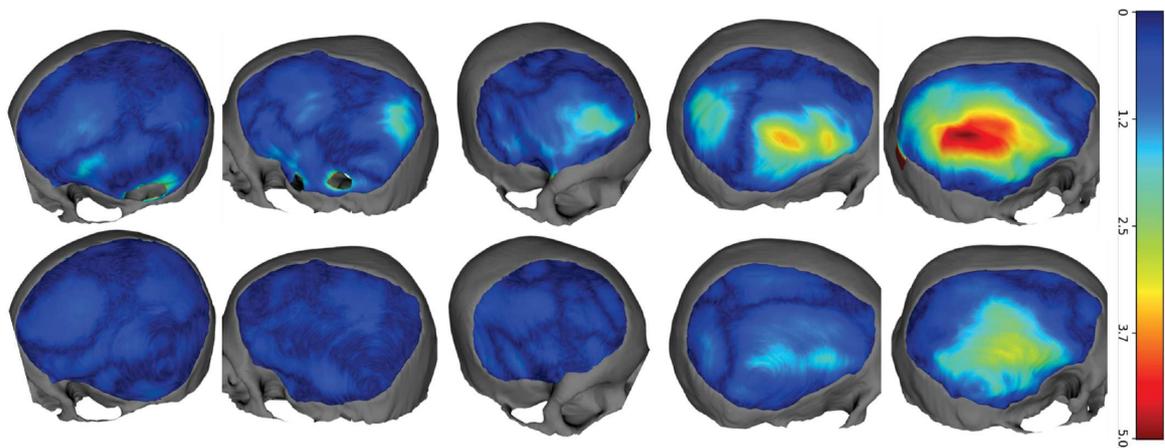
Fig. 5. Estimated skull patches of the baseline skull patch model (top) and the multi-branch model (bottom).
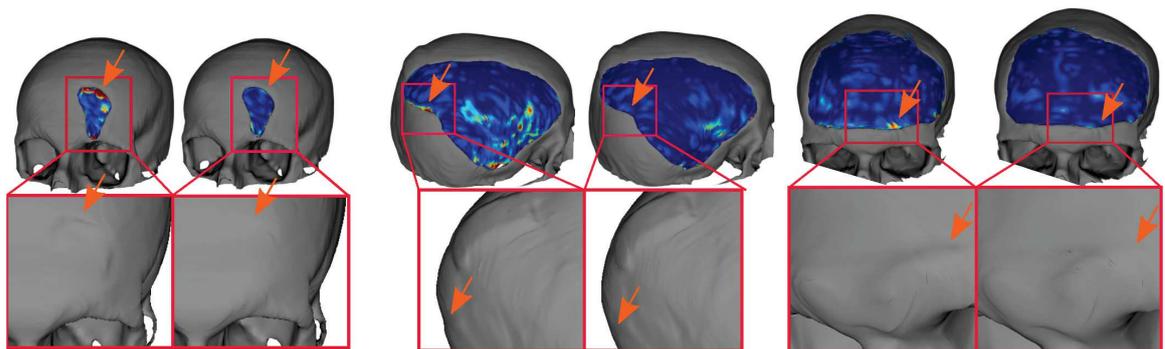
Fig. 6. Three example pairs of baseline skull patch model outputs and multi-branch model skull patch outputs, respectively, with color-coded Gaussian curvature error. The 3D models were rendered using the marching cubes algorithm and post-processed using quadratic decimation and normal smoothing. The multi-branch model can produce smoother results with lower curvature error. Note that we show the entire Gaussian curvature error maps for illustration while only defect border voxels are taken into account when computing the mean errors.

6

Fig. 7. Boxplots showing the error distributions of the evaluated models trained on corresponding datasets (Skullbreak - SB, in-house - IH). Average surface error and Dice coefficient of the implant estimates (top) and average outer surface error, border error, and Gaussian curvature error of the skull patch estimates (bottom). p_t and p_w values denote statistical significance of one-sided paired t-test and one-sided Wilcox test, respectively.

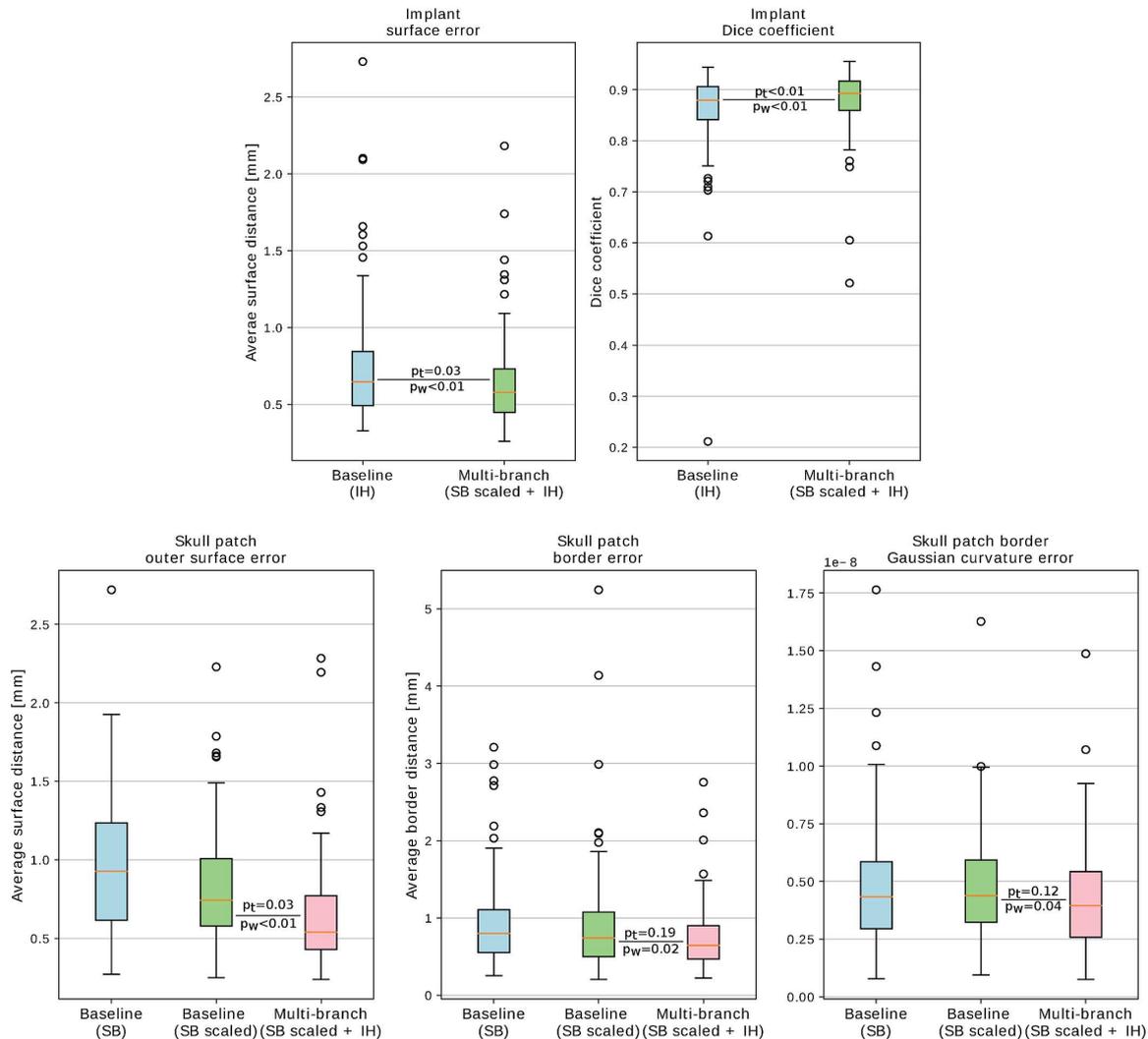general ability of the model to combine cranioplasty data from different sources and of different types can accelerate the adoption of the automatic reconstruction methods by allowing training on specific target datasets while exploiting the advantages of available synthetic datasets.

To our best knowledge, this was the first study that evaluated CNN-based skull reconstruction on a real clinical dataset of this size. The proposed multi-branch CNN cascade increased the reconstructed shape quality by allowing training on more data when compared to the individual baseline models. Although the results of this study are promising from a quantitative perspective, they will need to be evaluated next by an experienced implant designer to ascertain their clinical value.

Funding

This work was partially funded by TESCAN 3DIM company.

Ethics approval

All studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed consent

All data were evaluated retrospectively and processed in compliance with General Data Protection Rule (GDPR). Formal consent is not required for this type of study.

Declaration of competing interest

Michal Španěl is the CEO of TESCAN 3DIM company at the time of

8 Original Papers

writing this article.

Oldřich Kodym has no conflict of interest to declare.

Adam Herout has no conflict of interest to declare.

Appendix A. Correlation Analysis of Quantitative Metrics and Subjective Expert Score of Automatic Skull Reconstructions

This section illustrates how well the quantitative segmentation metrics can predict the usability of automatic skull reconstruction results in clinical practice. We created a dataset of automatically reconstructed defective skulls and submitted it to an expert with experience in the field of skull reconstruction and implant design for subjective quality evaluation. Comparing these subjective expert scores with metrics of similarity between the reconstructed and the original shape can give an idea of what to look for when evaluating the reconstructions.

A.1 Skull Data and Reconstruction

The skull data come from the SkullBreak and SkullFix datasets [18], so the ground truth original shapes are available. A CNN-based reconstruction of the missing shape [4] was performed on each skull. Because we would ideally want to cover for this analysis the whole quality spectrum from bad reconstructions to very good reconstructions, we included the following types of reconstructed cases:

- SkullFix test case reconstructions
- SkullFix additional test case reconstructions
- SkullBreak test case reconstructions
- SkullBreak training case reconstructions (to include several close-to-perfect reconstructions)
- SkullBreak test case reconstructions using generative model [4] (to include multiple different reconstructions for a single case, including visibly bad ones)

This resulted in a total of 35 skulls. The expert assigned a score on a scale from zero to ten to each of the reconstructions, where zero corresponded to unacceptable reconstruction and ten to a nearly perfect result.

A.2 Global Metrics

We first computed correlation coefficients between the subjective expert score and routinely used segmentation metrics, including volumetric Dice coefficient and average symmetric surface distance. We also included the surface distance computed at the outer surface of the skull, since it is the most important aspect for subsequent implant modeling steps [27]. The outer surface was used in the evaluation of some previous works [12] and also in this work because of the shape characteristics of in-house ground truth data.
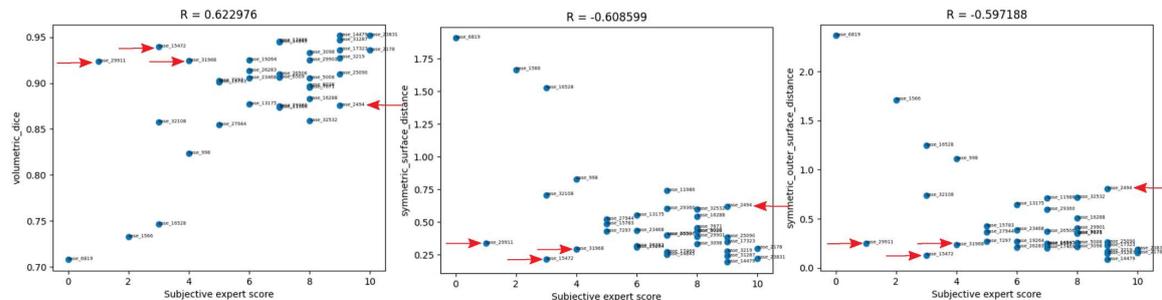
Fig. 8. Plots of the three global quantitative metrics plotted against the corresponding expert subjective score. Note that in some cases (highlighted by red arrows), the metrics failed to estimate the practical usability of the reconstruction result.

Figure 8 shows that these global metrics correlate with the expert subjective score with correlation coefficients around 0.6, confirming that they are appropriate for the comparison of different reconstruction methods. However, it can be noted that their correlation is weak when the subjective expert score is high, making it impossible to use them for discrimination between good and perfect results. Also, several cases satisfy the quantitative metrics while being seen as low-quality by experts and vice versa (see cases highlighted in red in Fig. 8).

A.3 Defect Border Metrics

The smoothness of the surface closest to the defect border has a significant impact on the aesthetic outcome of cranioplasty. We study two metrics that focus on this area: outer surface distance of the defect border and mean square error of Gaussian curvature. The defect border is defined as a set of outer surface voxels of the reconstructed skull patch shape in direct contact with the defective skull.

8

86

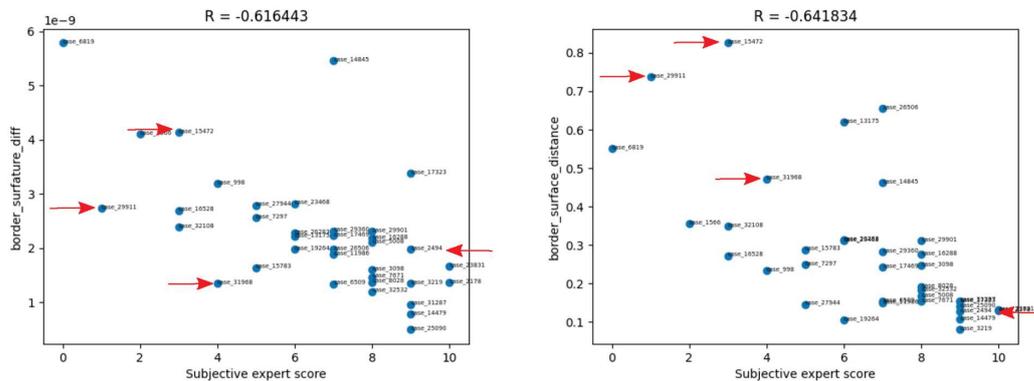**Fig. 9.** Plots of border Gaussian curvature error (left) and border surface distance (right) plotted against the corresponding expert subjective score. The same cases are highlighted as in the case of global metrics, showing that the border metrics convey different yet relevant information about the reconstruction result.

Figure 9 shows that both of these metrics correlate with the subjective expert score similarly or slightly more than the global metrics. Most importantly, it can be seen that the border metrics indeed convey different information. Although the quantitative border metrics do not always agree with the subjective quality score, the correlation with the expert score was higher in the cases where the correlation of the global metrics was low.

This study was performed using only one type of automatic reconstruction method and the results were evaluated by a single implant design expert, which leaves much room for more extensive studies. However, it can be concluded that to best gauge the quality of results of automatic skull reconstruction, different types of quantitative metrics should be combined together, and both global and border metrics should be taken into account.

References

[1] M.S. Gilardino, M. Karunanayake, T. Al-Humsi, A. Izadpanah, H. Al-Ajmi, J. Marcoux, J. Atkinson, J.P. Farmer, A comparison and cost analysis of cranioplasty techniques, J. Craniofac. Surg. 26 (1) (2015) 113–117, https://doi.org/10.1097/scs.0000000000001305.

[2] E.B. da Silva Júnior, A.H. de Aragão, M. de Paula Loureiro, C.S. Lobo, A.F. Oliveti, R.M. de Oliveira, R. Ramina, Cranioplasty with three-dimensional customised mould for polymethylmethacrylate implant: a series of 16 consecutive patients with cost-effectiveness consideration, 3D Printing in Medicine 7 (1) (2021), https://doi.org/10.1186/s41205-021-00096-7.

[3] C.H. Cheng, H.Y. Chuang, H.L. Lin, C.L. Liu, C.H. Yao, Surgical results of cranioplasty using three-dimensional printing technology, Clin. Neurol. Neurosurg. 168 (2018) 118–123, https://doi.org/10.1016/j.clineuro.2018.03.004.

[4] O. Kodym, M. Španěl, A. Herout, Skull shape reconstruction using cascaded convolutional networks, Comput. Biol. Med. 123 (2020) 103886, https://doi.org/10.1016/j.compbiomed.2020.103886.

[5] J. Li, J. Egger, Towards the automatization of cranial implant design in cranioplasty: first challenge, AutoImplant 2020, held in conjunction with MICCAI 2020, Lima, Peru, october 8, 2020, proceedings, Lect. Notes Comput. Sci. (2021), https://doi.org/10.1007/978-3-030-64327-0. Springer International Publishing.

[6] F. Matzkin, V. Newcombe, S. Stevenson, A. Khetani, T. Newman, R. Digby, A. Stevens, B. Glocker, E. Ferrante, Self-supervised skull reconstruction in brain CT images with decompressive craniectomy, in: Medical Image Computing and Computer Assisted Intervention – MICCAI 2020, 2020, pp. 390–399, https://doi.org/10.1007/978-3-030-59713-9_38.

[7] X. Chen, L. Xu, X. Li, J. Egger, Computer-aided implant design for the restoration of cranial defects, Sci. Rep. 7 (1) (2017), https://doi.org/10.1038/s41598-017-04454-6.

[8] Y.W. Chen, C.T. Shih, C.Y. Cheng, Y.C. Lin, The development of skull prosthesis through active contour model, J. Med. Syst. 41 (10) (2017), https://doi.org/10.1007/s10916-017-0808-2.

[9] Y. Volpe, R. Furferi, L. Governi, F. Uccheddu, M. Carfagni, F. Mussa, M. Scagnet, L. Genitori, Surgery of complex craniofacial defects: a single-step AM-based methodology, Comput. Methods Progr. Biomed. 165 (2018) 225–233, https://doi.org/10.1016/j.cmpb.2018.09.002.

[10] A. Marzola, L. Governi, L. Genitori, F. Mussa, Y. Volpe, R. Furferi, A semi-automatic hybrid approach for defective skulls reconstruction, Computer-Aided Design and Applications 17 (1) (2019) 190–204, https://doi.org/10.14733/cadaps.2020.190-204.

[11] A. Marzola, M. Servi, Y. Volpe, A reliable procedure for the construction of a statistical shape model of the cranial vault, in: Lecture Notes in Mechanical Engineering, Springer International Publishing, 2019, pp. 788–800, https://doi.org/10.1007/978-3-030-31154-4_67.

[12] M.A. Fuessinger, S. Schwarz, C.P. Cornelius, M.C. Metzger, E. Ellis, F. Probst, W. Semper-Hogg, M. Gass, S. Schlager, Planning of skull reconstruction based on a statistical shape model combined with geometric morphometrics, Int. J. Comput. Assist. Radiol.Surg. 13 (4) (2017) 519–529, https://doi.org/10.1007/s11548-017-1674-6.

[13] M.A. Fuessinger, S. Schwarz, J. Neubauer, C.P. Cornelius, M. Gass, P. Poxleitner, R. Zimmerer, M.C. Metzger, S. Schlager, Virtual reconstruction of bilateral midfacial defects by using statistical shape modeling, J. Cranio-Maxillofacial Surg. 47 (7) (2019) 1054–1059, https://doi.org/10.1016/j.jcms.2019.03.027.

[14] W. Semper-Hogg, M.A. Fuessinger, S. Schwarz, E. Ellis, C.P. Cornelius, F. Probst, M. C. Metzger, S. Schlager, Virtual reconstruction of midface defects using statistical shape models, J. Cranio-Maxillofacial Surg. 45 (4) (2017) 461–466, https://doi.org/10.1016/j.jcms.2016.12.020.

[15] A. Morais, J. Egger, V. Alves, Automated computer-aided design of cranial implants using a deep volumetric convolutional denoising autoencoder, in: Advances in Intelligent Systems and Computing, Springer International Publishing, 2019, pp. 151–160, https://doi.org/10.1007/978-3-030-16187-3_15. DOI 10.1007/978-3-030-16187-3_15, URL.

[16] G. von Campe, K. Pistracher, Patient specific implants (PSI), in: Towards the Automatization of Cranial Implant Design in Cranioplasty, 2020, pp. 1–9, https://doi.org/10.1007/978-3-030-64327-0_1.

[17] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N.G. Campeau, V.K. Venugopal, V. Mahajan, P. Rao, P. Warier, Development and Validation of Deep Learning Algorithms for Detection of Critical Findings in Head Ct Scans, 2018 arXiv: 1803.05854.

[18] O. Kodym, J. Li, A. Pepe, C. Gsaxner, S. Chilamkurthy, J. Egger, M. Španěl, SkullBreak/SkullFix – dataset for automatic cranial implant design and a benchmark for volumetric shape learning tasks, Data in Brief 35 (2021) 106902, https://doi.org/10.1016/j.dib.2021.106902.

[19] T.T. Le, L.G. Farkas, R.C. Ngim, L.S. Levin, C.R. Forrest, Proportionality in asian and north american caucasian faces using neoclassical facial canons as criteria, Aesthetic Plast. Surg. 26 (1) (2002) 64–69, https://doi.org/10.1007/s00266-001-0033-7. DOI 10.1007/s00266-001-0033-7, URL.

[20] P. Raghavan, D. Bulbeck, G. Pathmanathan, S.K. Rathee, Indian craniometric variability and affinities, Int. J. Evol. Biol. (2013) 1–25, https://doi.org/10.1155/2013/836738, 2013.

[21] F. Milletari, N. Navab, S.A. Ahmadi, V-net: fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, 2016, https://doi.org/10.1109/3dv.2016.79.

[22] V. Yeghiazaryan, I. Voiculescu, Family of boundary overlap metrics for the evaluation of medical image segmentation, J. Med. Imag. 5 (1) (2018) 1, https://doi.org/10.1117/1.jmi.5.1.015006.

[23] J. Li, A. Pepe, C. Gsaxner, G. von Campe, J. Egger, A baseline approach for AutoImplant: the MICCAI 2020 cranial implant design challenge, in: Tanveer Syeda-Mahmood (Ed.), Multimodal Learning for Clinical Decision Support and Clinical Image-Based Procedures, Springer, Cham, 2020, pp. 75–84. https://doi.org/10.1007/978-3-030-60946-7_8.

[24] H. Yamauchi, S. Gumhold, R. Zayer, H.P. Seidel, Mesh segmentation driven by Gaussian curvature, Vis. Comput. 21 (8–10) (2005) 659–668, https://doi.org/10.1007/s00371-005-0319-x.

9

8 Original Papers

[25] R. Goldman, Curvature formulas for implicit curves and surfaces, Comput. Aided Geomet. Des. 22 (7) (2005) 632–658, https://doi.org/10.1016/j. cagd.2005.06.005. DOI 10.1016/j.cagd. 2005.06.005, URL.

[26] O. Kodym, M. Španěl, A. Herout, Cranial defect reconstruction using cascaded CNN with alignment, in: Towards the Automatization of Cranial Implant Design in Cranioplasty, 2020, pp. 56–64, https://doi.org/10.1007/978-3-030-64327-0_7.

[27] F.M.M. Marreiros, Y. Heuzé, M. Verius, C. Unterhofer, W. Freysinger, W. Recheis, Custom implant design for large cranial defects, Int. J. Comput.Assist. Radiol.Surg. 11 (12) (2016) 2217–2230, https://doi.org/10.1007/s11548-016-1454-8.

88

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed
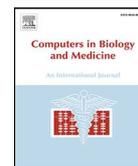
Skull shape reconstruction using cascaded convolutional networks

Oldřich Kodym *, Michal Španěl, Adam Herout

Department of Computer Graphics and Multimedia, Brno University of Technology, Božetěchova 2, 612 66 Brno, Czech Republic

ARTICLE INFO

ABSTRACT

Designing a cranial implant to restore the protective and aesthetic function of the patient's skull is a challenging process that requires a substantial amount of manual work, even for an experienced clinician. While computer-assisted approaches with various levels of required user interaction exist to aid this process, they are usually only validated on either a single type of simple synthetic defect or a very limited sample of real defects. The work presented in this paper aims to address two challenges: (i) design a fully automatic 3D shape reconstruction method that can address diverse shapes of real skull defects in various stages of healing and (ii) to provide an open dataset for optimization and validation of anatomical reconstruction methods on a set of synthetically broken skull shapes.

We propose an application of the multi-scale cascade architecture of convolutional neural networks to the reconstruction task. Such an architecture is able to tackle the issue of trade-off between the output resolution and the receptive field of the model imposed by GPU memory limitations. Furthermore, we experiment with both generative and discriminative models and study their behavior during the task of anatomical reconstruction.

The proposed method achieves an average surface error of $0.59\,\mathrm{mm}$ for our synthetic test dataset with as low as $0.48\,\mathrm{mm}$ for unilateral defects of parietal and temporal bone, matching state-of-the-art performance while being completely automatic. We also show that the model trained on our synthetic dataset is able to reconstruct real patient defects.

1. Introduction

Patient-specific implants (PSIs) are often used for the treatment of cranio-facial defects. Especially in cases of larger defects caused by trauma, tumour resection or decompressive craniectomy, it is usually required to reconstruct the original skull shape for aesthetic purposes and protection of intracranial structures against mechanical impact [1,2].

Current state-of-the-art methods usually comprise of using a patient CT scan to design the implant pre-operatively and then 3D printing of the result using bio-compatible materials such as titanium, porous polyethylene or polyether ether ketone [3,4]. Alternatively, implants can be cast in a 3D-printed mold from bone cement which can be loaded with antibiotics to decrease the risk of infection [5]. Such approaches lead to a reduction of operative time and improved patient results [6].

Provided that precise enough tissue segmentation is obtained from the CT data, the process of computer-aided design (CAD) of PSIs remains the most important step that affects the final quality and reproducibility of PSIs [7]. This presents a challenging and tedious task for the clinician or engineer designing the PSI. To ensure correct healing and prevent complications, the PSI must fit precisely to the defect border without any steps and with gaps of less than $0.8\,\mathrm{mm}$ between the implant and remaining tissue [8]. Furthermore, smoothness and symmetry of the anatomy should be preserved to ensure a correct aesthetic result [9]. The first step of the implant design is correct reconstruction of the missing shape of the skull anatomy from which the PSI is then derived.

1.1. Related work

A considerable number of CAD systems for skull reconstruction make heavy use of the natural facial symmetry by identifying the best symmetry plane and then mirroring the healthy part of the skull onto the defect area [10]. Recent automation of parts of this process led to an efficient and user-friendly way to provide an aesthetically correct result. Such methods, however, cannot account for bilateral defects reaching into both sides of the skull. Also, because the symmetry is usually not perfect in real cases, manual corrections are often needed to produce a correct reconstruction. Another group of methods is based on surface interpolation under the assumption of a nearly spherical shape
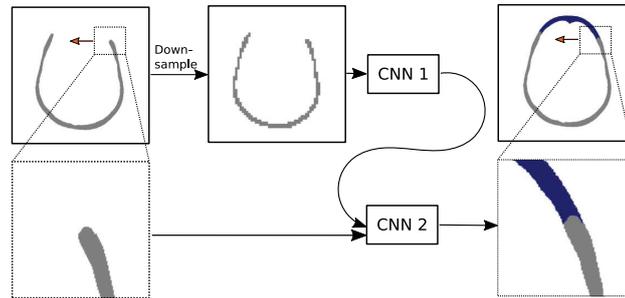

Fig. 1. The proposed skull reconstruction method fully automatically produces a binary shape of the missing part of the skull using sliding-window approach with coarse-resolution middle step.
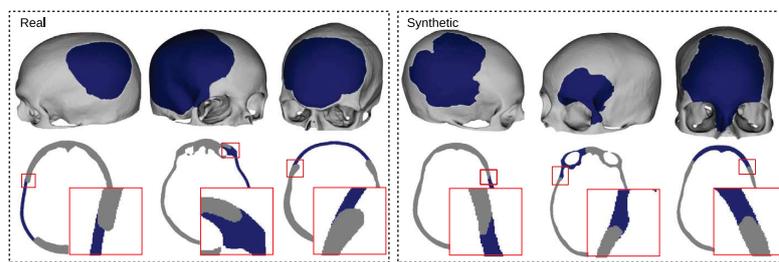

Fig. 2. Examples of 3D models and slices through defective skulls from real patients (left) and synthetically generated defective skulls (right). The real patient samples include defects reconstructed by an experienced clinician.
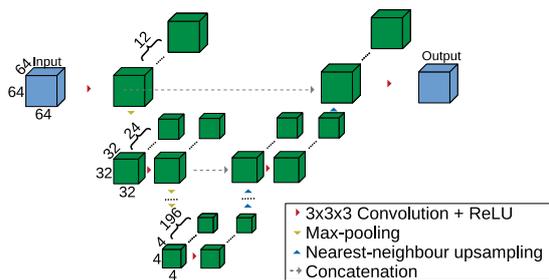

Fig. 3. Overview of the 3D CNN backbone architecture example for an input of size 64^3. Note that for an input of size 128^3, the CNN is deeper and the bottleneck tensor has 384 channels.

of the cranium [11,12]. Interpolation-based methods can guarantee desired continuity on the tissue-implant interface and also make it possible to modify the resulting fit by adjusting the parameters. Although these methods work well on smaller defects, they often struggle with larger defects because of a lack of constraints and they cannot model more complex anatomy shapes such as orbitals.

Current state-of-the-art methods usually exploit some kind of deformable models [13]. Statistical shape models combined with geometric morphometrics have been studied extensively in the context of skull reconstruction [14–16], achieving an average surface error of $0.47\,\mathrm{mm}$ for defects of the parietal and temporal area and $0.75\,\mathrm{mm}$ for small mid-facial defects, as measured on simple synthetically created defects against the original bone shape. These methods, while providing a good reconstruction result, rely on a clean, well-defined defect border, which is rarely the case in real patient cases with complex fractures in various stages of healing and bone resorption, as can be seen in the example slice in Fig. 1.

Another interesting approach has recently made use of convolutional denoising auto-encoders in the first attempt to use a deep learning-based shape completion for skull reconstruction [17], although only operating on a very coarse resolution. More details on the topic of automatic skull shape reconstruction can be found in the work of Buonamici et al. [18].

Using deep learning for a general 3D shape reconstruction (also shape inpainting, shape completion) is a well studied research topic in the literature. The basic approach is to represent the incomplete input shape as a binary voxel grid and train a 3D convolutional neural network (CNN) with a deterministic denoising auto-encoder architecture to output the completed binary shape [19]. These approaches cannot be utilized for bigger volumes due to GPU memory limitations. To tackle this, some authors exploit different data representations such as graphs or point clouds [20,21]. Another group of authors use the 3D CNN only for coarse shape estimation, refining the result in the post-processing step [22–24].

An orthogonal research direction in this area led to substituting the purely discriminative CNN models with generative models such as generative adversarial networks (GANs) or variational auto-encoders [20, 23,24], suggesting that the shape completion task actually has multiple correct solutions conditioned on a single input. This issue of one-to-many mapping has also been raised by authors in the context of anatomical shape reconstruction [25,26] where inter-expert variability of the resulting shape is also taken into account. However, the argument that the variability of the output should be enforced at the cost of precision measured against the original shape is in a conflict with the current literature on skull reconstruction where the original shape is considered to be the ground-truth.

1.2. Contributions

In this work, we design a cascaded CNN architecture for the estimation of a high-resolution 3D anatomy shape conditioned on the input defective skull. Although symmetry is used in the proposed
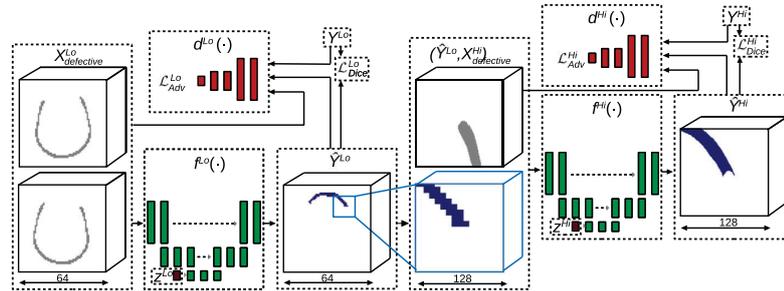
2

Fig. 4. Overview of the proposed 3D CNN cascade. The symmetrized low-resolution input X^{Lo} is fed into the first model f^{Lo} to produce the missing shape estimation $\hat{Y}^{Lo}$. Then, it is concatenated to the high-resolution input X^{Hi} and fed into the second model f^{Hi} to produce the final high-resolution missing shape $\hat{Y}^{Hi}$. Discriminator CNNs $d(\cdot)$ and latent vectors z are only used in the generative model.

method as an additional guiding signal, the method can successfully reconstruct defects reaching into both sides of the skull as well as into more complex anatomical regions such as orbitals. To the authors' best knowledge, this is the first deep learning-based method of 3D shape reconstruction that reaches a high enough resolution to be clinically viable for the skull reconstruction task. To address the issue of multiple possibly correct solutions, we also experiment with a probabilistic generative version of the proposed model. Finally, in order to improve the reproducibility of research in the area of automatic skull reconstruction, we introduce an open dataset of skull shapes with synthetic defects. The dataset mimics the variability in shape, position and bone resorption present among real patients and we show that a CNN model trained on this synthetic dataset also performs well on challenging real patient cases without any further pre-processing.

2. Materials

For training and validation of the proposed method, we used a public CQ500 dataset [27] as a source of head CT patient data. A total of 189 of the scans were rigidly aligned and the skulls were segmented and saved as 3D binary arrays of 512^3 voxels. Finally, 5 different defects were created on each segmented skull with an emphasis on simulating the variability in real defective patient skulls. We simulate the defects by subtracting randomly deformed combinations of spheres followed by morphologically rounding the defect edges to account for various genesis and healing processes of real defects. To allow for structured validation, the defects were categorized into *unilateral parietal*, *unilateral frontal* (the orbital area) and *bilateral* groups. In addition to these three groups, two more random defects were generated on each skull. The resulting dataset of 945 defective skulls with ground-truth original shapes along with further details is publicly available as the SkullBreak dataset[1] and we refer interested readers there for further information regarding the details of defective skulls generation process.

The synthetic dataset was split into 179 training and 10 testing skulls, resulting in 895 training and 50 testing defect shapes in total. To evaluate the ability of our approach to generalize, we also utilized an internal dataset of 9 real defective patients. For these patients, ground-truth skull reconstructions made by a clinician experienced in cranial implant design were available. Several samples from both datasets are shown in Fig. 2.

3. Methods

We formulate the skull reconstruction task as finding the missing part of the anatomy represented by binary volume $Y = X_{healthy} - X_{defective}$. Thus, we look for the function $f(\cdot)$ with parameters θ that maps the defective skull to an estimated shape $\hat{Y} = f_\theta(X_{defective})$ from distribution $P(\hat{Y}|X_{defective})$ of shapes that correctly complete it.

Fig. 5. The overall performance of the discriminative model on different groups of testing defective skulls. Average surface error [mm] for a simple input (green) and symmetrized input (blue).

3.1. Reconstruction model architecture

We use a combination of two CNN models with a 3D U-net [28] backbone to approximate the function f, with parameters θ being the trainable weights of the CNN. The individual models differ from the original 3D U-net in several ways. Instead of up-convolutions, we use nearest-neighbor up-sampling followed by regular convolution in the decoder part of the model, as this has been shown to improve the model training process and performance in some cases [29]. The number of down-sampling and up-sampling layers is such that the bottleneck tensor has spatial dimensions of 4^3 as shown in Fig. 3. This ensures that the output neurons of the CNN have a sufficient receptive field to correctly model the shape of missing anatomy in the case of defects with a large surface area.

Each of the models operates on a different resolution. The first model, denoted f^{Lo}, takes an input volume down-sampled to 64^3 voxels and is trained to output an estimate of missing anatomy $\hat{Y}^{Lo} = f^{Lo}(X_{defective}^{Lo})$ on an equivalent resolution of 3.2 mm per voxel. While this resolution is too low to model anatomy with enough precision, it can provide an initial estimate of the missing shape. The second model, denoted f^{Hi}, then takes a $128 \times 128 \times 128$ crop of the input data at the original high resolution concatenated to an up-sampled output of the first model. This model is trained to output the corresponding patch of the final missing anatomy estimate $\hat{Y}^{Hi} = f^{Hi}(\hat{Y}^{Lo}, X_{defective}^{Hi})$, which can be viewed as a super-resolution of the
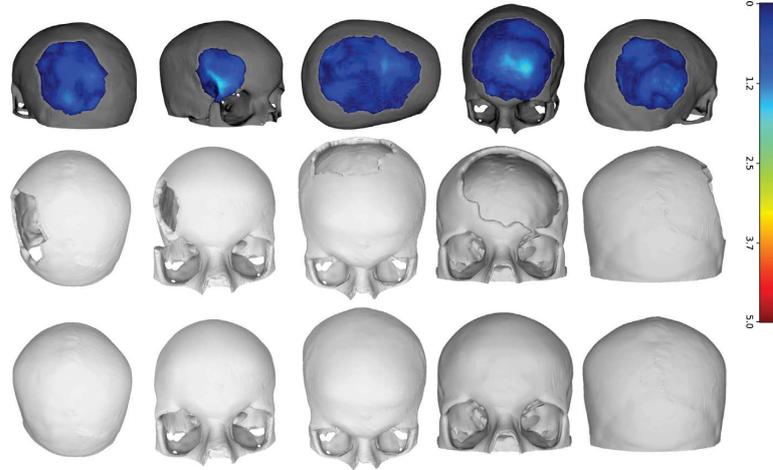
Fig. 6. Examples of results of the discriminative model reconstructions for parietal, frontal, bilateral and random defects, respectively. From top to bottom: Surface error maps, input synthetically broken skulls and reconstructed skull shapes. Note that the majority of the reconstructed surface reaches errors of less than one millimeter in all cases.

initial missing anatomy estimate conditioned on the remaining part of the skull at full resolution.

During inference, the first model provides enough contextual information about the overall shape of the defective skull while the second model can ensure precise contact at the defect border. The final estimate can be inferred by first computing the coarse estimate $\hat{Y}^{Lo}$ and then computing the final estimate $\hat{Y}^{Hi}$ using the sliding window approach, substantially reducing the memory footprint.

Symmetrized input. The chosen U-net architecture in the low-resolution CNN is in fact not well suited for transferring information from one side of the volume to the opposing side as this transfer can only happen in the deeper layers of the model where the shape information is already compressed. However, the ability to preserve the anatomical symmetry is a critical part of the method. To this end, we concatenated a sagitally flipped copy of the volume to the low-resolution CNN input. This makes it easier to propagate the symmetry information using convolutional kernels and skip connections of the U-net architecture. The effect of symmetrizing input is demonstrated in Section 4.

3.2. Optimization

We optimize the CNNs using training batches of size 2, which fully utilize the available GPU. An Adam optimizer is used as it is currently one of the most widely used optimization algorithms suitable for most deep learning applications [30]. Although we train both CNNs with their respective loss functions $\mathcal{L}^{Lo}$ and $\mathcal{L}^{Hi}$, we train the cascade in an end-to-end manner. The training samples for the first model $(Y^{Lo}, X^{Lo}_{defective})$ and the random training crops for the second model $(Y^{Hi}, X^{Hi}_{defective})$ are always sampled from the same skull volume.

Discriminative model. We first assume that the skull reconstruction task has a single correct ground-truth solution given by the original missing anatomy shape Y. This allows us to use a reconstruction loss similar to a segmentation task. We chose the soft Dice loss [31] due to its good performance in dealing with class imbalance. The two losses are defined as

$$\mathcal{L}^{Lo}_{Dice} = Dice(Y^{Lo}, f^{Lo}(X^{Lo}_{defective})), \tag{1}$$

$$\mathcal{L}^{Hi}_{Dice} = Dice\left(Y^{Hi}, f^{Hi}\left(f^{Lo}(X^{Lo}_{defective}), X^{Hi}_{defective}\right)\right), \tag{2}$$

and we optimize them iteratively for 300 000 training steps. While it is possible to optimize the whole cascade using only the $\mathcal{L}^{Hi}$ loss, we found that using the auxiliary loss $\mathcal{L}^{Lo}$ is necessary for correct model behavior.
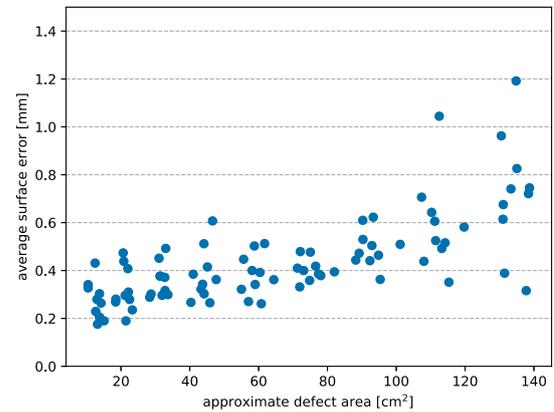
Fig. 7. The performance of the discriminative model in the context of a reconstructed surface area. While the variability of the model output errors increases with larger defects, the results are within an acceptable range even for a majority of the larger defects.

Generative model. To make the described reconstruction model generative, we make two modifications well known from GAN literature [32] to both CNNs. Namely, we add the adversarial loss function $\mathcal{L}_{Adv}$ in the form of a discriminator CNN $d(\cdot)$, which allows the model to learn the distribution $P(\hat{Y}|X_{defective})$, and inject a random latent vector into the reconstruction CNNs, which allows them to randomly sample from this distribution. We concatenate the random latent vector with the bottleneck tensor of both CNNs as shown in Fig. 4. The discriminator CNNs have the same architecture as the encoder part of the reconstruction CNNs with additional dense layers that output the discriminator scores. We use the improved Wasserstein GAN formulation with gradient penalty [33] during the training. Given a combination of the defective skull shape and the missing anatomy shape, the discriminator is trained to assign a low score $d(Y, X_{defective})$ to the ground-truth missing shape and a high score to the reconstructed missing shape $d(\hat{Y}, X_{defective})$ at both a low and high resolution, using the low-resolution discriminator d^{Lo} and high-resolution discriminator d^{Hi}. To optimize the reconstruction CNNs in this case, we use a combination of the reconstruction
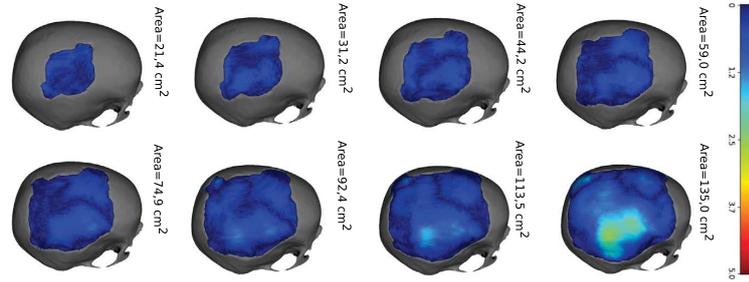
Fig. 8. Examples of results of the discriminative model reconstructions for a set of defects with different scales on a single test skull. An area where the model output deviates from the original shape by more than 2 mm can be observed in the last case.
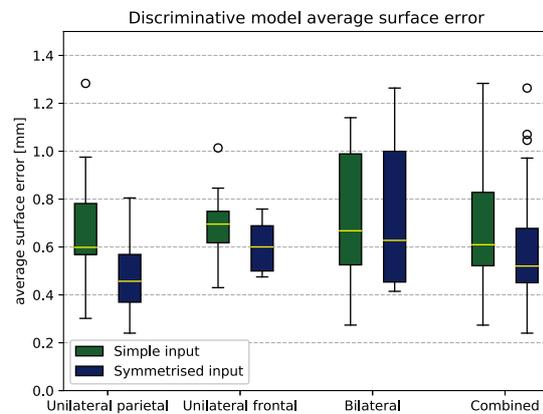

Fig. 9. The overall performance of the generative model on different groups of testing defective skulls. Average surface error [mm] for reconstruction with random latent vector (green) and for best-of-five reconstructions (blue).

Table 1
Average surface error [mm] for individual defect groups.

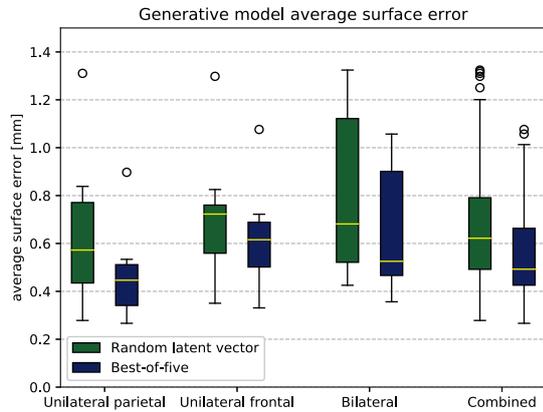| Method	Synthetic defects				Real defects
	UP	UF	Bi	Total	
Statistical shape models [15]	0.47	–	–	–	–
Discriminative + simple input	0.69	0.69	0.78	0.68	–
Discriminative + symmetrized input	0.48	**0.60**	0.73	0.59	0.80
Generative + symmetrized input (random)	0.63	0.71	0.81	0.68	–
Generative + symmetrized input (best of 5)	**0.46**	0.62	**0.65**	**0.56**	**0.69**

and adversarial loss, similarly to Wang et al. [24]. The losses of the reconstruction CNNs in this case are defined as

$$\mathcal{L}^{Lo} = \mathcal{L}^{Lo}_{Dice} + \alpha \mathcal{L}^{Lo}_{Adv}, \tag{3}$$

$$\mathcal{L}^{Hi} = \mathcal{L}^{Hi}_{Dice} + \alpha \mathcal{L}^{Hi}_{Adv}, \tag{4}$$

where α is set to 10^{-2}. We again optimize d^{Lo}, f^{Lo}, d^{Hi} and f^{Hi} iteratively for 300 000 training steps. For an overview of our method and both discriminative and generative models, see Fig. 4.

4. Experimental results

The experiments discussed in this section were run on a system with 11 GB Titan Xp GPU and a quad-core i5 processor with 24 GB RAM. The complete training of the models took approximately 8 days. After the model is trained, the method is able to fully reconstruct each skull in under 5 s, which is important for its efficient use in clinical practice. This is achieved by first inferring the low-resolution model on the full down-sampled volume and then sequentially inferring the high-resolution model on positions where the low-resolution model predicted a defect until the whole estimated defect area is processed. For visualization, the voxel grid was converted into a polygonal mesh which was then smoothed using a two-step smoothing algorithm [34].

We measured the precision of each method as the average symmetric unsigned distance between the surface voxels of the output reconstruction and the original anatomy shape which we considered to be ground-truth. We only measured the error on the outer surface of the skull because the inner surface is not relevant for cranial implant

design in the clinical practice. To get more insight into performance of our method, we divided the evaluation into four groups. The *unilateral-parietal*, *unilateral-frontal* and *bilateral* defect groups are described in Section 2 and the *combined* group includes all the defects, including random ones. Table 1 contains the average surface error for all models tested on each defect group.

We first evaluated the performance of the discriminative model and the effect of the symmetrized input on the error distribution in the testing set. The model with a simple input was able to reconstruct each testing skull successfully. However, we noticed that the errors in unilateral groups reached similar values as the bilateral group. This is in conflict with the expectation that while bilateral defects could allow for some variability in correctly completed shapes, the unilateral defects should be more directly constrained by the condition of symmetry and thus yield lower surface reconstruction errors. The effect of symmetrizing input as described in Section 3 was that the average measured error of the reconstructed unilateral defects dropped from 0.69 mm to 0.48 mm for parietal and from 0.69 mm to 0.60 mm for frontal defects. As expected, the bilateral defects group was less affected by the symmetrized input, although the error still slightly decreased since some bilateral defects are in fact partly constrained by the symmetry. The overall performance of the discriminative model for both simple and symmetrized input is shown in Fig. 5. The overall average surface error of the discriminative model with a symmetrized input for the whole testing set was 0.59 ± 0.21 mm. Several examples of the discriminative model reconstructions are shown in Fig. 6.

In order to explore the relation between the discriminative model performance and the area of the reconstructed defects, we created an extra set of nine cranial defects in each of the ten designated test skulls. The defects were created by subtracting the same shape with different scales from each skull (see Fig. 8 for their illustration). The surface area of the resulting skull defects ranged from 10 to 140 cm². The resulting surface errors of the discriminative model outputs are shown in Fig. 7 in the form of a scatter plot. While there is an apparent correlation between the measured surface error and the reconstructed surface area, the average surface error was under 0.7 mm for all defects up to an area of 100 cm². For even larger defects, the average surface error exceeded 1 mm in several cases. However, for majority of the cases, the surface errors of the results were still well under this value.
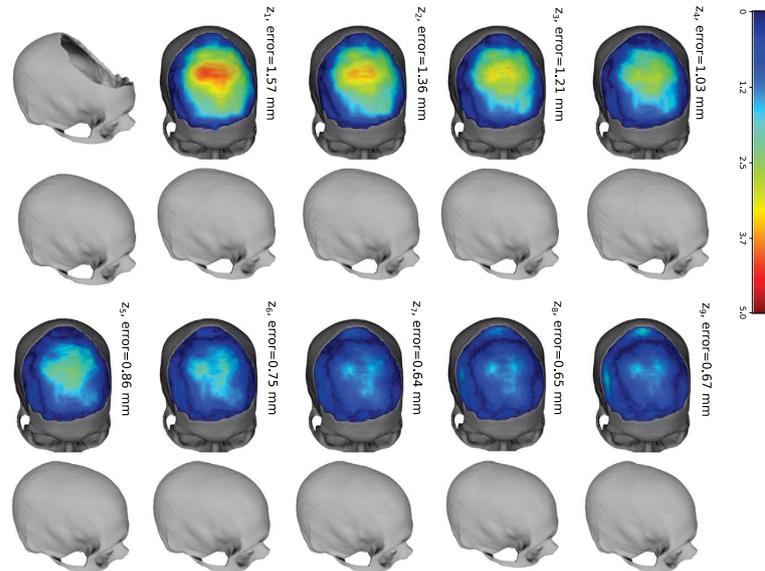
Fig. 10. The input defective skull (first row, left) and original skull shape (second row, left) and examples of output reconstructions resulting from linear interpolation in the latent space superimposed onto the input defective skull. Surface distance from the original shape can be seen decreasing in the middle part of the defect, however, it increases in areas near the defect border where errors are unacceptable for aesthetic reasons.
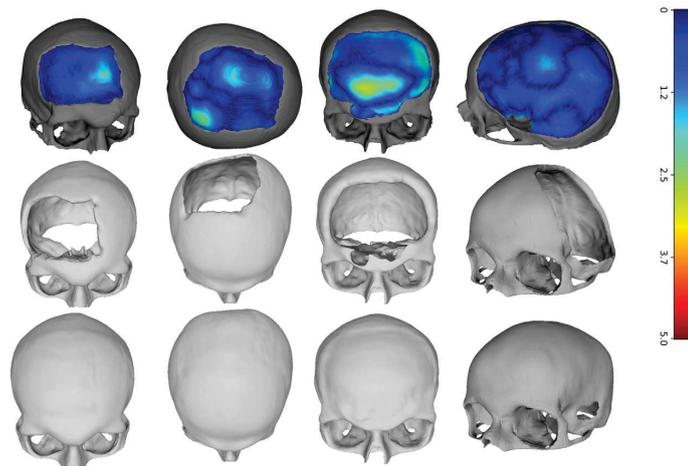
Fig. 11. Example outputs of the discriminative model for real patient data. Although some reconstruction faults can be seen in the last two cases, suggesting that real training data of target population should be added to the model in the future, the reconstruction is usually correct. The surface distance to the original shape is well below one millimeter on average.

Next, we evaluated the performance of the generative model with symmetrized input volumes and random input latent vectors z. The overall average surface error was 0.68 ± 0.28 mm. For each defect group, the error of the generative model was higher than that of the discriminative model. However, it should be noted that since now we consider multiple correct reconstructions for a single skull defect, the error measured against the ground-truth shape might not be a good indicator of the method's performance. The generative model allows us to sample multiple different outputs for a single input defective skull by changing the input latent vectors. Therefore, we also experimented with generating multiple reconstructions and measuring the best achieved

result. The overall average surface error when measuring the best-of-five sampled reconstructions for each testing skull was 0.56 ± 0.21 mm. The results for individual defect groups, as seen in Fig. 9, were similar to the discriminative model in this case. However, a reduction of the error can be noticed in the *bilateral* group, with the error reduced from 0.73 mm to 0.65 mm when compared to the discriminative model. This might once again be explained by the fact that due to weaker symmetry constraints in this group, the variability of acceptable reconstructions is greater. Therefore, generating subsequent different samples constrained on the same input increases the probability of generating at least one sample close to the original ground-truth shape.

6

To further illustrate the behavior of the generative model, we also conducted an experiment with latent space interpolation for one bilateral defect. We set both latent vectors for low-resolution z^{Lo} and high-resolution z^{Hi} to only contain constant values $c/10$ and we generated samples for $c = 1, 2 \ldots 9$. Our experiments showed that the generative model responds to these changes in total latent vector energy the most and we leave investigation into the limits of achievable anatomical variability in the output for future work. The resulting reconstructions of the generative model along with the ground-truth original shape are shown in Fig. 10. We also reported the measured surface errors against the ground-truth shape for each sample. It can be seen that the model is able to sample from the learned manifold of solutions, allowing for manipulating the reconstructed shape while still keeping a seamless connection to the original bone.

Finally, in order to evaluate the ability of our approach to generalize, we also tested the performance of the models trained exclusively on our synthetic dataset on an internal dataset of real defective patients without any fine-tuning of the model. Both models reconstructed the real defects mostly successfully. However, there was an expected increase in the surface error in both the discriminative model output and the best-of-five generative model output. In some cases, there were also visible faults such as slight depressions or even holes as seen in Fig. 11.

This could be partly attributed to the fact that the real testing patients come from a different geographic location, in which the anatomical variability of the skull is different [35]. Specifically, the differences in average shapes of the two datasets aligned using the same alignment method are illustrated in Fig. 13. The fact that these basic shape characteristics are learned by the low-resolution reconstruction model may lead to wrong estimation of the cranial volume in frontal part of the skull and even holes in parts which extend significantly beyond the anatomical variability observed in the training dataset. Overall, the outcome of this experiment is encouraging, although real defective patient scans from the target population should be added to the training process before evaluating the method performance in a real clinical setting.

5. Discussion

For deployment of reconstruction methods into the clinical workflow, several conditions must be met. First, symmetry of the skull should be preserved as well as possible, including in cases where the patient's skull itself is partly asymmetric and where the defect reaches partly into both sides of the skull. Second, the automatic reconstruction should fit very precisely to the defect borders. Although the models presented in this work will occasionally produce slightly asymmetric result or fail to avoid some depressions around the defect border, our results show that the proposed method can achieve an overall satisfactory performance in this regard, as illustrated by example reconstructions in Figs. 6 and 11. The measured average surface errors shown in Figs. 5 and 7 also show how the performance is affected by different shapes and sizes of the defects, including bilateral defects, orbital area reconstructions and defects with surface area of over $100 \, \text{cm}^2$. The implications of these results for the future implementation of the method into clinical practice should now be assessed by clinicians with experience in this area.

In the context of the current state of the art in the area of skull reconstruction, our approach differs from conventional mirroring-based and interpolation-based methods by its ability to reconstruct an arbitrary part of the skull present in the described dataset without requiring any parameter adjusting. Its ability to generalize to unseen skulls is, however, fully dependent on the variability of the training dataset used for model optimization. Fig. 12 demonstrates how using the model on a population where shapes of the skulls come from different distribution causes occasional faults and a slight increase in the average surface error of the reconstructions. Nevertheless, this issue will be mitigated
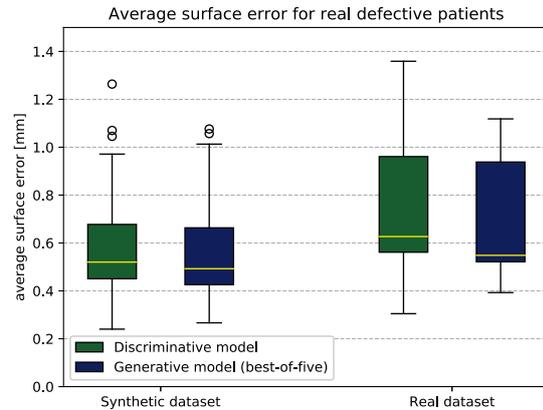

Fig. 12. Comparison of performance of the models' average surface error [mm] for reconstruction of synthetic defects and real defects. Results shown for discriminative model (green) and for best-of-five outputs of generative model (blue).

by introducing cases from the target population into the dataset and retraining the reconstruction model in the near future.

Methods based on statistical shape models also possess this dependency on training dataset variability and the potential ability to model any part of the skull. This makes them very similar to the approach proposed in this work in terms of possible target use cases. Fuessinger et al.[15] achieved an average surface error of $0.47 \, \text{mm}$ when reconstructing unilateral spherical defects of the cranial area with a radius of $5 \, \text{cm}$. This could be compared to the performance of our discriminative and generative models reaching 0.48 and $0.46 \, \text{mm}$ average surface error, respectively, on the unilateral parietal defect group. In contrast, our method does not require any manual cleaning of the defect border as the seamless fit of the reconstructed part to the rest of the skull is handled by the CNN model. It would be interesting to see the performance of the statistical shape model on more challenging parts of the introduced dataset such as defects of the orbital area and larger bilateral defects.

A more general comparison is currently limited by the lack of standardized datasets and methodology to evaluate the anatomical reconstruction methods. Especially in the case of bilateral defects in which symmetry cannot be used to uniquely define the correct output, we argue that although the absolute distance from the ground-truth shape might give an adequate estimate of how well a method performs, it should not be used as the single criterion of correct reconstruction. In addition to variability in cranium shape, modeling structures such as skull protuberances, sutures, or uneven surface is unnecessary for means of PSI design. Therefore, the most relevant metric to measure the reconstruction method performance would be the amount of time required by the operating expert to design clinically acceptable PSI from the initial reconstruction. However, this is infeasible without the method being deployed into clinical practice.

Since the discriminative model outputs reach lower average surface error than the randomly sampled outputs of the generative model, it can be concluded that it is more suitable for a completely automatic setting. However, the generative model could alternatively be used in a semi-automatic setting. In case the initial reconstruction is not satisfactory for further processing, several subsequent samples from the generative model could be offered to the expert to increase the chance of avoiding falling back to a less efficient conventional workflow.

Finally, the reconstruction method is not limited to skull reconstruction task or anatomical reconstruction in general. The method can potentially be applied to any shape completion task where both global contextual information as well as fine structural details need to be taken into account during the data volume reconstruction.
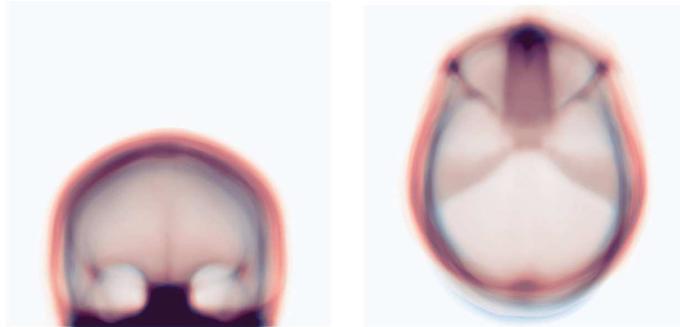
Fig. 13. Superimposed frontal (left) and axial (right) projections of the segmented skulls. The 10 testing cases of the synthetic dataset rendered in blue and the 7 testing cases of the internal dataset in red. A difference in several shape characteristics of the skulls can be observed.

6. Conclusions

This work presented a multi-scale cascaded CNN architecture for general shape completion applied to the reconstruction of missing skull anatomy in a fully automatic manner. We also showed that symmetrized input can increase the performance in this task and that both discriminative and generative models can be used successfully. The proposed method reaches enough precision and robustness to be considered in clinical practice. Validation was done on a synthetic dataset which closely mimics real patient cases and this dataset was made public.

The model trained exclusively on synthetic data also performs well on real defective patient cases, but adding samples from the real target population to the training should be considered in order to improve the results. Further testing with more patient data in clinical setting is now required to fully confirm its efficacy and identify any limitations.

Currently, the method is constrained to the cranial and orbital area. We plan to extend the method and the dataset to include maxilla and zygomatic bones as well in the future. Adapting the method for alternative data representations, such as point clouds or graphs, could also be explored as a way to improve processing speed and precision.

Declaration of competing interest

Michal Španěl is also a CEO of Tescan 3DIM company at the time of writing this paper.

Acknowledgments

References

[1] D.B. Kurland, A. Khaladj-Ghom, J.A. Stokum, B. Carusillo, J.K. Karimy, V. Gerzanich, J. Sahuquillo, J.M. Simard, Complications associated with decompressive craniectomy: A systematic review, Neurocrit. Care 23 (2) (2015) 292–304, http://dx.doi.org/10.1007/s12028-015-0144-7.

[2] M.-Y. Lee, C.-C. Chang, C.-C. Lin, L.-J. Lo, Y.-R. Chen, Custom implant design for patients with cranial defects, IEEE Eng. Med. Biol. Mag. 21 (2) (2002) 38–44, http://dx.doi.org/10.1109/MEMB.2002.1000184.

[3] A.L. Jardini, M.A. Larosa, R.M. Filho, C.A. de Carvalho Zavaglia, L.F. Bernardes, C.S. Lambert, D.R. Calderoni, P. Kharmandayan, Cranial reconstruction: 3D biomodel and custom-built implant created using additive manufacturing, J. Cranio-Maxillofac. Surg. 42 (8) (2014) 1877–1884, http://dx.doi.org/10.1016/j.jcms.2014.07.006.

[4] J. hyeon Oh, Recent advances in the reconstruction of cranio-maxillofacial defects using computer-aided design/computer-aided manufacturing, Maxillofac. Plast. Reconstr. Surg. 40 (1) (2018) http://dx.doi.org/10.1186/s40902-018-0141-9.

[5] H. van de Belt, D. Neut, W. Schenk, J.R. van Horn, H.C. van der Mei, H.J. Busscher, Infection of orthopedic implants and the use of antibiotic-loaded bone cements: A review, Acta Orthop. Scand. 72 (6) (2001) 557–571, http://dx.doi.org/10.1080/000164701317268978.

[6] K. Rudman, C. Hoekzema, J. Rhee, Computer-assisted innovations in craniofacial surgery, Facial Plast. Surg. 27 (04) (2011) 358–365, http://dx.doi.org/10.1055/s-0031-1283054.

[7] M. van Eijnatten, R. van Dijk, J. Dobbe, G. Streekstra, J. Koivisto, J. Wolff, CT image segmentation methods for bone used in medical additive manufacturing, Med. Eng. Phys. 51 (2018) 6–16, http://dx.doi.org/10.1016/j.medengphy.2017.10.008.

[8] R. Marsell, T.A. Einhorn, The biology of fracture healing, Injury 42 (6) (2011) 551–555, http://dx.doi.org/10.1016/j.injury.2011.03.031, Bone Regeneration in the 21st Century.

[9] M. Bashour, History and current concepts in the analysis of facial attractiveness, Plast. Reconstr. Surg. 118 (3) (2006) 741–756, http://dx.doi.org/10.1097/01.prs.0000233051.61512.65.

[10] X. Chen, L. Xu, X. Li, J. Egger, Computer-aided implant design for the restoration of cranial defects, Sci. Rep. 7 (1) (2017) http://dx.doi.org/10.1038/s41598-017-04454-6.

[11] Y.-W. Chen, C.-T. Shih, C.-Y. Cheng, Y.-C. Lin, The development of skull prosthesis through active contour model, J. Med. Syst. 41 (10) (2017) http://dx.doi.org/10.1007/s10916-017-0808-2.

[12] Y. Volpe, R. Furferi, L. Governi, F. Uccheddu, M. Carfagni, F. Mussa, M. Scagnet, L. Genitori, Surgery of complex craniofacial defects: A single-step AM-based methodology, Comput. Methods Programs Biomed. 165 (2018) 225–233, http://dx.doi.org/10.1016/j.cmpb.2018.09.002.

[13] A. Marzola, M. Servi, Y. Volpe, A reliable procedure for the construction of a statistical shape model of the cranial vault, in: C. Rizzi, A.O. Andrisano, F. Leali, F. Gherardini, F. Pini, A. Vergnano (Eds.), Design Tools and Methods in Industrial Engineering, Springer International Publishing, Cham, 2020, pp. 788–800.

[14] W. Semper-Hogg, M.A. Fuessinger, S. Schwarz, E. Ellis, C.-P. Cornelius, F. Probst, M.C. Metzger, S. Schlager, Virtual reconstruction of midface defects using statistical shape models, J. Cranio-Maxillofac. Surg. 45 (4) (2017) 461–466, http://dx.doi.org/10.1016/j.jcms.2016.12.020.

[15] M.A. Fuessinger, S. Schwarz, C.-P. Cornelius, M.C. Metzger, E. Ellis, F. Probst, W. Semper-Hogg, M. Gass, S. Schlager, Planning of skull reconstruction based on a statistical shape model combined with geometric morphometrics, Int. J. Comput. Assist. Radiol. Surg. 13 (4) (2017) 519–529, http://dx.doi.org/10.1007/s11548-017-1674-6.

[16] M.A. Fuessinger, S. Schwarz, J. Neubauer, C.-P. Cornelius, M. Gass, P. Poxleitner, R. Zimmerer, M.C. Metzger, S. Schlager, Virtual reconstruction of bilateral midfacial defects by using statistical shape modeling, J. Cranio-Maxillofac. Surg. 47 (7) (2019) 1054–1059, http://dx.doi.org/10.1016/j.jcms.2019.03.027.

[17] A. Morais, J. Egger, V. Alves, Automated computer-aided design of cranial implants using a deep volumetric convolutional denoising autoencoder, in: Advances in Intelligent Systems and Computing, Springer International Publishing, 2019, pp. 151–160, http://dx.doi.org/10.1007/978-3-030-16187-3_15.

[18] F. Buonamici, R. Furferi, L. Genitori, L. Governi, A. Marzola, F. Mussa, Y. Volpe, Reverse engineering techniques for virtual reconstruction of defective skulls: an overview of existing approaches, in: Proceedings of CAD2018, CAD Solutions LLC, 2018, http://dx.doi.org/10.14733/cadconfp.2018.6-10.

[19] A. Sharma, O. Grau, M. Fritz, VConv-DAE: Deep volumetric shape learning without object labels, in: G. Hua, H. Jégou (Eds.), Computer Vision – ECCV 2016 Workshops, Springer International Publishing, Cham, 2016, pp. 236–250.

[20] D. Stutz, A. Geiger, Learning 3D shape completion from laser scan data with weak supervision, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, http://dx.doi.org/10.1109/cvpr.2018.00209.

[21] W. Wang, D. Ceylan, R. Mech, U. Neumann, 3DN: 3D deformation network, in: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2019, http://dx.doi.org/10.1109/cvpr.2019.00113.

[22] A. Dai, C.R. Qi, M. Nießner, Shape completion using 3D-encoder-predictor CNNs and shape synthesis, in: 2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR, IEEE, 2017, http://dx.doi.org/10.1109/cvpr.2017.693.

[23] O. Litany, A. Bronstein, M. Bronstein, A. Makadia, Deformable shape completion with graph convolutional autoencoders, in: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, IEEE, 2018, http://dx.doi.org/10.1109/cvpr.2018.00202.

[24] W. Wang, Q. Huang, S. You, C. Yang, U. Neumann, Shape inpainting using 3D generative adversarial network and recurrent convolutional networks, in: 2017 IEEE International Conference on Computer Vision, ICCV, IEEE, 2017, http://dx.doi.org/10.1109/iccv.2017.252.

[25] A.H. Abdi, H. Borgard, P. Abolmaesumi, S. Fels, AnatomyGen: Deep anatomy generation from dense representation with applications in mandible synthesis, in: International Conference on Medical Imaging with Deep Learning, London, United Kingdom, 2019.

[26] A.H. Abdi, M. Pesteie, E. Prisman, P. Abolmaesumi, S. Fels, Variational Shape Completion for Virtual Planning of Jaw Reconstructive Surgery, in: Lecture Notes in Computer Science, Springer International Publishing, 2019, pp. 227–235, http://dx.doi.org/10.1007/978-3-030-32254-0_26.

[27] S. Chilamkurthy, R. Ghosh, S. Tanamala, M. Biviji, N.G. Campeau, V.K. Venugopal, V. Mahajan, P. Rao, P. Warier, Deep learning algorithms for detection of critical findings in head CT scans: a retrospective study, Lancet 392 (10162) (2018) 2388–2396, http://dx.doi.org/10.1016/s0140-6736(18)31645-3.

[28] Özgün. Çiçek, A. Abdulkadir, S.S. Lienkamp, T. Brox, O. Ronneberger, 3D U-Net: Learning dense volumetric segmentation from sparse annotation, in: Medical Image Computing and Computer-Assisted Intervention, MICCAI 2016, Springer International Publishing, 2016, pp. 424–432, http://dx.doi.org/10.1007/978-3-319-46723-8_49.

[29] M. Kolarik, R. Burget, K. Riha, Upsampling algorithms for autoencoder segmentation neural networks: A comparison study, in: 2019 11th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops, ICUMT, IEEE, 2019, http://dx.doi.org/10.1109/icumt48472.2019.8970918.

[30] S. Sun, Z. Cao, H. Zhu, J. Zhao, A survey of optimization methods from a machine learning perspective, IEEE Trans. Cybern. (2020) 1–14, http://dx.doi.org/10.1109/tcyb.2019.2950779.

[31] F. Milletari, N. Navab, S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, in: 2016 Fourth International Conference on 3D Vision, 3DV, IEEE, 2016, http://dx.doi.org/10.1109/3dv.2016.79.

[32] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial nets, in: Z. Ghahramani, M. Welling, C. Cortes, N.D. Lawrence, K.Q. Weinberger (Eds.), Advances in Neural Information Processing Systems, Vol. 27, Curran Associates, Inc., 2014, pp. 2672–2680.

[33] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, A.C. Courville, Improved training of Wasserstein GANs, in: I. Guyon, U.V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, R. Garnett (Eds.), Advances in Neural Information Processing Systems, Vol. 30, Curran Associates, Inc., 2017, pp. 5767–5777.

[34] A. Belyaev, Y. Ohtake, A comparison of mesh smoothing methods, in: Israel-Korea Bi-National Conference on Geometric Modeling and Computer Graphics, Tel Aviv University, 2003, pp. 83–87.

[35] Y.S.N. Jayaratne, R.A. Zwahlen, A systematic review of interethnic variability in facial dimensions, Plast. Reconstr. Surg. 129 (1) (2012) 164e–165e, http://dx.doi.org/10.1097/prs.0b013e3182362e3f.

Cranial Defect Reconstruction Using Cascaded CNN with Alignment

Oldřich Kodym$^{(\boxtimes)}$, Michal Španěl, and Adam Herout

Department of Computer Graphics and Multimedia, Brno University of Technology,
Božetěchova 2, 612 66 Brno, Czech Republic
`ikodym@fit.vutbr.cz`

Abstract. Designing a patient-specific cranial implant usually requires reconstructing the defective part of the skull using computer-aided design software, which is a tedious and time-demanding task. This lead to some recent advances in the field of automatic skull reconstruction with use of methods based on shape analysis or deep learning. The AutoImplant Challenge aims at providing a public platform for benchmarking skull reconstruction methods. The BUT submission to this challenge is based on skull alignment using landmark detection followed by a cascade of low-resolution and high-resolution reconstruction convolutional neural network. We demonstrate that the proposed method successfully reconstructs every skull in the standard test dataset and outperforms the baseline method in both overlap and distance metrics, achieving 0.920 DSC and 4.137 mm HD.

Keywords: Skull reconstruction · Shape completion · Cascaded convolutional networks

1 Introduction

Craniectomy is a procedure during which a specific part of the skull is resected and eventually replaced with a cranial implant. When designing the implant, the correct skull shape reconstruction is critical for satisfactory patient outcome. The shape of the implant should make it possible to restore the protective and aesthetic function of the skull and also fit very precisely along the border [7,8]. A successfully reconstructed skull should be mostly indistinguishable from a healthy skull. The original skull shape before the resection is therefore often used as the golden standard of the target reconstructed shape [9].

In case of unilateral defects, techniques based on mirroring the healthy part of the skull to the defect area are often used in combination with a Computer-Aided Design software (CAD) [2]. However, the assumption of perfectly symmetric skull does not hold in most cases and manual corrections are often required. To address these issues, recent methods aim to be completely or mostly automatic and to be able to reconstruct an arbitrary part of the skull, including

99

bilateral defects. One group of such methods is based on statistical shape models. In combination with geometric morphometrics, both unilateral and bilateral defects can be reconstructed with high precision [4,5]. Another group of methods that has been gaining considerable momentum in recent months is based on deep learning approaches. These methods usually make use of some form of volumetric convolutional neural networks (CNN) with an auto-encoder architecture, although output resolution may often be limited [10–12].

This paper presents a BUT submission to the MICCAI 2020 AutoImplant Challenge [9]. The proposed method is an adaptation of the cascaded reconstruction CNN architecture that has been recently applied to the SkullBreak dataset [6]. Furthermore, the method is extended by an automatic landmark-based registration and a detail-preserving morphological post-processing step. In our experiments, we show how different components of the method affect the reconstruction accuracy on a validation dataset of defective skulls. Finally, we report the results on the full testing dataset of the AutoImplant Challenge.

2 Proposed Method

The proposed method consists of several steps as illustrated in Fig. 1. The landmark detection step and the skull reconstruction step are handled by a 3D CNN model.

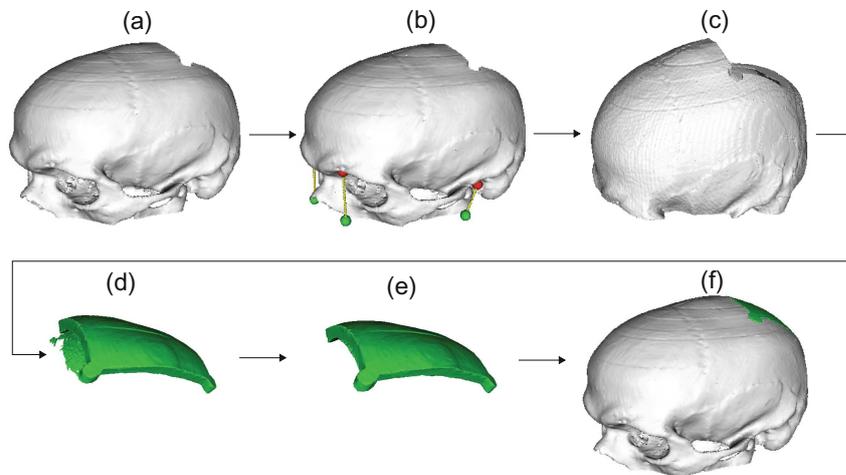**Fig. 1.** Overview of the proposed method. In the input skull volume (a) 4 landmarks are detected (b). The skull is transformed (c) so that the detected landmarks (red) are registered to the reference landmarks (green). Then, the skull is reconstructed by estimating the missing shape (d). Finally, the result is post-processed (e) and transformed back into the original skull coordinates (f). (Color figure online)

58   O. Kodym et al.

2.1   Skull Alignment

The defects in the AutoImplant dataset are generated on a static position inside the data volumes and the variability in their shapes and positions comes from the variability of positions of the skulls. When reconstructing the shapes with a volumetric CNN model, this introduces some difficulties. The reconstruction model needs to implicitly learn rotational and translational invariance and it also makes it cumbersome to exploit the symmetric properties of the skulls. To address this, we use scale and rigid transformations to normalize the scale and the position of the skulls.

Unlike the parameters of the scale transform that are known from the CT acquisition process, the parameters of the rigid transformation need to be inferred from the data. We use the positions of four anatomical landmarks, namely the left and right auditory meatus and left and right supraorbital notch (see Fig. 1b), to compute the transform. This allows us to avoid possible complications of using conventional registration methods, such as issues with substantial differences in initial positions of the data volumes and different anatomical regions present in the data.

We trained a simple 3D CNN model for landmark detection with a U-net architecture using the heatmap regression approach [13]. The detection model is illustrated in Fig. 2(left) and its training is further described in Sect. 3. After detecting the landmarks, we find the rigid transform that moves these landmarks onto reference landmarks placed on the xy plane using singular value decomposition [1]. Even if one landmark is not detected either because of the detection model failure or because of a skull defect, such missing detection can usually be identified [3] and the missing landmark position can be computed from the other three landmarks.

2.2   Skull Reconstruction

The skull reconstruction model takes the aligned binary defective skull data as an input and produces the missing part of the skull as an output. The model consists of two 3D CNNs with modified U-net architecture that are trained using the soft Dice loss. Both networks have additional max-pooling and up-sampling steps as compared to U-net to increase the field of view of the output neurons and only one convolutional layer at each resolution as shown in Fig. 2(right).

The first network takes a full data volume at a reduced resolution as an input and produces an estimate of the missing shape with the corresponding resolution. A laterally flipped copy of the volume is also concatenated to the input of this network to facilitate easier propagation of information from one side of the skull to the other [6]. The second network takes a single patch of the original resolution input concatenated to the up-sampled patch of the low-resolution estimate at the corresponding position and produces the final missing shape estimate in this patch. Both networks are trained using their respective resolution ground-truth. Each training step comprises of two updates. First, the

low-resolution network weights are updated using the low-resolution ground-truth. Next, both low- and high-resolution networks weights are updated using the high-resolution ground-truth. The patches are chosen randomly during the training. Evaluating the second network using a window sliding over all the positions in the low-resolution estimate produces the full missing shape at the original resolution.
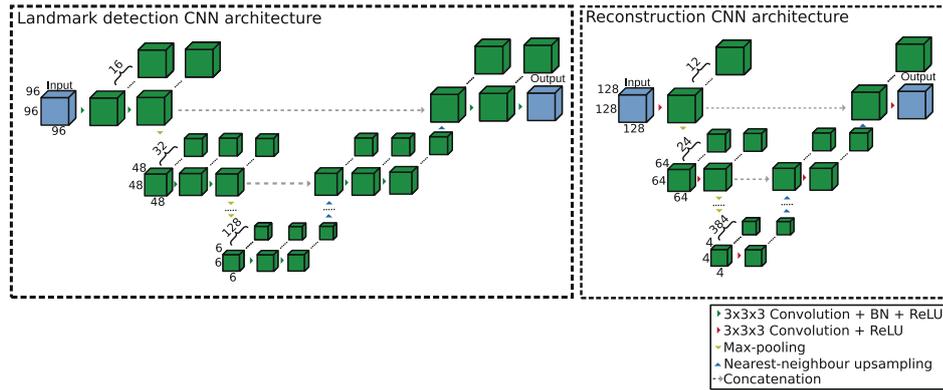
Fig. 2. Architectures of the 3D CNN models used for landmark detection (left) and shape reconstruction (right).

The architecture of both of the reconstruction networks is shown in Fig. 2(right) and the training details can be found in Sect. 3. The reconstruction model is described in further detail by Kodym et al. [6].

2.3   Shape Post-Processing

The reconstruction model will occasionally produce outputs that contain noise, such as disjoint objects or protuberances covering the healthy part of the skull as shown in Fig. 3(left). We make an assumption that the missing shapes should only consist of a single compact object. First, to isolate only the main missing shape, we use connected component analysis and discard all objects except the largest one. Second, we use morphological opening operation to remove any shape protuberances with less than desired minimum shape thickness.

However, the opening operation also tends to produce overly smooth shapes along the defect edges where it is desirable to keep the fine details produced by the reconstruction model. To address this, we keep both the original and morphologically open shapes. We then apply an additional morphological dilation to the open shape, producing a mask that is slightly bigger than the original shape but does not include the protuberances. Masking the original shape with such a mask results in a shape with the original fine details but without the larger protuberances as shown in Fig. 3(right).
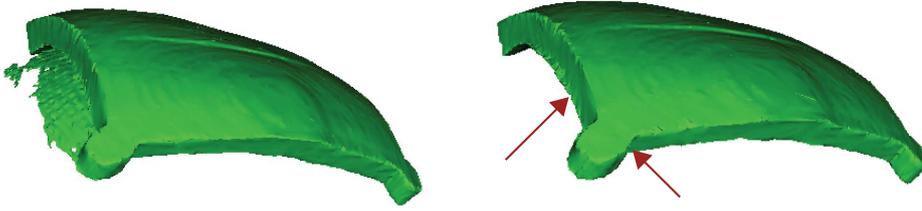
Fig. 3. Example of the detail-preserving morphological post-processing of the estimated missing shape. Note that the undesired protuberance is removed while the fine details are preserved along the object border.

3   Experiments

In this section, we describe the experiments and show the effect of individual method components on the reconstruction outputs. All the experiments were run on a system with Titan Xp GPU with 12 GB GRAM.

3.1   Landmark Detection

We manually annotated the four landmarks in all 100 training skull volumes. We trained the landmark detection CNN model on 90 samples, leaving 10 skulls for validation. The model was trained for 100 000 iterations using Adam optimizer with training step 10^{-4} and the dataset was strongly augmented using random rotations to ascertain that the model is able to detect the landmarks in cases of arbitrary patient positions inside the scanner.

The results of the landmark detection on the 10 validation cases can be seen in Fig. 4(left). The auditory meatus landmarks were detected with error of 1.22 ± 0.70 mm while the supraorbital notch landmarks achieved a slightly higher error of 1.84 ± 1.03 mm. An important observation is that the trained model also succeeded in detection of all four landmarks in all the 110 testing cases as well, and every skull could be aligned fully automatically without any manual intervention at test time.

3.2   Missing Shape Inference

Similarly to the landmark detection model, the reconstruction networks were also trained on 90 training samples. For the ablation experiments in this work, both low- and high-resolution networks were trained on batches of 4 samples using Adam optimizer with training step 10^{-4} for 50 000 iterations using resolution of 3.2 mm per voxel and 0.4 mm per voxel, respectively. All data volumes were padded to dimensions $512 \times 512 \times 512$ which means that the corresponding low-resolution samples had dimensions $64 \times 64 \times 64$. Random lateral flips were used to augment the dataset.

We trained three different reconstruction models. The *basic cascade model* is trained on the original provided challenge data. The mirrored input channel

is not used in the low-resolution network of this model as the sagittal plane is not known. The *aligned model* is trained on the data that have been previously aligned using the detected landmark positions. This also allows us to use the mirrored channel in this model. The *aligned and augmented* model is also trained on additional defective skulls that have been created from the training complete skulls. Five defects were created on each skull using random shapes similarly to the SkullBreak dataset, resulting in additional 450 training cases. We also created 10 additional validation cases using the same process.

The results of the reconstruction model on the validation cases are shown in Fig. 4(right). The basic cascade model had the worst performance on the validation cases, achieving average Dice score of 0.835. Simply aligning the data and adding the mirrored input to the low-resolution network in the aligned model had a substantial effect on the model performance, reaching 0.895 Dice score and showing the benefit of reducing the degrees of freedom of the defects during the reconstruction. However, both models overfit strongly to the training dataset with specific shape and position of the defects and were unable to generalize to the additional augmented validation cases where the distribution of defect shapes and positions is different. The aligned and augmented model trained on the additional defective cases, on the other hand, was able to both reconstruct the additional validation cases and increase the original data accuracy to Dice score 0.903.
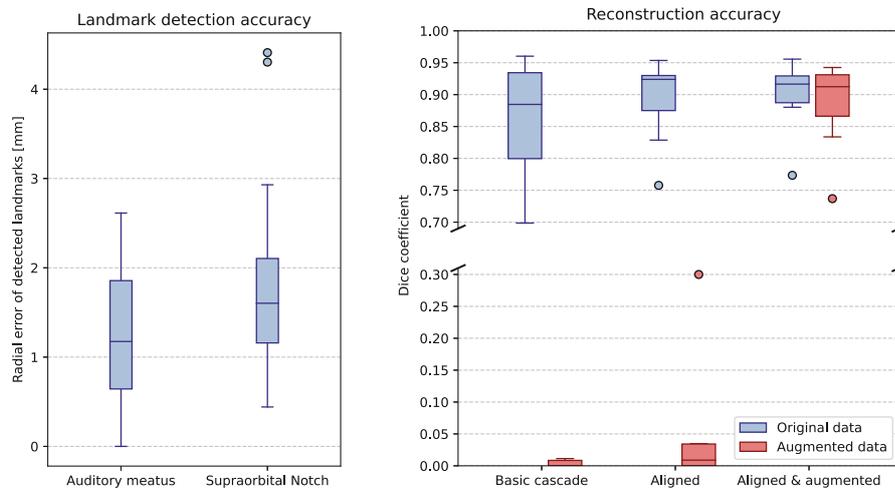
Fig. 4. Accuracy of the landmark detection (left) and the reconstruction models (right) on the validation cases.

4   Results

We aligned both subsets of the final 110 test cases of the AutoImplant challenge using the landmark detection model. For reconstruction, we used the aligned and

62   O. Kodym et al.

Table 1. The results of the proposed method on the AutoImplant Challenge test dataset in terms of Dice score and Hausdorff distance.

	Test case (100)	Test case (10)	Overall (110)
Mean DSC	0.920	0.910	0.919
Mean HD	4.137	4.707	4.189
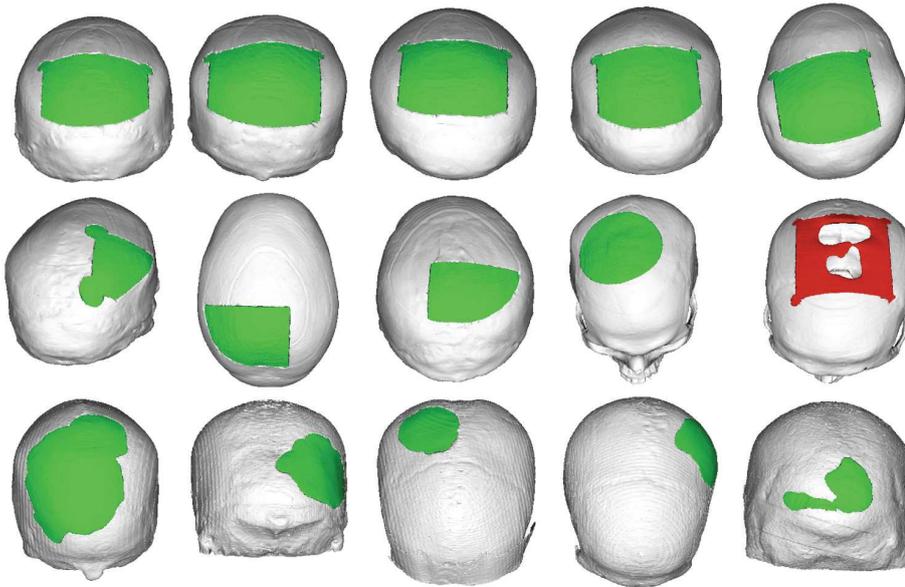
Fig. 5. Examples of the reconstruction results. From top row to bottom: The standard test set, the additional test set and the augmented validation set. Reconstruction failure could be observed in the last case of the additional test set in red color. (Color figure online)

augmented model that had been trained for 120 000 iterations. We also increased the first reconstruction network resolution to 1.6 mm per voxel, resulting in low-resolution volumes of dimensions $128 \times 128 \times 128$ voxels in the final evaluated model. To discard the occasional artifacts, we used the post-processing method described in Sect. 2.3. Both standard and additional subsets of the test dataset were reconstructed completely automatically without any manual interactions. The landmark detection model, the aligned training dataset and the augmented training dataset are publicly available[1].

The results of the proposed method on the challenge test dataset in terms of Dice coefficient and Hausdorff distance are shown in Table 1. Several qualitative examples of the reconstruction output on the standard subset, the additional subset and also the augmented validation dataset are shown in Fig. 5 where one case of reconstruction failure on the additional test set can also be observed.

---

[1] https://github.com/OldaKodym/BUT_autoimplant_public.

5 Conclusion

Our experiments showed that the skull alignment and data augmentation techniques we used increase the accuracy of the skull reconstruction. These are general concepts that could be applied to any other reconstruction model. Although we only encountered one failure case in our experiments, it hints at the fact that more defect shape augmentations should be used to increase robustness of the reconstruction model. It is currently unknown whether the achieved accuracy in terms of Dice coefficient and Hausdorff distance could warrant clinical applicability of the method. However, visual inspection of the reconstructed defects shows no visible artifacts in most cases.

While the reconstruction method reaches good accuracy, the final shape will usually have to be further edited by an experienced clinician in medical practice. Therefore, it would be beneficial to explore ways to include interactivity in the implant design method, possibly drawing inspiration from interactive convolutional networks that have been successfully applied to segmentation tasks. Another interesting research direction is leveraging different data representations such point clouds or level sets.

References

1. Besl, P., McKay, N.D.: A method for registration of 3-D shapes. IEEE Trans. Pattern Anal. Mach. Intell. **14**(2), 239–256 (1992). https://doi.org/10.1109/34.121791

2. Chen, X., Xu, L., Li, X., Egger, J.: Computer-aided implant design for the restoration of cranial defects. Sci. Rep. **7**(1), 1–10 (2017). https://doi.org/10.1038/s41598-017-04454-6

3. Drevický, D., Kodym, O.: Evaluating deep learning uncertainty measures in cephalometric landmark localization. In: Proceedings of the 13th International Joint Conference on Biomedical Engineering Systems and Technologies. SCITEPRESS - Science and Technology Publications (2020). https://doi.org/10.5220/0009375302130220

4. Fuessinger, M.A., et al.: Planning of skull reconstruction based on a statistical shape model combined with geometric morphometrics. Int. J. Comput. Assist. Radiol. Surg. **13**(4), 519–529 (2017). https://doi.org/10.1007/s11548-017-1674-6

5. Fuessinger, M.A., et al.: Virtual reconstruction of bilateral midfacial defects by using statistical shape modeling. J. Cranio-Maxillofac. Surg. **47**(7), 1054–1059 (2019). https://doi.org/10.1016/j.jcms.2019.03.027

6. Kodym, O., Španěl, M., Herout, A.: Skull shape reconstruction using cascaded convolutional networks. Comput. Biol. Med. **123**, 103886 (2020). https://doi.org/10.1016/j.compbiomed.2020.103886

7. Kurland, D.B., et al.: Complications associated with decompressive craniectomy: a systematic review. Neurocrit. Care **23**(2), 292–304 (2015). https://doi.org/10.1007/s12028-015-0144-7

64   O. Kodym et al.

8. Lee, M.Y., Chang, C.C., Lin, C.C., Lo, L.J., Chen, Y.R.: Custom implant design for patients with cranial defects. IEEE Eng. Med. Biol. Mag. **21**(2), 38–44 (2002). https://doi.org/10.1109/MEMB.2002.1000184

9. Li, J., Pepe, A., Gsaxner, C., Campe, G., Egger, J.: A baseline approach for autoimplant: the MICCAI 2020 cranial implant design challenge. In: Syeda-Mahmood, T., et al. (eds.) CLIP/ML-CDS -2020. LNCS, vol. 12445, pp. 75–84. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-60946-7_8

10. Li, J., Pepe, A., Gsaxner, C., Egger, J.: An online platform for automatic skull defect restoration and cranial implant design (2020)

11. Matzkin, F., et al.: Self-supervised skull reconstruction in brain CT images with decompressive craniectomy. In: Martel, A.L., et al. (eds.) MICCAI 2020. LNCS, vol. 12262, pp. 390–399. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-59713-9_38

12. Morais, A., Egger, J., Alves, V.: Automated computer-aided design of cranial implants using a deep volumetric convolutional denoising autoencoder. In: Rocha, Á., Adeli, H., Reis, L.P., Costanzo, S. (eds.) WorldCIST'19 2019. AISC, vol. 932, pp. 151–160. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-16187-3_15

13. Payer, C., Štern, D., Bischof, H., Urschler, M.: Regressing heatmaps for multiple landmark localization using CNNs. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 230–238. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_27

Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation

Ondrej Klima[a], Petr Kleparnik[a], Michal Spanel[b], and Pavel Zemcik[a]

[a]Department of Computer Graphics and Multimedia, Brno University of Technology,
Bozetechova 1/2, 612 66 Brno, Czech Republic
[b]3Dim Laboratory s.r.o, Kamenice 34, 625 00 Brno, Czech Republic

ABSTRACT

The reconstruction of a patient-specific 3D anatomy is the crucial step in the computer-aided preoperative planning based on plain X-ray images. In this paper, we propose a robust and fast reconstruction methods based on fitting the statistical shape and intensity model of a femoral bone onto a pair of calibrated X-ray images. We formulate the registration as a non-linear least squares problem, allowing for the involvement of Levenberg-Marquardt optimisation. The proposed methods have been tested on a set of 96 virtual X-ray images. The reconstruction accuracy was evaluated using the symmetric Hausdorff distance between reconstructed and ground-truth bones. The accuracy of the intensity-based method reached 1.18 ± 1.57 mm on average, the registration took 8.76 seconds on average.

Keywords: Preoperative planning, image registration, 2D/3D reconstruction, statistical shape and intensity model, Levenberg-Marquardt optimisation, GPU acceleration.

1. INTRODUCTION

In the field of the orthopaedic traumatology, surgical intervention is often preceded by a preoperative planning. The common aim of the preoperative planning is to get acquainted with the bone fracture and to at least roughly plan repositioning of the bone fragments. If the reposition is done virtually using 3D planning software, precise identification of the best shaped patient-specific bone plate is possible as well. Typically, the treatment of a traumatized bone requires Radiography (X-ray) and, in the difficult cases, the Computed Tomography (CT) examination is needed as well. The virtual planning is typically based on 3D models of bones or their fragments usually extracted from CT data sets captured with high level of detail. However, in comparison to the plain X-ray imaging, in case of CT, the patient is exposed to higher radiation doses during the CT examination.[1] In addition, the CT imaging is more time consuming and more expensive. Therefore, preoperative planning based on plain X-ray images has been brought into focus in recent years. For the purposes of the planning, it is important to reconstruct the 3D patient-specific anatomy. The reconstruction is usually achieved by a deformable 2D/3D registration of the shape prior into the set of co-registered X-ray images.

To satisfy requirements of the involvement in clinical planning software, the 2D/3D reconstruction method must be robust and fast enough for the use in urgent cases. As the registration is an iterative process, its performance depends mainly on the rate of convergence of the involved optimisation method. Another important performance factor is the level of parallelization of the time demanding parts of the registration pipeline. We propose three considerable fast methods. The first proposed intensity-based method is capable of recovering the surface and even internal structures of the reconstructed bone, bringing more benefits to the pre-operative planning. The next two proposed *Black & White* methods focus only on the surface reconstruction, but outperform the intensity-based method in speed or accuracy, depending on the global or local formulation of the registration. It is assumed that the bone background is segmented out from the original X-ray images by the user in a semiautomatic manner.

Send correspondence to Ondrej Klima or Michal Spanel.
Ondrej Klima: E-mail: iklima@fit.vutbr.cz, Telephone: +420 54114-1402
Michal Spanel: E-mail: spanel@3dim-laboratory.cz, Telephone: +420 728 670 984

The main contribution of the paper is the formulation of the registration in such a manner that it can be solved using the Levenberg-Marquardt optimisation algorithm,[2] which is a well established numerical method with a high rate of convergence and its involvement leads to a significant speed-up of the registration. For orientative comparison, Ehlke's intensity-based method[3] takes 1:41 minutes on average for one pelvic bone reconstruction, while our intensity-based method requires only 8.8 seconds on average to reconstruct a femoral bone. The second contribution is the formulation of the local similarity evaluation, leading to highly accurate deformable registration. The last contribution is the acceleration of certain parts of the registration pipeline using the graphics hardware.

The rest of the paper is organized as follows. There is a brief summary of the related literature with a particular focus on the intensity-based reconstruction methods in Section 2. The novel Levenberg-Marquardt based methods are proposed in Section 3 and their evaluation is reported in Section 4. Section 5 concludes the paper.

2. RELATED WORK

Two major categories of previously published 2D/3D reconstruction approaches can be distinguished. The first category comprises of methods based on features, usually edges or bone silhouettes extracted from the original X-ray images, and from the statistical shape models (SSM). The second category is formed by intensity-based methods, which work directly with pixels contained in the original X-ray images, and pixels rendered from the statistical appearance model. The categories differ in the anatomy features they are able to reconstruct, involved shape prior and their performance. The feature-based methods usually use a polygonal mesh SSM which makes them eligible only for the reconstruction of the bone shape, while the intensity-based methods involve a statistical appearance model and therefore, beyond the bone shape, they are commonly capable of reconstructing the bone densities and consequently the anatomy features such as compact and spongy bone. On the other hand, it requires considerably less effort to construct a plain shape model in comparison with training an appearance model. In addition, there are more stringent requirements for the training data of the appearance model, as CT data sets must be captured with the same X-ray energy. Also, the extraction of shape model features is more straightforward and efficient than the extraction of the bone densities from the appearance model. Methods from both categories require close manual initialization.

Yao proposed an intensity-based method for the reconstruction of pelvic and femoral bones. The major contribution of his work was the proposal of novel shape and appearance prior named statistical shape and intensity model (SSIM).[4] The SSIM model is based on a volumetric mesh and describes bone densities using analytical functions which allow efficient manipulation with the bone geometry in comparison with the voxel-based appearance models. Yao's work was continued by Sadowsky,[5] who focused on effective rendering of virtual X-ray images, also referred to as digitally reconstructed radiographs (DRRs). Sadowsky derived his rendering approach from the Projected Tetrahedra (PT)[6] algorithm, replaced the numerical evaluation of integrals of the rays intersecting the SSIM by the closed formula solution and involved the GPU acceleration of the proposed method. He further extended his approach[7] and exploited it for the reconstruction of the pelvic bone from X-ray images with limited field of view.[8] The rendering approach was partially adopted by Ehlke,[3] who focused on the full OpenGL acceleration of the DRR generation from the Yao's SSIM. Ehlke reused the closed formula solution for the ray integrals computation, but proposed novel approach for the tetrahedron thickness calculation instead of PT based method. He used the method for the single view registration of the pelvic bone. Gong involved the SSIM based deformable registration for the simultaneous fractured distal radius reduction and 2D/3D reconstruction.[9] Intensity-based reconstruction approaches using special appearance models were also investigated. Lamecker investigated the usability of the thickness images instead of DRRs for the intensity-based reconstruction of a pelvis.[10] Hurvitz constructed a statistical appearance model capturing whole CT data sets instead of modeling only the bone of interest.[11] The registration took advantage from authentically looking DRRs containing the anatomy of interest including the surrounding soft tissues and bone joints. Tang proposed a method based on a special shape model comprising of not intersecting spheres.[12]

Beyond the methods listed above, feature-based approaches focused on a femoral bone reconstruction were proposed by Zheng[13] and Baka.[14] Zheng presented a reconstruction of the proximal femur using a method based on 3D similarity metric and establishment of correspondences between the 2D bone silhouettes and the
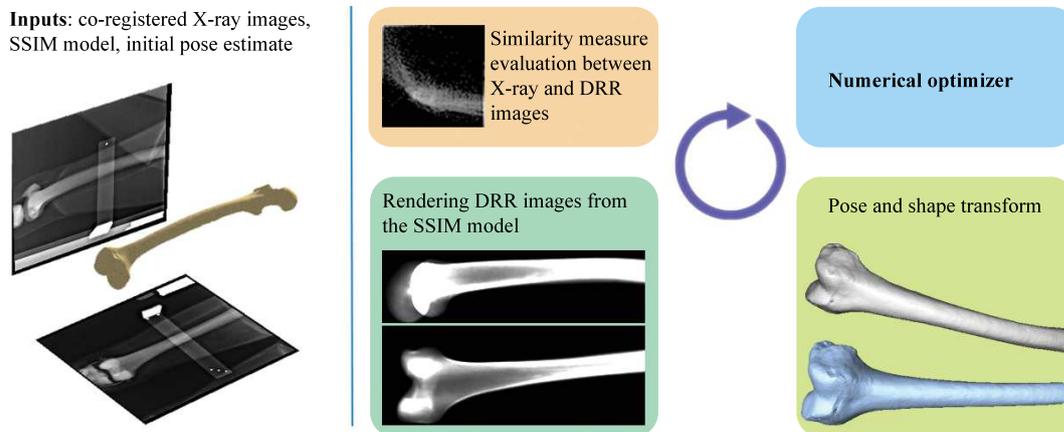
Figure 1. General scheme of the deformable 2D-3D registration process.

3D shape model. He assumed that the bone silhouettes are extracted from the original X-ray images by the user in a semiautomatic manner. Baka proposed a method combining the 3D similarity metric with the automatic extraction and selection of the relevant bone silhouettes from the X-ray images.

Yao also investigated factors affecting the deformable registration accuracy.[15] He found out that the best accuracy with respect to the running time is achieved when using two X-ray images. The experiments we performed revealed that the reconstruction error is minimal when the X-ray images are orthogonal. On the contrary, the view angles between the X-ray planes and the captured bone have no significant effect on the reconstruction accuracy. As expected, the reconstruction error is highly correlated with the X-ray images noise level, distortion and co-registration error.

The 2D/3D registration methods generally involve numerical optimisation. Most methods exploit various modifications of the gradient-descent algorithm.[3,16] Downhill Simplex optimiser was used by Sadowsky. Gong employed the Covariance Matrix Adaptation Evolution Strategy (CMA-ES).[17]

For the proposed intensity-based method, we have adopted Yao's SSIM appearance model[4] and partially Ehlke's accelerated reconstruction approach.[3] We replaced the gradient-descent based optimisation by the Levenberg-Marquardt method. The improved method significantly outperforms the Ehlke's original approach in the registration speed. To the best of our knowledge, no Levenberg-Marquardt algorithm based registration method has been proposed so far.

3. METHOD

The pipeline for the 2D/3D registration of the SSIM into the set of calibrated X-ray images is built as numerical optimisation. In each iteration, DRRs are rendered from the shape and intensity model. Differences between the original X-ray images and the corresponding DRRs are evaluated using an image similarity measure. The initial pose and shape parameters of the shape model are then adjusted to minimize the dissimilarities between original and rendered images. The patient-spefic bone model is reconstructed when the similarity between the DRR and X-ray images is maximal. The registration scheme is depicted in Figure 1. Accordingly to Yao's investigation of the accuracy factors, we reconstruct the 3D bone model from two orthogonal radiographs, usually taken from the anterior-posterior (AP) and lateral (LAT) view.

3.1 Statistical Shape and Intensity Model (SSIM)

We have largely adopted the Yao's SSIM appearance model which describes the shape variability of femur using a point distribution model (PDM)[18] and a reference tetrahedral mesh of the femoral bone. PDM is trained

Figure 2. A cross-section of a tetrahedral model of the proximal femur (left). Digitally reconstructed radiographs rendered from instances of the shape and intensity model. The value of the first principal components of the left and right bone corresponds to 2σ and -2σ respectively (middle). Binary masks rendered from the same instances of the PDM (right).

from tetrahedral meshes extracted from 22 CT data sets obtained from Virtual Skeleton Database (VSD).[19] The meshes have been brought into correspondence using the `Elastix` software[20] and aligned using the generalized Procrustes analysis (GPA) with the omitted re-scaling step. The linear model of the following form has been obtained by applying the Probabilistic Principal Component Analysis (PPCA):[21]

$$S = \phi_s \mathbf{b} + \overline{S} + \epsilon_s \tag{1}$$

where S is the volumetric bone model generated according to the given shape parameters $\mathbf{b}$, $\overline{S}$ is the mean bone shape, ϕ_s is the matrix of principal components and ϵ_s is a zero-mean Gaussian-distributed noise. The count n_b of the shape parameters, strictly lesser than the number of bones in the training set, can be chosen arbitrarily. In contrast to Yao, we have created the femoral volumetric mesh following the Si's Delaunay-based tetrahedralization.[22] The constructed tetrahedral model, illustrated in Figure 2 left, contains 104 thousand of tetrahedra and 26 thousand of vertices. According to Yao, the bone density is described in each tetrahedron independently using Bernstein polynomials:

$$D(\mu) = \sum_{\forall i,j,k,l \in \mathbb{Z} \wedge i+j+k+l=n} C_{i,j,k,l} B_{i,j,k,l}^n(\mu) \tag{2}$$

where $D(\mu)$ is a bone density in a certain point inside the tetrahedron, μ is a barycentric coordinate of that point, n is a degree of the Bernstein polynomial, $C_{i,j,k,l}$ are the polynomial coefficients and $B_{i,j,k,l}^n$ is the Bernstein basis function:

$$B_{i,j,k,l}^n(\mu) = \frac{n!}{i!j!k!l!} \mu_x^i \mu_y^j \mu_z^k \mu_w^l \tag{3}$$

We have involved polynomials of the 2nd degree resulting in requirement of 10 coefficients per tetrahedron. The coefficients have been obtained by solving an over-constrained system of linear equations as shown in Ref. 4. The generative model describing bone densities has been created using PPCA:

$$C = \phi_d e + \overline{C} + \epsilon_d \tag{4}$$

where C is a vector containing $C_{i,j,k,l}$ coefficients for each tetrahedron in the bone model, generated w.r. to the density parameters e, $\overline{C}$ is a vector of mean coefficients, ϕ_d is a matrix of principal components and ϵ_d is noise.

3.2 Digitally Reconstructed Radiographs

The rendering of virtual X-ray images is performed by a projection of the SSIM appearance model following the Ehlke's GPU accelerated approach. The X-ray beam passing through a bone is exponentially attenuated according to the Beer-Lambert law:

$$I_{out} = I_{in} e^{-\int_{w_{in}}^{w_{out}} \alpha(w)dw} \tag{5}$$

where I_{in} is the intensity of the beam entering the bone at the Cartesian coordinates w_{in}, I_{out} is the output intensity of the attenuated beam at the point w_{out} and $\alpha(w)$ is linear attenuation coefficient of a tissue. The overall attenuation encountered by the beam passing the single tetrahedron is determined by the closed-form expression:

$$\int_{w_{in}}^{w_{out}} D(\mu)d\mu = \|w_{out} - w_{in}\| \sum_{\forall i,j,k,l \in \mathbb{Z} \wedge i+j+k+l=n}^{n} C_{i,j,k,l} \int_{\mu_{in}}^{\mu_{out}} B_{i,j,k,l}^{n}(\mu)d\mu \tag{6}$$

where μ_{in} and μ_{out} are the barycentric coordinates of the ray entrance and exit respectively. Because the limits of the definite integrals of the Bernstein basis functions are in barycentric coordinates, the overall sum has to be multiplied by the actual distance between the w_{in} and w_{out} points. The definite integral of the Bernstein basis function has the following closed-form solution according to Sadowsky:[5]

$$\int_{\mu_{in}}^{\mu_{out}} B_{i,j,k,l}^{n}(\mu)d\mu = \frac{1}{n+1} \sum_{i' \le i, j' \le j, k' \le k, l' \le l} B_{i',j',k',l'}^{i'+j'+k'+l'}(\mu_{in}) B_{i-i',j-j',k-k',l-l'}^{i-i'+j-j'+k-k'+l-l'}(\mu_{out}) \tag{7}$$

For more details we refer to Ref. 5. The computation of the overall attenuation is performed using OpenGL fragment shaders as proposed by Ehlke.[3] Sample virtual X-ray images rendered from the different parts and instances of the constructed SSIM model are depicted in Figure 2 middle.

3.3 Intensity-based Registration

The intensity-based registration is performed by minimizing differences between the original radiographs and the DRRs rendered from SSIM. As the CT imaging is performed with higher X-ray energy than the plain radiography, the X-ray and DRR images may differ in contrast of corresponding tissues. Therefore, we involve the Normalized Mutual Information (NMI) similarity measure,[23] commonly used in inter-modal registration.

We set the similarity vector F to be:

$$F(\beta) = \big(\mathrm{NMI}_{AP}(\beta), \quad \mathrm{NMI}_{LAT}(\beta)\big) \tag{8}$$

where $\beta = (R, T, b)$ concatenates the rotation, translation and shape parameters of SSIM respectively. NMI_{AP} and NMI_{LAT} describe the similarity between the radiographs and corresponding DRRs rendered with respect to the β parameters. If the radiographs and DRRs are exactly the same images, then $F(\beta) = F_{max} = (2, \quad 2)$. The optimisation recovers the ideal parameter vector β in non-linear least squares manner:

$$\underset{\beta}{\mathrm{argmin}} = \big(F_{max} - F(\beta)\big)\big(F_{max} - F(\beta)\big)^{T} \tag{9}$$

The least squares problem is solved using the Levenberg-Marquardt algorithm. The Levenberg-Marquardt method performs local optimisation and therefore the close initial estimate of the SSIM pose has to be provided by the user. In each iteration, the change δ of the parameter vector β is obtained by solving the equation:

$$\big(J_F^T J_F + \lambda \mathrm{diag}(J_F^T J_F)\big)\delta = J_F^T \big(F_{max} - F(\beta)\big)^{T} \tag{10}$$

where λ is a damping factor and J_F is the Jacobian matrix containing partial derivatives of the similarity measures with respect to the rotation, translation and shape parameters:

$$J_F = \frac{\partial F}{\partial \beta} = \begin{pmatrix} \dfrac{\partial \mathrm{NMI}_{AP}}{\partial[r_x, r_y, r_z]} & \dfrac{\partial \mathrm{NMI}_{AP}}{\partial[t_x, t_y, t_z]} & \dfrac{\partial \mathrm{NMI}_{AP}}{\partial[b_1, b_2, \ldots, b_{n_b}]} \\ \dfrac{\partial \mathrm{NMI}_{LAT}}{\partial[r_x, r_y, r_z]} & \dfrac{\partial \mathrm{NMI}_{LAT}}{\partial[t_x, t_y, t_z]} & \dfrac{\partial \mathrm{NMI}_{LAT}}{\partial[b_1, b_2, \ldots, b_{n_b}]} \end{pmatrix} \tag{11}$$

As it is not possible to evaluate the J_F matrix using a closed-form solution, the finite differences approximation is used:

$$\frac{\partial f_p}{\partial \beta_q} \approx \frac{f_p(\beta_q + \epsilon) - f_p(\beta_q - \epsilon)}{2\epsilon} \tag{12}$$

where p, q denotes the J_F matrix row and column respectively. For the pose parameters, we set the $\epsilon_r = \epsilon_t = 1$. For the shape parameters **b** the best value of ϵ_b is investigated in Section 4.1. In case of pose parameters, the ϵ_r, ϵ_t are in units of [°] and [mm] respectively. For the shape parameters b_v, the ϵ_{b_v} are in units of standard deviations σ_k of the v-th principal component of the SSIM. Determining of the approximated J_F matrix of size $2 \times count(\beta)$ requires rendering of $2 * 2 * count(\beta)$ DRRs and consequently evaluation of the same number of NMI similarity measures. Therefore, the evaluation of the J_F matrix is the most time consuming part of the optimisation. The optimisation stops when the J_F is a zero matrix.

3.4 Black & White Registration (BW)

In contrast to the intensity-based registration, the *Black & White* pixel-based reconstruction involves a plain PDM model instead of the SSIM. In case of PDM, the DRR images are replaced by binary masks, as shown in Figure 2 right. The binary masks must be extracted from the original radiographs as well. The BW registration can be formulated in two ways, depending on involvement of the explicit image similarity measure.

3.4.1 Sum of Squared Differences (SSD) Approach

Beyond the binary form of the X-ray and DRR images, this approach differs from the intensity-based registration by involvement of the SSD image similarity measure instead of the NMI metric:

$$\text{SSD(X-ray, DRR)} = \sum_{x,y} \big(\text{X-ray}(x,y) - \text{DRR}(x,y)\big)^2 \tag{13}$$

where the x, y are the pixel coordinates. The usage of the SSD measure is advantageous for its straightforward parallelization and consequent acceleration using the `OpenGL` fragment shaders.

3.4.2 Pixel Differences (PD) Approach

In pixel differences approach the similarity between the X-ray and DRR images is evaluated directly by the Levenberg-Marquardt method. The vector F is reformulated to contain all pixels from both AP and LAT rendered binary masks, F_{max} contains pixels from the original radiograph masks:

$$F = \begin{pmatrix} \text{DRR}_{AP}\ (1,1) \\ \vdots \\ \text{DRR}_{AP}\ (w_{AP}, h_{AP}) \\ \\ \text{DRR}_{LAT}(1,1) \\ \vdots \\ \text{DRR}_{LAT}(w_{LAT}, h_{LAT}) \end{pmatrix}^T , \quad F_{max} = \begin{pmatrix} \text{X-ray}_{AP}\ (1,1) \\ \vdots \\ \text{X-ray}_{AP}\ (w_{AP}, h_{AP}) \\ \\ \text{X-ray}_{LAT}(1,1) \\ \vdots \\ \text{X-ray}_{LAT}(w_{LAT}, h_{LAT}) \end{pmatrix}^T \tag{14}$$

The Jacobian matrix is reformulated in a straightforward way. Instead of the two gradient vectors for AP and LAT view, the J_F is formed by gradient vectors for each pixel of the rendered binary masks. Despite the size of the approximated J_F matrix is changed to $w_{AP}h_{AP} + w_{LAT}h_{LAT} \times count(\beta)$, the number of required rendered images remains $2 * 2 * count(\beta)$ and no evaluation of an explicit similarity measure is performed.

Instead of the global similarity between the DRR and X-ray images, this formulation enables the optimisation to focus on the similarity of the local bone features. As more precise similarity information is available for the optimisation, the convergence of the Levenberg-Marquardt method is enhanced and consequently, the higher accuracy of the deformable registration is achieved.

3.5 Optimisation Scheme

For all the methods, the process of reconstruction consists of three subsequent optimisations.

In the first stage, the rigid 2D-3D registration of the mean shaped bone to the original X-ray images is performed. As the shape variability is Gaussian-distributed and the mean shape is *a priori* the most probable to occur, the principal components **b** are set to zero values. This step is involved to avoid getting stucked local minima. The vector β is reduced only to the rotation and translation parameters:

$$\beta_1 = ([r_x, r_y, r_z], [t_x, t_y, t_z]) \tag{15}$$

The second stage performs reduced deformable registration. As the time consumption caused by the J_F matrix evaluation linearly depends on the number of optimised parameters, the main purpose of the stage is to speed-up the registration. The pose is optimised simultaneously with a reduced subset of the first $u < n_b$ shape parameters:

$$\beta_2 = ([r_x, r_y, r_z], [t_x, t_y, t_z], [b_1, b_2, \ldots, b_u]) \tag{16}$$

To reach the maximal accuracy, the pose and all the shape parameters are optimised in the last stage:

$$\beta_3 = ([r_x, r_y, r_z], [t_x, t_y, t_z], \mathbf{b}) \tag{17}$$

In case of the intensity-based registration, the bone densities are set to the mean values during the whole registration.

4. RESULTS

The proposed methods have been evaluated on the set of virtual radiographs ray-casted from segmented CT images of femora. From each CT image, 12 virtual X-rays were rendered, rotated around longitudinal axis for $0, 30, 60, \ldots, 330°$, resulting in the data set of 96 images in total. The virtual radiographs were cropped to the average size of 205×477 pixels. The initial poses of the shape model were generated randomly with a uniform distribution. According to Baka,[14] the maximum difference between the initial and the ground-truth pose was limited to 10° rotation and 10 mm translation in each direction and along each axis.

Beyond the synthetic data set, the methods were evaluated using real X-ray images of phantom bones. The images were taken from AP and LAT view, their relative pose was recovered using a custom L-shaped marker. Sample phantom images are shown in Figure 5.

The optimisation scheme detailed in Section 3.5 was involved for each performed reconstruction. The subset of the first 5 principal components ($u = 5$, see Equation 16) was optimised in the second stage of the registration process. In case of intensity-based method, the joint histogram of size 64×64 bins is used for the NMI similarity measure evaluation.

4.1 Accuracy evaluation

To evaluate the registration accuracy, we measured the mean and maximum symmetric Hausdorff distance between the surfaces[24] of the ground-truth and the reconstructed bone model. We employed both leave-one-out and leave-all-in methodologies.

We investigated the influence of the parameter ϵ_b (see Equation 12) on the registration accuracy, the results are shown in Figure 3. The best accuracy was achieved when the value of the ϵ_b parameter was set between 1 and 1.5 standard deviation. For the rest of experiments, we set $\epsilon_b = 1\sigma$.

During the leave-one-out methodology, the bone model of currently used X-ray images was always discarded from the training data set of the used shape model. The shape models used for the evaluation were described by 18 principal components. Average results for the proposed methods are shown in Table 1.

During the leave-all-in methodology, the ground-truth tetrahedral model of the reconstructed bone was present in the PDM training set. For the leave-all-in evaluation, a shape model described by 20 components was used. This methodology is involved to reveal the reconstruction accuracy without the influence of PDM generality. Results of the leave-all-in methodology are shown in Table 2.
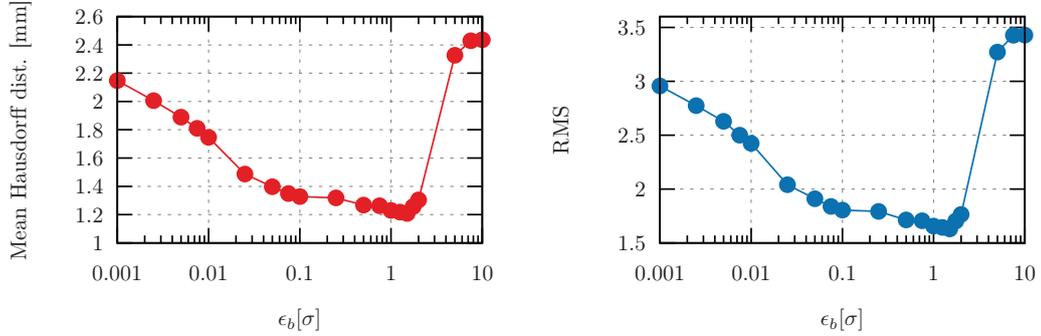
Figure 3. The influence of the ϵ_b parameter on the BW SSD method accuracy. The dependency of the average mean Hausdorff distance (left) and RMS (right) on the ϵ_b. The evaluation was performed using the leave-one-out methodology.

Method type	Mean [mm]	RMS	Maximum [mm]	r_x [°]	r_y [°]	r_z [°]	T [mm]
Intensity-based (3.3)	1.18	1.57	7.21	0.17	0.29	2.43	0.63
BW SSD (3.4.1)	1.23	1.65	7.69	0.25	0.33	2.84	0.78
BW PD (3.4.2)	1.02	1.35	7.10	0.16	0.18	1.55	0.55

Table 1. Average reconstruction accuracy reached using the leave-one-out methodology, described by mean and maximal symmetric Hausdorff distance, rotation and translation error. Corresponding performance results are shown in Table 4.

Method type	Mean [mm]	RMS	Maximum [mm]	r_x [°]	r_y [°]	r_z [°]	T [mm]
Intensity-based (3.3)	0.71	0.94	3.99	0.11	0.22	1.83	0.48
BW SSD (3.4.1)	0.77	1.02	4.48	0.18	0.18	2.53	0.57
BW PD (3.4.2)	0.43	0.51	1.78	0.09	0.09	1.04	0.36

Table 2. Accuracy of the methods reached using the leave-all-in methodology.

The results reveal that the best accuracy was reached by the BW PD method, while the accuracy of the BW SSD method was slightly worse than the results reached by the intensity-based method. The accuracy of the DW PD method is detailed using the graphs of the cumulative distributions of the Hausdorff distances in Figure 4. As the results show, the rotation around the longitudinal axis is difficult to recover accurately for all of the proposed methods.
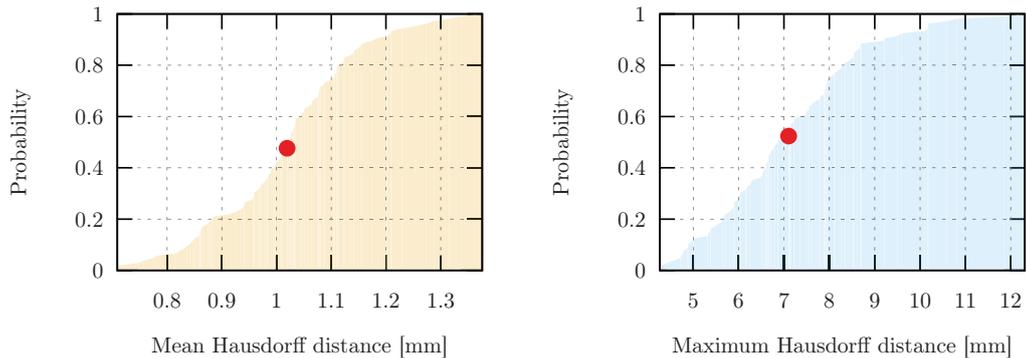

Figure 4. Detailed accuracy of the BW PD method. Cumulative distribution of the mean symmetric Hausdorff distance (left), cumulative distribution of the maximum error (right). The average case is highlighted by the red point.
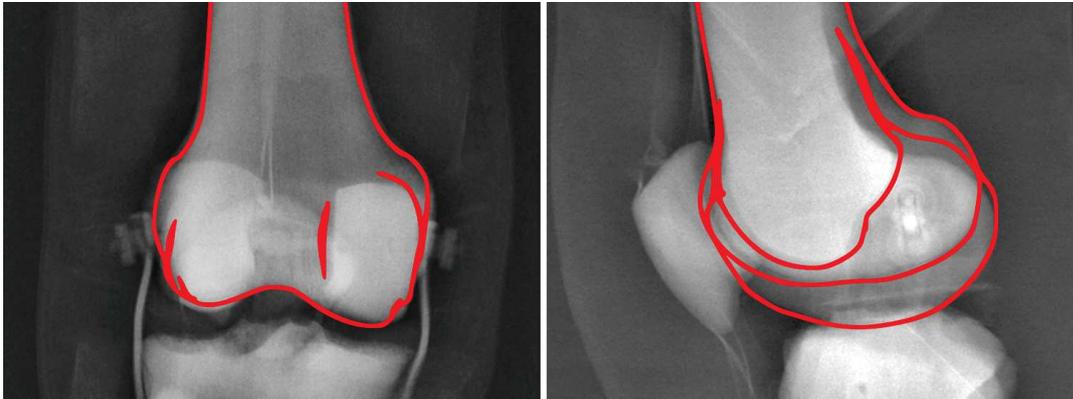
Figure 5. Silhouettes of the shape model fitted to the original X-ray images of a phantom. Radiographs are capturing distal femur from anterior-posterior (left) and lateral (right) view. The corresponding model of the reconstructed bone is shown in Figure 6.
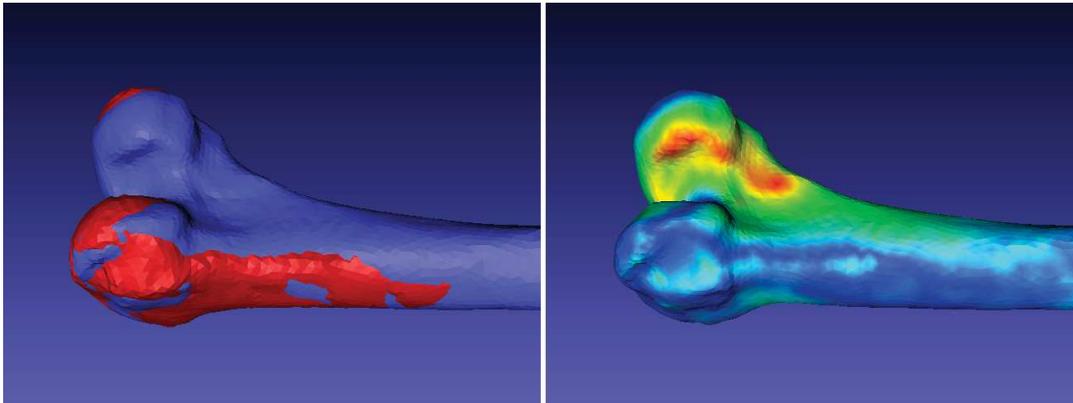

Figure 6. A comparison between the reconstructed and ground-truth surfaces of the phantom bone. The red colored ground-truth bone aligned to the reconstructed, blue colored bone (left). Heat-map visualization of the symmetric Hausdorff distance (right); the blue color highlights the most accurately reconstructed parts (0 mm difference), the red color shows the places with the highest registration error (6.62 mm). The original X-ray images are shown in Figure 5.

The distal femur of the experimental phantom was reconstructed using the BW PD (3.4.2) approach. The mean Hausdorff distance between the reconstructed and the ground-truth bone surface was 1.26 ± 1.66 mm, the maximal distance was 6.62 mm. The distance between the surfaces is visualized in Figure 6. The silhouettes of the reconstructed model back-projected to the original X-ray images are shown in Figure 5.

4.2 Speed performance evaluation

The speed performance of the proposed methods was evaluated on a desktop machine equipped with the NVidia 980 GTX Ti 6GB graphics card, Intel i5-4460 CPU and 24GB of memory. The evaluation was focused on the number of iterations, rendered images and amount of time needed for the reconstruction.

The rendering of one DRR image and one binary mask took 0.94 ms and 0.34 ms on average respectively. We also implemented the image similarity measures computation using the OpenGL fragment programs. The OpenGL implementation was chosen to minimize the data transfers between the CPU and GPU memory and
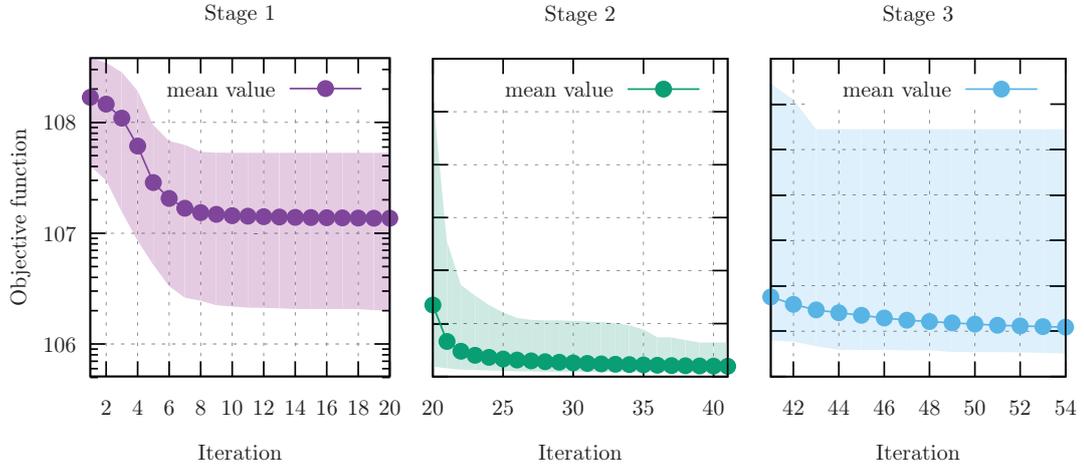
Figure 7. Convergence of the objective function (Equation 9) in particular stages of the optimisation: the rigid and reduced deformable registration stages (left), full deformable registration (right). The vertical axis are in logarithmic scale.

to increase the performance by the parallel computation of the similarity measures. The results of the OpenGL acceleration of the similarity metrics are shown in Table 3.

Measure type	CPU time [ms]	GPU time [ms]
NMI	2.51	1.10
SSD	2.03	0.23

Table 3. Average time needed for a single similarity measure evaluation using the CPU and OpenGL implementations. The measurements do not include the time needed to transfer the data between the GPU and CPU memory.

The acceleration led to more than twice faster evaluation of the NMI measure and more than eight times faster computation of the SSD metric. For the whole pipeline evaluation, we used the CPU implementation of the Levenberg-Marquardt method provided by the `dlib` library.[25] The speed performance results for the leave-one-out methodology are shown in Table 4.

Method type	Iterations			Images			Total time	
	st. 1	st. 2	st. 3	st. 1	st. 2	st. 3	CPU [s]	GPU [s]
Intensity-based	23.45	22.11	16.44	668.72	1079.07	1736.08	15.77	**8.76**
BW SSD	19.56	21.35	14.04	541.09	1041.39	1489.07	6.24	**3.19**
BW PD	34.89	16.13	13.79	934.42	794.99	1459.71	14.46	-

Table 4. Comparison of an average count of iterations, images and time required by the proposed methods. The total registration time was measured using CPU and GPU implementations of the similarity measures, including the time needed for the data transfers between the system memory and a video adapter. The DRR and binary mask rendering is always performed using GPU. The total GPU time for BW PD method is not available, as the method does not involve an explicit similarity measure and the CPU implementation of Levenberg-Marquardt method is used. Corresponding accuracy results are shown in Table 1. The average convergence graphs of the BW SSD method are shown in Figure 7.

The best speed-up was achieved using the BW SSD method. The OpenGL acceleration of the similarity measures computation led to the nearly double speed-up of the evaluated methods. The pixel differences approach faced worse performance due to the non-accelerated implementation of the Levenberg-Marquardt solver, as the

matrices F,F_{max} and J_F have significantly larger dimensions in comparison to the intensity-based and BW SSD methods (see Section 3.4.2).

5. CONCLUSIONS

We have proposed novel methods for the multiview 2D/3D reconstruction of the femoral bone. The formulation of the registration as a non-linear least squares problem and consequent optimisation using the well-established Levenberg-Marquardt algorithm leads to significantly faster and more reliable 2D/3D reconstruction method that requires a lower number of iterations to converge. Ehlke[3] reported that, for a pelvic bone reconstruction using his gradient-descent based optimisation aproach, approximately 6000 digitally reconstructed radiographs were rendered on average during the registration and the whole reconstruction took 1:41 minutes on average. In comparison, approximately 3500 digitally reconstructed radiographs were needed to reconstruct the femoral bone using our intensity-based approach on average, which results in reconstruction speed-up. The intensity-based registration took 8.76 seconds on average. It should be said that this comparison is rather orientative because of the different involved graphics hardware, number of X-ray images, size of the shape model, *et cetera*.

Due to its straightforward parallelization, the BW SSD method reaches the highest registration speed-up, as the average time of the reconstruction is less than 4 seconds. On the other hand, this method faces less registration accuracy caused by the global character of the SSD similarity measure which averages differences among the whole image. In the BW PD method, we overcome this drawback by the reformulation of the measure and the optimization focusing on similarities of local bone features. Here, the Levenberg-Marquardt method updates the optimised parameters with respect to derivatives of individual pixels of the X-ray and DRR binary masks instead of derivatives of the global similarity measures. This leads to very accurate registration method, the accuracy reached 1.02 ± 1.35 mm, so the BW PD method outperforms most methods that can be found in a brief summary of the state of the art methods presented by Baka.[14] Both the BW methods require precise X-ray images segmentation. Thanks to the presence of intensity information, the intensity-based method has the potential to work with original not segmented radiographs and at the same time to reach better accuracy than the BW SSD method.

Moreover, the least squares formulation of the registration allows straightforward extension to the multifragment 2D/3D reconstruction. Our future work will be focused on the simultaneous 2D/3D reconstruction and 3D reduction of a fractured femoral bone. The rendering part and the similarity measures part is distributed as open-source software and can be found at https://github.com/klepo/ssimrenderer.

ACKNOWLEDGMENTS

REFERENCES

[1] Wall, B. F. and Hart, D., "Revised radiation doses for typical X-ray examinations. report on a recent review of doses to patients from medical X-ray examinations in the uk by nrpb. national radiological protection board.," *The British Journal of Radiology* **70**, 437–439 (1997).

[2] Kelley, C. T., [*Iterative methods for optimization*], Frontiers in applied mathematics, SIAM, Philadelphia (1999).

[3] Ehlke, M., Ramm, H., Lamecker, H., Hege, H.-C., and Zachow, S., "Fast generation of virtual X-ray images for reconstruction of 3D anatomy," *IEEE Transactions on Visualization and Computer Graphics* **19**, 2673–2682 (2013).

[4] Yao, J. and Taylor, R. H., "Construction and simplification of bone density models," in [*Medical Imaging: Image Processing*], *Proc. SPIE* **4322**, 814–823, SPIE-The International Society for Optical Engineering (2001).

[5] Sadowsky, O., Cohen, J. D., and Taylor, J. H., "Rendering tetrahedral meshes with higher-order attenuation functions for digital radiograph reconstruction," in [*16th Visualization Conference*], 39–46, IEEE, Minneapolis (2005).

[6] Shirley, P. and Tuchman, A., "A polygonal approximation to direct scalar volume rendering," in [*Proceedings of the 1990 Workshop on Volume Visualization*], *VVS '90*, 63–70, ACM, New York (1990).

[7] Sadowsky, O., Cohen, J. D., and Taylor, J. H., "Projected tetrahedra revisited: A barycentric formulation applied to digital radiograph reconstruction using higher-order attenuation functions.," *IEEE Trans. Vis. Comput. Graph.* **12**, 461–473 (2006).

[8] Sadowsky, O., Chintalapani, G., and Taylor, R. H., "Deformable 2D-3D registration of the pelvis with a limited field of view, using shape statistics," in [*MICCAI*], *Lecture Notes in Computer Science* **4792**, 519–526, Springer Berlin Heidelberg (2007).

[9] Gong, R. H., Stewart, J., and Abolmaesumi, P., "Reduction of multi-fragment fractures of the distal radius using atlas-based 2D/3D registration," in [*Medical Imaging: Visualization, Image-Guided Procedures, and Modeling*], *Proc. SPIE* **7261**, SPIE-The International Society for Optical Engineering (2009).

[10] Lamecker, H., Wenckebach, T., and Hege, H.-C., "Atlas-based 3D-shape reconstruction from X-ray images," in [*18th International Conference on Pattern Recognition*], **1**, 371–374, ICPR (2006).

[11] Hurvitz, A. and Joskowicz, L., "Registration of a CT-like atlas to fluoroscopic X-ray images using intensity correspondences," *International Journal of Computer Assisted Radiology and Surgery* **3**, 493–504 (2008).

[12] Tang, T. S. and Ellis, R. E., "2D/3D deformable registration using a hybrid atlas," in [*MICCAI*], *Lecture Notes in Computer Science* **3750**, 223–230, Springer Berlin Heidelberg (2005).

[13] Zheng, G., Gollmer, S., Schumann, S., Dong, X., Feilkas, T., and Ballester, T. A. G., "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," *Medical Image Analysis* **13**, 883 – 899 (2009).

[14] Baka, N., Kaptein, B., de Bruijne, M., van Walsum, T., Giphart, J., Niessen, W., and Lelieveldt, B., "2D-3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models," *Medical Image Analysis* **15**, 840 – 850 (2011).

[15] Yao, J. and Taylor, R., "Assessing accuracy factors in deformable 2D/3D medical image registration using a statistical pelvis model," in [*Ninth IEEE International Conference on Computer Vision*], **2**, 1329–1334 (2003).

[16] Dworzak, J., Lamecker, H., von Berg, J., Klinder, T., Lorenz, C., Kainmuller, D., Seim, H., Hege, H.-C., and Zachow, S., "3D reconstruction of the human rib cage from 2D projection images using a statistical shape model.," *International Journal Computer Assisted Radiology and Surgery* **5**(2), 111–124 (2010).

[17] Hansen, N., "The CMA evolution strategy: A comparing review," in [*Towards a New Evolutionary Computation*], *Studies in Fuzziness and Soft Computing* **192**, 75–102, Springer Berlin Heidelberg (2006).

[18] Cootes, T., Cooper, D., Taylor, C., and Graham, J., "A trainable method of parametric shape description," *Image and Vision Computing* , 289–294 (1992).

[19] Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., and Buchler, P., "The virtual skeleton database: An open access repository for biomedical research and collaboration," *Journal of Medical Internet Research* (2013).

[20] Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J. P. W., "Elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging* **29**(1), 196–205 (2010).

[21] Bishop, C. M., [*Pattern Recognition and Machine Learning (Information Science and Statistics)*], Springer-Verlag New York, Inc., Secaucus, NJ, USA (2006).

[22] Si, H., "Tetgen, a Delaunay-based quality tetrahedral mesh generator," *ACM Trans. Math. Softw.* **41**, 11:1–11:36 (2015).

[23] Jan, J., [*Medical Image Processing, Reconstruction and Restoration - Concepts and Methods*], CRC Press, Taylor and Francis Group, Boca Raton, FL, USA (2006).

[24] Aspert, N., Santa-Cruz, D., and Ebrahimi, T., "Mesh: measuring errors between surfaces using the Hausdorff distance," in [*Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*], **1**, 705–708 (2002).

[25] King, D. E., "Dlib-ml: A machine learning toolkit," *Journal of Machine Learning Research* **10**, 1755–1758 (2009).

A Study on Performace of Levenberg-Marquardt and CMA-ES Optimization Methods for Atlas-based 2D/3D Reconstruction

Ondrej Klima [*], **Adam Chromy** [**], **Pavel Zemcik** [*],
Michal Spanel [***], **Petr Kleparnik** [*]

[*] *Department of Computer Graphics and Multimedia, Brno University
of Technology, Božetěchova 1/2, 612 66 Brno, Czech Republic
(e-mail: iklima@fit.vutbr.cz)*
[**] *Department of Control and Instrumentation, Brno University of
Technology, Technická 3082/12, 616 00 Brno, Czech Republic*
[***] *3Dim Laboratory s.r.o., Kamenice 34, 625 00 Brno, Czech Republic*

Abstract: In this study, we compare the performance of our previously proposed deformable 2D/3D registration approach based on the Levenberg-Marquardt optimization with methods exploiting Covariance Matrix Adaptation (CMA) and Covariance Matrix Self Adaptation (CMSA) evolution strategies. The aim of the registration is to reconstruct a patient-specific 3D bone model from a small set of plain 2D X-ray images what is achieved by fitting a deformable bone atlas onto the X-ray images. The comparison of different optimization methods is focused on both the robustness and the speed. The results were obtained using a large-scale data set of synthetic X-ray images. We show that our method is several times faster in comparison with the approaches based on evolution strategies while the robustness of the reconstruction is preserved. To speed-up the reconstruction process, certain parts of the registration pipeline are accelerated using graphics hardware. The median error of our proposed method was 1.12 mm and the median reconstruction time was 7.2 s. The median time reached by the CMA-ES and CMSA-ES methods was 48.5 s and 138.5 s respectively.

© 2016, IFAC (International Federation of Automatic Control) Hosting by Elsevier Ltd. All rights reserved.

Keywords: 2D/3D reconstruction, Levenberg-Marquardt optimization, CMA-ES, CMA-ES, Statistical Shape and Intensity Model

1. INTRODUCTION

An identification of the best fitting bone plate for a specific patient is a one of common tasks in the field of orthopedic surgery. The plate shape identification can be performed during a preoperative planning stage. The planning process typically exploits a 3D model of the injured bone which is usually derived from a Computed Tomography (CT) scan of the patient. In recent years, the possibilities of a preoperative planning based only on plain 2D X-ray images have been brought to focus, as the plain X-ray imaging exposes patients to lower doses of ionizing radiation and it is less expensive in comparison with the CT examination. The key moment of such planning is a reconstruction of the 3D bone model from a small count of 2D X-ray images. This reconstruction is achieved by a non-rigid 2D/3D registration of a deformable 3D femoral atlas onto the images. The 2D/3D registration is solved as a numerical optimization. The accuracy of the reconstruction depends on the choice of the optimization method. The registration speed is crucial, especially in urgent traumatology cases, and is closely related to the convergence rate of the chosen numerical optimization approach.

In Klima et al. (2015, 2016), we have proposed a 2D/3D registration method based on the Levenberg-Marquardt algorithm which is a highly effective non-linear least squares problem solver, see Kelley (1999). In this study, we compare the performance of our previously proposed method against approaches based on Covariance Matrix Adaptation Evolution Strategy (CMA-ES), see Hansen and Kern (2004), which is frequently used for the purpose of 2D/3D registration, and its Covariance Matrix Self Adaptation (CMSA-ES) variant proposed by Beyer and Sendhoff (2008). The results reveal that our Levenberg-Marquadt based method is several times faster than the methods using the evolution strategies while reaching the same reconstruction accuracy.

2. RELATED WORK

Most of previously published 2D/3D registration approaches can be classified as feature-based or intensity-based methods. Feature-based registration methods rely on bone edges, or silhouettes detected in the original X-ray images. The pre-trained bone atlas is usually a deformable Statistical Shape Model (SSM) which is capable to learn bone surface variations within a population. Consequently, such feature-based methods reconstruct only a shape of

121

a bone. In contrast, intensity-based methods proposed by Yao and Taylor (2001); Sadowsky et al. (2007); Gong et al. (2009); Ehlke et al. (2013) work with original X-ray images and use Statistical Shape and Intensity Model (SSIM). In addition to the bone shape, appearance models capture the bone densities. Hence the anatomical structures such as a compact and spongy bone can be reconstructed as well. In this study, we focus on the intensity-based methods.

Previous 2D/3D registration approaches use various optimization methods, i. e. modifications of the gradient-descent algorithm Ehlke et al. (2013). Strong attention is currently paid to the gradient-free CMA-ES method involved by Gong and Abolmaesumi (2008); Gong et al. (2009, 2011); Khallaghi et al. (2010); Otake et al. (2016); Ketcha et al. (2016); De Silva et al. (2016). The main benefit of the CMA-ES method is no requirement of the Jacobian matrix numerical approximation during the registration.

Many previously published comparisons of various atlas-based 2D/3D registration methods were only tentative. The main reason was the evaluation data sets were not shared between the authors and typically consisted only of a small number of cases. Moreover, the evaluations were performed on different hardware configurations and the experimental implementations were not equally sophisticated. The *main contribution* of this paper is a fair comparison of the mentioned methods on a reasonably large evaluation data set.

3. METHOD

Approaches evaluated in this study reconstruct bones from two co-registered orthogonal X-ray images. The background of the bone must be segmented out from the original X-ray images and a rough initial estimate of the appearance model pose must be provided interactively by a user. The rest of the registration pipeline is built as an iterative process. Virtual X-ray images, refered as Digitally Reconstructed Radiographs (DRRs), are rendered from the appearance model in each step of the registration using our OpenGL-based GPU implementation, proposed in Klima et al. (2016). The dissimilarities between original and virtual X-ray images are evaluated using Normalized Mutual Information (NMI) image similarity measure, detailed in Jan (2006). Due to performance reasons, evaluation of the NMI similarity metric is accelerated using GPU as well. The pose and the shape parameters of the appearance model are adjusted in each iteration until the dissimilarities are minimized. Finally, the bone is reconstructed as a specific instance of the 3D appearance model.

3.1 Statistical Shape and Intensity Model (SSIM)

We adopted SSIM appearance model based on volumetric meshes proposed by Yao and Taylor (2001). Tetrahedral meshes of femoral bones derived from CT scans were brought into correspondence using `Elastix` toolbox created by Klein et al. (2010). Registered meshes were aligned using Generalized Procrustes Analysis (GPA), while the size of models remained unchanged. A linear model describing the shape variability of femoral bones was obtained after applying the Probabilistic Principal Component Analysis (PPCA):

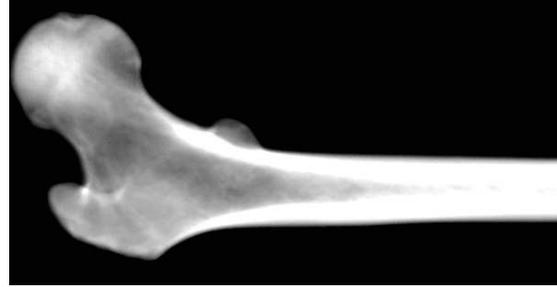$$S = \phi\mathbf{b} + \overline{S} + \epsilon \tag{1}$$

Figure 1. A digitally reconstructed radiograph of a proximal femur rendered from the involved SSIM model.

where S is a vector formed by concatenated coordinates of tetrahedral model vertices, $\overline{S}$ is a mean shape vector, ϕ is a matrix of principal components, $\mathbf{b}$ is a vector of independent shape parameters and ϵ is a zero-mean Gaussian noise. The bone density is described in each tetrahedron using Bernstein polynomials:

$$D(\mu) = \sum_{\forall i,j,k,l \in i+j+k+l=n} C_{i,j,k,l} B_{i,j,k,l}^n(\mu) \tag{2}$$

where μ are barycentric coordinates of a point inside the tetrahedron, $D(\mu)$ is a bone density at that point, $B_{i,j,k,l}^n$ is a Bernstein basis for n-th degree polynomial and $C_{i,j,k,l}$ is the corresponding coefficient. For more details about the SSIM models and their training we refer the reader to works published previously by Yao and Taylor (2001); Klima et al. (2016).

For the rendering of DRR images from the SSIM model, the `OpenGL` accelerated approach originally proposed by Ehlke et al. (2013) has been adopted. The integrals of Bernstein polynomials along the virtual rays intersecting the appearance model are evaluated using a closed form solution proposed by Sadowsky Sadowsky et al. (2007). A sample virtual X-ray rendered from the SSIM model is illustrated in Figure 1.

3.2 2D/3D Registration

We formulate the double-view 2D/3D registration as a non-linear least squares problem:

$$x^* = \arg\min_x \frac{1}{2} F(x)^T F(x) \tag{3}$$

where $x = (R, T, \mathbf{b})$ is a vector containing rotation, translation and shape parameters of the appearance model respectively and $F(x)$ is a column vector of residuals to be minimized defined as follows:

$$F(x) = \begin{bmatrix} \text{NMI}_1(x) - 2 \\ \text{NMI}_2(x) - 2 \end{bmatrix} \tag{4}$$

where $\text{NMI}_{\{1,2\}}(x)$ is a similarity between the original X-ray image and the DRR image rendered according to the parameters x, evaluated for the views 1 and 2 respectively. If the corresponding X-ray and DRR images are the same, the value of NMI similarity measure is equal to 2.

The speed of the reconstruction is determined by the total number of images rendered during the optimization, as the rendering and the similarity measure evaluation are the most time-demanding parts of the registration pipeline.

The total number of images depends on the convergence of the chosen optimization method and on the amount of virtual X-rays rendered in a single iteration.

The Levenberg-Marquardt method requires evaluation of the Jacobian matrix J_F. Size of the matrix is given as $N_{views} \times N_x$, where N_{views} is the number of X-ray views and N_x is the number of optimized parameters. As a closed formula solution of the J_F does not exists, the central differences approximation of the matrix is typically used. In consequence, two images are rendered for each J_F element. Because the study is focused on the two-view reconstruction, the count of images N_{LM} rendered in each iteration is approximately given by:

$$N_{\mathrm{LM}} \approx 4N_x \tag{5}$$

In our experiments, we use implementation of the Levenberg-Marquardt algorithm provided by the `dlib` library, proposed in King (2009).

The CMA-ES and CMSA-ES methods require to set parameters describing population size λ, parents count μ and an initial step size σ. We use the estimations $\lambda_{\mathrm{CMA}} = \max\left[5, \min(N_x, 4 + 3\lfloor \log N_x \rfloor)\right]$, $\sigma_{\mathrm{CMA}} = N_x^{-\frac{1}{2}}$ for the CMA-ES method and $\lambda_{\mathrm{CMSA}} = 4N_x$, $\sigma_{\mathrm{CMSA}} = 1$ for the CMSA-ES method. The $\mu = \left\lfloor \frac{\lambda}{2} \right\rfloor$ is same for both the methods. Consequently, the number of images rendered during a single iteration N_{CMA} and N_{CMSA} is equal to:

$$N_{\mathrm{CMA}} \approx 2\lambda_{\mathrm{CMA}} \tag{6}$$

$$N_{\mathrm{CMSA}} \approx 8N_x \tag{7}$$

The reference implementations of the CMA-ES and CMSA-ES methods from the `Shark` library by Igel et al. (2008) were adopted in this study. The optimizations are terminated when the values of the objective function in two subsequent iterations are equal. Comparison of the number of images is shown in Figure 2.
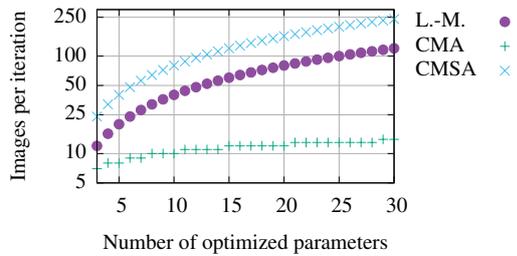
Figure 2. A growth of the number of images rendered in a single iteration with respect to the number of optimized parameters. The vertical axis is in a logarithmic scale.

3.3 Optimization Scheme

The whole reconstruction process comprises of 3 subsequent registrations:

(1) *Rigid registration* - only the pose of the mean shaped SSIM model is being optimized.
(2) *Reduced deformable registration* - the pose of the appearance model is optimized together with a subset of the first shape parameters.
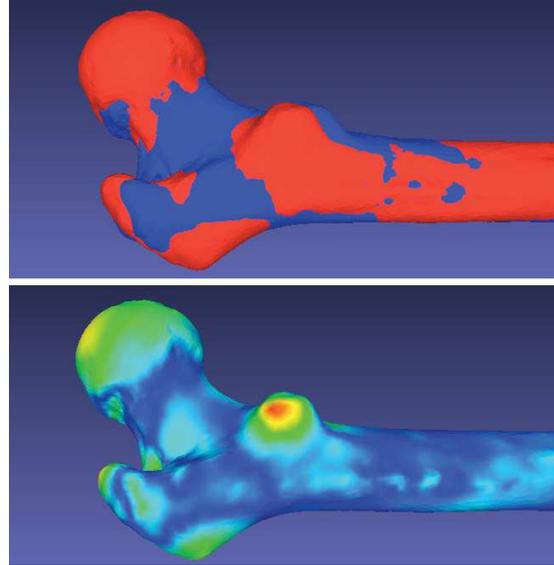
Figure 3. A proximal part of a femoral bone reconstructed with mean error of 1.12 mm. The reconstructed bone highlighted by a red color is aligned to the ground-truth bone model obtained from CT image (top). The reconstruction error is visualized by a heatmap, the red color shows places with the highest error of 5.68 mm (bottom).

(3) *Full deformable registration* - all shape parameters are optimized together with the pose.

The first and the second stages are involved to prevent the registration from getting stuck in a local minima and to speed-up the process by reducing the amount of simultaneously optimized parameters. The last stage is performed to reconstruct fine details of the bone. During the registration, the bone densities are set to mean values.

4. EXPERIMENTAL RESULTS

The evaluation was focused on the accuracy and time consumption of the registration procedure. We involved leave-one-out and leave-all-in evaluation methodologies. During the leave-all-in evaluation, the reconstructed bone was present in the training set of the appearance model and consequently, the accuracy of the registration was not affected by the appearance model generalization ability. In contrast, during the leave-one-out methodology, the reconstructed bone was discarded from the SSIM trainig set so the results reflect the real-world usage situation.

4.1 Evaluation Data Sets and Measurements

We used SSIM models created from 21/22 CT images obtained from `Virtual Skeleton Database`, proposed by Kistler et al. (2013), for leave-one-out/leave-all-in evaluations respectively. The reference tetrahedral mesh contained 26,000 vertices and 104,000 tetrahedra. The density was described using Bernstein polynomials of the 2nd degree,

resulting in 10 density coefficients per tetrahedron. The appearance models were described using 19/20 shape parameters for leave-one-out/leave-all-in methodologies respectively.

The methods were evaluated on a data set containing 100 orthogonal pairs of virtual X-ray images, ray-casted from CT images of 8 individuals. An average size of X-ray images was 201×474 pixels. The initial poses were generated randomly with a uniform distribution. The maximum error of initial pose was limited to $\pm 10 \, mm$ in translation and $\pm 10\,°$ in rotation.

The evaluations were performed on a desktop machine equipped with `NVidia GTX980Ti 6GB` video adapter, `Intel i5-4460` processor and 24GB of RAM.

As the evolution strategies are stochastic, we repeated CMA-ES and CMSA-ES reconstructions 10 times resulting in 1000 test cases for each evolution strategy. The reconstruction accuracy was evaluated using the mean symmetric Hausdorff distance, proposed by Aspert et al. (2002), measured between the reconstructed surface and the ground-truth bone models obtained from CT images. We considered the reconstruction successful when the RMS error between the surfaces was less than 3. Except two cases for the Levenberg-Marquardt optimization and three cases for CMSA-ES leave-all-in evaluation, all reconstructions

Table 1. Summary of time consumption in seconds.

Method	Min	Q1	Median	Q3	Max
Leave-one-out evaluation					
L.-M.	**3.54**	**5.84**	**7.15**	**8.88**	**15.49**
CMA-ES	14.61	39.95	48.51	61.07	137.59
CMSA-ES	38.06	111.54	138.47	172.01	331.90
Leave-all-in evaluation					
L.-M.	**4.11**	**7.79**	**8.92**	**10.74**	**16.49**
CMA-ES	13.04	42.23	51.42	66.05	133.33
CMSA-ES	32.72	119.39	145.14	179.05	344.07

Table 2. Avg. number of iterations per stage.

Method	*Leave-one-out*			*Leave-all-in*		
	St.1	St.2	St.3	St.1	St.2	St.3
L.-M.	**24**	**22**	**16**	**25**	**22**	**24**
CMA-ES	310	191	641	310	196	694
CMSA-ES	154	109	272	153	111	280

Table 3. Number of images rendered in each iteration and average total number of images for leave-one-out/leave-all-in evaluations.

Method	Images per iteration			Total images
	St. 1	St. 2	Stage 3	Stages 1 - 3
L.-M.	24	44	100 / 104	**1144 / 1452**
CMA-ES	**6**	**11**	**13 / 13**	8195 / 8700
CMSA-ES	48	88	200 / 208	23822 / 25107

were successful. The unsuccessful cases were discarded from the following plots.

In the first stage, only 6 parameters describing the appearance model pose were optimized. In the following stage, the pose parameters were optimized together with the first 5 shape parameters of the appearance model. During the full deformable registration, 25/26 parameters were optimized in total.

The measured number of iterations is shown in Table 2, number of images rendered per iteration and in total are shown in Table 3. The average time for one image rendering

Table 4. Summary of the mean Hausdorff distance distributions (values in millimeters).

Method	Min	Q1	Median	Q3	Max
Leave-one-out evaluation					
L.-M.	0.84	0.99	1.12	1.25	2.02
CMA-ES	0.81	0.98	1.09	1.20	1.72
CMSA-ES	0.81	0.99	1.09	1.21	1.76
Leave-all-in evaluation					
L.-M.	0.34	0.56	0.64	0.78	1.87
CMA-ES	0.28	0.47	0.56	0.63	1.08
CMSA-ES	0.22	0.50	0.59	0.69	1.48

Figure 4. Distributions of the reconstruction time. The vertical axis is in logaritmic scale.
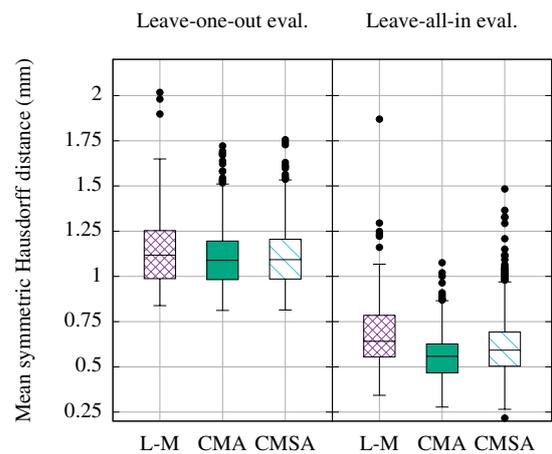

Figure 5. Distributions of the symmetric mean Hausdorff distance.

and for one metric evaluation was 0.94 ms and 1.10 ms respectively. The resulting reconstruction times are shown in Figure 7 and Table 1.

The corresponding results for reconstruction accuracy are shown in Figure 5 and summarized in Table 4. The case of Levenberg-Marquardt leave-one-out reconstruction, that reached a median accuracy, is visualized in Figure 3.

We also investigated a correlation of the results reached by the methods. The correlation of reconstruction accuracy is illustrated in Figure 6, the time-consumption correlation is shown in Figure 7. The corresponding coefficients are summarized in the Table 5.

Table 5. Correlation coefficients for leave-one-out/leave-all-in evaluations of reconstruction accuracy and time consumption.

Methods		Hausdorff d.	Time
L.-M.	CMA	0.5440 / 0.5002	0.3397 / 0.2764
L.-M.	CMSA	0.5828 / 0.5762	0.3463 / 0.2834
CMA	CMSA	0.8321 / 0.7324	0.4207 / 0.3491

4.2 Discussion on results

The results reveal that reconstruction based on the LM optimization is several times faster than CMA-ES and CMSA-ES based methods while reaching similar accuracy.

The median time of the registration using the LM method was more than 16-19 times lower in comparison to the
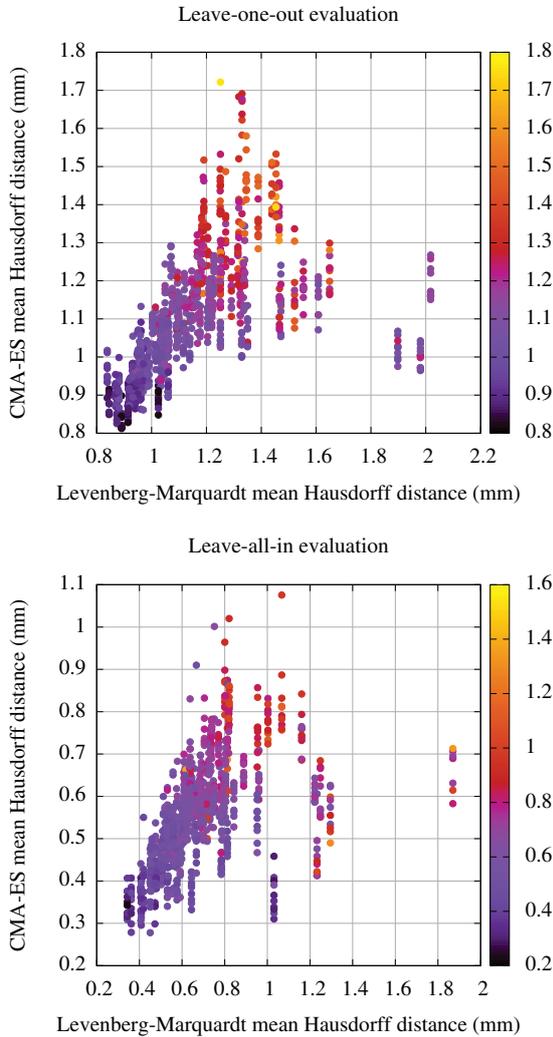

Figure 6. Correlation of the mean Hausdorff distance. The results of CMSA-ES are visualized using heatmap.
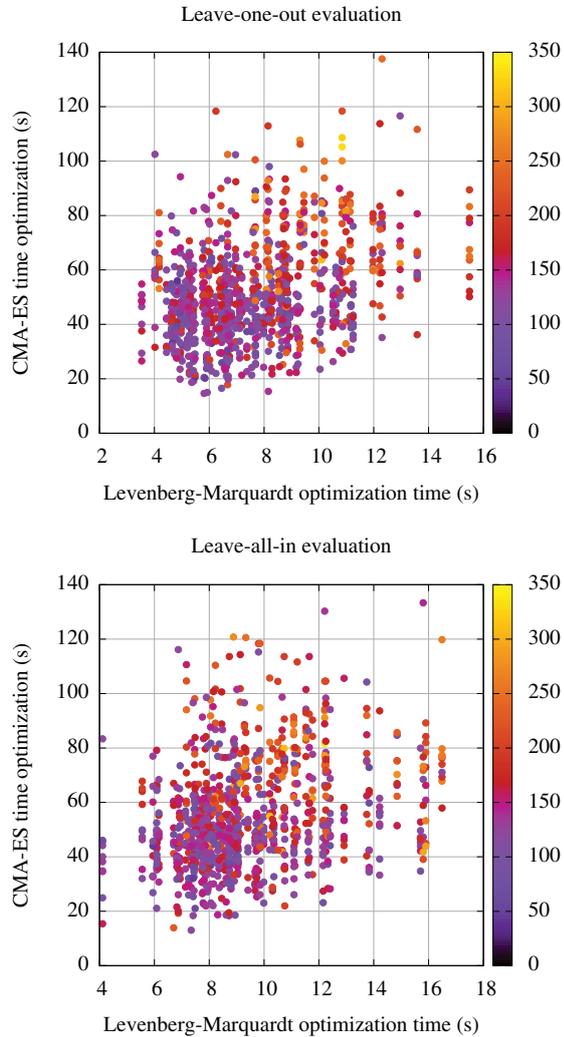

Figure 7. Correlation of the time consumptions. The heatmap shows CMSA-ES values.

Table 6. Speed-up factors between Levenberg-Marquardt and ES-based methods.

	Time		Images	
	CMA-ES	CMSA-ES	CMA-ES	CMSA-ES
Leave-one-out	6.78	19.37	7.16	20.82
Leave-all-in	5.76	16.27	5.99	17.29

CMSA-ES method and approximately 6-7 times lower in comparison to the CMA-ES method (see Table 6). Although the number of images rendered in a single iteration scales the best in case of CMA-ES method, the rate of convergence was significantly lower in comparison with the Levenberg-Marquardt algorithm. On average, the CMA-ES based registration required approximately 6-7 times more image renderings and metric evaluations to converge and consequently the registration time was proportionally longer. We also observed that the most iterations were spent on the refinement of fine details of the reconstruction.

The correlation was high in case of accuracy results and significantly lower in case of time consumption. This was caused by the stochastic nature of the evolution strategies, when the most while random amount of time was spent on a refinement of fine details, while the accuracy gain of fine details tuning was always low.

5. CONCLUSION

Although the CMA-ES optimization is more convenient for the usage in the 2D/3D registration, as the method does not require a non-trivial approximation of the Jacobian matrix, the Leveberg-Marquardt method provides several times faster registration and consequently is more suitable for the cases when the reconstruction time is crucial.

ACKNOWLEDGEMENTS

REFERENCES

Aspert, N., Santa-Cruz, D., and Ebrahimi, T. (2002). Mesh: measuring errors between surfaces using the hausdorff distance. In *Multimedia and Expo, 2002. ICME '02. Proceedings. 2002 IEEE International Conference on*, volume 1, 705–708.

Beyer, H.G. and Sendhoff, B. (2008). *Covariance Matrix Adaptation Revisited – The CMSA Evolution Strategy –*, 123–132. Berlin, Heidelberg. doi:10.1007/978-3-540-87700-4_13.

De Silva, T., Uneri, A., Ketcha, M., Reaungamornrat, S., Goerres, J., Vogt, S., Kleinszig, G., Wolinsky, J.P., and Siewerdsen, J. (2016). "levelcheck" localization of spinal vertebrae in intraoperative radiographs from preoperative mri. In *Proc. of CARS*, S54–S55. Springer.

Ehlke, M., Ramm, H., Lamecker, H., Hege, H.C., and Zachow, S. (2013). Fast generation of virtual x-ray images for reconstruction of 3d anatomy. *IEEE Transactions on Visualization and Computer Graphics*, 19, 2673–2682.

Gong, R.H. and Abolmaesumi, P. (2008). 2d/3d registration with the cma-es method. In *Medical Imaging*, 69181M–69181M. SPIE.

Gong, R.H., Stewart, J., and Abolmaesumi, P. (2009). Reduction of multi-fragment fractures of the distal radius using atlas-based 2d/3d registration. In *Medical Imaging: Visualization, Image-Guided Procedures, and Modeling*, volume 7261 of *Proc. SPIE*.

Gong, R.H., Stewart, J., and Abolmaesumi, P. (2011). Multiple-object 2-d–3-d registration for noninvasive pose identification of fracture fragments. *Biomedical Engineering, IEEE Transactions on*, 58(6), 1592–1601.

Hansen, N. and Kern, S. (2004). Evaluating the cma evolution strategy on multimodal test functions. *Parallel Problem Solving from Nature - PPSN VIII*, 282–291.

Igel, C., Heidrich-Meisner, V., and Glasmachers, T. (2008). Shark. *Journal of Machine Learning Research*, 9, 993–996.

Jan, J. (2006). *Medical Image Processing, Reconstruction and Restoration - Concepts and Methods*. CRC Press, Taylor and Francis Group, Boca Raton, FL, USA.

Kelley, C.T. (1999). *Iterative methods for optimization*. Frontiers in applied mathematics. SIAM, Philadelphia.

Ketcha, M.D., De Silva, T., Uneri, A., Kleinszig, G., Vogt, S., Wolinsky, J.P., and Siewerdsen, J.H. (2016). Automatic masking for robust 3d-2d image registration in image-guided spine surgery. doi:10.1117/12.2216913.

Khallaghi, S., Mousavi, P., Gong, R.H., Gill, S., Boisvert, J., Fichtinger, G., Pichora, D., Borschneck, D., and Abolmaesumi, P. (2010). Registration of a statistical shape model of the lumbar spine to 3d ultrasound images. In *MICCAI 2010*, 68–75.

King, D.E. (2009). Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10, 1755–1758.

Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., and Buchler, P. (2013). The virtual skeleton database: An open access repository for biomedical research and collaboration. *Journal of Medical Internet Research*.

Klein, S., Staring, M., Murphy, K., Viergever, M., and Pluim, J.P.W. (2010). Elastix: A toolbox for intensity-based medical image registration. *IEEE Transactions on Medical Imaging*, 29(1), 196–205.

Klima, O., Kleparnik, P., Spanel, M., and Zemcik, P. (2015). Gp-gpu accelerated intensity-based 2d/3d registration pipeline. Proceedings of Shape Symposium. Délemont: Swiss Institute for Computer Aided Surgery.

Klima, O., Kleparnik, P., Spanel, M., and Zemcik, P. (2016). Intensity-based femoral atlas 2d/3d registration using levenberg-marquardt optimisation. doi:10.1117/12.2216529.

Otake, Y., Esnault, M., Grupp, R., Kosugi, S., and Sato, Y. (2016). Robust patella motion tracking using intensity-based 2d-3d registration on dynamic bi-plane fluoroscopy: towards quantitative assessment in mpfl reconstruction surgery. doi:10.1117/12.2214699.

Sadowsky, O., Chintalapani, G., and Taylor, R.H. (2007). Deformable 2d-3d registration of the pelvis with a limited field of view, using shape statistics. In *MICCAI*, Lecture Notes in Computer Science, 519–526.

Yao, J. and Taylor, R.H. (2001). Construction and simplification of bone density models. In *Medical Imaging: Image Processing*, volume 4322 of *Proc. SPIE*, 814–823.

Virtual 2D-3D Fracture Reduction with Bone Length Recovery Using Statistical Shape Models

Ondřej Klíma[1]([✉]) [iD], Roman Madeja[2] [iD], Michal Španel[1,3] [iD], Martin Čuta[4] [iD], Pavel Zemčík[1] [iD], Pavel Stoklásek[5], and Aleš Mizera[5] [iD]

[1] IT4Innovations Centre of Excellence, Brno University of Technology,
Božetěchova 1/2, 612 66 Brno, Czech Republic
`{iklima,spanel,zemcik}@fit.vutbr.cz`
[2] University Hospital in Ostrava, 17. listopadu 1790, 708 52 Ostrava, Czech Republic
[3] 3Dim Laboratory s.r.o, Kamenice 34, 625 00 Brno, Czech Republic
[4] Laboratory of Morphology and Forensic Anthropology, Department of
Anthropology, Masaryk University, Kotlářská 2, 611 37 Brno, Czech Republic
[5] CEBIA-Tech, Faculty of Applied Informatics, Tomas Bata University in Zlín,
Nad Stráněmi 4511, 760 05 Zlín, Czech Republic
`http://www.3dim-laboratory.cz/`

Abstract. Computer-assisted 3D preoperative planning based on 2D stereo radiographs has been brought into focus recently in the field of orthopedic surgery. To enable planning, it is crucial to reconstruct a patient-specific 3D bone model from X-ray images. However, most of the existing studies deal only with uninjured bones, which limits their possible applications for planning. In this paper, we propose a method for the reconstruction of long bones with diaphyseal fractures from 2D radiographs of the individual fracture segments to 3D polygonal models of the intact bones. In comparison with previous studies, the main contribution is the ability to recover an accurate length of the target bone. The reconstruction is based on non-rigid 2D-3D registration of a single statistical shape model onto the radiographs of individual fragments, performed simultaneously with the virtual fracture reduction. The method was tested on a syntethic data set containing 96 virtual fractures and on real radiographs of dry cadaveric bones suffering peri-mortem injuries. The accuracy was evaluated using the Hausdorff distance between the reconstructed and ground-truth bone models. On the synthetic data set, the average surface error reached 1.48 ± 1.16 mm. The method was built into preoperative planning software designated for the selection of the best-fitting fixation material.

Keywords: Preoperative planning · Fracture reduction
Fixation devices · 2D-3D registration · Statistical shape model

1   Introduction

Plain radiography plays a key role in bone fracture diagnosis and treatment. In the case of surgical intervention, plain radiographs enable basic preoperative planning, such as bone fracture classification and the determination of an appropriate fixation technique for its stabilization. More advanced, computer-assisted planning of the osteosynthesis provides a virtual simulation of the intervention, which typically includes situating the fracture segments into anatomically correct and mechanically stable poses, measuring the bone morphology, or placing fixation devices [9]. The virtual simulations rely on 3D polygonal models of individual bone fragments, which are conventionally obtained from volumetric images provided by computed tomography (CT). However, during the CT examinations, the patients are exposed to substantially higher doses of radiation in comparison with plain radiography. Therefore, the indication of CT examinations is generally restricted only to cases of severe or complex fractures, while the treatment of rather common cases depends on plain radiographs. Nevertheless, computer-assisted planning can be still beneficial even for rather routine fractures, especially for long bone fractures of the lower limbs. One important contribution is the possibility of preoperative measurement of patient-specific bone morphology with aim of determining the features of the best-fitting fixation devices, such as the length of the intramedullary nail [8], the size of the bone plate, or the number and placement of bone screws. Therefore, a reconstruction of a 3D patient-specific anatomy based only on plain, clinically available radiographs instead of volumetric images is of great importance for the application of virtual planning in a broader spectrum of bone fracture treatment procedures.

In this paper, we propose a semi-automatic 3D virtual fracture reduction method, which is able to reconstruct a polygonal model of an intact bone from stereo radiographic images of the individual fracture segments. The method is focused on displaced diaphyseal fractures of the simple or wedge type.

2   Related Work

In the field of orthopedic surgery, a somewhat similar challenge of computer-assisted 3D preoperative planning based only on plain radiographs was recently addressed by several projects [1,2] focusing on total hip arthroplasty (THA), total knee arthroplasty (TKA), and lower extremity osteotomy. Other studies were focused on observing 3D joint kinematics from fluoroscopy sequences without the requirement of CT image acquisition [5,17,18]. Instead of CT scans, a non-rigid registration of 3D bone atlases onto the stereo radiographs was exploited to reconstruct polygonal models of the bones. As proposed in works such as [15], statistical shape models were involved as the atlases to perform a shape-constrained 2D-3D registration. With respect to the statistical shape models, this reconstruction approach is straightforward when the target bone is not suffering any injuries, which is fulfilled for the total joint arthroplasty or observation of joint kinematics. However, arbitrarily shaped fracture segments make reconstruction based on statistical shape models a challenging task.

The first attempt to reconstruct injured bones using statistical atlases was proposed in a study focused on the reduction of multi-fragment fractures of the distal radius [7]. The goal of the study was to obtain a polygonal model of an intact bone from plain radiographs of the fracture segments. The reconstruction, together with the fracture reduction, were achieved at the same time by a 2D-3D registration of a single statistical appearance model of an intact distal radius into individual fracture segments. Splitting the statistical appearance model into fracture segments was performed automatically by the registration. The method was evaluated *in silico* using simulated fractures, concluding that the atlas-based reconstruction may provide a more accurate distal radius template than the conventionally used mirrored model obtained from the contralateral limb.

A later study, using a similar principle of a multi-fragment 2D-3D registration of a statistical shape model, focused on diaphyseal fractures of the long bones of the lower limbs [16]. In contrast with the previous work, its aim was to determine the rotation alignment between the proximal and distal fragments along the longitudinal axis. In addition to the rotation angles, the study considered the reconstruction of surface models of the individual fracture segments. However, the approach was unable to perform virtual fracture reduction and to provide a model of the intact bone, as the method was unable to determine the correct length of the target bone. Moreover, the shape model had to be divided into fragments in advance, without further refinement during the registration process. The bone length also had to be provided manually in a study focused on automatic fracture reduction using statistical atlases, working with mesh models of fracture segments obtained from CT scans [3].

In this paper, we address the challenge of accurate bone length recovery. Unlike [16], the division of the statistical shape model into segments is performed automatically by the registration, enabling optimization of the shape model length. In consequence, the proposed method is able to perform virtual fracture reduction and provide a 3D model of the intact bone.

3 Method

The method is based on a multi-fragment registration of a statistical shape model into stereo radiographs of individual fracture segments, extended by simultaneous optimization of the shape model length.

3.1 Statistical Shape Models

The statistical shape models involved in this study work as elastic tetrahedral models of bones. As their elasticity is shape-constrained, it is ensured the models always represent anatomically reasonable bones. The shape models involved were created using a procedure detailed in [12]. As the models are based on probabilistic principal component analysis (PPCA), they are represented by the following generative model:

$$S = \phi \mathbf{b} + \overline{S} + \omega \tag{1}$$

Table 1. Characteristics of involved statistical shape models.

Statistical shape model	Size of training set	Modes of variation	Tetrahedral vertices	Tetrahedra
Femur	43 bones	41	20,843	93,480
Tibia	42 bones	40	22,003	106,436

$$b_1 = \pm 2\sigma \qquad\qquad b_2 = \pm 2\sigma$$
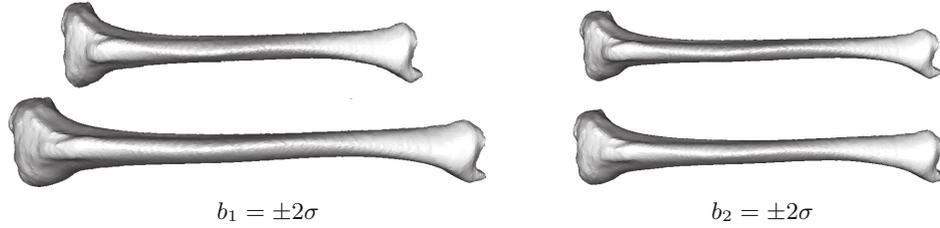
Fig. 1. Statistical shape model of the tibia. The instances were generated by setting the first *(left)* and the second *(right)* parameter to $\pm 2\sigma$. The rest of the modes were set to zero.

where S is a vector containing tetrahedral vertices of the model, the shape of which is determined by independent modes of variation $\mathbf{b}$; $\overline{S}$ is a tetrahedral model of the mean bone; ϕ is a matrix of the principal components; and ω describes zero-meaned Gaussian noise.

Two statistical shape models, representing the femur and tibia, were created using CT images of intact bones, provided by the University Hospital in Ostrava. The characteristics of the models are shown in Table 1. Both tetrahedral models include a polygonal surface, formed by $19,996$ faces and by a subset counting $10,000$ tetrahedral vertices.

As previously described in [3], the length of the femoral or tibial shaft is relatively independent of the shape of the joint regions. Considering the statistical shape models of the involved bones, the length of the shaft is controlled mainly by the first mode b_1, while features such as the size or shape of the joint regions are modeled in particular by the rest of the modes $b_2 \ldots b_n$ (Fig. 1). Therefore, it is impossible to determine the length of a bone based only on the shape of its distal and proximal parts.

3.2 Reconstruction

The reconstruction outcome comprises a model of a patient-specific intact bone, described by shape modes $\mathbf{b}$, and poses $p_{\text{prox}}, p_{\text{dist}}$ of both fracture segments, forming a vector $\mathcal{P} = (\mathbf{b}, p_{\text{prox}}, p_{\text{dist}})$. The results are obtained by minimization of the reprojection error, evaluated using a nonoverlapping area measure (NOA), together with a length criterion (LC):
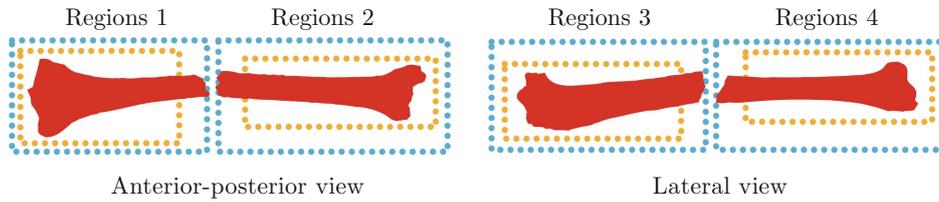
| Regions 1 | Regions 2 | Regions 3 | Regions 4 |

Anterior-posterior view   Lateral view

Fig. 2. The input radiographs are divided into different regions of interest. The regions related to nonoverlapping area evaluation *(orange)* border the maximal intact parts of the bones. The boundaries of regions for length estimation *(blue)* are determined with respect to the detachement point of the bone fragments. The regions are estimated as scaled-up bounding boxes of the fragment segmentations, except the sides nearest to the fracture, which are set interactively by the user. (Color figure online)

$$(\mathcal{P}^*) = \arg \min_{\mathcal{P}} \left[\text{NOA}(\mathcal{P}) + \text{LC}(\mathcal{P}) \right] \tag{2}$$

Both terms are evaluated using the input radiographs, though with different regions of interest (Fig. 2).

Nonoverlapping Area. The measure is evaluated between binary segmentations of the input digital radiographs (DR) and digitally reconstructed radiographs (DRR) [12] with a reprojected statistical shape model, using only the intact regions of the bones. The nonoverlapping area is defined as the area that the segmentations do not have in common (Fig. 3). As shown in [13], it can be evaluated as a sum of the squared pixel differences (PD) between the input and virtual segmentations:

$$\text{d}(\mathcal{P}, x, y) = \text{DR}(x, y) - \text{DRR}(\mathcal{P}, x, y) \tag{3}$$

$$\text{PD}(\mathcal{P}) = \left(\text{d}(x_1 \dots x_n, y_1 \dots y_m) \right) \tag{4}$$

$$\text{NOA}(\mathcal{P}) = \|\text{PD}(\mathcal{P})\|^2 \tag{5}$$

Instead of a count of different pixels, it is convenient to express the size of the nonoverlapping area relatively as $\frac{\text{NOA}(\mathcal{P})}{\text{NOA}(\mathcal{P}) + \text{OA}(\mathcal{P})}$, where $\text{OA}(\mathcal{P})$ is the size of the overlapping area. The measure is an intensity-based similarity metric in the sense that the evaluation is performed directly with the input and reprojected pixels, leading to correspondence-free registration [13]. In contrast, the feature-based methods [5,6] usually require establishing correspondences between the shape model vertices and the contours detected in the radiographs, which is a challenging and error-prone task.

Bone Length Recovery. As the method works with simple or wedge fractures, the injured bone is split into two main fragments. Each fragment is captured in two regions of interest forming a stereo pair, as shown in Fig. 2. The key idea of the recovery is to assign each vertex of the shape model to only one of the

Anterior-posterior view   Lateral view

Fig. 3. Nonoverlapping area between the input *(red)* and virtual *(green)* segmentations. The size of the depicted nonoverlapping area is 28.5%. (Color figure online)

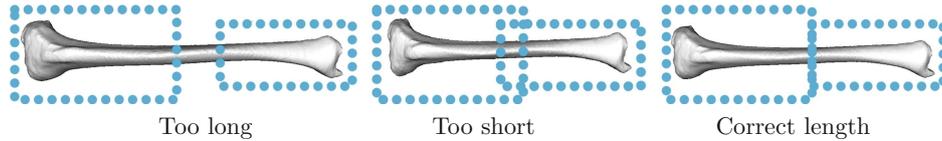Too long   Too short   Correct length

Fig. 4. Relation between vertex assignment and resulting bone length. No assignment of the vertices in the middle of the shaft to any of the fragments leads to a bone that is too long *(left)*. Assignment of the vertices to both fragments results in a bone that is too short *(middle)*. The correct length is ensured by assigning each vertex to exactly one fragment *(right)*.

main fragments. In consequence, each vertex should be reprojected in precisely two regions, which is achieved by minimizing the length criterion:

$$\mathrm{RV}(\mathcal{P}) = \big(\mathrm{r}(\mathcal{P}, v_1 \ldots v_n) - 2\big) \tag{6}$$

$$\mathrm{LC}(\mathcal{P}) = \|\mathrm{RV}(\mathcal{P})\|^2 \tag{7}$$

where $\mathrm{r}(\mathcal{P}, v)$ is the number of reprojections of the current vertex v. The relation between misassigned vertices and bone length is shown in Fig. 4. The regions of interest for the length recovery must be set with respect to a point of detachment (Fig. 2).

Optimization Scheme. The registration is solved as a non-linear least squares (NLS) problem, using a numerical Levenberg-Marquardt optimizer [10]. Although a computationally demanding approximation of the Jacobian matrix is required, due to its high rate of convergence, the optimizer is able to outperform stochastic gradient-free methods [14]. The Jacobian matrix J_F has the following form:

$$J_F = \begin{pmatrix} \partial \mathrm{PD}_{\mathrm{prox}}(\mathcal{P})/\partial p_{\mathrm{prox}} & 0 & \partial \mathrm{PD}_{\mathrm{prox}}(\mathcal{P})/\partial \mathbf{b} \\ 0 & \partial \mathrm{PD}_{\mathrm{dist}}(\mathcal{P})/\partial p_{\mathrm{dist}} & \partial \mathrm{PD}_{\mathrm{dist}}(\mathcal{P})/\partial \mathbf{b} \\ \partial \mathrm{RV}(\mathcal{P})/\partial p_{\mathrm{prox}} & \partial \mathrm{RV}(\mathcal{P})/\partial p_{\mathrm{dist}} & \partial \mathrm{RV}(\mathcal{P})/\partial \mathbf{b} \end{pmatrix} \tag{8}$$

where the partial derivatives are approximated using central differences as $\partial f(t)/\partial t \approx f(t+\epsilon) - f(t-\epsilon)/2\epsilon$.

The reconstruction is divided into three subsequent optimizations. At first, only poses $p_{\mathrm{prox}}, p_{\mathrm{dist}}$ are considered. Next, the first five shape modes $b_1 \ldots b_5$ are optimized together with the poses. Finally, all modes $\mathbf{b}$ are involved in the last stage. Before the optimization, a rough initial pose of the statistical shape

model together with the regions of interest must be set interactively by the user, or estimated from the segmentations, as described in the following sections. The binary segmentations of the input radiographs are performed manually. The modes of variation of the shape model are initialized to zeros.

4   Results

The accuracy and performance of the proposed method were evaluated on synthetic X-ray images of simulated fractures and on real radiographs of dry cadaveric bones suffering perimortem injuries. To evaluate the accuracy, the differences between the polygonal models reconstructed by the proposed method, and ground-truth surfaces obtained from CT data sets were measured using the symmetric Hausdorff distance [4]. The CT data sets of ground-truth bones were never included into the training sets of the statistical shape models. Following the reconstruction convergence criterion stated in [6], the method converged in each evaluated case, as the RMS error was always lower than 3 mm. The ϵ for the Jacobian matrix approximation was set to 1 mm or 1° in the case of pose parameters and to 1 standard deviation σ for shape modes $\mathbf{b}$, as previously proposed in [12].

The evaluations were performed using a 64-bit Windows 7 desktop machine, equipped with an Intel i5 processing unit, NVidia GTX 980Ti 6 GB graphics adapter and 24 GB DDR4 RAM.

4.1   Simulated Injuries

For the *in silico* evaluation of the fracture reduction, we adopted a data set of virtual X-ray images, previously presented in [12]. The virtual radiographs were ray-casted from 8 already segmented CT images of femoral bones obtained from the `Virtual Skeleton Database (VSD)` [11]. From each CT image, 12 virtual stereo pairs of orthogonal radiographs were created, resulting in 96 cases in total. As the bones were rotated 30° along the longitudinal axis between the individual renderings, the data set contained X-ray images captured even from arbitrary views, in addition to standard anterior-posterior and lateral radiographs. The source-image distance (SID) was set to exactly 1 m; the pixel spacing of the radiographs was set to 0.75 mm. To simulate transversal fractures of the femoral shaft, each radiograph was split into proximal and distal parts. A sample test case chosen from the evaluation data set is shown in Fig. 5.

Initial poses of the statistical shape model were generated randomly, with uniform distribution and maximum difference to the ground-truth poses limited to ±10 mm and ±10°, respectively.

Figure 5 shows the result of the virtual fracture reduction of the sample test case. As the virtual radiographs and the reference polygonal models were obtained from the same CT images, the reconstructed bones were compared directly with ground-truth surfaces. The accuracy evaluation for each bone, together with the size of the nonoverlapping area, the number of misassigned
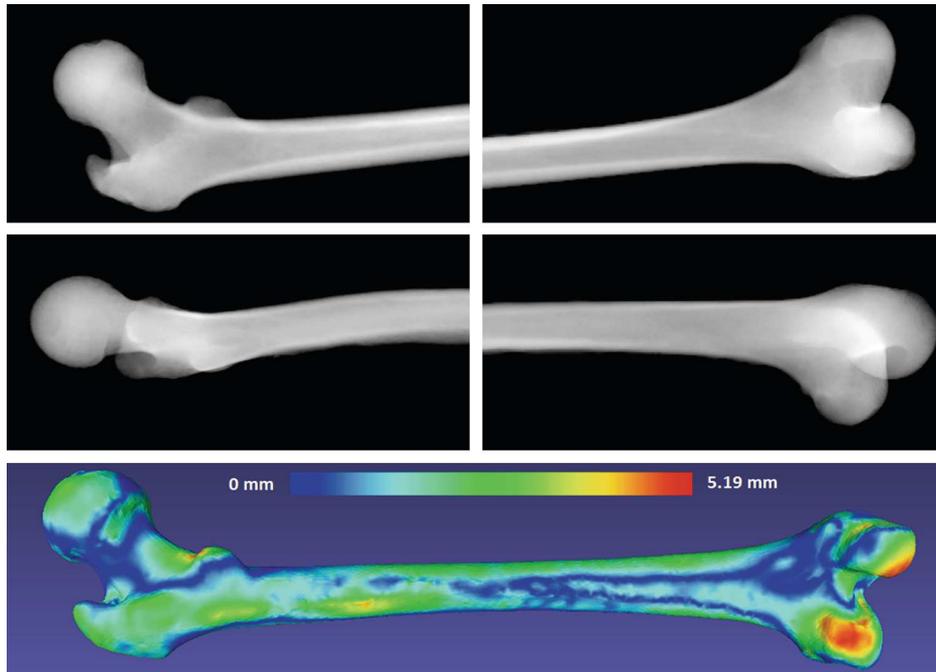
Fig. 5. A sample test case of stereo radiographs with simulative transversal fracture of the femoral shaft, chosen from the virtual data set *(top)*. Accuracy of the sample case reconstruction *(bottom)*. The heatmap shows the differences between reconstructed and ground-truth surfaces, evaluated using the symmetric Hausdorff distance.

tetrahedral vertices, and the length error, as well as the performance evaluation, including the overall reconstruction time, number of iterations in each stage and a total number of rendered images, are shown in Table 2. The results for each bone were averaged from 12 evaluations using different stereo radiographic pairs.

The virtual reduction method extends the *Black & White Pixel Differences* (BW-PD) approach proposed in [12], designated for a single-fragment 2D-3D reconstruction of the uninjured bones. Evaluated on the same synthetic data set, the BW-PD method reached an average accuracy of 1.02 ± 1.35 mm when reconstructing the uninjured bones, while the proposed method reached 1.48 ± 1.16 mm when performing virtual reduction of simulated shaft fractures.

4.2   Dry Cadaveric Bones Study

The cadaveric study involved archeological bones, two femoral and one tibial, suffering peri-mortem diaphyseal fractures. A sample bone from the study is shown in Fig. 6. The radiographs of individual fragments were taken sequentially, using a Kodak Carestream Directview DR 9500 System imaging system. Two CR X-ray cassettes with dimensions of 35×43 cm and 0.168 mm pixel spacing were exploited for the captures. The source-image distance was set to approximately
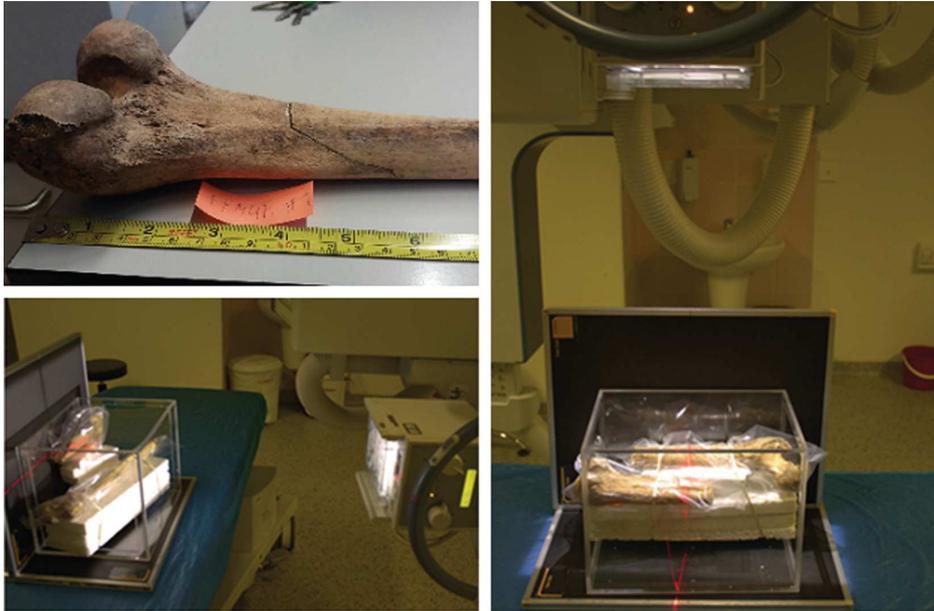
Fig. 6. Physically reduced dry cadaveric femur involved in the study *(top-left)*, capturing an anterior-posterior radiograph of the experimental setup *(right)*, taking a lateral radiograph *(bottom-left)*.

1 m. The radiographs were calibrated using a custom made radiostereometric biplanar calibration box, described in detail in [13]. The complete experimental set up for capturing radiographs is shown in Fig. 6. Individual bone fragments, sealed in a foil sleeve, were placed approximately in the center of the box, on Styrofoam underlays. Contrary to the synthetic data set, the radiographs were taken only from the anterior-posterior and lateral views. After capturing the radiographs, the fractures were actually reduced and fixed by gluing individual fragments together. Then, the reference polygonal models were obtained from CT images of the reduced bones. The poses of the statistical shape model were initialized interactively in a custom viewer.

In contrast with the *in silico* study, rigid registration of the reconstructed bones onto the reference models had to be performed before the Hausdorff distance evaluation. The results of the evaluation are shown in Table 2, revealing a slight decrease in accuracy for the cadaveric bones. The accuracy was affected by the manual segmentation and the real-world calibration of the radiographic images; the higher RMS error in comparison with the simulative data set was caused by certain degradations of the archeological bones involved. The higher number of misassigned tetrahedral vertices was related to a user estimation of the separation spot, which was, by contrast, ideal in the case of the *in silico* study.

Table 2. Results of the reconstruction accuracy and performance evaluation.

Bone	Nonoverlapping area (%)	Misassigned vertices	Length error (mm)	Mean Hausdorff distance (mm)	RMS error
VSD identif.	*Simulated fractures*				
226	2.34	11.0	0.57	1.28	0.99
230	2.38	6.3	1.06	1.23	0.95
238	2.60	10.3	0.49	1.61	1.27
254	2.56	13.3	2.16	1.54	1.20
5900	2.85	6.7	4.07	1.31	1.02
5953	2.60	3.7	2.77	1.41	1.09
6009	2.85	8.1	1.44	1.70	1.35
5939	3.33	15.0	0.94	1.78	1.43
	Perimortem fractured dry cadaveric bones				
Femur 1	3.41	86	3.8	1.89	2.16
Femur 2	2.33	50	3.1	1.38	1.70
Tibia	3.50	131	2.0	1.73	2.16

Bone	Overall time (mm:ss)	Iterations			Renderings
		Stage 1	Stage 2	Stage 3	
VSD identifier	*Simulated fractures*				
226	1:54.4	55.6	16.3	12.9	9,609
230	1:59.0	48.8	21.7	14.8	10,454
238	2:15.2	63.3	21.1	15.0	11,248
254	3:55.2	84.0	79.6	22.5	20,550
5900	3:04.5	80.2	66.3	12.7	15,387
5953	2:11.1	56.5	20.1	15.9	11,149
6009	2:57.6	66.1	58.5	16.3	15,343
5939	2:11.9	67.9	40.3	10.6	11,583
	Perimortem fractured dry cadaveric bones				
Femur 1	3:49.3	11	48	38	19,752
Femur 2	2:16.7	42	24	21	12,684
Tibia	1:13.0	17	34	13	9,160

4.3   Preoperative Planning Software

The method has been built into preoperative planning software, which provides a large database of 3D models capturing bone plates and intramedullary nails. The user is able to select the intended device from the database, place it interactively onto the reconstructed bone model, possibly to perform a virtual bending of the bone plate, and finally refine its pose using an automatic procedure. The application also provides the cutting planes of the obtained polygonal model as a tentative approximation of the fracture detachment sites, or measurements of required screw lengths (Fig. 7). A mutual pose of stereo radiographs is determined using a calibration marker, which is usually attached to a lower limb splint. The shape model is initially aligned with the longitudinal axes of fragments, which are reconstructed in 3D from the binary segmentations.
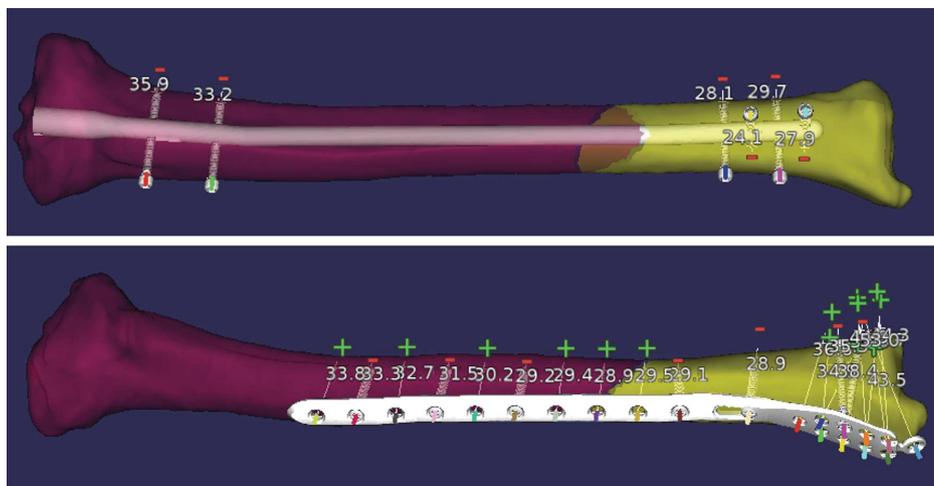
Fig. 7. A virtual simulation of intramedullary nailing of a tibial shaft fracture *(top)*, and a virtual placement of a distal tibial bone plate *(bottom)*. The bone model was reconstructed from radiographs of a real traumatology case.

5   Discussion and Conclusions

In this paper, we proposed a method for a virtual 2D-3D reduction of shaft fractures of the lower limbs. To the best of our knowledge, no other method considering multifragment 2D-3D reconstruction with a focus on accurate length estimation has been proposed so far. The accuracy of the method is comparable even with single-fragment reconstruction approaches, presented in a brief summary in [6]. The results revealed that the accuracy and performance are sufficient for involvement in preoperative planning software designed for the selection of the best-fitting fixation material. To omit the manual segmentation of input radiographs, which is a time-consuming and subjective task, the future work will focus on replacing the nonoverlapping area measure with density-based registration. We assume that the length estimation based on assigning the statistical shape model vertices to individual bone fragments is straightforwardly generalizable, even for application in virtual fracture reduction using 3D models of the fragments obtained from CT images, as proposed e.g. in [3]. The reconstruction method is distributed as open-source library and front-end application at https://github.com/klepo/libmultifragmentregister.

218   O. Klíma et al.

References

1. iJoint: 2D/3D reconstruction of patient-specific hip joint from conventional X-ray radiographs. http://www.istb.unibe.ch/research/information_processing_in_medical_interventions/ijoint/index_eng.html. Accessed 08 Aug 2018
2. iLeg: 2D/3D reconstruction of lower extremity from clinically available X-rays. http://www.istb.unibe.ch/research/information_processing_in_medical_interventions/ileg/index_eng.html. Accessed 08 Aug 2018
3. Albrecht, T., Vetter, T.: Automatic fracture reduction. In: Levine, J.A., Paulsen, R.R., Zhang, Y. (eds.) MeshMed 2012. LNCS, vol. 7599, pp. 22–29. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-33463-4_3
4. Aspert, N., Santa-Cruz, D., Ebrahimi, T.: MESH: measuring errors between surfaces using the Hausdorff distance. In: Proceedings of the 2002 IEEE International Conference on Multimedia and Expo. ICME 2002, vol. 1, pp. 705–708 (2002)
5. Baka, N., et al.: Statistical shape model-based femur kinematics from biplane fluoroscopy. IEEE Trans. Med. Imaging **31**(8), 1573–1583 (2012)
6. Baka, N., et al.: 2D–3D shape reconstruction of the distal femur from stereo X-ray imaging using statistical shape models. Med. Image Anal. **15**, 840–850 (2011)
7. Gong, R.H., Stewart, J., Abolmaesumi, P.: Reduction of multi-fragment fractures of the distal radius using atlas-based 2D/3D registration. Proc. SPIE **7261**, 7261–7261-9 (2009)
8. Issac, R.T., Gopalan, H., Abraham, M., John, C., Issac, S.M., Jacob, D.: Preoperative determination of tibial nail length: an anthropometric study. Chin. J. Traumatol. **19**(3), 151–155 (2016)
9. Jiménez-Delgado, J.J., Paulano-Godino, F., PulidoRam-Ramírez, R., Jiménez-Pérez, J.R.: Computer assisted preoperative planning of bone fracture reduction: simulation techniques and new trends. Med. Image Anal. **30**, 30–45 (2016)
10. Kelley, C.T.: Iterative Methods for Optimization. Frontiers in Applied Mathematics. SIAM, Philadelphia (1999)
11. Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., Büchler, P.: The virtual skeleton database: an open access repository for biomedical research and collaboration. J. Med. Internet Res. **15**(11), e245 (2013)
12. Klima, O., Kleparnik, P., Spanel, M., Zemcik, P.: Intensity-based femoral atlas 2D/3D registration using Levenberg-Marquardt optimisation. Proc. SPIE **9788**, 9788–9788-12 (2016)
13. Klima, O., et al.: Intensity-based nonoverlapping area registration supporting "Drop-Outs" in terms of model-based radiostereometric analysis. J. Healthc. Eng. **2018**, 1–10 (2018)
14. Klima, O., Chromy, A., Zemcik, P., Spanel, M., Kleparnik, P.: A study on performace of Levenberg-Marquardt and CMA-ES optimization methods for atlas-based 2D/3D reconstruction. IFAC-PapersOnLine **49**(25), 121–126 (2016). 14th IFAC Conference on Programmable Devices and Embedded Systems PDES 2016
15. Markelj, P., Tomaževič, D., Likar, B., Pernuš, F.: A review of 3D/2D registration methods for image-guided interventions. Med. Image Anal. **16**(3), 642–661 (2012). Computer Assisted Interventions
16. Schumann, S., Bieck, R., Bader, R., Heverhagen, J., Nolte, L.P., Zheng, G.: Radiographic reconstruction of lower-extremity bone fragments: a first trial. Int. J. Comput. Assist. Radiol. Surg. **11**(12), 2241–2251 (2016)

Virtual 2D-3D Fracture Reduction with Bone Length Recovery Using SSMs   219

17. Smoger, L.M., Shelburne, K.B., Cyr, A.J., Rullkoetter, P.J., Laz, P.J.: Statistical shape modeling predicts patellar bone geometry to enable stereo-radiographic kinematic tracking. J. Biomech. **58**, 187–194 (2017)
18. Valenti, M., et al.: Fluoroscopy-based tracking of femoral kinematics with statistical shape models. Int. J. Comput. Assist. Radiol. Surg. **11**(5), 757–765 (2016)