Robust and Independent 3D Statistical Shape Models for Medical Image Segmentation

zur Erlangung des akademischen Grades eines

Doktors der Ingenieurwissenschaften

von der KIT-Fakultät für Informatik
des Karlsruher Instituts für Technologie (KIT)

genehmigte

Dissertation

von

Tobias Norajitra

aus Karlsruhe

Tag der mündlichen Prüfung:	05. Juni 2018
Erster Gutachter:	Prof. Dr.-Ing. Rüdiger Dillmann
Zweiter Gutachter:	PD. Dr. Klaus Maier-Hein

Zusammenfassung

Die vorliegende Dissertation befasst sich mit der vollautomatischen Segmentierung medizinischer 3D-Bilddaten. Eine Einteilung der Bilder in relevante Objekt- und Hintergrundregionen ist ein notwendiger Schritt für eine nachfolgende ganzheitlichere Interpretation der Daten. Die manuelle Segmentierung solcher Daten ist sehr zeitaufwendig, und eine automatische Segmentierung wird durch Einschränkungen bei der Bildqualität, durch sichtbare Bildartefakte und durch eine variable Patientenanatomie erschwert.

In dieser Arbeit wird das Problem der automatischen Bilddatensegmentierung mithilfe einer methodischen Erweiterung 3D Statistischer Formmodelle (3D-SSM) behandelt. 3D-SSM zählen zu den bekanntesten und beliebtesten Segmentierungsmethoden. Sie verfolgen einen lern- und modellbasierten Ansatz, der Robustheit gegen schwierige Bildverhältnisse verleiht, und der erlernte Modellinformation automatisch mit neu segmentierten Bilddaten verknüpft.

Drei bekannte Schwächen von 3D-SSM werden in dieser Arbeit betrachtet: 1) 3D-SSM benötigen eine genaue initiale Modellplatzierung am Zielorgan, 2) sie sind empfindlich gegenüber wechselnden Modellposen während der Segmentierung, und 3) sie nutzen lediglich lokale Umgebungsinformation am segmentierten Organ, was Probleme verursacht, wenn sich Modellteile weiter vom Zielorgan entfernt befinden. Diese Schwachpunkte werden gezielt durch Einbinden nicht-lokaler Information bei der Erscheinungsmodellierung von Objektlandmarken und mithilfe einer omni-direktionalen, nicht-lokalen Landmarkensuche während der Segmentierung behandelt.

Diese Arbeit leistet folgenden Hauptbeitrag zur automatischen medizinischen Bildsegmentierung: 1) Einen neuen Typ von 3D-SSM, der eine omnidirektionale Suche nach Weichgewebelandmarken in 3D-Bilddaten zeiteffizient in ein Segmentierungsverfahren integriert. 2) Einen Ansatz, der die obigen Schwächen von 3D-SSM erstmals ganzheitlich adressiert. 3) Einen Ansatz, der 3D-SSM ohne weitere Hilfsmethoden als eigenständige Segmentierungstechnik einsetzbar macht. Den oben genannten Schwächen wird mit dem entwickelten Verfahren effektiv begegnet. Es liefert reproduzierbare Ergebnisse für wechselnde Modellplatzierungen im Bild, es findet gesuchte Objektlandmarken erfolgreich über weite Bilddistanzen und es ist ohne vorherige Platzierung am Zielorgan einsetzbar. Das Verfahren lässt sich somit anders als bisher sehr einfach auf eine Vielzahl von Organen, Bildmodalitäten und Szenarien anwenden und erzielt gleichzeitig eine höhere Segmentierungsgenauigkeit. Es wurde auf einer Reihe von proprietären und öffentlich verfügbaren Datensätzen getestet und mit Methoden vom aktuellen Stand der Technik verglichen.

Abstract

This thesis addresses the task of a fully-automatic and accurate segmentation of 3D medical image data. The labeling of object and background regions in medical images is a necessary first step towards a more holistic understanding of clinical data. Manual labeling of images is tedious and time consuming, and a reliable and accurate automatic segmentation is generally hard to obtain, due to low image contrast, noise, artifacts and a high anatomical variability of the segmented organs.

In this thesis, the problem of automatic image segmentation is approached by a methodological extension of 3D Statistical Shape Models (3D-SSM), a well-known and popular segmentation technique. 3D-SSM use an effective learning- and model-based approach to grant robustness against difficult imaging conditions and to link higher-level modeling information with newly segmented images.

Three limitations of 3D-SSM are addressed in this thesis, that strongly affect performance and applicability of 3D-SSM: 1) They require an accurate initial model placement at the segmented organs, 2) they are sensitive to changing model poses during segmentation, and 3) they only rely on local information at the segmented organs, which considerably fails if model parts are out of range during segmentation. These limitations are directly addressed by an incorporation of non-local information of landmark appearance, and by an unambiguous, non-local search for organ surface landmarks.

The main contributions of this work are: 1) A new 3D-SSM that incorporates a landmark-wise omni-directional search for soft tissue landmarks, used in a time efficient framework for volumetric image segmentation. 2) The first joint approach that addresses the aforementioned well-known issues of 3D-SSM. 3) A segmentation approach that makes 3D-SSM applicable as a standalone method without support from additional techniques. The above limitations are effectively addressed by the proposed method. Reproducible results are achieved from changing model pose, sought object landmarks are found across large distances, and the proposed 3D-SSM is applicable without need for previous model initialization at the organ of interest. As a consequence, the proposed method obtains an improved generalizability across different organs, modalities and use-cases, and it achieves an overall higher segmentation accuracy. The method has been tested and evaluated on a variety of proprietary and public datasets, in comparison with methods from the state-of-the-art.

Publications

Peer-reviewed International Conferences and Journals

- **T. Norajitra**, K.H. Maier-Hein, "3D Statistical Shape Models incorporating Landmark-wise Random Regression Forests for Omni-directional Landmark Detection." *IEEE Transactions on Medical Imaging,* vol. 36, no.1, pp. 155-168, 2017.

- **T. Norajitra**, H.P. Meinzer, K.H. Maier-Hein. "3D Statistical Shape Models incorporating 3D Random Forest Regression Voting for Robust CT Liver Segmentation." in *SPIE Medical Imaging,* International Society for Optics and Photonics, pp. 9414061-6, 2015.

- **T. Norajitra**, H.P. Meinzer, K.H. Maier-Hein. "3D Regression Voting on CT-Volumes of the Human Liver for SSM Surface Appearance Modeling." in *Shape - Symposium on Statistical Shape Models and Applications,* p. 13, 2014.

- **T. Norajitra**, H.P. Meinzer, K.H. Maier-Hein. "Active Shape Models incorporating Isolated Landmarks for Medical Image Annotation." in *SPIE Medical Imaging,* International Society for Optics and Photonics, pp. 90353J1-7, 2014.

- A. Giannakis, **T. Norajitra**, L. Kehler, J. Dinkel, O. Weinheimer, C. van Lunteren, M. Kreuter, K.H. Maier-Hein, C.P. Heußel. "Fully Automated Segmentation of Pulmonary Fibrosis using Different Software Tools." in *European Congress of Radiology,* European Society of Radiology, p. 368, 2018.

- M. Kreuter, L. Kehler, **T. Norajitra**, A. Giannakis, M. Wiesenfarth, R. Rubtsov, C. Bauer, K.H. Maier-Hein, C.P. Heußel. "Correlation of Functional Assessment and CT Scores in Acute Exacerbations in Patients with Idiopathic Pulmonary Fibrosis Using the Automated LUng FIbrosis Quantification Tool (LUFIT)." in *ILD Scientific Abstracts: Treatment and Acute Exacerbation,* American Thoracic Society, p. A1659, 2018.

- F. Rengier, O. Naas, **T. Norajitra**, M. Messerli, K. Kallenbach, K.H. Maier-Hein, H.U. Kauczor. "Quantitative MRI-based Three-Dimensional Volumetry of Dural Sac and Vertebral Bodies Improves Diagnosis of Dural Ectasia in Marfan Syndrome." in *European Congress of Radiology,* European Society of Radiology, p. 429, 2017.

- M. Götz, E. Heim, K. März, **T. Norajitra**, M. Hafezi, N. Fard, A. Mehrabi, M. Knoll, C. Weber, L. Maier-Hein, K.H. Maier-Hein. "A Learning-Based, Fully Automatic Liver Tumor Segmentation Pipeline based on Sparsely Snnotated Training Data." in *SPIE Medical Imaging,* International Society for Optics and Photonics, pp. 97841I1-6, 2016.

- P.F. Neher, M. Götz, **T. Norajitra**, K.H. Maier-Hein. "A Machine Learning Based Approach to Fiber Tractography Using Classifier Voting." in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI),* Springer, pp. 45-52, 2015.

- P.F. Neher, M. Götz, **T. Norajitra**, C. Weber, K.H. Maier-Hein. "A Machine Learning Based Approach to Fiber Tractography." in *Proceedings of International Society of Magnetic Resonance in Medicine (ISMRM),* p. 675, 2015.

Peer-reviewed National Conferences

- L. Kehler, C.P. Heußel, **T. Norajitra**, A. Giannakis, C. Bauer, K.H. Maier-Hein, M. Kreuter. "Vollautomatische Quantifizierung der Lungenfibrose mittels CT vor und nach Exazerbation einer IPF - das LUng FIbrosis quantification Tool (LUFIT)." in *Pneumologie,* vol. 72, no. 1, p. 284, 2018.

- F. Rengier, O. Naas, **T. Norajitra**, M. Messerli, K. Kallenbach, K.H. Maier-Hein, H.U. Kauczor. "3D-Quantifizierung der Duraektasie beim Marfan-Syndrom." in *RöFo-Fortschritte auf dem Gebiet der Röntgenstrahlen und der bildgebenden Verfahren,* vol. 189, no. 1, p. 306, 2017.

- **T. Norajitra**, S. Engelhardt, T. Held, S. Al-Maisary, I. Wolf, R. de Simone, H.P. Meinzer, K.H. Maier-Hein. "3D Statistische Formmodelle mit verteilter Erscheinungsmodellierung: Segmentierung per Abstimmung." in *Bildverarbeitung für die Medizin: Algorithmen-Systeme-Anwendungen. Proceedings des Workshops vom 13. bis 15. März in Berlin,* Springer, p. 56, 2016. *Best Paper Award.*

- E. Heim, T. Ross, **T. Norajitra**, M. Nolden, K. März, D. Kondermann, S. Speidel, K.H. Maier-Hein, L. Maier-Hein. "Crowdgestützte Organsegmentierung: Möglichkeiten und Grenzen." in *Proceedings of CURAC 14. Jahrestagung,* Dt. Gesellschaft für Computer- und Roboterassistierte Chirurgie e.V., pp. 37-42, 2015.

Acknowledgments

This thesis was prepared and written at the Department of Medical Image Computing at the German Cancer Research Center Heidelberg (DKFZ). A major part of my work was funded by the German Research Foundation (DFG) as part of project I02 in the CRC 'Cognition-guided Surgery', which is hereby gratefully acknowledged.

I would like to thank my thesis advisor Klaus Maier-Hein for his invaluable support and advice, and for his scientific input during my entire time as a PhD student at the DKFZ. Also, I would like to thank Rüdiger Dillmann at the Karlsruhe Institute of Technology (KIT) for accepting the supervision of my thesis, and for his helpful and highly regarded advice.

Many thanks to my colleagues and friends at the department for the various recreational and sports activities and cheerful times, and for the refreshing and joyful conversations covering all topics apart from our daily work. Particular regards to Jan for our classical music activities and for our culinary excursions, to Sascha for the good hours at the Unisport-Kraftkammer, to my roommates for the many pleasant talks, to the Waschbärbauch runners group, and to many others.

I would like to thank our former chief and founder of the MBI department, Pitt Meinzer, for his always inspiring cultural insights, for his plentiful advice in all matters, and for the humanistic values he holds in his high regards.

And finally, my best thanks to Julie, for all her support, for the good conversations and best times and experiences we enjoyed together.

Contents

List of Figures

List of Tables

CHAPTER 1

Introduction

1.1 Motivation

Since their introduction, medical imaging techniques have evolved into a key technology that has changed medicine and that is now an indispensable part of the clinic. The acquisition and reconstruction of tomographic data allows an unequaled and non-invasive representation of the human body, while different imaging modalities like ultrasound (US), computer tomography (CT), magnetic resonance imaging (MRI) or positron emission tomography (PET) deliver an imagery that can be adapted to a given clinical scenario for maximized insight. Even organs in motion and metabolic processes can be captured based on suitable acquisition protocols.

The obtained volumetric images provide an immediate visual representation of patient anatomy and tissue appearance as a valuable basis for diagnosis, treatment planning, follow-up and monitoring. However, a significant part of information within the data remains dormant unless it is revealed by further image processing. For example, quantifiable information of organ geometry such as volume and shape can be crucial for diagnosis and for therapy choice (e.g. in liver and heart surgery), and an exact geometrical modeling of target organs and adjacent risk structures is needed for a thorough treatment planning in surgery and radiotherapy. For an extraction of this kind of information, an *image segmentation* is required, which means that all relevant anatomical

structures must be identified, marked and delineated within the images.

For lack of a better solution, most images in the clinic are segmented manually by human operators using basic software tools. Due to the sheer size and the level of detail of the acquired image information, a manual segmentation becomes very tedious and time consuming. Also, the achieved accuracy of the segmentation is dependent on experience and skill of the operator. Therefore, solutions for a fully-automatic image segmentation are highly desirable, since they could operate autonomously without need for human interaction, blend in to the clinical workflow, and deliver a reproducible outcome that allows systematic analysis and high-level adjustment of results.

However, low contrast, noise, visible artifacts, as well as a high anatomical variability and various pathologies make an automatic and accurate image segmentation difficult. Also, the mere location of sought organs in an image is unknown by default, while irrelevant background structures crowd the image space. After decades, automatic segmentation is still under active research, and in addition to a large variety of basic algorithms, particular progress has been made with high-level approaches like model-based, atlas-based and deep-learning-based segmentation methods.

As of today, many propositions have been made on how to successfully employ these approaches on different image segmentation tasks. Applications have revealed that notable benefits and weaknesses arise from the underlying nature of these approaches, depending on the intended goals of the addressed tasks. Consequently, the methodologies are generally considered to be complementary tools in the field of medical image segmentation.

3D Statistical Shape Models (3D-SSM) are among the most prominent model-based approaches, and they have been used with great success in several applications. 3D-SSM are characterized by a combination of traits that are particularly appealing for certain scenarios. First, they offer a higher-level representation of segmentation results that is intrinsically backed by trained prior models of object shape and appearance. Each new segmentation that is obtained by 3D-SSM can automatically be put in context with previously analyzed object classes, thus linking higher-level modeling information with low-level image information. This provides the means for an assessment of object typicality and for a direct comparison of object geometries based on well-defined shape correspondences.

Furthermore, higher-level information is provided where low level information is missing, e.g. in the presence of impaired imaging conditions, artifacts or occlusions. In addition to this, the built-in shape constraint of 3D-SSM

grants further robustness while precluding ill-shaped segmentation errors. In contrast to many other methods, 3D-SSM rely on an explicit modeling of the specific target organs, which enables additional means for a thorough organ-wise optimization.

Currently, 3D-SSM only reach their full potential at the expense of a profound organ-wise training optimization. Also, 3D-SSM cannot reach their best performance autonomously, instead they need support from a number of elaborate auxiliary methods. This leads to highly complex segmentation systems which are controlled by a large number of free parameters that have to be optimized for each specific application. Unfortunately, a tweaked system cannot easily be adjusted to new and unexpected image data, nor can it directly be translated to other clinical cases. Usually, an application of these systems then remains limited to the original use-cases they were once designed for.

As a reason for the described inflexibility of 3D-SSM, a number of basic limitations can clearly be identified: restricted use of spatial information, weak learning and an inflexible one-dimensional landmark detection. So far, these basic limitations of 3D-SSM have only been subject to preliminary research. But today, the rapidly developing field of machine learning has led to promising techniques like strong and flexible learning, extensive feature description and global landmark detection. These techniques have already demonstrated their potential in computer vision, and they promise to be an effective remedy for the described limitations.

1.2 Objectives

The main objective of this thesis is the methodological extension of 3D Statistical Shape Models (3D-SSM), with the goal of a promoted overall availability of high-level modeling and segmentation functionality in different use-case scenarios. The result should be a new type of 3D-SSM that provides both high flexibility and accuracy for the fully-automatic segmentation of medical image data. The new 3D-SSM is meant to operate autonomously without support from auxiliary methods, and it should easily be trainable for changing use-cases. During segmentation, the new standalone method is supposed to achieve an immediate higher accuracy without need for a tailor-made system design. In particular, the following characteristics should apply:

- **Independence from auxiliary methods**: The segmentation outcome of the new 3D-SSM should not rely on an application of additional, exter-

nal methods. Especially, accurate results should no longer be dependent from an exact initial placement of the employed surface model. Thus, sophisticated model initialization procedures should no longer be a prerequisite for high accuracy in future applications.

- **Fully-automatic training**: During training, no manual intervention should be required for a sustained high accuracy across different use-cases. From underlying training data, image features should automatically be determined that best describe the appearance of a sought target organ. For a successful learning of different use-cases, no further adjustment of training parameters should be required.

- **Accuracy *and* generalizability**: A high overall accuracy should be achieved without loss of generalizability. The method should produce similar results in comparison with optimized, use-case tailored hybrid systems.

- **Clinical workflow compatibility**: Segmentations for single target organs should be available within minutes. Both quick visual results and time-efficient background processing of larger clinical datasets should be provided this way. Fine-grain parallelization and an optional integration of algorithms for quicker learning should facilitate training on large datasets e.g. for ongoing studies.

1.3 Contribution

This work contributes methodological extensions to 3D Statistical Shape Models (3D-SSM) that address previous well-known problems of this popular and widely-used segmentation technique. The contributed extensions aim at higher intrinsic search and model fitting capabilities of 3D-SSM, and at a higher availability of segmentation performance from 3D-SSM in different segmentation scenarios.

This work proposes the following methodological contributions:

- A new 3D-SSM is presented that incorporates a landmark-wise omnidirectional search for arbitrary soft tissue landmarks, using a time efficient segmentation framework for application in volumetric image data.

- For the first time, three well-known problems of 3D-SSM are jointly addressed by the proposed method, i.e. dependence of results from previous

model initialization (*Model Initialization Problem*); dependence of results from model pose changes and proneness to be stuck in local minima during segmentation (*Visibility Problem*); and a limited search range that fails to attract 3D-SSM to the sought organ of interest from growing distances (*Capture Range Problem*).

- A segmentation approach is presented that generalizes well across different organs and modalities, only relying on the proposed 3D-SSM as a self-sufficient segmentation technique without need for support from other methods. In particular, the proposed 3D-SSM can be used as a single technique for multi-organ segmentation tasks without additional effort for previous organ localization in larger image volumes.

1.4 Structure of this thesis

This thesis is structured as follows: Chapter 2 will present an overview of the state-of-the-art for 3D-SSM and their application in image segmentation. The overview will focus on techniques in the field of medical image analysis, but references from other fields like computer vision will be included as needed. The chapter will encompass the essential concepts that contribute to the successful modeling, training and application of 3D-SSM. In particular, different strategies and approaches will be detailed that address the modeling of object appearance, and the model adaptation process during segmentation.

In Chapter 3, the training of shape prior models for a number of common organ structures will be presented, that build the foundation for the later extensions of appearance modeling, of landmark search and of model fitting in this work. Methods for shape model construction will be chosen for the proposed method, that allow an incorporation of techniques on the basis of explicitly modeled surface landmarks. The extensions to 3D-SSM in this work will be proposed independently from the underlying shape modeling techniques, except for a claimed landmark-based shape representation. Thus, the proposed extensions can be combined with a large variety of other landmark-based approaches that address different domains and issues of 3D-SSM.

In Chapter 4, the proposed use of a non-local landmark context appearance modeling will be presented. The aforementioned well-known problems of 3D-SSM will be analyzed in detail, and conceptual solutions will be motivated to address these problems. Classification- and regression-based strategies for object modeling and detection will be discussed. The use of a flexible, randomized

3D Haar-like feature descriptor will be described, including its time-efficient computation and normalization, and its role in the greater modeling and detection framework. Finally, the proposed use of landmark-wise trained random regression forests in 3D-SSM will be described, incorporating flexible and robust machine learning for the intended modeling and detection of arbitrary soft tissue landmarks in volume images. In addition, the use of extremely randomized trees and faster sample impurity computation will be proposed to speed up training.

In Chapter 5, the previous techniques of shape prior modeling, non-local landmark appearance modeling and omni-directional landmark detection will be put together to provide a time efficient model-based segmentation approach. The choice of a voting scheme from different possible strategies will be motivated, thus enabling a robust aggregation of forest predictions on the sought landmark positions, based on information from distributed viewpoints in the image. The omni-directional landmark detectors will be embedded in a multi-scale approach, where the use of sparse surface modeling, a multi-resolution model fitting and a final profile-based segmentation for all landmarks will provide a time efficient, robust and accurate applicability for the considered 3D applications.

Chapter 6 will present a experimental validation of the proposed method in a variety of segmentation experiments, for different organs of interest, imaging modalities and use-case scenarios. The goal of the experiments was to examine the effects of the proposed techniques on model initialization, visibility and capture range problems of 3D-SSM, and to demonstrate the generalizability of the proposed method to different use cases. In a comparison with a typical previous 3D-SSM, the robustness of the proposed 3D-SSM was tested against random model pose changes around the organ of interest. Capture range experiments were conducted to determine the distance from which the proposed method is attracted to the underlying organs of interest. Search experiments were performed where the quality of segmentation results was compared when started under different conditions of model initialization, regarding the proneness to locally optimal solutions during segmentation. And finally, the proposed method was compared with various methods from other groups on publicly available data, comprising a liver segmentation task and a multi-organ task for liver, spleen and kidney segmentation on CT data. In addition, experiments were conducted for liver segmentation in T1-weighted MRI data, for left ventricle segmentation in ultrasound data, and for a segmentation of pathological lungs in MDCT data, to further demonstrate the generalizability

of the proposed method to challenging and heterogeneous imaging modalities and to a clinical dataset with pronounced presence of pathologies. Finally, a runtime and robustness analysis of the employed random regression forests was conducted, where influence of a number of training parameters on segmentation accuracy, on convergence, on precision of forest predictions and on training duration was examined.

In Chapters 7 and 8, based on the shown experimental results, the general characteristics, benefits and limitations of the proposed method will be discussed and concluded, regarding its solution for the addressed problems of 3D-SSM, its application under varying imaging conditions, encountered anatomical and pathological variabilities, and its implications for model based segmentation, also in comparison with other approaches from the field of medical image segmentation.

CHAPTER 2

State of the Art

2.1 Model-Based Image Segmentation

A number of fundamental tasks constitute the greater field of medical image processing. They have coined the major disciplines of segmentation, registration and modeling, for which numerous methods were developed during several decades of active research.

Image segmentation deals with the separation of images in different object and background regions. For the task, separate methodologies have evolved which build on their own strategies and ideas to perform a successful segmentation. Due to the notion of an extended model-based segmentation in this thesis, the following chapter will focus on an overview of current 3D Statistical Shape Model (3D-SSM) approaches. A detailed overview of the field of medical image processing, and of other segmentation methodologies can e.g. be found in [1].

3D-SSM are among the most prominent model-based approaches, and they have been used with great success in various applications. 3D-SSM rely on a surface representation of distinct organs of interest in order to provide a targeted segmentation of these structures. In addition, 3D-SSM incorporate prior knowledge of shape and appearance which is learned for given organs of interest from previously annotated training images. The learned prior is then adopted during segmentation in order to fit a surface model to the learned

target organ in new unknown images. Despite the advent of more and more promising voxel classification and labeling techniques, many approaches still rely on a comparable shape representation at some point in order to improve segmentation accuracy. A comprehensive survey of 3D-SSM and their application can be found in [2].

Also, 3D-SSM are characterized by a combination of traits that are particularly appealing for certain scenarios. First, they offer a higher-level representation of segmentation results that is intrinsically backed by trained prior models of object shape and appearance. Each new segmentation that is obtained by 3D-SSM can automatically be put in context with previously analyzed object classes, thus linking higher-level modeling information with low-level image information. This provides the means for an assessment of object typicality and for a direct comparison of object geometries based on well-defined shape correspondences.

Starting from a set of annotated training images, a number of tasks and challenges arises en route to a 3D-SSM that is applicable for image segmentation. First, a geometric representation must be provided that is suitable for a modeling of the considered object classes. Different representation techniques have been proposed that offer particular benefits for chosen application scenarios. These techniques will be presented in Sect. 2.2.

Furthermore, shape prior information that is retrieved from the underlying training shapes needs to be properly encoded and modeled in order to be applicable for shape analysis and segmentation. To provide a comparable basis for different shapes and to capture only meaningful shape variation from training data, the latter typically needs to be transformed to a common shape domain. Techniques for a suitable processing of the original training shapes and for the subsequent modeling of shape prior will be outlined in Sect. 2.3.

Learning object class appearance from image information for successful segmentation is another key feature of 3D-SSM. Similarly to the previous shape prior modeling, different challenges and design choices arise during this process. For a meaningful modeling of object appearance, various feature descriptors exist which respond to certain object characteristics as they are encountered in the underlying images. They build the foundation for a learning of object appearance from the training data, for which various learning techniques have been proposed. These topics will be detailed in Sect. 2.4.

3D-SSM typically perform an iterative search for an optimal model placement at the organ of interest to perform a segmentation. For best results, a best possible model initialization near the organ of interest has to be provided

prior to the iterative segmentation process. A number of well-known problems arise from the notion of a model initialization that provides boundary conditions for a subsequent, highly dependent search. As a response to the arising need for compensatory measures, various techniques have been proposed in the literature. These topics and techniques will be described in Sect. 2.5.

The majority of 3D-SSM that are used for general segmentation purposes build on a landmark-based shape prior model. A landmark-based representation offers various benefits, and it goes particularly well with the algorithmic development in this thesis. Therefore, the remainder of this chapter will focus on developments that build on this type. The proposed enhancements for 3D-SSM will however not be limited to this object representation type alone. Further details on the developments regarding all types of 3D-SSM can be found in [2].

2.2 Surface Representation

3D-SSM rely on a geometric modeling of organ surfaces that naturally provides an organ delineation in the image volumes during segmentation. They also serve as a basis for the learned shape and appearance prior during training, and for an incorporation of the latter in the model fitting process during segmentation.

In this regard, different surface representation techniques exist that have been employed in a variety of image analysis tasks. They provide distinct advantages concerning compactness, simplicity, topological flexibility or geometrical expressiveness of the resulting surface model. Generally speaking, some techniques offer convenient solutions for a broad variety of segmentation tasks, whereas others provide particular solutions e.g. regarding anatomical structures of complex geometry and peculiar shape variation.

In general, the choice of surface representation affects 3D-SSM at various stages during their training and their application. Specifically, this means that requirements are introduced on the retrieval of corresponding surfaces from labeled training data, on the statistical modeling of shape variation, or on surface detection and model adaptation during segmentation. The different representations will shortly be detailed in this regard in the remainder of this section.

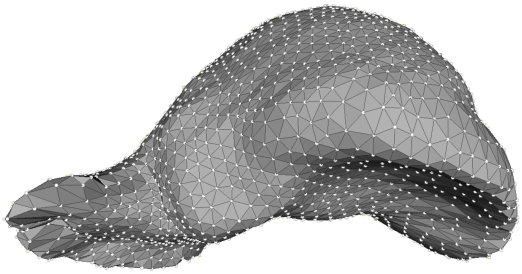


Figure 1: Landmark-based surface model of the liver. An explicit representation of surface landmarks embedded in a triangular mesh.

2.2.1 Landmark-based Models

The most widespread surface representation, called the *Point Distribution Model (PDM)*, is based on an explicit modeling of distcint surface landmarks $\vec{x}_i = (x_{11}, x_{12}, x_{13})^T$ that are usually integrated in a triangular mesh structure. Consequently, given shapes $\vec{X}_i$ can be encoded through a concatenation of their landmark coordinates:

$$\vec{X}_i = (x_{11}, x_{12}, x_{13}, \cdots, x_{n1}, x_{n2}, x_{n3}) \in \mathbb{R}^{3n}, i = 1, \cdots, k$$

The PDM has originally been proposed for statistical shape models in [3], and thanks to its simplicity, its computational efficiency and its generality, this type of representation has remained highly popular ever since. Accordingly, it has been a foundation for many successful applications in image segmentation. Due to the general focus on image segmentation tasks in this thesis, the PDM will also build the foundation for all presented developments. Henceforth, emphasis will be put on shape and appearance prior modeling techniques that are of particular relevance for this type of surface representation.

2.2.2 Medial Models

Another class of surface representations is based on medial models. They describe a shape of interest by an internal centerline structure in conjunction with distributed radial information to capture surface geometry. This kind of representation was originally proposed for shape modeling in 2D domain [4], where connecting vectors between centerline and surface were defined for description of 2D shapes. The concept was later extended to 3D domain [5], and

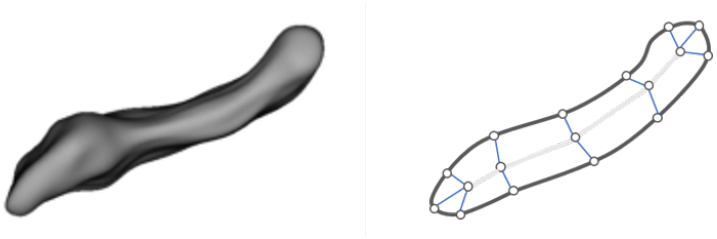


Figure 2: Spherical harmonics (left) and medial model (right) surface representations. Shapes are represented by sets of basic functions and by centerlines with added radial information.

to the continuous modeling of surface geometry [6]. Apart from the increased compactness of the medial model representation, it is apparently ideally suited for the modeling of structures that exhibit convoluted tubular geometries, e.g. the cochlea [7].

2.2.3 Fourier Descriptors, Spherical Harmonics, Splines

Furthermore, methods have been proposed that provide a compact representation through a finite number of descriptive elementary functions. Their compact representation allows a higher-level interpretation, and they promise higher topological flexibility during training. Also, their parameterization-based representation can be used for a mapping to the higher parameter space, where shape correspondences are more easily established, which is a prerequisite for shape model building. In [8], the Fourier transformation known from signal processing, is extended to the surface domain. Similarly, a set of basic functions is employed in [9], [10] called *spherical harmonics* (*SPHARM*), which offer high-level representation of deformable surface models for segmentation and shape analysis as in [11]. Also based on a spherical composition, a hierarchical approach is proposed in [12] that describes surfaces through vibrational modes of spherical meshes. Obviously, both techniques are limited to the description of spherical topologies. This is overcome by an extension of SPHARM called surface harmonics from [13]. A set of wavelet functions for shape description was proposed first in 2D [14], and later in 3D domain [15], [16] using spherical wavelets.

Another, highly compact description is based on *non-uniform rational B-Splines (NURBS)* [17], which describe a surface through a sparse number of control points and their resulting spline interpolation. While their explicit and

compact description allows convenient user interaction through manipulation of control points, the level of surface detail is rather limited. For a shape analysis apart from segmentation, the use of shape descriptors was proposed in [18].

2.2.4 Level-Sets

In [19], level-sets were proposed as an implicit representation of object geometries. Object boundaries are defined by the zero-level sets of a higher dimensional level-set function. The major benefit of this modeling type is that it provides topological independence for modeled objects, which can have arbitrary, changing topologies during application. However in general, the use of level-sets for image segmentation is not straightforward and often requires an application specific design of the underlying level-set function.

The original notion of level-sets for object representation from [19] was consequently adopted for use in image segmentation by [20], where a model fitting scheme and the use of predefined image features were proposed. Subsequently, shape prior was added to the level-set formulation in [21], [22], where prior knowledge was represented in image domain through the use of distance maps. Due to the non-linear distance mapping, ill-shaped solutions could not be prohibited, which was later tackled by [23] through a representation of shape prior in the linear LogOdds space. Another problem with level sets and shape priors arises from a time consuming estimation of optimal pose parameters during segmentation. In [24], a solution was proposed based on intrinsic coordinate systems in each training shape, following the idea described in [25]. With it, translation, rotation and scale of shapes could directly be incorporated in the evolution equation, making time consuming optimization of registration parameters obsolete.

Based on their implicit surface formulation, on the requirements during their algorithmic design and on the peculiarities during their application, level-sets can be regarded an alternative concept that is entirely different from the aforementioned explicit representation techniques. An extensive review on level-sets can for example be found in [26].

2.3 Shape Prior Modeling

Shape can commonly be defined as the part of object geometry that is invariant under translation, rotation and scale. To incorporate prior knowledge of shape

variation from a class of objects, a shape prior model is trained on previously annotated data. During model training, a shape distribution is estimated from the encountered objects in the training data by means of statistical modeling. The trained shape prior model can be used to generate plausible shapes that are in accordance with the observed shape variation in the training data.

The shape prior model grants a generally strong robustness for 3D-SSM during segmentation. Erroneous surface geometries are excluded by the learned shape constraint, while convergence of the segmentation process is promoted towards plausible solutions. Also, the shape of any solution that is hypothesized during segmentation can be projected back to the learned shape distribution, where its plausibility can be assessed.

Training shapes in a common shape representation as described in the previous sections serve as input for model training. A landmark-based surface representation builds the foundation for the techniques that were developed in this thesis. The following parts of this section will therefore mostly be related to this type of surface representation. However, numerous similar concepts have been proposed in the literature for the other representation types.

Prior to shape model training, the provided surface shapes typically undergo further processing in order to remove non-essential information from the training data and to ensure conformity of the surface shapes. Also, an optional augmentation of training samples can be performed to increase information gain from limited training data. Processing of training shapes and shape prior modeling will be the subjects of the following sections.

2.3.1 Shape Alignment and Normalization

Before suitable information on shape variation can be extracted from the initially provided training shapes, an alignment and a scale normalization have to be conducted on the latter. Intuitively for most applications, shape variation is considered as the disparity between distinct shapes that remains after a removal of differences in translation, rotation and scale. Consequently, learning and modeling shape distribution from a set of training shapes requires an alignment of all input surfaces by affine registration first.

An efficient and convenient method for alignment of landmark based shapes is provided by the generalized procrustes analysis (GPA) [27], [28]. The GPA works under the assumption of an equal number of corresponding landmarks that are distributed on the surface of each shape. This requirement is also a prerequisite for the subsequent statistical shape analysis, and it is ensured

throughout the model training process by a dedicated correspondence optimization step that will be the topic of the following section.

The procrustes analysis delivers a closed form solution that minimizes the euclidean distance between two given shapes. Based on the pairwise alignment of shapes, the GPA produces an alignment of all training shapes with an optimized mean shape, i.e. the one that minimizes squared distances to the aligned shapes. Since the initially unknown optimal mean cannot be identified analytically, it is determined iteratively through shape alignment and repeated updates of the mean shape.

In order for this process to converge, at least a scale normalization of the mean shape (usually $|\bar{x}| = 1$) has to be maintained throughout the process [3]. The remaining shapes can then either be scaled implicitly during the procrustes analysis, or explicitly, to a hypersphere ($|x_i| = 1$) or into the tangent space of the mean shape through a posterior scaling by $1/x_i\bar{x}$ for each shape. It can be shown that only for the latter, additional non-linearities in the resulting shape distribution can be avoided (cf. [29]).

Further propositions on shape alignment were made in the literature. In [30], an optimization of the L_1 and L_∞ norms was proposed during the procrustes analysis, thus addressing sensitivity to outliers. In [31], an optimization based on the minimum description length (MDL) was proposed, which is compatible with correspondence optimization methods that also rely on MDL optimization. In [32], a group-wise shape alignment was proposed by establishing transitivity on the pairwise transforms, which addresses previous sensitivity to noise from occlusions or false landmark correspondences.

2.3.2 Correspondence Optimization

For landmark-based surface representations, a statistical modeling of shape variation is generally performed in a common parameter space spanned by the landmark coordinates of the training shapes. Obviously, embedding all shapes in this joint parameter space requires a uniform representation based on an equal number of corresponding landmarks on each single shape. Shape variation can then be derived from the positional variations of any set of corresponding landmarks on all training samples.

This means that the actual placement of corresponding landmarks on the different shapes has a direct influence on the validity of the shape variation that is captured therefrom, and the quality of the trained shape prior model is consequently dependent from a meaningful placement of landmark positions

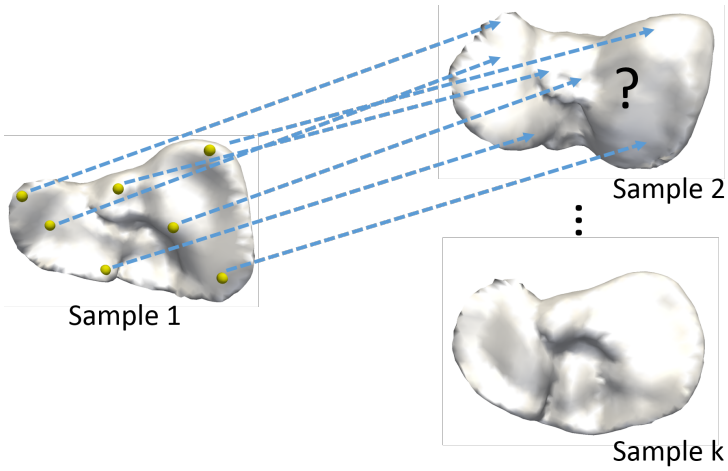


Figure 3: Landmark correspondence problem. Find a set of corresponding landmarks across different surface shapes.

on each shape.

The optimal selection of corresponding landmark positions on the training data is a non-trivial problem. First, the large number of required landmarks on multiple training shapes renders a manual placement non-feasible, particularly so for 3D surface shapes. And second, the actual meaningfulness of landmark placements is in general undefined, and in theory, requirements for optimality have to be considered with regard to the purpose of the subsequent shape prior modeling.

Consequently, a number of challenges arises, that altogether contribute to the complexity of a suitable correspondence optimization for 3D-SSM: the automated placement of large numbers of landmarks on all training shapes; the convenient and accurate selection of meaningful landmark positions on different shapes; and the establishment of an optimality criterion for the chosen landmark correspondences.

For an automated placement of large sets of landmarks on multiple training shapes, different strategies exist. For instance, all landmarks can be transfered from a template shape to all other training shapes via mesh-to-mesh registration. Alternatively, an image registration can be performed on all training images. Then after surface mesh creation on any labeled image, landmarks can directly be transfered to all other cases automatically. Another strategy can be to perform a segmentation of unlabeled images, either using a de-

formable template mesh or a presegmented atlas. Finally, all training shapes can be projected to a joint parameter space that serves as a common reference frame for all projected shapes (parameterization-based correspondence optimization). There, landmarks can then conveniently be placed on all training shapes simultaneously.

All of these approaches yield their own advantages and problems regarding the convenience, the meaningfulness and an intended optimality of the landmarking. These traits play an important role in the model building process, since they decide about the availability of an accurate shape prior modeling in various applications. Accordingly, extra measures can be required for the different approaches in order to improve the quality of their found correspondences.

Regarding the optimality of the landmark correspondences, various demands can be raised. At least, a homeomorphic mapping of the corresponding training meshes is naturally desired, where landmark positions do not produce ill-defined foldings on the corresponding surface parts of the remaining shapes. Another qualitative demand can be to remove any unfavorable biases of the correspondence optimization that emerge from the actual mesh structure of the training shapes, or from the choice of arbitrary reference templates at the beginning of the process (selection bias).

Apart from these qualitative optimality criteria, various quantitative measures exist that regularize landmark correspondences. They aim either at an optimization on a per shape basis (e.g. through minimization of landmark distances, or through conformity on shape features), or they impose optimality on the whole set of training shapes simultaneously (group-wise optimization).

According to these basic ideas, various propositions have been made in the literature for correspondence optimization. They will briefly be reviewed in the following subsections.

2.3.2.1 Mesh-to-mesh registration

In [33]–[35], correspondences via mesh-to-mesh registration were proposed, using the iterative closest point algorithm (ICP) [36] or the Softassign Procrustes [37]. For removal of the selection bias from the arbitrarily chosen reference shape, a binary tree for the pairwise registration of shapes was proposed in [38], based on a symmetric ICP algorithm. In order to impose optimality to the correspondences, feature preserving registration was proposed with regard to crest lines, normal vectors and local shape information [39]–[41].

The confinement to similarity transformations during shape registration

does not naturally enable the direct selection of common landmark positions on two registrated shapes. Instead, an inaccurate selection of landmark positions is performed based on minimum surface distance. Especially in case of large, elastic shape variations, this can negatively affect the quality of the landmark correspondences, resulting in non-homeomorphic mappings and ill-defined surface meshes.

One straightforward solution for this problem can be a non-rigid registration based on splines, e.g. using B-Splines [39] or multi-resolution Octree Splines [42]. Another alternative is the selection of sparse correspondences based on shape features [43]–[46] or on a sparse manual landmarking [47]–[51]. From there, the remaining landmark correspondences can then be established such that mesh quality is preserved, e.g. through thin-plate spline registration, followed by physical-model based mesh adaptation [47] or by a markov random field regularization of the spline deformations to ensure homeomorphic mapping [51]. Apart from manual landmarking, automatic feature detection can be involved, e.g. based on similarity, structural and prior information [43], based on local surface geometry and geodesic distances [44], or using pattern recognition through clustering, classifier training or dynamic programming [45], [46].

2.3.2.2 Template- and atlas-based matching

One major drawback of the mesh-to-mesh registration based approaches described above is the necessity for non-rigid surface matching techniques. It can instead be reasonable to rely on mesh adaptation techniques that are later used during segmentation anyway. Or, if manual annotations of training images are not available, the fitting of a deformable model to these images can be a reasonable option, thereby retrieving landmark correspondences.

In [52]–[55], the matching of a template shape to pre-segmented images was proposed, and alternatively in [56], [57], the template-based segmentation of unlabeled images. In order to remove the selection bias from the chosen template, an optimal shape selection was introduced in [55], involving additional shape modifications for improved fitting. Apart from this, measures that rely on physical deformation models, multi-resolution model fitting and gradient vector flow are employed to improve the adaptation accuracy of the template shape.

Analogously for the use of atlases, registration to pre-segmented images [58]–[60] and atlas-based segmentation of unlabeled images [57], [61], [62] were proposed. The proposed techniques involve rigid pre-registration followed by thin-plate splines or B-splines [57], [58], [61], relying on mutual information

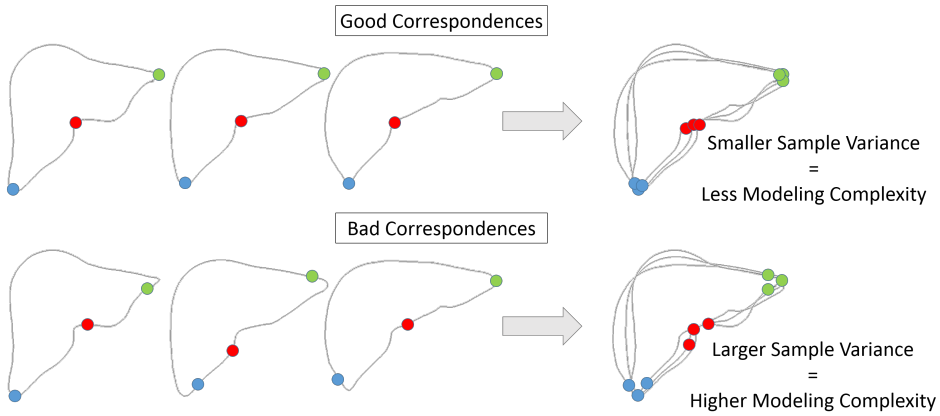


Figure 4: Population-based correspondence optimization. Approaches of this type assume that the best correspondences are those which explain shape variation with least modeling complexity.

and normalized mutual information (NMI), or on label consistency and on the kappa-metric for the registration of multi-label images [59], [60]. For a removal of the selection-bias from the employed atlas, an atlas-selection based on minimization of overall deformation was proposed [62].

2.3.2.3 Parameterization-based Correspondence Optimization

Although template- and atlas-based matching techniques are appealing due to the re-use of established methods from segmentation and image registration, they do not essentially address the original problem of optimizing landmark correspondences. They rather focus on a mere matching of shapes and atlases to similar objects in image domain. However, apart from tackling selection-biases, they do not consider any optimality criteria regarding the resulting landmark distribution on the matched objects.

Therefore, correspondence optimization is still most naturally approached in the surface domain of the underlying training shapes. To avoid the aforementioned problems of the mesh-to-mesh registration approaches in terms of non-rigid shape matching and limited availability of sparse surface features, a parameterization based approach can be the better choice. The basic idea is to provide a parameterization for the different training shapes, thus transforming each shape into a common parameter space. There, landmarks can

be defined and be distributed across all shapes in a convenient and unbiased manner, without need for a non-rigid registration of selected shapes.

Several methods have been proposed that rely on different parameter spaces and optimization strategies. In [10], a parameterization via spherical harmonics was proposed. There, all shapes are aligned by their first-degree ellipsoids, and landmark correspondences can immediately be transfered to all shapes. In [40], [63], harmonic maps were employed for a mapping of cartilage to 2D topological disks, with shape alignment based on disk rotation minimizing landmark distances that were determined via ICP on the training shapes. While diffeomorphism across all shapes is established this way, an arbitrary landmarking is employed that does not aim for further optimality.

A natural approach to achieve higher quality of the chosen correspondences is to take into account anatomical or geometrical key structures, which are extracted manually [64], [65] or automatically [66]. Examples are sulci of the brain [64], user-selected image patches, feature lines [65], [67], landmark features [66], or geometrical characteristics like surface normals and curvature features. Optimization of the landmark correspondences is carried out either during the initial parameterization of all shapes [64]–[66], or iteratively through re-parameterization of the initial shape parameterizations [67]–[69].

Apart from the above propositions where optimality is introduced rather intuitively through conformity across distinct shapes, an entirely different optimization policy can be to aim at the joint optimality of all shapes (population-based / goup-wise optimization). A suitable objective for a more global correspondence optimization is given by the quality of the shape prior model that itself relies on the quality of the chosen landmark correspondences. Therefor, a feasible criterion is based on the compactness of the resulting linear shape prior model. Generally for the covariance matrix Σ, this means an overall smaller aggregation of eigenvalues on fewer eigenmodes, thus favoring models that explain shape variation more efficiently than other, more complex models. Although being a basic assumption, this well-known principle holds true for many real-world problems and applications.

Following this rationale, different optimality criteria were proposed in the literature. In [70], an optimality criterion based on the determinant of the covariance matrix was proposed. There, only a limited choice of landmark positions was supported. In [71], an optimization of the minimum description length (MDL) was proposed. The latter is an information theoretical measure that follows the aforementioned Occam's Razor principle of favoring simpler hypotheses over the more complex ones. Here, the main challenge lies in the

increased computational complexity of the MDL cost function. As a suitable alternative, a simplified version of the MDL was proposed in [72], which supports a more efficient computation. In [73], an entropy based optimization was proposed that is also considered equivalent to the MDL. This approach also introduces a particle based alternative to the previously used spherical parameterization, thus enabling topological independence. Another alternative was proposed by [74], relying on a 2D disk parameterization and on piecewise bilinear maps. Furthermore, an MDL-based group-wise image registration was proposed in [75].

For the actual optimization of the proposed cost functions, genetic algorithms [70], [71] and a Nelder-Mead heuristic [76], [77] were proposed. Significant speedup was achieved by [78] based on a gradient descent optimization of the simplified MDL.

Summarized, the parameterization-based approaches provide an effective and convenient means for a direct optimization of correspondences across multiple training shapes. They implement well-defined cost functions that directly aim at optimality of the trained shape prior models. Approaches have been proposed for objects of arbitrary topology, and although computationally the methods generally come at a higher expense, the availability of faster optimization algorithms and the importance of model quality for the later applications make a use of these approaches worthwhile. They can provide a robust shape prior during segmentation that accurately represents the considered object classes in terms of specificity and generalizability, and that can deliver meaningful higher-level information from newly segmented objects.

2.3.3 Modeling Shape Distribution

After meaningful landmark correspondences have been established on the training shapes, shape prior knowledge can be learned from the training data. The learned shape prior can then be used to generate plausible shapes and to fit them to the sought objects in new image data. Or it can be used to asses similarity of shapes that are retrieved during segmentation or during any other application involving a representation of the modeled objects.

Mathematically, shape prior can be represented by any shape distribution that is suited to explain shape variation that is encountered in the training data. Different basic assumptions can be made on the underlying distribution. In conjunction with landmark-based SSM, the assumption of a normal distribution was proposed in [3]. Shape variation is explained by the mean shape

$\bar{X}$ and by the covariance Σ of the training shapes

$$\vec{X}_i = (x_{11}, x_{12}, x_{13}, \cdots, x_{n1}, x_{n2}, x_{n3}) \in \mathbb{R}^{3n}, i = 1, \cdots, k$$

A dimensionality reduction is performed using the principal component analysis (PCA), which results in a smaller number of eigenvectors of Σ that are used for shape description within the entire shape space $\mathbb{R}^{3n}$.

Numerically, this is preferably done prior to a calculation of the covariance matrix, using the singular value decomposition (SVD) due to its greater stability. In order to address outlier sensitivity of conventional eigenanalysis and SVD, a robust PCA was proposed in [79].

As a result, shapes $\vec{X}_j$ can be generated in accordance with the modeled distribution:

$$\vec{X}_j = \bar{X} + \sum_{i=1}^{c} \alpha_i \Phi_i = \bar{X} + P\vec{b}$$

Based on the eigenmodes P of the shape distribution, the plausibility of future shapes $\vec{x}$ can be assessed using the Mahalanobis Distance

$$d(\vec{X}_j, \bar{X}) = \sqrt{(\vec{X}_j - \bar{X})^T \Sigma^{-1} (\vec{X}_j - \bar{X})}$$

Likewise, the generation of surfaces can be constrained to plausible shapes by limiting α_i from the parameter vector $\vec{b}$ separately, or by applying a hyper-ellipsoidal constraint to $\vec{b}$.

The above relies more or less on the absence of non-linearities in the underlying shape distribution, and plausibility is not assured for the larger discrepancies in the presence of considerable non-linearities. Some works have explicitly addressed the more specific selection of plausible shapes thereby taking non-linearities into account. In [80], a gaussian mixture model was proposed to estimate the probability density function of shapes in the parameter space of $\vec{b}$. Similarly in [81], a hierarchical PDM based on gaussian models was proposed. In [82], a tabular constraint of shape parameters was employed based on the previously encountered training data.

Alternatively, non-linear shape deformations can directly be modeled in a non-linear shape prior, and different propositions have been made to this end in the literature. In [83], non-linear PCA was employed based on polynomial regression (PRPDM), and later using a multi-layer perceptron (MLPPDM) [84]. In [85], a highly-expressive non-linear modeling was proposed based on kernel transformation (Kernel PCA).

Another important aspect of shape prior modeling concerns the range of the shape deformation during a change of modes. In the linear PDM based on the PCA, each mode of variation generally affects shape on a global level (*global support*). However for many applications, a description of purely local shape variations can be desirable (*local support*), e.g. for the representation of pathological alterations at dedicated organ subparts. For the principal components of the linear PCA, this means that they are desired to have a delimited effect on local landmark clusters (*sparsity*).

For linear shape prior models, this can be achieved through redefinition of the principal components, e.g. by the orthogonality preserving Orthomax rotation of components [86], or approximately through quasi-orthogonal redefinition as proposed in [87]. In [88], modes were selected towards a maximization of autocorrelation factor (MAF).

Alternatively, an independent component analysis (ICA) can be applied for non-linear shape prior modeling [89]. There, a mixed distribution of non-gaussian random variables is decomposed in its independent components after decorrelation of input signals based on the inverse correlation matrix, and e.g. through an iterative rotation towards maximization of non-normality in the marginal distributions [90]. It can be shown that ICA and MAF are equivalent. An example application can be found in [91], [92] for pattern description in cardiac MRI. Since no orthogonality is established on the components, additional measures have to be taken during an iterative segmentation scheme in order to find an optimal fit to potential landmark locations. Also naturally, shape selection from the modeled distribution cannot simply be made based on principal component variance alone.

2.3.4 Model Training

For 3D-SSM, robustness during segmentation results from the specificity of the employed shape prior model. Segmented surfaces are constrained to plausible shapes and thus excluding ill-shaped delineation errors. However, specificity can also be accompanied by a loss of generalizability to newly encountered shapes.

Often, the problem arises when only little training data is available due to the high effort for manual annotation of training cases. Then, little information on shape variation is provided for suitable application in future cases. To tackle this problem, an artificial augmentation of training data can be performed,

where plausible shape variation is added to the underlying data.

The prerequisite is to only impose plausible shape variation during augmentation. In [93], plausible shape variation is generated using vibrational modes that model physical deformation from a finite element model. In [94], a simpler alternative is proposed where additional variation is added landmark-wise in the covariance matrix of the training samples. To prevent implausible variations of isolated landmarks, covariance among neighboring landmarks is increased simultaneously. This approach was later adopted by [95].

In [96], non-rigid local deformations of surface landmarks were proposed. Plausibility was ensured through topology preserving landmark shifts and through a rejection of implausible deformations that are not in accordance with a probability map constructed from the original training data. In [97], different approaches were evaluated regarding their effect on the generalizability of the resulting models. It was shown that overall best results were achieved using non-rigid deformations and the FEM approach from [93].

Another idea to tackle the problem of limited training data is to decompose the modeling of a global shape to smaller local shapes of lower complexity. The overall decreased complexity of the subparts then promotes a modeling on smaller training set sizes. One illustrative example for this is the decomposition of tubular structures in a cross-sectional part and in a part that describes variations of the tubular axis itself. This was proposed in [14], [98] for the modeling of aortic aneurysms where higher generalizability was achieved.

Alternatively, in [55], a separate modeling of different surface parts was proposed. Implausible shape variation was then prevented by a regularization of the parameter curves for the different parts based on their similarity to parameter curves observed in the original training data.

Summarized, techniques that increase the expressiveness of the shape prior modeling during training show great promise if limited training data is intended to be used to a greater extent. However, a possible loss of modeling specificity has to be considered when shape variation is artificially induced that is not explicitly backed by underlying training data. This is important since specificity is what provides most part of the robustness while using shape priors during segmentation.

Altogether, there is no general recommendation for regular use of these techniques, and their engagement remains a design choice to be considered separately for different clinical applications. In case of modeling scenarios where complex geometries are composed of simpler subparts, a use of the presented decomposition techniques generally seems to be advisable.

2.3.5 Summary

A large variety of methods and strategies exist that tackle the problems that arise during shape model training. They cover a multitude of conditions that can be encountered in different applications, regarding object geometry and topology, availability and representation of training data, assumed shape distribution and anatomical variability. Approaches have been proposed that induce modeling optimality for the later application of the trained models. Particularly, they build the foundation for a shape prior modeling that can provide additional segmentation robustness in 3D applications.

Some of the proposed techniques have been developed for general use in differing applications, whereas others focus on more particular use-case scenarios. Altogether, the presented techniques largely rely on an explicit modeling of object landmarks, which makes them compatible with various algorithms that were developed on this basis. This also means that methods for object appearance modeling of for model fitting optimization can in large parts be developed independently from the previous shape prior modeling. Their notions can then easily be extended to additional scenarios that profit from particular shape modeling techniques.

While the role of the shape prior model is to add constraints to the hypotheses that are found in the segmentation process, appearance prior models are employed to provide meaningful candidates for this process by searching the image domain. The challenges and techniques that arise from the task of appearance prior modeling for 3D-SSM will be presented in the following section.

2.4 Appearance Prior Modeling

To generate plausible hypotheses during segmentation that match sought objects in images, information on object appearance is learned from annotated data during 3D-SSM training. This follows the greater top-down bottom-up approach of 3D-SSM, where higher-level modeling information from the learned shape and appearance priors is linked with low-level image information during segmentation to provide higher robustness.

For a successful modeling and learning of object appearance to this end, different strategies exist that describe object characteristics in the underlying image data. In general, they yield benefits for different objects and modalities, depending on the actually encountered appearance of the considered objects in

the images. Also, their choice can have a strong influence on the applicability of different search and model fitting strategies during segmentation.

The previously trained models are directly employed for an estimation of plausible model placements during segmentation. Generally, hypotheses are either generated through search of plausible solutions in the image domain, and then rating these by their posterior probabilities in the previously modeled distributions. Or, a generative model is trained for the creation of hypotheses that are fitted to the underlying image during segmentation, in order to determine the most plausible solution.

Existing concepts for appearance modeling can roughly be categorized by the actual object characteristics or parts they actually describe. A first type builds on an exclusive description of object boundaries, i.e. contours in 2D and surfaces in 3D domain. This type probably constituted the first type of appearance modeling used for object description and segmentation in image data. Due to its focus on outer boundaries of the considered objects of interest, these appearance models naturally support a search for the visible object delineations during segmentation. The boundary-based appearance models will be detailed in Sect. 2.4.1.

Another way to describe appearance of objects is based on their texture. In contrast to the boundary-based approaches, texture-based appearance models consider information from the entire interior region of the considered objects. This goes naturally well with objects that involve characteristic recognizable features or distinctive texture patterns. Since these approaches generally involve information from within the entire object, they can make use of this information in terms of a global constraint for texture search during segmentation. The texture-based appearance models will be detailed in Sect. 2.4.2.

More recently, an object description found entrance to image segmentation tasks that is based on capturing the characteristics at distinct object landmarks. In contrast to the boundary- and texture-based approaches where discrete landmark positions are used as sample points for boundary and texture appearance of the described objects, those modeling approaches focus on the distinct description of these separate object landmarks. This approach is theoretically well suited for an unambiguous detection of the modeled landmark positions during segmentation, which should have significant implications on model-based segmentation in general. However, these implications have not yet been investigated and applications remain largely limited to selected objects in 2D applications. Previous work on landmark-based appearance models will be presented in Sect. 2.4.3.

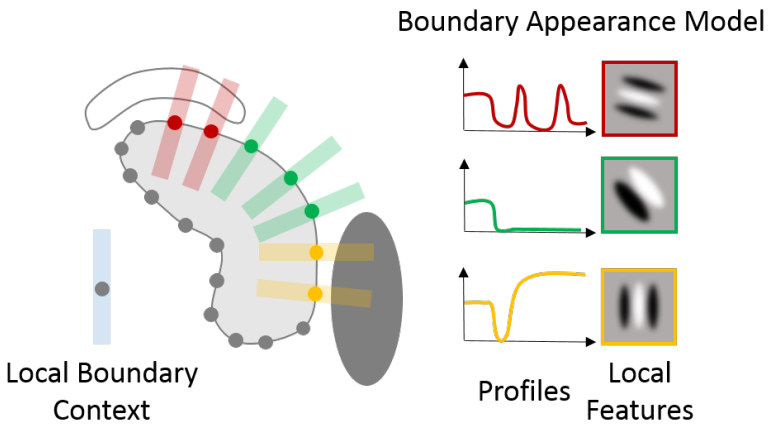


Figure 5: Boundary-based appearance modeling. The modeled objects of interest are described by their local boundary appearance. Local information is encoded in directed linear profiles and in feature response vectors from multi-scale oriented wavelet filter banks.

2.4.1 Boundary-based Modeling

Various approaches have been proposed for an appearance modeling of object boundaries in image data. They comprise simple and more complex feature sets for the description of boundary appearance. They make different assumptions on the underlying sample distribution of the modeled appearance, and they employ a variety of methods to estimate posterior probabilities of possible boundary candidates during segmentation.

Originally, a boundary appearance modeling based on intensity and gradient profiles was proposed in [99], [100]. Boundary appearance was modeled by a gaussian distribution of sampled profiles from training data. To estimate posterior probabilities of new samples during segmentation, the Mahalanobis distance has been shown to produce favorable results [100]. Also in these works, normalization of the profiles to unity scale was proposed in order to compensate for intensity fluctuation from illumination changes. The normalized profiles have also been shown to deliver the best results in the presented applications.

Instead of a separate modeling of landmark appearance, a boundary description across all landmarks via concatenation of landmark-wise profiles was proposed in [101]. It was assumed that separate profiles are not generally uncorrelated, which can better be captured by the concatenated profile model. In conjunction with a more complex quasi-newtonian optimization of landmark

candidates during segmentation, more accurate results were achieved in the presented scenario. However, the proposed optimization strategy was accompanied by slower execution times, and the chosen boundary representation did not allow an assessment of the local quality of fit apart from the employed global cost function.

A more extensive modeling was proposed for a 2D-SSM in [102], where a gaussian mixture model was used based on a large set of Gabor wavelet features (called *jet*). Features were used in various scales and rotations, thus providing a rich description of boundary appearance. Distribution of the local jets was estimated via expectation maximization. In [103], the use of rotation-invariant Gabor features was proposed for use in 2D-SSM, to compensate for the rotational field of view of the employed ultrasound probe. In [104], steerable filters were proposed for an appearance modeling in 2D-SSM. The employed steerable filters represent another alternative for a rich set of features, which in addition offer an efficient evaluation of rotated versions of the provided features. Canonical correlation analysis (CCA) was used to model the local landmark appearance.

In [105], it was reasoned that linear modeling of landmark appearance was insufficient. Therefore, a k-nearest neighbor classifier was proposed that estimates posterior probabilities from a non-linear distribution of foreground and background profiles. Higher accuracy was demonstrated in this approach. Later, further non-linear models were proposed using steerable filters [106] and profile histograms [82]. Also in contrast to previous work, a stronger AdaBoost classifier was proposed for the automatic selection of meaningful features from the larger feature sets, and for the aggregation of weak classifiers to a stronger overall classifier.

The aforementioned approaches have in common that they define features and feature sets once for use during training, and that they do not preselect any features from these sets to be excluded or to be focused on during training. In contrast, an additional feature selection was proposed by other groups, where prior knowledge is used on meaningful features for specific applications. Selection of useful features was proposed via boundary likelihood tests [107] and via simulated search of landmark positions [108]. In [109], [110], heuristics were used based on typical intensity profiles that are encountered during liver segmentation in CT volumes. Obviously, this kind of feature selection requires considerable extra effort and is highly use-case dependent. On the other hand, notable performance was achieved through the extra effort in a particular application [110].

One problem that typically affects classifier-based profile modeling is the small number of profile samples available at each surface landmark. To solve this problem, a clustering of training samples can be applied. Larger numbers of samples are concentrated in each cluster and thus made available to each surface landmark. Different propositions were made as how to cluster samples and how to assign landmarks to clusters. In [111], a surface-wide fuzzy clustering with c-means was proposed. Assignment of landmarks from chosen boundary sections to eligible clusters was done manually. In [112], a clustering with k-means was proposed, where each landmark was assigned one cluster that holds the most training samples from that landmark. Alternatively, assignment to that cluster which produces the best accumulated fitting costs for the considered landmark was proposed in [113]. Another way of clustering was presented in [114], where spatial relationship between surface landmarks was considered. In the process, profiles were blurred across the corresponding boundary sections.

Summarized, boundary-based appearance modeling provides a natural and intuitive means of object description regarding segmentation tasks. It means a direct modeling of object parts that need to be delineated during segmentation. The modeling type naturally corresponds to a linear search for object boundaries, where the problem of object detection and delineation is reduced to identifying the most likely intersection point of search lines with the sought object boundary. This formulation of the segmentation task has proven highly effective in past applications, and various boundary descriptors have been proposed that can provide rich information for the delineation.

Despite their popularity, the boundary-based descriptors also exhibit some significant disadvantages. First, the focus on boundary appearance usually means that useful information can only be found in the vicinity of the boundary. Thus, detection quickly fails for boundary parts that are out of reach during segmentation (capture range problem). Furthermore, boundary-based descriptors do not distinguish between neighboring landmark positions on the same object boundary, which allows ambiguous results during segmentation. In addition, the above intersection points of the search profiles with the object boundaries continuously change during segmentation, depending on the origin of the search profiles on the model surface of 3D-SSM. This regularly leads to unforeseen segmentation errors. However, due to the simplicity and the effectiveness of the boundary-based appearance models, they are probably the most widespread type among 3D-SSM that are used in image segmentation tasks.

2.4.2 Region-based Modeling

In contrast to the use of boundary-based features, propositions were made to model object appearance based on regional characteristics. They involve features that are suitable for a description of spatial intensity patterns. Due to the higher dimensionality of the considered spatial domain, a limitation of the complexity of these features naturally becomes important. Also, regional intensities tend to be more sensitive to unforeseen intensity variations. Therefore, suitable measures for spatial intensity normalization have to be addressed, if additional robustness is desired.

In [115], [116], an intensity vector was sampled from a reference frame inside the modeled object. All training shapes were first warped to their mean using a triangulation algorithm. Intensity vectors were then sampled from the warped training images in a predefined region covered by the mean shape. A normalization of the sampled vectors was proposed through projection to the tangent space of the mean vector. Object appearance was modeled by a normal distribution of normalized vectors, using the Mahalanobis distance for posterior estimation during segmentation. As an alternative to the linear tangent space normalization, a non-linear intensity normalization was proposed in [117]. A global non-linear transformation was applied to remove shift and skewness from encountered intensity distributions to approximate normal distribution.

In [118], the use of a regional description based on direction and strength of gradients was proposed. The previous intensity vector was replaced by a gradient vector where gradients had undergone a previous non-linear normalization. Main motivation was that the use of a gradient-based representation is generally more robust against illumination changes. The approach was shown to be beneficial only in the presence of changing illumination. In [119], the gradient based description was extended by additional descriptors involving edge and corner strength, which lead to improved results in the presented scenarios. It was assumed that additional descriptors might be of further benefit.

In [120], a rich feature set known from texture analysis was proposed to describe the appearance of patches around landmark positions in 2D-SSM. For the feature descriptor, histogram moments from various filter responses were considered, produced by a multi-scale gradient filter bank. From the rich set of features, a landmark-wise selection of optimal features was performed by feature forward and backward selection. Search for suitable landmark positions was conducted through a meaningful separation of classified foreground and background voxels by the employed surface model. In addition to the automatic selection scheme, complexity of regional description was reduced to

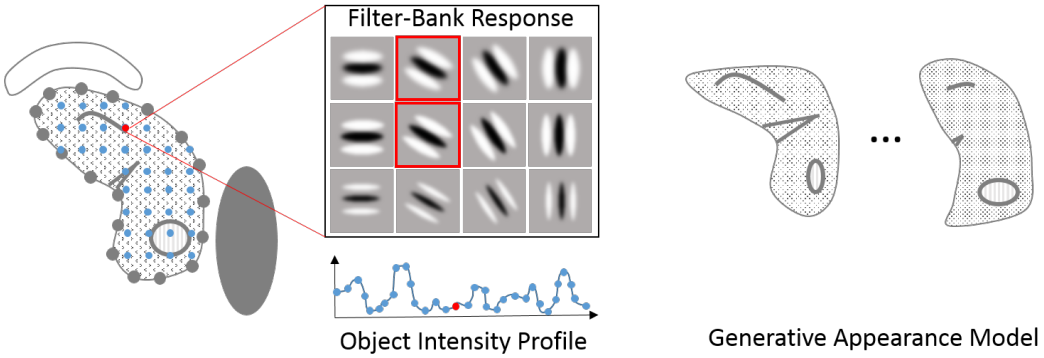


Figure 6: Region-based appearance modeling. The modeled objects of interest are described by their global or local regional appearance. Information on texture and on salient features of an object of interest is encoded by global intensity and gradient profiles, and by use of multi-scale oriented wavelet filter banks. A generative model of object appearance can be trained to generate joint solutions of shape and appearance during model fitting in unknown images (Active appearance model approach).

patches around the modeled surface landmarks. Additionally in [121], the use of Cartesian Differential Invariants was proposed to reach rotational invariance of the employed feature set. In [122], the use of a Gabor filter bank was proposed for appearance modeling of the prostate in 3D ultrasound images. To reduce complexity, 2D filters were applied in orthogonal view planes of the volumetric data. Locally trained support vector machines were used for a classification of foreground and background voxels to guide a 3D-SSM during segmentation. In [123], the use of wavelet and wedgelet features was proposed for global texture model adaptation.

To reduce complexity of the modeled object texture appearance, image resampling has shown to be effective [117], [124], [125]. Other methods were proposed to reduced complexity in comparison with texture sampling. In [126], regional description by an object-wide intensity histogram was proposed instead of a voxel-wise feature sampling. Quality of fit between appearances from the segmented region and from the previously learned appearance was calculated by the L_α distance of the histograms' cumulative distribution functions. In [127], an additional weighting of voxels was proposed depending on their distance from the object boundary. In [128], appearance modeling was reduced to texture patches around previously identified feature points.

In [129], liver tissue appearance in CT images was modeled using 2D Haar-like features - i.e. a large set of randomized box integral features - in addition to a conventional profile-based boundary modeling. Learning from training data was performed by a cascade of AdaBoost classifiers, and classifier outputs were later used in a level-set function for model-adaptation. In [130], Haar-like features were proposed to describe regional appearance for prostate, bladder and rectum in 3D CT. In addition, auto context classifiers were used to increase robustness. That is, a cascade of classifiers is trained where each classifier learns from the output of its predecessor in addition to training samples from image domain. Finally, a boundary regression was proposed using 3D-Haarlike features and random forests. The trained forests were used to estimate organ boundaries from image positions determined by the results of the trained classifiers. A deformable surface model was then employed to perform segmentation of the estimated organ boundaries.

Summarized, region-based appearance models have been shown to successfully incorporate information more globally, i.e. from larger object parts, in contrast to the boundary-based approaches. Unlike other 3D-SSM approaches, they allow a combined optimization of shape and appearance during segmentation as done by the presented active appearance model approaches. Also theoretically, these approaches are not necessarily subject to the problems of boundary-based approaches where results depend from the visibility of linear search scopes during segmentation. This is because active appearance models were basically designed to directly learn optimal displacements for later segmentation using a regression approach. But this model adaptation scheme is only applied by the presented active appearance models and not by other region-based approaches.

In general, applications benefit the most from the region-based approaches if confined objects with characteristic inner features are considered for segmentation. Consequently, an application of these approaches largely remains limited to certain organ segmentation tasks [2], like prostate, left cardiac ventricle and hippocampus segmentation in MRI and ultrasound data. In contrast, region-based approaches are rather rarely encountered in the more general image annotation tasks, i.e. apart from specialized use in certain clinical scenarios.

2.4.3 Landmark-based Modeling

In more recent approaches, an appearance modeling at distinct feature points or landmarks was proposed, instead of an appearance modeling at boundaries

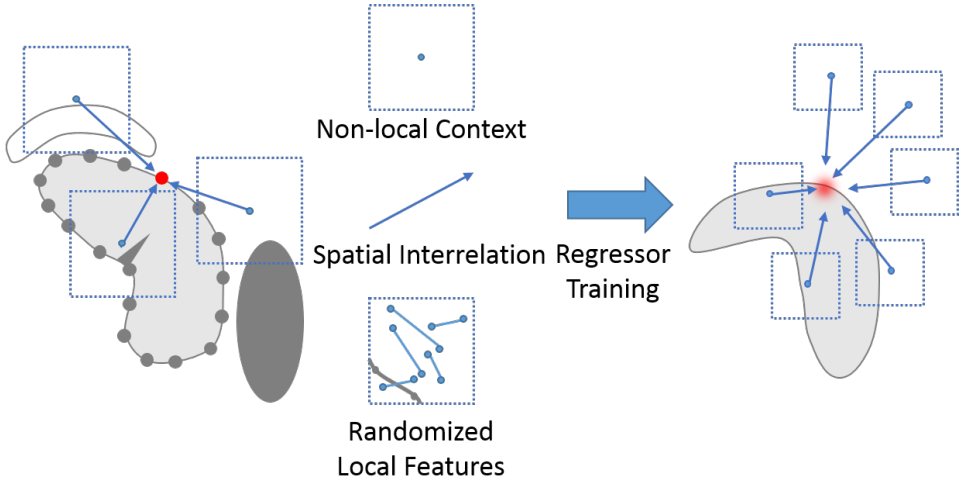


Figure 7: Landmark-based appearance modeling. The modeled objects of interest are described by the distinct appearance of their surface and salient feature landmarks. Non-local information is gathered around the landmark positions by means of flexible, highly randomized features. A regressor can learn the spatial interrelation of appearance around the modeled landmarks from these features.

or foreground and background regions. They involve features that produce a unique response at separate landmark positions. This is in contrast to boundary and region based features that do usually not disambiguate adjacent landmark positions, e.g. at boundary sections of similar appearance.

Often, large sets of randomly generated features are used to produce a localized response at specific key points. A popular randomized feature descriptor was introduced in [131], where the use of *Haar-like features* was proposed. Feature response was generated by subtraction of two box integrals of random size and position within a centered patch at a given image position. A high expressiveness of the resulting feature descriptor was reached thanks to the rich sets of randomly generated features. Also, fast feature evaluation in constant time was enabled by using precalculated integral images as a lookup table for the calculated box integrals.

In [132], the high specificity of randomized 2D Haar-like features was shown for description of facial landmarks. A highly localized feature response was produced at the examined facial landmark position, in contrast to the compared profile- and Gabor-filter based descriptions. An Adaboost classifier was used for automatic selection of meaningful features from the random feature set, and for an estimation of posterior probability of the sought landmark position during segmentation.

Alternatively in [104], steerable filters were proposed for an appearance modeling of facial landmarks and of hand-bone contour points in 2D. A method called canonical correlation analysis (CCA) was used to model local landmark appearance and to control model adaptation during segmentation.

In [133], [134] an appearance modeling based on 2D-Haarlike features was proposed for the left ventricle in 2D ultrasound images. A learning technique called image based boosting regression was employed to both predict an initial bounding box around the left ventricle (LV), and to predict LV shapes that deliver plausible segmentation results. In [29], [135], 2D Haar-like features were employed for a landmark-wise appearance modeling of the femur in 2D radiographs. High segmentation accuracy could be achieved through a landmark detection using random regression forests. In [136], an image driven model adaptation scheme was proposed for the segmentation of bones in 2D radiographs. There, global solutions could be provided for the detection of multiple landmark positions through an optimization of a convex cost function. Also, feasibility was shown for test cases of bone segmentation in CT volumes. However, no efficient and accurate solution could be presented for 3D domain at that time.

In [137], randomized 3D Haar-like features and a feature description based on voxel-wise image intensities and gradients was proposed for kidney segmentation in CT volumes. After a rough bounding box detection of the kidney using random regression forests, a refinement of bounding box position and a final kidney delineation by a deformable ellipsoid was performed. In this approach, 3D Haar-like features were only involved during the rough pre-localization of the kidneys.

In [138], the use of another random feature descriptor using *random ferns* was proposed for the description of facial appearance in 2D images. A (*fern*) is a set of random features that compare image intensities at pair-wise voxel positions. Classification and regression tasks can then be learned and be performed efficiently using a semi-naive version of Bayes' theorem. Depending on the chosen size for the ferns, conditional probabilities can efficiently be determined on the generated features. The method was used for an iterative shape regression of facial shapes during segmentation. In contrast to forest-based techniques, random ferns allow a parallel evaluation of all features at runtime, which enables a particularly fast inference. On the other hand, the presented technique relies on weak regressors, which makes measures like cascading necessary in order to boost performance [139]. Also, efficiency and performance of the random ferns depend on size and number of the ferns, while the expressive-

ness of the ferns essentially relies on an extra selection of meaningful features during training. Altogether, this gives rise to an additional optimization of the training process, and measures for a meaningful selection of features for the ferns have to be employed for the actual task at hand. Different selection techniques have been proposed based on regression error minimization [140], [141], or on a maximized correlation of the selected features with the regression target in conjunction with a minimization of inter-feature correlation [138].

2.4.4 Summary

Propositions from the presented boundary-based, region-based and landmark-based appearance modeling strategies have so far revealed some insights in the strengths and weaknesses of the different types. In summary, boundary-based appearance modeling is the most widespread type that is used for segmentation in a large variety of use-case scenarios. Its focus on a boundary-based object description fits most applications, and weaker basic assumptions are being made in contrast to the region-based approaches. There, benefits are probably limited to cases with pronounced presence of inner object details. Furthermore, the simplicity of the modeling and search schemes for the boundary-based type notably facilitates its direct applicability in new applications. Finally, the availability of rich feature sets like the proposed Gabor filter banks has been shown to be applicable even under difficult imaging conditions as encoutnered e.g. in ultrasound segmentation.

However, there is a strong downside of the boundary-based modeling approaches regarding image segmentation, as detailed in Sect. 2.4.1. That is, the spatial ambiguity of the used boundary descriptors and the corresponding linear boundary search make the resulting 3D-SSM highly sensitive to changing conditions during segmentation. Changes of model pose that continuously occur throughout the iterative segmentation process, and the choice of the initial model placement prior to the segmentation greatly affect final results. Consequently, the course of the segmentation is highly unpredictable, and unexpected errors arise on a regular basis.

The presented landmark-based appearance modeling approaches do in theory offer the means to cope with these well-known problems that are associated with the most widespread type of appearance modeling in 3D-SSM. Their focus on a description of object appearance in terms of distinguishable landmarks provides the kind of spatial distinctiveness that is required for an unambiguous and reproducible detection of different object parts during seg-

mentation. However, previous work has so far focused on other, single aspects of this relatively new type of appearance modeling. High accuracy for bone segmentation, global optimization of landmark search in 2D domain, and 3D soft tissue segmentation using 2D feature descriptors were shown. While the proposed concepts were so far limited to 2D domain or bone segmentation, most importantly, possible benefits regarding the aforementioned problems of previous 3D-SSM have not yet been examined.

2.5 Image Segmentation with 3D-SSM

During segmentation, top-down information from the learned shape and appearance priors and bottom-up information from an evaluation of underlying image features are combined in order to obtain meaningful results. To this end, a surface model from the previously learned shape distribution is fitted to the sought organ in the underlying image, while the learned appearance prior is used to determine the posterior probability of found solutions with regard to their appearance in the image.

Essentially, the model fitting procedure can be considered an optimization problem where shape and appearance posteriors are to be maximized in order to obtain an optimal solution of the segmentation problem. Different strategies exist for this optimization, and according propositions have been made in the literature.

Generally, optimization is performed in two steps: a model initialization and a subsequent model fitting procedure. During the model initialization, an approximate solution is determined where pose and sometimes shape parameters of the sought target organ are roughly identified prior to the actual segmentation. From this initial localization of the organ of interest, suitable boundary conditions can be provided for the subsequent model fitting procedure.

The model fitting can then either be performed in an iterative optimization scheme where an optimized fit of the employed surface model to the underlying organ is obtained through stepwise evolution of pose and shape parameters. Or, an instant solution can directly be derived from the initial model placement. Methodically in contrast to the iterative model fitting, no update of boundary conditions is performed during this approach.

In Sect. 2.5.1, the task of finding a suitable model initialization for the subsequent segmentation will be outlined together with existing strategies in

the literature. Sect. 2.5.2 will then further elaborate on the different challenges and solutions for a robust and accurate model fitting.

2.5.1 Model Initialization

To provide a suitable model initialization for the subsequent model fitting prior to segmentation, the actual organ of interest must first be localized in the underlying images. Usually, the position of visible organs in these images is unknown in advance, and significant anatomical variation and positional changes from breathing motion or varying body posture can be encountered. In addition, images often exhibit varying fields-of-view. Sometimes, pathologies and previous treatments cause large changes of patient anatomy.

In addition to a mere localization of considered organs, the model initialization step must also provide an accurate model placement for 3D-SSM to increase the chances of success for the subsequent model fitting. As described previously, 3D-SSM are in general highly sensitive against the choice of their initial model placement, due to the ambiguity of the employed unidirectional search during segmentation.

Often in selected scenarios, an organ detection is only required for one particular organ of interest. In these cases, a model initialization can be set up with a focus on that organ. The resulting initialization methods rely on common basic assumptions about the general organ position and size in the images at hand, and about intensity distribution of the underlying imaging modality. Typical examples are the works from [142]–[145], where techniques like thresholding, morphological operators, voxel counting and connected component analysis are employed.

In other works, the use of more generic methods for a detection in these particular scenarios was proposed. The methods conduct an automatic search for organ location that is less dependent from limiting basic assumptions. This makes the resulting model initialization generally more robust, and allows an application in other segmentation scenarios. In [110], the Generalized Hough Transform [146] was proposed to conduct an exhaustive search for a suitable model placement in the image volume.

Another example is the use of an evolutionary optimization algorithm to determine a suitable model placement as proposed in [147]. Shape parameters and rigid transformation parameters of a 3D-SSM were randomly altered and then selected based on a fitness function that takes into account estimates from a previously trained appearance model.

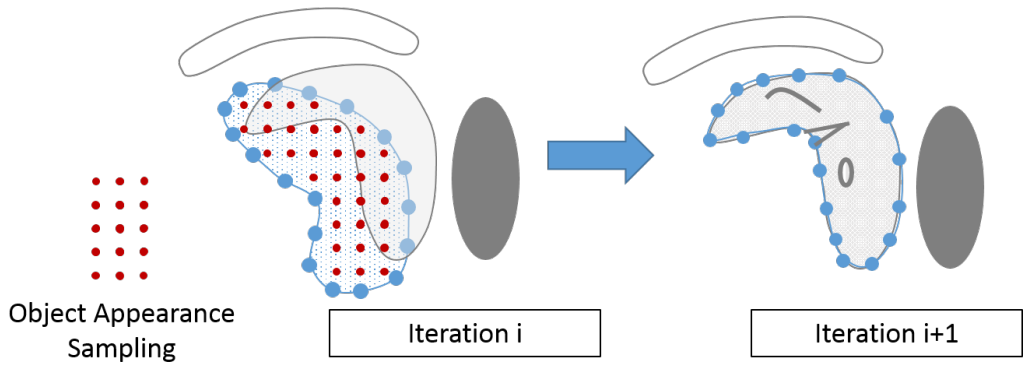


Figure 8: Iterative model fitting via active appearance model search. Pose, shape and appearance parameters are estimated via regression from sampled residuals in the image. A joint generative model of shape and appearance is fitted to the unknown image, in order to find a meaningful estimate for the encountered object of interest.

Despite their higher robustness, an application of these methods is still limited to certain organs and use-cases. Also, these methods are computationally expensive, often, they consume the same amount of time as the subsequent segmentation. In this regard, the use of machine learning can be a better solution. According methods have been shown to provide robust detection of various organs in the use-cases they were trained on, and they are known to be computationally fast. In [129], the classifier output that was also used for appearance model training served as a basis for a detection of the liver in CT volumes, by selecting the largest region in the thresholded classifier output for the whole image. A cascade of boosted CART classifiers was used for this purpose.

Instead of an entire scanning of the image by a trained classifier, the use of a regressor was proposed in [133], [134] to perform a pose estimation of the left ventricle in 2D ultrasound images. A number of regression estimates is performed on different bounding boxes to determine position and size of the shown ventricle. Multiple regression estimates from randomly altered bounding boxes are averaged to obtain the final pose estimate of the ventricle.

Another regression-based approach for single organ detection was proposed in [135]. Detection of the femur in 2D radiographs was performed by the trained regressor, but instead of a bounding box detection, an initial position of selected landmarks on the femur contour was estimated in the images.

Although being fast and robust, an applicability of the above detection and initialization approaches was limited to single organ detection in smaller 2D images, and no solution was presented for larger image volumes where greater variation of anatomical context and possible organ locations is more often encountered.

Other propositions have addressed the task of a more general detection of multiple organs in larger image volumes. In [148], [149], Marginal Space Learning (MSL) was proposed where probabilistic boosting tree classifiers were employed for bounding box detection in CT volumes. To reduce the computational cost for a repeated scanning of the entire image volume using several trained classifiers, a limited search in the marginal spaces of the cascaded classifier outputs was proposed. The approach was later also adopted for a prediction of organ shape in addition to the bounding box location.

In [150], for the same task, an organ classification approach was proposed using a faster and more accurate version of decision forest (*Entangled Decision Forests*) which takes contextual information into account during training and application. The approach was even faster than MSL, but unlike MSL, it did not involve shape prior to add robustness against labeling noise. Similarly to the single-organ case, the use of regressors was proposed in [151]–[153], to profit from reduced scanning effort and from improved accuracy for detection of multiple anatomical structures in the larger image volumes.

In [154], the use of additional information on organ shape and location interrelation in a probabilistic atlas was proposed to improve detection and segmentation. An overall improvement of performance could be achieved with this approach, and in addition, the segmentation of heterogeneous images was supported thanks to the incorporation of higher-level information. Robust and accurate placement for the employed statistical shape models could be provided using the proposed contextual information. However, no conclusive results could be produced by the employed 3D-SSM, and a graph-cut method had to be applied in order to retrieve an accurate segmentation.

Summarized, various propositions have been made for organ detection and model initialization at single or multiple organs of interest. For isolated clinical applications, specific solutions are often preferred that are based on techniques from basic image processing, due to their high availability. However, high-level methods have also been proposed for that purpose, based on robust classifier and regressor approaches. For multi-organ segmentation tasks, such high-level solutions are obviously the only possible choice. The proposed classification- and regression-based strategies have proved to be particularly fast. The use

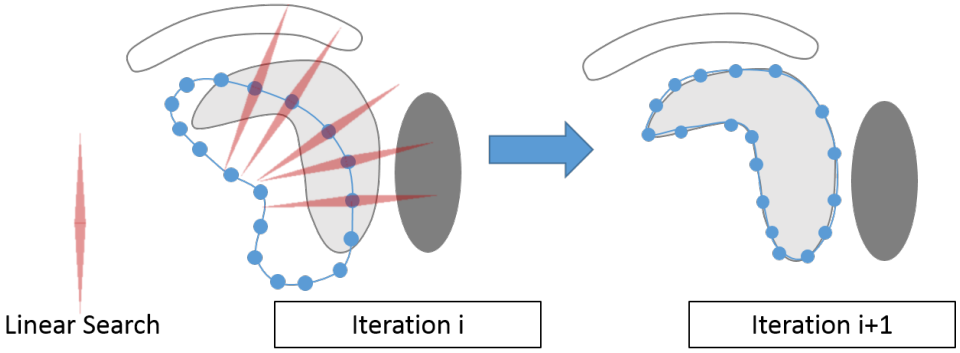


Figure 9: Iterative model fitting via profile-based search. Pose and shape parameters are estimated from landmark-wise displacements as provided by the linear search. A discriminative model of boundary- or region-based appearance provides plausible displacements within the linear search scopes.

of combined classification and regression approaches and of contextual information from shape and location interrelation have been shown to produce a robust and accurate organ detection. However, it has not yet been possible to provide a model initialization that could serve as a basis for an error-free subsequent application of 3D-SSM. In this regard, better insight can be gained from an understanding of the nature of the model fitting and its interplay with the previous initialization of 3D-SSM.

2.5.2 Model Fitting

From an initial placement of the surface model, the final adaptation to the organ of interest is carried out during model fitting. The basic task is to determine an optimal model placement that maximizes the learned posterior probability regarding shape and appearance of the modeled organ of interest. Essentially, this optimization can be conducted in different target domains.

One way is to optimize model parameters directly as it was proposed for active appearance models in [115], [116]. There, a regressor is trained to predict adjustments to shape and appearance parameters in order to minimize the residual prediction error. Automatically in such an approach, shape constraints are always imposed throughout the model fitting process. However, this also

means that an introduction of additional freedom from the learned shape prior is not immediately possible.

Another strategy is to optimize model parameters indirectly based on a previous detection of suitable landmark positions on the sought organ surface. This optimization scheme corresponds to the original formulation of statistical shape models in [3], and has since then remained the most widespread optimization strategy for shape model fitting (cf. [155] for an comprehensive overview of applications).

Here, a meaningful shape constraint is imposed after a separate landmark detection step by choosing parameters that minimize the residuals to the found landmark candidates. Since this optimization scheme is based on a separate detection of landmark positions, it is not solely bound to an adjustment of model parameters and additional techniques can be introduced that allow an extra deformation of the surface model to increase freedom of the model adaptation process. This often becomes necessary if the encountered shape variation exceeds the variability learned from limited training data.

Additionally, the use of landmark-wise detectors opens the possibility of treating landmarks differently during the model fitting across the modeled surface e.g. based on reliability and confidence of the detector. Also, an intuitive qualitative inspection of results is easily possible in this approach.

On the downside, a separate detection during the model fitting is also accompanied with a number of challenges. On the one hand, a time-efficient search strategy is required. The standard procedure for this is a linear search perpendicular to the model surface, which can lead to the problem of missing landmark positions at surface parts that are not covered during the linear search. As a result, the outcome of the model fitting becomes dependent from the quality of the previous model initialization.

Consequently, the need for highly accurate and complex initialization methods arises. A possibility to compensate for this problem is to actively steer the model fitting process as in [144]. There, modeled surface parts are treated differently using a rule-based system. It could be shown that the sensitivity against model initialization can be greatly reduced this way. However, this procedure is accompanied by a large number of free parameters and by significant fine tuning, which hampers the general applicability of this approach.

As an alternative to the limited profile-based search, another detection strategy was proposed for 2D bone segmentation using a landmark detection based on a trained regressor [29], [135]. Apart from the high accuracy that was achieved for the bone segmentation, the possible benefits regarding the visi-

bility problem and the dependency from previous model initialization remain unclear. However in [136], it was furthermore demonstrated that a globally optimal solution could be derived for 2D bone segmentation based on the formulation of the detection problem as a convex optimization task.

Another problem with approaches that rely on landmark detectors is that they do not necessarily guarantee well-defined landmark displacements during the model update process, and a global selection of landmark positions can be required that is meaningful across larger parts of the surface model. Also, consistency has to be preserved across the surface model if additional freedom from shape constraints is allowed during the model fitting.

To this end, additional measures were proposed in the literature. For example in [147], an additional free shift of landmarks perpendicular to the model surface was employed through a shape-constrained deformable model. There, a surface-wide regularization of the linear displacements could be imposed through a graph cut based optimization. In [156], the freedom of the shape-constrained model deformation was further extended by the introduction of omnidirectional landmark shifts. For regularization, a Markov Random Field optimization was proposed using a fast primal dual solver. Despite the use of an extra regularization, the appearance of surface inconsistencies during the free model fitting could not always be guaranteed.

A third possible optimization strategy during the model fitting is to directly optimize on the landmark positions themselves without taking shape parameters into account (shape regression). To this end, a regressor is trained that predicts shapes from varying model positions around the assumed organ of interest. This means that landmark displacements are learned surface-wide between the shifted positions and the target surface. Therefore, the predicted landmark shifts during segmentation are assumed to be well-defined as long as inconsistencies do not enter the model fitting process at any point. Furthermore, joint shape predictions have shown to be computationally efficient e.g. by using fast parallel predictors like random ferns [138], [139].

Instead of varied model positions around the sought organ, a regressor can also be trained from a variety of image patches that contain the considered organ of interest as proposed in [133], [134]. In any case, landmark predictions are made globally on a per shape basis, which means that no immediate influence is possible on selected landmarks. Also since predictions are exclusively determined by the overall outcome of the regressor, detection results cannot be interpreted as intuitively as through separate landmark detectors, and cannot be directly counteracted on a per-landmark basis.

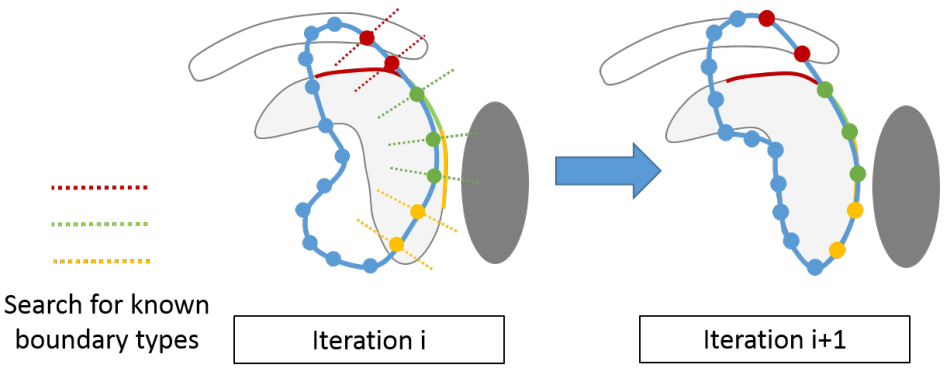


Figure 10: Visibility problem for boundary-based appearance models. From an unfavorable model pose, the typical unidirectional search often fails to provide valid landmark positions during model fitting. An ambiguous boundary appearance allows landmark shifts on the target surface, and landmarks end up in false positions, producing locally optimal solutions.

2.5.3 Visibility, Capture Range and Local Optimization

Regarding the optimization process that 3D-SSM carry out during segmentation, certain basic limitations arise that have a strong repercussion on performance and applicability of 3D-SSM in general. Depending on the chosen model initialization, 3D-SSM are highly prone to be stuck in local optima at some point of the subsequent segmentation. The dependency of results from the previous initialization has been reported regardless of the underlying model adaptation strategies, i.e. for separate detection of landmark positions and for regression-based pose and shape estimates.

Since initial conditions have an irreversible effect on the outcome of the subsequent search, it often means that errors have already taken place prior to the segmentation process. These sources of error are hard to anticipate and to reconstruct from the highly unpredictable course of the subsequent model fitting. Altogether, increased uncertainty and a constant source of segmentation error is caused by the described interplay between model initialization and subsequent model fitting.

Approaches for a general segmentation of various organs strongly build on the performance of the first model initialization step [148], [149], [154], whereas the typical limited search is performed during the subsequent model fitting. With an increasing sophistication of the employed techniques, a notable degree

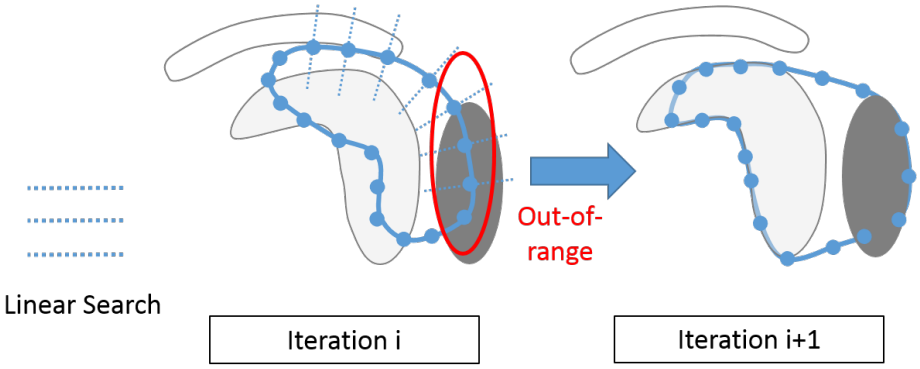


Figure 11: Capture range problem for statistical shape models. Due to the limited range of the trained landmark search, model parts are no longer attracted to the sought organ of interest, if they move out of range at some point during model fitting.

of algorithmic complexity and an accumulation of basic assumptions can often be observed. Both generally limit the direct applicability of the proposed systems to different segmentation tasks. Another way to tackle the problem can be to additionally constrain displacement at different surface parts, as demonstrated in [144]. There it was shown for liver segmentation that the influence from model initialization could greatly be reduced, by introducing a rule-based system where different surface parts are governed by a set of control parameters. However, this fine-tuning of the model fitting through a large set of free parameters must be repeated for new use-cases, which makes an adaptation of such approaches to other scenarios difficult.

The original model fitting scheme for 3D-SSM which conducts search for suitable landmark positions during segmentation, allows a more intuitive understanding of the problem that different local optima are being found under changing conditions. One likely reason for this problem is the ambiguity of the search that produces changing results depending on the intersections of the linear search profiles with the actual object boundary.

This problem usually occurs in conjunction with a boundary-based appearance modeling (cf. Sect. 2.4.1) and has coined the term *visibility problem* for 3D-SSM. In [156], this problem was identified as a reason for high-curvature surface parts that are regularly missed during segmentation, due to lacking visibility of landmark positions at these object parts during the typical unidirectional search of 3D-SSM.

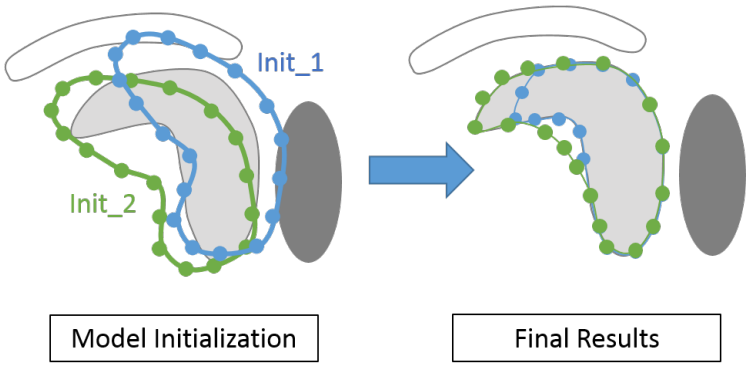


Figure 12: Model initialization problem for statistical shape models. The outcome of the model fitting is often highly unpredictable, and model initialization is a regular source of error. Optimization is difficult, since the sufficiency of the initialization always remains unclear. An unfavorable model initialization is usually hard to compensate in the subsequent model fitting.

As reasoned earlier (cf. Sect. 2.4.4), recent appearance modeling approaches that build on a landmark-based object description can offer the means of a separate and unambiguous detection of different object parts during segmentation. Previous work has shown high spatial distinctiveness [132] and high accuracy [29], [135] of landmark-based appearance modeling in 2D applications. Furthermore, the use of trained regressors for 3D tasks has shown promise for a successful incorporation of non-local image information [130], [151]–[153], which should also be applicable for segmentation strategies using 3D-SSM.

Together, these techniques should provide the means of a model fitting scheme that is based on a robust and omni-directional search for distinct landmark positions. With this, it should be possible to tackle the aforementioned visibility problem of 3D-SSM, and thus to lift the previous influence from model initialization and to improve performance and applicability of 3D-SSM through a more global search for optimal solutions during segmentation.

2.5.4 Summary

Various approaches have been proposed for an image segmentation using 3D-SSM. They follow the different basic strategies for a model fitting of 3D-SSM to unknown image data, i.e. iterative search for landmark positions, direct

estimation of pose and shape parameters, or global estimation of surface-wide landmark displacements.

Traditionally, an iterative search for landmark positions has been in use with boundary-based appearance modeling types, while pose and shape parameter estimation is typically associated with a region-based object description, specifically with active appearance model approaches.

Recently, with the upcoming use of landmark-based object descriptors, both iterative search for landmark positions and global estimation of landmark displacements have been proposed for a model fitting based on that appearance modeling type. These approaches have mostly been applied to 2D segmentation tasks, where spatial distinctiveness of object description and high accuracy have been demonstrated.

An accurate model initialization at the organ of interest is required for 3D-SSM to conduct a subsequent error-free segmentation. As a reason for the well-known sensitivity of 3D-SSM from the previous model initialization, the visibility problem can be identified, where segmentation results are dependent from changing views of linear search scopes on landmark positions during model fitting.

In addition to solutions for single use-case scenarios by means of low-level image processing and specific basic assumptions, more general high-level methods have been developed that aim at a robust and accurate model initialization for a segmentation with 3D-SSM. Nevertheless, the simplicity of the subsequent model fitting of the employed 3D-SSM has shown to be limiting for overall accuracy, which makes correctional measures necessary. Altogether, the high sophistication and complexity of the proposed methods leads to a worse adaptability to other use-cases than the presented scenarios.

The use of a landmark-based object description during segmentation has the particular advantage that it allows an unambiguous detection of distinct landmark positions during segmentation. This should provide the means to directly target the visibility problem of 3D-SSM, where segmentation results are dependent from changing views of linear search scopes on landmark positions during model fitting.

Thus, it should be possible to reduce the strong influence from the previous model initialization on segmentation from 3D-SSM, which has always strongly affected the overall performance and applicability of 3D-SSM (cf. Sect. 2.5.3). Despite the importance of these basic problems for 3D-SSM in general and despite their far-reaching implications, benefits and possibilities of the new techniques have never been examined with a focus on these problems before.

2.6 Conclusions

A large number of strategies, approaches and techniques have been proposed that cover the core aspects of 3D-SSM, i.e. surface representation, shape and appearance prior modeling, model initialization and model fitting. They provide the means for an application of 3D-SSM in general segmentation scenarios and in more specific use-cases.

However, important and fundamental problems of 3D-SSM have so far remained unsolved - i.e. the visibility problem, sensitivity against previous model initialization and susceptibility to local optima during segmentation. They have far-reaching consequences for the general performance and applicability of 3D-SSM. The aforementioned problems have not yet been addressed on a conceptual level by previous work. Instead, compensatory measures have been proposed in the literature which add complexity and hamper an application of 3D-SSM to different use cases.

The more recent concepts of landmark-based object description and incorporation of more robust machine learning techniques during segmentation using 3D-SSM should offer the means for a more direct approach to the aforementioned issues. These techniques have already been shown to provide beneficial traits like high accuracy and direct disambiguation of separate object parts. While applied mostly for 2D applications and for more specific use-cases like bone and ventricle segmentation, these techniques have not been involved with a focus on the more basic problems of 3D-SSM and of their application in general segmentation tasks.

The developments in this thesis will specifically target these topics. The intended extensions to 3D-SSM will build on the presented recent techniques for an appearance modeling using landmark-based object description, for a global use of image information, and for an incorporation of robust machine learning for detection tasks during segmentation. The goal of these means will be to directly address the visibility and initialization problem of 3D-SSM, thus making 3D-SSM more robust against influence from the previous model initialization, and enabling a more global search for optimal solutions during segmentation. These additions to 3D-SSM are intended to be applicable to general purpose segmentation tasks, thus adding to the aforementioned methods that have already shown other benefits of machine-learning based landmark detection for selected 2D segmentation scenarios.

Surface representation and shape prior modeling techniques will be adopted for the developments in this thesis, that allow an incorporation of the aforementioned techniques on a per-landmark basis. They will be chosen to allow

a profound shape prior modeling that also supports a straightforward application of the proposed techniques to most common organ segmentation tasks. The strong focus on a development of landmark-based techniques will also allow an incorporation of other shape modeling techniques, which work on a per-landmark basis and which are suitable for an application of the proposed techniques in more peculiar use-cases.

CHAPTER 3

Shape Model Construction

As described in Sect. 2.2 and Sect. 2.3, different strategies exist for the geometrical representation of organ surfaces and for the training of a shape prior model that generates meaningful surface models during segmentation. The originally proposed explicit surface modeling [3] based on distinct landmarks and mesh triangulations has remained highly popular ever since, and it is also well suited for the proposed enhancements of this thesis.

Apart from its convenient and efficient data representation, this modeling type also allows a fine-grain division of appearance model training and of landmark detection during segmentation. This means that organ appearance can be learned in parallel on a per-landmark basis, thus promoting distributed computation. Also prospectively, segmentation confidence can be assessed separately at distinct landmark positions, and training effort can be focused on particularly demanding landmarks. In parts during segmentation, detection can be narrowed down to fewer landmarks, in order to achieve faster computation. The representation also allows convenient inspection, analysis and understanding of segmentation performance at separate landmark positions. Finally, it promotes compatibility of the proposed enhancements with a vast number of propositions made for this most widespread type of model representation.

For the statistical modeling of shape distribution from a set of training meshes, different approaches yield varying degrees of descriptive power, as de-

scribed in Sect. 2.3. Some methods promise to be of particular use for specific problems and for peculiar anatomical structures, like highly deformable and eccentric shapes, tubular organs, coupled multi-part structures, and severe, shape-altering pathologies. In the more general organ segmentation scope, many approaches rely on a point distribution model (PDM) that is based on a linear shape modeling via principal component analysis (PCA). Despite its strong basic assumption of a linear distribution of shapes, this PDM has been part of solutions where an excellent segmentation accuracy was achieved for deformable soft tissue organs like the liver [110]. Altogether, this qualifies for the projected extensions of 3D-SSM. Since shape modeling, appearance modeling and model fitting are well separated concepts in the original formulation of 3D-SSM, independent propositions can be made in this thesis for a non-local appearance modeling and a more proficient model fitting. These can easily be combined with other shape modeling techniques in the future in order to address more particular problems.

A number of procedural steps are necessary to retrieve a shape prior model from a set of annotated training images. The latter usually contain manual segmentations of specific target organs that are represented by a labeling of object and background voxels. Starting from the object labelings, the following operations are subsequently performed during shape model creation: generation and refinement of surface meshes from the object voxels, optimal placement of an equal number of corresponding landmarks across the surface meshes (*correspondence optimization*), and statistical analysis of shape variation from the corresponding surfaces. These steps will be outlined in the following.

3.1 Surface Mesh Generation

In this first step towards the desired shape prior model, a transition has to be made from the voxel-based representation of the input segmentations to the chosen triangle mesh representation that is used in all following steps (cf. Sect. 2.2.1 and 2.6). The latter raise a number of requirements for the mathematical topology and for the mesh quality of the resulting surfaces, particularly as a prerequisite for the subsequent correspondence optimization step. The meshes must be of genus zero to allow a projection to the unity sphere (*spherical parameterization*), where the placement of an equal number of corresponding surface landmarks on all projected training meshes is particularly convenient (cf. Sect. 2.3.2.3). Consequently, the meshes have to be cleaned from detached

surface parts and from any openings that alter mesh topology. Furthermore, the surface models must be free from unwanted discretization artifacts imposed by the original voxel structure of the input segmentations.

The required processing of binary data and its transformation to the surface meshes is common knowledge and largely builds on the use of basic processing tools. The encountered basic problems and common solutions will shortly be outlined in the following. Further detail can e.g. be found in [157].

Generally, it can be easier to first ensure the aforementioned requirements on the initial binary segmentations, and then to generate the surface meshes afterwards. Otherwise, particular measures have to be taken for the initially generated meshes, e.g. size-preserving smoothing, removal of erroneous and isolated mesh parts and the like. On the input segmentations, inconsistencies are removed that are usually added during manual segmentation, like isolated regions outside the organ of interest, holes or missing parts on single slices of the segmented volume. The removal of these flaws can be achieved using connected component analysis and local median filtering. For some organs, a morphological closing is recommendable for cut-out tubular structures like blood vessels that cross the segmentation volume. Finally, an eccentric anisotropic spacing is often encountered at larger slice distances, which leads to an adoption of characteristic staircase effects by the created surfaces meshes. One effective remedy for this unwanted phenomenon is a Gaussian smoothing of the binary segmentation.

After this kind of basic preprocessing, the segmentations are suited for an extraction of the desired surface meshes. The usual way to do this is the construction of an iso-surface at the mid value of the Gaussian-smoothed segmentations, using the marching cubes algorithm for instance [158]. Finally, the landmark count of the resulting triangle meshes is reduced, for which various remeshing algorithms are available [159], [160].

3.2 Correspondence Optimization

In the process of the chosen statistical shape prior modeling, the given training meshes are considered to be part of a normal distribution of a set of multivariate random variables. In order for these training meshes to be embedded in a common random vector space, an equal number of corresponding surface landmarks has to be placed on all training shapes in a consistent and meaningful manner.

The way in which these corresponding landmarks are placed on the dis-

tinct training meshes has a significant impact on the descriptive quality of the resulting shape prior model. As outlined in Sect. 2.3, different approaches exist that aim at a beneficial placement of corresponding surface landmarks (*Correspondence Optimization*).

For the given purpose, an automatic and particularly time efficient method from [78] is employed. To attain a high quality of landmark correspondences, the method aims at an optimization of the minimum description length (*MDL*) of the produced shape prior model. Thus, the optimization process itself focuses on a high compactness and descriptive accuracy of the resulting shape model. The method is well suited for the task of image segmentation, where a high degree of specificity and generalizability of the employed shape prior model is desired in order to provide a robust shape constraint during model fitting. Furthermore, the optimization of the MDL cost function is conducted in a gradient descent approach which grants a high computational efficiency of the whole process.

Altogether, this makes an automatic, accurate and highly efficient solution for this crucial step of automatic shape model creation. However, the benefits come at the expense of a reduced topological flexibility of the modeled surface shapes. The reason for this is that a spherical parameterization of training meshes is involved in the process, where surfaces are mapped to the unity sphere for a convenient placement of reference landmarks on all shapes. In order for this mapping to succeed, the meshes must be of a genus zero topology, which prohibits holes and tunnelings in the modeled organ surface.

For the general task of a multi-organ segmentation in medical images, soft tissue structures and bones of abdomen, thorax, pelvic area, head-neck region and of the extremities are of major interest. The majority of these structures of interest can directly be modeled by genus zero surfaces, e.g. liver, spleen, kidneys, stomach, pancreas, prostate, bladder, lungs, heart, the thyroid glands, the brain and its subparts, ribs, and all major bones of the extremities. The same applies for the spines where the enclosed spinal cord can be included in the modeled volume and additionally be separated after segmentation if necessary. The hip bone contains several openings and can accordingly not be represented by a genus zero mesh. Tubular structures like the aorta and the rectum can in theory be modeled by a closed surface if upper and lower boundaries are delimited properly. However, a more meaningful modeling has reportedly been achieved for a separate representation of center line and diameter. Due to the highly displaceable nature of the intestines, these are best being treated as a greater region of interest that can again easily be modeled

and segmented by a closed surface. This also applies for other more general anatomical regions of interest, like the torso, the thoracic and abdominal regions, and the head.

For the aforementioned structures that cannot be covered by a genus zero surface, various alternatives exist that can be employed for a modeling of these particular cases [2], [161].

3.3 Statistical Shape Analysis

As motivated at the beginning of this chapter, a statistical shape modeling is applied based on the principal component analysis. However, the presented techniques in this thesis were developed independently from the underlying shape prior modeling, and they will also be compatible with other landmark-based shape modeling techniques that were proposed for various purposes in the literature (cf. Sect. 2.3 and 2.6).

The PCA assumes that the training meshes be samples of a normal distribution that hypothetically describe the variety of encountered shapes for a given organ of interest. From this, a generative shape model (*Point Distribution Model*) is constructed that allows the arbitrary creation of surfaces in agreement with the previously learned shape distribution. The PDM was first introduced in [162] for use as a robust shape constraint in image segmentation. The underlying basic principles and their actual application during segmentation will shortly be outlined in the following.

At first, after the previous landmark correspondence optimization, all training shapes are represented by an equal number of meaningful surface landmarks $\mathbf{x_j}$. Each of the k training shapes $\vec{X}_i$ is then regarded as a random vector of surface landmark coordinates x_{j1}, x_{j2}, x_{j3} (cf. Sect. 2.3.3):

$$\vec{X}_i = (x_{11}, x_{12}, x_{13}, \cdots, x_{n1}, x_{n2}, x_{n3}) \in \mathbb{R}^{3n}, i = 1, \cdots, k$$

Shape is generally considered to be the difference in geometry that remains after a removal of translation, rotation, scaling and sheering from distinct surfaces. Therefore, all training shapes have to be aligned by an affine transformation before the actual shape information can be retrieved from the residuals of the aligned meshes. To this end, a procrustes analysis is usually applied [28], which delivers a closed-form solution for the alignment of all shapes through affine transformation. For a normalization of scale, the aligned shapes are pro-

jected to the tangent space of the normalized mean shape where the averaging property of the latter is preserved (cf. Sect. 2.3.1).

The aligned training samples now reflect the assumed normal distribution of shapes that is described by the covariance matrix S of the k training meshes $\vec{X}_i$:

$$S = \frac{1}{k-1} \sum_{i=1}^{k} (\vec{X}_i - \bar{X})(\vec{X}_i - \bar{X})^T$$

with regard to their mean shape

$$\bar{X} = \frac{1}{k} \sum_{i=1}^{k} \vec{X}_i$$

As motivated in [162], the high dimensionality of the shape space ($\mathbb{R}^{3n}$, n in a magnitude of > 1000) and the correlation of the landmark variables x_{ik} prohibit an efficient and convenient generation of new shapes X_j from the learned distribution. Therefore, a dimensionality reduction is carried out by the principal componenent analysis (PCA), where the given shape distribution is explained by the fewer eigenvectors (*eigenmodes*) of covariance matrix S instead of the full set of random landmark variables x_{ik}. The eigenvectors $\Phi_1, \cdots, \Phi_{k-1}$ ($k \ll 3n$ in general) of S provide a compact description of the principal modes of shape variation across the observed training samples. Thus, they define a new set of uncorrelated random variables $\alpha_1, \cdots, \alpha_{k-1}$ (*shape parameters*) which necessarily describe the given shape distribution, with variances $\sigma_1^2, \cdots, \sigma_{k-1}^2$ equaling the respective eigenvalues $\lambda_1, \cdots, \lambda_{k-1}$. For an increasing sample size k, it is the accumulated variance from the expectedly declining eigenvalues that defines an upper bound for model complexity. Therefore, it is sufficient to rely on the largest c eigenvalues that cover the majority (usually between 95% and 98%) of the total variance from all $k-1$ eigenvalues (after ordering the λ_i):

$$\sum_{i=1}^{c} \lambda_i \geq r \cdot \sum_{i=1}^{k-1}, \quad \lambda_1 \geq \cdots \geq \lambda_{k-1},$$
$$c < k, \quad r \in [\sim 0.95, \sim 0.98]$$

Instead of an eigen-analysis of the covariance matrix, a singular value decomposition (SVD) of the mean-free shape matrix $(\vec{X}_1 - \bar{X} \cdots \vec{X}_k - \bar{X})$ is usually employed for higher numerical stability.

This directly leads to the generative model of shapes, where arbitrary surfaces $\vec{X}_j$ from the learned distribution can be formed by a linear combination of the eigenmodes Φ_i, with chosen weights for the random shape parameters α_i:

$$\vec{X}_j = \bar{X} + \sum_{i=1}^{k-1} \alpha_i \Phi_i \approx \bar{X} + \sum_{i=1}^{c} \alpha_i \Phi_i$$

During shape creation, the variances σ_i^2 of the shape parameters α_i can be taken into account, e.g. separately by a parameter-wise limitation:

$$-3\sqrt{\lambda_i} \leq \alpha_i \leq +3\sqrt{\lambda_i}$$

Or by a global, hyper-ellipsoidal constraint:

$$\sum_{i=1}^{c} \frac{\alpha_i^2}{\lambda_i} \leq t$$

Each shape from the learned distribution is uniquely identified by its shape parameters, hereafter summarized by the joint shape vector $\vec{b} = (\alpha_1, \cdots, \alpha_c)^T$. To be available during image segmentation, any shape has to be mapped from shape space to the image domain by an affine transformation T. With $P = (\phi_1 \cdots \phi_c)$, each surface $\vec{X}_j$ in image domain is then defined by its $\vec{b}$ and T:

$$\vec{X}_j = T(\bar{X} + \sum_{i=1}^{c} \alpha_i \phi_i) = T(\bar{X} + P\vec{b})$$

For the image segmentation, surface shapes are iteratively generated to best fit the sought target organ (*model fitting*). At each iteration t, a search for new landmark positions from the current surface $\vec{X}_t$ is conducted, thus delivering a hypothetically optimal set of landmark displacements $d\vec{X}$. The newly found landmark positions $\vec{X}_t + d\vec{X}$ are then approximated by an update of shape parameters $\vec{b} + d\vec{b}$ and of the affine transform T_{t+1}:

$$\vec{X}_{t+1} = \vec{X}_t + d\vec{X} \approx T_{t+1}(\bar{X} + P(\vec{b} + d\vec{b}))$$

First, the new mapping T_{t+1} is determined by a procrustes matching of the current surface $\vec{X}_t$ and the updated landmark positions $\vec{X}_t + d\vec{X}$. Then, the shape vector is updated by $\vec{b} + d\vec{b}$ to fit the residual between the new mapping of the former shape $T_{t+1}(\vec{X}_t)$ and the landmark positions $\vec{X}_t + d\vec{X}$. Using the

orthogonality of P, $\vec{dx} \approx P \cdot \vec{db}$ and $\vec{dx} = \vec{X}_t + d\vec{X} - T_{t+1}(\vec{X}_t)$, this leads to a solution for the updated shape parameters:

$$\vec{db} = P^T(\vec{X}_t + d\vec{X} - T_{t+1}(\vec{X}_t))$$

Thus, the model fitting yields an iterative optimization process of T and $\vec{b}$ in order to maximize the shape posterior $p(\vec{X}|I)$ in image I.

CHAPTER 4

Landmark-based Context Appearance Modeling

As described in Sect. 2.5, 3D-SSM exhibit a number of intrinsic basic limitations that generally hamper their application and that notably increase the effort necessary to reach high accuracy in various applications. As a hypothesis, these problems go back to the nature of the underlying appearance modeling, and to the characteristics of landmark search and model fitting. Combined, these intertwined concepts lead to the well-known and often observed problems regarding applicability and performance of 3D-SSM, and they cause greater implications for 3D-SSM and for their use in medical image segmentation.

In the following, a statement of the problem will be presented based on the limitations detailed in Sect. 2.5.3, and the choice of another type of appearance modeling, landmark search and model fitting scheme will be motivated, in order to directly address the aforementioned problems.

4.1 Problem Statement and Approach

Three general problems are found in 3D-SSM during their application in image segmentation tasks (also cf. Fig. 10 - 12 in Sect. 4):

1. Their solution during segmentation strongly depends on the choice of the previous model initialization, which serves as a starting condition for the subsequent model fitting. (*Model Initialization Problem*)

2. During the iterative model fitting, parts of the surface model can move out of detection range regarding meaningful landmark positions on the target surface. As a result, attraction of the surface model to the organ of interest fails, which leads to segmentation errors. (*Capture Range Problem*)

3. The visibility of meaningful landmark positions depends on the current model pose during the iterative model fitting process. In conjunction with the limited capture range of 3D-SSM, this means that search for meaningful solutions of the segmentation problem is strongly limited to locally optimal solutions. As a result, lack of visibility at parts of the target surface cause 3D-SSM to get stuck at suboptimal solutions. (*Visibility Problem*)

The first problem, dependence from model initialization, can be regarded as a result of the other problems. The visibility problem makes the final segmentation outcome dependent from the initial model placement relative to the sought organ of interest. That is, throughout the iterative model fitting, visibility of landmark positions changes continuously, and is altogether dependent from the initial visibility of positions at the actual model initialization. Furthermore, the limited capture range promotes the occurrence of segmentation errors from suboptimal model initialization, if surface model parts are at more remote positions at the beginning of the model fitting. And finally, the limited capture range gives rise to the requirement of model placement in the vicinity of the target surface altogether.

Regarding appearance modeling and model fitting, the following traits can be identified that are expected to cause visibility and capture problems of 3D-SSM. First, typical boundary-based approaches only provide an ambiguous feature description that is suited for an identification of larger boundary sections during segmentation (cf. Fig. 5). The resulting ambiguity is a prerequisite for the visibility problem.

Second, boundary- and region-based approaches generally only incorporate local information for a description of object or surface parts, which gives rise to the capture range problem, where meaningful information is only available within limited range of the modeled objects.

And finally, the resulting search strategies based on unidirectional linear profiles and on residual sampling for Active Appearance Model fitting translate the aforementioned limitations to the model fitting process. Their sampling range is limited to the vicinity of the surface model during segmentation, and

in case of the profile-based search, visibility of landmark positions is directly limited to the one-dimensional scope. Also, visibility and model initialization problems have been reported for Active Appearance Models (e.g. [163]).

Therefore, based on the recent previous advances in landmark-based appearance modeling techniques (cf. Sect. 2.4.3), the proposed appearance modeling and model fitting approach will be aimed at the following characteristics as a direct response to the aforementioned conceptual limitations:

1. A non-local appearance modeling that incorporates global information from widely distributed viewpoints in the image domain.

2. An unambiguous modeling of object appearance, based on a separate description of distinct surface landmarks on the model surface.

3. An omni-directional and locally unconstrained search for distinct landmark positions on the sought organ surface.

The expectation of the chosen approach is that it should allow a detection of landmark positions regardless of model pose and from larger distances during segmentation. As a result, 3D-SSM should be enabled to find solutions during segmentation more reproducibly, the should be attracted to the sought target surface from larger distances, and they should generally produce solutions that are closer to the global optimum of the segmentation problem.

The employed concepts of a non-local appearance modeling and of a feature description required to disambiguate object landmarks from their non-local context will be described in the remainder of this chapter. The use of the learned landmark-based appearance model for an omni-directional landmark detection will be detailed in chapter Sect. 5.

4.2 Non-local Landmark Appearance Model

In contrast to the local boundary or regional appearance modeling of previous 3D-SSM as described in Sect. 2.4.4 (*local appearance modeling*), a non-local appearance modeling of object landmarks will build the foundation for an omni-directional landmark detection of the proposed 3D-SSM during segmentation. Information will be captured at widespread locations around the modeled surface landmarks, instead of a limitation to image information near or inside the modeled surface.

Main task of the non-local appearance model training is an automatic selection of features to accurately describe the appearance of the considered surface landmarks. While for the boundary- and region-based appearance modeling approaches, feature descriptors were employed that reflect local object characteristics well, the use of a feature descriptor for non-local modeling is required to provide a different kind of expressiveness and flexibility to capture distributed image information. As described in Sect. 2.4.3, randomized features are well suited for this task. Their use for the proposed appearance modeling in 3D-SSM will be detailed in Sect. 4.3.

Also, accuracy and robustness must be provided for the trained landmark detectors during segmentation, and the proposed appearance modeling must also generalize to differing use-cases. Enabling robust and accurate inference during model training is expected to be achieved by means of a suitable machine learning, which will be described in Sect. 4.4.

Finally, distributed non-local modeling of object appearance and the subsequent inference of sought positions in unknown image data can be performed either using a classification-based approach, or a regression-based approach. The difference between the two concepts will be detailed in the following.

4.2.1 Classification-based Modeling

The majority of statistical shape model based approaches is based on a formulation of appearance modeling and of object detection as a classification task (cf. Sect. 2.4). Voxel affinity to objects or boundaries is provided from annotated images by means of positive and negative training samples. Samples are used to train a classifier which performs a discriminative task of object / boundary voxels and background voxels during segmentation.

Such an approach can also be used for a non-local modeling and detection of landmarks in image data [164]. The image space is then scanned in a voxel-wise manner in order to classify foreground and background voxels during segmentation. One general disadvantage of these approaches is the exhaustive nature of the search due to the repetitive scanning of larger image areas.

Another problem of these approaches is the presence of labeling noise that is regularly observed for purely classification-based approaches. Labeling noise is usually caused by false positive classification, and a regularization is required to allow robust and accurate inference of sought landmark positions, e.g. by means of graphical models [164]. This would however not hamper use in 3D-SSM, since they already impose regularization based on their intrinsic shape constraint.

Also, classification based approaches are largely based on information from negative non-object samples, which are available in abundance from underlying training images. By contrast, only little information on positive samples is provided for a modeling of single landmarks or object boundaries. However, in order to limit the sparsity of positive samples as opposed to the large number of negatives, a cut-off decision has to be made at some point in order to limit the number of negative samples. Alternatively, clustering techniques can be employed to increase the number of positive samples available for classifier training, as detailed in Sect. 2.4.1.

Apart from their widespread use and robust performance in many applications, classification-based approaches clearly exhibit a number of disadvantages that are encountered regarding design of landmark-based appearance modeling techniques. A comprehensive examination of the described traits related to the use of classifiers can e.g. be found in [165]. The use of regression-based approaches as a natural alternative has several appealing advantages, as discussed in the following.

4.2.2 Regression-based Modeling

In addition to the classification-based approaches, regression-based techniques have later been proposed to formulate detection tasks in medical image analysis [151]. They go back to similar propositions in computer vision [132], [166], [167]. The idea of regression-based detection is outlined in Fig. 2.4.3, where a spatial context function is learned from distributed patch positions around sought landmark positions, based on an expressive set of image features.

For the first time in [151], regressors were proposed for detection tasks in medical images. Immediate incentives were the more natural approach to the detection task, where relative position estimates are made from distributed image positions, instead of an exhaustive scanning of larger image regions. Also, the aforementioned problem of sparsity of positive samples does not apply to the regression approaches.

Furthermore, in contrast to the labeling noise of classification-based approaches, where false positives can hamper a conclusive extraction of landmark positions without additional regularization, false position estimates produced by a regressor do not necessarily impede further detection. On the opposite, many position estimates are often at least closer to the sought position, which can be utilized to produce new subsequent estimates from the first rough position estimates [153].

This ideally suits the iterative nature of the 3D-SSM model fitting, where predictions are refined iteratively while they close in on the segmented target organ. Meanwhile, rough outliers are removed by the learned shape constraint. Also, if a regression is opted for, the use of classifier output can still be considered to add information e.g. on the confidence of the regressor results. Finally, the use of a regressor has been shown to significantly improve landmark detection accuracy [153].

In [29], [135], the idea of a regression-based detection was first introduced for shape-model based bones segmentation and facial landmark detection in 2D image data. Later, global optimization of regression votes were proposed also for 2D bone segmentation [136]. Despite their promising results in the presented applications, the use of regression-based landmark detectors remained limited to the aforementioned 2D applications, in addition to follow-up application on bones in 2D datasets [168]. In particular, a use of these techniques to tackle the well-known problems of 3D-SSM in volume segmentation tasks, as described in Sect. 4.1, have never been addressed.

As a major reason for the limitation of landmark regressors to 2D domain was the higher modeling complexity and computational burden for 3D-SSM, as suggested in [136]. Also, the more general case of soft tissue landmark modeling adds further challenges regarding higher shape variability and reduced contrast [147]. Whereas free model deformation as proposed by the constrained local model approach from [29], [135] is expected to work for more rigid structures in 2D domain, an application for highly deformable soft tissue in 3D domain without additional regularization is unclear. Also, an extended spatial context, variable organ positioning and changing fields of view of 3D data differs from the previous radiographic and facial picture task.

In contrast to the previous applications for 2D bone segmentation, the proposed method will extend the above concept of appearance modeling to the more general case of soft tissue landmark modeling and segmentation in volumetric image data for different imaging modalities. In [2], 3D-SSM were categorized based on their underlying appearance modeling and model fitting strategies. Since its proposition for 2D active shape models in [29], [135], no similar categorization of the regression-based context modeling approach for 3D-SSM has been presented so far. Therefore in the following, the proposed approach will be outlined in the broader context of appearance modeling for 3D-SSM.

The use of an expressive feature set has proved an essential prerequisite for a regression-based landmark modeling and detection [132]. A family of suit-

able feature descriptors, and their incorporation for the proposed appearance modeling will be outlined in the following section.

4.3 Randomized Feature Description

For the use of regression based detection tasks, a number of feature sets has been established that has proved to possess the flexibility and expressiveness that is required for description of image context at changing viewpoints [132], [138], [140], [169]. They all have in common the use of structural primitives on which variation is induced by randomization within a predefined image patch. By contrast, the use of steerable filters has also been proposed for a detection of facial landmarks and of hand-bone contour points [104]. The concept of steerable filters has also successfully been employed in other detection tasks, like the classification based approach of Marginal Space Learning [148], [149]. However, the use of the randomized features has probably been more popular, in conjunction with machine learning approaches that work well with randomized feature sets, like random forests or random ferns.

The use of random regression forests also offers appealing benefits regarding the goals of the proposed method and has often been applied with success in detection tasks. Therefore, the following sections will focus on a use of randomized feature descriptors in conjunction with random regression forests for the proposed appearance modeling.

4.3.1 General Considerations

In any case, flexibility is promoted through an additional introduction of structural randomness. Features are defined by a random positioning of (usually two) box elements or voxels within the patch, and sometimes by an additional random scaling of the structured elements. Randomization can be limited to one moving element with a fixed element at the center point of interest like in [153], or to a random pair of voxels [138], [140]. Or, both elements can be configured at random, like in [169], [170].

From the broad variety of potential feature descriptors proposed previously (e.g. by [140], [153], [169], [170]), randomized 3D Haar-like features possess all of the desired traits: they use freely arrangeable structured elements that have proved a particularly high degree of flexibility and expressiveness [171]. They can reflect highly variable anatomical key structures, and their randomization grants the ability to capture image information at arbitrary viewpoints in the

image. For a repeated evaluation of features during model fitting of 3D-SSM, a fast evaluation of the employed feature descriptors is important to allow feasible execution times during segmentation. Despite their higher complexity compared with features based on random voxel pairs [138], [140], they can be evaluated in constant time by means of precomputed integral images.

Although 3D Haar-like features have been shown to be robust against object rotation to some degree [171], they are not rotation-invariant. This means that the recognition of learned features may not work in images that display major rotations for any reasons. This can for example occur if coordinate axes of the underlying image geometry are defined differently across the images. In these cases, the use of header information from these images allows a retrospective rectification of the alternate reference frames. To this end, basic software tools can be applied, and no further measures have to be taken.

Another reason for an appearance of major rotations is the variability of patient positioning during image acquisition. In fact, most clinical scenarios do involve image acquisition protocols that arrange for a general patient positioning which is usually either a dorsal or a ventral bedding of the patient, or an axis alignment with a specific anatomical target region. Minor variations can occur, like e.g. casual inclinations of a patient's body. Such cases are then in general sufficiently represented within the underlying training images. Due to their robustness with regard to non-linear learning tasks, the employed random regression forests are then able to cope with the described minor variations in the datasets.

If a handling of more general cases should be required in the future, various solutions can be considered in this regard, like a rigid pre-registration of input images, an augmentation of the training data with artificial rotations, confidence-weighted selection of specifically trained forests, or the use of rotation-invariant features.

The following subsection will describe the randomized 3D Haar-like feature descriptor and its time-efficient evaluation. The selection of actually meaningful 3D Haar-like features from underlying training images is then part of the forest-based appearance model training which is detailed in Sect. 4.4.

4.3.2 3D Haar-like Features

Haar-like features were first introduced in [131] for facial detection in 2D computer vision. The idea was to provide an extensive set of simple, yet highly representative and time-efficient features for the description of human faces in

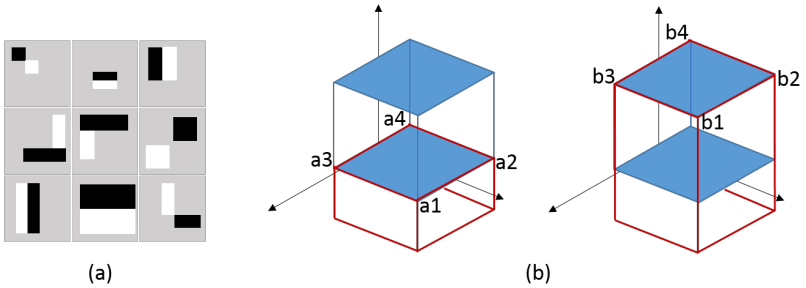


Figure 13: Example set of randomized Haar-like features. Each feature is generated as a set of two random box integrals (black and white) which are calculated on the underlying image intensities (a). Subtraction of the calculated box integrals yields the corresponding feature response. Each feature is evaluated in constant time using pre-calculated integral images for integral computation (b). The shown box integral in image space (blue cuboid) is calculated from voxels $a1, \cdots a4$ and $b1, \cdots b4$ in the integral image. $a1-a2-a3+a4$ and $b1 - b2 - b3 + b4$ yield the red box integrals (b, left and right), which are subtracted to get the blue box integral.

gray-value images.

Features are composed of either two, three or four adjacent rectangles, that are all together arranged variably within a predefined image patch. If applied to an image, feature response is produced by the contrasting rectangles, i.e. for each rectangle the integral of intensities in the underlying image region is calculated, and the response is derived from the difference of the opposing integrals, as described in [131]. Depending on the feature at hand, a strong response is obtained at places that reflect the structure of the given feature. On the opposite, there is a neutral response within homogeneous image regions, due to the area zeroing wavelet structure of the Haar-like features.

At a glance, the total feature set contains edge- and ridge-features of different scales and in horizontal, vertical and diagonal orientation. For an actual detection task, meaningful features then have to be learned from all possible configurations. For the 2D facial detection in [131], this was done e.g. by a cascade of AdaBoost classifiers, and the learned features reflected characteristic image regions like the greater eye-region and the mid-face, and facial details like nose and mouth.

Thanks to their high efficiency and their flexibility, Haar-like features were subsequently adopted for extensive use in other tasks, e.g. for identification,

description and matching of characteristic key-points in the SURF-algorithm [172].

For the recognition tasks in said applications, the employed Haar-like features merely served local purposes. They were used to either speedup local image filtering operations (e.g. fast computation of multi-scale Hessian determinants), for fast local neighborhood description at distinct keypoints in the style of oriented gradients, or for a representation of characteristic local features within objects of interest. Whereas, the global application of these features is usually limited to a rejection of true-negatives from the image background.

A detailed definition of 2D Haar-like features, and ideas for their time efficient calculation using 2D integral images is provided in the original work [131]. Rather little information is available in the literature on an extension of these features to 3D domain and on a time-efficient computation in 3D. Therefore, we will provide a definition of 3D Haar-like features, and details on their constant-time evaluation in the following, including an algorithm for a time-efficient computation of the required 3D integral images. To the best of our knowledge, we have for the first time proposed 3D Haar-like features for a description and detection of arbitrary landmarks on soft tissue surfaces and for use in different imaging modalities.

To extend the original concepts to 3D domain, each possible Haar-like feature is defined by two randomized cuboids C_1, C_2 within a predefined reference patch P that is centered at an evaluated reference voxel $v = (v_1, v_2, v_3)$:

$$C_i = (x_{\min}, x_{\max}, y_{\min}, y_{\max}, z_{\min}, z_{\max}), \quad (x_i, y_i, z_i) \in P$$

In order to control the degree of image context that is considered for detection by the employed feature descriptor, a maximum patch size can be chosen that serves as an outer bound for the generated cuboids.

The feature response $f_{c1,c2}(v)$ as stated above is derived from the intensity integrals Σ_{c1}, Σ_{c2} over C_1 and C_2:

$$\Sigma_{c_i} = \sum_{i=x_{\min}}^{x_{\max}} \sum_{j=y_{\min}}^{y_{\max}} \sum_{k=z_{\min}}^{z_{\max}} I(i,j,k)$$
$$f_{c1,c2}(v) = \Sigma_{c1} - \Sigma_{c2}$$

Instead of the above summation of all underlying intensities, the computation of each integral Σ_c can be achieved in constant time by a total of 7

additions. This is possible using a pre-calculated *integral image* $\Sigma(I) := \Sigma_I$ of the original image I, as detailed for the 2D case in [171]. For the 3D case, at each index i, j, k, the integral image contains the sum of intensities in the cuboid $C_{ijk} = (0, i, 0, j, 0, k)$ that is spanned in I between the origin $(0, 0, 0)$ and (i, j, k):

$$\Sigma_I(i, j, k) = \sum_{x=0}^{i} \sum_{y=0}^{j} \sum_{z=0}^{k} I(x, y, z)$$

The integral Σ_c of an arbitrary cuboid $C = (x_{\min}, x_{\max}, y_{\min}, y_{\max}, z_{\min}, z_{\max})$ can as well be composed by such origin-spanning cuboid volumes at each corner point of C (cf. Fig. 13):

$$\begin{aligned}\Sigma_{\mathrm{c}} = \quad & \Sigma_I(x_{\max}, y_{\max}, z_{\max}) - \Sigma_I(x_{\min} - 1, y_{\max}, z_{\max}) \\ & -\Sigma_I(x_{\max}, y_{\min} - 1, z_{\max}) + \Sigma_I(x_{\min} - 1, y_{\min} - 1, z_{\max}) \\ & -\Sigma_I(x_{\max}, y_{\max}, z_{\min} - 1) + \Sigma_I(x_{\min} - 1, y_{\max}, z_{\min} - 1) \\ & +\Sigma_I(x_{\max}, y_{\min} - 1, z_{\min} - 1) - \Sigma_I(x_{\min} - 1, y_{\min} - 1, z_{\min} - 1)\end{aligned}$$

Each $\Sigma_I(x, y, z)$ can then simply be looked-up in the pre-computed integral image. The computation of the integral image plays an important role, especially because its computation time adds to the total time needed for each new segmentation, or even twice if a feature normalization is used as described below. Little information is provided on time efficient computation of the 3D integral image in the literature. For the proposed method, the following algorithm was devised for the integral image computation in linear time with regard to its voxel count.

Each cuboid $\Sigma_I(x, y, z)$ is composed by the cuboid $\Sigma_I(x-1, y, z)$ of the previously visited voxel and by the next cuboid-complementing slice $\Sigma_{\mathrm{slice}}(x, y, z)$. The latter can be obtained by an addition of the previously calculated slice $\Sigma_{\mathrm{slice}}(x, y, z-1)$ and the newly calculated line $\Sigma_{\mathrm{line}}(x, y, z)$. Only requirement is holding and initializing the required slice and line sums. This way, it is possible to compute typically sized integral images in a matter of seconds on standard desktop computers ($512 \times 512 \times 200$ voxels in $<3s$, single-threaded). All details are shown in Algorithm 1.

4.3.3 Feature Normalization

Unlike the previous works in 2D, a normalization of the Haar-like features will be used by the proposed method, in order to make it applicable to more het-

Algorithm 1 Calculate the integral image $\Sigma(I)$ of image I

Require: 3D gray-valued image $I(x, y, z)$
Ensure: Integral Image $\Sigma_I(x, y, z) = \sum_{i=0}^{x} \sum_{j=0}^{y} \sum_{k=0}^{z} I(i, j, k)$
Initialize slice sums: $\Sigma_{slice}(x, y) = 0 \quad \forall x, y$
for all z **do**
  Re-initialize line sums: $\Sigma_{line}(x) = 0 \quad \forall x$
  for all y **do**
    Re-initialize cuboid sum: $\Sigma_{cuboid} = 0$
    for all x **do**
      $\Sigma_{line}(x)$ += $I(x, y, z)$
      $\Sigma_{slice}(x, y)$ += $\Sigma_{line}(x)$
      $\Sigma_I(x, y, z) = \Sigma_{cuboid} + \Sigma_{slice}(x, y)$
      $\Sigma_{cuboid} = \Sigma_I(x, y, z)$
    end for
  end for
end for

erogeneous imaging modalities. Various techniques for linear and non-linear normalization on image regions like the Haar-like feature patches were proposed in the literature (cf. Sect.2.4.2). A simple, robust and fast normalization based on patch variance of image intensities was proposed in [131]. Like the fast evaluation of the Haar-like features, normalization can also be done in constant time with the help of integral images. Instead of traversing patch voxels twice for computation of mean and variance, the sum of squared voxels in the patch can be used for variance calculation:

$$\sigma^2 = \mu^2 - \frac{1}{N} \sum_{i=1}^{N} x_i^2$$

The integral image for the previous box integral calculation is also used for constant time calculation of the squared mean. Analogously, another integral image of squared intensities is used to get the sum of squared voxels.

The randomized 3D Haar-like descriptor is well suited for a representation of distributed image features. It can be used to capture image information from the surroundings of a modeled organ of interest, and thus allows a non-local modeling of landmark appearance for 3D-SSMs. To this end, image features must be selected from the underlying vast feature pool that are actually

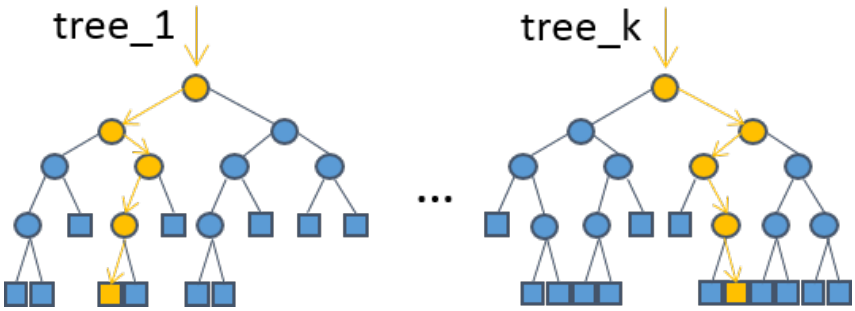


Figure 14: Trained random regression forests. During training, trees are constructed by splitting sample sets into left and right subsets at each tree node. Samples are split according to a splitting criterion that maximizes sample homogeneity. Predictions are made on the sample distributions in the leaf nodes, which are aggregated across all trees based on confidence of the predictions, in order to derive a robust forest estimate.

meaningful for detection. Furthermore, the learned, distributed appearance information must be reconnected to the modeled landmark positions. Both is a task for the automatic, non-local appearance model training and will be discussed in the following section.

4.4 Model Training and Inference

As motivated earlier, landmark-wise trained random regression forests are well suited to learn the spatial context function as part of the proposed appearance modeling (cf. Sect. 4.2), and to infer landmark positions from the learned model during segmentation. They offer a high degree of flexibility and robustness regarding their application to different use-cases, which fits the goal of a 3D-SSM that is more readily available for use in different segmentation tasks. Their main characteristics, their training and their application will be detailed in the following.

4.4.1 Random Regression Forests

Random forests were originally introduced in [173] as an extension of decision trees that had earlier only been used with limited success. Significant boost in performance was reached through bagging, that is, predictors were trained and

averaged on random subsets of samples and features (*bootstrap aggregating* or *bagging*), to compensate for over-fitting of the decision trees. An ensemble of the trained tree predictors is called a random forest.

After their introduction [173], random forests have widely gained popularity as a robust and effective tool in various machine learning applications [174]. In medical imaging tasks, random forests have first been applied for classification-based object detection and segmentation (cf. e.g. [175], [176]), before additional benefits from a use of random forests in regression tasks during detection were demonstrated [153], [169], [177].

Random forests owe their popularity to several beneficial traits. They perform well in non-linear classification and regression tasks, and they are known for their easy and robust application where parameter settings for high performance can easily be found in various use-case scenarios, while they determine meaningful image features automatically during training. For further details on random forests, please refer to [174].

For previous approaches relying on a landmark-based modeling of object appearance by means of regression, different approaches were proposed (cf. Sect. 2.4.3), like canonical correlation analysis [104], boosting regression [133], [134], random ferns [139], and random regression forests [29], [135].

Random regression forests have succeeded in various anatomical detection tasks in CT data, like for bounding box detection around organs of interest [137], [169] or for a detection of a set of anatomical landmarks [153]. Random regression forests also served as a highly accurate foundation for the previous constrained local model approaches in 2D domain [29], [135]. Although an application for the intended modeling and detection of arbitrary soft tissue surface landmarks has not been examined so far, the use of random regression forests is promising for the intended landmark modeling of the proposed method. Furthermore, their well-known straightforwardness regarding an application in different use-cases fits the goals of the proposed method. An incorporation of random regression forests is expected to make the proposed appearance modeling generalize well to different use-cases.

In the following, the use of the proposed random regression forests for the above appearance modeling will shortly be outlined. They will in large parts be similar to the 2D case from [29], [135]. However, in contrast to these works, two additions will be made for the proposed method. Instead of the original ensemble trees from [173], we propose the use of extremely randomized trees [178], which have been shown to be considerably faster during training, and which have proved to be more robust in certain applications. Also, we propose

the use of a splitting criterion based on impurity of displacements which is sometimes encountered in the literature (cf. e.g. [167]), and which may also speed-up training in contrast to the criterion based on sample entropy [29], [135]. Training of the ensemble tree predictors and their use for inference will be detailed in the following section.

4.4.2 Sample Split and Tree Construction

Each tree has a binary topology as shown in Fig. 14, where tree nodes split incoming training samples into smaller subsets. Thus, the overall learning task is divided into easier sub-problems by each split.

To reach a simplification of sub-problems, each split aims to increase the homogeneity of samples in the resulting subsets, i.e. purity of class labels (classification) or of output values (regression). Purity is reached by a *splitting criterion* that is optimized during each split. Thus ideally, the splitting criterion is designed to minimize sample impurity. A splitting function f_{split} (e.g. hyperplane, ellipsoid) splits the samples in a random subspace from k random features $F = \{f_1, \cdots, f_k\}$ of the whole feature space.

If samples are split by hyperplanes $f_i = x, \quad f_i \in F, x \in \mathbb{R}$ in the feature space over F, then the possible splits are given by split functions

$$f_{\text{split}}(i, x) := f_{i,x}, \quad i = 1, \cdots, k, \quad x \in \mathbb{R}$$

Each $f_{i,x}$ splits the arriving training samples $A = \{a_1, \cdots, a_n\}$ into subsets A_L and A_R:

$$A_L(f_{i,x}) = \{a_i \in A | f_i(a_i) \leq x\}$$
$$A_R(f_{i,x}) = \{a_i \in A | f_i(a_i) > x\}$$

Splitting functions $f_{i,x}$ for each possible split can e.g. be found by sorting all samples by f_i first and then repeatedly choose $f_{i,x}$ that split A in ascending order.

The objective is to select a feature f_i that minimizes the splitting criterion over the subsets. In case of the proposed regression, the goal is to minimize the impurity over all vector samples in A_L and A_R in terms of sample dispersion in the subsets. Different measures for sample impurity are commonly used, e.g. based on entropy $\log(|\Sigma(A)|)$ of sample covariance [29], [135], or based on total sample variance $\text{Tr}(\Sigma(A))$ [167], which can speed up training. In conjunction with the proposed use of extremely randomized trees described below, the use

of a sample variance based criterion will also be examined for the proposed method, in addition to the use of the entropy based criterion.

Impurities $S_{f_{\text{split}}}(A)$ are defined for each split of A:

$$S_{f_{\text{split}}}(A) = |A_{L,f_{\text{split}}}| \cdot \log|\Sigma(A_{L,f_{\text{split}}})| + |A_{R,f_{\text{split}}}| \cdot \log|\Sigma(A_{R,f_{\text{split}}})|$$

and

$$S_{f_{\text{split}}}(A) = |A_{L,f_{\text{split}}}| \cdot \text{Tr}(\Sigma(A_{L,f_{\text{split}}})) + |A_{R,f_{\text{split}}}| \cdot \text{Tr}(\Sigma(A_{R,f_{\text{split}}}))$$

To reduce computation time of the splits for larger sample sets in the upper tree part, the optimal split is determined on a smaller random subset of all samples at each node, as proposed in [29], [135].

As motivated earlier, the use of extremely randomized trees will be examined regarding their use in the proposed regression based landmark appearance modeling [178]. Instead of finding an optimal splitting function $f_{i,x}$, where i and x are chosen to minimize sample impurity $S_{f_{\text{split}}}(A)$, an optimal feature f_i is determined with random instances of x, thus performing a random split in each dimension of the parameter space instead of search for an optimum split for each considered feature. This means that complexity of the node splitting will be reduced by the full search for optimal feature values in each dimension of the considered feature subspace.

Splitting stops if a sufficient purity of samples is reached or if a minimum number of samples is encountered in a subset. After the final split, the remaining data in the leaf nodes serve as a basis for later prediction (*leaf prediction*). As described above, several trees are trained for a random forest in this manner, and randomness is introduced during training by bagging of training samples and by using random feature subsets at each node.

4.4.3 Leaf Prediction and Aggregation

During segmentation, for any tested location around the sought landmark, a relative position estimate for the unknown landmark position can be inferred from the trained forest. The tested sample is processed by each of the trained ensemble trees, where the sample is evaluated by the previously built split nodes as it travels down each tree (cf. Fig. 14). Its image context is evaluated at each node based on the selected split feature at that node, and based on the split decision for the tested sample, it is passed to one of the child nodes.

Depending on the previous split decisions, the test sample arrives at a leaf node that contains all samples that are expected to be taken from the same context of the modeled landmark during training.

Thus, a position estimate can be derived from the corresponding samples in the leaf node, based on their displacements to the modeled landmark position. All available information about the sample distribution in the leaf node can be used for a meaningful prediction. Usually, mean and covariance of the sampled distribution are incorporated during leaf prediction [29], [169], which is expected to also fit the requirements of the proposed method. Also, more ample modeling approaches on the sample distributions in the leaf nodes have been proposed [179], e.g. based on Gaussian mixture modeling.

Based on the bagging of tree predictors in random forests, reliable leaf predictions are usually available in at least some leaves of the tree ensemble, provided enough trees have been trained. Therefore, robust predictions can be derived taking into account estimates from all leaf nodes, and weighting them by a confidence measure in each node, e.g. based on sample impurity. Thus, all estimates can be aggregated to obtain a robust overall estimate from the trained forest [173]. Different strategies to derive position estimates from the random forest predictions during segmentation which will be discussed in chapter 5.

4.5 Summary of Training and Detection Framework

Summarized at this point, the flexible Haar-like feature description, the non-local appearance modeling and the landmark-wise trained random regression forests provide an omni-directional detection of suitable landmark positions during segmentation. With it, the proposed techniques provide the means for a methodological independence of the proposed 3D-SSM with regard to their training and their application.

Landmark appearance is trained without using manually tailored features, and meaningful characteristics for visible anatomical structures are determined automatically from a pool of randomized features during training. The employed random forest regressors introduce the intrinsic capability for a robust and flexible learning to the proposed 3D-SSM.

In particular, the proposed method takes advantage of the robust and easy applicability of random forests known for varying applications from the liter-

ature. Furthermore, landmark appearance is now based on non-local information learned from distributed image locations. The extent of the information coverage during training can arbitrarily be adjusted in the process, thus providing a devised detection range as needed for the later segmentation.

During application, the proposed method is expected to profit from a non-local analysis of distributed image information, and based on this, from an omni-directional landmark detection and from a larger capture range during the iterative model fitting, based on more widely available information on sought object positions at larger distances. From this, two major benefits arise with regard to the application of the proposed 3D-SSM.

First, the proposed method is intended to be employed from arbitrary starting placements in the image. That is, based on the extent of information learned during training, landmarks are supposed to be detected from remote image positions. As explained earlier, the detection accuracy of the landmark estimates is expected to increase as the surface model closes in on the target structure during the iterative model fitting. At the same time, misleading predictions that arise in the process should be compensated by the learned shape prior model.

A second benefit arises for the reproducibility of the segmentation results using the proposed 3D-SSM. The omni-directional detection of landmark positions is intended to produce unambiguous results that are robust against changes of model pose, and significantly more constant results are expected for a model fitting from changing initializations. This should generally lift the predominant influence from model initialization on 3D-SSM, and also remove the importance of an accurate fit of the initial placement with the organ of interest.

Altogether, the proposed techniques are expected to target the well-known weaknesses of 3D-SSM summarized in Sect. 4.1, i.e. strong dependence of results from the previous model initialization, limited capture range at the sought organ of interest, and a high sensitivity against model pose changes, which makes 3D-SSM prone to be stuck in locally optimal solutions during segmentation. By addressing these issues, the proposed 3D-SSM is expected to be better applicable across different use-cases, i.e. to reach higher performance without support from elaborate model initialization methods or from techniques that compensate for sensitivity against locally optimal solutions.

In the following sections, the introduced techniques are all put together for an actual application of the proposed 3D-SSM to clinical data. That is, predictions from the random forest ensemble will be incorporated from distributed

image positions, in order to derive robust position estimates for the modeled landmark positions in unknown images. Furthermore, a multi-scale approach will be applied that serves as a framework for a robust and computationally efficient model fitting during segmentation. For this purpose, a sparsification of surface landmarks and a multi-resolution image pyramid will be introduced to the model fitting process

CHAPTER 5

Model-Based Segmentation Approach

Using the developed techniques from the previous sections, a segmentation procedure will be established for the proposed 3D-SSM. Apart from the benefits that result from the newly employed techniques, the segmentation is based on the usual two step approach for 3D-SSM. That is, a model surface is initially placed in the image that contains a sought target organ. From there, the iterative search for suitable landmark positions is started, which guides the surface model to perform the intended segmentation.

As a difference to the previous approaches for 3D-SSM, the initial model placement will be allowed to be at an arbitrary location in the image. Thus, no upstream method for an accurate and precise model initialization will be required. The omni-directional landmark detection will then guide the subsequent model fitting across increased distances in the underlying images. In the subsequent finer resolutions of the model fitting, the proposed omni-directional landmark detectors are expected to grant a strong robustness against the continuous model pose changes that occur during model adaptation of 3D-SSM, thus providing a high reproducibility of results regardless of model pose, and allowing a more extensive non-local search for optimal solutions during segmentation. This will also be in contrast to previous 3D-SSM, where choice of model initialization and subsequent changes of model pose during model fitting have a strong impact on the final results, and where 3D-SSM regularly get stuck in locally optimal solutions.

Apart from the differences, a multi-scale approach will be followed that is usually employed during model fitting of 3D-SSM. At the different stages of this procedure, varying aims are pursued. That is, in the early stages, position and shape of the target organ will roughly be covered by the surface model. In the subsequent steps, the segmentation result will gradually be refined, until final results can be obtained for the actual image.

In contrast to the typical multi-resolution approaches of previous 3D-SSM, the proposed method introduces a long distance model fitting at the first stage of the model fitting, which leads the surface model from remote image positions to the sought organ of interest. At this stage, global information from across the image domain has to be analyzed and be incorporated in order to bridge the initial large distances during this first model fitting step.

An increased computational complexity is introduced by the employed omni-directional landmark detectors, due to repeated landmark-wise evaluation of voxel positions by the trained random regression forests in each iteration. to To speedup the segmentation process, a sparsification approach will be established that performs detection on a smaller subset of surface landmarks. The sparse surface modeling, the step-wise refinement of results through a multi-resolution approach, and the long distance model fitting as a particular first step of the segmentation will be detailed in the following sections.

5.1 Omni-directional Landmark Detection

This section will outline the use of the previously trained context appearance model for an omni-directional landmark detection during segmentation. Based on the learned image context for each modeled landmark, the trained random regression forests will provide a robust and unambiguous inference of sought landmark positions during segmentation.

In accordance with the learned distributed information of the context appearance model, position estimates will be inferred from different image locations to provide a robust voting for sought landmark positions. This adds to the robustness that is also achieved by means of an aggregation of estimates within the trained forests. Several voting strategies exist for that purpose, some of which have e.g. been proposed by previous approaches in 2D [29], [135]. Unlike these works, a different search and model fitting scheme builds the foundation for the proposed method, which makes the previous voting

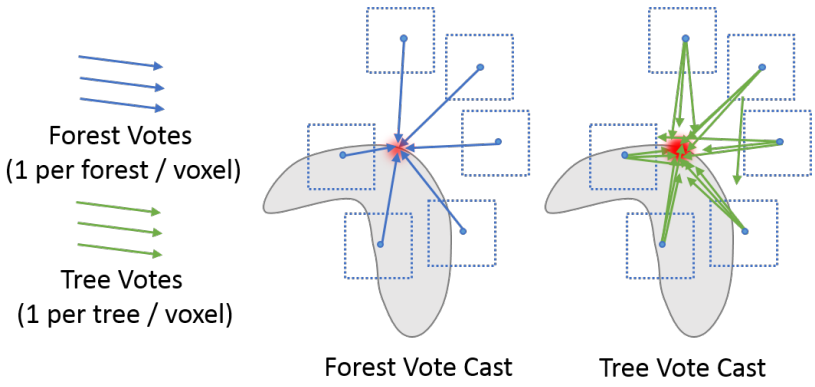


Figure 15: Different voting strategies for the trained random regression forests. Estimates from leaf predictions in the trained ensemble trees are expected to accumulate at the sought landmark position. Left figure: tree predictions are aggregated through confidence weighting in the trained random forests, casting more robust votes to a smaller focal region around the estimated landmark positions. Right figure: tree predictions are directly cast to the image, with or without confidence weighting, thus mapping information from forest leaves to the image domain, resulting in smoother cost functions.

strategies less suited for an application with the proposed method. This will be further discussed in Sect. 5.1.1.

The goal of the proposed landmark-based object description and of the landmark-wise omni-directional search was to address the visibility and model initialization problems of 3D-SSM (cf. Sect. 2.5.3). To the best of our knowledge, these problems of 3D-SSM for medical image segmentation are addressed for the first time by means of an unambiguous appearance modeling and of an omni-directional landmark detection for 3D-SSM in this thesis. The intended use of the presented landmark detection for model fitting of 3D-SSM, and the major differences to previous approaches will be outlined in Sect. 5.1.2.

5.1.1 Voting for Landmark Positions

As motivated above, different schemes can be applied for a voting of sought landmark positions from distributed surrounding locations. In [29], [167], the use of accumulator arrays was proposed where votes are stored in an underlying image or voting map, instead of working directly with the original vector votes as proposed e.g. by [169]. This allows the mapping of votes from different

distributions to the underlying voting maps, and furthermore, it fits the use of voxel-based cost functions that are usually employed by shape-model based approaches [2]. Therefore, the use of accumulator arrays is also well-suited for the proposed method.

However, unlike the previous approaches from [29], [135], another voting strategy will be chosen for the proposed method, due to differences in the underlying nature of the employed search and model fitting approaches. Unlike typical 3D-SSM approaches, the constrained local model approach (CLM) as proposed by [29], [135] performs a joined optimization of pose and shape parameters, and of the landmark-wise cost functions across all landmarks of the employed shape model. This means that instead of a landmark-wise selection of optimal candidates, shape and pose parameters are chosen as to determine a global solution that considers cost functions from all landmark positions.

Accordingly, strategies were proposed by [29] for a direct mapping of weighted or unweighted tree votes, of sample distributions or of separate training samples from the leaf nodes to the accumulator array. Estimates are not aggregated in the forests prior to their mapping, which makes sense for the CLM since the added uncertainty projection from the unaggregated votes provides an overall smooth cost function for optimization through the CLM (cf. Fig. 15, right).

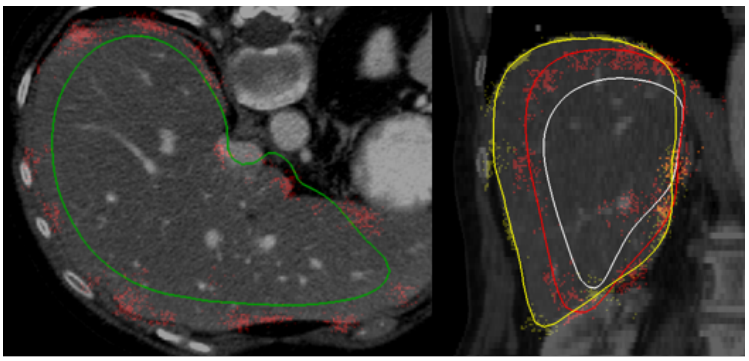


Figure 16: Voting map creation during model fitting. Each cluster of votes represents a position estimate for one particular model landmark. (a) Landmark-wise votes (red) cast from a momentary model position (green contour). (b) Votes and model updates for two subsequent iterations i_n: i_0 (white contour, red votes), i_1 (red contour, yellow votes) i_2 (yellow contour).

The proposed method however builds on a 3D-SSM approach that relies on an accurate landmark-wise detection of optimal candidate positions during

segmentation. Instead of a smooth cost function as used by the approaches from [29], [135], an optimized estimation of landmark-wise positions is preferable, where the additional uncertainties from above are resolved prior to the vote cast, by means of a tree aggregation of the trained regression forests. The proposed voting strategy therefore incorporates the original notion of a bagging of tree predictors in random forests, and votes of a generally higher confidence are produced by the forest predictions. Nevertheless, the voting maps that are produced by the chosen approach still qualify for use as a cost function. Thus, an incorporation of additional optimization schemes for 3D-SSM is possible if desired (cf. [2]).

Let $V = \{v_1, \cdots, v_n\}$ denote the set of voxels from which estimates on the sought landmark position are infered during segmentation. At each $v_j \in V$, the trained random regression forest produce estimates $\vec{d_i}, \quad i = 1, \cdots, k$ for the sought landmark position. Each $\vec{d_i}$ is weighted by a confidence measure $C_i(v_j)$ which is defined for all leaf nodes in the forest, and tree predictions are aggregated to obtain the overall forest prediction $\vec{D}(v_j)$ for v_j:

$$\vec{D}(v_j) = \frac{\sum_{i=1}^{k} C_i(v_j)\vec{d_i}(v_j)}{\sum_{i=1}^{k} C_i}$$

The $C_i(v_j)$ represent the uncertainties in each leaf node, for which we choose the impurity of leaf nodes as in [29]:

$$C_i(v_j) = |\Sigma_i|^{-1}$$

where $|\Sigma_i|$ denotes the determinant of sample covariance or, for the additionally proposed total variance based measure (cf. Sect. 4.4) the sum of sample variance, in the i-th leaf node that contributes a prediction on $\vec{D}(v_j)$.

$\vec{D}(v_j)$ then yields the *confidence-weighted mean* of the different tree estimates in the forest. From all forest predictions $\vec{D}(v_j)$ at the different v_j, the 3D voting map $M(i, j, k)$ is created:

$$M(i, j, k) = \sum_{l=1}^{n} 1_{i,j,k}(P(v_l))$$

where $1_{i,j,k}P(v_l)$ denotes the voxel membership of the vector-valued predictions

$P(v_l)$:

$$1_{i,j,k}(P(v_l)) = \begin{cases} 1 & \text{if } P(v_l) \text{ is part of voxel } (i,j,k) \text{ in } M \\ 0 & \text{otherwise} \end{cases}$$

In contrast to tree-wise votes as in [29], [135], a forest-vote is cast on the voting map for each v_j by the chosen approach (cf. Fig. 15, left). Votes from the tested voxel positions tend to accumulate in a narrow region around the sought landmark positions in the image. After a Gaussian smoothing of each vote cluster, a position candidate for each landmark is retrieved from the regional maximum of the vote clusters.

After the extraction of landmark positions from the voting maps, the resulting position estimates are available for the subsequent model fitting (cf. Sect. 3.3).

5.1.2 Adding to the Model Fitting of 3D-SSM

The previous sections have described the use of a non-local context-based appearance modeling and of an accumulated regression voting for the proposed 3D-SSM. These concepts are intended to provide an omni-directional landmark-wise search during segmentation, based on the trained random regression forests. Specifically, the proposed method builds on the ability of its landmark detectors to separately pinpoint positions of the modeled surface landmarks in underlying images.

During segmentation, the landmark-wise trained regression forests make predictions on suitable landmark positions based on information from distributed image locations. The resulting position estimates point towards plausible landmark positions in the image and thus provide an input for the model fitting in each iteration. As the modeled surface closes in on the organ of interest, estimates from the vicinity of the sought landmark positions and thus more accurate detection results are produced as the model fitting progresses.

The visibility of landmark positions for each modeled surface landmark is essentially determined by the provided detection radius during appearance model training, which makes landmark visibility in essence independent from model pose. This should provide the means for the proposed method to address the model initialization, capture range and visibility problems of 3D-SSM.

With its incorporation of the above techniques, the proposed method is the first 3D-SSM that addresses these well-known problems of 3D-SSM by means of

unambiguous and spatially extended landmark detection during model fitting. This is e.g. in contrast to original 2D approaches from [29], [135], where landmarks are detected jointly across the shape model by the proposed constrained local model (CLM) approach. In particular, detection results are dependent from a joint optimization with the model pose in the CLM. By contrast, the proposed method builds on a separate detection in each modeled landmark independently from pose and shape, with the intended goal of an unambiguous landmark detection from differing model poses. Also, this differs from other regression-based approaches which rely on a joint estimation of shape and pose parameters or shape-wide landmark displacements like e.g. random ferns (cf. Sect. 2.4.3).

5.2 Sparse Surface Modeling

As motivated at the beginning of the chapter, regarding the high number of forest predictions derived at various image locations and in each iteration, a lowering of computational costs becomes essential. To tackle the resulting computational complexity, a sparse surface modeling is applied where landmark detection is conducted only for a sparse subset of distributed landmarks across the modeled surface.

For an adaptation of the surface model throughout the fitting process, meaningful updates for all landmark positions are required, as described in Sect. 3. For this purpose, the required surface-wide position updates for all landmarks will be derived from the sparse subset based on an approach that is similar to [180]. That is, a transfer of position updates from the detected sparse landmark positions to the in-between surface landmarks will be made based on geodesic distances on the surface model.

Be $L_s^t = \{\vec{l_1^t}, \cdots, \vec{l_k^t}\}$ sparse landmark positions at iteration t and

$$L_s^{t+1} = \{\vec{l_1^{t+1}}, \cdots, \vec{l_k^{t+1}}\}, \quad \vec{l_i^{t+1}} = \vec{l_i^t} + \vec{dl_i^t}$$

position updates for iteration $t+1$. For each non-sparse landmark $\vec{r_i^t}$, a set of sparse landmarks from the neighborhood are selected using

$$N(\vec{r_i^t}) = \{\vec{l_i^t} \in L_s^t \mid \|\vec{l_i^t} - \vec{r_i^t}\| \leq \chi\}$$

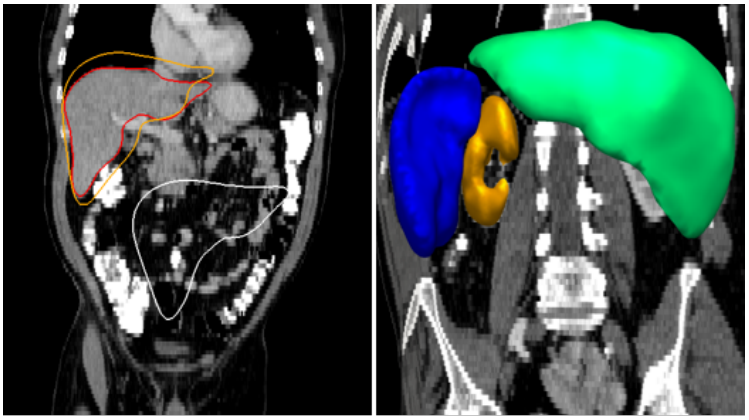
within a predefined *geodesic distance* χ. From the selected positions, updates are then propagated to $\vec{r_i^t}$ after a normalization on the underlying geodesic

distances:

$$r_i^{\vec{t+1}} = \frac{\sum_{l_i^{\vec{t}} \in N(r_i^{\vec{t}})} \|\vec{l_i^t} - \vec{r_i^t}\| l_i^{\vec{t+1}}}{\sum_{l_i^{\vec{t}} \in N(r_i^{\vec{t}})} \|\vec{l_i^t} - \vec{r_i^t}\|}$$

In contrast to [180], where updates are propagated based on a gaussian kernel weighting, the proposed interpolation of position updates clearly produces favorable results for the proposed 3D-SSM.

5.3 Multi-Resolution Search



(a) (b)

Figure 17: Image segmentation by the proposed method. (a) Long-distance model fitting: arbitrary initialization (white), long-distance result (orange) and final segmentation (red). (b) Exemplary segmentations of multiple organs produced by the proposed method: liver (green), spleen (blue) and left kidney (orange).

As part of the multi-scale segmentation approach (cf. Sect. 5), a hierarchical search is conducted based on sub-sampled versions of the original image, following the notion of a multi-scale image pyramid [181], [182], proposed for a faster and more robust shape model fitting during segmentation, while removing unnecessary clutter from image details that are not needed until the later stages of the segmentation.

Resampled versions of the original image are created with isotropic spacings that are doubled at each step of the pyramid, and versions of the non-local

appearance model are trained in all resolutions. Since the 3D-SSM is expected to close in on the organ of interest with increasing resolutions, detection range is reduced stepwise with the trained versions of the appearance model.

Thus, focus shifts from a remote detection of the organ of interest to an accurate detection of nearby landmarks. The coarser resolutions emphasize image features that describe the greater anatomical context needed to guide the model fitting from distant image locations to the target organ. The finer resolutions gradually reveal the local detail needed to cover the shape of the target organ and to finally derive a segmentation.

Image segmentation starts with a long-range model fitting on the coarsest resolution, where the mean shape of the 3D-SSM is guided from remote image locations to the organ of interest (cf. Fig. 17). The subsequent finer model fitting stage adjusts shape and pose parameters to provide a more accurate fit to the organ of interest. Finally, the most accurate model fitting stage is applied, followed by a profile-based adaptation for all surface landmarks such that a free surface deformation can be applied as proposed in [147]. This way, shapes that exhibit high anatomical variability can be covered that are otherwise excluded from the hard shape constraint.

Altogether, the presented segmentation approach shows characteristics that are substantially different from previous multi-resolution approaches. The proposed 3D-SSM is expected to achieve a high reproducibility of results throughout all stages of the multi-resolution approach, and at remote locations from the organ of interest, based on the intended robustness of landmark detection against changes of model pose during segmentation.

The choice of overlapping landmark detection ranges during appearance model training is expected to provide high robustness throughout all stages of the proposed model fitting. At the same time, the proposed omni-directional search should guide the model fitting to solutions that are generally closer to the global optimum of the segmentation, compared with the typical unidirectional search schemes of previous 3D-SSM. The expected reproducibility of results and to the improved non-local search of the proposed 3D-SSM should remove the need for an optimal model initialization, which significantly facilitates application and optimization of 3D-SSM towards higher performance.

This is in contrast to previous 3D-SSM, where the model initialization step is required to provide a suitable fit with the organ of interest, in order to provide optimal results during the subsequent model fitting. The typical model fitting of previous 3D-SSM is highly unpredictable in its course, and is prone to be stuck in locally optimal solutions during segmentation. Furthermore,

the capture range of these 3D-SSM is typically strongly limited, especially at high curvature surface parts. Other than to provide a best possible initial fit with the organ of interest, optimality criteria for the model initialization with regard to the outcome of the subsequent model fitting are generally unclear. This gives rise to more complex methods for model initialization and to compensatory measures during segmentation, but still, model initialization remains an unpredictable source of error. Altogether, these problems affect the general performance and applicability of previous 3D-SSM, which are now directly addressed by the chosen approach of the proposed method.

In the following chapter, the proposed method will be evaluated with regard to its robustness against changes of model pose during adaptation of 3D-SSM, its added global capture range for 3D-SSM within the image space, its capability to perform a more non-local search for optimal solutions during segmentation, and its generalizability to different segmentation tasks and use-case scenarios.

CHAPTER 6

Validation and Application of the Proposed Method

For an evaluation of the proposed method, segmentation experiments were conducted for different organs on various clinical image datasets. The experiments were laid out to reflect the overall goals of this thesis, i.e. higher generalizability for 3D-SSM by means of a robust and flexible appearance model training, and of a stronger independence from auxiliary methods. Prior to the presentation of the conducted experiments, some introductory information will be provided in the following section that is helpful for an understanding of these experiments.

6.1 Fundamentals of the Evaluation

The experiments in this thesis will in large part be based on a measurement of the segmentation accuracy produced by the proposed method and by other state-of-the-art approaches. For this purpose, meaningful and widely-used accuracy measures have been proposed in the literature. They will be described in Sect. 6.1.1.

Also, as a part of its generalizability, the proposed method is applicable to different use-cases without thorough optimization of forest training and voting parameters. Thanks to the straightforward nature of the employed random regression forests, a comprehensible set of parameters is easily found that can then be reused in different applications. The parameter set that was chosen

for the experiments of this thesis will be described in Sect. 6.1.2.

Different attributes and qualities of the proposed method contribute to the overall goals of this thesis. They take effect on several stages during training and during application. To better understand how these aspects of the proposed methods are reflected by the conducted experiments, Sect. 6.1.3 will present an outline of all experiments together with an explanation of their purpose.

6.1.1 Segmentation Performance Measuring

6.1.1.1 Analytical and empirical performance measuring

Various approaches exist for the performance evaluation of segmentation algorithms [183]. Some conduct a performance analysis at a methodological level, where algorithmic prerequisites, basic assumptions and complexities of each segmentation method are assessed. They allow a comparative evaluation of the conditions and prerequisites that need to be fulfilled prior to successful segmentation, which can be highly relevant for the applicability of a particular method in clinical applications. Such a qualitative analysis will be part of the evaluation of the proposed method in Sect. 6.3.3.

In addition to the qualitative analysis of the overall segmentation method, an evaluation of its partial techniques can be conducted. Such an analysis can provide additional insight to the conceptual capabilities of the examined methods, however it is not always suited for a direct comparison of different methods.

Finally, a qualitative analysis of results from different segmentation methods can be performed. Such an analysis aims at the plausibility of the results and at their geometric properties like homogeneity and smoothness of the segmentations. It allows a direct comparison of different methods and a better understanding for occurring types of segmentation errors that are relevant for clinical applications.

6.1.1.2 Quantitative measures of segmentation accuracy

Instead of a qualitative analysis of segmentation results, a number of quantitative measures can be introduced that allow an evaluation based on manual expert annotations. These are usually provided by one or more medical experts and reflect the expected ground truth from a clinical point of view. On a dedicated test dataset, the quantitative measures can then be used to assess

the agreement of automatically produced results with the underlying manual annotations.

Several well-known and widely-used quantitative evaluation measures have been proposed in the literature [184], [185]. They are either based on a voxel-wise comparison of volumetric segmentations or on a comparison of the resulting surface delineations. The volume-based measures do a comparison by analyzing the overlap of segmentation volumes and their discrepancy. The surface-based measures on the other hand quantify surface discrepancies between two segmentations based on point-wise differences. Each measure is able to reflect certain error characteristics of the segmentations, which allows deeper insight into the nature of encountered errors.

In the following, A and B denote two binary segmentations of one original image that are to be compared, where voxels are labeled to belong either to the target object or to its background. Then, one quantitative measure for the overlap of A and B is given by the *Dice coefficient*:

$$\mathrm{Dice}(A, B) = \frac{2 \cdot |A \cap B|}{|A| + |B|}$$

where $|A|$ and $|B|$ are the measured volumes of the binary segmentations. Another measure that quantifies volume overlap is given by the *Tanimoto coefficient* (also known as *Jaccard coefficient*):

$$\mathrm{Tanimoto}(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Both coefficients map the measured volume overlaps to the range of $[0; 1]$, where 0 means no overlap and where 1 reflects the identity of the volumes. Based on Dice and Tanimoto coefficients, the according volume overlap error percentage (VOE) is sometimes used:

$$\mathrm{VOE_D} = 1 - \mathrm{Dice}$$

and

$$\mathrm{VOE_T} = 1 - \mathrm{Tanimoto}$$

When compared, the Tanimoto coefficient is more sensitive to small overlap errors than the Dice coefficient. Apart from this, $\mathrm{VOE_T}$ is a metric since it also satisfies the triangle inequality.

In addition to these overlap-based measures, segmentation errors can also be assessed based on the difference of segmentation volume. This can be done

using the *relative volumetric distance* measure:

$$\mathrm{RVD}(A, B) = \frac{|A| - |B|}{|B|}$$

where $|B|$ can be a ground truth segmentation that is used as a reference for $|A|$. The RVD reflects volumetric errors which can be of particular interest in applications that rely on organ volume as a clinical parameter. Apart from this, the RVD is complementary to the above overlap measures, i.e. a low RVD does not necessarily imply a low overlap error of the segmentations.

In addition to the above volume-based measures, the use of surface-based measures allows an assessment of segmentation error based on the agreement of the segmented surfaces.

In the following, S_A and S_B denote the surfaces of A and B, and V_A be a discrete set of vertices in S_A. Then, the *average surface distance (asd)* describes the average landmark-wise deviation of S_A from S_B:

$$\mathrm{asd}(S_A, S_B) = \frac{1}{|S_A|} \sum_{v \in V_A} \min_{x \in S_B} ||v - x||$$

Distances are determined for each modeled surface landmark v on one mesh S_A with regard to the nearest points x on the other surface S_B. The *root mean square distance (rmsd)* is based on squared surface distances instead:

$$\mathrm{rmsd}(S_A, S_B) = \sqrt{\frac{1}{|S_A|} \sum_{v \in V_A} \min_{x \in S_B} ||v - x||^2}$$

In contrast to the average surface distance, the root mean square distance is more sensitive to outliers. Thus, it additionally pronounces outlier deviations from the ground truth surface. For an isolated outlier measurement, the *Hausdorff distance* can be used that works with the maximum distance between S_A and S_B:

$$\mathrm{hd}(S_A, S_B) = \max_{v \in V_A} \left(\min_{x \in S_B} ||v - x|| \right)$$

The above surface-based measures are asymmetric with regard to an origin and a target surfaces S_A and S_B (e.g. a reference segmentation). For a lesser biased measurement regarding the choice of A and B, a symmetric version of

each measure can be used based on their reciprocal application:

$$\mathrm{ASD}(S_A, S_B) = \frac{1}{2}(\mathrm{asd}(S_A, S_B) + \mathrm{asd}(S_B, S_A))$$
$$\mathrm{RMSD}(S_A, S_B) = \sqrt{\frac{1}{2}(\mathrm{rmsd}(S_A, S_B)^2 + \mathrm{rmsd}(S_B, S_A)^2)}$$
$$\mathrm{HD}(S_A, S_B) = \max\left(\mathrm{hd}(S_A, S_B), \mathrm{hd}(S_B, S_A)\right)$$

As for the purpose of the presented quantitative accuracy measures, there are also particularities to be considered. First, these measures merely offer a global representation of segmentation error, where local errors are masked behind the global measure. At most, the maximum surface distance indicates the presence of larger discrepancies at certain surface locations. However, these measures offer a valuable error and accuracy quantification, and in combination with a qualitative analysis of segmentation results, they are ideally suited for a thorough evaluation of different algorithms.

Second, these measures usually reflect the agreement of segmentation results with a ground truth that is defined only by an individual group of medical experts. This means that good accuracy scores do represent meaningfulness based on best practices in the medical community. However, they do not take into account the limited reproducibility of annotation by human raters, or the disagreement across different raters (*inter- and intra-observer variance*).

In [186], a scoring system was proposed based on quantitative performance measures. There, the inter-observer variability is determined using annotations from a second rater. Then, scores are awarded for the achieved segmentation accuracy that relate to the underlying inter-observer variability.

6.1.2 Training and Segmentation Parameters

One key feature of the proposed 3D-SSM is a higher generalizability regarding its application across different use-cases. Generalizability is on the one hand implemented during the proposed appearance model training, which strongly builds on robustness and flexibility of the employed random regression forests in conjunction with the underlying Haar-like feature descriptor. The straightforward nature of random forests regarding their training and their application is widely known, and is expressed by an easy and uncomplicated parameterization during training and during application.

The effect of the involved parameters on the detection results is highly intuitive, and strong performance can be reached in a wide parameter range.

Regarding the proposed method, this holds true for a sufficient number of training samples from an adequate range in image domain, for a reasonable number of tree predictions during segmentation, and for a sufficiently large image context reflected by the employed feature descriptor. The robustness of the trained landmark regressors will also be part of the experiments in Sect. 6.5.

During training, forest parameters comprise the radius and density of samples taken from the underlying training images during forest training (cf. Sect. 4.4), the number of ensemble trees that are trained in each forest (cf. Sect. 4.4), and the maximum patch size of the created 3D Haar-like features (cf. Sect. 4.3.2). During segmentation, forest parameters involve the vote count and the area in which votes are raised during the regression voting (cf. Sect. 5.1.1). As described in Sect. 4.3.2), a variance normalization of Haar-like features can compensate for intensity inconsistencies encountered during segmentation, which strongly applies to the segmentation of MRI data (cf. Sect. 6.4.2).

Table 1: Forest training and voting parameters for the proposed method.

Level of detail	0	1	2
Spacing	4mm	2mm	1mm
Sampling Radius	60mm	30mm	15mm
Samples / LM / Img	2000	1000	400
Feature Patch Size	120mm	30mm	15mm
Number of Trees	12	12	12
Maximum Tree Depth	max	max	max
Stop Splitting at Sample Count	4	4	4
Number of Random Features / Node	100	100	100
Voting Range (mm^3)	$36 \times 36 \times 36$	$18 \times 18 \times 18$	$9 \times 9 \times 9$
Votes / LM	$9 \times 9 \times 9$	$9 \times 9 \times 9$	$9 \times 9 \times 9$

Tab. 1 shows the parameter values that were chosen for use in all experiments of this thesis. The rationale behind their choice was based on an increasing spatial context and sample coverage required for the coarser resolutions, and on an immediately well-working voting strategy that allows robust landmark detection during segmentation.

The choice of parameters for the different resolutions largely followed the doubling of spacings in the employed multi-resolution pyramid (cf. Sect. 5.3). One exception was made for the patch size of the coarsest resolution. To intuitively support the more global organ detection during the long distance model fitting with structural features found at larger distances, the patch size was additionally doubled. The chosen training parameters were used in all experiments of this thesis.

Regarding application of the proposed method, generalizability is provided by the ability to be employed as a single, standalone segmentation technique. This means in particular that no initialization techniques have to be provided for use of the proposed method in different cases, and that a parameter tuning of underlying techniques is limited to the segmentation parameters of the proposed method only.

For the casting of votes during segmentation, it was found that votes from densely packed regions (i.e. with votes cast from a coherent region of neighboring voxels) immediately provided better results than more widely spread votes (i.e. from regions in which no votes are cast from some of the voxels). Apart from this it was found that in single cases, votes that were taken from inside the surface model provided better results than votes that were taken from inside and outside the modeled surface around each landmark position. Thus, the voting region was shifted inwards in normal direction of the surface model by an offset of 8mm. Also, to provide for votes best being cast from inside the object of interest, the first resolution of the model fitting was started from a half-size down-scaled shape on the organ position found by the long-distance model fitting.

6.1.3 Overview of the Experiments

In the segmentation experiments of this thesis, the proposed method was evaluated with regard to the following intended goals: to reach an independence from previous model initialization for the proposed 3D-SSM; to make 3D-SSM less prone to be stuck in local optima during segmentation, thus improving overall accuracy; and to enable robust and flexible standalone applicability of 3D-SSM across different use-cases without necessity for additional auxiliary methods.

To assess the independence of the proposed 3D-SSM from the previous model initialization, a series of segmentation experiments was conducted. There, the sensitivity of results against changes of model initialization was assessed

Table 2: Number of sparse landmarks and total number of landmarks for liver, spleen, kidney, left ventricle and lung models.

Modeled Organ	Sparse Model Landmarks	All landmarks
Liver-DKFZ	119	2562
Liver-Challenge	105	2600
Liver-Multiorgan	112	2600
Spleen	104	2263
Left Kidney	123	1659
Liver-MRI	110	2500
Ultrasound-LV	235	2033
MDCT-LeftLung	163	1498
MDCT-RightLung	158	1497

for the proposed method in comparison with a typical profile-based 3D-SSM. Then, the capture range of the proposed model fitting was examined, i.e. the maximum distance from which the proposed 3D-SSM is reliably attracted to the underlying organ of interest. Both robustness against changing initializations and large capture range build the foundation for an application of the proposed method as a standalone technique. These experiments will be presented in Sect. 6.2.

To examine how the proposed omni-directional landmark detectors help avoid local optima during segmentation, and whether more global solutions can be found instead, liver segmentation experiments were carried out on CT image data. The high shape variability of the liver stresses the visibility problem of 3D-SSM, which makes this task well suited to demonstrate the benefits of the omni-directional search. Also, many sophisticated approaches have been proposed for the task and are thus available for comparison with the proposed method. These experiments will be presented in Sect. 6.3.

To assess the standalone applicability of the proposed method, it was employed in a variety of segmentation scenarios with unchanged training parameters, and without the use of auxiliary methods for model initialization, image preprocessing, model steering, or additional segmentation techniques. Its overall accuracy was assessed in comparison with other state-of-the-art segmentation approaches. With the chosen segmentation tasks, a variety of

conditions were covered that are typically encountered in clinical scenarios. These experiments will be presented in Sect. 6.4.

To demonstrate the ability of the proposed method for a segmentation of distributed organs in larger image volumes, the proposed method was employed for liver, spleen and left kidney segmentation in a public multi-atlas labeling challenge. Organ detection in image volumes of differing size, heterogeneous imaging quality, and variability of the underlying morphology are the challenges that were encountered in this case. This experiment will be described in Sect. 6.4.1.

To demonstrate the generalizability of the proposed method across different imaging modalities, a liver segmentation was conducted on MRI image data. Apart from the importance and from the extensive use of this modality for clinical diagnostics, the underlying image data serves as an example for strong local and global inconsistencies of image intensities. This experiment will be presented in Sect. 6.4.2. For similar reasons, the proposed method was further employed in a left cardiac ventricle segmentation task on 4D ultrasound image data. Motion artifacts, occlusions and strong noise make segmentation in this imaging modality particularly challenging. This experiment will be presented in Sect. 6.4.3.

In Sect. 6.4.4, an application of the proposed method in a typical clinical use case will be presented. There, the proposed method was used for a lung segmentation in high resolution CT data in the presence of a pronounced appearance-altering pathology. Performance was assessed through qualitative evaluation and through a histogram-based analysis of results, in comparison with an elaborate region-based lung segmentation method.

Speeding up the appearance model training can become particularly important when facing the sheer plurality of possible clinical applications. Therefore, speedup options of the employed random regression forest training were examined in Sect. 6.5. There, the robustness of results from the trained random regression forests was assessed for parameter changes that reduce training times, and for architectural modifications that speedup the underlying forest splitting. Furthermore, the well-known robustness of the employed random regression forests is of importance for the generalizability of the proposed method. To demonstrate this robustness also for the proposed 3D-SSM, the influence of training parameters was additionally examined that carry the most weight during an application of the proposed method, i.e the image context reflected by the underlying feature descriptor, and robustness of tree aggregation from the previously trained tree ensembles during segmentation.

6.2 Independence from Model Initialization

The goal of the experiments in this section was to evaluate how the proposed method copes with the known sensitivity of 3D-SSM from previous model initialization. The dependency of 3D-SSM from an accurate model initialization near the sought organ of interest is a widely known problem. It means on the one hand that segmentation results of 3D-SSM generally vary with changing model initializations, and that non-suitable starting conditions cause segmentation errors. The reason for this is that the outcome of the typical unidirectional search strongly depends from the model pose at the beginning of the search. The dependency from model initialization means on the other hand that a positioning of the model surface near the sought organ is essential for the mere applicability of 3D-SSM. Without it, the segmentation is usually guaranteed to fail.

To demonstrate how the proposed method copes with both parts of the problem, two experiments were conducted on a CT dataset of the human liver. First, the degree of model pose invariance was evaluated that is gained from the proposed omni-directional search, i.e. the reproducibility of results for pose changes of the underlying surface model. For this purpose, the precision of results of the proposed method was assessed for changing initial placements around the liver. This experiment will be described in Sect. 6.2.2.

Second, the capture range of the proposed model fitting was examined, i.e. the distance from which the 3D-SSM is reliably attracted to the sought organ of interest. A larger capture range was expected by the proposed method due to its non-local appearance modeling and the employed long distance model fitting, thus allowing an application of the proposed method without an accurate model initialization. This experiment will be presented in Sect. 6.2.3.

6.2.1 Data Material

The dataset that was used in the experiments of this section was taken from a larger collective of 99 abdominal CT scans (DKFZ_ABDOM) (cf. [187]). Images were acquired as part of surgical planning and clinical studies at the German Cancer Research Center Heidelberg. Images were recorded in the central venous phase after application of contrast agent, using scanners from different manufacturers. The underlying image geometry exhibits an in-plane resolution of ~ 0.55 mm $- 0.80$ mm, and an inter-slice distance of 3.0 mm and of 5.0 mm in some cases.

For the initialization experiments in this section, a subset of 45 CT volumes

(DKFZ_LIVER) from the DKFZ_ABDOM dataset was used. The volumes covered the entire abdomen, and additionally in some cases part of the thoracic region and the pelvic region. More than half of the volumes contain visible liver lesions that mainly affect the tissue appearance of the liver. No surgical alterations of the liver were encountered in the dataset. Gold standard annotations were produced on the dataset by manual slice-by-slice segmentation from radiological experts.

6.2.2 Model Pose Invariance

Based on its omni-directional search, the proposed method is expected to reach a higher degree of invariance under model pose changes during segmentation. That is, landmark detection results are supposed to be largely reproducible regardless of pose changes of the employed surface model during segmentation. As a result, higher robustness is expected from the proposed method against an influence from changing model initializations prior to the segmentation.

To evaluate the robustness, an experiment was conducted where the proposed method was started from randomized model placements around a predefined optimal model placement. Then, the reproducibility of segmentation results from the changing model initializations was evaluated, by measuring their dispersion in comparison with the underlying ground truth. The experiment was carried out for increasing degrees of degradation in order to evaluate the robustness of the proposed method against a growing influence from erratic model initializations. To exclude the case that a reproducibility of results is primarily reached in the first long distance model fitting step where no adaptation of shape parameters is performed, segmentation of the proposed method were started from the first LOD, where shape and pose parameters are fitted based on the conducted omni-directional search.

The results of the proposed method were compared with a typical 3D-SSM that solely relies on local appearance modeling and on a unidirectional search during segmentation [147], [188]. A detailed description of the compared method and of its general characteristics can be found in Sect. 6.3.2.1.

Liver models were trained for the proposed method on 5 cross validation sets of a subpart of the DKFZ_LIVER dataset, where 5 images were left out during training for testing in each set. The predefined optimal placement that served as a starting point for the pose randomization was obtained from the pose parameters of the left-out training shape. Then, translation, rotation and scale of the assumed optimum were randomly altered to generate a number of

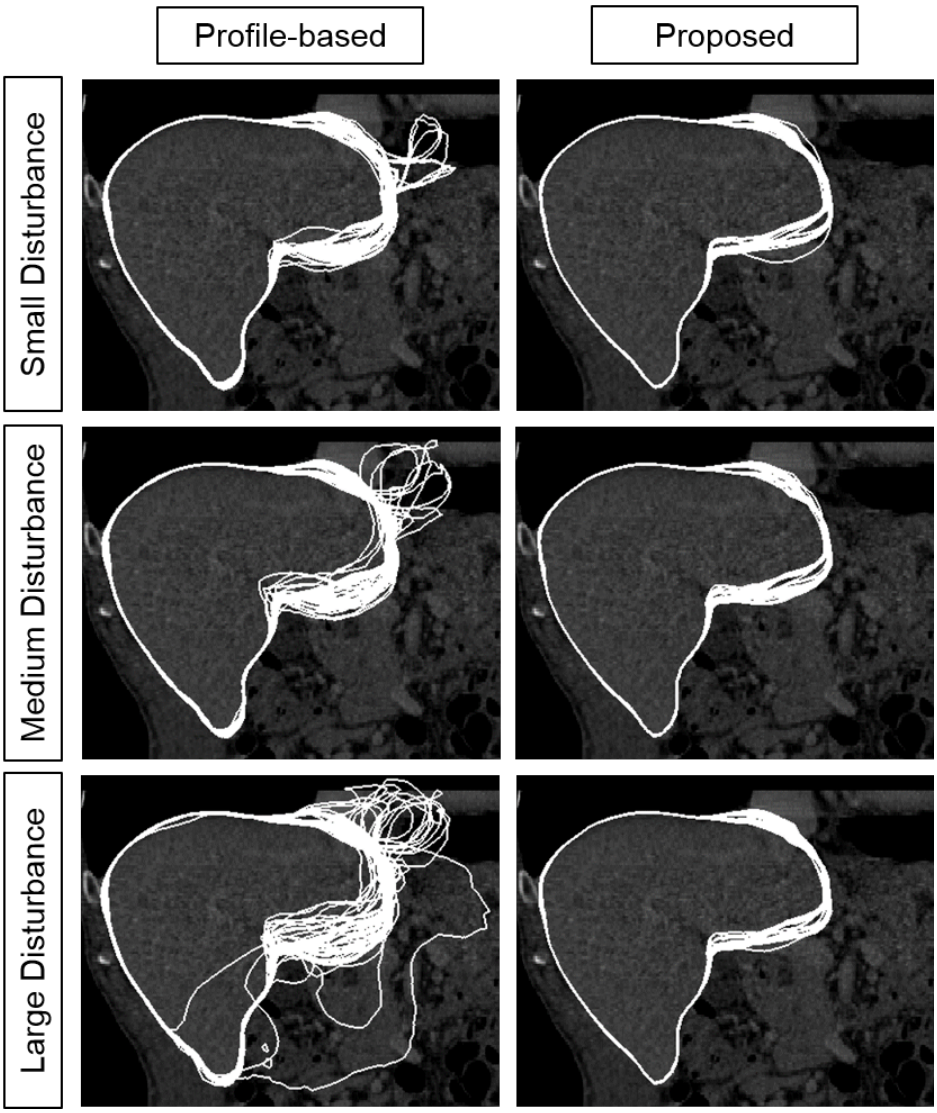


Figure 18: Influence of random pose changes on model fitting results (qualitative). When started without long distance model fitting from 150 increasingly degraded model initializations around the liver (top to bottom row), the proposed method displayed strong robustness against the induced pose changes (right column) in contrast to the profile-based 3D-SSM (left column). Three degradation levels were applied: small (translation t<6.0mm, rotation r<0.4 and scaling s<0.1, n=50), medium (t<8.0mm, r<0.6, s<0.15, n=50) and large (t<12.0mm, r<0.8, s<0.2, n=50).

Table 3: Random degradation of model initializations. Three levels of degradation from the assumed optimal positioning were randomly applied to the model initialization of the proposed method.

Magnitude of degradation	Small	Medium	Large
Translation	≤ 6 mm	≤ 8 mm	≤ 12 mm
Rotation	≤ 0.4 rad	≤ 0.6 rad	≤ 0.8 rad
Scale	$\leq 10\%$	$\leq 15\%$	$\leq 20\%$

150 degraded starting placements (50 at each degradation level). The three levels of degradation are specified in Tab. 3.

Fig. 18 shows cross sections of the results from all model initializations at the different degradation levels, i.e. for the proposed method on the right side and from the profile-based 3D-SSM on the left side. Results are shown from top to bottom for the three increasing degradation levels.

The segmentations from the profile-based method show a growing number of larger segmentation errors throughout the aggravated stages of the experiment (cf. Fig. 18 left column). Also, several locations can be identified where visible imprecisions of results arise. Precision declines visibly with growing magnitude of the dislocations. In direct comparison, the results of the proposed method (Fig. 18 right column) exhibit lesser segmentation errors and an overall higher precision of results at different parts of the segmentation. In particular, no increase of visible segmentation errors and imprecisions can be observed with growing degradation of the model initializations.

Fig. 19 shows the quantitative segmentation results for all 45 cases in all degradation levels, as measured by the Jaccard coefficient for volume overlap. Each column represents 50 segmentations from varied model poses around the liver. The results show that the proposed method produced highly constant results regardless of the magnitude of pose variations. Three exceptions were observed, i.e. cases 26, 28 and 33. In cases 26 and 28 with few exceptions, segmentation generally failed, regardless of model initialization. Thus, these cases clearly stand out as outlier segmentations for the proposed method. Case 33 shows a few outliers that occur depending on the chosen model pose. However, apart from the single outlier cases, the remaining segmentations of case 33 show a very high precision.

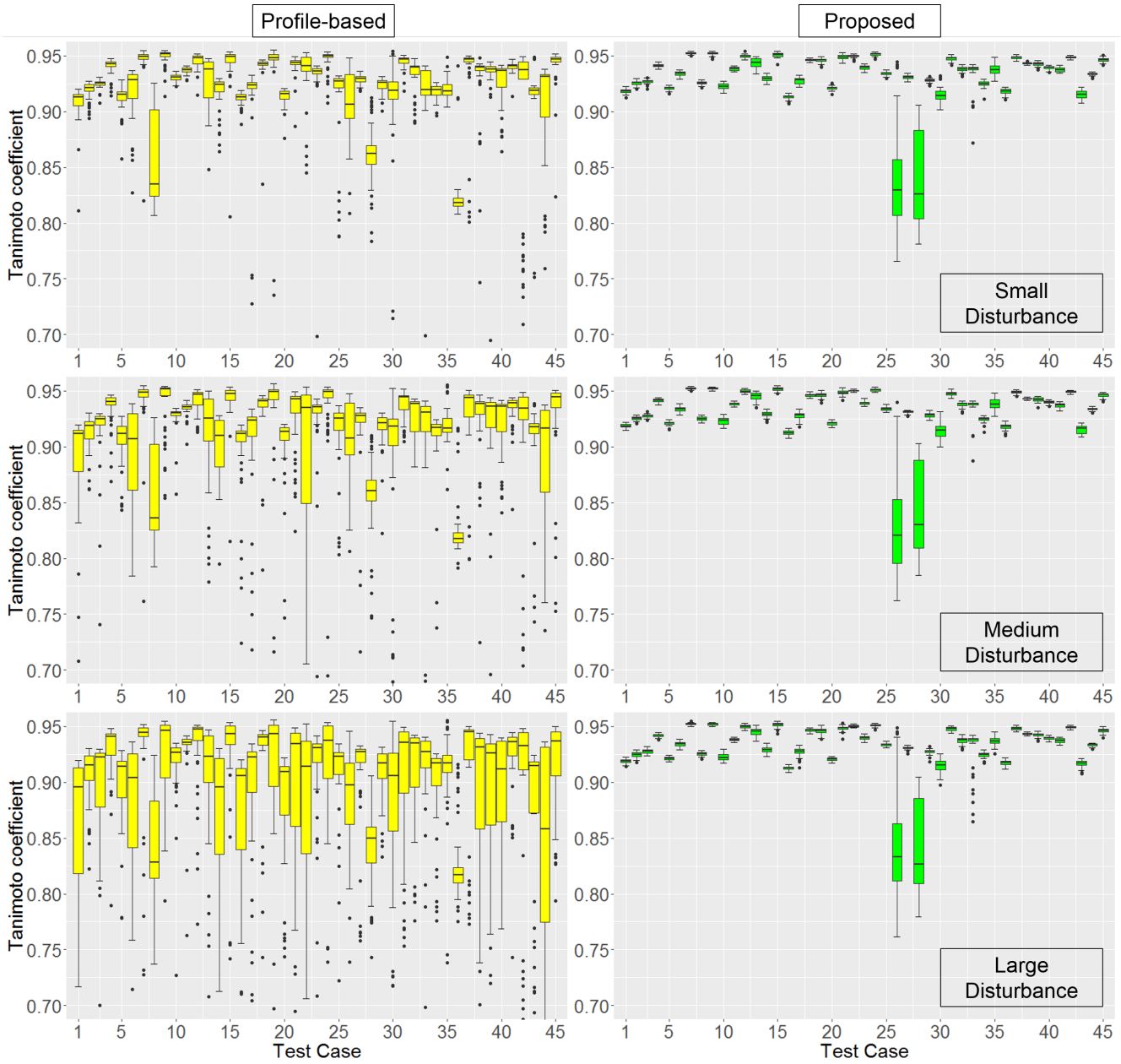


Figure 19: Influence of random pose changes on model fitting results (quantitative). When started without long distance model fitting from 150 increasingly degraded model initializations around the liver (top to bottom row), the proposed method displayed strong robustness against the induced pose changes (right column) in contrast to the profile-based 3D-SSM (left column). Three degradation levels were applied: small (translation t<6.0mm, rotation r<0.4 and scaling s<0.1, n=50), medium (t<8.0mm, r<0.6, s<0.15, n=50) and large (t<12.0mm, r<0.8, s<0.2, n=50).

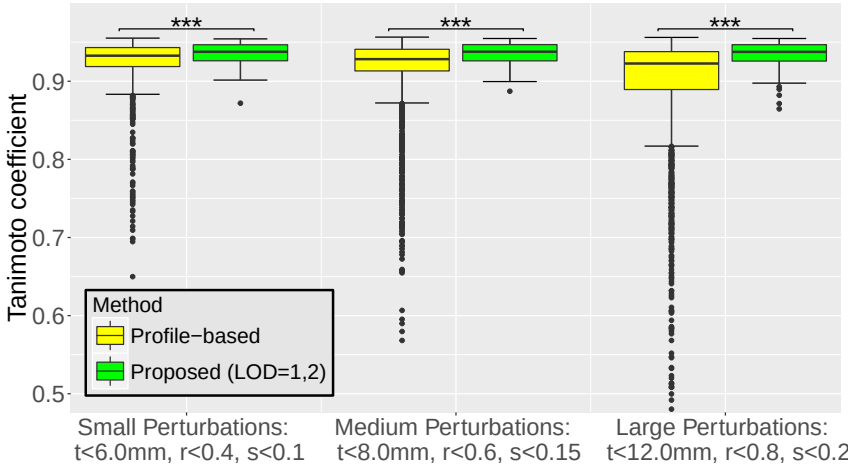


Figure 20: Overall improvement of robustness against model pose changes by the proposed method. Total spread of results for the profile-based method (yellow) and for the proposed method (green), when started without long distance model fitting from n=150 randomly degraded model initializations. To show the influence of initialization on segmentation accuracy, cases are excluded where segmentation generally failed regardless of model initialization (i.e. cases 8, 28 and 36 for the profile-based method, and cases 26 and 28 for the proposed method, cf. Fig. 19), without affecting significance of the results. ***: $p < 0.001$, as assessed with Wilcoxon's signed rank test.

By contrast, the profile-based method shows an overall strong dependence of results against pose changes, which increases strongly with larger pose changes. In every degradation level, large spread of results and occurrence of outliers were observed in most of the 45 cases. Like for the proposed method, clear outlier cases can be identified, where segmentation generally fails regardless of initialization, i.e. cases 8, 28 and 36. Finally, it can be observed that under the variation of model pose, a hypothetical optimum of segmentations for the profile-based method can be identified, i.e. the best achieved segmentations in each of the 45 columns, across all degradation levels. It is notable that the results of the proposed method are not only very precise, but they are also in many cases at least near or at the hypothetical global optimum of results of the proposed method (cf. right and left boxplot pairs in Fig. 19 at any degradation level for the separate 45 cases).

Fig. 20 shows the overall influence of model initialization on the spread of results for both methods. To exclude the influence of the aforementioned outlier cases that fail regardless of model initialization, they are not added to the plots (8, 28 and 36 for the profile-based method, 26 and 28 for the proposed method). Three pairs of yellow and green boxplots reflect the volume overlap

based on the Jaccard-Coefficient of the profile-based 3D-SSM (yellow) and of the proposed method (green). From left to right, the boxplot pairs represent the three increasing degradation levels.

These results are in accordance with the observed qualitative results from Fig. 18. The overall segmentation error from all dislocated initializations is significantly higher for the profile based method than for the proposed method. Also, the larger outliers shown in Fig. 18, which were produced by the profile-based method, are represented in Fig. 20. Errors and outliers increase for the profile-based method with growing dislocations (yellow boxplots from left to right), in contrast to the proposed method (green boxplots from left to right). The same applies for the overall precision of results, which is higher for the proposed method compared with the profile-based method, and which grows significantly only for the profile-based method through the different stages of the experiment.

Altogether, the results from the profile-based method vividly illustrate the known strong sensitivity of 3D-SSM to model pose changes and to changing model initializations. In comparison, the results of the proposed method have shown a strong increase of robustness against model pose changes during segmentation. They display a significantly higher accuracy, and a strong stability regardless of the chosen magnitude for the initial pose variations. Also, the results of the proposed method tend to be near the hypothetically observed optimum of results that were achieved by the proposed method under optimal model pose conditions.

6.2.3 Capture Range

The second experiment was conducted to evaluate the capture range of the proposed method by using it to perform segmentations from spatially distributed model initializations across the image domain. The experiment was intended to determine from which distance the proposed method was successfully attracted to the sought organ of interest during the course of the model fitting. The achieved capture range of the proposed method was then compared with the capture range from the profile-based 3D-SSM as detailed in the previous experiment.

Altered model initializations were generated similarly to the first experiment, but in a distributed image-wide pattern, to obtain a global coverage of accuracy measurements. Focus was put on the largest coherent area in

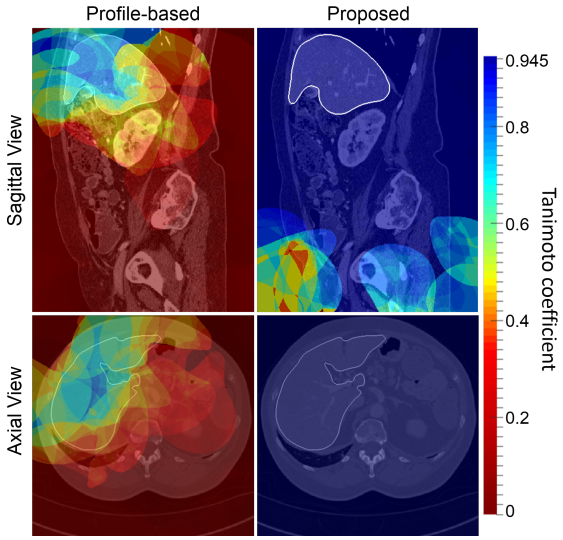


Figure 21: Quantitative comparison of model capture range around the liver. Left column: the capture range map (cf. Sect. 6.2.3) of the profile-based method shows a quick loss of accuracy (blue to red coloring) for initializations with a growing distance from the true liver surface (white contour), as measured by the Tanimoto coefficient (right hand color legend). Right column: high accuracy is achieved by the proposed method from initializations in a wide area around the liver (dark blue area).

which the proposed model fitting could be started while still sustaining a high reproducibility of results.

The experiment was conducted for the liver, for the spleen and for the right kidney. To perform segmentations of the liver, one of the trained cross validation models from the previous section was selected for which a test image was available in the DKFZ_LIVER dataset that was sufficiently large for use in the capture range experiment. In case of spleen and right kidney, trained models from the multi-organ segmentation experiments in Sect. 6.4.1 were chosen to conduct the segmentations. Here, as a sufficiently large image volume for the capture range experiment, an image was chosen from the DKFZ_LIVER dataset for which gold standard annotations of spleen and right kidney were available outside the training set of the selected models.

The chosen volumes spanned $512 \times 512 \times 394$ voxels with in-plane spacing of 0.59 mm $\times$ 0.59 mm and inter-slice distance of 1 mm in case of the liver, and $512 \times 512 \times 414$ voxels with in-plane spacing of 0.71 mm $\times$ 0.71 mm and inter-slice distance of 1 mm in case of spleen and right kidney. Both images covered the abdominal and pelvic regions. In this scenario, the capture range of the proposed long distance model fitting could be tested to a considerable

extent. Also, the applicability of the proposed method from an arbitrary model placement in the image could be demonstrated, without relying on a previous model initialization. Finally, this setting also allowed insight to the actual capture range in comparison with the chosen detection range during training.

Distributed model placements were sampled for each organ in a regular grid pattern throughout the image volume. Different numbers of initializations were involved for the organs depending on their size, i.e. less placements were required for the liver to cover the entire image due to its bigger size ($n = 150$ for the liver, $n = 250$ for the spleen, $n = 255$ for the right kidney). Also, random rotations were applied to all model initializations. Finally, segmentations were conducted from the generated model initializations by both methods.

For a performance evaluation at the distributed starting positions, a *capture range map* was generated. There for each model initialization, the Tanimoto coefficient of the subsequent segmentation was determined in comparison with the ground truth, an then mapped back to all voxels that were originally covered by the initial model placement. Tanimoto scores were accumulated for voxels where several model placements initially overlapped prior to segmentation. The accumulated scores were then averaged over the number of overlapping surfaces.

Figs. 21 - 23 show cross-sections of the capture range maps that were produced by both methods for the liver (cf. Fig. 21), for the spleen (cf. Fig. 22) and for the right kidney (cf. Fig. 23). The capture range maps of the profile-based 3D-SSM show unsteady results for all three organs from the distributed initializations (cf. Fig. 21 - 23, left columns). Furthermore, the capture range of the profile-based 3D-SSM was strictly limited to a local neighborhood around these organs. In contrast, the capture range maps of the proposed method show that steady results were obtained across a large and coherent area of the entire image (cf. Fig. 21 - 23, right columns). Thus, a large gain of capture range was reached by the proposed method over the local capture range of the profile-based 3D-SSM.

Different capture ranges of the proposed method were observed for the examined organs. For the liver, the achieved capture range roughly spanned the upper two thirds of the image volume. In accordance with the limited sampling radius chosen during training (cf. Tab. 1), the capture range of the long distance model fitting faded in the lower third of the image. There were still distant areas that exhibited sustained accuracy, due to parts of the surface model that were still within capture range as discussed in Sect. 5.3. From these positions, the subsequent model fitting lead the surface model towards

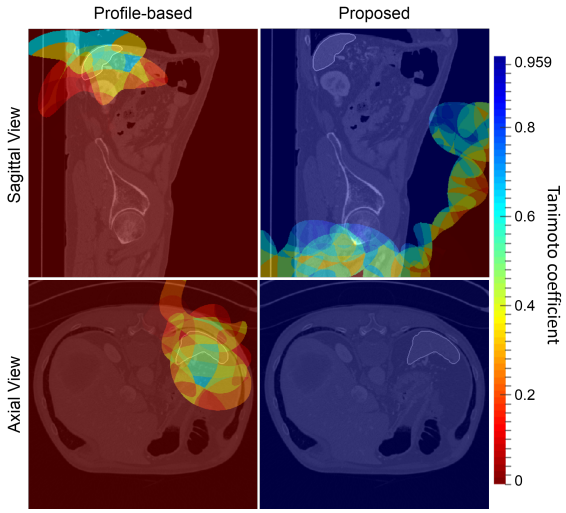


Figure 22: Quantitative comparison of model capture range around the spleen. Left column: the capture range map (cf. Sect. 6.2.3) of the profile-based method shows a quick loss of accuracy (blue to red coloring) for initializations with a growing distance from the true spleen surface (white contour), as measured by the Tanimoto coefficient (right hand color legend). Right column: high accuracy is achieved by the proposed method from initializations within a wide area around the spleen (dark blue area).

the target organ during segmentation. However, the reliability for starting positions in this part of the image was compromised, and it vanished for the corner regions. Altogether, the observed capture range clearly exceeded the range of samples provided during training.

For the spleen, a similar capture range was observed as in the liver case. Similarly for the spleen, the actual capture range clearly exceeded the range of samples provided during training. The results for the right kidney differed from the other organs, i.e. a smaller capture range was produced. In particular, the range rather matched the range of training samples that was provided during training. Still, a constant high reproducibility of results was reached when the model fitting was started from within the most part of the abdominal area, while it faded in the pelvic region of the image.

6.3 Segmentation Performance

The goal of the following experiments was to evaluate the benefit of the proposed omni-directional search on the segmentation performance of 3D-SSM. Their purpose was to address the visibility problem of 3D-SSM, where the

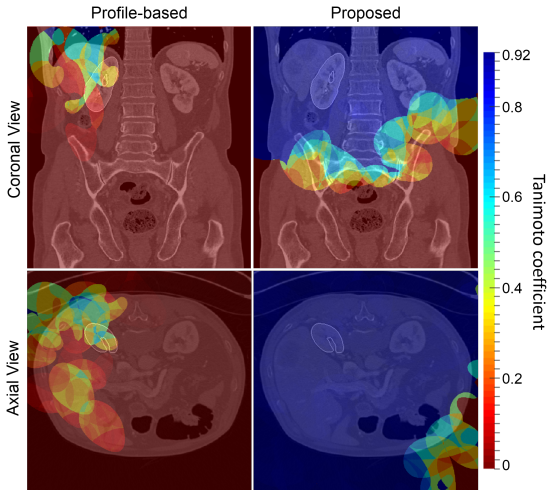


Figure 23: Quantitative comparison of model capture range around the kidney. Left column: the capture range map (cf. Sect. 6.2.3) of the profile-based method shows a quick loss of accuracy (blue to red coloring) for initializations with a growing distance from the true surface of the right kidney (white contour), as measured by the Tanimoto coefficient (right hand color legend). Right column: high accuracy is achieved by the proposed method from initializations within a wide area around the right kidney (dark blue area).

unidirectional detection of meaningful landmark positions was again strongly dependent from model pose during segmentation. As a consequence, 3D-SSM were prone to be stuck in local optima during the iterative segmentation process. The omni-directional landmark detectors of the proposed method were intended to perform a more global search for landmark positions from changing model pose, and thus to better avoid local optima during segmentation.

To assess this effect, liver segmentation experiments were conducted on the DKFZ_LIVER dataset. Results of the proposed method were compared with the profile-based 3D-SSM from the previous experiments (cf. Sect. 6.2). The chosen reference 3D-SSM and its main characteristics as a typical profile-based 3D-SSM will be detailed in Sect. 6.3.2.1. The experiments and their findings will be detailed in Sect. 6.3.2.

Finally, the resulting performance of the proposed method was compared with current liver segmentation approaches in a public segmentation challenge. Achieved results and a description of the competing methods of the challenge will be presented in Sect. 6.3.3.

Table 4: Parameter setting for the profile-based reference method. In addition, a model initialization method with an optimized parameter setting is used prior to the segmentation with the shown model-fitting parameters. Cf. Sect. 6.3.2.1 and [147], [157] for further details.

Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
4	$d_{\max} < 4.0$mm	-	-	6	10
3	$d_{\max} < 2.0$mm	-	-	6	10
3	$d_{\max} < 0.3$mm	0.01	1	6	10
2	$d_{\max} < 0.4$mm	0.02	2	6	10
2	$d_{\max} < 1.0$mm	0.05	2	6	10
2	$I = 50$	0.10	2	6	10
1	$I = 50$	0.10	2	6	10
0	$I = 20$	0.10	2	6	10

6.3.1 Data Material

For the first part of liver segmentation experiments, the DKFZ_LIVER dataset was used as described in Sect. 6.2.1. The dataset comprises 45 CT volumes (DKFZ_LIVER) of the human liver, with in-plane resolutions ranging from 0.55 mm to 0.80 mm and with an interslice distance of 3.0 mm and of 5.0 mm in some cases. Contrast agent was applied, and images were acquired at the central venous phase.

For the participation of the proposed method in the MICCAI 2007 liver segmentation challenge *'Sliver'*, the publicly available challenge dataset was used (MICCAI_LIVER) (cf. [186] for details). The dataset comprises 30 contrast-enhanced, abdominal CT images, which are also part of the DKFZ_ABDOM dataset as described in Sect. 6.2.1. 10 images were chosen by the challenge organizers for testing and for evaluation.

6.3.2 Non-Local Search Capability

6.3.2.1 Reference Method

As a reference method for comparison with the proposed method, a profile-based 3D-SSM was chosen [147], [188] of a type that is often encountered in many image segmentation applications. In contrast to the proposed non-local landmark appearance modeling, to the employed random forest regressors, and to the omni-directional search during segmentation, the chosen reference method solely relies on a local boundary appearance modeling, on a kNN

classifier training and on a unidirectional search during segmentation. Thus with regard to the common visibility problem of previous 3D-SSM, the effect of the proposed omni-directional landmark detectors can directly be assessed from the comparison.

Apart from this, the compared method does not introduce compensatory measures to take influence on search and on model fitting at different surface parts. However, an anisotropic diffusion filtering step was applied to the underlying CT images, which has a homogenizing effect on the sampled line profiles. The filtering was not applied to the input data of the proposed method.

The chosen reference approach includes an automatic model initialization step, based on an evolutionary optimization of model shape and pose parameters. Parameters are vector-encoded, and randomly perturbed during evolution based on a Gaussian distribution. Solutions are selected based on their fitness value which is defined on landmark-wise posterior probabilities $p_i(b|g_i)$ as determined by the previously trained kNN-classifier. An optimal set of parameters for the evolutionary optimizer was determined experimentally in the previous study of [157], comprising an optimization of 10 shape parameters, a population size of 1000, an evolutionary optimization throughout 25 iterations with convergence speed of $\nu = 10$, an initial Gaussian perturbation factors of $\sigma_0 = 0.6$ which was linearly decreased to a final perturbation factor of $\sigma = 0.905$, and a scaling factor for the translation perturbation of $\sigma_{\text{trans}} = 10$ (cf. [147], [157] for further details).

During segmentation, the reference method employs the same profile-based model adaptation like the final model fitting step of the proposed method, where subsequently to the previous omni-directional search of the sparse model, a linear search for all surface landmarks is performed using estimates from the previously trained kNN-classifier. This also includes the surface mesh deformation approach that is based on the proposed physical model and on the graph-cut based candidate selection as described in [157].

Tab. 5, Sect. 6.1.2, and Tab. 4 show the parameter settings that were used for a segmentation with the proposed method and with the chosen reference method. Parameters were determined empirically to obtain best results for both methods, which was done for the proposed method as part of this thesis, and for the reference method in the earlier studies of [147], [157].

6.3.2.2 Experimental Setup

To evaluate how both methods cope with the presence of local optima during segmentation, three cross validation experiments were performed on the

Table 5: Parameter setting for liver segmentation on the DKFZ_-LIVER dataset with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	$d_{\max} < 1.0$mm	-	-	22.0	1
1	$d_{\max} < 1.0$mm	8.0	-	15.0	15
2	$d_{\max} < 1.0$mm	-	-	15.0	39
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
3	I=50	-	-	6	34
2	I=50	-	-	6	34
1	I=50	-	-	12	34
2	I=50	0.1	1	6	34
2	I=50	0.1	1	6	15
1	I=50	0.1	1	6	15
0	I=50	0.1	1	6	15

DKFZ_LIVER dataset, where 9 sets of 5 images were excluded from training each time for subsequent segmentation testing. First, both methods were started from the automatic model initialization as provided by the profile-based method. For the proposed method in this case, the long distance model fitting was omitted, and segmentation was run from the first detailed stage (LOD=1) of the multi-resolution approach. This allowed a comparison of how both methods cope with suboptimal solutions under the same initial conditions for the subsequent iterative search, while differences between the automatic model initialization of the profile-based method and the detection outcome of the proposed long distance model fitting were removed.

In a second setting, the proposed method was started like in the previous setting, but from a 0.5 downscaled pose of the provided model initialization. The rationale behind this was that a large capture range was observed for the proposed model fitting in LOD=1, when started from rough placements inside the organ of interest. This trait represents an additional degree of freedom of the proposed model fitting which can be utilized during segmentation. The downscaling then acts as a step from the provided model initialization towards a more neutral initial placement.

In the final setting, the proposed method was applied throughout all LODs, including the initial long distance model fitting. The latter then provides intrinsic starting conditions for the subsequent higher LODs, as intended by the employed multi-scale approach. Altogether, the three experiments allow an

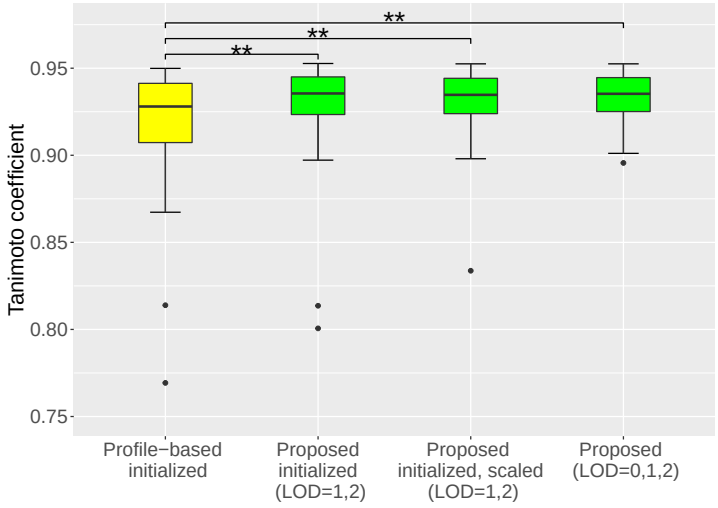


Figure 24: Improvement of accuracy by the proposed omni-directional search. In a 9-fold cross-validation on 45 CT volumes, significantly better solutions were found for the liver segmentation task by the proposed method (green) in comparison with the profile-based 3D-SSM (yellow), when started without long distance model fitting from the same initialization (Proposed initialized), after initial down-scaling by factor 0.5 (Proposed initialized, scaled), and from the outcome of the previous long distance model fitting (Proposed). $** : p < 0.01$ as assessed with Wilcoxon's signed rank test.

assessment of how the iterative omni-directional search in the higher levels of detail (LOD=1,2) of the proposed method copes with local optima during the presented liver segmentation: once under the same conditions as the profile-based method, with and without utilizing the intrinsic scaling step; and once using its own long distance model fitting to provide suitable starting conditions.

6.3.2.3 Results

Fig. 24 shows the achieved Tanimoto coefficients of the profile-based method and of the proposed method in all three settings. Overall in each setting, solutions were found by the proposed omni-directional search that were significantly closer to the global optimum in comparison with the solutions from the profile-based method. Apart from one respectively two outliers produced by both methods, the proposed method found better solutions (with Tanimoto coefficients ≥ 0.92) for the worst cases of the profile-based method (with Tanimoto coefficients of < 0.89). This roughly accounted for the lower quartile of cases from the profile-based method (10 cases out of 45).

The same benefits were reached regardless of the starting conditions of the

proposed omni-directional search, and segmentation accuracy of the proposed method remained highly similar under the differing conditions, except for the occurrence of outliers. Regarding the latter, an improvement was achieved by the applied scaling step (one outlier with a Tanimoto coefficient of ~ 0.83) in comparison with the segmentations that were started from the provided model initialization (two outliers with Tanimoto coefficients of < 0.82), and more so by using the long distance model fitting of the proposed method (one outlier with a Tanimoto coefficient of ~ 0.89). All improvements were significant with a p-value < 0.01 as assessed with Wilcoxon's signed rank test.

Fig. 25 shows cross section examples of the segmentations as produced by the profile-based method (yellow) and by the proposed method (green), in comparison with the ground truth (red contour). The upper row represents cases where a good solution was found by both methods. The middle row represents cases where solutions found by the proposed method were closer to the global optimum than the solutions produced by the compared method. This example illustrates a typical case where the profile-based 3D-SSM gets caught in a local optimum during the iterative model fitting. Part of the surface model gets attracted to erroneous landmark positions nearby, whereas more meaningful positions remain out of scope. The bottom row represents cases where no suitable solution was found by either method (3 out of 45 cases). In contrast to the aforementioned cases, the depicted segmentation error is caused by a very low contrast between the liver and the adjacent stomach, which makes a correct delineation highly challenging.

6.3.3 Performance Comparison

In the second experiment, the performance of the proposed method was compared with other approaches in the public liver segmentation challenge SLIVER. The goal of this comparison was to put the performance of the proposed method into perspective with a variety of state-of-the-art liver segmentation methods, and to compare their underlying system designs. Thus, better insight can be gained on how optimal results are commonly achieved in specific segmentation scenarios like the presented one, and how the compared algorithms can be assessed with regard to their generalizability to other tasks.

In the SLIVER challenge, more or less than 50 contributions have been made for the automatic liver segmentation task since 2007. For the performance analysis of this section, approaches were selected that performed best at the time when the proposed method participated in the challenge (as of

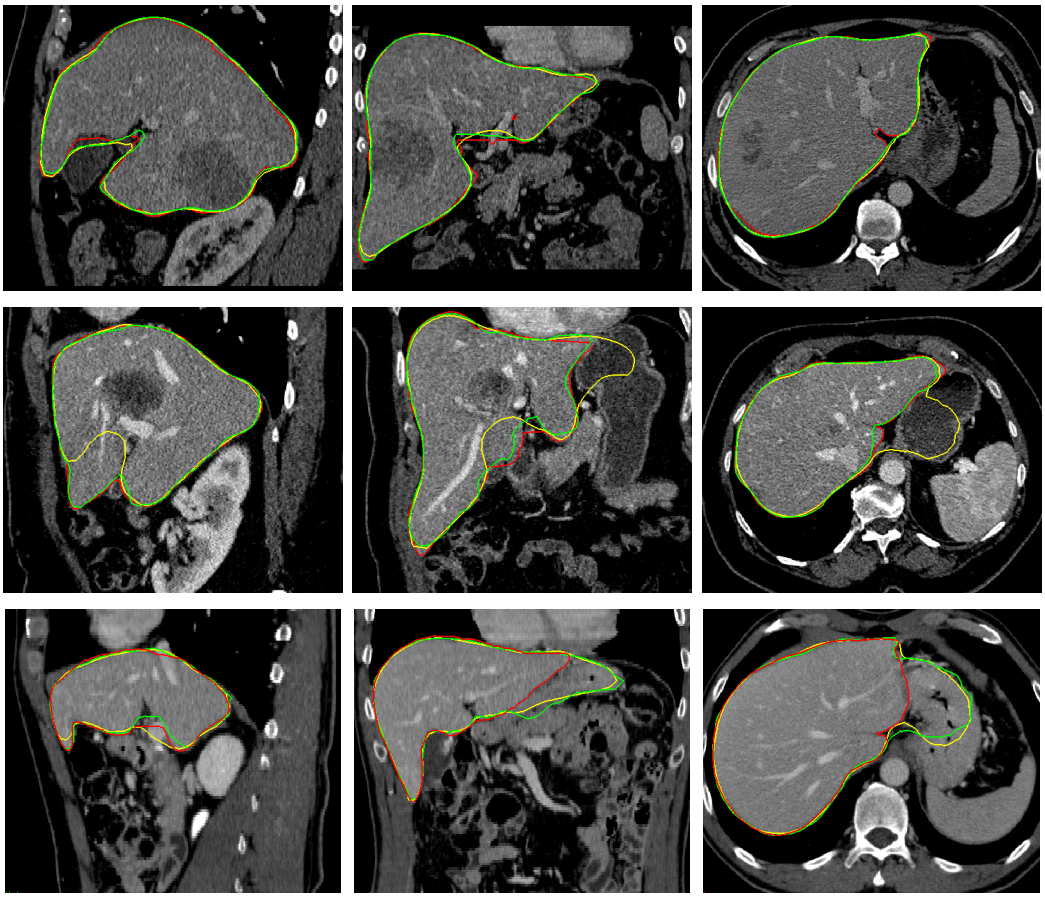


Figure 25: Qualitative comparison of CT liver segmentations. Results of the proposed method (green contour) and of the compared profile-based method [147] (yellow contour), in comparison with the ground truth (red contour). From left to right: sagittal, coronal and axial slices of results. Top row: an example where both methods achieved good results. Middle row: an example for those cases where a better solution was found using the omni-directional search of the proposed method. Bottom row: a difficult case where no suitable solution was found by neither method (3 out of 45 cases). All cases are shown with an intensity window of [-65, 385] HU.

Table 6: Parameter setting for liver segmentation on the MICCAI_LIVER dataset with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	$d_{\max} < 1.0$mm	-	-	15.0	1
1	$d_{\max} < 1.0$mm	8.0	-	15.0	15
2	$d_{\max} < 1.0$mm	8.0	-	15.0	61
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
3	I=50	-	-	6	39
2	I=50	-	-	6	39
1	I=50	-	-	12	39
2	I=50	0.1	1	6	39
1	I=50	0.1	1	12	39
0	I=50	0.1	1	12	39

May 2016). To reach high performance in the presented liver segmentation task, some of the methods rely on single segmentation techniques, whereas others involve complex hybrid systems. The majority of all methods are algorithmically designed for the liver segmentation task at some point, and only few methods are expected to be applicable to other tasks just by changing the segmentation parameters that had been optimized for the SLIVER challenge.

For the challenge participation, the segmentation parameters of the proposed method were determined empirically for best performance on the liver segmentation task of the challenge. Tab. 6 shows the resulting parameter setting for the proposed method. For model training, a subset of the DKFZ_LIVER dataset was used where the test cases of the MICCAI_LIVER dataset had been excluded (cf. Sect. 6.3.1). In addition, 27 CT liver annotations were added from another set of abdominal CT images (UHP_LIVER), which were acquired during clinical routine at the University Hospital of Heidelberg, resulting in a total of 62 liver cases. Annotations on the second dataset were produced by a medical student in accordance with the annotations of the DKFZ_LIVER dataset.

Tab. 7 and 8 show the segmentation results and the achieved challenge scores of the proposed method and of the competitors on the test dataset. Tab. 7 lists the mean volume- and surface-based quality measures from Sect. 6.1.1.2 for all competitors. For each of these volume- and surface-based evaluation measures, Tab. 8 presents a corresponding score from the scoring system

Table 7: Quantitative results from the MICCAI liver segmentation challenge *SLIVER*. Volumetric overlap error *VOE* (based on the Tanimoto metric), volumetric difference *VD*, average surface distance *AvgD*, root mean square distance *RMSD*, and maximum surface distance *MaxD*. Methods are sorted by the achieved total score in the challenge (cf. Tab. 8). The best averaged values in each category are highlighted.

Segmentation method	VOE [%]	VD [%]	AvgD [mm]	RMSD [mm]	MaxD [mm]
Al-Shaikhli et al. [143]	6.44	1.53	0.95	**1.58**	**15.92**
Li et al. [142]	6.07	**0.00**	0.97	1.82	17.49
Lu et al. [189]	**5.90**	2.70	**0.91**	1.88	18.94
Proposed	**5.90**	1.17	0.98	2.08	21.63
Kainmueller et al. [110]	6.09	-2.86	0.95	1.87	18.69
Wimmer et al. [129]	6.47	1.04	1.02	2.00	18.32
Linguraru et al. [190]	6.37	2.26	1.00	1.92	20.75
Wang et al. [191]	7.57	-1.83	1.23	2.08	19.07
Wu et al. [145]	7.87	1.31	1.29	2.50	23.56
Erdt et al. [144]	8.62	1.32	1.54	3.13	25.90
Heimann et al. [147]	7.73	1.66	1.39	3.25	30.07

of the SLIVER challenge. The score takes the inter-observer variability of a second human rater into account (cf. Sect. 6.1.1.2). That is, for each quantitative performance measure, 100 points are assigned if the measure matches those of the first expert rater, 75 points if the discrepancies to the first rater are comparable with those from the less-experienced secondary rater, 50 points for double discrepancies than from the secondary rater etc. The highest scores for each performance measure are highlighted in the table. Note that in case of the volumetric distance measure, a total value closer to zero does not necessarily imply a higher score in the scoring system, if for example larger positive and negative discrepancies level themselves out. Thus, it is more desirable to achieve a high accuracy of the liver volume estimate for every patient, rather than estimating a closer mean for all patients in the dataset.

The first five methods in Tab. 8 achieved total scores of more than 77 points. Their performances differ across the examined performance measures. The overall best volume overlaps and volumetric distances are achieved by the proposed method. Also best in volume overlap, in conjunction with the best

Table 8: Inter-rater scores for the participants of the SLIVER segmentation challenge. 75 points for results that reach the quality of a secondary human rater based on the ground truth of a primary human rater. The best scores are highlighted in each category. Note that the highest score for the volumetric distance (VD).

Segmentation method	VOE-Scr	VD-Scr	AvgD-Scr	RMSD-Scr	MaxD-Scr	Total Score
Al-Shaikhli et al. [143]	74.9	89.7	76.2	**78.1**	**79.1**	**79.6**
Li et al. [142]	76.3	88.0	75.8	78.0	77.0	77.8
Lu et al. [189]	**77.0**	85.6	**77.3**	73.8	75.1	77.8
Proposed	**77.0**	**91.6**	75.6	71.1	71.5	77.4
Kainmueller et al. [110]	76.2	84.7	76.3	74.0	75.4	77.3
Wimmer et al. [129]	74.7	86.4	74.5	72.3	75.9	76.8
Linguraru et al. [190]	75.1	85.0	74.9	73.4	72.7	76.2
Wang et al. [191]	70.4	88.7	69.2	71.1	74.9	74.9
Wu et al. [145]	69.3	85.7	67.9	65.3	69.0	71.4
Erdt et al. [144]	66.3	91.6	61.6	56.5	65.9	68.4
Heimann et al. [147]	69.8	87.9	65.2	54.9	60.4	67.6

average surface distance, favorable results are obtained by the method from [189]. Best root mean square and maximum surface distances are scored by the two methods from [142], [143], with an overall best performance achieved by the method from [143].

In the group of the automatically learning based 3D-SSM, higher performance was reached with boosted regressors as proposed in [129] and with regression forests by the proposed method, in comparison with the previous kNN classifiers [147]. The use of thorough use-case tailoring for 3D-SSM as in [110] has outperformed the more generalizable machine-learning based approaches for a long time. However, competitive results could be achieved by the proposed methodology.

Tab. 9 shows an overview of algorithmic components from competitors that rely on a 3D-SSM as their major method of segmentation. Except for their intrinsic 3D-SSM-based techniques, they refrain from additional segmentation methods to reach higher accuracy. In the listing, techniques are highlighted that are expected to cause significant extra effort in addition to the usual parameter adaptation of 3D-SSM to other use-cases, or where chances of suc-

Table 9: 3D-SSM participants in the MICCAI liver segmentation challenge *SLIVER*. Components are highlighted that are expected to cause significant extra effort in addition to the known parameter tuning of the typical 3D-SSM model fitting.

3D-SSM	Appearance Modeling		Model fitting	Model initialization
	Feat.Selection	Training		
Kainmueller et al. [110]	Manual	Manual	Profile-based	Gen.Hough Transf.
Erdt et al. [144]	Manual	Automatic	Rule-based	Heuristic
Heimann et al. [147]	Manual	Automatic	Profile-based	Evolut.Optim.
Wimmer et al. [129]	Automatic	Automatic	Level-Set	Boosted Class.+Heuristic
Proposed	Automatic	Automatic	Regression-based	—

cess are generally unclear. In detail, the 3D-SSMs from [110], [144] rely on an expert selection of image features after an analysis of tissue intensities and histograms of the underlying image data. Alternatively, an optimal set of features is identified through experimental segmentation of test images in [147]. During the appearance model training, the method from [110] increases segmentation performance by a heuristic surface-dependent modeling of liver appearance while building on expert knowledge on the specifics of the shape-model based liver segmentation into account. During model fitting, the method from [144] employs a rule-based approach, where different parts of the surface model can be governed separately by adapting a suitable set of control parameters. Thereby, the influence from unfavorable starting placements on the outcome of the model fitting can be compensated. Finally, for a suitable model initialization, some techniques rely on heuristic basic assumptions [110], [144], while in [147], an initialization technique is used where an immediate application to other organs is unclear. However in this regard, more generalizable methods for model initialization have been proposed that could be chosen to remove this limitation. Only the 3D-SSM from [129] and the proposed method are entirely based on techniques that support an easier application to different use-cases. Both methods rely on flexible machine learning and on an automatic selection of meaningful features during training for this purpose. The method from [129] employs an additional model initialization step in conjunction with minor heuristics. The initialization step should however in theory be applicable to other organs as well.

Tab. 10 shows a similar listing for hybrid segmentation systems that rely on an apt combination of different algorithms for the liver segmentation task, together with the 3D-SSM based approaches from Tab. 9. The highlighted entries indicate components that are expected to undergo significant modifications prior to an application in other segmentation scenarios. The changes are considered to require substantial expert knowledge on algorithm design, in addition to a larger number of free parameters and strong basic assumptions. All methods are ordered by their achieved total score in the SLIVER challenge. The approaches from [142], [189] employ a combination of different algorithms, which includes image preprocessing, heuristic detection of the liver, a convolutional neural network designed for the classification task [189], a 3D-SSM based shape constraint [142], and a graph cut segmentation. The method from [143] incorporates complexity through its level set formulation designed for the task, apart from a heuristic liver detection. The 3D-SSM from [110] includes heuristic tailoring of liver appearance prior to its application (cf. Tab. 9).

Altogether in addition to its independence from model initialization and its intended easier applicability, the proposed method has also shown the potential to reach a high segmentation accuracy in specific use-cases like the presented liver segmentation task, by means of the search capabilities of the employed omni-directional landmark detectors.

6.4 Generalizability

6.4.1 Multi-Organ Segmentation

The generalizability of the proposed method was first evaluated in a multi-organ segmentation task. The segmentation of several different organs in 3D images is highly relevant for many clinical applications. Important applications include target and risk structure annotation for radiotherapy planning [192]–[194], assessment of patient geometry as a boundary condition for path planning in robotic surgery [195], [196], anatomical subdivision for the detection of anomalies like tumors [197], and support for higher level feature extraction in radiomics [198].

Different methodologies for the task of multi-organ segmentation have been proposed in the literature. They comprise multi-atlas segmentation techniques

Table 10: Overview of the examined challenge participants in the MICCAI liver segmentation challenge *SLIVER*. Components are highlighted that are expected to require significant changes to the underlying algorithmic components in order to be applicable to other use-cases.

Method	Shape constraint	Graphcut	Voxel Classifier/ Image Partitioning	Preprocessing	Organ detection
Al-Shaikhli et al. [143]	Level-Set	—	—	—	Heuristic
Li et al. [142]	SSM	×	—	×	Heuristic
Lu et al. [189]	—	×	CNN	×	—
Proposed	SSM	—	—	—	—
Kainmueller et al. [110]	SSM	—	—	—	Gen. Hough Transform
Wimmer et al. [129]	SSM/Level-Set	—	Boosted Classifiers	—	Boosted Class. +Heuristic
Linguraru et al. [190]	Level-Set/Act.Cont.	(×)	—	×	Norm. Mutual Inform./ Prob. Atlas
Wang et al. [191]	Level-Set	—	Prob. Atlas - MAP	×	Norm. Cross Corr./ Prob. Atlas
Wu et al. [145]	—	×	Supervoxels (SLIC)	×	Heuristic
Erdt et al. [144]	SSM	—	—	—	Heuristic
Heimann et al. [147]	SSM	—	—	—	Evolut. Optimization

[199], deep learning applications [194], and shape priors in conjunction with a previous organ detection and model initialization. The latter is often provided by atlas-registration techniques or learning-based organ detection methods [149], [154]. These shape-prior based approaches are typical examples for the use of combined high-level techniques as a basis for the subsequent application of 3D-SSM during segmentation. They follow the rationale of administering optimal conditions for 3D-SSM which are generally highly sensitive to the provided conditions. Particularly in [154], the employed 3D-SSM were mainly utilized to support the overall segmentation by the built-in shape prior, and a graph-cut method was subsequently used to perform the final surface delineation.

The application of the proposed method in the multi-organ segmentation task involves a different approach. Instead of a complex support from methods that are supposed to provide optimal conditions for the previous susceptible 3D-SSM, the proposed method builds on its more robust search and model fitting capabilities to autonomously perform the segmentation task. This makes an application of the proposed method independent from the previous complex initialization methods. However, an application for multi-organ segmentation

also rises a number of challenges for the non-local appearance modeling and for the trained landmark detectors of the proposed method. They must detect organs at variable positions and with changing neighborhoods, and they must cope with morphological variabilities, with differing image sizes and heterogeneous imaging qualities. Regarding landmark detection and surface delineation accuracy, the proposed method is required to generalize across the differing organs without need for modifications of the appearance model training or of the model fitting procedure.

For the actual segmentation task, the proposed method took part in the *MICCAI 2015 Challenge "Multi-Atlas labeling beyond the cranial vault"* [200]. There, different state-of-the-art multi-atlas registration approaches competed in an abdominal organ segmentation task. The participating methods are dedicated to the presented multi-organ segmentation task and involve a variety of current registration and labeling techniques. Although the challenge specifically addressed an application of multi-atlas approaches, it was also open for participation of other methodologies like the proposed 3D-SSM. Its application among the atlas-based approaches allowed more detailed insights in the segmentation characteristics of these completely different methodologies.

6.4.1.1 Data Material

A set of 50 abdominal CT images (MICCAI_MULT) was made publicly available as part of the *MICCAI 2015 Challenge "Multi-Atlas labeling beyond the cranial vault"*. Images were acquired based on a colorectal cancer chemotherapy trial, and a retrospective ventral hernia study at the Vanderbilt University Medical Center (VUMC) (cf. [200]). Various scanners were involved for image acquisition. Varying fields of view were encountered in the dataset, ranging from $280 \times 280 \times 280$ mm^3 $-$ $500 \times 500 \times 650$ mm^3, comprising volumes of $512 \times 512 \times 85 - 512 \times 512 \times 198$ voxels across the pelvic and abdominal regions. Images comprise in-plane resolutions of $0.54 - 0.98$ mm voxel spacing and slice distances between $2.5 - 5.0$ mm.

Different pathologies and anomalies were present in the dataset, such as tumors, kidney atrophies, unexpected organ positions caused by sudden inhale during acquisition, and a previous liver resection. Also in contrast to the datasets from the previous experiments, this dataset exhibited more heterogeneous image qualities in terms of contrast, noise and inter-slice distance.

For the challenge, gold standard annotations of the abdominal organs had been created through manual labeling by trained raters under the supervision of medical experts from radiology and radiotherapy. 30 training and 20 test

images were selected from the labeled datasts such that random images from all scanner types were represented in both sets.

6.4.1.2 Challenge Outline

The presented segmentation challenge was designed with a special focus on multi-atlas approaches. These methods involve a registration of multiple annotated atlases to the target images. Segmentations are then retrieved from the atlas labels that have been mapped to the target image domain. Segmentations from typical multi-atlas based methods naturally build on a voxel-based representation, i.e. they focus on a meaningful membership of voxels to a segmentation volume rather than on a meaningful surface definition.

Two categories of segmentation tasks were offered by the challenge organizers. The *standard competition* focused on an evaluation of the labeling step after successful image registration. Registration of the training atlases to the test images was to be used as provided by the challenge organizers. In the *free competition*, the joint performance from the registration and label fusion steps was evaluated. No predefined registration had to be used, and segmentations were freely produced by all methods on the test dataset. This was therefore also the category for non-atlas based participants such as the proposed method.

The challenge ranked results from the different competitors based on volume overlap performance as measured by the Dice coefficient. Apart from the Dice coefficient for the ranking, average surface distance and Hausdorff distance (cf. Sect. 6.1.1.2) were also evaluated. The underlying manual annotations of the challenge dataset put an emphasis on a multi-atlas based segmentation. They were defined such that many of the inner details of the labeled organs were excluded from the larger organ. This concerns e.g. larger parts of the hepatic vessel tree and the renal calix. Atlas-based approaches take these details naturally into account during segmentation, since the labeled atlases that are used for target registration also contain these details. Model-based methods usually exclude inner details from their surface representation, which means that their results systematically include parts in the segmentation that are excluded in the ground truth. This systematically affects the achieved accuracy measures, which has to be considered in the presented challenge scores.

6.4.1.3 Challenge Participants

Several multi-atlas based approaches were proposed for the presented multi-organ segmentation task:

- **CLSIMPLEJLF: Corrective learning and simple joint label fusion [201].** The method had been particularly successful in two previous segmentation challenges. It won the 2012 Multi-Atlas Labeling Challenge for brain image segmentation, and it was one of the leading methods in the 2013 MICCAI Segmentation: Algorithms, Theory and Applications (SATA) challenge. It is based on a label fusion by means of a weighted voting strategy. Label-fusion approaches have previously been shown to be particularly successful in many image segmentation tasks (cf. [202]), and brain segmentation results at inter-rater reliability levels were reported using multi-atlas label fusion [201]. For the weighting of the label fusion, the use of spatial weighting patterns was proposed, which take intensity similarities between atlas and target images into account [203]. Furthermore, in the proposed corrective learning algorithm, an AdaBoost classifier was trained to discriminate between true and false labelings of the previous primary segmentation method, thus learning from its segmentation errors for later correction at runtime. For this purpose, results from the primary segmentation method in conjunction with underlying ground truth annotations were used as training data [201].

- **PATH: An extension of CLSIMPLEJLF, introducing auto context for corrective learning [204].** To take contextual information into account during the corrective learning step, an auto context classifier was proposed as introduced in [205]. It involves a cascade of trained classifiers where the output of the preceding stage is taken as an additional input image for the subsequent training stage. Thus, common segmentation errors of the trained classifiers are being corrected iteratively, by taking both the original image patterns and contextual information from the preceding classifier results into account. A significant improvement of segmentation performance had been reported by using this approach for the corrective learning as used by PATH [204]. PATH was the overall winner of the challenge's standard registration category as described above.

- **IMI_deeds_SSC: self similarity context and dense displacement sampling [206].** The method involves an extension of self similarity descriptors, which were proposed in [207] to provide a multimodal feature representation by means of self-repeating structural patterns. The descriptor is used to define a similarity metric for the atlas-registration. That is, a dense displacement field is optimized based on

the chosen similarity metric, and smoothness constraints are imposed through efficient inference on a Markov Random Field. IMI_deeds_SSC was the overall winner of the challenge's free category as described above.

- **Deepseg: Combined multi-atlas registration and convolutional neural network approach [208].** A multi-atlas labeling technique [209] was used to provide an initial region of interest within each of the encountered organs. Based on the initialization, a previously trained CNN was employed in a region growing fashion. New voxel candidates were selected in each iteration for classification by the trained CNN, based on a dilation of regions that had previously been assigned object labels by the CNN with a probability larger than 0.5. A final segmentation was retrieved after thresholding and connected component analysis.

Altogether, the described challenge participants represented a current variety of state-of-the-art multi-atlas segmentation methods, incorporating effective and successful concepts from label fusion, corrective and context learning, innovative feature description and integration of deep in the multi-atlas labeling approach. Details on the participation of the proposed method in the segmentation challenge will be presented in the following.

6.4.1.4 Experimental Setup

The proposed method was used for spleen, liver and left kidney segmentation in the presented challenge. For the spleen, a model of the proposed method was trained on the MICCAI_MULT dataset and additionally on 35 spleen annotations of the DKFZ_LIVER dataset (total of 65 training examples). For the liver, a model was trained on 28 training examples from the the MICCAI_MULT dataset, and from an additional 25 images of the DKFZ_LIVER dataset, where the vena cava had been removed from the underlying training examples by manual annotation, in order to match the requirements of the challenge for the liver segmentations (total of 53 training examples). For two training examples from the MICCAI_MULT, the employed surface model training was not immediately applicable due to particularities of the surface geometry that impeded spherical parameterization of these meshes (cf. Sect. 3). For the kidney, a model was trained on 25 training examples of the MICCAI_MULT dataset, on 15 manual annotations of the kidney from the DKFZ_LIVER dataset, and on 5 manual annotations of the UHP_LIVER dataset (cf. Sect. 6.3.3) (total of 45 training examples). 5 training examples were excluded from

Table 11: Parameter setting for spleen segmentation on the MICCAI_MULT dataset with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	$d_{\max} < 1.0$mm	-	-	8.0	1
1	$d_{\max} < 1.0$mm	8.0	0.5	8.0	1
2	$d_{\max} < 1.0$mm	7.0	-	8.0	15
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
2	I=100	-	1	6	10
1	I=50	-	1	6	10
1	I=50	0.1	2	6	10
2	I=50	0.03	2	6	10
1	I=50	0.1	2	6	10
0	I=50	0.1	2	6	10

the MICCAI_MULT dataset for the same reasons as described for the above liver training examples.

Like in the remaining experiments of this thesis, the proposed method was used without previous model initialization. During segmentation, based on the forest parameters from Tab. 11 - 13. Model fitting was conducted organ-wise with segmentation parameters shown in Tab. 1, starting with the liver from the center of the image volume, followed by the spleen and by the left kidney. Results of the proposed method were compared with the results from the other competitors based on the provided evaluation system of the challenge.

6.4.1.5 Segmentation Results

Fig. 26 shows the Dice ranking of the challenge participants in the examined organ categories of the free competition, and additionally the results of the winner in the standard competition (PATH) [211]. In detail, the proposed method achieved the best rank for the spleen segmentation, the second best rank for the liver segmentation, and the fourth rank for the left kidney segmentation. IMI_deeds [206] achieved the best rank in the liver segmentation category, and the best rank in the left kidney segmentation category. It was also the overall winner of the challenge's free competition. The winner of the standard competition, PATH [211], reached third ranks in the liver, spleen and left kidney segmentation categories. Deepseg [208] reached the second best re-

Table 12: Parameter setting for liver segmentation on the MICCAI_MULT dataset with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	$d_{\max} < 1.0$mm	-	-	15.0	1
1	$d_{\max} < 1.0$mm	8.0	-	15.0	15
2	$d_{\max} < 1.0$mm	8.0	-	15.0	15
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
3	I=75	-	1	6	61
2	I=50	-	-	6	15
2	I=50	0.04	2	12	15
1	I=50	0.05	2	12	15
0	I=50	0.05	2	12	15

Table 13: Parameter setting for left kidney segmentation on the MICCAI_MULT dataset with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	$d_{\max} < 1.0$mm	-	-	6.0	1
1	$d_{\max} < 1.0$mm	6.0	0.5	6.0	5
2	$d_{\max} < 1.0$mm	-	-	6.0	15
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
2	I=50	-	1	12	5
1	I=50	-	1	12	5
1	I=50	0.1	2	12	5
0	I=50	0.1	2	12	5

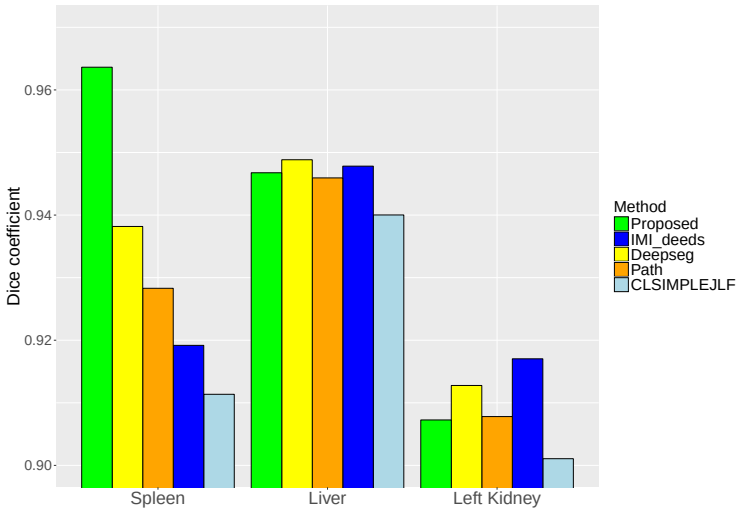


Figure 26: Results from the MICCAI 2015 Multi-atlas Labeling Challenge [200]. Dice coefficients in the *free competition* category for spleen, liver and left kidney segmentation: IMI_deedsSSC [206] (dark blue, overall winner in this category), deepseg [208] (yellow), CLSIMPLEJLF [210] (light blue) and the proposed method (green); also, the winner of the *standard registration* category: PATH [211] (orange).

sults in the spleen and left kidney segmentation categories, and has afterwards also achieved the best results in the liver segmentation category.

At a glance, the recent approaches - IMI_deeds, Path, Deepseg and the proposed method - clearly outperformed the label fusion approach CLSIMPLEJLF that had previously been employed with particular success in brain segmentation tasks (cf. Sect. 6.4.1.3). The spleen and kidney segmentation results show a clearer ranking (Dice coefficients ranging from 91.1% - 96.4% for the spleen, and 90.1% - 91.7% for the left kidney) than the liver segmentation results, where similar scores were reached by all competitors, except by CLSIMPLEJLF (Dice coefficients ranging from 94.6% - 94.9%). The achieved quantitative scores result from characteristic aspects of the segmentations as produced by the different methods, i.e. local segmentation errors, varying levels of object detail, general delineation accuracy, and handling of outlier cases.

A qualitative comparison of results from the different methods is shown in Figs. 27 - 30. They illustrate the aforementioned differences between the challenge participants for spleen segmentation (Fig. 27), liver segmentation (Fig. 28), left kidney segmentation and for the segmentation of the earlier

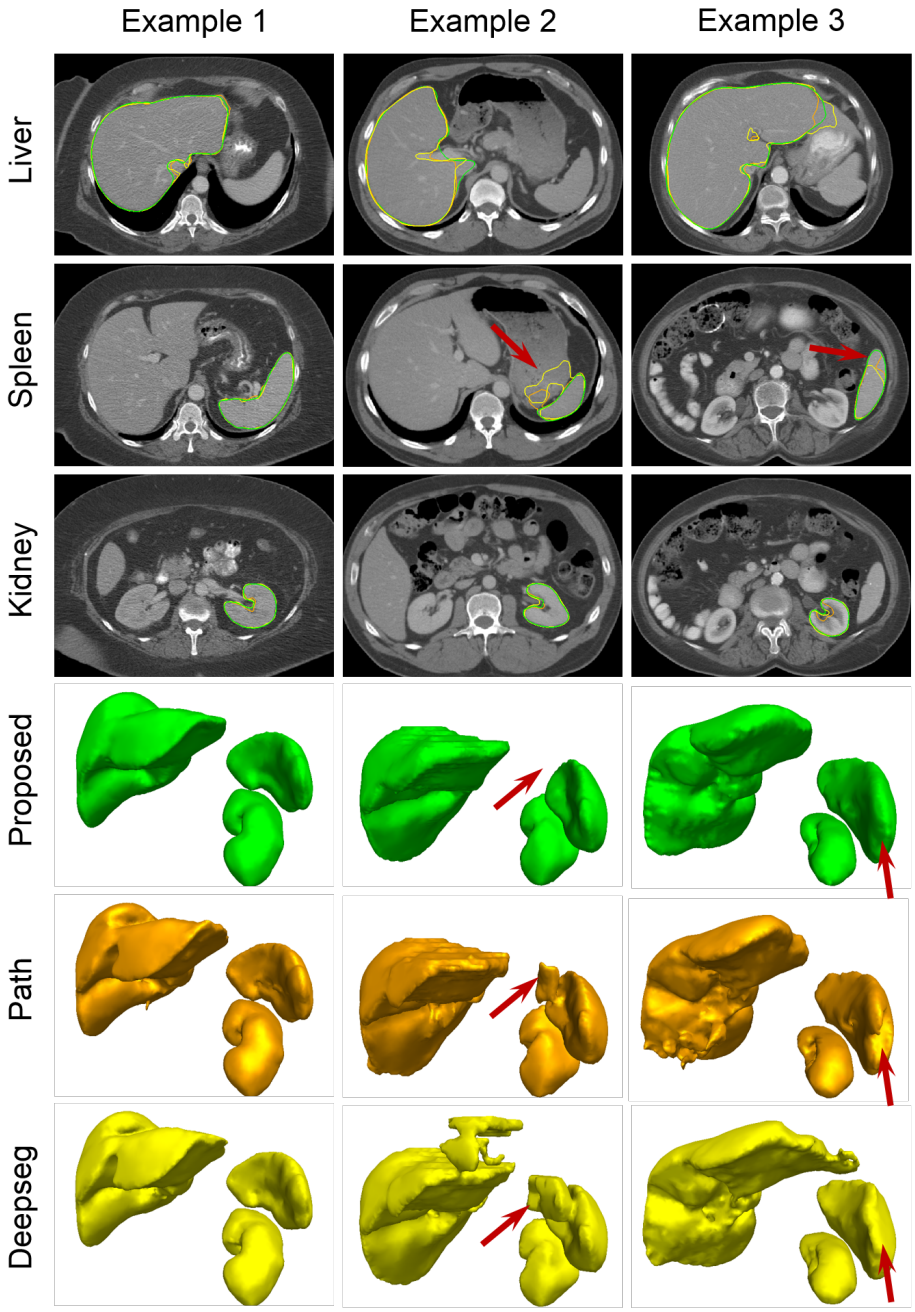


Figure 27: Differences of spleen segmentations from the competitors. Left column: good segmentation example as produced by the different methods. Middle column: More meaningful segmentation due to the explicit shape prior modeling of the proposed method. Right column: better cavity segmentation by the proposed method. All cases are shown with an intensity window of [-550, 710] HU.

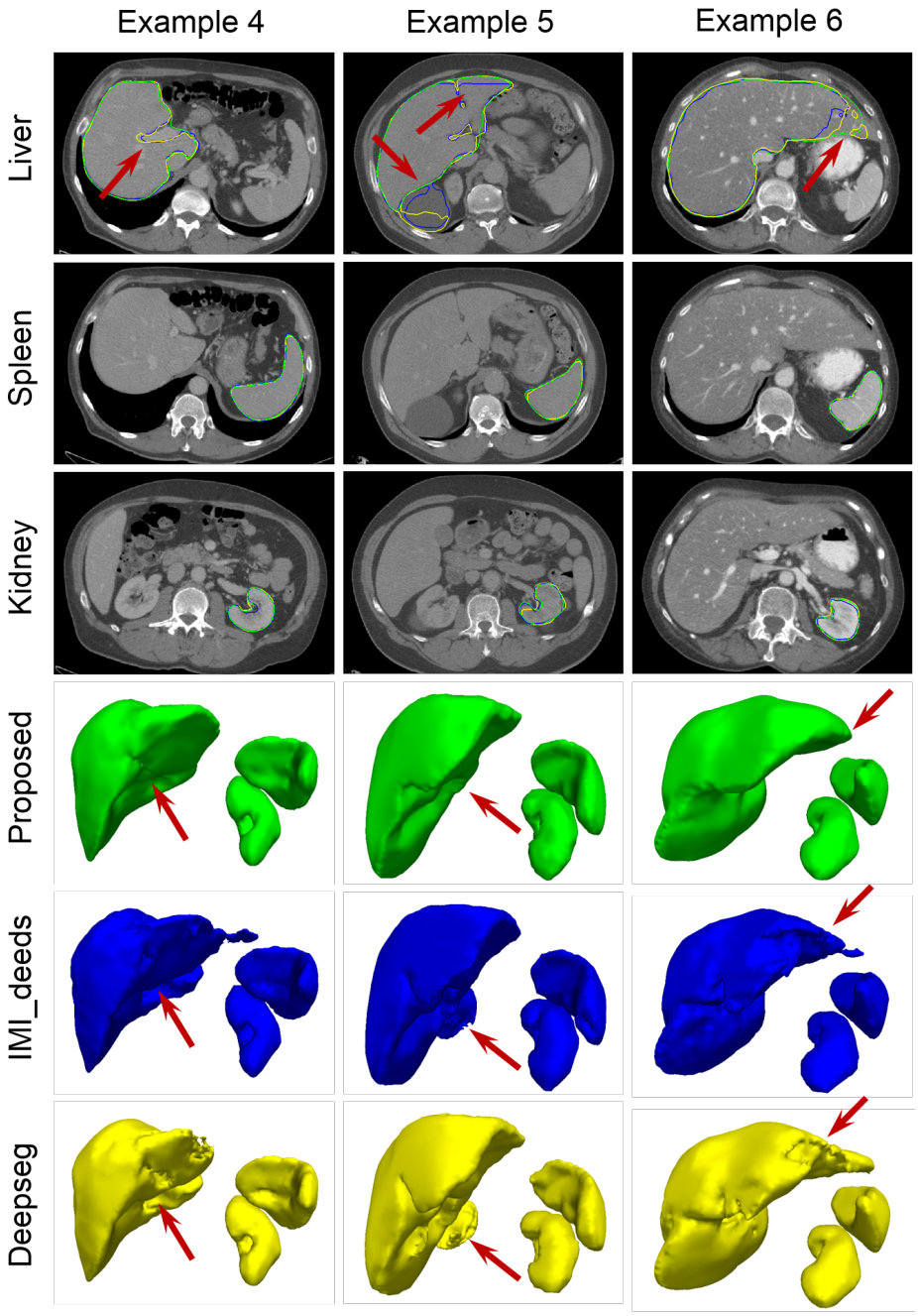


Figure 28: Differences of liver segmentations from the competitors. Left column: detailed vessel exclusion by IMI_DEEDS and Deepseg in contrast to the proposed method. Middle column: More meaningful segmentation by the proposed method thanks its explicit shape prior modeling. Right column: liver lobe sometimes missed by all methods. All cases are shown with an intensity window of [-550, 710] HU.

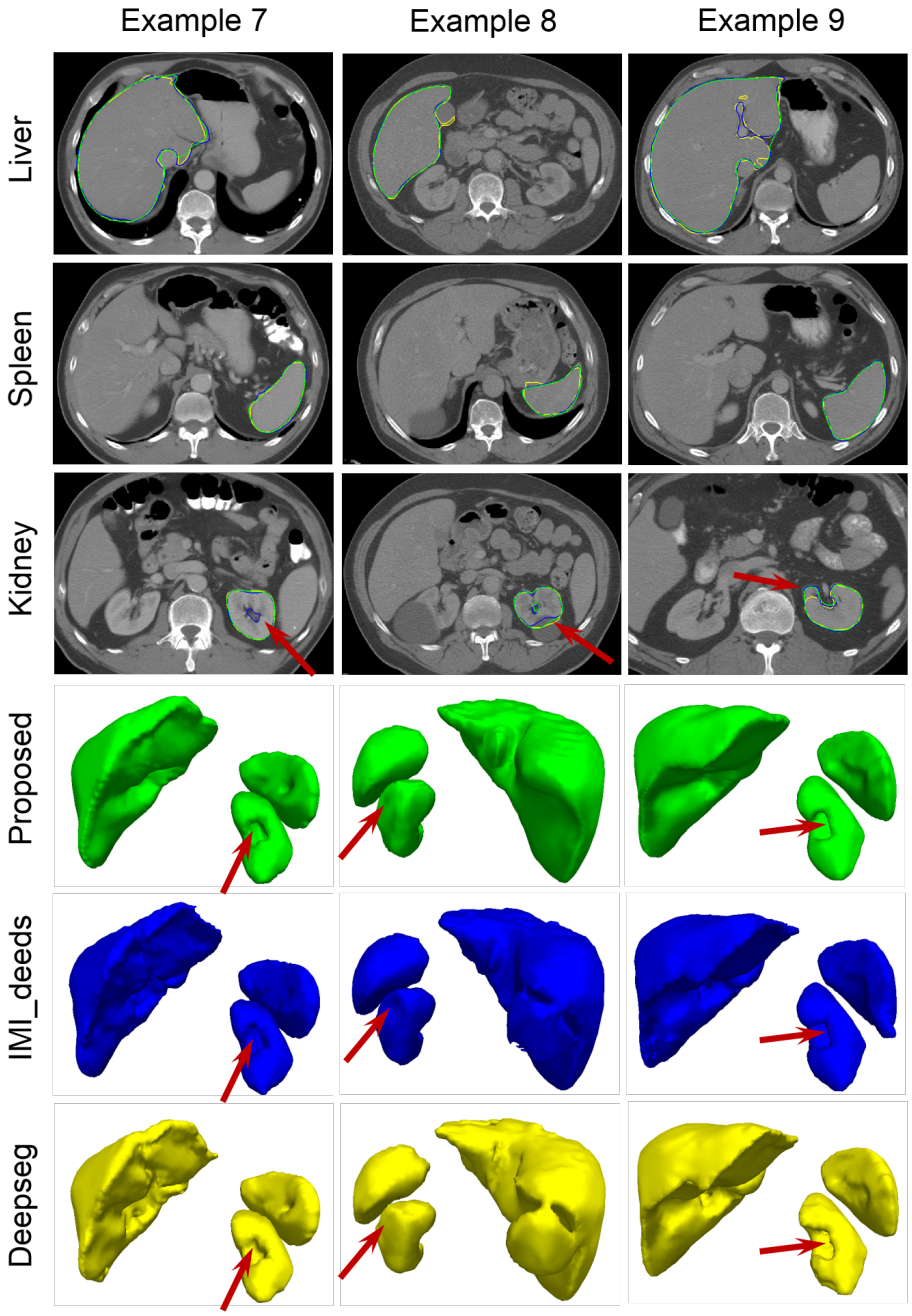


Figure 29: Differences of left kidney segmentations from the competitors. Left column: Detailed exclusion of the inner cavity from IMI_DEEDS. Middle column: Excellent tumor separation from IMI_DEEDS. Right column: Favorable segmentation of higher curvature parts by IMI_DEEDS and by the proposed method. All cases are shown with an intensity window of [-550, 710] HU.

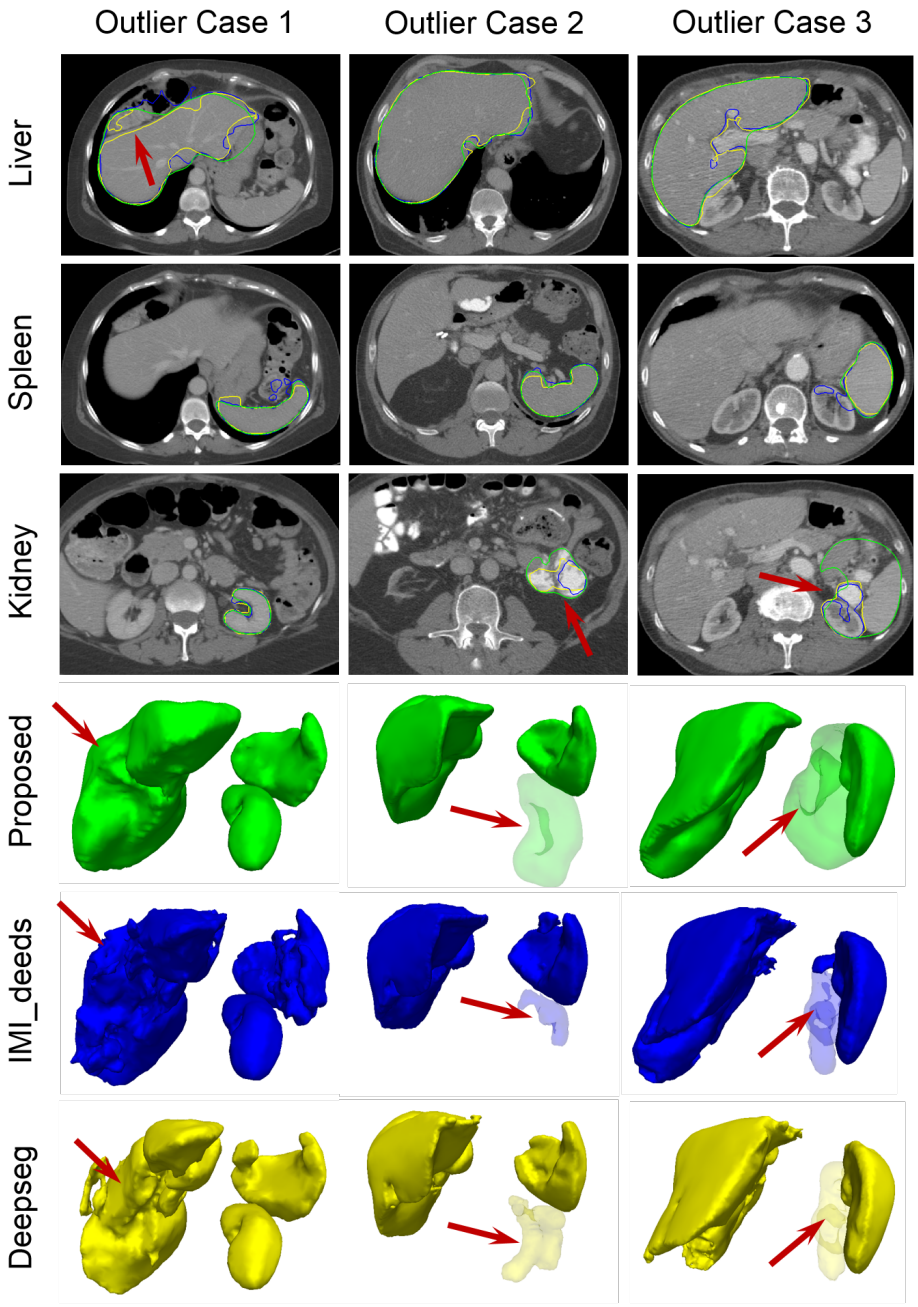


Figure 30: Difficult outlier cases in the challenge dataset. Left column: resected liver case which was probably best handled by Deepseg. Middle column: atrophic kidneys, where a valid solution is unclear. Right column: elevated kidney positioning due to influence from breathhold attempt. All cases are shown with an intensity window of [-550, 710] HU.

described outlier cases (Fig. 29). The proposed method is always compared with the (other) two best performing methods (cf. Fig. 26).

In case of the spleen, the left column of Fig. 27 shows an example for cases where a highly accurate segmentation was produced by all methods. The middle row shows a typical segmentation error of the multi-atlas approaches, where segmented regions leaked into neighboring objects, sometimes across well-defined object boundaries. These errors were also observed during liver segmentation by the multi-atlas approaches. They did not occur for segmentations by the proposed method, due to the intrinsic shape prior of the underlying 3D-SSM. The right column of Fig. 27 shows another typical spleen segmentation error that was sometimes produced by the multi-atlas approaches, where surface concavities were partly missed during segmentation. This is also a known problem of 3D-SSM, but in case of the spleen, it was mainly displayed by the multi-atlas approaches. The problem was prominent during liver segmentation as well, including the proposed method.

In case of the liver, the left column of Fig. 28 shows the aforementioned problem of 3D-SSM (cf. Sect. 6.4.1.2), where inner details of the segmented organ were missing from the trained surface model and were thus regularly included by the proposed method. In the shown case, this affects the intrahepatic vessels, which were in contrast very well detailed by Deepseg and by IMI_deeds due to their atlas-based label representation. An additional step that performed a threshold-based segmentation of the portal vein was limited in its success, leading to an increase of the mean Dice coefficient of 0.05%. The same problem also occurred during kidney segmentation. Otherwise, the case is an example for high accuracy achieved by all methods. The middle row of Fig. 28 shows typical errors from segmentation leaks into neighboring objects, as described above for the spleen segmentation. Results of the proposed method were not affected by this kind of segmentation error. The right row of Fig. 28 shows the aforementioned under-segmentation of surface concavities that was sometimes produced by all methods. In case of the liver, Deepseg was generally least affected by this type of segmentation error.

In case of the kidney, the left column of Fig. 29 shows a case where inner details of the organ were segmented differently by the challenge competitors. Parts of the renal cavity could sometimes not entirely be covered by the surface model of the proposed method. Unlike the proposed method and Deepseg, IMI_deeds produced a very large degree of detail in the kidney segmentations. The middle column of Fig. 29 shows an encountered tumor case in the challenge dataset. In contrast to Deepseg and to the proposed method, IMI_deeds

achieved an excellent separation of tumor and kidney tissue. The right column of Fig. 29 shows a typical case of surface cavity under-segmentation as produced by Deepseg, and minor delineation inaccuracies as produced by both multi-atlas approaches.

Three major outlier cases were encountered in the challenge dataset that were particularly difficult to segment by all segmentation methods. The left column of Fig. 30 shows a resected liver case. The untypical shape of the liver, and the unexpected non-liver tissue in place of the removed liver part lead to false segmentations from all participants. The middle column of Fig. 30 shows a kidney atrophy case, where the mere location of the left kidney was highly unclear. As a result, intestinal parts were segmented instead by the different methods. The right column of Fig. 30 shows a particular positioning of the kidneys right below the diaphragm, which was probably caused by a strong inhale of the patient at acquisition time. Only parts of the kidney were segmented by each method, and neighboring structures were partially included in the segmentation. IMI_deeds clearly produced the least segmentation error in this case, followed by Deepseg. The proposed method produced the largest over-segmentation of neighboring structures in the outlier case.

6.4.1.6 Summary and Conclusion

Summarized, for spleen and liver, the proposed method achieved similar or better segmentation results than the presented state-of-the-art multi-atlas approaches. In case of the kidney, the proposed method achieved similar quantitative results like the standard segmentation category winner PATH, while better scores were reached by IMI_deeds and by Deepseg. It can be assumed that for all participants, the overall scores for the liver and kidney segmentations were notably affected by the shown outlier cases (cf. Fig. 30).

Apart from surface cavities that were sometimes missed by all methods, the proposed method produced the best overall surface delineations due to its model-based shape prior, by successfully preventing the occurrence of erroneous segmentation leaks. For the same reason, the results of the proposed method were visually more meaningful than the segmentations from the other competitors. As a downside of its model-based approach, inner objects details were not well represented by the proposed method. Also, the shape prior model lead to a worse handling of the shown outlier cases, because no meaningful shape could be hypothesized from the encountered morphological irregularities, and instead, a worse overall solution was chosen in these cases.

6.4.2 MRI Liver Segmentation

In the segmentation experiment of this section, the proposed method was applied for liver segmentation in T1-weighted MRI images. The purpose of this experiment was to examine the generalizability of the proposed method to imaging modalities that comprise larger inconsistencies of the underlying intensities. In particular, the robustness of the employed Haar-like feature normalization can be assessed (cf. Sect. 4.3.2). For this purpose, MRI data segmentation is well suited: in contrast to CT data, intensities of variable scale, as well as local and global inhomogeneities are present in MRI image data. Also, pronounced soft tissue contrast is found in MRI data compared with high contrast of bone structures in CT data. Finally due to its wide area of clinical applications, the modality is of particular interest for the clinical applicability of the proposed method.

6.4.2.1 Clinical Scenario

The presented liver segmentation task was part of a clinical study (HELENA) in which the effects of an intermittent calorie restriction were examined on a number of metabolic parameters and on tissue composition of patients. 150 overweight patients were separated into groups that followed selected dietary plans, i.e. intermittent calorie restriction, continuous calorie restriction and a control diet, for the duration of one year. In the course of the study, a number of parameters were examined, comprising gene expression from subcutaneous adipose tissue biopsies, changes of metabolic, hormonal, inflammatory and metagenomic parameters, and image-derived parameters like subcutaneous, visceral and hepatic fat content (cf. [212] for further details).

The goal of the study was to compare possible benefits and the overall effectiveness of an intermittent calorie restriction diet with the other dietary models. The chosen set of genetic, metabolic and image-derived biomarkers served as a basis for a broader assessment of these effects across the different patient groups. For a retrieval of image-based parameters, MRI datasets were acquired from baseline and follow-up measurements of 150 overweight patients (50% male, 50% female). Different protocols were used to assess fat distribution and liver fat content of the patients.

Fat accumulation in the liver is expected to be strongly connected with observed overall metabolic and body composition changes in the body under dietary effects. It is directly measurable with the employed imaging protocols, and therefore as a first step to derive this measurement from the acquired

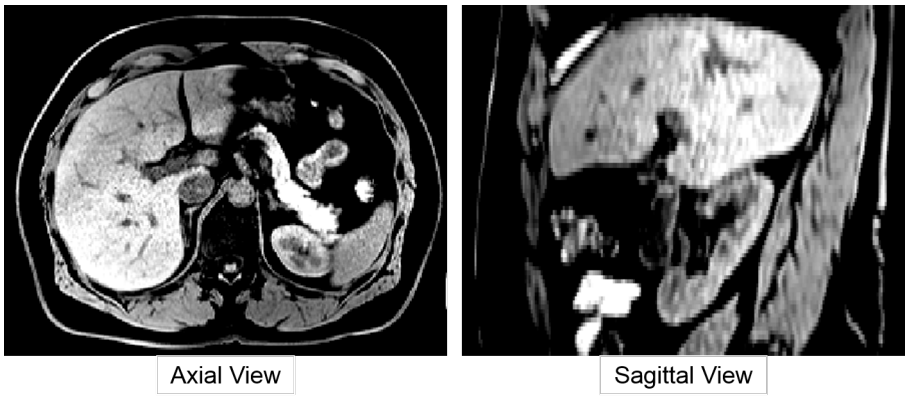


Figure 31: Typical bias field inconsistencies in the HELENA dataset. Bias field with gradient from dorsal to ventral as encountered in the dataset. A windowing function of [71,156] was chosen for visualization.

images, an automatic segmentation of the liver is required, in order to give access to the underlying information of interest.

For training and testing of the proposed method in the presented liver segmentation scenario, a subset of the data collective was selected for which manual annotations were provided. The dataset and the applied image acquisition protocol will be described in the following part.

6.4.2.2 Data Material

A dataset of thoraco-abdominal MRI volumes from 32 patients (LIVER_DIXON) was used in this experiment as part of the dataset acquired for the HELENA study. A T1-weighted 3D-VIBE two-point DIXON sequence was used for image acquisition, to allow an assessment of adipose tissue distribution in the body. Images were acquired at the Department for Diagnostic and Interventional Radiology, University Hospital Heidelberg, using a 1.5 Tesla MRI scanner with 70 cm bore design (Siemens MAGNETOM Aera). As described above, the acquisition protocol was suited for an assessment of thoracoabdominal, visceral and liver fat content. Images comprise constant in-plane resolution of 1.4×1.4 mm spacing and inter-slice distance of 3.0 mm. Challenges of the segmented dataset were pronounced bias field inconsistencies and visible liver lesions in the underlying images (cf. Fig. 31). Manual annotations of the liver were created by a medical student, via slice-by-slice segmentation.

Table 14: Parameter setting for liver MRI segmentation with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	$d_{\max} < 1.0$mm	-	-	22.0	1
1	$d_{\max} < 1.0$mm	-	-	15.0	15
2	$d_{\max} < 1.0$mm	-	-	15.0	34
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
3	I=50	-	-	6	34
2	I=50	-	-	6	34
1	I=50	-	-	12	34
2	I=50	0.1	1	6	15
1	I=50	0.1	1	6	15
0	I=50	0.1	1	6	15

6.4.2.3 Experimental Setup

An 8-fold cross validation of liver segmentation was conducted for the proposed method on the presented image dataset. Liver models were trained on the training sets using variance normalization for the employed Haar-like feature descriptor (cf. Sect. 4.3.2). Like previously, segmentation was conducted without model initialization from the image center. Similar model fitting parameters were adopted as used in the CT liver segmentation experiments in Sect. 6.3.2.2 (cf. Tab. 14).

6.4.2.4 Results

Tab. 15 shows quantitative segmentation results of the proposed method in the test datasets based on average surface distance (AvgD), maximum surface distance (MaxD), root mean square distanc (RMSD), Dice- and Jaccard coefficients. An overall high segmentation accuracy was reached by the proposed method, with a mean average surface distance of 1.27 mm, a mean Tanimoto coefficient of 0.92 and a mean Dice coefficient of 0.96. Compared with the CT liver segmentation results from Sect. 6.3, results were slightly worse, with a mean Tanimoto coefficient of 0.93, and mean Dice coefficient of 0.97 on the CT dataset.

Feature normalization was a prerequisite for a meaningful segmentation of the MRI data by the proposed method, otherwise segmentation was prone

Table 15: Quantitative results of the proposed method for liver segmentation in MRI data. Average surface distance (AvgD), maximum surface distance (MaxD), root mean square distance (RMSD), Tanimoto and Dice coefficient. Segmentations were evaluated on a set of 32 images by 8-fold cross validation.

No.	AvgD [mm]	MaxD [mm]	RMSD [mm]	Tanimoto-Coefficient	Dice-Coefficient
1	1.16	2.12	23.03	0.92	0.96
2	1.76	4.29	45.37	0.93	0.96
3	1.22	2.14	22.18	0.92	0.96
4	1.11	2.23	21.83	0.93	0.96
5	1.02	2.02	20.89	0.94	0.97
6	1.01	1.89	15.60	0.92	0.96
7	1.20	2.57	25.01	0.94	0.97
8	1.24	2.40	21.09	0.90	0.95
9	1.03	2.05	21.35	0.93	0.96
10	1.02	2.05	19.69	0.93	0.96
11	1.26	2.44	21.83	0.92	0.96
12	1.28	2.66	22.36	0.92	0.96
13	1.24	2.52	22.11	0.92	0.96
14	2.12	6.11	46.66	0.90	0.95
15	1.20	2.45	32.62	0.92	0.96
16	1.35	3.14	29.61	0.93	0.96
17	1.23	2.44	24.10	0.93	0.96
18	1.15	2.20	21.23	0.92	0.96
19	1.31	2.76	31.60	0.92	0.96
20	1.53	3.13	33.91	0.91	0.95
21	1.00	1.94	18.76	0.93	0.97
22	1.36	2.27	18.92	0.92	0.96
23	1.46	3.60	40.12	0.92	0.96
24	0.96	1.84	18.28	0.93	0.97
25	1.52	2.85	26.53	0.91	0.95
26	1.17	2.43	26.64	0.93	0.96
27	1.16	2.39	27.33	0.92	0.96
28	2.24	5.36	47.50	0.89	0.94
29	1.08	2.10	20.32	0.93	0.96
30	1.15	2.09	23.24	0.92	0.96
31	1.10	2.44	25.35	0.93	0.97
32	0.88	1.68	14.79	0.93	0.97
Avg	1.27 ± 0.30	2.64 ± 0.97	25.91 ± 8.60	0.92 ± 0.01	0.96 ± 0.01

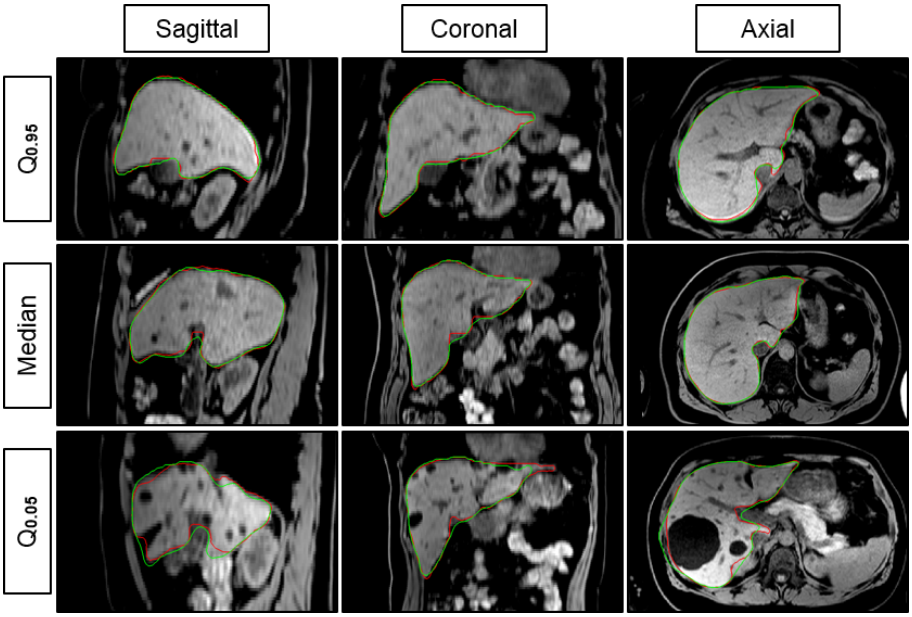


Figure 32: Qualitative evaluation of liver segmentations in T1-weighted MRI data. Exemplary results of the proposed method (green contour) in comparison with the manual ground truth (red contour), from $Q_{0.05}$ quantile (best results), median (average results) and $Q_{0.95}$ quantile of average surface distance error (failed segmentations).

to fail. In all cases, the employed feature normalization provided robustness against the encountered intensity inconsistencies from the MRI acquisition. Fig. 6.4.2 shows exemplary segmentation results of the proposed method in comparison with the underlying manual annotations. Generally, a high delineation accuracy was achieved. Segmentation errors in the more challenging cases were found at the lower liver lobe near the rib cage and intestine, or when a thin part of the liver lobe entered between heart and stomach.

A number of liver segmentation methods for MRI data were proposed in the literature. They involve the use of atlas-based approaches [213], [214], of 2D or 3D level-sets [215], [216], random walker [217], watershed [218], [219] or active contours [220]–[223], or combination thereof [224]. According to the literature, good results are not often achieved, and the use of semi-automatic approaches is regularly referred to as a solution towards higher accuracy [225].

A comparison of the different published results is difficult due to differences of image acquisition protocols, of clinical use-cases and of visible pathologies in the dataset. The methods from [214], [216] have published results on T1-weighted two-point Dixon imaging, comparable with the dataset used for experiments with the proposed method in this section. The work from [216]

builds on a level-set technique, incorporating image denoising, distance transformation, the use of probability maps and of intensity priors that are learned from training data. Healthy and fatty livers were automatically segmented. An overall mean Dice coefficient of 0.94 ± 0.02 was reached on a set of 20 non-fatty livers, and a mean Dice of 0.89 ± 0.06 on 20 fatty livers using their approach. In [214], a method based on multi-atlas segmentation and morphological operations was proposed for segmentation of image data acquired from 20 female patients, prior and after implementation of a diet plan. A mean Dice coefficient of 0.943 ± 0.023 was achieved using their approach. In comparison, the proposed method has achieved a mean Dice of 0.96 ± 0.01 on the 32 patients of the LIVER_DIXON dataset.

Other approaches have presented results for other acquisition protocols and use-case scenarios. In [219], 17 datasets of fatty and healthy livers acquired on different machines were automatically segmented using an improved watershed algorithm. A mean Dice of 95.0 ± 1.0 (mean Jaccard coefficient of 0.91 ± 0.02) was reached on the presented dataset. In [220], a geodesic active contour was used for automatic segmentation of 23 T1-MRI with visible tumors, where a mean Dice coefficient of 93.6 ± 1.7 was reported. In [222], another geodesic active contour was proposed for segmentation of T1-MRI from 10 patients, where a mean Dice coefficient of 91.0Â $\pm$ 2.8 was reported. Probably the best overall results were reported by [225], where a semi-automatic approach by means of Laplacian mesh optimization was proposed for segmentation of 20 contrast-enhanced MRI volumes acquired with LAVA-3D sequence for fat suppression. A Tanimoto coefficient of 0.92 ± 0.14, an average surface distance of 1.52 ± 0.26 mm and a maximum surface distance of 2.60 ± 0.55 mm were reported. Other approaches have either been used for 2D segmentation, or less accurate overall results were reported.

6.4.2.5 Summary and Conclusion

In this section, an application of the proposed method for liver segmentation in T1-MRI data was presented. The method was applied without algorithmic adaptations, without changes of the model training and without previous model initialization, and similar segmentation parameters were used like in the liver segmentation experiments from Sect. 6.3.2. Also, the proposed feature normalization was applied in both the CT liver and the MRI liver segmentation experiments (cf. Sect. 4.3.2).

In a comparison with results that had been reported by other works on MRI liver segmentation in the literature, the proposed method has achieved better

results than the state-of-the-art methods for images acquired by the two-point Dixon technique. Also, no better results were reported by any of the other presented approaches, which is however hard to compare due to the different imaging conditions and use-cases presented in these latter works.

The shown results have demonstrated that the proposed method also generalizes well across imaging modalities that exhibit larger inhomogeneities of their intensity distributions. Results have also shown that the proposed feature normalization was robust against the encountered bias field inconsistencies in the LIVER_HELENA dataset. Finally, the achieved accuracies in the presented cases indicate that the benefit of the proposed omni-directional landmark detectors to find solutions that are closer to the global optimum during segmentation, also apply to other more heterogeneous imaging modalities than the presented liver CT scenario in Sect. 6.3.2.

Since no conclusive breakthroughs have been reported in the field of MRI data segmentation as of yet, the shown results are promising for a development of techniques that can promote higher segmentation accuracy for this challenging modality type.

6.4.3 Left Ventricle Ultrasound Segmentation

In this section, an application of the proposed method for left cardiac ventricle segmentation in 3D+t ultrasound images will be presented. This scenario is an example for a largely different task that is particularly challenging in several regards.

In these images, strong noise, motion artifacts and partial occlusions of anatomical structures are regularly encountered. These effects are challenges that can put the learning capabilities of the regression forests to the test. Also, they cause notable variabilities of visible anatomical structures and of their appearance. This is a particular challenge for the flexibility and for the expressiveness of the employed Haar-like feature descriptor. Here, it is of interest whether the observed variabilities can be compensated by the randomness of the feature descriptor, and whether meaningful features can successfully be learned during training.

The dataset that was used in this experiment will be described in Sect. 6.4.3.2. The results that were obtained by the proposed method, and a comparison with other approaches for left ventricle segmentation in ultrasound images will be presented in Sect. 6.4.3.4.

6.4.3.1 Clinical Scenario

The presented task of left ventricle segmentation in ultrasound images originates from a clinical application in cardiac surgery. By means of a fast intra-operative acquisition of ultrasound images, a quick assessment of ventricle geometry and function becomes available during surgical intervention. This is particular important in case of time critical interventions, and it is also highly useful for real-time evaluation and monitoring of the applied surgical measures. Ultrasound imaging of the heart is also an important diagnostic tool for patient examination prior to surgery.

Based on the acquisition of time-resolved ultrasound data, ventricle motion can be monitored and be evaluated across the cardiac cycle, allowing to reveal types of pathological anomalies and their severity. In addition, the computerized image analysis at single time frames also allows a precise assessment of ventricular geometry during the cardiac cycle, thus providing valuable insight in functional parameters and potential biomechanic disorders of the heart.

In this scenario, a model-based segmentation of ventricle can be particularly beneficial, since it provides an immediate surface-based representation of the ventricle, which allows visual inspection in the operating room, thus providing an immediate qualitative assessment of ventricle geometry. Furthermore, a segmentation with 3D-SSM also allows an incorporation of higher modeling information on cardiac motion by means of a learning-based motion modeling of the heart. The use of this kind of modeling information can in turn be useful to promote robustness and accuracy for ventricle segmentation, particularly in case of challenging imaging modalities such as ultrasound imaging.

However, conflicting demands from depth information acquisition and spatial resolution, strong noise, the capturing of motion artifacts, variable view angles and a stronger influence from experience level of different ultrasound operators make an automatic segmentation of the considered cardiac ultrasound data highly challenging (cf. Fig. 33). In addition to the robustness and accuracy requirements, segmentation is desired to be fast in order to provide instant information for intra-operative use. Altogether, the presented scenario of ventricle segmentation in ultrasound data is a well-known and challenging task. Despite numerous improvements in the development of segmentation techniques, the problem remains challenging and is still subject of ongoing research.

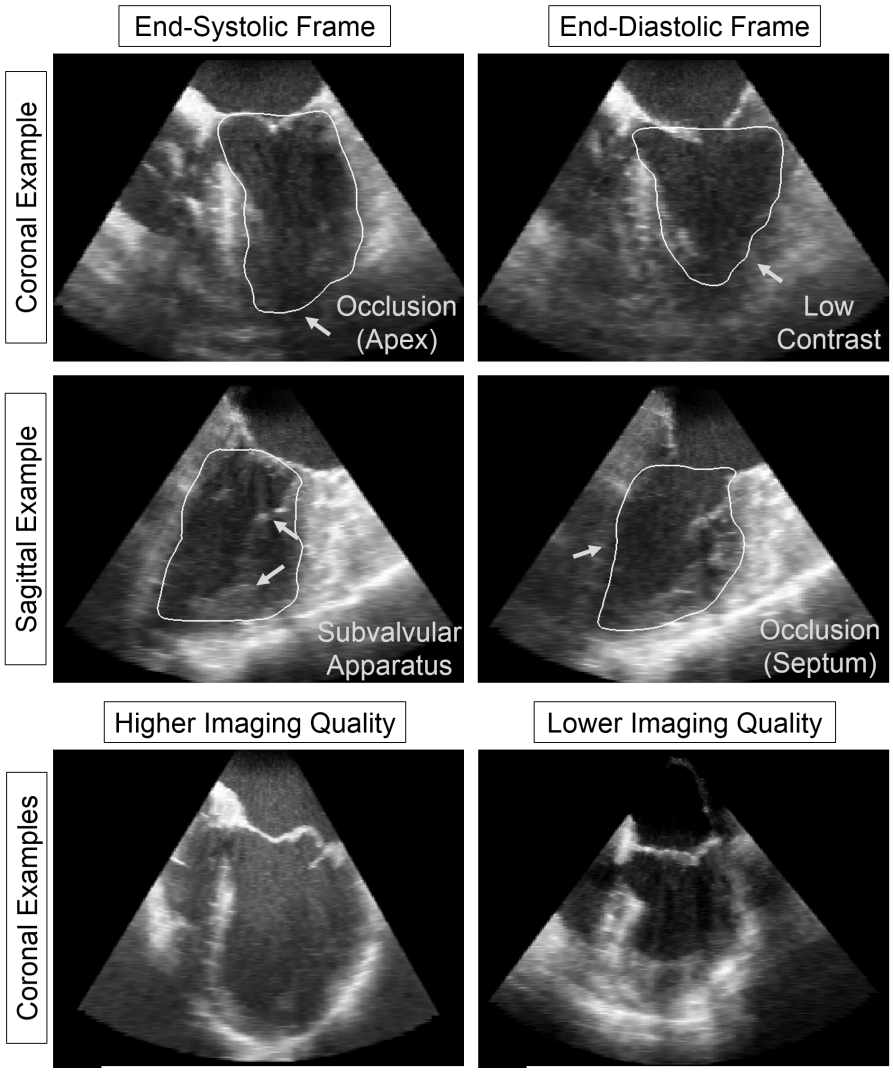


Figure 33: Challenges of the left ventricle ultrasound segmentation task. Wall occlusions, ventricle deformation, intra-ventricular substructures, low contrast, noise, artifacts and changes of imaging quality make left ventricle segmentation in ultrasound images a challenging task (ground truth shown as white contours).

Table 16: Parameter setting for left ventricle segmentation with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
1	$d_{\max} < 1.0$mm	-	-	3.0	5
2	$d_{\max} < 1.0$mm	10.0	-	3.0	5
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
2	I=50	-	1	6	5
1	I=50	-	1	6	5
0	I=50	-	1	6	5

6.4.3.2 Data Material

The image dataset used for segmentation in this section comprises 35 ECG-triggered transesophageal echocardiagrams (4D-TEE) which were acquired from 19 patients prior to cardiac surgery at the University Hospital of Heidelberg (UHP_VENTRICLE). Patients from the selected group suffered from cases of mitral valve insufficiency. Images were acquired in four chamber view of the left ventricle using ECG triggering for data extraction at end-diastolic and end-systolic time points of the cardiac cycle. A Philips IE 33 ultrasound system was used for image acquisition, equipped with a X7-2t matrix array transducer. Image resolution was isotropic with a spacing of 0.5 mm. Typical view-angle changes were encountered in the dataset, caused by ventricular motion, by operator variability and by cardiac-esophageal variation. Also, strong noise, motion artifacts and time-varying occlusions of the ventricle wall were present in the images (cf. Fig. 33 for an overview of the encountered difficulties in the UHP_VENTRICLE dataset).

6.4.3.3 Experimental Setup

A cross-validation experiment was conducted on the UHP_VENTRICLE dataset, where two patients were randomly excluded from model training for segmentation testing in each set. Models were trained on 9 sets, with 4 images used for testing in all sets, except for one set where a diastolic time frame of a patient was not available in the dataset. Training and voting parameters were applied for the proposed method like in the other experiments, as detailed in Tab. 1. Parameters for the model fitting are shown in Tab. 16. Due to the smaller size of the ultrasound images in comparison with the other image vol-

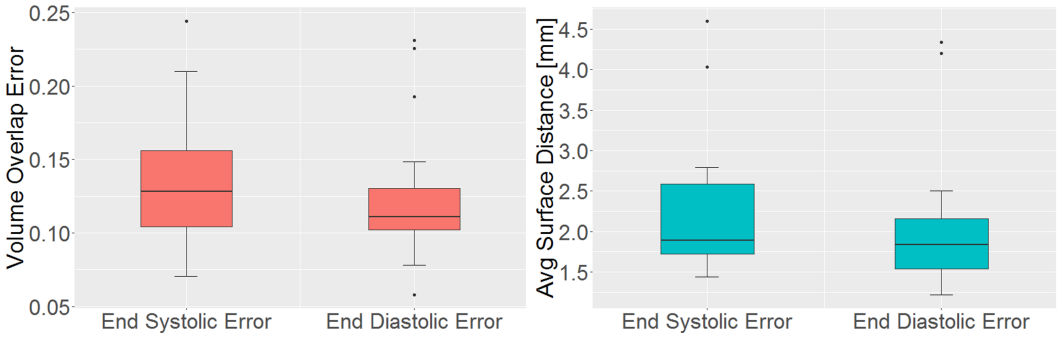


Figure 34: End-systolic and end-diastolic errors on the UHP_VENTRICLE dataset as produced by the proposed method. End-systolic (left column) and end-diastolic (right column) time frames: results of the proposed method (green contour) in comparison with the ground truth (red contour).

umes of this thesis, the trained capture range of the model fitting at LOD=1 was already covering an ample part of the encountered volumes, thus providing enough capture range to find the ventricle position in the image. Segmentations were evaluated with the quantitative performance measures from Sect. 6.1.1.2.

6.4.3.4 Results

Fig. 34 shows the quantitative results of the proposed method for the diastolic and for the systolic time frames, based on the Dice volume overlap error and on the average surface distance metrics. Mean overlap errors of 0.136 ± 0.047 (systolic) and of 0.126 ± 0.049 (diastolic), and mean average errors of 2.26 ± 0.87 mm (systolic) and of 2.08 ± 0.89 mm (diastolic) were produced by the proposed method.

Fig. 35 shows representative segmentation results produced by the proposed method at end-systolic and end-diastolic time frames. Cases comprise high accuracy in both time steps (upper left figures), smaller deviations in both time steps (upper right figures), lower accuracy in one of the time steps, more often in the end-systolic frame (lower left figures), and one failed segmentation at both time frames (lower right figures). Segmentation problems generally occurred due to shadowing at ventricle septum and apex, and due to interference from the subvalvular apparatus at the apex.

The produced results of the proposed method are also in the range of re-

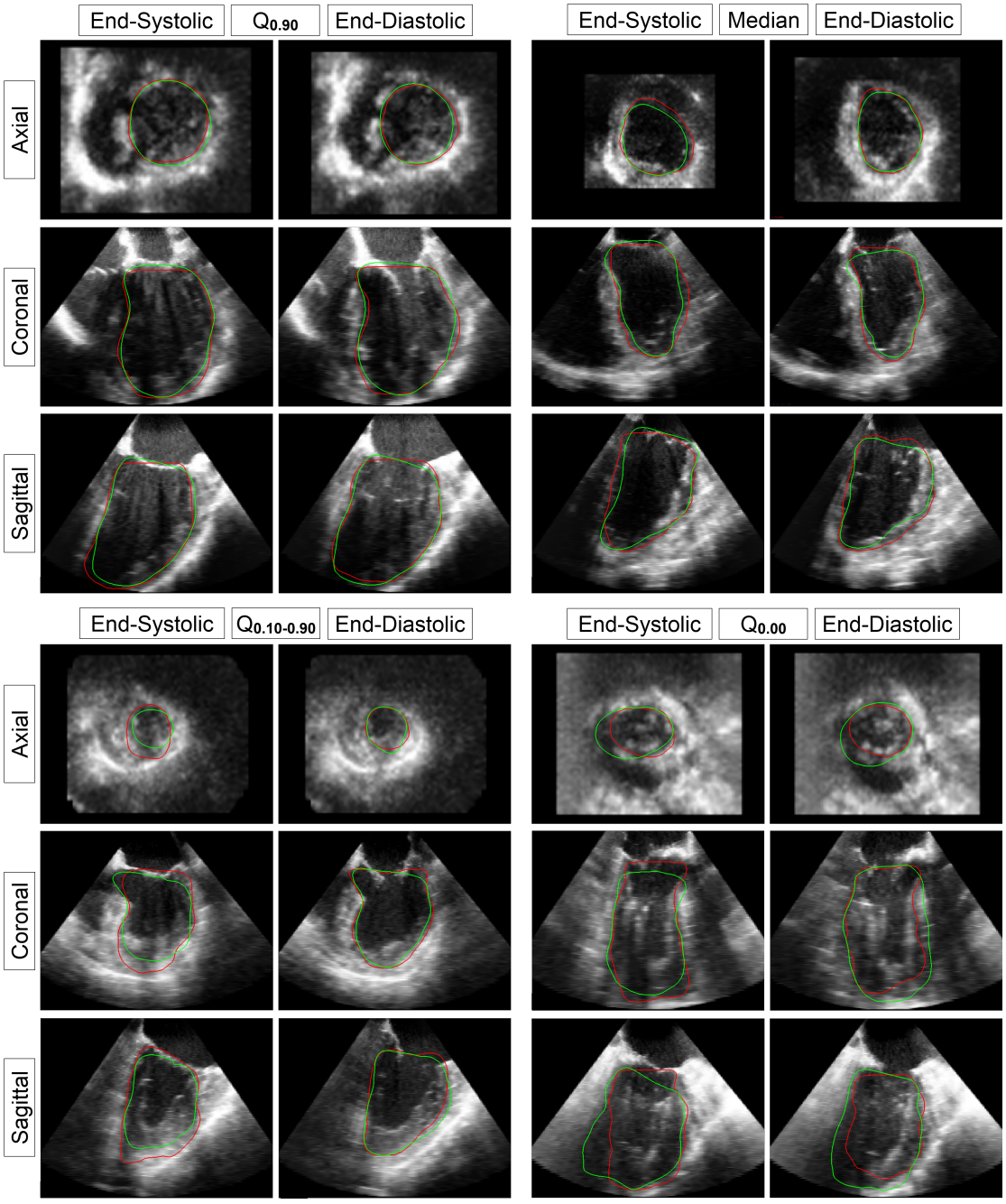


Figure 35: Segmentation examples of the left ventricle in the UHP-_VENTRICLE dataset. Results of the proposed method (green contour) in comparison with the ground truth (red contour) for the 90th percentile of all achieved Dice coefficients (best results, top left), for the median (average results, top right), for the 10th / 90th percentile at the end-systolic / end-diastolic time frames (bottom left), and for the worst result (failed segmentation, bottom right).

cently reported state-of-the-art results. Several approaches have been proposed for the fully-automatic segmentation of the left ventricle in 3D+T ultrasound data. Some of them have been evaluated on the public dataset of the MICCAI 2014 Segmentation Challenge 'CETUS' [226]. There, segmentations were evaluated on a set of ultrasound images acquired at systolic and diastolic time frames for 30 patients. Their results will shortly be outlined in the following. For calculation of their overlap errors, a modified Dice measure was proposed for the overlap comparison of surface-based segmentation representations.

In [227], a B-Spline Explicit Active Surface Approach was proposed to fit an ellipsoid to the ventricle during segmentation. A motion estimation was involved based on global optical flow and temporally consistent, recurrent block-matching. They reached mean overlap errors of 0.144 ± 0.057 (systolic) and of 0.106 ± 0.041 (diastolic), and mean average errors of 2.43 ± 0.91 mm (systolic) and of 2.26 ± 0.73 mm (diastolic). In [228], the B-Spline Explicit Active Surface approach was extended by the use of shape priors, which has produced the best results so far, i.e. overlap errors of 0.125 ± 0.046 (systolic) and of 0.091 ± 0.034 (diastolic), and mean average errors of 1.98 ± 0.66 mm (systolic) and of 1.81 ± 0.59 mm (diastolic)

In [229], a random forest auto context classifier was used, based on features of image intensities and of geodesic distances to anatomical keypoints, like ventricle center, mitral valve and myocardium. The use of the auto context implicitly induced shape information during the labeling process, that was previously learned by the classifier cascade. They reached mean overlap errors of 0.158 ± 0.057 (systolic) and of 0.130 ± 0.048 (diastolic), and mean average errors of 2.54 ± 0.75 mm (systolic) and of 2.44 ± 0.95 mm (diastolic).

In [230], a joint detection and segmentation approach based on Hough forests was proposed. Hough forest predictions on the ventricle center position were produced on voxel positions that were classified as part of the ventricle wall. After back-projection of the predictions to original wall positions, label patches stored during forest training are composed to a probabilistic map of the ventricle, from which the segmentation is obtained. They reached mean overlap errors of 0.162 ± 0.062 (systolic) and of 0.107 ± 0.031 (diastolic), and mean average errors of 2.91 ± 1.01 mm (systolic) and of 2.14 ± 0.68 mm (diastolic).

In [231], a deformable mesh was applied where global and local deformation was estimated by a Kalman Filter for the different time frames. Input for the filter was generated from profile-based edge detection of the surface model. They reached mean overlap errors of 0.156 ± 0.050 (systolic) and of $0.115 \pm$

0.038 (diastolic), and mean average errors of 2.92 ± 0.93 mm (systolic) and of 2.62 ± 0.95 mm (diastolic).

In [163], an Active Appearance Model (AAM) was used, where the surface model was initialized using a set of strategies, of which the best fit was selected based on the residuals of the AAM. They reached mean overlap errors of 0.165 ± 0.079 (systolic) and of 0.121 ± 0.054 (diastolic), and mean average errors of 2.79 ± 1.24 mm (systolic) and of 2.44 ± 0.91 mm (diastolic).

For the presented methods it can be summarized that the regression-based Hough Forest approach [230] outperformed the classification-based auto-context approach [229]. And that the use of explicit surface models in combination with motion estimation for joint segmentation of the two time frames performed best [227], particularly so with an added shape prior for more accurate segmentation [228]. Results on the UHP_VENTRICLE dataset used for the segmentations in this section do not allow a direct comparison with the above results. However, it can be concluded that the results of the proposed method on a typical clinical dataset are well in the range of the presented state-of-the-art methods for ventricle segmentation in the clinical dataset of the CETUS challenge. Methodologically, the proposed method is probably most similar to the explicit model-based segmentation approaches from [227], [228]. By contrast, they also incorporate motion estimation to profit from cross-information from the different time frames.

6.4.3.5 Summary and Conclusion

The presented ventricle segmentation task in 3D+T ultrasound data served as an additional example for the pronounced generalizability of the proposed method. In the presented case, the proposed method was applied without modifications to a set of challenging conditions that strongly deviate from the previous CT and MRI segmentation scenarios. Imaging conditions introduced large differences to the previously examined modalities, in terms of contrast, noise, pronounced structural occlusions and visibility of anatomical subparts that interfered with the task of a meaningful surface delineation.

Application of the proposed method was straightforward, without modifications of algorithms and training procedures, and without use of additional methods for image pre-processing or model initialization that are often needed for an adaptation to heterogeneous and diverse conditions like in the presented case. Also, a set of segmentation parameters was used that follows a very similar rationale like in the previous segmentation scenarios.

As a result, a segmentation performance was reached with error measures

on the UHP_VENTRICLE dataset that are similar to the results produced by state-of-the-art approaches on the dataset of the CETUS challenge. Apart from the different dataset sizes (18 patients in the UHP_VENTRICLE dataset and 30 patients in the CETUS challenge), similar findings on image quality and on the presence of pathologies were described by the challenge organizers [226].

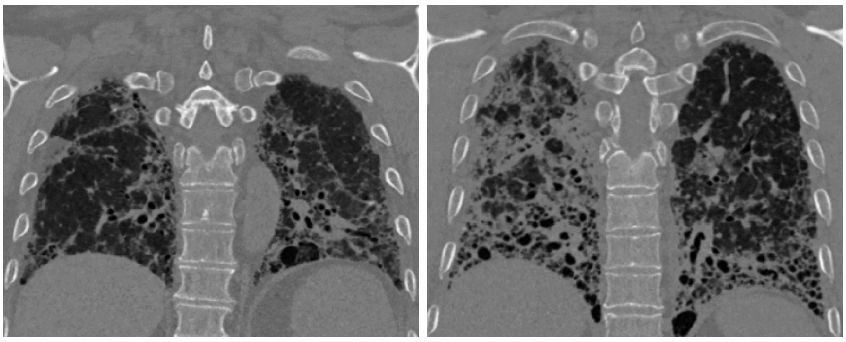
Altogether, the shown segmentation experiment underlines an immediate and successful applicability of the proposed method to highly heterogeneous conditions, without prior modifications or adaptations. Results were produced by the proposed method that are promising for the presented clinical scenario on the UHP_VENTRICLE dataset, in particular, regarding the comparably low average surface distance errors that were produced by the proposed method for both time frames. The landmark-based foundation of the proposed method also allows an extension of the proposed method by additional motion modeling and correction components, which seems promising regarding the encountered problems of missing image information from motion-induced shadowing of the ventricle wall.

6.4.4 Segmentation of Pathological Lungs in MDCT Data

In this experiment, a typical clinical application of the proposed method will be presented. The shown use-case scenario is concerned with the diagnosis and the treatment of idiopathic pulmonary fibrosis (IPF), a disease of unknown origin, which is often characterized by a sudden and severe exacerbation of the disease. The choice of treatment options significantly depends from the state of the disease and from the risk of exacerbation for examined patients.

In the following, an application of the proposed method on image data from IPF patients in an ongoing clinical study will be presented. For this study, a dataset of 420 CT volumes was acquired from patients suffering from mild to severe cases of lung fibrosis. The automatically generated segmentations of the proposed method built the foundation for a subsequent histogram analysis of the lung parenchyma. The analysis was intended to measure the amount of fibrotic tissue within the parenchyma, thus providing a possible clinical marker for the severity and for the progression of the disease.

Regarding the generalizability of the proposed method, a robust performance in the presence of pronounced appearance altering pathologies was of major interest in this experiment. Also, whereas the previous experiments were focused on a translation of segmentation performance to different use-cases and



(a) (b)

Figure 36: Examples of idiopathic lung fibrosis. (a) moderate case and (b) severe case. The discrimination of fibrotic tissue from the lung periphery becomes increasingly difficult. In addition, the overall appearance of the lung parenchyma is subject to heavy alteration.

imaging modalities, the experiment in this section served as a demonstration for the straightforward and robust applicability of the proposed method in an extensive clinical setting, which is another relevant trait for algorithms with regard to their generalizability to different use-cases.

6.4.4.1 Clinical Scenario

The clinical study behind this experiment was aimed at an examination of idiopathic pulmonary fibrosis (IPF). The pathology is characterized by fibrotic tissue growth within the lung's functional part (*parenchyma*), which can impede respiratory function. A CT image acquisition has been established as a standard tool for imaging based examination of patients with IPF. Different kinds of fibrotic tissue exist, with characteristic variations in density and texture. Mild to severe cases can be observed, and often, the origin of the disease remains unknown. Fig. 36 shows examples of increasingly severe lung fibrosis.

Apart from a visual inspection and evaluation by radiological experts, an image-derived assessment of fibrotic tissue that takes holistic image information into account for diagnosis, monitoring and evaluation of IPF, is highly desirable. The segmentation of the lung parenchyma is a crucial first step for the subsequent tissue analysis. It provides a separation of the relevant image regions from irrelevant surrounding tissue, which is a prerequisite for the targeted analysis of the lung parenchyma.

The segmentation of fibrotic lungs yields particular challenges (cf. Fig. 36c). The attachment of fibrosis to the pleura leads to regions where the parenchyma is hardly distinguishable from its outer surroundings. Purely density-based segmentation approaches often suffer from an under-segmentation of fibrotic tissue in the produced segmentation. Since the excluded fibrotic parts are highly relevant for an analysis regarding IPF, important information is lost from the overall content of the image analysis. Furthermore, the appearance of the fibrosis strongly depends on its type and on its severity, which causes an overall high heterogeneity in the images.

With its shape prior, the proposed method is expected to provide an additional means for a delineation of the lung surface by incorporating higher-level information in the delineation task. An improved separation of fibrotic tissue from adjacent non-lung tissue parts means an additional inclusion of relevant information that can be gained from the segmentation. From a successful segmentation of the lung parenchyma and a subsequent analysis of the visible fibrotic tissue, severity and progression of the disease can finally be assessed.

6.4.4.2 Data Material

A clinical dataset of 420 multi-detector CT (MDCT) volumes of the thorax from diseased patients was used for the experiments in this section (LUFIT_LUNG). Data was acquired from 127 patients suffering from idiopathic lung fibrosis at baseline and at follow-up dates, allowing patient monitoring regarding progressive disease. Cases were grouped in patients with stable disease (72 patients, median age 71 years, male 83%, smoking history 100%) and patients that had undergone one or more acute exacerbations (39 patients, median age 69 years, male 79%, smoking history 97%).

Images were acquired by a 4-slice Siemens Volume Zoom scanner, and by a 64-slice Siemens Somatom Definition AS scanner. For acquisition, thin-section MDCT were acquired in spiral mode scanning, covering the whole chest of patients in supine position. Patients were in addition instructed to perform an inspiratory breath-hold. Image resolution was isotropic with a spacing of ~ 0.7mm.

9 different reconstruction kernels were involved in the generation of the images from the dataset. They comprised various kernel classes used for filtered back-projection or for iterative reconstruction of the raw data (cf. [232] for further reading). The kernels have an impact on contrast and on noise levels in the resulting images. Thus, they emphasize the visibility of different tissue types associated with presence and development of lung fibrosis. For two pa-

Table 17: Parameter setting for IPF lung segmentation with the proposed method. Regression voting (top) and profile-based adaptation (bottom).

LOD	Stopping Criterion	δ	s	σ	Shape Parameters
0	I=10	-	-	15.0	1
1	I=10	-	-	15.0	37
2	I=10	-	-	15.0	37
Resolution	Stopping Criterion	γ	Δ	Radius	Shape Parameters
3	I=50	-	-	12	37
3	I=50	0.1	1	12	37
2	I=50	0.1	1	12	37
2	I=50	0.1	1	12	15
1	I=50	0.1	1	12	15

tients, a contrast agent was applied. The differing imaging conditions notably affected the overall appearance and the quality of the images.

Medium-soft kernels have been shown to produce more accurate and less noisy images than hard kernels [233]. Therefore, analysis was focused on datasets reconstructed with medium-soft kernels, and a training dataset for the proposed method was chosen based on I40 iterative reconstruction algorithms. However, since these reconstructions were not available for all patients in the dataset, a segmentation of images reconstructed by other kernels was also performed in the presented experiment. They comprise 10 images of reconstruction with B40 kernel, three images with B40s, one image with B40f, one image with B35f, one image with B46, one image with B31, one image with I30 and one image with I70). From a radiological point of view, the use of I70 kernel reconstructed images is preferred for a visual inspection and differentiation of fibrosis tissue. Therefore in addition, all 420 cases were also segmented for the I70 kernel reconstructed images, in order to examine robustness against the choice of a sharp kernel for image reconstruction in contrast to the medium-soft kernels.

Model training was performed on manual annotations of left and right lungs in another dataset of 39 lung CT images (LUFIT_TRAINING), that were acquired under the same conditions as the LUFIT_LUNG dataset. Images were provided by manual slice-by-slice annotation, carried out by a medical expert. CT images from iterative reconstruction with I40f were used for model training.

6.4.4.3 Experimental Setup

One goal of the segmentations in this section was to examine if the proposed method as a shape prior based approach allows a better inclusion of fibrotic tissue in the segmentation volume of the LUFIT_LUNG dataset. Since ground truth labeling of fibrosis in the lung was not feasible for the 420 datasets, another approach was chosen for an indication of additional fibrosis segmentation in the dataset.

The idea was to compare results of the proposed method on the LUFIT_LUNG dataset with results from an accurate intensity-based lung segmentation method. Based on earlier reports [234], it was expected that the compared method would still succeed in the minor or moderate cases, but was prone to fail for the more severe cases of lung fibrosis. In these cases, a strong decrease of segmented lung volume should be observed, whereas the proposed method was expected to be robust against volume loss in these cases.

According to this idea, a lung segmentation approach was chosen as a reference method that had proved to provide a highly accurate lung segmentation in absence of severe pathologies [235] (*Yacta*). The method scored third in the LOLA 2011 lung segmentation challenge. It relies on a number of segmentation steps, including tracheabronchial tree segmentation, seed point detection, threshold based region growing, and correction using morphological operators and other basic image processing techniques. A mean Jaccard coefficient of 0.970 was reached for the LOLA 2011 lung segmentation task.

After application of both methods on the LUFIT_LUNG dataset, a comparison of segmentation volumes was performed to identify cases of fibrosis where the shape model based approach of the proposed method was more successful in a separation of fibrosis from the lung periphery. Although this approach was no proved criterion for an additional segmentation of lung fibrosis, it was still expected to serve as a good indicator.

In a second setup, the lung segmentation of the proposed method was used to derive a possible clinical marker for an exacerbation of lung fibrosis. In [236], the 40th and 80th percentiles (p40, p80) were identified as potential biomarkers for an increase of two types of pathological changes in the parenchyma, i.e. ground-glass (p40) and fibrosis (p80). Thus, an increase of p40 and p80 was expected for rapid pathological changes of the lung parenchyma in case of an exacerbation. Since the percentiles are directly affected by the amount of segmented fibrosis in the lung, successful segmentations of the more severe cases by the proposed method were expected to better reveal significant changes of p40 and p80 in case of an exacerbation. To this end, information on known

exacerbation of patients was compared with the observed changes of p40 and p80 based on the segmentation of the proposed method. Finally, a correlation analysis of the p40 and p80 biomarkers, with ground-glass and fibrosis percentage results from a commercial software (*Imbio*) was performed to further underline the meaningfulness of p40 and p80 as potential biomarkers.

All segmentation of the proposed method were performed without previous model initialization from the image center, using model fitting parameters from Tab. 17. No changes of model training, algorithmic adaptations or additional methods were used for an application of the proposed method in the lung segmentation scenario.

6.4.4.4 Fibrosis Segmentation

After segmentation of all 420 images in the LUFIT_LUNG dataset, a mean volumetric difference of 235.33 ± 161.96 ml between the proposed method and Yacta was found (p-value < 0.001 as assessed with Wilcoxon's signed rank test). Visual inspection of the data confirmed that the found volumetric differences reflected an additional segmentation of fibrosis in the parenchyma by the proposed method. Also, the found volumetric difference provided a suitable means for a selection of mild, moderate and severe cases from the dataset. Segmentation took $\sim$ 8 minutes per image and 3 days in total on one Intel core i7 by the proposed method.

Fig. 37 shows representative examples for the Q_{95}, Q_{90}, Q_{50}, and Q_{10} percentiles of the volumetric difference. The shown example illustrate the delineation of severe fibrosis from the lung periphery by the proposed method, despite the underlying tissue similarity. The examples show that the intensity-based Yacta cannot achieve a separation of these cases, despite its still remarkable exclusion of high-density parts from the parenchyma. Also, the tracheabronchial tree segmentation was sometimes affected by the presence of the lung fibrosis, leading to an additional exclusion of high-density tissue from the segmentation volume. Altogether, a benefit of additional fibrosis segmentation was observed for the moderate to the severe cases, which roughly contributed half of the examined patients in the data collective.

Regarding segmentation of the aforementioned sharp I70 kernel reconstructed images for all cases, segmentation results of the proposed method were only slightly preferable on the I40 reconstructed images (cf. Fig. 38). Apart from this, no notable differences were produced by the proposed method on the I70 kernel reconstructed images, or on the other encountered images from differing reconstruction kernels (cf. Sect. 6.4.4.2).

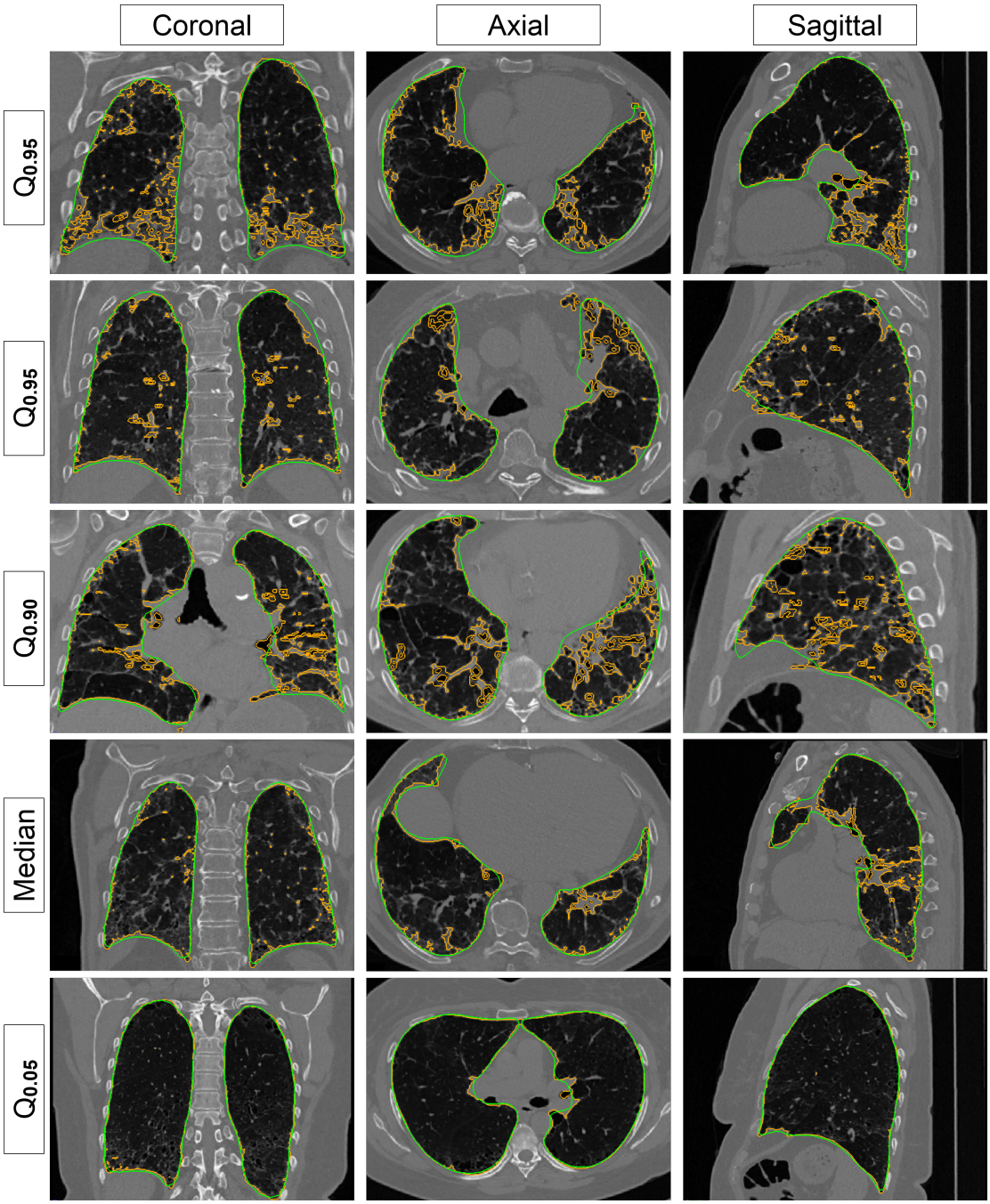


Figure 37: Lung segmentation examples from the LUFIT_LUNG dataset. Results of the proposed method (green) and from the intensity-based reference approach Yacta (orange). Examples from the Q95 percentile (very severe cases), from the Q90 percentile (severe cases), from the median (moderate cases) and from the Q5 percentile (mild cases), based on the volumetric difference between the proposed method and Yacta.

Problems of the proposed method were encountered in terms of under-segmentation of pronounced concavities of the lung (cf. Fig. 39a), which was most often encountered at the lung base. Also, a difficult outlier was encountered, as shown in Fig. 39(b), caused by a catheter that was inserted near the lung base, which lead the model fitting of the proposed method to a false upward placement.

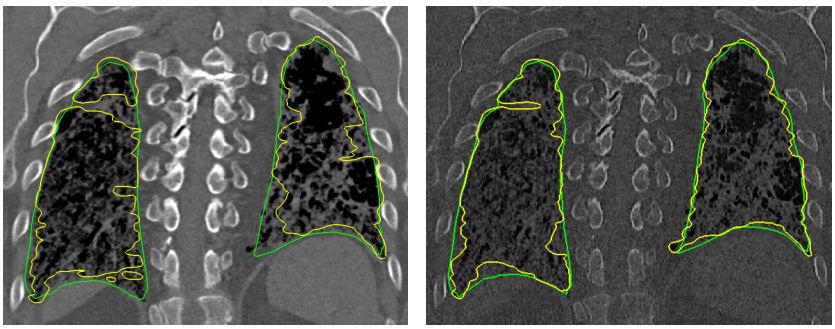
6.4.4.5 Clinical Biomarker Evaluation

To evaluate the expressiveness of the 40th and 80th histogram percentiles (p40 and p80) as disease markers that are obtained from the segmentations of the proposed method, a regression analysis was conducted on the underlying dataset. There, the statistical relationship between the histogram percentiles and two clinical parameters from the dataset was examined: the functional vital capacity (FVC) which represents the respiratory function of a patient; and the incidence of an acute exacerbation i.e. a dramatic aggravation of the disease. To this end, a linear mixed-effects model was fitted to the data using restricted and standard maximum likelihood estimation. There, a decrease in FVC and an increase in the acute exacerbation probability was observed ($p < 0.05$). Furthermore, both percentile markers were compared with the estimated fibrosis percentage from the fibrosis annotation software IMBIO. A positive correlation between p40 and p80 from the proposed method, and the estimated fibrosis percentage from IMBIO was found ($r = 0.918$ for p80, $r = 0.784$ for p40). Although the accuracy of IMBIO could not be considered to be a gold standard, the image-based estimation results from IMBIO did confirm the expressiveness of the histogram-based results of the proposed method.

6.4.4.6 Summary and Conclusion

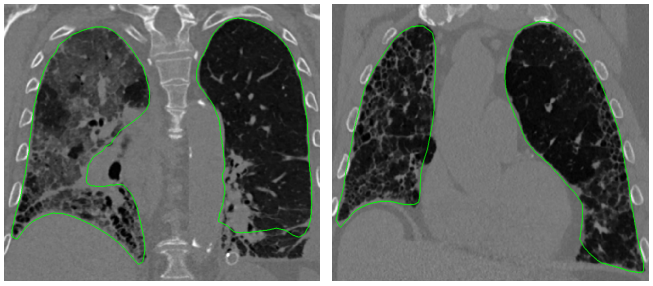
In this section, a clinical application example for use of the proposed method on a larger pathological dataset was presented. The shown cases of lung fibrosis cause strong global changes of organ appearance, and they hamper a delineation of the lung surface due to the high similarity of tissue ingrowth and surrounding non-lung tissue. Particularly, the more severe cases have previously been reported to cause problems for density-based approaches [234], like the chosen reference method Yacta in the shown experiments.

Based on a ranking of volumetric differences to the reference method Yacta, a qualitative analysis of the 420 cases of lung fibrosis segmentation has shown the proposed method achieved an accurate separation of lung fibrosis at the



(a) (b)

Figure 38: Segmentation example of a case of severe fibrosis in two different imaging conditions. (a) I40 kernel and (b) I70 kernel. Overall, fibrotic tissue was more effectively separated from the adjacent pleura by the proposed method (green contour) than by a region-based method (yellow contour) that had been adapted for fibrotic lung segmentation. The proposed method worked slightly better on the I40-reconstructed image, which had also been used during model training. Segmentations of the compared method worked better on the contrast-enhancing I70 kernel.



(a) (b)

Figure 39: Cases that were difficult to segment by the proposed method. (a) 2D cross-section of a segmentation outlier that was caused by a catheter passing at the base of the left lung. (b) Example of cavity under-segmentation by the proposed method.

lung periphery also in severe cases of the disease. Yacta as an intensity-based approach has still shown notable robustness in the moderate cases of lung fibrosis. In the severe cases of lung fibrosis, it was evident that density-based information alone was not sufficient to accurately separate the fibrosis from the lung periphery. Here, the use of shape prior from the proposed method has shown to be beneficial across the LUFIT_LUNG dataset.

As reported in many earlier cases, the segmentation of high curvature parts was also a problem for the proposed method in case of the lung. It can be assumed that part of the reason for this problem is linked to the missing flexibility of surface deformation of 3D-SSM based approaches [156]. An incorporation of a final deformation step where landmark shifts are regularized to allow a consistent, more flexible model adaptation could help solve the described problem.

Altogether, the presented scenario of IPF lung segmentation has reconfirmed the generalizability of the proposed method, in terms of straightforward and robust applicability for different use-cases. No algorithmic modifications or adaptations of the training process, and no additional methods for model initialization or other refinements were required for successful application under the shown clinical conditions. This sets the proposed method apart from previous model-based approaches where tailored solutions were required for successful use of 3D-SSM for lung segmentation [237], [238]. However, the use of additional measures seems worthwhile to address the aforementioned limitation of 3D-SSM regarding concavity segmentation. With its landmark-based foundation, its integration in the general 3D-SSM framework and its immediate robust performance, the proposed method offers a convenient platform for future extensions, e.g. towards a more flexible surface deformation.

Finally, the more robust segmentation of severe cases of IPF has allowed the extraction of the p40 and p80 biomarkers in those cases that are particularly relevant for use of these markers. The presented findings on the extracted biomarkers have produced first evidence that these markers are associated with exacerbation of the disease and with functional parameters of the lung.

6.5 Runtime and Robustness

In the following experiments, different choices and parameters for the random regression forest training were examined. When large datasets and use case varieties are encountered, a reduction of training time becomes particularly important. Several options exist for a speedup of the regression forest training,

which will be considered in this section. Furthermore, the proposed method is expected to profit from the well-known robustness of random forests, in terms of a generally easier application to different use-cases without an elaborate parameter tuning. Therefore, the goal of the following experiments was to demonstrate that a robust performance can immediately be achieved by the trained random regression forests for a larger range of parameters.

For an analysis of training duration and random regression forest robustness, parameters were of interest from an application point of view, based on the underlying use-case scenario and the encountered image material. For an application of the proposed method, this concerns the amount and the extent of image information that is considered during training and during segmentation, and furthermore, the number of trees that are used by the trained forests during landmark detection. Furthermore as proposed in Sect. 4.4, the use of extremely randomized trees and of a faster splitting criterion was examined (*Extra-Tree*). It was reported earlier that these measures can reduce training time and increase robustness in certain applications [178].

Apart from this, limiting tree depth is another possibility to speed up training. However, this is an option that is rarely regarded in the context of medical image segmentation using random regression forests, and that should generally not be required due to the robustness of the bagged trees against over-fitting (cf. 4.4 and [173]). Similarly, a large variety of other improvements and tuning options are available for random forests. Although it can be assumed that accuracy can be improved by various measures, an optimization of segmentation accuracy for particular use-cases is beyond the scope of this thesis and will be left for future work.

6.5.1 Experimental Setup

Based on the stated goals above, the effect of the chosen forest parameters on training runtime, on segmentation accuracy and on the convergence of detection results for increasing numbers of trees was evaluated. The CT liver segmentation on the DKFZ_LIVER dataset from Sect. 6.3.2 served as a test scenario for the following experiments. The experiments were carried out in three setups.

First, the effect of sample count, and the use of extremely randomized trees (*Extra-Tree*) in conjunction with the faster variance based impurity measure on training duration and on segmentation accuracy was evaluated. This experiment served to determine the sensitivity of the tested random regression

forest types against a reduction of training samples, and to determine possible speedup of training without loss of segmentation accuracy in the examined scenario. Sample counts were reduced by lowering the bagging size for each tree, i.e. the number of random samples that are passed to each tree during training. Samples were taken without restriction from the underlying training images to provide constant coverage in the image domain.

For speedup measurement, training was conducted using both the default trees and the faster extremely randomized trees while exponentially reducing the employed size of training samples. In this manner, 32 versions of the liver model from Sect. 6.3 were trained in the three levels of detail (LOD) of the proposed method. That is, two cross-validation sets at 8 exponentially dropping reduction levels, for each the default trees and the extremely randomized trees. Training was performed on an Intel core i7 (hexacore, parallelized). For the measurement of segmentation accuracy, model fitting was carried out in all LODs on the 10 test cases. Segmentation accuracy was determined using the Dice coefficient. All segmentation experiments were conducted for the default trees and for the extremely randomized trees in all sample reduction levels.

In a second setup, the effect of the above parameters on the stability of model fitting results and on the convergence of the random regression forest predictions was evaluated. Model fitting was carried out separately in each LOD for all sample reduction levels used in the first experiment. Model fitting was performed repeatedly, each time with an incremented count of tree estimates.

Then, convergence and overall spread of the model fitting results was assessed by measuring the volume overlap of all model fitting results with their mean shape using the Dice coefficient. For a sufficient number of training samples, model fitting results are expected to converge to a common solution when additional tree predictions are incrementally included during detection. An insufficient number of training samples should lead to overall worse results and to an increase of dispersion.

To examine the effect of the sample reduction separately for each LOD, the reduction was only applied to the current LOD of interest, while no further sample reduction was applied in the preceding LODs. Thus, the effect of sample reduction in a particular LOD could be determined without interference from sample reduction in previous LODs.

Finally, the same experiment was repeated measuring the influence of the Haar-like feature patch size instead of the previous sample reduction (cf. Sect. 4.3.2). Evaluation was carried out for increasing patch sizes of the feature

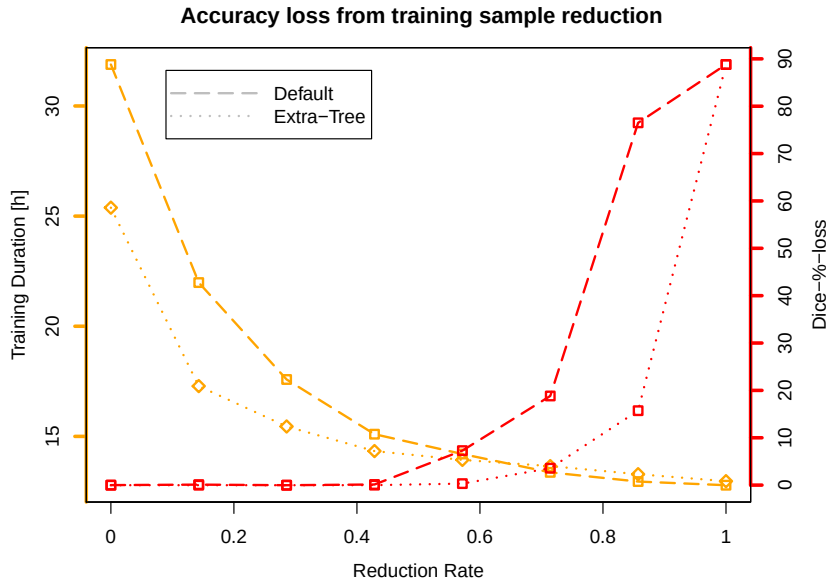


Figure 40: Landmark detector training: runtime and accuracy analysis. Training duration [h] (orange) and Dice-%-loss (red) for default trees (dashed line) and for extremely randomized trees (dotted line). Speedup was achieved through exponential reduction of training samples (from left to right). Considerable speedup was reached in the first 4 reduction steps without significant loss of accuracy.

descriptor, measuring precision and convergence of the model fitting results separately in each LOD for incremental tree counts. Similarly to the second experiment, optimum patch sizes were chosen in the preceding LODs to allow a separate assessment in each LOD. Again, two test sets of 5 images each were excluded from training to allow cross-validation of the forest predictions. Only the fast splitting tree version was involved in these experiments, since results will show that the faster version reaches an equivalent performance like their slower counterparts. On the two cross validation sets, 30 liver models were trained in LODs=0,1,2 with 10 patch sizes for each set.

6.5.2 Regression Forest Performance

Fig. 40 shows the total training duration for both tree versions for all reduction levels (orange contours). With increasing sample reduction, the training duration expectedly dropped at an exponential rate for both tree versions. A significant speedup was reached within the first four reduction steps, with

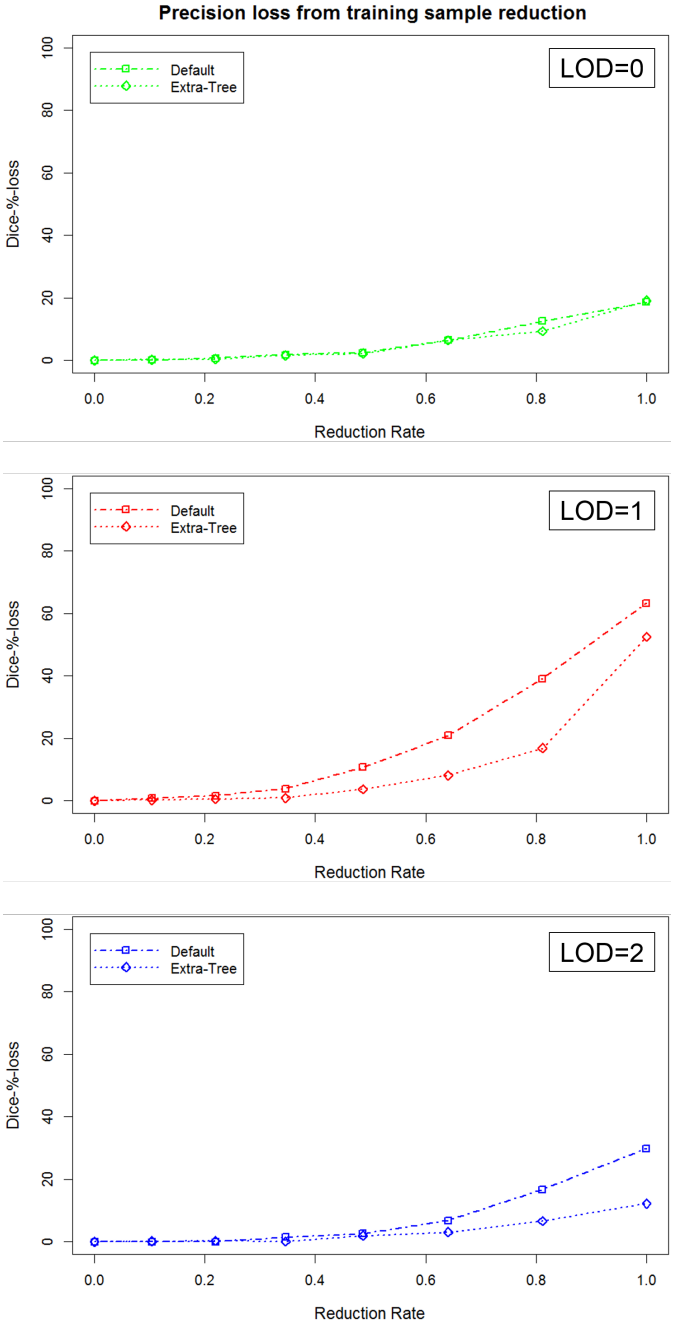


Figure 41: Effect of training sample reduction on random forest precision. The top, middle and bottom plots show the loss of precision for model fitting results in each LOD, when exponentially less training samples were used for the random forest training (from left to right). Model fitting was started several times for an increasing number of trees (n=1,$\cdots$,12). Overall precision of results around the mean outcome of the model fitting was measured using the Dice-coefficient.

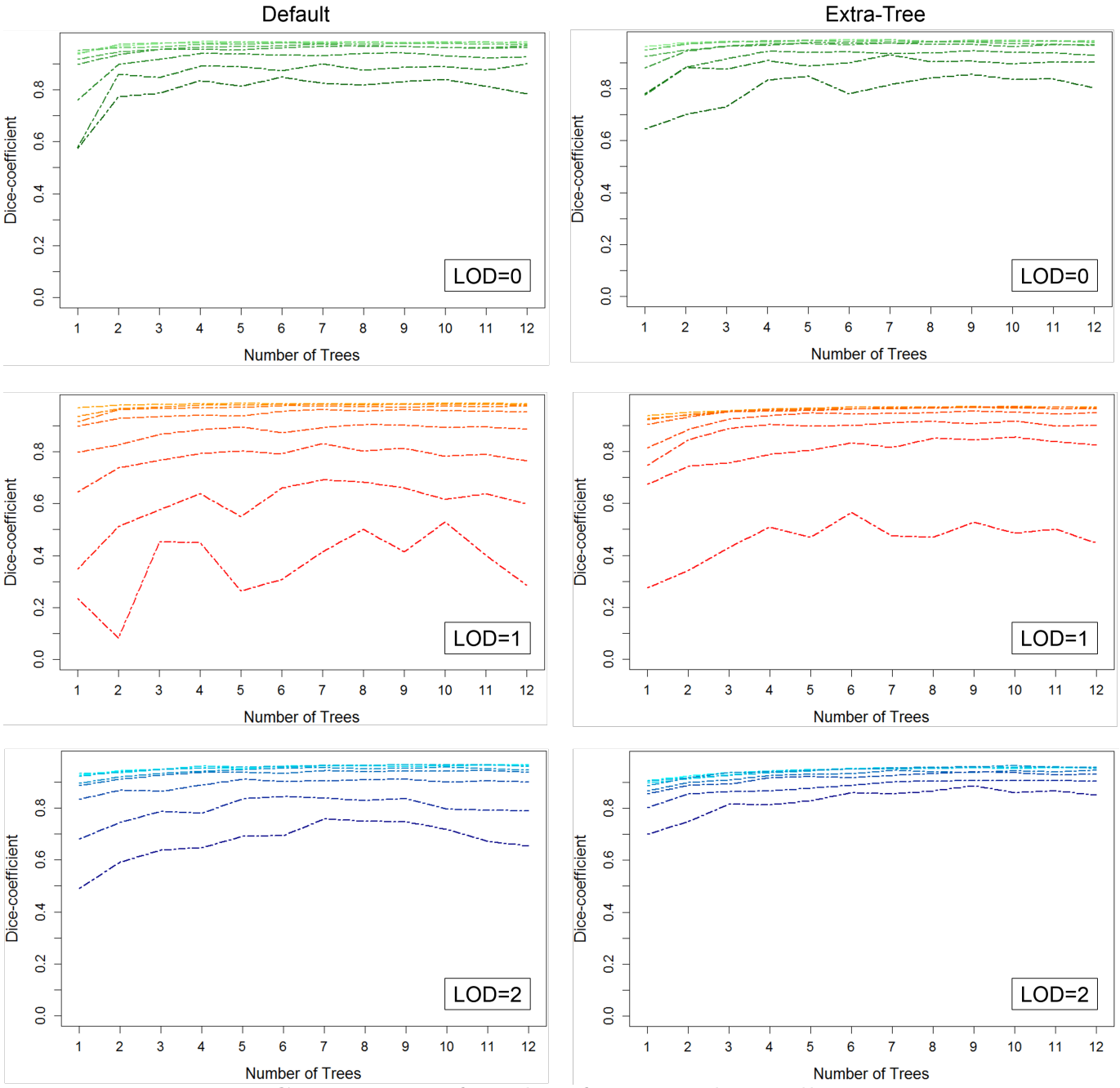


Figure 42: Convergence of random forest results in all training sample reduction levels. Each plot shows the convergence of model fitting results towards a common solution, when increasing numbers of trees are used during landmark detection. Convergence was measured in all LODs using the Dice-coefficient (top, middle and bottom rows) for the default regression trees (left column) and for extremely randomized trees (right column), based on the volume overlap of results with the mean outcome of the model fitting.

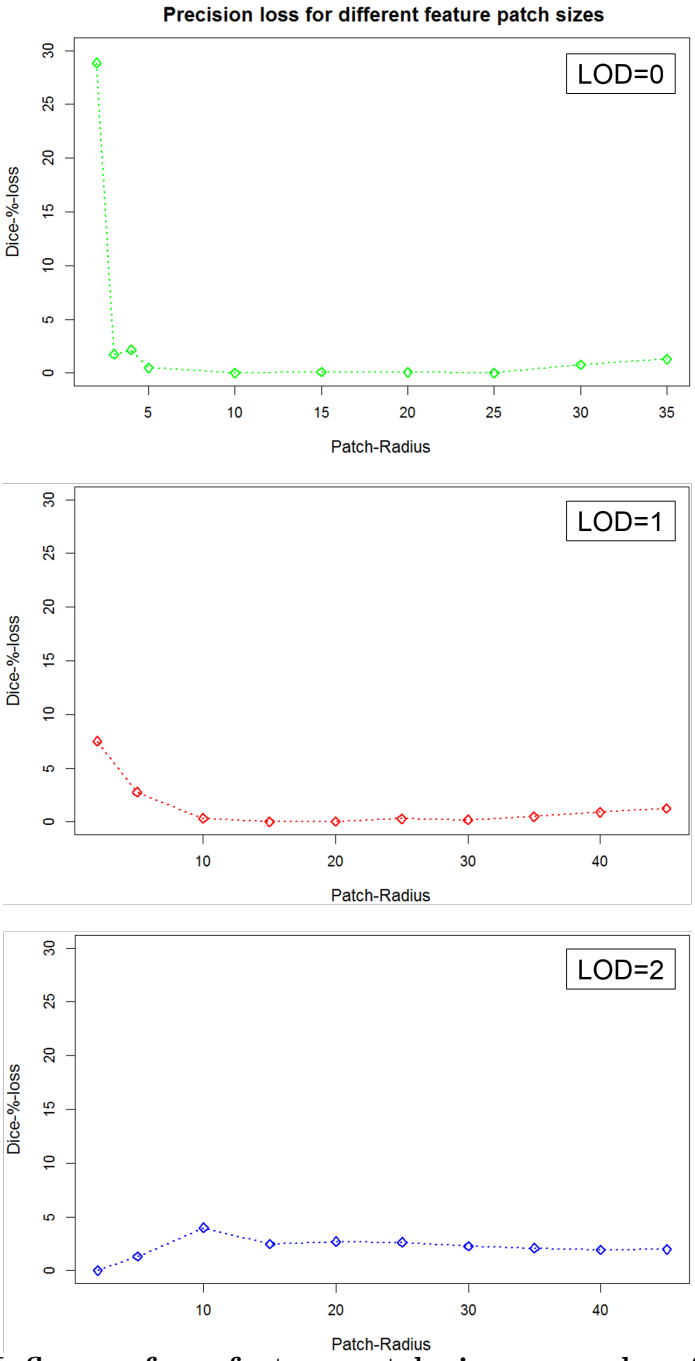


Figure 43: Influence from feature patch size on random forest precision. The top, middle and bottom plots show the loss of precision for model fitting results in each LOD, when different feature patch sizes are used for the training of extremely randomized trees (from left to right). Model fitting was started several times for an increasing number of trees (n=1,$\cdots$,12). Overall precision of results around the mean outcome of the model fitting was measured using the Dice-coefficient.

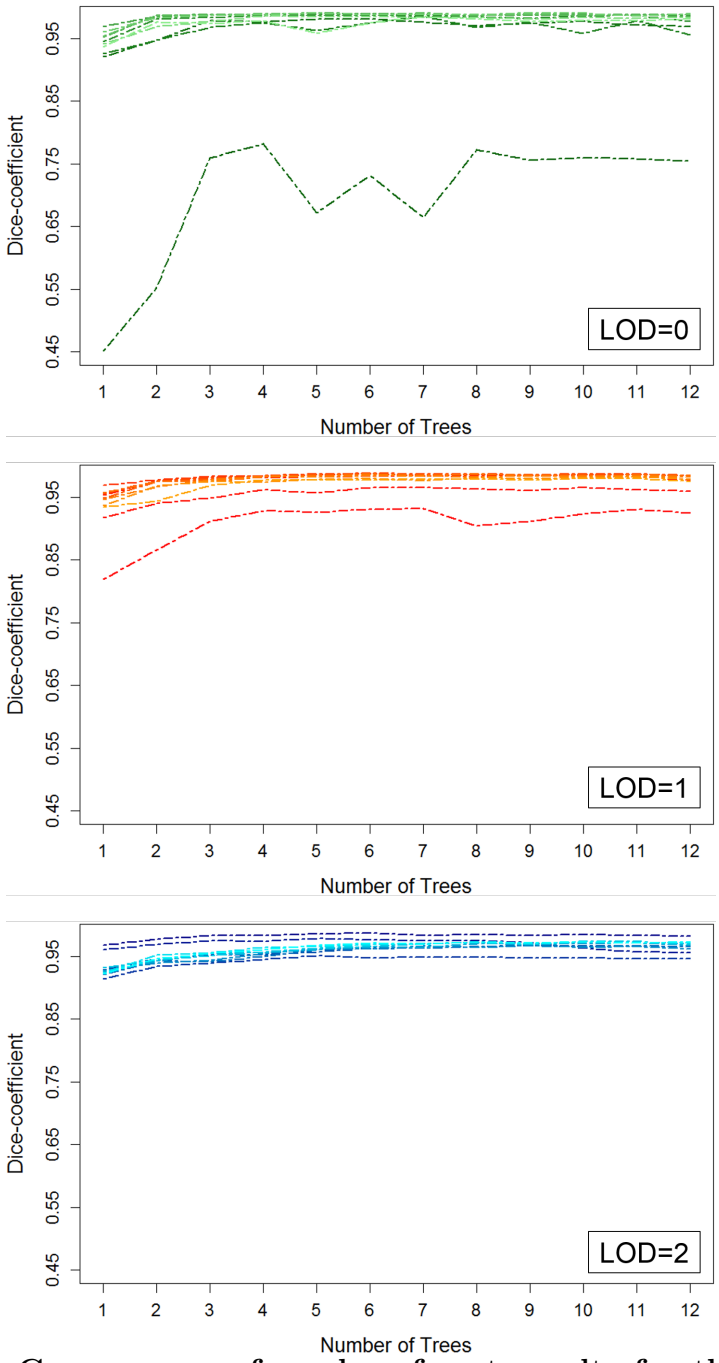


Figure 44: Convergence of random forest results for the different feature patch sizes. Each plot shows the convergence of model fitting results towards a common solution produced by the extremely randomized trees, when increasing numbers of trees are used during landmark detection. Convergence was measured in all LODs using the Dice-coefficient (top, middle and bottom rows), based on the volume overlap of results with the mean outcome of the model fitting.

training duration dropping from $\sim$ 32 to 16 hours for the default trees, and from $\sim$ 25 to 15 hours for the fast splitting trees. No loss of segmentation accuracy (Fig. 40, red contours) was observed throughout the first four reduction steps. Beyond this point from reduction level 5, accuracy declined notably for the default forests. In contrast, decline started significantly later for the extremely randomized trees, with a clear loss of accuracy at level 6.

Figs. 41 and 42 show the results of the second experiment. Overall, no larger loss of precision was observed for the first 5 reduction steps at all LODs for both forest types, except for a loss by the default type at LOD=1 and reduction level 4 (Fig. 41). Precision loss grew notably beyond this point. Loss was very similar for both forest types at LOD=0. In LODs=1,2, extremely randomized trees suffered considerably less from precision loss than their counterpart. At LOD=1 for both types, and at LOD=2 for the default forest type, a generally higher loss rate was observed in comparison with the other results.

Results for the forest convergence in Fig. 42 reflect the findings of the precision loss analysis. Convergence is clearly flawed for the 3 largest reduction levels at LOD=0 (both forest types), for the largest 3 reduction levels at LOD=1 (largest 4 reduction levels for the default type), and for the largest two reduction levels at LOD=2 (largest 3 reduction levels for the default type). All other forest results show convergence towards a common solution, with a final overall precision of $\sim$ 95% Dice.

Figs. 43 - 44 show the results of the third experiment. Each patch size is represented by its patch radius, i.e. patch size equals the center voxel plus two times radius in each dimension. For each LOD, a stable patch size interval can be identified where no significant loss of precision was encountered. That is, at LOD=0 between patch radii 5 and 25, at LOD=1 between 10 and 30, and at LOD=2 between 15 and 45. Note that at LOD=2, a decrease of precision loss was observed for patch radii < 10, unlike at the other LODs. Thus, the Dice-%-loss appears elevated between 15 and 45, however, absolute precision was not lower than at the other LODs.

Results for the forest convergence in Fig. 44 again reflect the findings of the precision loss analysis. Convergence is clearly flawed for the 3 smallest patch radii at LOD=0, for the largest 2 smallest reduction levels at LOD=1, and for patch radii 5 and 10 at LOD=2. All other forest results show convergence towards a common solution, with a final overall precision of $\sim$ 95% Dice.

6.5.3 Summary and Conclusion

A high performance of the trained landmark detectors was reached in the presented scenario, for a large range of the considered training parameter settings. A high quality of convergence was achieved for the regression forest results, within a wide range for patch size and training sample count. Same convergence was guaranteed for an ensemble tree count of 8 and larger, regardless of the chosen forest type. High accuracy and precision of model fitting results were achieved within the determined range.

The use of extremely randomized trees instead of the default forest type has shown to be beneficial in terms of training duration and robustness. Also, training duration could significantly be reduced without loss of accuracy using the proposed measures. Altogether, the well-known robustness of the employed random regression forests was confirmed by the experiments with the proposed method.

From a practical point of view, this significantly facilitates training and application of the proposed method to different use-cases, by means of a straightforward choice of training parameters and reliable and strong performance for the proposed landmark detectors. Although there is no guarantee for equivalent results in all future cases, the shown robustness and simplicity of the considered random regression forests, as well as many examples for their success in other applications underline their contribution to a higher generalizability of the proposed 3D-SSM.

CHAPTER 7

Discussion of Methods and Results

Well-known problems of strong dependence from model initialization, of pose-dependent landmark visibility and of limited capture range for 3D-SSM were successfully addressed by the proposed method. Highly reproducible results were achieved for arbitrary pose changes and from large distances during segmentation, and the previously strong influence from model initialization was effectively alleviated. Susceptibility to locally optimal solutions during segmentation was significantly decreased by the proposed measures, leading to an overall reduced segmentation error and to a generally higher accuracy for 3D-SSM. High robustness and easy applicability of the employed random regression forests, and a strong independence from model initialization significantly facilitate an application of 3D-SSM to different use-cases. Altogether, this increases the availability of typical benefits from 3D-SSM in terms of their higher-level information modeling, to a broader scope of medical imaging applications.

In this chapter, the findings of this thesis will be discussed with regard to the achieved independence from model initialization (Sect. 7.1), to a widespread applicability for 3D-SSM (Sect. 7.2), to the capabilities and limitations considering imaging conditions (Sect. 7.3), encountered anatomical variabilities and pathologies (Sect. 7.4), to the computational performance (Sect. 7.5), and to the general implications of the proposed method for model-based image segmentation (Sect. 7.7).

7.1 Segmentation without Initialization

The results from Sect. 6.2 demonstrate that the proposed method has reached a strong independence from the model initialization step during segmentation. Two criteria were met that provide an independence of the proposed 3D-SSM from its model initialization: first, a high reproducibility of results was achieved for arbitrary variations of the initial placement (cf. Sect. 6.2.2). And second, the high reproducibility of results was sustained for starting positions that were widely distributed across the image (cf. Sect. 6.2.3).

The reached reproducibility of results can be attributed to the proposed omni-directional landmark detectors and to their ability to pinpoint separate surface landmarks from a 3D scope. This is in contrast to the exclusive 1D search of previous 3D-SSM, where a reproducible detection of landmark positions is highly unlikely for changes of model pose, e.g. from different model initializations. However, a comparably small local spread of results could also be observed for the proposed method in the corresponding experiments. This is likely to be caused by the profile-based finalization of the previous sparse model segmentation for all surface landmarks as described in Sect. 5.3.

The final profile-based step of the proposed model fitting was incorporated as a solution to several problems that are encountered when incorporating landmark-wise detectors in 3D-SSM. First, the computational complexity during segmentation renders an application for all landmarks in all stages of the model fitting unfeasible. This was also reported earlier in [136], concerning an extension of a regression-voting optimization to 3D domain. Unfortunately, no solution was proposed afterwards for an efficient extension to higher dimension. Secondly, the free and omni-directional deformation of landmarks for 3D-SSM quickly leads to overlapping landmark shifts, that require further regularization. In [29], [135], additional freedom from the shape constraint was allowed in the later stages of the CLM model fitting. However, for the presented 2D contour models, the problem of landmark intersections is much less prominent, due to the rigid shape deformations of the considered objects, and due to the smaller number of contour landmarks in comparison to 3D surface models. Therefore, no regularization of intersections was proposed.

The use of the proposed sparse modeling during the regression-voting based model fitting, and of the profile-based approach for the last part of the model fitting has shown to provide a good solution to the aforementioned problems. Regarding the final profile-based fitting, the proposed method builds on the high reproducibility of results provided by the landmark detectors in the previous stages of the model fitting.

As a solution to the observed smaller spreads of results during the profile-based model fitting, an introduction of free deformation for the omni-directional landmark detectors could be of interest. Approaches which use 3D shape deformations were e.g. proposed in [156], [239]. The approach from [239] however does not constrain model displacements in the underlying deformation field, and the absence of shape model regularization gives rise to surface smoothing and remeshing operations. The approach from [156] introduces a MRF-based regularization of omni-directional landmark displacements, which fits the proposed method. Well defined displacements could however also not be guaranteed in all cases. Therefore, compensatory measures can be introduced that take effect at different parts of the surface model [110], which might hamper generalizability to other cases.

The high reproducibility of results means that model initialization is no longer a greater source of error for the proposed model fitting. Thus, the influence from model initialization can practically be neglected for an application of the proposed method. For 3D-SSM in general, this is a significant trait since their former sensitivity to the model initialization necessarily lead to an increased optimization effort, which includes highly complex measures for an accurate initialization and for the robustness of the subsequent model fitting. In this regard now, only the performance of the employed model fitting has to be considered during an application of the proposed method. Due to its considerable capture range, the proposed method does not require any placement at the organ of interest, which makes additional methods for previous organ detection obsolete. In practice, this considerably facilitates an application of 3D-SSM. As illustrated in the capture range experiments, the non-local appearance modeling of the proposed method allowed an effective use of widespread image information. This was shown in direct contrast to the local appearance modeling of the compared 3D-SSM, where a drastic decrease of usable information was observed with growing distance from the sought organs of interest.

In the experiments from Sect. 6.2.3, the achieved capture range for liver and spleen had actually exceeded the previously chosen sampling range during training. In contrast, the capture range for the kidney was rather in accordance with the sampling range. This could be attributed to the more constant locations of liver and spleen, when compared with the highly variable location of the kidney. The latter seemed to counteract the effect of the salient landmark regions that would otherwise provide meaningful information in the absence of local training samples (Sect. 5.3).

With its robustness against changes of model pose and its capture range, the proposed method is the first 3D-SSM that reaches the described degree of independence from model initialization. In [144], a similar reproducibility was reached for a 3D-SSM of the liver. However, for this purpose, a rule-based model-fitting was introduced where different surface parts were controlled by a set of numerous parameters during segmentation. The presented free parameters are expected to significantly impede an applicability to other use-cases. Also, the method only worked in the vicinity of the liver and required a rough initialization near the target organ. The shape regression machine from [133] showed sensitivity of results from initial pose estimation, and the approach was limited to 2D domain. In [239], model fitting of a deformable surface model was proposed by means of boundary regression. Application without previous model initialization was achieved, however, information on unambiguous landmark positions got lost in the boundary regression, and reproducibility of model fitting for model pose changes is unclear. Furthermore, surface regularization like smoothing and remeshing was regularly applied, which substantially alters surface topology and thus did not preserve point-to-point correspondences like 3D-SSM based approaches.

7.2 Wide-spread Applicability

The experiment in Sect. 6.3.3 has shown that the segmentation accuracy of 3D-SSM can significantly be improved by the proposed omni-directional landmark detectors. This was reflected by a generally lower number of false boundary segmentations, and by a more detailed delineation of higher curvature parts of the liver. There, the landmark detectors were beneficial in providing landmark positions that effectively promoted a more favorable model fitting during segmentation. This is in contrast to the purely profile-based 3D-SSM, where model fitting was more often mislead to boundaries outside the liver. In practice, such segmentation errors can often only be avoided by a more accurate thus elaborate model initialization.

The shown improvements of accuracy also support the applicability of the proposed method to different cases, as it is easier to reach higher accuracy in these cases without the necessity for measures that compensate for segmentation errors from the previous limitations of 3D-SSM. These measures typically increase complexity by involving additional free parameters that are difficult to optimize for many specific cases (cf. e.g. [110], [144]). By contrast, the benefits of accuracy for the proposed 3D-SSM were reached by means of the

employed landmark detectors, where the patent robustness and flexibility of random regression forests greatly facilitate training and application to different scenarios. Finally, the proposed method acts as a standalone technique without using additional methods for an accurate model initialization. This also significantly reduces complexity in terms of extra algorithms, additional free parameters and limiting basic assumptions. As a consequence, the proposed enhancements can foster the role of 3D-SSM towards a more general-purpose segmentation methodology.

Concerning the whole model training process, no adaptations or algorithmic changes were required for the presented applications. Still, there are model fitting parameters of the proposed method that needed optimization in the shown experiments, so as to reach best performance. The tuning of the landmark detectors was relatively easy and followed an intuitive rationale. The basic strategy was to increase the degree of freedom of the shape prior model step-wise from LOD 0 through 2, in order to bring the model fitting as close to the sought organ boundary as possible, i.e. to the predicted positions of the landmark detectors. This is then supposed to provide best conditions for the final profile-based adaptation step. In addition, three choices have proved effective to improve results for single outlier cases that are usually encountered in each dataset (also cf. Sect. 6.3.2): to employ a shift of vote displacements or a scaling step at any of the LODs, and to limit the degree of freedom of the shape prior model at any of the LODs 1 and 2.

The profile-based adaptation step of the proposed method remains as a more finicky part regarding its optimization. In contrast to the intuitive nature of the regression voting, it is still subject to the described unpredictable nature of the profile-based search. Although larger principles for improved accuracy can be identified, they are not so easily understood and definitely require expert knowledge, in contrast to the regression voting step. Nevertheless, the known robustness against locally optimal solutions, the lifted major influence from model pose and from initialization, and the absence of many additional measures for the proposed method notably facilitate an application to different cases. This is in contrast to known previous measures to compensate for the original limitations of 3D-SSM, where fine grain tailoring of appearance modeling and of model fitting at different surface parts have shown to deliver the best remedies so far [110], [144] (cf. Sect. 6.3 for further details).

This was illustrated by the application of the proposed method in the public segmentation challenge 'SLIVER' (cf. Sect. 6.3). There in its role as an adaptable standalone tool, the proposed method has shown to produce simi-

lar results like customized segmentation approaches that rely on a variety of techniques to reach high performance. With its results, the proposed method entered the range of the best-performing methods. Higher performance was only reached by methods that employ an elaborate design and an apt combination of techniques that necessarily impede an immediate application to other use-cases.

Regarding the other 3D-SSMs of the challenge, the proposed method counts towards a rather small number of approaches that also aim at a higher generalizability in addition to accuracy (cf. Sect. 6.3.3). The proposed method extends the family of these generalizable 3D-SSM into the performance range of highly specialized segmentation approaches. In this endeavor, it ties in with the previous strong learning-based approach from [129]. Also, the proposed method has reached a similar performance like the highly successful, expert tailored 3D-SSM from [110].

With its combined flexibility and performance, the proposed method can serve as a convenient platform for future development. In terms of mere segmentation performance, additional improvements can be added to take further steps in the direction of a more generalizable and accurate 3D-SSM. The proposed method can also build the foundation for special applications with a strong focus on high accuracy in particularly challenging image data.

In the experiment from Sect. 6.4.1, the proposed method was successfully applied to the multi-organ segmentation task, without necessity for a previous adaptation of the model training. In this scenario, the simultaneous detection and segmentation of distributed organs had to be provided by the proposed method. This is in contrast to the often encountered task of single organ segmentation from a more localized image context. Also during application, large variations of the underlying image geometry had to be coped with, in terms of field of view, size and spatial resolution of the images. These challenges were successfully handled by the proposed method, which was in accordance with the model fitting results shown in Sect. 6.2.3.

In the presented labeling challenge, the proposed method reached a similar performance like state-of-the-art multi-atlas registration methods. This is of particular relevance for the proposed method and for 3D-SSM in general. It was shown that 3D-SSM can generalize well beyond the role of a dedicated tool for single organ segmentation. In case of the proposed method, this was reached by transforming 3D-SSM from being less robust and highly dependent to a more robust and autonomous technique. It was demonstrated that these 3D-SSM can autonomously produce results of comparable quality when de-

tached from their involvement in complex multi-organ segmentation systems.

Regarding the challenge results for the liver and spleen segmentation, lower overall dice scores were reached for the left kidney by all competitors, due to the shown outliers from Sect. 6.4.1.5, Fig. 30. The first case - an extraordinary kidney positioning from breath-hold disruption - was expectedly not covered by the trained appearance model of the proposed method. This is because the modeled neighborhood of the kidney was fundamentally altered, from an expected adjacency of visceral cavity, spleen and bowel, to an adjacency of spine, diaphragm and lung. The same problem also applied to the other approaches due to their data-driven nature, although some of them managed to reach better results. For all approaches in the challenge, a representation of such cases in the training data can be regarded as a prerequisite for their successful segmentation. The second case - an atrophic kidney - was indiscernible in the underlying image, hence it was not segmented by any of the challenge competitors. Apart from these cases, the resected liver in the challenge dataset was only handled with partial success across the challenge competitors.

Finally in comparison with the voxel-wise results from the atlas-based methods, the more well-defined surface delineations of the proposed method can promote a higher plausibility of results in the clinic. However, if more of the inner details of the segmented organs are required, additional measures will have to be taken, e.g. for liver vessel and for renal cavity segmentation.

Regarding an application to other anatomical structures, adaptations to the surface and shape modeling components of the proposed method are expected to be required in particular cases. For example, the segmentation of tubular structures can be of interest in different medical applications. Examples are the aorta, the inferior vena cava, the esophagus and the rectum. It was shown earlier in the literature that decomposition approaches are much better suited for shape model construction in these cases [98]. Other additions to the shape prior modeling of the proposed method are expected for non-tubular, but highly eccentric shapes, e.g. pancreas and stomach. They are known to cause distortions in the spherical parameterizations, which were used for shape modeling in this thesis.

Non parametric approaches can therefore be preferred. For example, particle-based correspondence optimization techniques have been shown to provide high topological and geometrical flexibility, also for thin structure parts [240]. In [241], a correspondence optimization based on non-rigid mesh registration was shown to be applicable to thin, elongated structures and large intra-class variability like the pancreas. These approaches are also equally well suited

for a modeling of non-genus 0 topologies, like hip-bone or vertebrae. For a modeling of non-linear shape variation of structures like the pancreas, various techniques have been proposed in the literature (cf. Sect. 2.3.3). Kernel-PCA has been proposed as an effective non-linear shape modeling technique [85], which has been shown to provide a suitable means for shape modeling of more complex shape variation [241]. Altogether, the aforementioned techniques for a shape modeling of more arbitrarily shaped geometries explicitly allow a correspondence optimization and a shape prior modeling based on landmarks. Therefore, they also fit the landmark-based appearance modeling and detection techniques that were developed in this thesis.

7.3 Imaging Heterogeneity

The applicability of the proposed method was demonstrated for a variety of imaging conditions. The datasets in the abdominal organ segmentation experiments (Sect. 6.2 - 6.4.1) largely differed in slice distance and in-plane resolution, contrast, and in the displayed field of view. The datasets in the left ventricle and lung segmentation experiments introduced variability in terms of noise, occlusion, field of view, image artifacts and intensity distribution, due to the employed modalities and reconstruction kernels of the data. Furthermore, the effectiveness of the proposed Haar-like feature normalization was successfully shown for liver segmentation in T2 MRI.

The shown results reconfirm the known robustness of regression forest based methods in medical imaging applications. Nevertheless, without further development, previous regression-based approaches have often put focus on a single domain, i.e. bone segmentation in x-ray images [135], [136] or CT images [130], [239]. Successful application for more heterogeneous data from dual energy x-ray absorptiometry was e.g. proposed in [168].

Limitations of the proposed techniques are expected for larger changes of intensity distribution that are not reflected during the previous model training, e.g. for use of contrast agent in part of the data, or in case of changing acquisition protocols. In general, an application can be desirable in cases where encountered imaging conditions are unknown in advance, e.g. in large datasets from mixed image acquisition. To this end, additional research will be required, and techniques for transfer learning, for optimized model training and for an on-line model selection will be of particular interest.

Recent developments have e.g. focused on design of high-level features that are particularly robust against strong local heterogeneities like metal artifacts

[242]. The recent proposition regarding model transfer learning from [243] has shown that a regression forest based shape model matching can be adapted online to different data annotation styles. These techniques show great promise for the goal of an automatic analysis of heterogeneous image data using 3D-SSM like the proposed method in the future.

7.4 Anatomical Variability

The dependency of model- and learning-based segmentation methods from a representative quantity of training data is a common problem. For 3D-SSM, which incorporate both modeling and learning paradigms, the handling of scarce anatomical aberrations can become a particular limitation. In addition to the disparity from information learned in the previous appearance model training, the explicit shape constraint from the underlying shape prior can limit the flexibility of 3D-SSM to cope with unexpected cases.

Several pathologies and anatomical aberrations were encountered in the images during the experiments of this thesis. They comprise visible lesions of variable size in liver and kidney, lung fibrosis, liver resection, kidney atrophy, elevated hemidiaphragm and attempted breath-hold.

Overall, the proposed method has demonstrated a high robustness against pathologies that mainly affect the appearance of the modeled target organs (i.e. liver and kidney lesions, lung fibrosis). There, the pathological variabilities of the underlying image intensities were successfully learned during the proposed appearance model training. In particular, the encountered cases of lung fibrosis in the segmentation experiments from Sect. 6.4.4 are good examples for a strong alteration of overall organ appearance by a pathology. Also in many cases, no clear separation of the lung from its adjacent tissue was provided. Here, the incorporated shape and appearance priors of the proposed 3D-SSM have shown to be beneficial for a separation of object and background.

The proposed method has also succeeded in cases where organs other than the actual organ of interest were unexpectedly missing. The atrophic kidney from Sect. 6.4.1 represents an example that was encountered in this thesis.

According to the aforementioned problems of explicit shape prior in cases of unexpected anatomical aberrations, difficulties were encountered by the proposed method for pathologies that affect more than tissue appearance. Encountered examples were the liver resection (cf. Sect. 6.4.1), kidney displacement (cf. Sect. 6.4.1), elevated hemidiaphgram (cf. Sect. 6.4.4), and kidney atrophy (cf. Sect. 6.4.1). In these abnormal cases, search for a meaningful

shape lead to a generally worse solution than the compared multi-atlas labeling approaches. Also, non-linear shape variation of pathological organs is known to cause problems for typical shape prior modeling techniques. This usually adds to the problem of under-representation of pathological cases in datasets used for model training.

Different strategies have recently been proposed to address these problems. For example, deformation of a surface model was proposed in [244], to cover pathological shapes at relevant locations identified by previous landmark detection. Another approach was proposed in [245], where a low rank sparse decomposition technique [246] was proposed to recover shape and appearance prior from the presence of non-linear gross errors. Promising results were achieved by these approaches. Their focus on landmarking and on shape prior models also fit the landmark-based techniques of the proposed method. Generally, modeling multi-modal distributions of shape can be a reasonable means for tackling pathological shape variation (cf. Sect. 2.3.3).

However, with the growing interest in an automatic analysis of clinical image data, the diversity of encountered pathologies and imaging modalities is expected to increase drastically. With growing size and heterogeneity of the data, a visual inspection of segmentations becomes unfeasible, and it will at some point be necessary to assess uncertainties of segmentations produced by 3D-SSM. Propositions have been made to assess goodness of fit based on the confidence of trained appearance models [247]–[250]. Random regression forest based approaches like the proposed method store basic information on sample distribution within the leaf nodes of the trained forests. They can cast this information to the image space in various ways (cf. Sect. 5.1.1), which similarly allows to asses landmark detection confidence during and after segmentation.

7.5 Runtime and Robustness

In the conducted runtime and performance experiments, a significant speedup of the random regression forest training without performance loss was reached for the proposed method by the tested measures, with a total speedup of ~ 17 hours ($\sim 53\%$ of 32 hours total). The trained forests have shown high stability of performance during the reduction of training samples. The choice of extremely randomized trees as a faster alternative has been shown to deliver equivalent results like their default counterpart. Furthermore, extremely randomized trees have shown a generally higher robustness against an influence from sample reduction during training. Both measures are recommendable for

a reduction of training times in the shown scenario. Extremely randomized trees have not been examined by other related regression-based approaches [239], [251], and their use might be beneficial in various scenarios.

Fast convergence with a high precision of results was observed for both forest types within a large range of training samples, patch sizes and tree counts. Convergence was quickly reached for tree counts between 4 and 8 trees, depending on the observed setting of LOD, sample count and forest type. All tests beyond this point revealed no further improvement for up to 12 trees. Onset of performance loss for suboptimal parameters was always observed simultaneously for segmentation accuracy, for tree precision and for convergence of forest results. Generally, the observed precision of results produced by the different ensemble trees seemed to be a strong indicator for a possible loss of segmentation accuracy. For the proposed multi-stage model fitting, this means that for a loss of convergence, dispersion of model fitting results is induced by the regression forests. The final profile-based stage of the proposed method is known to be sensitive against these effects, which is then immediately reflected by the shown results. In case of convergence, high precision of results from the regression forests provides the means for accuracy in the final profile-based adaptation.

Similar tree counts were reported to produce best results for related approaches [239], [251]. Depending on the examined settings of the presented scenarios, no more than 10-15 trees were required for optimal results. Whereas the experiments in this thesis were aimed at runtime and robustness of the model training, a more extensive performance evaluation was presented for face and hand landmark detection in [251]. It was found that as a general rule, segmentation accuracy of the presented scenarios improved up to a certain point with a growing number of trees, of considered split features, of tree depth, and with an allowed smaller number of samples in each leaf.

Furthermore in [251], it was reported that voting information from more distributed votes could compensate for a small tree count, which was not observed for applications of the proposed method. In comparison with the overall appearance of hand bones and facial images, a more heterogeneous neighborhood of organs was encountered in the examined segmentation scenarios of the proposed method, which made votes from the direct landmark neighborhood preferable over more widely distributed votes. Apart from the above, further implications from the presented scenarios were rather difficult to obtain, due to the underlying differences of model fitting schemes, target structures and imaging modalities.

Altogether, the conducted sensitivity analysis has reconfirmed the known robustness of the employed random forests for the proposed method. This is of significance for the intended high generalizability of the proposed standalone 3D-SSM. The latter strongly relies on the straightforward translation of the landmark detector training to varying use-cases. In practice, the presented findings mean that within a wide range of reasonably chosen forest parameters, a constant, high performance of the proposed method is to be expected. This greatly reduces the required optimization effort prior to an application of the proposed method, which significantly facilitates translation to different segmentation scenarios. Finally, much shorter training times can be reached with little risk of performance loss, which additionally promotes the applicability of the proposed method.

During application, a segmentation runtime was reached by the proposed method that allows an inspection of results within minutes and a fully automated background application on larger datasets within days. For the segmentation of both lungs in the 418 CT volumes presented in Sect. 6.4.4, 3 days were required on an Intel Core i7 Hexacore CPU (multi-threaded).

This makes the proposed method generally well suited for a retrospective analysis of clinical data. The annotation of images from a rather big data context becomes similarly feasible if additional computational power is employed. Also, in a server-based application of the proposed method, annotation times of less than 3 days were reached for 4 abdominal organs in $\sim$ 20000 clinical CT volumes.

7.6 Implications for Model-based Image Segmentation

The shown results of the proposed method have a general impact on the widespread availability of 3D-SSM as a model-based segmentation methodology. 3D-SSM are involved in a large variety of medical scenarios, where they are used for image segmentation and where higher-level modeling information is incorporated for image analysis tasks. During segmentation, 3D-SSM are generally used in conjunction with additional, auxiliary techniques in order to be successfully applicable and to promote good results. Sometimes, 3D-SSM are only used as a helper tool for the subsequent refinement by other techniques, like for the final graph cut algorithm in [154]. In other cases, only the robust shape prior is utilized as a part of other methods to perform a segmentation

[142], [143]. Or, simply a surface model is employed to carry out the segmentation, based on the results from previously applied techniques [130].

Summarized, 3D-SSM and their components are often found as a part of more complex systems, and they are rarely used as a self sufficient technique during segmentation. With the facilitated use of a more proficient 3D-SSM as described above, the necessity for auxiliary methods can be removed that are otherwise needed to compensate the limitations of previous 3D-SSM.

Altogether, this helps to promote the availability of higher-level information modeling from 3D-SSM to different use-cases, where each new segmentation can immediately be put in context with previously analyzed object classes, thus linking higher-level modeling information with low-level image information. Different kinds of higher-level information have been proposed for use in medical applications. They also define current directions of research for statistical shape models.

For example, various works have addressed the joint modeling of image structures, to capture information on shape and spatial interrelation of several objects, using medial models [252]–[255] or conditional models [256], [257]. Information on object interrelation can be used for robust, simultaneous segmentation of objects [258]–[260] or for an analysis of object shape and pose interrelation [261]. In [257], object sub-composition has been proposed as another domain for modeling shape interrelation using conditional shape modeling.

Another use of higher-level modeling information is focused on motion modeling. Information on organ motion is learned an encoded in temporal models. They can be used for model-based registration of 3D+t image data, or to increase robustness of model-based segmentation in presence of heavy motion artifacts. Furthermore, pathologies can be revealed from patient-specific motion patterns. Exemplary fields of application involve respiratory motion [262], respiration induced liver motion [263], or cardiac motion [264], [265].

Apart from facilitating use of 3D-SSM in applications like the aforementioned scenarios, the proposed concept of unambiguous landmark detection in medical images also has certain additional benefits. For example, an omnidirectional detection can deliver candidates during segmentation that are more meaningful for an adaptation of temporal modeling techniques. Whereas solutions from typical previous 3D-SSM can again be strongly dependent from the visibility of landmark positions during the unidirectional search. This generally applies to any scenario where search results also have an impact on higher-level interpretation of the final model fitting outcome.

7.7 Contributions of the Proposed Method

The proposed techniques of context appearance modeling and of landmark-wise omni-directional detection have significantly improved the intrinsic capabilities of 3D-SSM. By the use of these techniques, the proposed 3D-SSM has displayed a set of new characteristics that differ from previous 3D-SSM. The high robustness against model placements and the high reproducibility of results lead to a predictable and intuitive model fitting for 3D-SSM. Furthermore, no model initialization is required for an application of the proposed method. This makes segmentation with 3D-SSM much easier in contrast to the unpredictable and error-prone nature of typical previous 3D-SSM. Also, good results can be obtained more easily.

The proposed method complements the previous approaches from [29], [135]. They are the first who have introduced the use of regression voting to the domain of model-based medical image segmentation. In the meantime, they have tuned their method into a fast and highly accurate expert system for bone segmentation in x-ray images and for facial detection. The proposed use of the constrained local model (CLM) allows a joint detection of landmarks, and the presented 2D image segmentation tasks fit the degree of freedom from the CLM model constraint well.

Unlike the work from [29], [135], the development of the proposed method was directly aimed at the well-known problems of 3D-SSM that particularly occur during soft tissue segmentation in volumetric image data, i.e. sensitivity from model pose and from model initialization, limited capture range and sensitivity against locally optimal solutions during segmentation. In contrast to the 2D femur and facial segmentation problems from [29], [135], the examined scenarios of the proposed method are characterized by higher shape variability, and by lower contrast of the considered soft tissue organs and their neighborhood, and furthermore, by a significantly larger size and variability of the shown field of view in the images. These conditions strongly promote the presence of the aforementioned problems for 3D-SSM. The higher modeling complexity in 3D domain makes computational performance a stronger issue. And finally, the problem of high shape variability of 3D soft tissue organs makes a free surface deformation from the rigid shape constraint much more challenging.

Altogether, this hampers an immediate application of the proposed CLM in 3D domain. Actually, in the joint shape and appearance optimization of the CLM, detection results and shape constraint are strongly intertwined. This can impede a separate and unambiguous detection of landmark positions as a

solution for the described problems. The problems were however well addressed by means of the separate multi-stage landmark detectors and the profile-based deformable model fitting of the proposed method.

In addition to the proposed method, the regression-based methods from [239] and from [136] were developed for segmentation in 3D domain. In contrast to [239], the proposed method builds on a 3D-SSM, whereas no statistical shape model was involved in their approach. Therefore, smoothing and remeshing steps were required during segmentation, and in contrast with a 3D-SSM based segmentation, no information from prior modeling is available for further use and deduction after segmentation. More importantly, the method from [239] builds on a detection of nearest boundaries instead of the omni-directional landmark detection by the proposed method. As a result, unambiguous results are not guaranteed, and after the first rough positioning of the surface model at the organ of interest, proper visibility of boundaries had first to be established through an extra rotational optimization step of the model fitting. Unlike the method from [239], not only an application without model initialization was shown for the proposed method, but also high robustness against pose changes during segmentation, and a high reproducibility of results.

Finally, a global optimization of regression votes across different patches was introduced for 2D bone segmentation in [136]. Optimality was reached via convex optimization of the regression-based cost functions. Although the global optimization scheme was promising for the landmark detection task in 3D-SSM, problem of high computational complexity was reported for 3D domain in a preliminary study on CT data. Regression-based landmark detection was proposed in subsequent studies for detection of intervertebral disks in MRI data [266] and for a landmark-based initialization of probabilistic atlases for hip joint segmentation in CT data [267]. However, no time-efficient solution of the original regression approach for 3D-SSM was proposed, and the use of classifiers, of probabilistic atlases and of a previous 3D-SSM was preferred in the presented applications.

CHAPTER 8

Conclusions

The proposed techniques of non-local appearance modeling and of landmark-wise regression voting have notably extended the intrinsic segmentation capabilities of previous 3D Statistical Shape Models (3D-SSM).

This essentially concerns the well-known model initialization problem of 3D-SSM, which is effectively addressed by the proposed method. It reaches a high robustness of results against positional changes prior to the segmentation, within a large capture range around the organs of interest. With the proposed method, 3D-SSM obtain the ability to perform image segmentation autonomously from arbitrary image positions, without need for a previous model initialization. Generalizability of the proposed method across different use-cases was supported by the robustness and flexibility of the trained random regression forests. Also, the presented model fitting without need for previous initialization significantly facilitates an application of 3D-SSM in medical image segmentation tasks. From the state-of-the-art, the proposed method is the first 3D-SSM that has demonstrated a combination of these benefits.

The proposed landmark detectors have also shown to promote a higher segmentation accuracy in comparison with the previous unidirectional search. With its combined accuracy and generalizability, the proposed standalone 3D-SSM has successfully extended the capabilities of 3D-SSM when used as a generic segmentation methodology, and higher segmentation accuracy can more easily be achieved with the proposed 3D-SSM. In the presented cases,

similar performance was reached like other, more specialized segmentation approaches.

The robustness of the proposed method was demonstrated in a variety of clinical segmentation tasks under diverse conditions. Without modifications, the proposed method was successfully applied to different anatomical structures, to changing imaging conditions, and to different pathologies of varying severity. Similar results were produced in comparison with other state-of-the-art approaches on the same datasets and from the literature. Seamless and time efficient background application on a larger clinical dataset, and an extraction of meaningful biomarkers from segmentations of the proposed method were demonstrated.

Finally, effective measures for a reduction of training times without performance loss were proposed and validated experimentally. In addition, the robustness of the employed random regression forests was shown for a wide range of training parameters. Both findings underline the generalizability of the employed regressors in terms of tuning-free straightforward application and of upscaled training.

The translation of computational methods to different clinical scenarios often requires considerable effort and modifications of underlying algorithms prior to a successful application. Often, these adaptations arise from the basic limitations of the used algorithms, like the problems for 3D-SSM that were addressed in this thesis. The shown solutions for these problems allow a more easily achievable performance of 3D-SSM in different scenarios, and also promote the availability of higher-level modeling information from 3D-SSM for clinical use-cases. Altogether, the shown benefits can be considered a significant step towards a more widespread translation of 3D-SSM for image analysis tasks in the clinic.

Future work will focus on incorporating the aforementioned higher-level modeling information into the landmark detection framework of the proposed 3D-SSM. The omni-directional search can offer more meaningful landmark positions for scenarios where higher-level information is based on landmark correspondences, e.g. in applications for shape and motion analysis. To fully exploit this benefit of a meaningful landmark extraction using 3D-SSM, new model deformation schemes will be required where constraints for omni-directional landmark shifts can be imposed during the final stages of the model fitting. Also, a combination of deep learning and model-based image processing promises to be beneficial for both fields of research. Popular deep learning approaches in medical image analysis, like convolutional neural networks and generative

adversarial networks, generally refrain from explicit, model-based shape constraints on image outputs. However, they can be expected to strongly benefit from a pronounced modeling of shape priors, in terms of more accurate and geometrically more meaningful results. Overall, deep learning promises high speed and accuracy for image processing tasks, and it also allows an elegant synthesis of different sources of information during training and application. This naturally suits the original modeling paradigm of 3D-SSM, and thus further encourages an incorporation of deep learning for training and application of future model-based approaches.

Bibliography

[1] M. Sonka and J. M. Fitzpatrick, "Handbook of medical imaging", in *Medical image processing and analysis*, SPIE - The international society for optical engineering, vol. 2, 2000 (Cited on page 9).

[2] T. Heimann and H.-P. Meinzer, "Statistical shape models for 3D medical image segmentation: a review", *Medical Image Analysis*, vol. 13, no. 4, pp. 543–563, 2009 (Cited on pages 10, 11, 33, 55, 64, 82 and 83).

[3] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Active shape models - their training and application", *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, 1995 (Cited on pages 12, 16, 22, 42 and 51).

[4] S. M. Pizer, D. S. Fritsch, P. A. Yushkevich, V. E. Johnson, and E. L. Chaney, "Segmentation, registration, and measurement of shape variation via image object shape", *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 851–865, 1999 (Cited on page 12).

[5] S. M. Pizer, P. T. Fletcher, S. Joshi, A. Thall, J. Z. Chen, Y. Fridman, D. S. Fritsch, A. G. Gash, J. M. Glotzer, M. R. Jiroutek, C. Lu, K. E. Muller, G. Tracton, P. Yushkevich, and E. L. Chaney, "Deformable m-reps for 3D medical image segmentation", *International Journal of Computer Vision*, vol. 55, no. 2-3, pp. 85–106, 2003 (Cited on page 12).

[6] P. A. Yushkevich, H. Zhang, and J. C. Gee, "Continuous medial representation for anatomical structures", *IEEE Transactions on Medical Imaging*, vol. 25, no. 12, pp. 1547–1564, 2006 (Cited on page 13).

[7] J. Gaa, L. A. Kahrs, S. Müller, O. Majdani, and T. Ortmaier, "Cochlear shape description and analyzing via medial models", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2015, pp. 9 413 451–6 (Cited on page 13).

[8] L. H. Staib and J. S. Duncan, "Model-based deformable surface finding for medical images", *IEEE Transactions on Medical Imaging*, vol. 15, no. 5, pp. 720–731, 1996 (Cited on page 13).

[9] G. Székely, A. Kelemen, C. Brechbühler, and G. Gerig, "Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models", *Medical Image Analysis*, vol. 1, no. 1, pp. 19–34, 1996 (Cited on page 13).

[10] A. Kelemen, G. Székely, and G. Gerig, "Elastic model-based segmentation of 3-D neuroradiological data sets", *IEEE Transactions on Medical Imaging*, vol. 18, no. 10, pp. 828–839, 1999 (Cited on pages 13 and 21).

[11] M. Styner, I. Oguz, S. Xu, C. Brechbühler, D. Pantazis, J. J. Levitt, M. E. Shenton, and G. Gerig, "Framework for the statistical shape analysis of brain structures using SPHARM-PDM", *The Insight Journal*, no. 1071, p. 242, 2006 (Cited on page 13).

[12] C. Nikou, G. Bueno, F. Heitz, and J.-P. Armspach, "A joint physics-based statistical deformable model for multimodal brain image analysis", *IEEE Transactions on Medical Imaging*, vol. 20, no. 10, pp. 1026–1037, 2001 (Cited on page 13).

[13] A. Matheny and D. B. Goldgof, "The use of three-and four-dimensional surface harmonics for rigid and nonrigid shape recovery and representation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 10, pp. 967–981, 1995 (Cited on page 13).

[14] C. Davatzikos, X. Tao, and D. Shen, "Hierarchical active shape models, using the wavelet transform", *IEEE Transactions on Medical Imaging*, vol. 22, no. 3, pp. 414–423, 2003 (Cited on pages 13 and 25).

[15] D. Nain, S. Haker, A. Bobick, and A. Tannenbaum, "Multiscale 3-D shape representation and segmentation using spherical wavelets", *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 598–618, 2007 (Cited on page 13).

[16] P. Yu, P. E. Grant, Y. Qi, X. Han, F. Ségonne, R. Pienaar, E. Busa, J. Pacheco, N. Makris, R. L. Buckner, *et al.*, "Cortical surface shape analysis based on spherical wavelets", *IEEE Transactions on Medical Imaging*, vol. 26, no. 4, pp. 582–597, 2007 (Cited on page 13).

[17] B. Tsagaan, A. Shimizu, H. Kobatake, and K. Miyakawa, "An automated segmentation method of kidney using statistical information", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2002, pp. 556–563 (Cited on page 13).

[18] P. Golland, W. E. L. Grimson, M. E. Shenton, and R. Kikinis, "Detection and analysis of statistical differences in anatomical shape", *Medical Image Analysis*, vol. 9, no. 1, pp. 69–86, 2005 (Cited on page 14).

[19] S. Osher and J. A. Sethian, "Fronts propagating with curvature-dependent speed: algorithms based on Hamilton-Jacobi formulations", *Journal of Computational Physics*, vol. 79, no. 1, pp. 12–49, 1988 (Cited on page 14).

[20] R. Malladi, J. A. Sethian, and B. C. Vemuri, "Shape modeling with front propagation: A level set approach", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 158–175, 1995 (Cited on page 14).

[21] M. E. Leventon, W. E. L. Grimson, and O. Faugeras, "Statistical shape influence in geodesic active contours", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2000, pp. 316–323 (Cited on page 14).

[22] A. Tsai, A. Yezzi, W. Wells, C. Tempany, D. Tucker, A. Fan, W. E. Grimson, and A. Willsky, "A shape-based approach to the segmentation of medical imagery using level sets", *IEEE Transactions on Medical Imaging*, vol. 22, no. 2, pp. 137–154, 2003 (Cited on page 14).

[23] K. M. Pohl, J. Fisher, M. Shenton, R. W. McCarley, W. E. L. Grimson, R. Kikinis, and W. M. Wells, "Logarithm odds maps for shape representation", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2006, pp. 955–963 (Cited on page 14).

[24] C. Arrieta, C. Sing-Long, S. Uribe, M. Andia, P. Irarrazaval, and C. Tejos, "Level set segmentation with shape prior knowledge using intrinsic rotation, translation and scaling alignment", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 1568–1571 (Cited on page 14).

[25] D. Cremers, S. J. Osher, and S. Soatto, "Kernel density estimation and intrinsic alignment for shape priors in level set segmentation", *International Journal of Computer Vision*, vol. 69, no. 3, pp. 335–351, 2006 (Cited on page 14).

[26] D. Cremers, M. Rousson, and R. Deriche, "A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape", *International Journal of Computer Vision*, vol. 72, no. 2, pp. 195–215, 2007 (Cited on page 14).

[27] C. Goodall, "Procrustes methods in the statistical analysis of shape", *Journal of the Royal Statistical Society, Series B (Method.)*, pp. 285–339, 1991 (Cited on page 15).

[28] J. C. Gower, "Generalized procrustes analysis", *Psychometrika*, vol. 40, no. 1, pp. 33–51, 1975 (Cited on pages 15 and 55).

[29] T. F. Cootes, M. C. Ionita, C. Lindner, and P. Sauer, "Robust and accurate shape model fitting using random forest regression voting", in *European Conference on Computer Vision (ECCV)*, Springer, 2012, pp. 278–291 (Cited on pages 16, 35, 42, 46, 64, 72, 73, 74, 75, 80, 81, 82, 83, 84, 85, 168 and 180).

[30] R. Larsen and H. Eiriksson, "Robust and resistant 2D shape alignment", Tech. Rep., 2001 (Cited on page 16).

[31] A. Ericsson and J. Karlsson, "Aligning shapes by minimising the description length", in *Scandinavian Conference on Image Analysis*, Springer, 2005, pp. 3–15 (Cited on page 16).

[32] F. Bernard, J. Thunberg, P. Gemmar, F. Hertel, A. Husch, and J. Goncalves, "A solution for multi-alignment by transformation synchronisation", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2015, pp. 2161–2169 (Cited on page 16).

[33] P. Heinze, D. Meister, R. Kober, J. Raczkowsky, and H Worn, "Atlas-based segmentation of pathological knee joints", *Studies in Health Technology and Informatics*, pp. 198–203, 2002 (Cited on page 18).

[34] K. Josephson, A. Ericsson, and J. Karlsson, "Segmentation of medical images using three-dimensional active shape models", in *Scandinavian Conference on Image Analysis*, Springer, 2005, pp. 719–728 (Cited on page 18).

[35] F. Vos, P. W. de Bruin, J. Aubel, G. J. Streekstra, M. Maas, L. J. van Vliet, and A. M. Vossepoel, "A statistical shape model without using landmarks", in *International Conference on Pattern Recognition (ICPR)*, vol. 3, 2004, pp. 714–717 (Cited on page 18).

[36] P. J. Besl and N. D. McKay, "Method for registration of 3-D shapes", in *Robotics-DL tentative*, International Society for Optics and Photonics, 1992, pp. 586–606 (Cited on page 18).

[37] A. Rangarajan, H. Chui, and F. Bookstein, "The softassign procrustes matching algorithm", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 1997, pp. 29–42 (Cited on page 18).

[38] A. D. Brett and C. J. Taylor, "A method of automated landmark generation for automated 3D PDM construction", *Image and Vision Computing*, vol. 18, no. 9, pp. 739–748, 2000 (Cited on page 18).

[39] G. Subsol, J.-P. Thirion, and N. Ayache, "A scheme for automatically building three-dimensional morphometric anatomical atlases: application to a skull atlas", *Medical Image Analysis*, vol. 2, no. 1, pp. 37–60, 1998 (Cited on pages 18 and 19).

[40] A. D. Brett and C. J. Taylor, "Automated construction of 3D shape models using harmonic maps", in *Medical Image Understanding and Analysis*, London, 2000, pp. 175–78 (Cited on pages 18 and 21).

[41] A. Caunce and C. J. Taylor, "Building 3D sulcal models using local geometry", *Medical Image Analysis*, vol. 5, no. 1, pp. 69–80, 2001 (Cited on page 18).

[42] M. Fleute, S. Lavallée, and R. Julliard, "Incorporating a statistically based shape model into a system for computer-assisted anterior cruciate ligament surgery", *Medical Image Analysis*, vol. 3, no. 3, pp. 209–222, 1999 (Cited on page 19).

[43] C. R. Shelton, "Morphable surface models", *International Journal of Computer Vision*, vol. 38, no. 1, pp. 75–91, 2000 (Cited on page 19).

[44] Y. Wang, B. S. Peterson, and L. H. Staib, "3D brain surface matching based on geodesics and local geometry", *Computer Vision and Image Understanding*, vol. 89, no. 2, pp. 252–271, 2003 (Cited on page 19).

[45] L. Ferrarini, H. Olofsen, W. M. Palm, M. A. Van Buchem, J. H. Reiber, and F. Admiraal-Behloul, "GAMEs: growing and adaptive meshes for fully automatic shape modeling and analysis", *Medical Image Analysis*, vol. 11, no. 3, pp. 302–314, 2007 (Cited on page 19).

[46] A. Pitiot, H. Delingette, and P. M. Thompson, "Learning shape correspondence for nD curves", *International Journal of Computer Vision*, vol. 71, no. 1, pp. 71–88, 2007 (Cited on page 19).

[47] C. Lorenz and N. Krahnstöver, "Generation of point-based 3D statistical shape models for anatomical objects", *Computer Vision and Image Understanding*, vol. 77, no. 2, pp. 175–191, 2000 (Cited on page 19).

[48] T. J. Hutton, B. Buxton, and P. Hammond, "Dense surface point distribution models of the human face", in *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis (MMBIA)*, 2001, pp. 153–160 (Cited on page 19).

[49] K. B. Hilger, R. R. Paulsen, and R. Larsen, "Markov random field restoration of point correspondences for active shape modeling", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2004, pp. 1862–1869 (Cited on page 19).

[50] R. Paulsen, R. Larsen, C. Nielsen, S. Laugesen, and B. Ersboll, "Building and testing a statistical shape model of the human ear canal", in *International Conference on International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2002, pp. 373–380 (Cited on page 19).

[51] R. R. Paulsen and K. B. Hilger, "Shape modelling using markov random field restoration of point correspondences", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2003, pp. 1–12 (Cited on page 19).

[52] D. Shen and C. Davatzikos, "Adaptive-focus statistical shape model for segmentation of 3D MR structures", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2000, pp. 23–39 (Cited on page 19).

[53] M. R. Kaus, V. Pekar, C. Lorenz, R. Truyen, S. Lobregt, and J. Weese, "Automated 3-D PDM construction from segmented images using deformable models", *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 1005–1013, 2003 (Cited on page 19).

[54] Y. Shang and O. Dössel, "Statistical 3D shape-model guided segmentation of cardiac images", in *Computers in Cardiology*, 2004, pp. 553–556 (Cited on page 19).

[55] Z. Zhao, S. R. Aylward, and E. K. Teoh, "A novel 3D partitioned active shape model for segmentation of brain MR images", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2005, pp. 221–228 (Cited on pages 19 and 25).

[56] M. Fleute, S. Lavallée, and L. Desbat, "Integrated approach for matching statistical shape models with intra-operative 2D and 3D data", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2002, pp. 364–372 (Cited on page 19).

[57] G. Heitz, T. Rohlfing, and C. R. Maurer Jr, "Statistical shape model generation using nonrigid deformation of a template mesh", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2005, pp. 1411–1421 (Cited on page 19).

[58] A. F. Frangi, W. J. Niessen, D. Rueckert, and J. A. Schnabel, "Automatic 3D ASM construction via atlas-based landmarking and volumetric elastic registration", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2001, pp. 78–91 (Cited on page 19).

[59] A. F. Frangi, D. Rueckert, J. A. Schnabel, and W. J. Niessen, "Automatic construction of multiple-object three-dimensional statistical shape models: Application to cardiac modeling", *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1151–1166, 2002 (Cited on pages 19 and 20).

[60] S. Ordas, L. Boisrobert, M. Bossa, M. Huguet, M. Laucelli, S. Olmos, and A. Frangi, "Grid-enabled automatic construction of a two-chamber cardiac PDM from a large database of dynamic 3D shapes", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2004, pp. 416–419 (Cited on pages 19 and 20).

[61] H. Park, P. H. Bland, and C. R. Meyer, "Construction of an abdominal probabilistic atlas and its application in segmentation", *IEEE Transactions on Medical Imaging*, vol. 22, no. 4, pp. 483–492, 2003 (Cited on page 19).

[62] D. Rueckert, A. F. Frangi, and J. A. Schnabel, "Automatic construction of 3-D statistical deformation models of the brain using nonrigid registration", *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 1014–1025, 2003 (Cited on pages 19 and 20).

[63] A. D. Brett and C. J. Taylor, "Construction of 3D shape models of femoral articular cartilage using harmonic maps", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2000, pp. 41–92 (Cited on page 21).

[64] P. Thompson, C. Schwartz, and A. Toga, "High-resolution random mesh algorithms for creating a probabilistic 3D surface atlas of the human brain", *NeuroImage*, vol. 3, no. 1, pp. 19–34, 1996 (Cited on page 21).

[65] H. Lamecker, T. Lange, and M. Seebass, "A statistical shape model for the liver", in *International Conference on International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2002, pp. 421–427 (Cited on page 21).

[66] E. Praun, W. Sweldens, and P. Schröder, "Consistent mesh parameterizations", in *Conference on Computer Graphics and Interactive Techniques*, ACM, 2001, pp. 179–184 (Cited on page 21).

[67] P. M. Thompson and A. W. Toga, "Detection, visualization and animation of abnormal anatomic structure with a deformable probabilistic brain atlas based on random vector field transformations", *Medical Image Analysis*, vol. 1, no. 4, pp. 271–294, 1997 (Cited on page 21).

[68] D. Meier and E. Fisher, "Parameter space warping: shape-based correspondence between morphologically different objects", *IEEE Transactions on Medical Imaging*, vol. 21, no. 1, pp. 31–47, 2002 (Cited on page 21).

[69] Y. Wang, M.-C. Chiang, and P. M. Thompson, "Automated surface matching using mutual information applied to Riemann surface structures", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2005, pp. 666–674 (Cited on page 21).

[70] A. C. Kotcheff and C. J. Taylor, "Automatic construction of eigenshape models by direct optimization", *Medical Image Analysis*, vol. 2, no. 4, pp. 303–314, 1998 (Cited on pages 21 and 22).

[71] R. H. Davies, C. J. Twining, T. F. Cootes, J. C. Waterton, and C. J. Taylor, "A minimum description length approach to statistical shape modeling", *IEEE Transactions on Medical Imaging*, vol. 21, no. 5, pp. 525–537, 2002 (Cited on pages 21 and 22).

[72] H. H. Thodberg, "Minimum description length shape and appearance models", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2003, pp. 51–62 (Cited on page 22).

[73] J. Cates, P. Fletcher, M. Styner, M. Shenton, and R. Whitaker, "Shape modeling and analysis with entropy-based particle systems", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2007, pp. 333–345 (Cited on page 22).

[74] P. Horkaew and G.-Z. Yang, "Optimal deformable surface models for 3D Medical Image Analysis", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2003, pp. 13–24 (Cited on page 22).

[75] C. J. Twining, T. Cootes, S. Marsland, V. Petrovic, R. Schestowitz, and C. J. Taylor, "A unified information-theoretic approach to groupwise non-rigid registration and model building", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2005, pp. 1–14 (Cited on page 22).

[76] R. H. Davies, C. J. Twining, P. D. Allen, T. F. Cootes, and C. J. Taylor, "Building optimal 2D statistical shape models", *Image and Vision Computing*, vol. 21, no. 13, pp. 1171–1182, 2003 (Cited on page 22).

[77] R. Davies, C. Twining, T. Cootes, J. Waterton, and C. Taylor, "3D statistical shape models using direct optimisation of description length", *European Conference on Computer Vision (ECCV)*, pp. 1–17, 2002 (Cited on page 22).

[78] T. Heimann, I. Wolf, T. Williams, and H.-P. Meinzer, "3D Active shape models using gradient descent optimization of description length", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, vol. 3565, 2005, pp. 566–577 (Cited on pages 22 and 54).

[79] F. De La Torre and M. J. Black, "A framework for robust subspace learning", *International Journal of Computer Vision*, vol. 54, no. 1, pp. 117–142, 2003 (Cited on page 23).

[80] T. F. Cootes and C. J. Taylor, "A mixture model for representing shape variation", *Image and Vision Computing*, vol. 17, no. 8, pp. 567–573, 1999 (Cited on page 23).

[81] T. Heap and D. C. Hogg, "Improving Specificity in PDMs using a Hierarchical Approach.", in *British Machine Vision Conference (BMVC)*, vol. 1, 1997, pp. 80–89 (Cited on page 23).

[82] Y. Li and W. Ito, "Shape parameter optimization for adaboosted active shape model", in *IEEE International Conference on Computer Vision (ICCV)*, vol. 1, 2005, pp. 251–258 (Cited on pages 23 and 29).

[83] P. D. Sozou, T. F. Cootes, C. J. Taylor, and E. Di Mauro, "Non-linear generalization of point distribution models using polynomial regression", *Image and Vision Computing*, vol. 13, no. 5, pp. 451–457, 1995 (Cited on page 23).

[84] P. D. Sozou, T. F. Cootes, C. J. Taylor, E. Di Mauro, and A. Lanitis, "Non-linear point distribution modelling using a multi-layer perceptron", *Image and Vision Computing*, vol. 15, no. 6, pp. 457–463, 1997 (Cited on page 23).

[85] C. J. Twining and C. J. Taylor, "The use of kernel principal component analysis to model data distributions", *Pattern Recognition*, vol. 36, no. 1, pp. 217–227, 2003 (Cited on pages 23 and 174).

[86] M. B. Stegmann, K. Sjöstrand, and R. Larsen, "Sparse modeling of landmark and texture variability using the orthomax criterion", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2006, 61441G1–12 (Cited on page 24).

[87] K. Sjöstrand, M. B. Stegmann, and R. Larsen, "Sparse principal component analysis in medical shape modeling", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2006, pp. 61444X1–12 (Cited on page 24).

[88] K. B. Hilger, R. Larsen, and M. C. Wrobel, "Growth modeling of human mandibles using non-Euclidean metrics", *Medical Image Analysis*, vol. 7, no. 4, pp. 425–433, 2003 (Cited on page 24).

[89] A. F. Frangi, J. H. Reiber, B. P. Lelieveldt, *et al.*, "Independent component analysis in statistical shape models", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2003, pp. 375–383 (Cited on page 24).

[90] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent component analysis*. John Wiley & Sons, 2004, vol. 46 (Cited on page 24).

[91] M. Üzümcü, A. Frangi, M. Sonka, J. Reiber, and B. Lelieveldt, "ICA vs. PCA active appearance models: Application to cardiac MR segmentation", in *International Conference on International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2003, pp. 451–458 (Cited on page 24).

[92] A. Suinesiaputra, M. Üzümcü, A. Frangi, T. Kaandorp, J. Reiber, and B. Lelieveldt, "Detecting regional abnormal cardiac contraction in short-axis MR images using independent component analysis", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2004, pp. 737–744 (Cited on page 24).

[93] T. F. Cootes and C. J. Taylor, "Combining point distribution models with shape models based on finite element analysis", *Image and Vision Computing*, vol. 13, no. 5, pp. 403–409, 1995 (Cited on page 25).

[94] ——, "Data driven refinement of active shape model search.", in *British Machine Vision Conference (BMVC)*, 1996, pp. 1–10 (Cited on page-25).

[95] Y. Wang and L. H. Staib, "Boundary finding with prior shape and smoothness models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 7, pp. 738–743, 2000 (Cited on page 25).

[96] J. Lötjönen, K. Antila, E Lamminmäki, J. Koikkalainen, M. Lilja, and T Cootes, "Artificial enlargement of a training set for statistical shape models: Application to cardiac images", in *International Workshop on Functional Imaging and Modeling of the Heart*, Springer, 2005, pp. 92–101 (Cited on page 25).

[97] J. Koikkalainen, T. Tolli, K. Lauerma, K. Antila, E. Mattila, M. Lilja, and J. Lotjonen, "Methods of artificial enlargement of the training set for statistical shape models", *IEEE Transactions on Medical Imaging*, vol. 27, no. 11, pp. 1643–1654, 2008 (Cited on page 25).

[98] M. de Bruijne, B. van Ginneken, M. A. Viergever, and W. J. Niessen, "Adapting active shape models for 3D segmentation of tubular structures in medical images", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, vol. 2732, Springer, 2003, pp. 136–147 (Cited on pages 25 and 173).

[99] T. F. Cootes and C. J. Taylor, "Using grey-level models to improve active shape model search", in *International Conference on Pattern Recognition: Computer Vision & Image Processing.*, vol. 1, 1994, pp. 63–67 (Cited on page 28).

[100] G. Behiels, F. Maes, D. Vandermeulen, and P. Suetens, "Evaluation of image features and search strategies for segmentation of bone structures in radiographs using active shape models", *Medical Image Analysis*, vol. 6, no. 1, pp. 47–62, 2002 (Cited on page 28).

[101] J. Haslam, C. J. Taylor, and T. F. Cootes, "A probabilistic fitness measure for deformable template models.", in *British Machine Vision Conference (BMVC)*, 1994, pp. 1–10 (Cited on page 28).

[102] F. Jiao, S. Li, H.-Y. Shum, and D. Schuurmans, "Face alignment using statistical models and wavelet features", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2003, pp. I–I (Cited on page 29).

[103] D. Shen, Y. Zhan, and C. Davatzikos, "Segmentation of prostate boundaries from ultrasound images using statistical shape model", *IEEE Transactions on Medical Imaging*, vol. 22, no. 4, pp. 539–551, 2003 (Cited on page 29).

[104] G. Langs, P. Peloschek, R. Donner, M. Reiter, and H. Bischof, "Active feature models", in *International Conference on Pattern Recognition (ICPR)*, vol. 1, 2006, pp. 417–420 (Cited on pages 29, 35, 65 and 72).

[105] M. de Bruijne, B. van Ginneken, W. J. Niessen, J. A. Maintz, and M. A. Viergever, "Active-shape-model-based segmentation of abdominal aortic aneurysms in CTA images", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2002, pp. 463–474 (Cited on page 29).

[106] S. Li, L. Zhu, and T. Jiang, "Active shape model segmentation using local edge structures and AdaBoost", in *International Workshop on Medical Imaging and Virtual Reality*, Springer, vol. 3150, 2004, pp. 121–128 (Cited on page 29).

[107] J. Bailleul, S. Ruan, D. Bloyet, and B. Romaniuk, "Segmentation of anatomical structures from 3D brain MRI using automatically-built statistical shape models", in *International Conference on Image Processing (ICIP)*, vol. 4, 2004, pp. 2741–2744 (Cited on page 29).

[108] J. Peters, O. Ecabert, and J. Weese, "Feature optimization via simulated search for model-based heart segmentation", in *International Congress Series*, Elsevier, vol. 1281, 2005, pp. 33–38 (Cited on page 29).

[109] H. Lamecker, T. Lange, and M. Seebass, *Segmentation of the liver using a 3D statistical shape model*. Konrad-Zuse-Zentrum für Informationstechnik Berlin, 2004 (Cited on page 29).

[110] D. Kainmüller, T. Lange, and H. Lamecker, "Shape constrained automatic segmentation of the liver based on a heuristic intensity model", in *MICCAI Workshop on 3D Segmentation in the Clinic: A Grand Challenge*, Springer, 2007, pp. 109–116 (Cited on pages 29, 38, 52, 116, 117, 118, 119, 120, 169, 170, 171 and 172).

[111] M. Brejl and M. Sonka, "Object localization and border detection criteria design in edge-based image segmentation: automated learning from examples", *IEEE Transactions on Medical Imaging*, vol. 19, no. 10, pp. 973–985, 2000 (Cited on page 30).

[112] M. R. Kaus, J. Von Berg, J. Weese, W. Niessen, and V. Pekar, "Automated segmentation of the left ventricle in cardiac MRI", *Medical Image Analysis*, vol. 8, no. 3, pp. 245–254, 2004 (Cited on page 30).

[113] J. Stough, P. Pizer, E. L. Chaney, and M. Rao, "Clustering on image boundary regions for deformable model segmentation", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2004, pp. 436–439 (Cited on page 30).

[114] S. Ho and G. Gerig, "Profile scale-spaces for multiscale image match", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2004, pp. 176–183 (Cited on page 30).

[115] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", in *European Conference on Computer Vision (ECCV)*, Springer, 1998, pp. 484–498 (Cited on pages 31 and 41).

[116] T. F. Cootes, G. J. Edwards, and C. J. Taylor, "Active appearance models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 681–685, 2001 (Cited on pages 31 and 41).

[117] J. G. Bosch, S. C. Mitchell, B. P. Lelieveldt, F. Nijland, O. Kamp, M. Sonka, and J. H. Reiber, "Automatic segmentation of echocardiographic sequences by active appearance motion models", *IEEE Transactions on Medical Imaging*, vol. 21, no. 11, pp. 1374–1383, 2002 (Cited on pages 31 and 32).

[118] T. F. Cootes and C. J. Taylor, "On representing edge structure for model matching", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, 2001, pp. 1114–1119 (Cited on page 31).

[119] I. M. Scott, T. F. Cootes, and C. J. Taylor, "Improving appearance model matching using local image structure", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2003, pp. 258–269 (Cited on page 31).

[120] B. V. Ginneken, A. F. Frangi, J. J. Staal, B. M. T. H. Romeny, and M. A. Viergever, "Active shape model segmentation with optimal features", *IEEE Transactions on Medical Imaging*, vol. 21, no. 8, pp. 924–933, 2002 (Cited on page 31).

[121] S. Ordas, L. Boisrobert, M. Huguet, and A. Frangi, "Active shape models with invariant optimal features (IOF-ASM) application to cardiac MRI segmentation", in *Computers in Cardiology*, 2003, pp. 633–636 (Cited on page 32).

[122] Y. Zhan and D. Shen, "Deformable segmentation of 3-D ultrasound prostate images using statistical texture matching method", *IEEE Transactions on Medical Imaging*, vol. 25, no. 3, pp. 256–272, 2006 (Cited on page 32).

[123] R. Larsen, M. B. Stegmann, S. Darkner, S. Forchhammer, T. F. Cootes, and B. K. Ersbøll, "Texture enhanced appearance models", *Computer Vision and Image Understanding*, vol. 106, no. 1, pp. 20–30, 2007 (Cited on page 32).

[124] S. C. Mitchell, J. G. Bosch, B. P. Lelieveldt, R. J. Van der Geest, J. H. Reiber, and M. Sonka, "3-D active appearance models: segmentation of cardiac MR and ultrasound images", *IEEE Transactions on Medical Imaging*, vol. 21, no. 9, pp. 1167–1178, 2002 (Cited on page 32).

[125] R. Beichel, G. Gotschuli, E. Sorantin, F. Leberl, and M. Sonka, "Diaphragm dome surface segmentation in CT data sets: A 3D active appearance model approach", in *SPIE Medical Imaging*, International Society for Optics and Photonics, vol. 4684, 2002, pp. 475–484 (Cited on page 32).

[126] D. Freedman, R. J. Radke, T. Zhang, Y. Jeong, D. M. Lovelock, and G. T. Chen, "Model-based segmentation of medical imagery by matching distributions", *IEEE Transactions on Medical Imaging*, vol. 24, no. 3, pp. 281–292, 2005 (Cited on page 32).

[127] R. E. Broadhurst, J. Stough, S. M. Pizer, and E. L. Chaney, "A statistical appearance model based on intensity quantile histograms", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2006, pp. 422–425 (Cited on page 32).

[128] D. Cristinacce and T. F. Cootes, "Feature detection and tracking with constrained local models.", in *British Machine Vision Conference (BMVC)*, vol. 1, 2006, p. 3 (Cited on page 32).

[129] A. Wimmer, G. Soza, and J. Hornegger, "A generic probabilistic active shape model for organ segmentation", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2009, pp. 26–33 (Cited on pages 33, 39, 116, 117, 118, 120 and 172).

[130] Y. Shao, Y. Gao, Q. Wang, X. Yang, and D. Shen, "Locally-constrained boundary regression for segmentation of prostate and rectum in the planning CT images", *Medical Image Analysis*, vol. 26, no. 1, pp. 345–356, 2015 (Cited on pages 33, 46, 174 and 179).

[131] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE Computer Society, vol. 1, 2001, pp. 511–518 (Cited on pages 34, 66, 67, 68 and 70).

[132] L. Zhang, H. Ai, S. Xin, C. Huang, S. Tsukiji, and S. Lao, "Robust face alignment based on local texture classifiers", in *IEEE International Conference on Image Processing (ICIP)*, vol. 2, 2005, pp. 354–357 (Cited on pages 34, 46, 63, 64 and 65).

[133] S. K. Zhou and D. Comaniciu, "Shape regression machine", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, vol. 4584, 2007, ch. 2, pp. 13–25 (Cited on pages 35, 39, 43, 72 and 170).

[134] S. K. Zhou, "Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram", *Medical Image Analysis*, vol. 14, no. 4, pp. 563–581, 2010 (Cited on pages 35, 39, 43 and 72).

[135] C. Lindner, S Thiagarajah, J. M. Wilkinson, G. A. Wallis, T. F. Cootes, arcOGEN Consortium, *et al.*, "Accurate fully automatic femur segmentation in pelvic radiographs using regression voting", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2012, pp. 353–360 (Cited on pages 35, 39, 42, 46, 64, 72, 73, 74, 80, 82, 83, 84, 85, 168, 174 and 180).

[136] C. Chen, W. Xie, J. Franke, P. Grutzner, L.-P. Nolte, and G. Zheng, "Automatic X-ray landmark detection and shape segmentation via data-driven joint estimation of image displacements", *Medical Image Analysis*, vol. 18, no. 3, pp. 487–499, 2014 (Cited on pages 35, 43, 64, 168, 174 and 181).

[137] R. Cuingnet, R. Prevost, D. Lesage, L. Cohen, B. Mory, and R. Ardon, "Automatic detection and segmentation of kidneys in 3D CT images using random forests", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2012, pp. 66–74 (Cited on pages 35 and 72).

[138] X. Cao, Y. Wei, F. Wen, and J. Sun, "Face alignment by explicit shape regression", *International Journal of Computer Vision*, vol. 107, no. 2, pp. 177–190, 2014 (Cited on pages 35, 36, 43, 65 and 66).

[139] J. Cheng, W. Xiong, Y. Gu, S. C. Chia, and Y. Wang, "Cascaded shape regression for automatic prostate segmentation from extracorporeal ultrasound images", in *Augmented Reality Environments for Medical Imaging and Computer-Assisted Interventions*, Springer, 2013, pp. 65–74 (Cited on pages 35, 43 and 72).

[140] M. Ozuysal, M. Calonder, V. Lepetit, and P. Fua, "Fast keypoint recognition using random ferns", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 448–461, 2010 (Cited on pages-36, 65 and 66).

[141] P. Dollár, P. Welinder, and P. Perona, "Cascaded pose regression", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 1078–1085 (Cited on page 36).

[142] G. Li, X. Chen, F. Shi, W. Zhu, J. Tian, and D. Xiang, "Automatic liver segmentation based on shape constraints and deformable graph cut in CT images", *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5315–5329, 2015 (Cited on pages 38, 116, 117, 119, 120 and 179).

[143] S. D. S. Al-Shaikhli, M. Y. Yang, and B. Rosenhahn, "Automatic 3D liver segmentation using sparse representation of global and local image information via level set formulation", *arXiv preprint 1508.01521*, 2015 (Cited on pages 38, 116, 117, 119, 120 and 179).

[144] M. Erdt and M. Kirschner, "Fast automatic liver segmentation combining learned shape priors with observed shape deviation", in *IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, IEEE, 2010, pp. 249–254 (Cited on pages 38, 42, 45, 116, 117, 118, 120, 170 and 171).

[145] W. Wu, Z. Zhou, S. Wu, and Y. Zhang, "Automatic liver segmentation on volumetric CT images using supervoxel-based graph cuts", *Computational and Mathematical Methods in Medicine*, 2016 (Cited on pages 38, 116, 117 and 120).

[146] D. Ballard, "Generalizing the hough transform to detect arbitrary shapes", *Pattern Recognition*, vol. 13, no. 2, pp. 111–122, 1981 (Cited on page 38).

[147] T. Heimann, S. Münzing, H.-P. Meinzer, and I. Wolf, "A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, vol. 4584, Springer, 2007, ch. 1, pp. 1–12 (Cited on pages 38, 43, 64, 87, 99, 109, 110, 114, 116, 117, 118 and 120).

[148] Y. Zheng, B. Georgescu, and D. Comaniciu, "Marginal space learning for efficient detection of 2D/3D anatomical structures in medical images", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, vol. 5636, Springer, 2009, ch. 34, pp. 411–422 (Cited on pages 40, 44 and 65).

[149] Y. Zheng and D. Comaniciu, "Marginal Space Learning", in *Marginal Space Learning for Medical Image Analysis*. Springer New York, 2014, ch. 2, pp. 25–65 (Cited on pages 40, 44, 65 and 120).

[150] A. Montillo, J. Shotton, J. Winn, J. E. Iglesias, D. Metaxas, and A. Criminisi, "Entangled decision forests and their application for semantic segmentation of CT images", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2011, pp. 184–196 (Cited on page 40).

[151] A Criminisi, J. Shotton, and E Konukoglu, "Decision forests for classification, regression, density estimation, manifold learning and semi-supervised learning", *Microsoft Research Cambridge, Technical Report MSRTR-2011-114*, vol. 5, no. 6, p. 12, 2011 (Cited on pages 40, 46 and 63).

[152] B. Glocker, O. Pauly, E. Konukoglu, and A. Criminisi, "Joint classification-regression forests for spatially structured multi-object segmentation", in *European Conference on Computer Vision (ECCV)*, Springer, 2012, pp. 870–881 (Cited on pages 40 and 46).

[153] R. Donner, B. H. Menze, H. Bischof, and G. Langs, "Global localization of 3D anatomical structures by pre-filtered Hough Forests and discrete optimization", *Medical Image Analysis*, vol. 17, no. 8, pp. 1304–1314, 2013 (Cited on pages 40, 46, 63, 64, 65 and 72).

[154] T. Okada, M. G. Linguraru, M. Hori, R. M. Summers, N. Tomiyama, and Y. Sato, "Abdominal multi-organ segmentation from CT images using conditional shape-location and unsupervised intensity priors", *Medical Image Analysis*, vol. 26, no. 1, pp. 1–18, 2015 (Cited on pages 40, 44, 120 and 178).

[155] T. Heimann, B. van Ginneken, M. Styner, Y. Arzhaeva, V. Aurich, C. Bauer, A. Beck, C. Becker, R. Beichel, G. Bekes, F. Bello, G. Binnig, H. Bischof, A. Bornik, P. Cashman, Y. Chi, A. Cordova, B. Dawant, M. Fidrich, J. Furst, D. Furukawa, L. Grenacher, J. Hornegger, D. Kainmüller, R. Kitney, H. Kobatake, H. Lamecker, T. Lange, J Lee, B. Lennon, R Li, S Li, H.-P. Meinzer, G. Nemeth, D. Raicu, A.-M. Rau, E. van Rikxoort, M. Rousson, L. Rusko, K. Saddi, G. Schmidt, D. Seghers, A. Shimizu, P. Slagmolen, E. Sorantin, G. Soza, R. Susomboon, J. Waite, A. Wimmer, and I. Wolf, "Comparison and evaluation of methods for liver segmentation from CT datasets", *IEEE Transac-*

tions on Medical Imaging, vol. 28, no. 8, pp. 1251–1265, 2009 (Cited on page 42).

[156] D. Kainmüller, H. Lamecker, M. O. Heller, B. Weber, H.-C. Hege, and S. Zachow, "Omnidirectional displacements for deformable surfaces", *Medical Image Analysis*, vol. 17, no. 4, pp. 429–441, 2013 (Cited on pages 43, 45, 157 and 169).

[157] T. Heimann, "Statistical shape models for 3D medical image segmentation", PhD thesis, University of Heidelberg, 2007 (Cited on pages 53, 109 and 110).

[158] W. E. Lorensen and H. E. Cline, "Marching cubes: a high resolution 3D surface construction algorithm", in *ACM Siggraph Computer Graphics*, ACM, vol. 21, 1987, pp. 163–169 (Cited on page 53).

[159] W. J. Schroeder, B. Lorensen, and K. Martin, *The visualization toolkit: an object-oriented approach to 3D graphics*. Kitware, 2004 (Cited on page 53).

[160] S. Valette, J. M. Chassery, and R. Prost, "Generic remeshing of 3D triangular meshes with metric-dependent discrete Voronoi diagrams", *IEEE Transactions on Visualization and Computer Graphics*, vol. 14, no. 2, pp. 369–381, 2008 (Cited on page 53).

[161] M. Erdt, M. Kirschner, and S. Wesarg, "Simultaneous segmentation and correspondence establishment for statistical shape models", in *Modelling the Physiological Human*. Springer, 2009, vol. 5903, ch. 3, pp. 25–35 (Cited on page 55).

[162] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, "Training models of shape from sets of examples", in *British Machine Vision Conference (BMVC)*, Springer, 1992, pp. 9–18 (Cited on pages 55 and 56).

[163] M. Van Stralen, A. Haak, K. Leung, G. van Burken, and J. G. Bosch, "Segmentation of multi-center 3D left ventricular echocardiograms by active appearance models", in *MICCAI Challenge on Echocardiographic 3D Ultrasound Segmentation (CETUS)*, Springer, 2014, pp. 73–80 (Cited on pages 61 and 147).

[164] R. Donner, E. Birngruber, H. Steiner, H. Bischof, and G. Langs, "Localization of 3D anatomical structures using random forests and discrete optimization.", in *MICCAI Workshop on Medical Computer Vision*, Springer, 2010, pp. 86–95 (Cited on page 62).

[165] S. K. Zhou, J. Zhang, and Y. Zheng, "Discriminative learning for anatomical structure detection and segmentation", in *Ensemble Machine Learning*, Springer, 2012, pp. 273–306 (Cited on page 63).

[166] X. Wang, G. Doretto, T. Sebastian, J. Rittscher, and P. Tu, "Shape and appearance context modeling", in *IEEE International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8 (Cited on page 63).

[167] J. Gall and V. Lempitsky, "Class-specific hough forests for object detection", in *Decision forests for computer vision and Medical Image Analysis*, Springer, 2013, pp. 143–157 (Cited on pages 63, 73 and 81).

[168] P. Bromiley, J. Adams, and T. Cootes, "Localisation of vertebrae on DXA images using constrained local models with random forest regression voting", in *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, Springer, 2015, pp. 159–171 (Cited on pages 64 and 174).

[169] A. Criminisi, J. Shotton, D. Robertson, and E. Konukoglu, "Regression forests for efficient anatomy detection and localization in CT studies", in *International MICCAI Workshop on Medical Computer Vision*, Springer, 2010, pp. 106–117 (Cited on pages 65, 72, 75 and 81).

[170] T. Ebner, D. Stern, R. Donner, H. Bischof, and M. Urschler, "Towards automatic bone age estimation from MRI: localization of 3D anatomical landmarks.", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2014, pp. 421–428 (Cited on page 65).

[171] P. Viola and M. J. Jones, "Robust real-time face detection", *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, 2004 (Cited on pages 65, 66 and 69).

[172] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-up robust features (SURF)", *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, 2008 (Cited on page 68).

[173] L. Breiman, "Random forests", *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001 (Cited on pages 71, 72, 75 and 158).

[174] A. Criminisi, J. Shotton, E. Konukoglu, *et al.*, "Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning", *Foundations and Trends in Computer Graphics and Vision*, vol. 7, no. 2–3, pp. 81–227, 2012 (Cited on page 72).

[175] Y. Zhan, X. S. Zhou, Z. Peng, and A. Krishnan, "Active scheduling of organ detection and segmentation in whole-body medical images", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2008, pp. 313–321 (Cited on page 72).

[176] A. Criminisi, J. Shotton, and S. Bucciarelli, "Decision forests with long-range spatial context for organ localization in CT volumes", in *MICCAI Workshop on Probabilistic Models for Medical Image Analysis*, Springer, vol. 1, 2009, pp. 146–158 (Cited on page 72).

[177] A. Montillo and H. Ling, "Age regression from faces using random forests", in *IEEE International Conference on Image Processing (ICIP)*, 2009, pp. 2465–2468 (Cited on page 72).

[178] P. Geurts, D. Ernst, and L. Wehenkel, "Extremely randomized trees", *Machine Learning*, vol. 63, no. 1, pp. 3–42, 2006 (Cited on pages 72, 74 and 158).

[179] J. Gall, N. Razavi, and L. Van Gool, "An introduction to random forests for multi-class object detection", *Outdoor and large-scale real-world scene analysis*, pp. 243–263, 2012 (Cited on page 75).

[180] H. C. van Assen, M. G. Danilouchkine, A. F. Frangi, S. Ordas, J. J. M. Westenberg, J. H. C. Reiber, and B. P. F. Lelieveldt, "Spasm: A 3-D-asm for segmentation of sparse and arbitrary oriented cardiac MRI data", *Medical Image Analysis*, vol. 10, pp. 286–303, 2006 (Cited on pages 85 and 86).

[181] P. J. Burt, "The pyramid as a structure for efficient computation", in *Multiresolution Image Processing and Analysis*, ser. Springer Series in Information Sciences. Springer, 1984, vol. 12, ch. 2, pp. 6–35 (Cited on page 86).

[182] T. Cootes, C. Taylor, and A Lanitis, "Multi-resolution search with active shape models", in *International Conference on Pattern Recognition: Computer Vision & Image Processing.*, vol. 1, 1994, pp. 610–612 (Cited on page 86).

[183] Y. J. Zhang, "A survey on evaluation methods for image segmentation", *Pattern Recognition*, vol. 29, no. 8, pp. 1335–1346, 1996 (Cited on page-90).

[184] W. J. Niessen, C. J. Bouma, K. L. Vincken, and M. A. Viergever, "Error metrics for quantitative evaluation of medical image segmentation", in *Performance Characterization in Computer Vision*, Springer, 2000, pp. 275–284 (Cited on page 91).

[185] G. Gerig, M. Jomier, and M. Chakos, "Valmet: A new validation tool for assessing and improving 3D object segmentation", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2001, pp. 516–523 (Cited on page 91).

[186] B. Van Ginneken, T. Heimann, and M. Styner, "3D segmentation in the clinic: a grand challenge", *3D segmentation in the clinic: a grand challenge*, pp. 7–15, 2007 (Cited on pages 93 and 109).

[187] T. Heimann, I. Wolf, and H.-P. Meinzer, "Active shape models for a fully automated 3D segmentation of the liver–an evaluation on clinical data", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2006, pp. 41–48 (Cited on page 98).

[188] T. Heimann, H.-P. Meinzer, and I. Wolf, "A statistical deformable model for the segmentation of liver CT volumes", *3D Segmentation in the clinic: A grand challenge*, pp. 161–166, 2007 (Cited on pages 99 and 109).

[189] F. Lu, F. Wu, P. Hu, Z. Peng, and D. Kong, "Automatic 3D liver location and segmentation via convolutional neural networks and graph cut", *arXiv preprint 1605.03012*, 2016 (Cited on pages 116, 117, 119 and 120).

[190] M. G. Linguraru, W. J. Richbourg, J. M. Watt, V. Pamulapati, and R. M. Summers, "Liver and tumor segmentation and analysis from CT of diseased patients via a generic affine invariant shape parameterization and graph cuts", in *MICCAI Workshop on Computational and Clinical Challenges in Abdominal Imaging*, Springer, 2011, pp. 198–206 (Cited on pages 116, 117 and 120).

[191] J. Wang, Y. Cheng, C. Guo, Y. Wang, and S. Tamura, "Shape–intensity prior level set combining probabilistic atlas and probability map constrains for automatic liver segmentation from abdominal CT images", *International Journal of Computer Assisted Radiology and Surgery*, pp. 1–10, 2015 (Cited on pages 116, 117 and 120).

[192] V. Gupta, Y. Wang, A. Romero, A. Myronenko, P. Jordan, B. Heijmen, and M Hoogeman, "SU-E-J-208: fast and accurate auto-segmentation of abdominal organs at risk for online adaptive radiotherapy", *Medical Physics*, vol. 41, no. 6, pp. 205–205, 2014 (Cited on page 119).

[193] R. Trullo, C. Petitjean, S. Ruan, B. Dubray, D. Nie, and D. Shen, "Segmentation of organs at risk in thoracic CT images using a sharpmask architecture and conditional random fields", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2017, pp. 1003–1006 (Cited on page 119).

[194] K. Men, J. Dai, and Y. Li, "Automatic segmentation of the clinical target volume and organs at risk in the planning CT for rectal cancer using deep dilated convolutional neural networks", *Medical Physics*, vol. 44, no. 12, pp. 6377–6389, 2017 (Cited on pages 119 and 120).

[195] J. Hutzl, D. Oertel, and H. Wörn, "Knowledge-based direction prediction to optimize the null-space parameter of a redundant robot in a telemanipulation scenario", in *IEEE International Symposium on Robotic and Sensors Environments (ROSE)*, 2014, pp. 25–30 (Cited on page-119).

[196] J. Hutzl, "Parameteranalyse eines redundanten Roboters mit Pivotpunkt-Restriktion", PhD thesis, Karlsruhe Institute of Technology (KIT), 2017 (Cited on page 119).

[197] M. Götz, E. Heim, K. März, T. Norajitra, M. Hafezi, N. Fard, A. Mehrabi, M. Knoll, C. Weber, L. Maier-Hein, *et al.*, "A learning-based, fully automatic liver tumor segmentation pipeline based on sparsely annotated training data", in *SPIE Medical Imaging*, International Society for Optics and Photonics, 2016, pp. 97841I1–6 (Cited on page 119).

[198] P. Lambin, E. Rios-Velazquez, R. Leijenaar, S. Carvalho, R. G. van Stiphout, P. Granton, C. M. Zegers, R. Gillies, R. Boellard, A. Dekker, *et al.*, "Radiomics: extracting more information from medical images using advanced feature analysis", *European Journal of Cancer*, vol. 48, no. 4, pp. 441–446, 2012 (Cited on page 119).

[199] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: a survey", *Medical Image Analysis*, vol. 24, no. 1, pp. 205–219, 2015 (Cited on page 120).

[200] B. Landman, Z. Xu, J. E. Iglesias, M. Styner, T. R. Langerak, and A. Klein, *Multi-Atlas Labeling Beyond the Cranial Vault - Workshop and Challenge. MICCAI 2015, Munich, Germany.* Accessed: 2018-01-17 (Cited on pages 121 and 127).

[201] H. Wang, S. R. Das, J. W. Suh, M. Altinay, J. Pluta, C. Craige, B. Avants, P. A. Yushkevich, A. D. N. Initiative, *et al.*, "A learning-based wrapper method to correct systematic errors in automatic image segmentation: consistently improved performance in hippocampus, cortex and brain segmentation", *NeuroImage*, vol. 55, no. 3, pp. 968–985, 2011 (Cited on page 123).

[202] H. Wang and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion and corrective learning - an open source implementation", *Frontiers in Neuroinformatics*, vol. 7, 2013 (Cited on page 123).

[203] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, 2013 (Cited on page 123).

[204] H. Wang, Y. Cao, and T. Syeda-Mahmood, "An experimental study on combining the auto-context model with corrective learning for canine leg muscle segmentation", in *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2015, pp. 1106–1109 (Cited on page 123).

[205] Z. Tu and X. Bai, "Auto-context and its application to high-level vision tasks and 3D brain image segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 10, pp. 1744–1757, 2010 (Cited on page 123).

[206] M. P. Heinrich, O. Maier, and H. Handels, "Multi-modal multi-atlas segmentation using discrete optimisation and self-similarities", *VISCERAL Organ Segmentation Challenge at ISBI, CEUR Workshop Proceedings*, p. 27, 2015 (Cited on pages 123, 125 and 127).

[207] E. Shechtman and M. Irani, "Matching local self-similarities across images and videos", in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2007, pp. 1–8 (Cited on page 123).

[208] M. Larsson, Y. Zhang, and F. Kahl, "DeepSeg: Abdominal organ segmentation using deep convolutional neural networks", in *Swedish Symposium on Image Analysis*, 2016 (Cited on pages 124, 125 and 127).

[209] F. Kahl, J. Alvén, O. Enqvist, F. Fejne, J. Ulén, J. Fredriksson, M. Landgren, and V. Larsson, "Good features for reliable registration in multi-atlas segmentation.", in *VISCERAL Organ Segmentation Challenge at ISBI, CEUR Workshop Proceedings*, 2015, pp. 12–17 (Cited on page 124).

[210] Z. Xu, R. P. Burke, C. P. Lee, R. B. Baucom, B. K. Poulose, R. G. Abramson, and B. A. Landman, "Efficient multi-atlas abdominal segmentation on clinically acquired CT with SIMPLE context learning", *Medical Image Analysis*, vol. 24, no. 1, pp. 18–27, 2015 (Cited on page 127).

[211] H. Wang, Y. Cao, and T. Syeda-Mahmood, "Multi-atlas segmentation with learning-based label fusion", in *International Workshop on Machine Learning in Medical Imaging*, Springer, 2014, pp. 256–263 (Cited on pages 125 and 127).

[212] R. Schübel, M. E. Graf, J. Nattenmüller, D. Nabers, D. Sookthai, L. F. Gruner, T. Johnson, C. L. Schlett, O. von Stackelberg, R. Kirsten, *et al.*, "The effects of intermittent calorie restriction on metabolic health: Rationale and study design of the HELENA Trial", *Contemporary Clinical Trials*, vol. 51, pp. 28–33, 2016 (Cited on page 134).

[213] O. Goksel, T. Gass, and G. Szekely, "Segmentation and landmark localization based on multiple atlases", in *VISCERAL Organ Segmentation Challenge at ISBI, CEUR Workshop Proceedings*, 2014, pp. 37–43 (Cited on page 138).

[214] J. Shen, T. Baum, C. Cordes, B. Ott, T. Skurk, H. Kooijman, E. J. Rummeny, H. Hauner, B. H. Menze, and D. C. Karampinos, "Automatic segmentation of abdominal organs and adipose tissue compartments in water-fat MRI: Application to weight-loss in obesity", *European Journal of Radiology*, vol. 85, no. 9, pp. 1613–1621, 2016 (Cited on pages 138 and 139).

[215] G. Chen, L. Gu, L. Qian, and J. Xu, "An improved level set for liver segmentation and perfusion analysis in MRIs", *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 1, pp. 94–103, 2009 (Cited on page 138).

[216] O. Gloger, K. Toennies, and J.-P. Kuehn, "Fully automatic liver volumetry using 3D level set segmentation for differentiated liver tissue types in multiple contrast MR datasets", in *Scandinavian Conference on Image Analysis*, Springer, 2011, pp. 512–523 (Cited on page 138).

[217] J. Thoma, F. Ozdemir, and O. Goksel, "Automatic segmentation of abdominal MRI using selective sampling and random walker", in *Medical Computer Vision and Bayesian and Graphical Models for Biomedical Imaging*, Springer, 2016, pp. 83–93 (Cited on page 138).

[218] H. Masoumi, A. Behrad, M. A. Pourmina, and A. Roosta, "Automatic liver segmentation in MRI images using an iterative watershed algorithm and artificial neural network", *Biomedical Signal Processing and Control*, vol. 7, no. 5, pp. 429–437, 2012 (Cited on page 138).

[219] F. López-Mir, V. Naranjo, J. Angulo, M Alcañiz, and L Luna, "Liver segmentation in MRI: a fully automatic method based on stochastic partitions", *Computer Methods and Programs in Biomedicine*, vol. 114, no. 1, pp. 11–28, 2014 (Cited on pages 138 and 139).

[220] H. T. Huynh, I. Karademir, A. Oto, and K. Suzuki, "Computerized liver volumetry on MRI by using 3D geodesic active contour segmentation", *American Journal of Roentgenology*, vol. 202, no. 1, pp. 152–159, 2014 (Cited on pages 138 and 139).

[221] J. Oh, D. R. Martin, and X. Hu, "Partitioned edge-function-scaled region-based active contour (p-ESRAC): automated liver segmentation in multiphase contrast-enhanced MRI", *Medical Physics*, vol. 41, no. 4, pp. 0 419 141–11, 2014 (Cited on page 138).

[222] T.-N. Le, H. T. Huynh, *et al.*, "Fully automatic scheme for measuring liver volume in 3D MR images", *Biomedical Materials and Engineering*, vol. 26, no. S1, pp. 361–369, 2015 (Cited on pages 138 and 139).

[223] A. Bereciartua, A. Picon, A. Galdran, and P. Iriondo, "Automatic 3D model-based method for liver segmentation in MRI based on active contours and total variation minimization", *Biomedical Signal Processing and Control*, vol. 20, pp. 71–77, 2015 (Cited on page 138).

[224] H. T. Huynh, N. Le-Trong, A. Oto, K. Suzuki, *et al.*, "Fully automated MR liver volumetry using watershed segmentation coupled with active contouring", *International Journal of Computer Assisted Radiology and Surgery*, vol. 12, no. 2, pp. 235–243, 2017 (Cited on page 138).

[225] G. Chartrand, T. Cresson, R. Chav, A. Gotra, A. Tang, and J. A. De Guise, "Liver segmentation on CT and MR using laplacian mesh optimization", *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 9, pp. 2110–2121, 2017 (Cited on pages 138 and 139).

[226] O. Bernard, J. G. Bosch, B. Heyde, M. Alessandrini, D. Barbosa, S. Camarasu-Pop, F. Cervenansky, S. Valette, O. Mirea, M. Bernier, *et al.*, "Standardized evaluation system for left ventricular segmentation algorithms in 3D echocardiography", *IEEE Transactions on Medical Imaging*, vol. 35, no. 4, pp. 967–977, 2016 (Cited on pages 146 and 148).

[227] D. Barbosa, D. Friboulet, J. D'hooge, and O. Bernard, "Fast tracking of the left ventricle using global anatomical affine optical flow and local recursive block matching", in *MICCAI Challenge on Endocardial 3D Ultrasound Segmentation (CETUS)*, Springer, 2014, pp. 17–24 (Cited on pages 146 and 147).

[228] J. Pedrosa, S. Queirós, O. Bernard, J. Engvall, T. Edvardsen, E. Nagel, and J. D'hooge, "Fast and fully automatic left ventricular segmentation and tracking in echocardiography using shape-based B-spline explicit active surfaces", *IEEE Transactions on Medical Imaging*, vol. 36, no. 11, pp. 2287–2296, 2017 (Cited on pages 146 and 147).

[229] K. Keraudren, O. Oktay, W. Shi, J. V. Hajnal, and D. Rueckert, "Endocardial 3D ultrasound segmentation using autocontext random forests", in *MICCAI Challenge on Echocardiographic 3D Ultrasound Segmentation (CETUS)*, Springer, 2014, pp. 41–48 (Cited on pages 146 and 147).

[230] F. Milletari, M. Yigitsoy, and N. Navab, "Left ventricle segmentation in cardiac ultrasound using hough-forests with implicit shape and appearance priors", in *MICCAI Challenge on Endocardial 3D Ultrasound Segmentation (CETUS)*, Springer, Citeseer, 2014 (Cited on pages 146 and 147).

[231] E. Smistad and F. Lindseth, "Real-time tracking of the left ventricle in 3D ultrasound using Kalman filter and mean value coordinates", in *MICCAI Challenge on Echocardiographic 3D Ultrasound Segmentation (CETUS)*, Springer, 2014, pp. 65–72 (Cited on page 146).

[232] D. Völgyes, M. Pedersen, A. Stray-Pedersen, D. Waaler, and A. C. T. Martinsen, "How different iterative and filtered back projection kernels affect computed tomography numbers and low contrast detectability", *Journal of Computer Assisted Tomography*, vol. 41, no. 1, pp. 75–81, 2017 (Cited on page 150).

[233] N. Sverzellati, E. Calabrò, A. Chetta, G. Concari, A. Larici, M. Mereu, R. Cobelli, M. De Filippo, and M. Zompatori, "Visual score and quantitative CT indices in pulmonary fibrosis: relationship with physiologic impairment", *La radiologia medica*, vol. 112, no. 8, pp. 1160–1172, 2007 (Cited on page 151).

[234] J. Wang, F. Li, and Q. Li, "Automated segmentation of lungs with severe interstitial lung disease in CT", *Medical Physics*, vol. 36, no. 10, pp. 4592–4599, 2009 (Cited on pages 152 and 155).

[235] O. Weinheimer, T. Achenbach, C. P. Heussel, and C. Düber, "Automatic lung segmentation in MDCT images", in *Fourth International Workshop on Pulmonary Image Analysis*, 2011, pp. 241–255 (Cited on page 152).

[236] D. Colombi, J. Dinkel, O. Weinheimer, B. Obermayer, T. Buzan, D. Nabers, C. Bauer, U. Oltmanns, K. Palmowski, F. Herth, *et al.*, "Visual vs fully automatic histogram-based assessment of idiopathic pulmonary fibrosis (IPF) progression using sequential multidetector computed tomography (MDCT)", *PLOS One*, vol. 10, no. 6, e0130653, 2015 (Cited on page 152).

[237] S. Sun, C. Bauer, and R. Beichel, "Automated 3-D segmentation of lungs with lung cancer in CT data using a novel robust active shape model approach", *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 449–460, 2012 (Cited on page 157).

[238] A. Soliman, F. Khalifa, A. Elnakib, M. A. El-Ghar, N. Dunlap, B. Wang, G. Gimel'farb, R. Keynton, and A. El-Baz, "Accurate lungs segmentation on CT chest images by adaptive appearance-guided shape modeling", *IEEE Transactions on Medical Imaging*, vol. 36, no. 1, pp. 263–276, 2017 (Cited on page 157).

[239] Y. Gao, Y. Shao, J. Lian, A. Z. Wang, R. C. Chen, and D. Shen, "Accurate segmentation of CT male pelvic organs via regression-based deformable models and multi-task random forests", *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1532–1543, 2016 (Cited on pages-169, 170, 174, 177 and 181).

[240] I. Oguz, J. Cates, M. Datar, B. Paniagua, T. Fletcher, C. Vachet, M. Styner, and R. Whitaker, "Entropy-based particle correspondence for shape populations", *International Journal of Computer Assisted Radiology and Surgery*, vol. 11, no. 7, pp. 1221–1232, 2016 (Cited on page 173).

[241] M. Kirschner, "The probabilistic active shape model: From model construction to flexible medical image segmentation", PhD thesis, Technische Universität, Darmstadt, 2013 (Cited on pages 173 and 174).

[242] J. Zhang, Y. Gao, L. Wang, Z. Tang, J. J. Xia, and D. Shen, "Automatic craniomaxillofacial landmark digitization via segmentation-guided partially-joint regression forest model and multiscale statistical features", *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 9, pp. 1820–1829, 2016 (Cited on page 175).

[243] C. Lindner, D. Waring, B. Thiruvenkatachari, K. OBrien, and T. Cootes, "Adaptable landmark localisation: Applying model transfer learning to a shape model matching system", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2017, pp. 144–151 (Cited on page 175).

[244] B. Ibragimov, R. Korez, B. Likar, F. Pernus, L. Xing, and T. Vrtovec, "Segmentation of Pathological Structures by Landmark-Assisted Deformable Models", *IEEE Transactions on Medical Imaging*, vol. 36, pp. 1457–1469, 2017 (Cited on page 176).

[245] C. Shi, Y. Cheng, J. Wang, Y. Wang, K. Mori, and S. Tamura, "Low-rank and sparse decomposition based shape model and probabilistic atlas for automatic pathological organ segmentation", *Medical Image Analysis*, vol. 38, pp. 30–49, 2017 (Cited on page 176).

[246] V. Chandrasekaran, S. Sanghavi, P. A. Parrilo, and A. S. Willsky, "Sparse and low-rank matrix decompositions", *IFAC Proceedings Volumes*, vol. 42, no. 10, pp. 1493–1498, 2009 (Cited on page 176).

[247] F. M. Sukno, S. Ordas, C. Butakoff, S. Cruz, and A. F. Frangi, "Active shape models with invariant optimal features: Application to facial analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 7, pp. 1105–1117, 2007 (Cited on page 176).

[248] C. Tobon-Gomez, F. M. Sukno, C. Butakoff, M. Huguet, and A. F. Frangi, "Automatic training and reliability estimation for 3D ASM applied to cardiac MRI segmentation", *Physics in Medicine and Biology*, vol. 57, no. 13, p. 4155, 2012 (Cited on page 176).

[249] A. Mansoor, J. J. Cerrolaza, R. A. Avery, and M. G. Linguraru, "Partitioned shape modeling with on-the-fly sparse appearance learning for anterior visual pathway segmentation", in *Workshop on Clinical Image-Based Procedures*, Springer, 2015, pp. 104–112 (Cited on page 176).

[250] A. Mansoor, J. J. Cerrolaza, R. Idrees, E. Biggs, M. A. Alsharid, R. A. Avery, and M. G. Linguraru, "Deep learning guided partitioned shape model for anterior visual pathway segmentation", *IEEE Transactions on Medical Imaging*, vol. 35, no. 8, pp. 1856–1865, 2016 (Cited on page-176).

[251] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes, "Robust and accurate shape model matching using random forest regression-voting", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, no. 9, pp. 1862–1874, 2015 (Cited on page 177).

[252] H. Blum and R. N. Nagel, "Shape description using weighted symmetric axis features", *Pattern Recognition*, vol. 10, no. 3, pp. 167–180, 1978 (Cited on page 179).

[253] J.-Y. Jeong, S. M. Pizer, and S. Ray, "Statistics on anatomic objects reflecting inter-object relations", in *MICCAI Workshop on Mathematical Foundations of Computational Anatomy: Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability*, Springer, 2006, pp. 136–145 (Cited on page 179).

[254] K. Siddiqi and S. Pizer, *Medial representations: mathematics, algorithms and applications*. Springer Science & Business Media, 2008, vol. 37 (Cited on page 179).

[255] J. Damon and E. Gasparovic, "Modeling multi-object configurations via medial/skeletal linking structures", *International Journal of Computer Vision*, vol. 124, no. 3, pp. 255–272, 2017 (Cited on page 179).

[256] M. De Bruijne, M. T. Lund, L. B. Tankó, P. C. Pettersen, and M. Nielsen, "Quantitative vertebral morphometry using neighbor-conditional shape models", *Medical Image Analysis*, vol. 11, no. 5, pp. 503–512, 2007 (Cited on page 179).

[257] M. Pereañez, K. Lekadir, I. Castro-Mateos, J. M. Pozo, Á. Lazáry, and A. F. Frangi, "Accurate segmentation of vertebral bodies and processes using statistical shape decomposition and conditional models", *IEEE Transactions on Medical Imaging*, vol. 34, no. 8, pp. 1627–1639, 2015 (Cited on page 179).

[258] A. Tsai, W. Wells, C. Tempany, E. Grimson, and A. Willsky, "Mutual information in coupled multi-shape model for medical image segmentation", *Medical Image Analysis*, vol. 8, no. 4, pp. 429–445, 2004 (Cited on page 179).

[259] A. Litvin and W. C. Karl, "Coupled shape distribution-based segmentation of multiple objects.", in *Biennial International Conference on Information Processing in Medical Imaging (IPMI)*, Springer, 2005, pp. 345–356 (Cited on page 179).

[260] P. Yan, A. A. Kassim, W. Shen, and M. Shah, "Modeling interaction for segmentation of neighboring structures", *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 2, pp. 252–262, 2009 (Cited on page 179).

[261] K. Gorczowski, M. Styner, J. Y. Jeong, J. Marron, J. Piven, H. C. Hazlett, S. M. Pizer, and G. Gerig, "Multi-object analysis of volume, pose, and shape using statistical discrimination", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 4, pp. 652–661, 2010 (Cited on page 179).

[262] J. Ruhaak, T. Polzin, S. Heldmann, I. Simpson, H. Handels, J. Modersitzki, and M. P. Heinrich, "Estimation of large motion in lung CT by integrating regularized keypoint correspondences into dense deformable registration", *IEEE Transactions on Medical Imaging*, vol. 36, no. 8, pp. 1746–1757, 2017 (Cited on page 179).

[263] F. Preiswerk, V. De Luca, P. Arnold, Z. Celicanin, L. Petrusca, C. Tanner, O. Bieri, R. Salomir, and P. C. Cattin, "Model-guided respiratory organ motion prediction of the liver from 2D ultrasound", *Medical Image Analysis*, vol. 18, no. 5, pp. 740–751, 2014 (Cited on page 179).

[264] C. Metz, N. Baka, H. Kirisli, M. Schaap, T. van Walsum, S. Klein, L. Neefjes, N. Mollet, B. Lelieveldt, M. de Bruijne, *et al.*, "Conditional shape models for cardiac motion estimation", in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Springer, 2010, pp. 452–459 (Cited on page 179).

[265] C. Metz, S. Klein, M. Schaap, T. van Walsum, and W. J. Niessen, "Nonrigid registration of dynamic medical imaging data using nD+t B-splines and a groupwise optimization approach", *Medical Image Analysis*, vol. 15, no. 2, pp. 238–249, 2011 (Cited on page 179).

[266] C. Chen, D. Belavy, W. Yu, C. Chu, G. Armbrecht, M. Bansmann, D. Felsenberg, and G. Zheng, "Localization and segmentation of 3D intervertebral discs in MR images by data driven estimation", *IEEE Transactions on Medical Imaging*, vol. 34, no. 8, pp. 1719–1729, 2015 (Cited on page 181).

[267] C. Chu, C. Chen, L. Liu, and G. Zheng, "Facts: fully automatic CT segmentation of a hip joint", *Annals of Biomedical Engineering*, vol. 43, no. 5, pp. 1247–1259, 2015 (Cited on page 181).