Thesis for the Degree of Licentiate of Engineering

Improving Multi-Atlas Segmentation Methods for Medical Images

Jennifer Alvén

CHALMERS

Department of Electrical Engineering
Chalmers University of Technology
Göteborg, Sweden 2017

Improving Multi-Atlas Segmentation Methods for Medical Images
Jennifer Alvén

Technical report no R007/2017 ISSN 1403-266X
Computer Vision and Image Analysis group
Department of Electrical Engineering
Chalmers University of Technology
SE–412 96 Göteborg, Sweden

Cover:
Illustration of a successful multi-atlas segmentation of the 1st lumbar vertebra in a whole-body CT image. Left figure: Manual labelling delineated by a medical expert. Central figure: Automatic segmentation computed with means of the method proposed in thesis paper IV. Right figure: Manual and automatic labellings overlaid. See Paper IV for more details.

Typeset by the author using LaTeX.

Chalmers Reproservice
Göteborg, Sweden 2017

Improving Multi-Atlas Segmentation Methods for Medical Images
Jennifer Alvén
Department of Electrical Engineering
Chalmers University of Technology

Abstract

Semantic segmentation of organs or tissues, *i.e.* delineating anatomically or physiologically meaningful boundaries, is an essential task in medical image analysis. One particular class of automatic segmentation algorithms has proved to excel at a diverse set of medical applications, namely multi-atlas segmentation. However, these multi-atlas methods exhibit several issues recognized in the literature. Firstly, multi-atlas segmentation requires several computationally expensive image registrations. In addition, the registration procedure needs to be executed with a high accuracy in order to enable competitive segmentation results. Secondly, up-to-date multi-atlas frameworks require large sets of labelled data to model all possible anatomical variations. Unfortunately, acquisition of manually annotated medical data is time-consuming which needless to say limits the applicability. Finally, standard multi-atlas approaches pose no explicit constraints on the output shape and thus allow for implausibly segmented anatomies.

This thesis includes four papers addressing the difficulties associated with multi-atlas segmentation in several ways; by speeding up and increasing the accuracy of feature-based registration methods, by incorporating explicit shape models into the label fusion framework using robust optimization techniques and by refining the solutions with means of machine learning algorithms, such as random decision forests and convolutional neural networks, taking both performance and data-efficiency into account. The proposed improvements are evaluated on three medical segmentation tasks with vastly different characteristics; pericardium segmentation in cardiac CTA images, region parcellation in brain MRI and multi-organ segmentation in whole-body CT images. Extensive experimental comparisons to previously published methods show promising results on par or better than state-of-the-art as of date.

Keywords: Supervised learning, semantic segmentation, medical image segmentation, multi-atlas segmentation, image registration, feature-based registration, label fusion, convolutional neural networks, random decision forests, conditional random fields.

Acknowledgements

First and foremost, I would like to offer my special thanks to my supervisor Fredrik Kahl for sharing interesting and novel ideas, for encouraging autonomy and ambition and for the helpful guidance through the academic jungle. I would also like to express my great appreciation to my ever so encouraging co-supervisor Olof Enqvist. Thank you for sharing reassuring wisdom as well as code snippets in time of need. I also wish to acknowledge my fellow doctoral students at the department of Electrical Engineering for making life at work more enjoyable; thanks for sharing laughter as well as frustration. I would especially like to express my gratitude to Måns Larsson, Carl Toft and Erik Stenborg; thanks for being supportive PhD colleagues and for being part of a helpful atmosphere free from competition.

Further, I wish to acknowledge:

My current and former roommates; Frida Fejne, Bushra Riaz and Fatemeh Shokrollahi Yancheshmeh. Thanks for the company and the never-ending patience.

Current, temporary and former members of the Computer Vision and Image Analysis group; Yuhang Zhang, Behrooz Nasihatkon, Jesús Briales García, Carl Olsson and Artur Chodorowski. Thanks for great collaboration and support.

The WiSE team; Sabine Reinfeldt, Hana Dobsicek Trefna, Helene Lindström, Yvonne Jonsson, Malin Ulfvarson and Elin Björklund. Thanks for acting as great female role models and for offering much needed advice in a male dominated world.

All medical research partners. I would especially like to acknowledge the Scapis team; David Molnar, Göran Bergström and Ola Hjelmgren. Thanks for time and effort spent on producing high-quality medical data.

Current and former PhDs at the Centre of Mathematical Sciences at Lund University whom I have collaborated with on various projects; Johannes Ulén, Johan Fredriksson, Matilda Landgren and Viktor Larsson.

Fellow researchers and administrative staff at MedTech West and the department of Electrical Engineering as well as former students at Chalmers University of Technology.

Finally, I would like to express my deepest gratitude to Jonas Ingesson, my former teacher in mathematics, to my soon-to-be husband Daniel Gustafsson and to family and friends; none mentioned, none forgotten.

Included publications

Paper I J. Alvén, A. Norlén, O. Enqvist and F. Kahl. "Überatlas: Fast and Robust Registration for Multi-atlas Segmentation". *Pattern Recognition Letters,* 80:245–255, 2016. Extended version of paper (a).

Paper II A. Norlén, J. Alvén, D. Molnar, O. Enqvist, R. Rossi Norrlund, J. Brandberg, G. Bergström and F. Kahl. "Automatic Pericardium Segmentation and Quantification of Epicardial Fat from Computed Tomography Angiography". *Journal of Medical Imaging,* 3(3), 2016.

Paper III F. Fejne, M. Landgren, J. Alvén, J. Ulén, J. Fredriksson, V. Larsson and F. Kahl. "Multi-atlas Segmentation Using Robust Feature-Based Registration". In *Cloud-Based Benchmarking of Medical Image Analysis,* Springer International Publishing, 203–218, 2017. Extended version of paper (b).

Paper IV J. Alvén, F. Kahl, M. Landgren, V. Larsson, J. Ulén and O. Enqvist. "Shape-Aware Label Fusion for Multi-Atlas Frameworks". Submitted to *Pattern Recognition Letters.* Extended version of paper (c).

Subsidiary publications

(a) J. Alvén, A. Norlén, O. Enqvist and F. Kahl. "Überatlas: Robust Speed-Up of Feature-Based Registration and Multi-Atlas Segmentatio". *Scandinavian Conference on Image Analysis (SCIA),* 92–102, 2015. Received the "Best Student Paper Award" at SCIA 2015.

(b) 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". *VISCERAL Challenge@ ISBI,* 12–17, 2015.

(c) J. Alvén, F. Kahl, M. Landgren, V. Larsson and J. Ulén. "Shape-Aware Multi-Atlas Segmentation". *IAPR International Conference on Pattern Recognition (ICPR),* 1101–1106, 2016. Received the "IBM Best Student Paper Award (Track: Biomedical Image Analysis and Applications)" at ICPR 2016.

Abbreviations

General

BMI	**B**ody **M**ass **I**ndex
CAD	**C**omputer-**A**ided **D**iagnosis
CAS	**C**omputer-**A**ssisted **S**urgery
CNN	**C**onvolutional **N**eural **N**etwork
CRF	**C**onditional **R**andom **F**ield
CT(A)	**C**omputed **T**omography (**A**ngiography)
EF(V)	**E**picardial **F**at (**V**olume)
HU	**H**ounsfield **U**nits
MRF	**M**arkov **R**andom **F**ield
MR(I)	**M**agnetic **R**esonance (**I**maging)
(N)MI	(**N**ormalized) **M**utual **I**nformation
SCAPIS	**S**wedish **CA**rdio**P**ulmonary bio**I**mage **S**tudy
SAD	**S**um of **A**bsolute **D**istances
SSD	**S**um of **S**quared **D**istances
SVM	**S**upport **V**ector **M**achine
TPS	**T**hin **P**late **S**plines
VISCERAL	**VIS**ual **C**oncept **E**xtraction challenge in **RA**dio**L**ogy

Methods

ADMM	**A**lternating **D**irection **M**ethod of **M**ultipliers
DRAMMS	**D**eformable **R**egistration via **A**ttribute **M**atching and **M**utual-**S**aliency weighting
ICP	**I**terative **C**losest **P**oint
IRLS	**I**teratively **R**eweighted **L**east **S**quares
MAPER	**M**ulti-**A**tlas **P**ropagation with **E**nhanced **R**egistration
RANSAC	**RA**ndom **SA**mple **C**onsensus
SIFT	**S**cale-**I**nvariant **F**eature **T**ransform
SIMPLE	**S**elective and **I**terative **M**ethod for **P**erformance **L**evel **E**stimation
STAPLE	**S**imultaneous **T**ruth **A**nd **P**erformance **L**evel **E**stimation
SURF	**S**peeded **U**p **R**obust **F**eatures

Contents

II Included Publications

Part I

Introductory Chapters

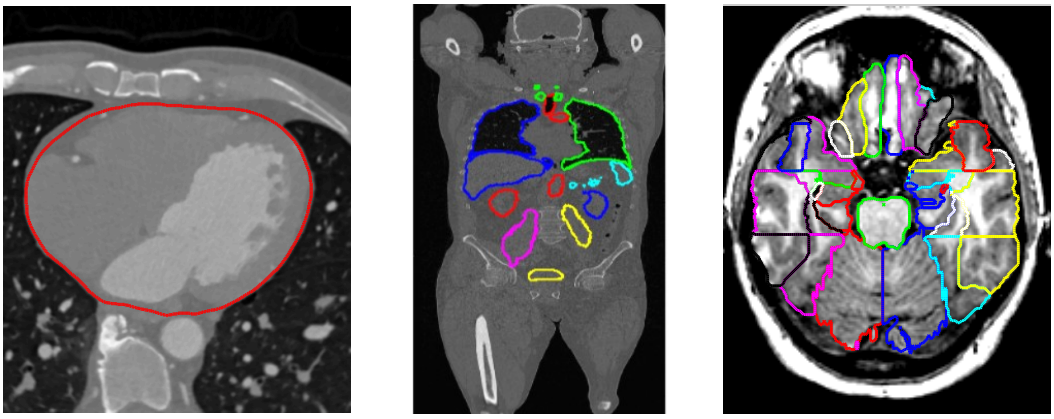
Chapter 1

Introduction

Medical imaging, *i.e.* techniques for producing visual representations of the interior (human) body, allows scientists and clinicians to examine, diagnose and treat diseases with means of non-invasive radiology. Medical images, acquired with *e.g.* ultrasound, magnetic resonance imaging (MRI) or non-enhanced/enhanced computed tomography (CT/CTA), provide information essential for understanding and modeling healthy as well as diseased anatomy. Decades of successful development of imaging techniques have brought an increased image quality capturing fine anatomical and functional details while the amount of images acquired on a daily basis is steadily growing. The demand for automatic tools for analysis has increased along this development, since manual techniques for inspection cannot effectively and accurately process the huge amount of high-quality data [1].

The focus of this thesis is semantic segmentation of anatomical structures in medical 3D images such as CT, CTA and MRI. Semantic segmentation, *i.e.* dividing an image into meaningful parts by assigning each voxel (a 3D pixel) a label, is an essential problem in medical image analysis and thus utterly well-studied. Commonly, the labels are predetermined and correspond to biologically meaningful object classes, such as different organs or tissue types. The set of labels might correspond to anatomically derived objects embedded in a "background" (*e.g.* different organs in whole-body CT), or physiologically (functionally) derived sub-regions densely covering large parts of the image (*e.g.* region parcellation in brain MRI). See Figure 1.1 for three examples of medical segmentation problems.

Segmentation of medical images has numerous applications. Delineated organ and tissue boundaries are used for both diagnostic and visualization purposes. Examples of sub-problems are detection and localization of tumors and other pathologies, tissue volume quantification and organ localization. Further, segmentation results are useful for a wide spectrum of applications such as computer-aided diagnosis (CAD) systems, radiotherapy planning and in computer-assisted surgery (CAS), *e.g.* surgery planning, virtual surgery simulation, intra-surgery navigation and robotic surgery [6, 7].



(a) (b) (c)

Figure 1.1: *Slices of medical 3D images and manual labellings (coloured contours) from the three different datasets considered in the included thesis papers. (a) Slice of a* Scapis *[2] cardiac CTA image plus pericardium ("heart sack") labelling. (b) Slice of a* Visceral *[3] whole-body CT image plus organ labellings,* e.g. *lungs, liver, kidneys* etc. *(c) Slice of a* Hammers *[4, 5] brain MRI plus region labellings,* e.g. *hippocampus, amygdala* etc.

Manual delineation of anatomical structures is time-consuming and the quality is highly determined by the expert's skill set. Further, the interobserver variability is usually high. Thus, manual annotation of images is not feasible for applications such as large-scale studies or computer-assisted surgery. Compared to manual methods, automatic segmentation methods are typically fast, cheap, reliable and scale well. Automatic methods able to accurately obtaining boundaries of organs and tissues are therefore highly requested in medical research and by clinical care [8].

Medical images offer several challenges compared to their non-medical counterparts. Typically, medical images contain both low contrast details as well as a moderate to a high level of noise. Inter- and intra-patient variability and imaging ambiguities such as motion artifacts and partial volume effects further increase the difficulty. Compared to neighbouring research fields such as image classification and computer vision, manually labelled data is rarely abundant. However, common challenges associated with 2D images, such as (partial) occlusion and light source ambiguities, are usually avoided when processing medical 3D images. Due to these distinct differences (comparing medical images to "standard" 2D images), the research field includes several segmentation methods specifically adapted for medical imaging [7].

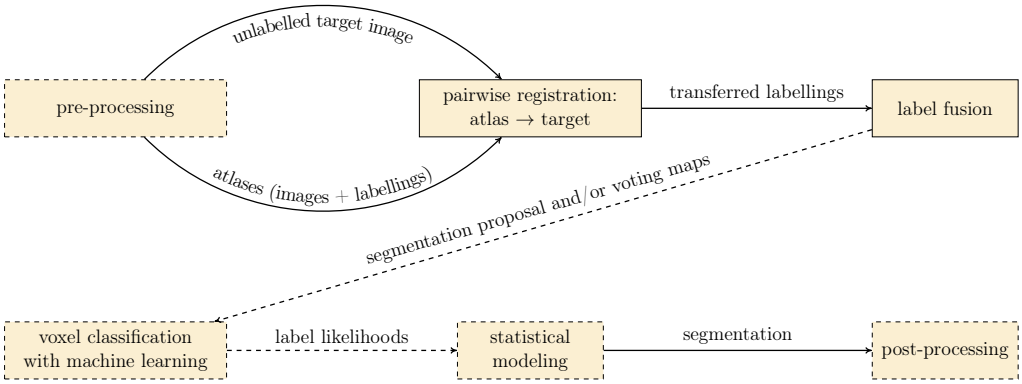


Figure 1.2: *Schematic summary of the multi-atlas framework. Mandatory sub-steps in the pipeline constitute pairwise registration of atlas images to an unlabelled target image followed by label propagation and label fusion. Different label fusion schemes may provide voxelwise label likelihoods and/or a segmentation proposal. Optional sub-steps (dashed blocks and arrows) such as local voxel classification and statistical modeling as well as pre-/post-processing may be included or left out.*

In recent years, one particular class of segmentation algorithms called *multi-atlas segmentation* has proved to excel at several segmentation tasks and across different modalities (medical imaging techniques). The multi-atlas framework has been comprehensively used on a diverse set of applications, *e.g.* brain MRI [9–13], knee MRI [14], cardiac CT [15, 16] and CTA [17, 18], thoracic CT [19, 20], abdominal CT [21–24] and whole-body CT [25]. For more applications on medical segmentation using multi-atlas approaches, see the recent survey in [8].

Multi-atlas segmentation relies on a set of atlases (images with corresponding manual labellings), which are separately registered (*i.e.* aligned) to an unlabelled target image. The images are typically registered using a global linear transformation followed by a local, elastic transformation if refinement is necessary. See Section 2.2.1 for details regarding this procedure. Each atlas labelling is transferred to the coordinate frame of the target image according to the pairwise registration. The transferred labellings are combined into one segmentation proposal by label fusion, see Section 2.2.2. Some fusion schemes produce a final segmentation output, while others produce voting maps (*i.e.* voxelwise label likelihoods) that can be further processed. In some frameworks, the segmentation proposal is further refined by using machine learning techniques and/or statistical modeling. In Section 2.3, two standard machine learning tools for voxel classification, random decision forests and convolutional neural networks, respectively are described in detail. In Section 2.4, a probabilistic graphical model suitable for image segmentation, conditional random fields, is presented. See Figure 1.2 for a schematic summary of the multi-atlas pipeline.

Table 1.1: *Summary of datasets used for training, validation and testing in the included thesis papers. Background class is excluded from the number of different classes.*

Name	Modality	Task	# of classes	Papers
Scapis [2]	cardiac CTA	pericardium segmentation	1	I, II
Visceral [3]	whole-body CT	multi-organ segmentation	20	III, IV
Hammers [4, 5]	brain MRI	brain region parcellation	83	I, IV

1.1 Thesis aim and scope

The included thesis papers present possible improvements for the multi-atlas segmentation framework. The intended usage is organ (or region) segmentation of medical 3D images. Three major research questions are addressed:

(i) How can we improve performance and precision of medical segmentation algorithms in order to meet the requirements on timing and accuracy posed by *e.g.* computer-aided diagnosis and surgery as well as medical research?

(ii) How can we guarantee anatomically meaningful segmentation results while still allowing for generalizability and scalability?

(iii) How can we reduce the reliance on access to large sets of manually labelled data when developing competitive segmentation methods?

Typically, current segmentation methods greatly depend on modality and application, leading to task-specific methods of little use for dissimilar segmentation tasks. In this thesis, the proposed methods aim to achieve the opposite, *i.e.* generalizing well across a diverse set of applications and imaging techniques, by considering three significantly different datasets, see Table 1.1 and Figure 1.1.

1.2 Thesis outline

The thesis is divided into two parts. Part I constitutes the introductory chapters; Chapter 2 briefly compiles theory and methods necessary for understanding the remainder of the thesis, Chapter 3 summarizes the main contribution for each of the included thesis papers and Chapter 4 provides a concluding discussion and potential future research directions. Part II comprises the four included thesis papers.

Chapter 2

Preliminaries

The following sections briefly compile theory, concepts, methods and tools made use of in the included thesis papers and can with ease be skipped by experienced readers. The chapter is structured as follows: Section 2.1 briefly lists some reoccurring key concepts and is intended to be used as a dictionary for inexpert readers. Multi-atlas segmentation, including the two essential concepts image registration and label fusion, is presented in Section 2.2. Two standard machine learning methods, random decision forests and convolutional neural networks, applied in some of the included thesis papers are summarized in Section 2.3. Finally, the theoretical building blocks for the conditional random fields model are accounted for in Section 2.4.

2.1 Basic concepts

Atlas: The term *atlas* refers to an image pair consisting of an intensity image and a corresponding manual labelling.

Classes: The *classes* are a predefined set of objects relevant for the application, such as ”kidney”, ”pancreas”, ”liver” *etc*. (for abdominal organ segmentation).

Classification: Image classification means assigning one or more discrete classes (such as dog, cat *etc*.) to an entire image, while voxelwise classification refer to compute a label for each voxel, *e.g*. heart voxel, lung voxel, liver voxel *etc*.

Ground truth/Gold standard: A manual labelling, delineated by a physician or other medical expert, is usually referred to as the *ground truth* labelling. In medical applications, the term *gold standard* is sometimes used instead (indicating the lack of objective truth when it comes to medical image segmentation).

Image: In this thesis, an *image* refers to a 3D matrix where the elements contain gray-scale intensity levels measured by a medical imaging instrument such as a MR scanner or a CT scanner.

Label: A voxel *label* indicates which object class the specific voxel belongs to. Commonly, labels are represented by different integer values. For binary segmentation problems, zero (black) typically corresponds to *background class* while one (white) corresponds to *foreground/object/organ class.*

Labelling: An image *labelling* refers to an integer matrix of the same dimension as the corresponding image, where each voxel has been assigned a label (either manually or automatically).

Modality: The type of imaging technique, *i.e.* type of scanner or probe, that has been used to acquire a medical image is sometimes referred to as the *modality*, *e.g.* ultrasound, CT, MRI *etc.*

Probability map: In some of the included thesis papers, the term *probability map* is used for denoting a voxelwise label likelihood derived *e.g.* from the multi-atlas voting map or by machine learning techniques.

Segmentation: The words labelling and *segmentation* can be used interchangeably, however, a segmentation typically refers to an image labelling acquired with (semi-)automatic methods.

Target (image): The *target*, or *target image*, refers to the unlabelled image that is to be segmented. The terms *fixed image* or *reference image* may be used interchangeably.

Voting map: In this thesis, a voxelwise label likelihood (unnormalized) inferred from propagated labellings via label fusion is named *voting map*. See Section 2.2 for more details.

Voxel: A matrix element in a volumetric image, *i.e.* a 3D pixel, is sometimes referred to as a *voxel* (VOlume piXEL).

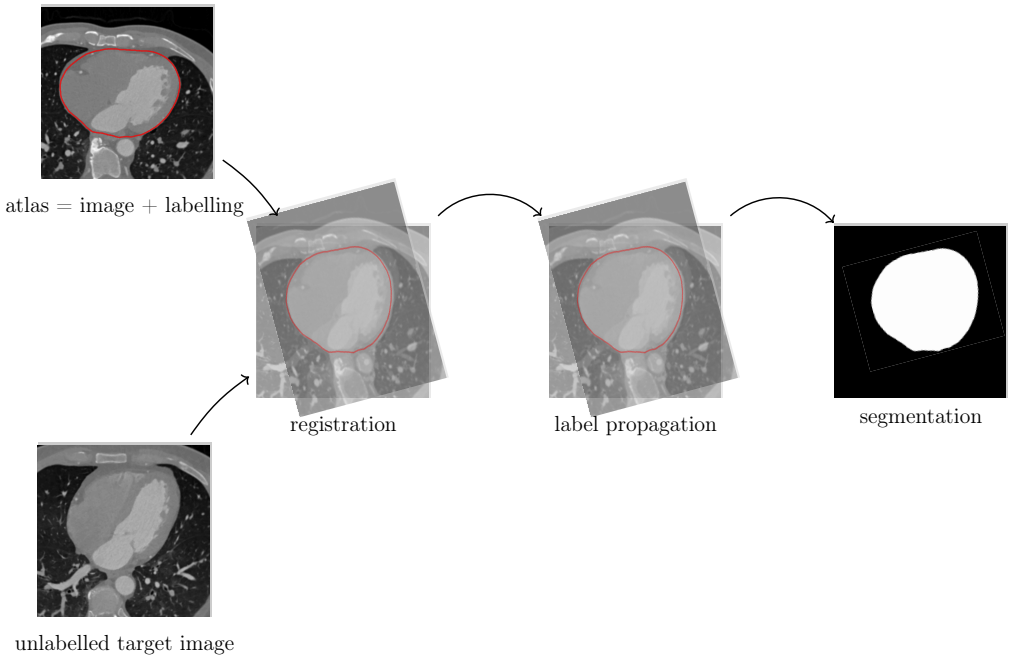


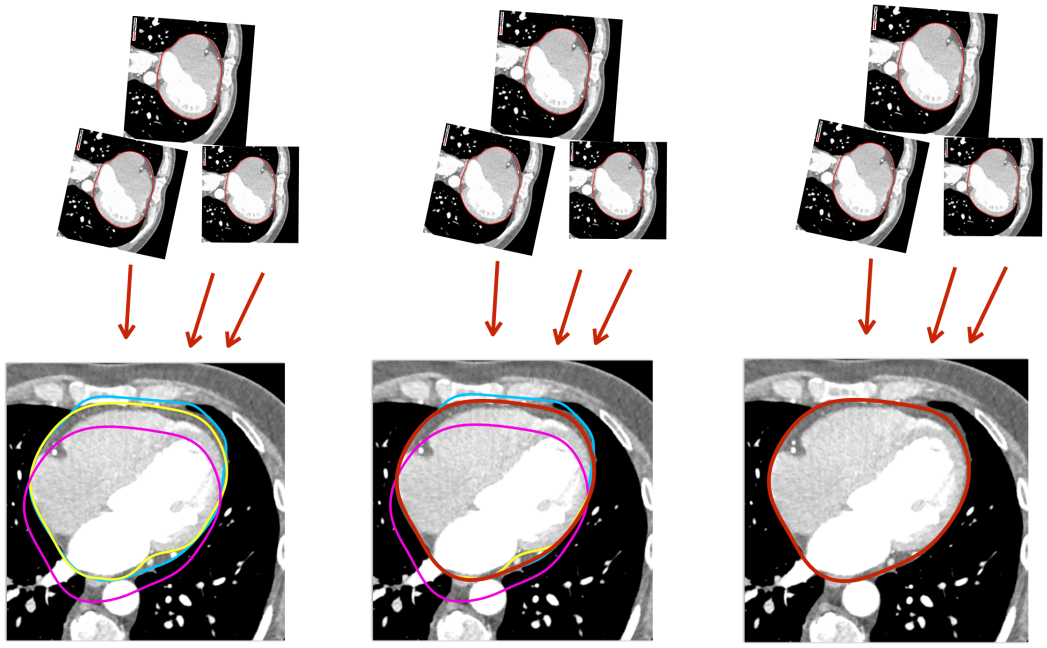
Figure 2.1: *Example of single-atlas segmentation (registration-based segmentation) of the pericardium in a* Scapis *cardiac CTA slice.*

2.2 Multi-atlas segmentation

Multi-atlas segmentation [26–28], proposed over a decade ago, is one of the most widely used methods for segmentation in medical applications. For an extensive summary of the research field, see the recent survey in [8].

Multi-atlas segmentation is an extension of single-atlas segmentation. An *atlas* means an image paired with a corresponding labelling. Single-atlas segmentation relies on registering one atlas image to the unlabelled target image and transferring the labelling according to the computed transformation. Thus, the inferred target image segmentation equals the aligned labelling. For that reason, single-atlas segmentation is also called registration-based segmentation. Figure 2.1 exemplifies single-atlas segmentation of the pericardium ("heart sack") in a slice of a Scapis cardiac CTA. Refer to Section 2.2.1 for details regarding image registration.

Two or more single-atlas segmentations can be combined into a multi-atlas segmentation. The motivation behind using several atlases is *e.g.* to capture all possible anatomical variations and to increase the robustness to imperfect registration results. Thus, multi-atlas segmentation involves registration of several atlas images to the unlabelled target image. According to the pairwise atlas-target registrations, each atlas labelling is propagated to the target image space and thereafter combined via *label fusion*.



(a) (b) (c)

Figure 2.2: *Example of a multi-atlas segmentation of the pericardium in a slice of a* Scapis *cardiac CTA image using three atlases. (a) The atlas images are registered to the unlabelled target image and the labellings are transferred accordingly (the contours of the labellings are marked as yellow, cyan and magenta respectively). (b) The transferred labellings are combined into one segmentation proposal (red contour) by label fusion. (c) The inferred segmentation accurately delineates the pericardium compared to the three individual single-atlas segmentations.*

In some label fusion approaches, the final segmentation is directly inferred by fusing the transferred labels. For other approaches, label fusion rather serves to combine the transferred labellings into a *voting map*, *i.e.* a voxelwise likelihood for each label, that may be used in a subsequent analysis step. See Section 2.2.2 for more details regarding label fusion. Figure 2.2 depicts an example of a coarse multi-atlas segmentation (Scapis pericardium segmentation) using three atlases.

There are several multi-atlas approaches using varying refinement techniques beyond label fusion. The transferred labels, the voting map and/or the fused segmentation proposal may serve as either data input or spatial initialization for *e.g.* machine learning classifiers, see Section 2.3, or a conditional random fields model, see Section 2.4. Also, pre- and postprocessing of the input (*i.e.* the target image and the atlases) and the output (*i.e.* the segmentation), such as filtering, are commonly included in multi-atlas frameworks.

2.2.1 Image registration

To register an atlas image to a target images means computing a transformation that aligns the atlas image to the target image. Image registration algorithms aim to align a source image, $\mathcal{I}_s$, to a target image, $\mathcal{I}_t$, by solving an optimization problem of the form

$$\mathbf{T}^* = \arg\min_{\mathbf{T}} \left[\rho_1(\mathcal{I}_t, \mathbf{T} \circ \mathcal{I}_s) + \rho_2(\mathbf{T})\right], \tag{2.1}$$

where $\mathbf{T}$ is a coordinate transformation from source image voxels to target image voxels; $\mathbf{T} \circ \mathcal{I}_s$ means mapping the source image voxels to the target image space. The level of alignment of the target image and the warped source image is quantified by the first term, ρ_1, while the second term, ρ_2, aims to regularize the transformation, *e.g.* by penalizing implausible deformations and/or by introducing prior knowledge of the deformation. The form of the regularization term should be influenced by the choice of transformation.

Thus, image registration allows for several design choices; type of (i) transformation, (ii) objective function and (iii) optimization method. For a comprehensive overview of different registrations methods and their design choices, see the surveys in [29, 30].

Transformation types

Preferably, the type of transformation is determined by the application. In multi-atlas approaches, the images are typically first aligned using an affine transformation. The affine transformation translates, rotates, scales, reflects and/or shears the image globally. Mathematically, it can be described as a composition of a linear map $\mathbf{A}$ and a translation $\mathbf{t}$:

$$\mathbf{T}(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{t}, \tag{2.2}$$

where $\mathbf{x}$ is the voxel coordinates.

To capture the local nonlinear deformations commonly present in medical applications, the affine transformation is sometimes followed by a non-rigid registration using a nonlinear dense transformation. The deformation is elastic and warps the image locally by using a displacement field $\mathbf{U}$ (that varies with voxels):

$$\mathbf{T}(\mathbf{x}) = \mathbf{x} + \mathbf{U}(\mathbf{x}). \tag{2.3}$$

However, estimating an accurate non-rigid transformation tend to be more computationally demanding than the linear counterpart. Thus, non-rigid registration may be omitted in applications such as computer-assisted surgery or large-scale studies due to timing issues.

Objective functions and optimization methods

The choice of objective function, and thereby also the optimization method, is highly influenced by the image registration approach. Roughly speaking, there are two different approaches to image registration; intensity-based registration and feature-based registration. Of course, there are hybrid methods combining advantages of both approaches such as Dramms [31] (Deformable Registration via Attribute Matching and Mutual-Saliency weighting) and the block-matching strategy in [32, 33].

Using intensity-based methods is a popular choice in medical applications, *e.g.* [34–36], due to their capability of producing accurate registrations, even between images in different modalities. Unfortunately, intensity-based registration methods tend to be computationally demanding and sensitive to initialization; the objective functions are usually computed over the entire image domain and optimized locally (increasing the risk of getting trapped in a sub-optimal, local minimum).

Feature-based methods, using sparse point correspondences between images for establishing coordinate transformations, are typically faster and more robust to initialization and large deformations. The objective functions are typically quantifying residual errors of the mapped point correspondences. This class of objective functions enables efficient computations and optimization methods able to find a global (approximate) minimum. However, these methods risk failing due to the difficulty in detecting salient features in medical images; distinctive features are crucial for establishing correct point-to-point correspondences between the images. Therefore, the accuracy of (sparse) feature-based methods is generally assumed to be inferior to intensity-based methods.

Intensity-based registration. Intensity-based registration methods rely on comparing voxelvise characteristics such as intensities, colors, depths *etc.* directly. Typically, these methods use local optimization or multiresolution strategies for minimizing an objective function such as sum of squared distances, sum of absolute distances, cross-correlation or (normalized) mutual information, (N)MI, [37]. See the comparisons in [38, 39] for different optimization strategies. The non-rigid transformation is commonly represented by deformations derived from physical models such as the diffusion model in [40] (Demons) or by interpolation-based models such as radial basis functions, *e.g.* thin plate splines (TPS) [41], or free-form deformations, *e.g.* cubic B-splines [42]. However, there are numerous nonlinear deformation models in the image registration literature, see the survey in [30].

Feature-based registration. Despite being a popular choice in *e.g.* computer vision and remote sensing, feature-based registration is less common in medical image analysis due to the difficulty of detecting distinctive features in medical images. However, Svärm *et al.* [43] showed that feature-based registration based on robust optimization outperforms several intensity-based methods when applied to whole-body CT and brain MRI.

Sparse feature-based registration methods rely on established point-to-point correspondences between images for estimating coordinate transformations. In order to establish correct correspondences, one needs to (i) detect distinctive feature points in each image and (ii) match detected feature points by taking their similarity in appearance into account. There are numerous hand-crafted feature detectors where the prime examples are Sift [44] (using difference-of-Gaussians) and Surf [45] (using integral images). Feature detectors are paired with a *descriptor*, a histogram aiming to provide a unique description of the feature point and its neighbourhood. These descriptors are computed locally and include image characteristics such as intensity information, gradients, higher order derivatives and/or wavelets. Preferably, the descriptor should be invariant to scale, pose, contrast and, for some applications, rotation. Recently, automatically learned feature detectors and descriptors have proved to excel at several applications, *e.g.* detectors and descriptors learned with convolutional neural networks [46, 47].

Once having detected and described a set of features points for the images that are to be registered, one needs to robustly match the descriptors in order to derive point-to-point correspondences. Usually, a metric measuring the distance (*e.g.* Euclidean distance) between the descriptors is used to rank the quality of match hypotheses. A one-to-one correspondence is derived by *e.g.* choosing the nearest neighbour in the descriptor space (either computed in one direction, non-symmetrically, or compute in both directions, non-symmetrically), perhaps combined with a criterion such as in [44] (comparing ratios between nearest and second nearest neighbour). However, more advanced classification tools such as convolutional neural networks can be used for matching as well [48].

Given the match hypotheses, iterative algorithms such as Ransac [49] can be used to estimate the parameters of an affine transformation approximately aligning the two images to be registered. If a non-rigid registration should follow, matches that are inconsistent with this affine transformation, so called *outliers*, are usually sorted out. A non-rigid deformation may be represented by interpolation-based techniques, *e.g.* B-splines as in [50] or thin plate splines as in [51]. There are also methods simultaneously establishing one-one-point correspondences while estimating the mapping, such as modified variants of the Iterative Closest Point (ICP) method [52], *e.g.* the registration method in [53].

2.2.2 Label fusion

In single-atlas segmentation, the final segmentation equals the transferred labels of the one atlas image used. In multi-atlas segmentation, there are several propagated atlas labellings that need to be combined into one unique segmentation proposal.

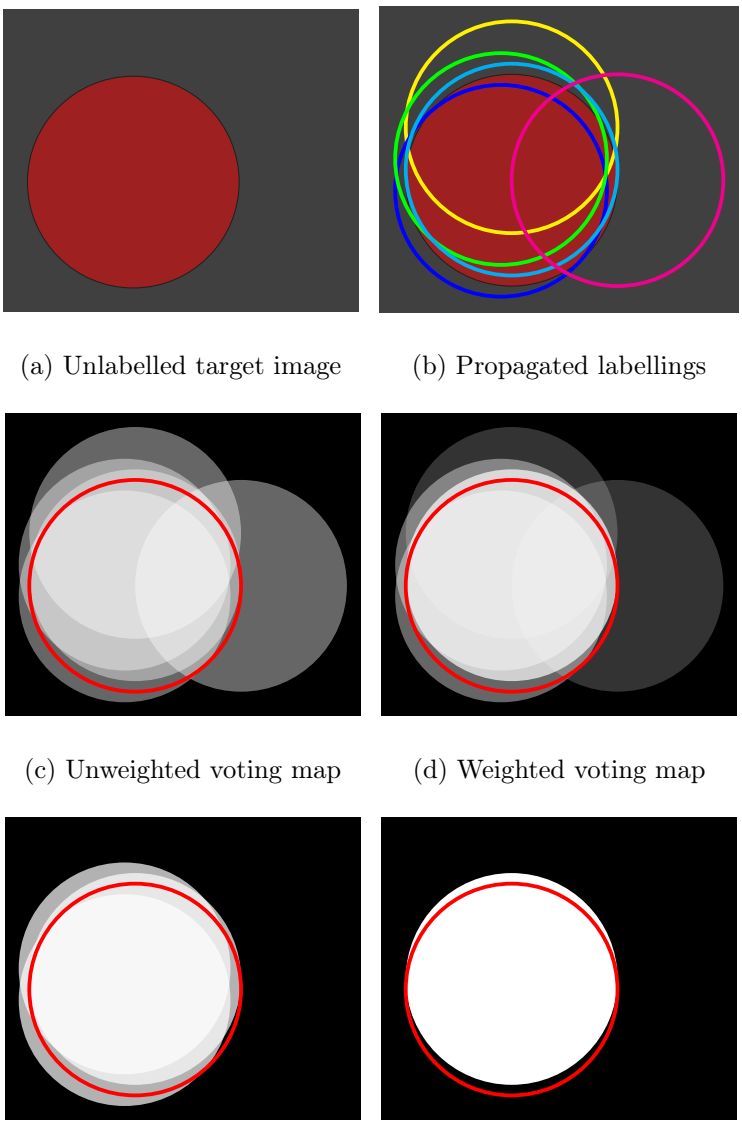
Each transferred atlas labelling can viewed as a vote, for each voxel indicating whether that particular atlas estimates the voxel to be inside/at the organ boundary or not. By summarizing all votes in one image a voting map is obtained. The voting map can be regarded as an unnormalized voxelwise label likelihood over the entire image. From this voting map, the final segmentation can be inferred by *e.g.* thresholding or statistical reasoning. The process of combining several transferred atlas labellings into one voting map is referred to as label fusion. For some label fusion schemes, the output simply equals the voting map while other fusion strategies output the final inferred segmentation proposal.

The simplest fusion scheme is unweighted voting, *e.g.* [26–28], meaning that each registered atlas is assigned the same weight, see Figure 2.3c. Typically, methods using unweighted voting maps infer the final segmentation by *majority voting*. As the name implies, majority voting means that the most frequent label is assigned to each voxel.

It is common to sift out promising atlas candidates and only fuse this restricted subset. This process, known as *atlas selection*, has proven to improve the computational efficiency (by decreasing the amount of registrations that need to be computed) and accuracy (by ignoring irrelevant anatomies). Atlas selection can be done either before pairwise registration, *e.g.* [54], by choosing atlas images believed to best represent the anatomical shape variation, or after, *e.g.* [55], by choosing the atlas images which are more similar to the target image and/or are believed to boost the algorithm performance. Common similarity metrics used for atlas selection are sum of squared distances, cross-correlation and non-image data such as age difference. The most simple case of atlas selection is *best atlas selection* [26], where merely one atlas is chosen. Note that best atlas selection is a special cases of single-atlas segmentation where the one atlas is chosen according to *e.g.* image similarity. See Figures 2.3e and 2.3f for examples of atlas selection and best atlas selection respectively.

Atlas selection may be regarded as an extreme case of weighted voting, *i.e.* fusing propagated labels by assigning each atlas different weights, see Figure 2.3d. The atlas weights can be derived globally, as in [55, 56], or locally (patchwise or voxelwise) as in [25, 57–60].

There are numerous additional sophisticated fusion schemes including ideas from statistics and machine learning. Among others, there are strategies using *e.g.* probabilistic reasoning regarding predicted performance [24, 55, 61, 62], generative probabilistic models [63] and convolutional neural networks [64].



(a) Unlabelled target image (b) Propagated labellings

(c) Unweighted voting map (d) Weighted voting map

(e) Atlas selection (f) Best atlas selection

Figure 2.3: *Toy example visualizing different label fusion strategies. (a) An unlabelled target image depicting a red, circular shape on a gray background. (b) Five atlases are registered to the unlabelled target image and labellings (coloured contours) are propagated accordingly. (c) Unweighted voting fuses the labellings directly by assigning each atlas the exact same weight. The red contour indicates the location of the true boundary (ground truth labelling). (d) Weighted voting assigns different weights to each atlas based on* e.g. *image similarity. (e) Atlas selection sifts out promising atlas candidates before/after registration according to* e.g. *image similarity. (f) Best atlas selection is equivalent to single-atlas segmentation using only one atlas chosen with respect to* e.g. *image similarity.*

2.3 Machine learning tools for voxel classification

As previously mentioned in Section 2.2, machine learning classifiers can be utilized as an additional refinement step in multi-atlas frameworks, before or after label fusion. Typically, a classifier is fed with data input, such as the unprocessed image and/or features derived by processing the image, and outputs a voxelwise label likelihood over the image. The label fusion output, *e.g.* voting maps and/or segmentation proposals, may serve as either data input or spatial initialization (*i.e.* defining the region of interest) for such classifiers. The output of a classifier, the voxelwise label likelihood, can either be thresholded in order to infer a final segmentation or it can further processed, for instance with means of a conditional random field model. Included thesis papers in this thesis make use of two type of machine learning classifiers; random decision forests and convolutional neural networks. Therefore, a brief overview of the techniques follows below.

2.3.1 Random decision forests

Random decision forests [65, 66] (short: random forests) are a machine learning technique suitable for classification tasks. It is a computationally efficient method, appropriate for binary classification tasks as well as multi-class problems, and it generalizes well to unseen data. The technique has successfully been used as an additional refinement step in multi-atlas pipelines and can be applied both before and after label fusion, *cf.* [67–70].

When applied to an unlabelled image voxel, a random decision forest is fed a set of features, *i.e.* characteristics derived from the image, as input and outputs an estimated conditional probability over labels, $\hat{P}(l|f)$, where l denotes the voxel label and f denotes a vector consisting of the input features. In that manner, random decision forests may be used in order to estimate voxelwise probabilities for each label, *i.e.* a likelihood estimate for each voxel belonging to a certain class. The random forest training and classification is done voxelwise, that is, no spatial dependencies are encoded.

Typically, features such as image intensities, gradients and/or higher order derivatives are used. It is also common to pre-process the image, *e.g.* by filtering, and include these pre-processed intensities as features. If the random forest classifier is part of a larger multi-atlas framework, transferred labellings and/or the result of label fusion may be used as features as well. It is good practice to normalize each feature before training to have zero mean and unit standard deviation with respect to the training set.

Decision trees

A random decision forest consists of a set of decision trees, binary trees where each node is associated with its own splitting (decision) function. A common choice of splitting function is a separating hyperplane of the same dimension as the input feature vector. The parameters of the hyperplane are learned during training and usually chosen such that the information gain (*i.e.* the confidence) is maximized and/or the entropy (*i.e.* the unpredictability) is minimized.

When classifying an unlabelled voxel, the input data point begins at the rote node. Depending on the result of the current splitting function (*i.e.* the decision), the data point is either passed to the right or to the left child node. The subsequent nodes will continue passing the data point along the tree until it reaches a leaf node. The leaf nodes contain posterior distributions over labels, learned during training, and thus output a conditional probability for the data point belonging to a certain class.

In Figure 2.4a training of a binary decision tree is visualized. In this specific example, 20 data points are used for training. There are two classes, blue and red, and two different features have been extracted for each data point. That is, the classification problem is two-dimensional. The binary decision tree has in total six nodes; one root node, two decision nodes and three leaf nodes. Below the leaf nodes, the estimated posterior distribution for the two different classes (for that particular leaf) is given. In Figure 2.4b classification of one unlabelled data point is visualized. The data point is passed along the three according to the decision nodes, and the estimated posterior distribution over the classes is decided by the leaf node the data point end up in. For this particular example, the data point would be classified as "red", since the estimated posterior distribution is the largest for this class.

Random forests

Decision trees tend to overfit training data, *i.e.* they have a low bias but a high variance. Therefore, random forests consist of several decision trees where each decision tree is trained on a random subset of the training data (referred to as *tree bagging*). The estimated posterior probability is typically computed as the average over all trees:

$$\hat{P}(l|f) = \frac{1}{T}\sum_{t=1}^{T}\hat{P}_t(l|f), \tag{2.4}$$

where l denotes the label, f denotes the feature vector and T equals the number of trees. To further reduce variance by decorrelating the trees, only a subset of the features is randomly chosen at each tree node.

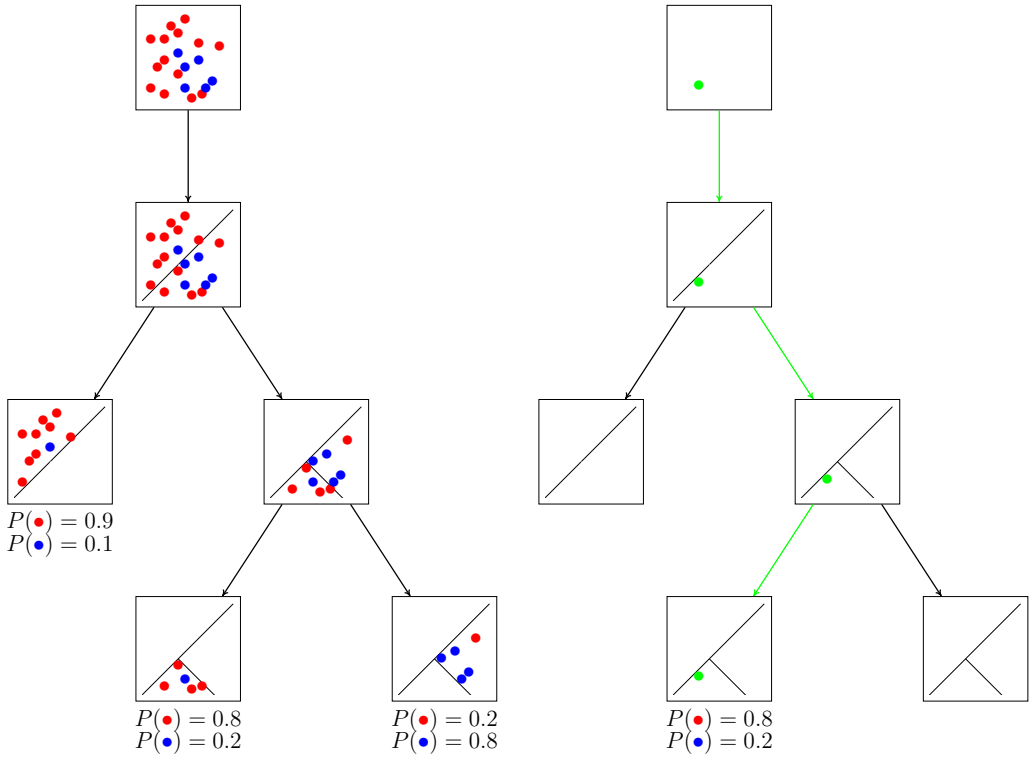


Figure 2.4: *Example of a binary decision tree consisting of six nodes; one root node, two decision nodes and three leaf nodes. (a) The decision tree is trained on 20 data points belonging to two different classes, "red" and "blue". For each data point, two different features have been computed. The two decision nodes (containing splitting functions equaling separating hyperplanes) are trained to divide the data into three different distributions (the leaf nodes). Each leaf node provides a posterior distribution over the classes for test data points ending up in that particular leaf node. (b) Features for an unlabelled data point (green) are computed and the data point is passed along the decision tree according to the splitting functions. The unlabelled data point ends up in the middle leaf node and is thus classified as "red".*

2.3.2 Convolutional neural networks

Convolutional neural networks (CNNs) constitute a class of machine learning tools for *e.g.* classification in image, video and natural language processing. Despite being introduced already in the 70s [71] by the name "Neocognitron", CNNs have received a great deal of attention from the image analysis and computer vision research community the last decade. The popularity stems from recent success on problems such as image classification [72] and object detection [73]. The success can predominantly be explained by an increased computational power of modern GPUs (Graphical Processing Units) and the access to large annotated datasets. Below follows a brief introduction to the technique, see the overview in [74] and the survey in [75] for more details.

Due to their ability to learn complex connections between input and output data, CNN-based methods have also been successfully applied to image segmentation tasks. In particular, so called *fully convolutional networks* [76–79] tend to produce results excelling at a variety of segmentation problems. Due to the promising results, CNN-based segmentation methods have emerged in the field of medical image analysis as well. So far, CNNs have been applied to *e.g.* breast electron microscopy images [80], knee MRI [81], abdominal CT [82] and brain MRI [79, 83, 84].

Architecture

CNNs are feed-forward artificial networks consisting of trailing computational layers where connections enable the result from one layer to be forwarded to a subsequent layer for further processing. CNNs are so called *universal function approximators*, *i.e.* the they are (in theory) able to model any function. To enable this capacity, CNNs contain thousands or millions of parameters that are automatically learned during training. In contrast to other image classification algorithms, pre-processing of the input data is typically not required when using CNNs; any needed image processing is learned automatically.

CNNs consist of one input layer, one output layer and one or more hidden layers. In CNNs constructed for classification or segmentation problems, the output layer typically equals conditional probabilities over predefined object classes, *cf.* the output of random decision forests in Section 2.3.1. For image classification problems, the CNN input usually equals the entire image to be classified and outputs a likelihood for image subjects, *e.g.* whether the image depicts a dog, a cat or a horse. Similar CNNs constructed for voxelwise classification rather take a smaller patch, centered at the voxel to be classified, as input and output a label likelihood for that specific voxel. So called fully convolutional networks can handle all input sizes; depending on the size of the input, the output is either label likelihoods for an entire image, for a smaller patch or for a single voxel.

The purpose of the hidden layers is to map the given input to the desired output. To enable modeling of any complex function, the hidden layers contain several different building blocks such as sets of learnable filters (convolutional layers), downsampling layers (pooling layers) and decision functions (nonlinear activation functions). Typically, CNNs consist of a set of trailing convolutional layers terminated with nonlinearities and layered with pooling layers. However, there are numerous proposed architectures in the literature. It is generally assumed that networks containing many small convolutional layers (*i.e.* deep networks) are more likely to produce good results than networks containing a few large convolutional layers (*i.e.* wide, shallow networks), but the findings so far are inconclusive [85].

Convolutional layers. The purpose of the convolutional layers is to extract image characteristics such as blobs, corners, lines *etc.* with means of automatically learned filters. Depending on the depth and width of the network, *i.e.* the amount of subsequent layers and their size, the learned filters may be able to recognize more complex features such as *e.g.* human faces. In contrast to hand-crafted feature detectors such as Sift or Surf, the CNN filter weights are automatically learned during training and thus not designed with any prior knowledge in mind. The convolutional property enables translation invariance, *i.e.* each region of the image is processed in the exact same manner.

Pooling layers. The pooling layers aim to downsample the image (and subsequent filter responses) in order to reduce the parameter space preventing undesired effects such as overfitting and unnecessary high computational complexity. A common choice of pooling is so called max-pooling, *i.e.* applying a maximum (dilation) filter. Note that pooling layers in principle equal convolutional layers with fixed (non-learnable) filter weights.

Nonlinear activation functions. Nonlinearities are important to enable the universal function approximator property; using only linear combinations of convolutional layers would enable nothing but linear maps from input to output. The nonlinearities also restrict unbounded layer outputs to a certain range, and thus help avoiding an accumulation of large values in some sections of the network. There is a wide selection of activation functions such as the rectified linear units (ReLUs) [86], sigmoid units (rarely used in practice), tanh units and Maxout units [87]. The nonlinear softmax unit, mapping arbitrary numbers to probabilities, is particularly useful in the output layer of classification/segmentation networks.

Fully connected layers. Before the output layer, there are sometimes one, two or more fully connected layers. Standard CNNs used for *e.g.* image classification include fully connected layers, while fully convolutional networks do not. The fully connected layers aim to map a large set of multidimensional filter responses to a more manageable 1D histogram. For instance, a CNN constructed for distinguishing two image classes typically terminates with fully connected layers mapping the filter responses to a histogram of size two. Applying the softmax operator to this histogram gives a conditional probability estimate for the two classes.

Fully convolutional networks. Standard CNNs (using fully connected layers) are not particularly efficient when dealing with voxelvise classification tasks such as segmentation; these networks can not be trained on nor be applied to images of arbitrary sizes. Moreover, the fully connected layers omit spatial relationships and are computationally demanding. However, another class of networks, fully convolutional networks, is better suited for segmentation tasks. Fully convolutional networks drop the terminating fully connected layers. Instead, they solely use convolutional layers for filtering, downsampling, upsampling and "defiltering" the image. These networks are capable of processing images of arbitrary sizes, and they are computationally more efficient than their fully connected counterparts. The output is typically label likelihoods over an image of the same size as the input.

Training

CNNs are trained using local optimization methods, common choices are stochastic gradient descent or mini-batch gradient descent combined with adaptive learning rate, batch normalization [88] and/or Nesterov's momentum [89]. Despite complex architectures and a huge amount of parameters, the gradients can be efficiently computed using the backpropagation algorithm, first proposed in [90–92]. Training is done in epochs, where all training samples are utilized in each epoch. For networks using a terminating softmax unit, voxelwise cross-entropy is used as objective function. Another choice of objective function is the max-margin hinge loss allowing for a support vector machine (SVM) classifier.

Due to the large amount of learnable parameters, an important consideration during training is to prevent overfitting. There are several techniques for this, *e.g.* dropout [93], artificially augmented data sets (to increase the amount of training data), filter weight regularization and early stopping [94]. When faced with a new classification/segmentation task, it can be beneficial to use a pre-trained CNN, espcially if training data is limited. Pre-training can be done either using other (preferably similar) datasets or with means of unsupervised training as in [95]. Pre-training facilitates learning by enabling the network to re-use filters that have already learned to recognize certain features.

2.4 Conditional random fields

Conditional random fields (CRFs), a variant of Markov random fields (MRFs) [96–98], is a class of probabilistic graphical models suitable for modeling spatial context such as smooth segmentation boundaries, coherent shapes *etc*. CRFs may be regarded as implicit shape models; they do not directly enforce an explicit (parameterized) shape model but still encourage spatial smoothness between neighbouring voxels. By also considering the classification of neighbours when assigning a label to a voxel, noisy or implausible boundaries can be avoided. CRFs have successfully been used in multi-atlas frameworks, for *e.g.* label fusion or postprocessing, in chest radiographs [99], knee MRI [14], abdominal CT [100] and brain MRI [10, 101, 102].

When using CRFs for computing segmentations, the labelling problem is posed as an optimization problem that is solved either exactly (if possible) or approximately. More specifically, the image is regarded as an observation of a conditional random field and the labelling (*i.e.* the realization of the field) is inferred by solving an energy minimization problem.

2.4.1 Mathematical model

Let $l_p \in \mathcal{L}$ be a variable indicating what class a voxel, indexed by $p \in \mathcal{P}$, is assigned to and let $i_p \in \mathcal{I}$ denote its observed intensity fo the voxel. Here, $\mathcal{I}$ denotes the image, $\mathcal{L}$ denotes the labelling and $\mathcal{P}$ denotes the set of all voxel indices. The optimal segmentation is inferred as the labelling that maximizes the posterior probability given by

$$P(\mathcal{L} \mid \mathcal{I}; \boldsymbol{\theta}) = \frac{1}{Z} e^{-E(\mathcal{L}, \mathcal{I}; \boldsymbol{\theta})}, \tag{2.5}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \ldots)$ are tunable parameters and Z is the partition function (*i.e.* the normalizing constant). The parameters are either fixed (*e.g.* derived by prior assumptions) or learned during training.

In most image applications, the energy E is assumed to decompose over unary and pairwise potentials. If so, the energy can be expressed as

$$E(\mathcal{L}, \mathcal{I}; \boldsymbol{\theta}) = \sum_{p \in \mathcal{P}} \phi_p(l_p, \mathcal{I}; \boldsymbol{\theta}) + \sum_{(p,q) \in \mathcal{N}} \phi_{p,q}(l_p, l_q, \mathcal{I}; \boldsymbol{\theta}), \tag{2.6}$$

where the set of all pairwise neighbours is denoted as $\mathcal{N}$. The unary potential ϕ_p may also be referred to as the unary cost, unary energy or the data cost. Similarly, the pairwise potential $\phi_{p,q}$ may be referred to as the pairwise cost, pairwise energy or regularization/coherence cost. In some applications, it may be beneficial to include potentials of higher orders (*i.e.* cliques including three or more neighbours), as in [102].

The neighbourhood of a voxel is defined by the voxel connectivity. In 3D applications, common choices are 6-connectivity (neighbours are defined by connected faces), 18-connectivity (neighbours are defined by connected faces and edges) or 26-connectivity (neighbours are defined by connected faces, edges and corners). However, larger neighbourhoods are also allowed. Further, one may incorporate the distance between voxels directly in the potentials, letting the pairwise energy depend smoothly on voxel distances (dense CRFs). If so, the second term in Equation (2.6) is summarized over all possible voxel combinations.

The unary cost, also known as the data cost, is usually dependent on conditional probabilities learned from data, such as the label likelihoods computed by *e.g.* a multi-atlas voting map or a machine learning classifier. A typical choice is

$$\phi_p = \theta_1 \log(\hat{P}(l_p \mid \mathcal{I})), \tag{2.7}$$

where $\hat{P}(l_p \mid \mathcal{I})$ equals the previously estimated likelihood (*i.e.* the normalized voting map or the classifier output).

The pairwise cost is an interaction term that regularizes the solution. In the simplest case, the pairwise costs are set to a fixed constant for all neighbours assigned with different labels, neighbours with the same labels are not penalized. This is called a *Potts model*:

$$\phi_{p,q} = \mathbb{1}_{l_p \neq l_q} \theta_2, \tag{2.8}$$

where $\mathbb{1}_{l_p \neq l_q}$ denotes the indicator function equaling one if $l_p \neq l_q$, *i.e.* if the neighbours are assigned different labels. However, more complex pairwise potentials taking the neighbouring intensities into account as well are usually beneficial. A common choice of the pairwise energy, consisting of two terms both penalizing neighbouring voxels being labelled differently, is given by

$$\phi_{p,q} = \mathbb{1}_{l_p \neq l_q} \left(\theta_2 + \theta_3 e^{-d(i_p, i_q)} \right), \tag{2.9}$$

where $d(\cdot, \cdot)$ is a metric measuring *e.g.* the contrast of the neighbouring voxels.

Unfortunately, the choices of pairwise interaction term in Equations (2.8) and (2.9) may lead to a bias towards shorter segmentation boundaries, *i.e.* a *shrinking bias*. However, there are several proposed solutions in the literature, *cf.* [103, 104]

2.4.2 Inference

A function on the form in Equation (2.6) can be formulated as a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ is the set of nodes (*i.e.* voxels) and $\mathcal{E}$ is the set of edges connecting neighbouring voxels. If the segmentation problem is binary and if the energy in Equation (2.6) is submodular, the globally optimal labelling can be computed exactly and in polynomial time using graph cuts [105]. Otherwise, methods such as alpha expansion [106], mean field inference or linear programming relaxations may be used to solve the minimization problem approximately and thus infer the labelling.

Chapter 3

Thesis contribution

As detailed in Section 1.1: excellent medical segmentation algorithms are characterized by speed, allowing for scalability, and an accuracy comparable to an expert radiologist. They should produce plausible organ (or region) shapes while generalizing well to unseen and rarely occurring anatomies. Preferably, training the segmentation algorithm should be data-efficient since manually labelled data typically is scarce in the medical community. Thus, these are all aspects considered in the included papers:

Paper I mainly concerns speeding up the image registration procedure.

Paper II mainly concerns improving machine learning techniques for voxel classification taking accuracy and data-efficiency into account.

Paper III mainly concerns increasing the image registration accuracy.

Paper IV mainly concerns improving label fusion taking plausible organ shapes into account.

This chapter is structured as follows: each section constitutes an overview of one out of the four included thesis papers. The sections provide summaries of the main algorithmic contributions as well as schematic visualizations of each paper's version of the multi-atlas pipeline, *cf*. Figure 1.2. Also, the contributions of the thesis author are stated for each paper respectively.

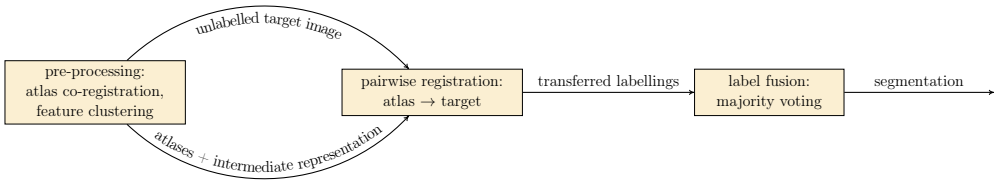


Figure 3.1: *Schematic summary of the multi-atlas framework in Paper I.*

3.1 Paper I

J. Alvén, A. Norlén, O. Enqvist and F. Kahl. "Überatlas: Fast and Robust Registration for Multi-Atlas Segmentation". *Pattern Recognition Letters,* 80:245–255, 2016.

Multi-atlas segmentation has the disadvantage of requiring multiple atlas registrations to capture the full range of possible anatomical variation. In general, image registration is computationally heavy which consequently limits the practical size of the atlas set. To speed up the registration procedure, and thus allowing for larger atlas sets, the paper proposes an intermediate representation of the atlas set. The intermediate representation consists of feature points that are similar and consistently detected throughout the atlas set. This intermediate representation may be used for simultaneously finding point correspondences and affine transformations to a target image from an arbitrarily large set of atlas images.

The main idea is to cluster extracted feature points from the atlas set to form the intermediate representation. To make sure the feature points in a cluster describe the same anatomical feature, the clustering procedure takes both descriptor distances and spatial distances (according to an offline spatial co-registration of the atlases) into account. At running time, one only needs to register the target image once, and point correspondences to all images in the atlas set are automatically obtained. Once good point correspondences are obtained to all the atlases, one can quickly and robustly compute an affine transformation for each atlas individually.

For a schematic overview of all steps included in the framework, see Figure 3.1.

Author contribution. I implemented most of the framework including the clustering algorithm, atlas co-registration and the iteratively reweighted least squares algorithm. I also carried out all the Elastix experiments. The remainder of the implementations, experiments as well as the writing were joint work. All authors contributed to the main idea.

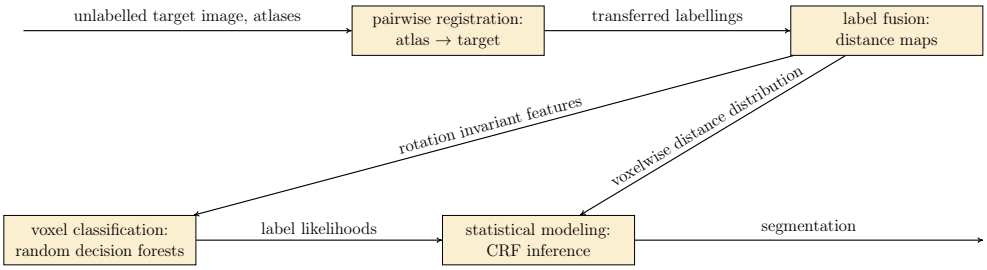


Figure 3.2: *Schematic summary of the multi-atlas framework in Paper II.*

3.2 Paper II

A. Norlén, J. Alvén, D. Molnar, O. Enqvist, R. Rossi Norrlund, J. Brandberg, G. Bergström and F. Kahl. "Automatic Pericardium Segmentation and Quantification of Epicardial Fat from Computed Tomography Angiography". *Journal of Medical Imaging,* 3(3), 2016.

For some applications, standard multi-atlas segmentation without refinement works rather poorly and serves as a decent initialization for a local boundary detector. One such example is segmentation of the pericardium, which is merely visible in CTA scans. Local classification of voxels based on machine learning techniques may help improve the results. Though, machine learning tools are dependent on large sets of labelled data, which are rarely occurring in medical applications. The paper addresses the problem of overcoming a shortage of labelled data when applying a random forest classifier to pericardium segmentation.

The primary algorithmic contribution of this paper is the incorporation of a generalized formulation of multi-atlas segmentation based on distance maps into a random forest classification framework. More specific, transferred atlas labellings define a voxelwise distribution over distances to the organ boundary. This distribution is utilized in two manners. Firstly, it serves as a global initialization for the organ boundary search space. Secondly, it provides a local coordinate system enabling alignment of extracted features to the organ boundary. Rotation invariant features greatly simplify the voxel classification task (reducing the 3D boundary detection problem to 1D line search) but also normalize the training data leading to more efficient use of the labelled data set. In this manner, the random decision forest classifier learns recognizing organ boundaries irrespective of the orientation relative the image coordinate axes.

For a schematic overview of all steps included in the framework, see Figure 3.2.

Author contribution. I carried out all baseline experiments and contributed with some ideas. Norlén carried out most of the algorithm implementations. The rest of the experiments as well as the writing were joint work. Norlén, Enqvist and Kahl proposed the main idea.

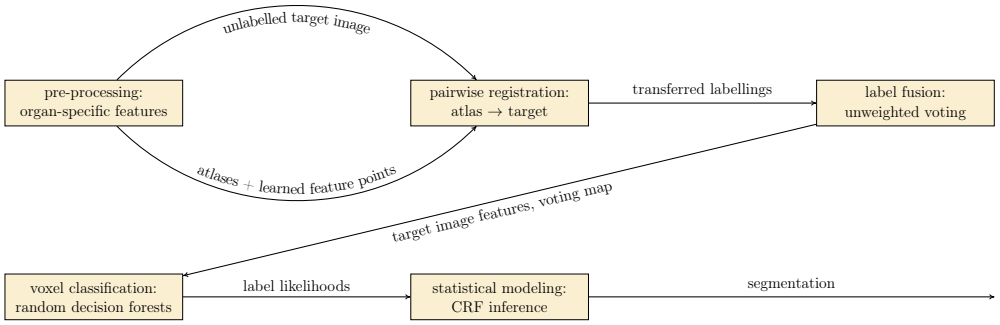


Figure 3.3: *Schematic summary of the multi-atlas framework in Paper III.*

3.3 Paper III

F. Fejne, M. Landgren, J. Alvén, J. Ulén, J. Fredriksson, V. Larsson and F. Kahl. "Multi-atlas Segmentation Using Robust Feature-Based Registration". In *Cloud-Based Benchmarking of Medical Image Analysis,* Springer International Publishing, 203–218, 2017.

For a successful multi-atlas segmentation, one needs to register the atlas images to the target image as accurately as possible. As detailed in Section 2.2.1, there are two different approaches to image registration. Intensity-based methods are popular methods for medical applications, but lack speed and risk producing sub-optimal solutions. Sparse feature-based methods are faster and more robust, but have failed to gain popularity in medical image analysis, mainly due to the difficulty in detecting distinctive feature points and thereby establishing correct one-to-one point correspondences, leading to a high rate of outlier matches.

To improve the establishment of point-to-point correspondences, and thus allowing for feature-based registration methods to be applied to medical applications, the paper proposes using only a subset of the detected atlas feature points. These feature points are organ-specific and sifted out during an offline pre-processing step. The selected feature points should be likely to give inlier matches based on residual errors measured by offline co-registration of the atlases. Thus, this adapted feature-based method reduces the risk of establishing incorrect point-to-point correspondences by reliably identifying organ-specific feature points among the atlas images.

For a schematic overview of all steps included in the framework, see Figure 3.3.

Author contribution. Implementations, experiments as well as the writing were joint work. I mostly contributed to the implementation of the modified feature-based registration method and to the writing of the book chapter. Kahl proposed the main idea.

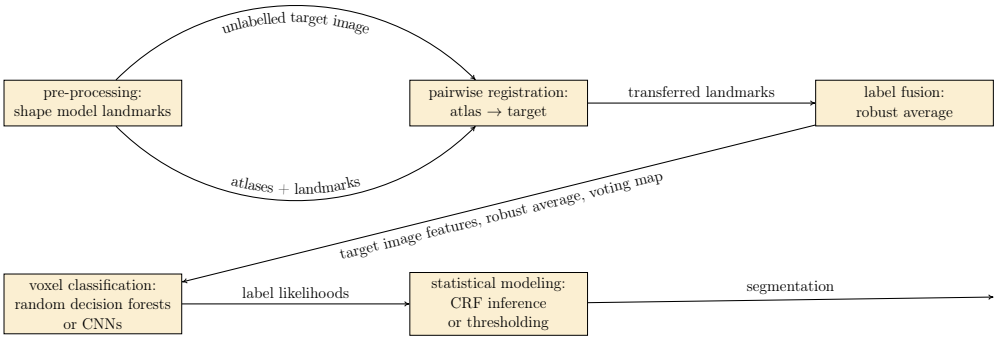


Figure 3.4: *Schematic summary of the multi-atlas framework in Paper IV.*

3.4 Paper IV

J. Alvén, F. Kahl, M. Landgren, V. Larsson, J. Ulén and O. Enqvist. "Shape-Aware Label Fusion for Multi-Atlas Frameworks". Submitted to *Pattern Recognition Letters*.

Good segmentation algorithms should generalize well to unseen or rarely occurring anatomies while still producing plausible organ (or region) shapes. Fortunately, multi-atlas frameworks tend to generalize well. However, traditional multi-atlas label fusion puts no explicit constraints on the output shape. On the contrary, standard label fusion combines transferred labels locally by merely considering the current voxel and/or spatially neighbouring voxels. In order to guarantee a preserved topology and to prevent disjoint organ shapes or lost structures, one needs to include global shape regularization. Unfortunately, most methods with explicit shape constraints fail generalizing as well as multi-atlas methods do.

This paper incorporates a shape prior into label fusion without losing the generalizability of multi-atlas methods. Instead of fusing the labels at the voxel level, each transferred labelling is regarded as a shape model estimate. The shape model is a point distribution model of the organ surface consisting of landmark correspondences established offline. Online, pairwise registrations provide coordinate estimates for these landmarks in the target image. These estimates are used for computing an average shape by using robust optimization techniques. In this manner, an awareness of the overall shape is directly incorporated into the label fusion preventing implausible results while keeping robustness to outlier registrations.

For a schematic overview of all steps included in the framework, see Figure 3.4.

Author contribution. Implementations, experiments as well as the writing were joint work. I mostly contributed to (i) implementations related to CNNs and establishing landmarks, (ii) running the experiments and (iii) the writing of the paper. I, Kahl and Enqvist proposed the main idea.

Chapter 4

Concluding discussion

This thesis addresses three major research questions:

(i) How can we improve performance and precision of medical segmentation algorithms in order to meet the requirements on timing and accuracy posed by *e.g.* computer-aided diagnosis and surgery as well as medical research?

(ii) How can we guarantee anatomically meaningful segmentation results while still allowing for generalizability and scalability?

(iii) How can we reduce the reliance on access to large sets of manually labelled data when developing competitive segmentation methods?

Below follows a brief discussion regarding how successful the included papers are in tackling these questions as well as a proposal of three future research topics each addressing one of these problems respectively.

4.1 Discussion

In the matter of segmentation accuracy, included papers seemingly manage to meet the objectives; each paper presents algorithms performing better or on par with compared methods. However, these conclusions need to be seen in the light of the difficulties of objectively evaluating and comparing medical segmentation algorithms; choice of similarity metrics, evaluation data and tuning parameters may greatly impact the results and thereby drawn conclusions.

Merely one paper out of four, Paper III, evaluates the proposed algorithm directly on unseen testing data provided by a public benchmark (the VISCERAL dataset), enabling online comparisons to competing methods. Thus, unbiased conclusions regarding performance can be drawn directly from a benchmark leaderboard containing pre-determined accuracy metrics.

Two papers out of the remaining three, Paper I and Paper IV, do validate the proposed methods on publicly accessible data (the Visceral and the Hammers datasets); however, evaluation is done by splitting the provided training data into smaller sets and by running public or re-implemented versions of the competing methods. Re-running or re-implementing competing methods may lead to (perhaps) sub-optimal design choices merely following recommendations in corresponding papers, *e.g.* in the presence of tunable hyperparameters. Moreover, implementing and/or installing, tuning and running baseline methods are time-consuming and implementations of current state-of-the-art may not even be publicly accessible. Due to this, only a fraction of previously published methods are used for comparison, which of course is crippling for evaluations said to be meticulous.

Objective comparisons were particularly challenging for Paper II; merely one dataset (Scapis) is considered due to the very task-specific objective (delineating the pericardium). Unfortunately, neither datasets nor implemented versions of previously proposed methods for pericardium segmentation are (as of date) publicly available. Further, these difficulties highlights the advantages of general-purpose algorithms independent of application and modality; only considering one specific segmentation task makes comparisons highly inconvenient in the absence of benchmark databases.

One paper out of four, Paper I, addresses running time as an explicit research objective. The paper does succeed in speeding up parts of the multi-atlas framework, however, some time-consuming steps are heavily overlooked (*e.g.* image warping and non-rigid registration). Implicitly, all four papers more or less address running times by successfully using fast feature-based registration previously assumed to be unfit for medical applications. Various techniques for improving unique point-to-point matches are proposed; feature clustering (Paper I), boundary-proximate features (Paper II) and organ-specific features (Paper III). However, it would surely be informative to compare these different techniques for establishing reliable point correspondences more rigorously.

Regarding segmentation plausibility, Paper IV does present a convincing evaluation of the qualitative shape with means of two different evaluation metrics. However, qualitative shape is highly subjective and thus difficult to quantify; for example a thorough visual inspection by a medical expert would benefit the comparison. For some cases, the shape prior did seem to impact the segmentation negatively, leading to over-regularized boundaries, which may infuse doubt regarding the generalizability. Additionally, one may dispute the choice of merely including the shape-regularized segmentation as input to a classifier, and not directly enforcing the refined solution to cohere with the shape prior. In retrospect, one could consider executing a comparison to a similar classifier merely trained on the image.

Finally, limited access to labelled data is obviously an issue in all included papers; the Scapis and Hammers datasets consist of 30 manually annotated images while the Visceral dataset consists of 20 training atlases. Although standard solutions such as cross-validation and data augmentation are included in several papers, merely one paper (Paper II) addresses the concern explicitly. However, it remains unclear to what extent the paper manages to increase the data-efficiency.

4.2 Future directions

Deep learning in multi-atlas frameworks. The recent success of end-to-end deep learning in various research fields, such as self-driving vehicles [107], robotics [108] and speech recognition [109], shows that deep algorithms are indeed capable of producing promising results for applications posing great demands on accuracy, online running time and/or safety; requirements not quite different from medical applications.

Inspired by these efforts, a deep multi-atlas framework trained end-to-end would surely be capable of producing competetive results as well as attracting interest from the medical image analysis community. So far, various researchers have proposed deep learning-based methods for feature detection, description and matching [46–48], 2D image registration [110] as well as label fusion [64]. Future papers that extend these, or similar, ideas to 3D and that assemble the pieces into a deep multi-atlas framework, enabling end-to-end training, would likely revolutionize the field of multi-atlas segmentation. Again, one bottleneck of this approach would definitely be the lack of labelled medical data.

Geometric priors in CRF inference. Conditional random field models are useful for posing implicit shape constraints on the output segmentation while still generalizing well to unseen data. Moreover, recent advances in the field of deep learning enables end-to-end training of CNN segmentation networks coupled with CRFs, *e.g.* [111, 112].

However, there are several medical segmentation tasks that could benefit from enforcing anatomical constraints more definitely, *e.g.* relative position or shape topology. Previous work has successfully included such constraints in CRF frameworks, *e.g.* star-shaped or convex shapes as in [113–115] or relative position of multiple regions as in [116, 117]. Enabling end-to-end training of networks coupled with CRFs enforcing these geometric priors is yet to be done, and will surely boost the qualitative performance of deep segmentation algorithms.

Weakly supervised learning. Most segmentation frameworks, among them standard multi-atlas segmentation, require completely annotated 3D volumes in order to produce meaningful results. However, using deep segmentation methods, such as fully convolutional networks, opens up for the possibility of training competitive algorithms on partially annotated 3D volumes. Recent weakly- and/or semi-supervised methods [118,119], propose using *e.g.* image level labels, bounding boxes and/or partially labelled images for training 2D segmentation algorithms. Successfully extending these ideas to medical 3D images would surely be received gratefully from the medical community.

Enabling training data consisting of partially annotated volumes would allow for, for instance, using automatically computed segmentations of inconsistent quality as ground truth, as well as using ground truth volumes consisting of a fraction of the amount of manually labelled slices required for producing a complete 3D labelling. However, future research following this intriguing direction will surely need to address novel problems, *e.g.* how shape constraints should be incorporated and learned using partially incomplete organ boundaries and how 3D segmentation algorithms should be evaluated reliably on incomplete data such as 2D slices.

Bibliography

[1] P. Suetens, *Fundamentals of medical imaging.* Cambridge University Press, 2017.

[2] G. Bergström, G. Berglund, A. Blomberg, J. Brandberg, G. Engström, J. Engvall, M. Eriksson, U. Faire, A. Flinck, M. G. Hansson *et al.*, "The Swedish cardiopulmonary bioimage study: Objectives and design," *Journal of Internal Medicine*, vol. 278, no. 6, pp. 645–659, 2015.

[3] O. A. Jimenez del Toro, H. Müller, M. Krenn, K. Gruenberg, A. A. Taha, M. Winterstein, I. Eggel, A. Foncubierta-Rodríguez, O. Goksel, A. Jakab *et al.*, "Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: Visceral anatomy benchmarks," *IEEE Transactions on Medical Imaging*, vol. 35, no. 11, pp. 2459–2475, 2016.

[4] A. Hammers, R. Allom, M. J. Koepp, S. L. Free, R. Myers, L. Lemieux, T. N. Mitchell, D. J. Brooks, and J. S. Duncan, "Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe," *Human Brain Mapping*, vol. 19, no. 4, pp. 224–247, 2003.

[5] I. S. Gousias, D. Rueckert, R. A. Heckemann, L. E. Dyet, J. P. Boardman, A. D. Edwards, and A. Hammers, "Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest," *NeuroImage*, vol. 40, no. 2, pp. 672–684, 2008.

[6] R. H. Taylor, A. Menciassi, G. Fichtinger, and P. Dario, "Medical robotics and computer-integrated surgery," in *Handbook of Robotics.* Springer, 2008, pp. 1199–1222.

[7] D. L. Pham, C. Xu, and J. L. Prince, "Current methods in medical image segmentation 1," *Annual Review of Biomedical Engineering*, vol. 2, no. 1, pp. 315–337, 2000.

[8] 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.

[9] T. Rohlfing and C. R. Maurer, "Shape-based averaging," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 153–161, 2007.

[10] F. van der Lijn, T. den Heijer, M. M. Breteler, and W. J. Niessen, "Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts," *NeuroImage*, vol. 43, no. 4, pp. 708–720, 2008.

[11] M. Chupin, E. Gérardin, R. Cuingnet, C. Boutet, L. Lemieux, S. Lehéricy, H. Benali, L. Garnero, O. Colliot, Alzheimer's Disease Neuroimaging Initiative *et al.*, "Fully automatic hippocampus segmentation and classification in alzheimer's disease and mild cognitive impairment applied on data from ADNI," *Hippocampus*, vol. 19, no. 6, p. 579, 2009.

[12] R. A. Heckemann, S. Keihaninejad, P. Aljabar, D. Rueckert, J. V. Hajnal, A. Hammers, Alzheimer's Disease Neuroimaging Initiative *et al.*, "Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation," *NeuroImage*, vol. 51, no. 1, pp. 221–227, 2010.

[13] S. Parisot, W. Wells, S. Chemouny, H. Duffau, and N. Paragios, "Uncertainty-driven efficiently-sampled sparse graphical models for concurrent tumor segmentation and atlas registration," in *IEEE International Conference on Computer Vision*, 2013, pp. 641–648.

[14] J.-G. Lee, S. Gumus, C. H. Moon, C. K. Kwoh, and K. T. Bae, "Fully automated segmentation of cartilage from the MR images of knee using a multi-atlas and local structural analysis method," *Medical Physics*, vol. 41, no. 9, 2014.

[15] R. Shahzad, D. Bos, C. Metz, A. Rossi, H. Kirişli, A. van der Lugt, S. Klein, J. Witteman, P. de Feyter, W. Niessen *et al.*, "Automatic quantification of epicardial fat volume on non-enhanced cardiac CT scans using a multi-atlas segmentation approach," *Medical Physics*, vol. 40, no. 9, 2013.

[16] X. Ding, D. Terzopoulos, M. Diaz-Zamudio, D. S. Berman, P. J. Slomka, and D. Dey, "Automated pericardium delineation and epicardial fat volume quantification from noncontrast CT," *Medical Physics*, vol. 42, no. 9, pp. 5015–5026, 2015.

[17] H. A. Kirisli, M. Schaap, S. Klein, L. A. Neefjes, A. C. Weustink, T. Van Walsum, and W. J. Niessen, "Fully automatic cardiac segmentation from 3D CTA data: A multi-atlas based approach," in *Proceedings of SPIE: Medical Imaging*, 2010.

[18] J. V. Spearman, F. G. Meinel, U. J. Schoepf, P. Apfaltrer, J. R. Silverman, A. W. Krazinski, C. Canstein, C. N. De Cecco, P. Costello, and L. L. Geyer, "Automated quantification of epicardial adipose tissue using CT angiography: Evaluation of a prototype software," *European Radiology*, vol. 24, no. 2, pp. 519–526, 2014.

[19] X. Han, M. Hoogeman, P. Levendag, L. Hibbard, D. Teguh, P. Voet, A. Cowen, and T. Wolf, "Atlas-based auto-segmentation of head and neck CT images," in *Conference on Medical Image Computing and Computer-Assisted Intervention*, 2008, pp. 434–441.

[20] L. Wang, K. C. Chen, Y. Gao, F. Shi, S. Liao, G. Li, S. G. Shen, J. Yan, P. K. Lee, B. Chow *et al.*, "Automated bone segmentation from dental CBCT images using patch-based sparse representation and convex optimization," *Medical Physics*, vol. 41, no. 4, 2014.

[21] C. Chu, M. Oda, T. Kitasaka, K. Misawa, M. Fujiwara, Y. Hayashi, Y. Nimura, D. Rueckert, and K. Mori, "Multi-organ segmentation based on spatially-divided probabilistic atlas from 3D abdominal CT images," in *Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 165–172.

[22] G. Sanroma, G. Wu, Y. Gao, and D. Shen, "Learning-based atlas selection for multiple-atlas segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 3111–3117.

[23] 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.

[24] 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.

[25] R. Wolz, C. Chu, K. Misawa, M. Fujiwara, K. Mori, and D. Rueckert, "Automated abdominal multi-organ segmentation with subject-specific atlas generation," *IEEE Transactions on Medical Imaging*, vol. 32, no. 9, pp. 1723–1730, 2013.

[26] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, 2004.

[27] A. Klein, B. Mensh, S. Ghosh, J. Tourville, and J. Hirsch, "Mindboggle: Automated brain labeling with multiple atlases," *BMC Medical Imaging*, vol. 5, no. 1, p. 7, 2005.

[28] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[29] F. Khalifa, G. M. Beache, G. Gimel'farb, J. S. Suri, and A. S. El-Baz, "State-of-the-art medical image registration methodologies: A survey," in *Multi Modality State-of-the-art Medical Image Segmentation and Registration Methodologies*. Springer, 2011, pp. 235–280.

[30] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[31] Y. Ou, A. Sotiras, N. Paragios, and C. Davatzikos, "DRAMMS: Deformable registration via attribute matching and mutual-saliency weighting," *Medical Image Analysis*, vol. 15, no. 4, pp. 622–639, 2011.

[32] S. Ourselin, A. Roche, G. Subsol, X. Pennec, and N. Ayache, "Reconstructing a 3D structure from serial histological sections," *Image and Vision Computing*, vol. 19, no. 1, pp. 25–31, 2001.

[33] S. Ourselin, R. Stefanescu, and X. Pennec, "Robust registration of multimodal images: Towards real-time clinical applications," in *Conference on Medical Image Computing and Computer-Assisted Intervention*, 2002, pp. 140–147.

[34] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, 2009.

[35] S. Klein, M. Staring, K. Murphy, M. A. Viergever, and J. P. Pluim, "Elastix: A toolbox for intensity-based medical image registration," *IEEE Transactions on Medical Imaging*, vol. 29, no. 1, pp. 196–205, 2010.

[36] B. B. Avants, N. J. Tustison, M. Stauffer, G. Song, B. Wu, and J. C. Gee, "The insight toolkit image registration framework," *Frontiers in Neuroinformatics*, vol. 8, 2014.

[37] J. P. Pluim, J. A. Maintz, and M. A. Viergever, "Mutual-information-based registration of medical images: A survey," *IEEE Transactions on Medical Imaging*, vol. 22, no. 8, pp. 986–1004, 2003.

[38] F. Maes, D. Vandermeulen, and P. Suetens, "Comparative evaluation of multiresolution optimization strategies for multimodality image registration by maximization of mutual information," *Medical Image Analysis*, vol. 3, no. 4, pp. 373–386, 1999.

[39] S. Klein, M. Staring, and J. P. Pluim, "Evaluation of optimization methods for nonrigid medical image registration using mutual information and B-splines," *IEEE Transactions on Image Processing*, vol. 16, no. 12, pp. 2879–2890, 2007.

[40] J.-P. Thirion, "Image matching as a diffusion process: An analogy with Maxwell's demons," *Medical Image Analysis*, vol. 2, no. 3, pp. 243–260, 1998.

[41] F. L. Bookstein, "Principal warps: Thin-plate splines and the decomposition of deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, no. 6, pp. 567–585, 1989.

[42] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes, "Nonrigid registration using free-form deformations: Application to breast MR images," *IEEE Transactions on Medical Imaging*, vol. 18, no. 8, pp. 712–721, 1999.

[43] L. Svärm, O. Enqvist, F. Kahl, and M. Oskarsson, "Improving robustness for inter-subject medical image registration using a feature-based approach," in *International Symposium on Biomedical Imaging*, 2015, pp. 824–828.

[44] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[45] 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.

[46] P. Sermanet, D. Eigen, X. Zhang, M. Mathieu, R. Fergus, and Y. LeCun, "Overfeat: Integrated recognition, localization and detection using convolutional networks," *arXiv preprint arXiv:1312.6229*, 2013.

[47] E. Simo-Serra, E. Trulls, L. Ferraz, I. Kokkinos, P. Fua, and F. Moreno-Noguer, "Discriminative learning of deep convolutional feature point descriptors," in *IEEE International Conference on Computer Vision*, 2015, pp. 118–126.

[48] P. Fischer, A. Dosovitskiy, and T. Brox, "Descriptor matching with convolutional neural networks: A comparison to SIFT," *arXiv preprint arXiv:1405.5769*, 2014.

[49] M. A. Fischler and R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[50] S. Lee, G. Wolberg, and S. Y. Shin, "Scattered data interpolation with multilevel B-splines," *IEEE Transactions on Visualization and Computer Graphics*, vol. 3, no. 3, pp. 228–244, 1997.

[51] H. Chui and A. Rangarajan, "A new point matching algorithm for non-rigid registration," *Computer Vision and Image Understanding*, vol. 89, no. 2, pp. 114–141, 2003.

[52] P. J. Besl, N. D. McKay *et al.*, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 2, pp. 239–256, 1992.

[53] C. V. Stewart, C.-L. Tsai, and B. Roysam, "The dual-bootstrap iterative closest point algorithm with application to retinal image registration," *IEEE Transactions on Medical Imaging*, vol. 22, no. 11, pp. 1379–1394, 2003.

[54] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, 2009.

[55] T. R. Langerak, U. A. van der Heide, A. N. Kotte, M. A. Viergever, M. Van Vulpen, and J. P. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE Transactions on Medical Imaging*, vol. 29, no. 12, pp. 2000–2008, 2010.

[56] X. Artaechevarria, A. Muñoz-Barrutia, and C. Ortiz-de Solorzano, "Efficient classifier generation and weighted voting for atlas-based segmentation: Two small steps faster and closer to the combination oracle," *SPIE Medical Imaging: Image Processing*, vol. 6914, no. 3, pp. 69 141W–1, 2008.

[57] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de Solórzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1266–1277, 2009.

[58] 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.

[59] G. Wu, Q. Wang, D. Zhang, F. Nie, H. Huang, and D. Shen, "A generative probability model of joint label fusion for multi-atlas based brain segmentation," *Medical Image Analysis*, vol. 18, no. 6, pp. 881–890, 2014.

[60] Y. Song, G. Wu, Q. Sun, K. Bahrami, C. Li, and D. Shen, "Progressive label fusion framework for multi-atlas segmentation by dictionary evolution," in *Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 190–197.

[61] T. Rohlfing, D. B. Russakoff, and C. R. Maurer, "Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 8, pp. 983–994, 2004.

[62] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Transactions on Medical Imaging*, vol. 23, no. 7, pp. 903–921, 2004.

[63] M. R. Sabuncu, B. T. Yeo, K. Van Leemput, B. Fischl, and P. Golland, "A generative model for image segmentation based on label fusion," *IEEE Transactions on Medical Imaging*, vol. 29, no. 10, pp. 1714–1729, 2010.

[64] H. Yang, J. Sun, H. Li, L. Wang, and Z. Xu, "Deep fusion net for multi-atlas segmentation: Application to cardiac MR images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 521–528.

[65] L. Breiman, "Random forests," *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.

[66] A. Criminisi and J. Shotton, *Decision forests for computer vision and medical image analysis*. Springer Science & Business Media, 2013.

[67] X. Han, "Learning-boosted label fusion for multi-atlas auto-segmentation," in *International Workshop on Machine Learning in Medical Imaging*, 2013, pp. 17–24.

[68] E. Konukoglu, B. Glocker, D. Zikic, and A. Criminisi, "Neighbourhood approximation using randomized forests," *Medical Image Analysis*, vol. 17, no. 7, pp. 790–804, 2013.

[69] D. Zikic, B. Glocker, and A. Criminisi, "Encoding atlases by randomized classification forests for efficient multi-atlas label propagation," *Medical Image Analysis*, vol. 18, no. 8, pp. 1262–1273, 2014.

[70] 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*, 2014, pp. 256–263.

[71] K. Fukushima, "Neural network model for a mechanism of pattern recognition unaffected by shift in position- neocognitron," *Electron. & Commun. Japan*, vol. 62, no. 10, pp. 11–18, 1979.

[72] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems*, 2012, pp. 1097–1105.

[73] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[74] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[75] J. Schmidhuber, "Deep learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[76] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[77] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *IEEE International Conference on Computer Vision*, 2015, pp. 1520–1528.

[78] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 234–241.

[79] T. Brosch, L. Y. Tang, Y. Yoo, D. K. Li, A. Traboulsee, and R. Tam, "Deep 3D convolutional encoder networks with shortcuts for multiscale feature integration applied to multiple sclerosis lesion segmentation," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1229–1239, 2016.

[80] D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber, "Deep neural networks segment neuronal membranes in electron microscopy images," in *Advances in Neural Information Processing Systems*, 2012, pp. 2843–2851.

[81] A. Prasoon, K. Petersen, C. Igel, F. Lauze, E. Dam, and M. Nielsen, "Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network," in *Conference on Medical Image Computing and Computer-Assisted Intervention*, 2013, pp. 246–253.

[82] H. R. Roth, L. Lu, A. Farag, H.-C. Shin, J. Liu, E. B. Turkbey, and R. M. Summers, "Deeporgan: Multi-level deep convolutional networks for automated pancreas segmentation," in *Conference on Medical Image Computing and Computer-Assisted Intervention*, 2015, pp. 556–564.

[83] W. Zhang, R. Li, H. Deng, L. Wang, W. Lin, S. Ji, and D. Shen, "Deep convolutional neural networks for multi-modality isointense infant brain image segmentation," *NeuroImage*, vol. 108, pp. 214–224, 2015.

[84] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[85] Z. Wu, C. Shen, and A. v. d. Hengel, "Wider or deeper: Revisiting the resnet model for visual recognition," *arXiv preprint arXiv:1611.10080*, 2016.

[86] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *International Conference on Machine Learning*, 2010, pp. 807–814.

[87] I. J. Goodfellow, D. Warde-Farley, M. Mirza, A. Courville, and Y. Bengio, "Maxout networks," *arXiv preprint arXiv:1302.4389*, 2013.

[88] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *International Conference on Machine Learning*, 2015, pp. 448–456.

[89] I. Sutskever, J. Martens, G. Dahl, and G. Hinton, "On the importance of initialization and momentum in deep learning," in *International Conference on Machine Learning*, 2013, pp. 1139–1147.

[90] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural Computation*, vol. 1, no. 4, pp. 541–551, 1989.

[91] Y. LeCun, B. E. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. E. Hubbard, and L. D. Jackel, "Handwritten digit recognition with a backpropagation network," in *Advances in Neural Information Processing Systems*, 1990, pp. 396–404.

[92] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[93] N. Srivastava, G. E. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting." *Journal of Machine Learning Research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[94] L. Prechelt, "Early stopping – but when?" in *Neural Networks: Tricks of the Trade*. Springer, 2012, pp. 53–67.

[95] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[96] A. Blake, P. Kohli, and C. Rother, *Markov random fields for vision and image processing*. MIT Press, 2011.

[97] C. Sutton, A. McCallum *et al.*, "An introduction to conditional random fields," *Foundations and Trends ® in Machine Learning*, vol. 4, no. 4, pp. 267–373, 2012.

[98] C. Wang, N. Komodakis, and N. Paragios, "Markov random field modeling, inference & learning in computer vision & image understanding: A survey," *Computer Vision and Image Understanding*, vol. 117, no. 11, pp. 1610–1627, 2013.

[99] S. Candemir, S. Jaeger, K. Palaniappan, J. P. Musco, R. K. Singh, Z. Xue, A. Karargyris, S. Antani, G. Thoma, and C. J. McDonald, "Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration," *IEEE Transactions on Medical Imaging*, vol. 33, no. 2, pp. 577–590, 2014.

[100] C. Platero and M. C. Tobar, "A multiatlas segmentation using graph cuts with applications to liver segmentation in CT scans," *Computational and Mathematical Methods in Medicine*, vol. 2014, 2014.

[101] R. Wolz, R. A. Heckemann, P. Aljabar, J. V. Hajnal, A. Hammers, J. Lötjönen, D. Rueckert, Alzheimer's Disease Neuroimaging Initiative *et al.*, "Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI," *NeuroImage*, vol. 52, no. 1, pp. 109–118, 2010.

[102] C. Platero and M. C. Tobar, "A label fusion method using conditional random fields with higher-order potentials: Application to hippocampal segmentation," *Artificial Intelligence in Medicine*, vol. 64, no. 2, pp. 117–129, 2015.

[103] V. Kolmogorov and Y. Boykov, "What metrics can be approximated by geo-cuts, or global optimization of length/area and flux," in *IEEE International Conference on Computer Vision*, vol. 1, 2005, pp. 564–571.

[104] A. K. Sinop and L. Grady, "A seeded image segmentation framework unifying graph cuts and random walker which yields a new algorithm," in *IEEE International Conference on Computer Vision*, 2007, pp. 1–8.

[105] V. Kolmogorov and R. Zabin, "What energy functions can be minimized via graph cuts?" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 2, pp. 147–159, 2004.

[106] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.

[107] M. Bojarski, D. Del Testa, D. Dworakowski, B. Firner, B. Flepp, P. Goyal, L. D. Jackel, M. Monfort, U. Muller, J. Zhang *et al.*, "End to end learning for self-driving cars," *arXiv preprint arXiv:1604.07316*, 2016.

[108] S. Levine, C. Finn, T. Darrell, and P. Abbeel, "End-to-end training of deep visuomotor policies," *Journal of Machine Learning Research*, vol. 17, no. 39, pp. 1–40, 2016.

[109] D. Amodei, S. Ananthanarayanan, R. Anubhai, J. Bai, E. Battenberg, C. Case, J. Casper, B. Catanzaro, Q. Cheng, G. Chen *et al.*, "Deep speech 2: End-to-end speech recognition in english and mandarin," in *International Conference on Machine Learning*, 2016, pp. 173–182.

[110] A. Kanazawa, D. W. Jacobs, and M. Chandraker, "Warpnet: Weakly supervised matching for single-view reconstruction," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 3253–3261.

[111] A. G. Schwing and R. Urtasun, "Fully connected deep structured networks," *arXiv preprint arXiv:1503.02351*, 2015.

[112] S. Zheng, S. Jayasumana, B. Romera-Paredes, V. Vineet, Z. Su, D. Du, C. Huang, and P. H. Torr, "Conditional random fields as recurrent neural networks," in *IEEE International Conference on Computer Vision*, 2015, pp. 1529–1537.

[113] O. Veksler, "Star shape prior for graph-cut image segmentation," in *European Conference on Computer Vision*, 2008, pp. 454–467.

[114] V. Gulshan, C. Rother, A. Criminisi, A. Blake, and A. Zisserman, "Geodesic star convexity for interactive image segmentation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3129–3136.

[115] L. Gorelick, O. Veksler, Y. Boykov, and C. Nieuwenhuis, "Convexity shape prior for segmentation," in *European Conference on Computer Vision*, 2014, pp. 675–690.

[116] A. Delong and Y. Boykov, "Globally optimal segmentation of multi-region objects," in *IEEE International Conference on Computer Vision*, 2009, pp. 285–292.

[117] J. Ulén, P. Strandmark, and F. Kahl, "An efficient optimization framework for multi-region segmentation based on Lagrangian duality," *IEEE Transactions on Medical Imaging*, vol. 32, no. 2, pp. 178–188, 2013.

[118] G. Papandreou, L.-C. Chen, K. P. Murphy, and A. L. Yuille, "Weakly-and semi-supervised learning of a deep convolutional network for semantic image segmentation," in *IEEE International Conference on Computer Vision*, 2015, pp. 1742–1750.

[119] J. Xu, A. G. Schwing, and R. Urtasun, "Learning to segment under various forms of weak supervision," in *IAPR International Conference on Pattern Recognition*, 2015, pp. 3781–3790.