

# Segmentation and Quantification of Spinal Cord Gray Matter–White Matter Structures in Magnetic Resonance Images

**Inaugural dissertation**

to  
be awarded the degree of Dr. sc. med.

presented at  
the Faculty of Medicine  
of the University of Basel

by  
Antal Horváth  
from Basel, Switzerland

Basel, 2020

Approved by the Faculty of Medicine  
on application of

Prof. Dr. Philippe C. Cattin, University of Basel, *faculty representative*  
PD Dr. Katrin Parmar, University Hospital Basel, *co-referee*  
Prof. Dr. Jean-Philippe Thiran, Swiss Federal Institute of Technology  
Lausanne, *external expert*  
Dr. Simon Pezold, University of Basel, *further advisor*

Basel, 16<sup>th</sup> of December 2019

Prof. Dr. Primo Schär  
*Dean*

*To my parents Monika Horváth-Lüthy and György Horváth.*





# Contents

<b>Acknowledgments</b>	<b>vii</b>
<b>Foreword</b>	<b>ix</b>
<b>Summary</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Motivation . . . . .	2
1.2 Contribution . . . . .	3
1.3 Outline . . . . .	5
<b>2 Background</b>	<b>7</b>
2.1 Medical and Clinical Background . . . . .	7
2.1.1 Anatomy . . . . .	7
2.1.2 Multiple Sclerosis . . . . .	9
2.2 Technical Background . . . . .	13
2.2.1 Mathematical Terminology and Notations . . . . .	13
2.2.2 Computer Vision . . . . .	27
2.2.3 Digital Images . . . . .	29
2.2.4 Variational Principle . . . . .	31
2.2.5 Variational Derivative . . . . .	37
2.2.6 Segmentation Representations . . . . .	38
2.2.7 Segmentation Algorithms . . . . .	43
2.2.8 The Mumford-Shah Functional . . . . .	47
2.2.9 Including Appearance Models . . . . .	64
2.2.10 Supervised Deep Learning . . . . .	66
2.2.11 Validation and Comparison of Segmentation Models . . . . .	71
<b>3 The Underlying Project</b>	<b>75</b>
3.1 The Search for an Optimal MR Sequence . . . . .	75
3.2 Clinical Validation . . . . .	87
<b>4 High-Order Slice Interpolation for Medical Images</b>	<b>91</b>
<b>5 A Principled Approach to Combining Inversion Recovery Images</b>	<b>103</b>

## *Contents*

<b>6</b>	<b>Variational Segmentation of the White and Gray Matter in the Spinal Cord Using a Shape Prior</b>	<b>109</b>
6.A	Adaptation of the Continuous Cut Model to AMIRA . . . . .	123
<b>7</b>	<b>Novel Reliable Automatic Pipeline for Spinal Cord Gray Matter Quantification</b>	<b>127</b>
<b>8</b>	<b>Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA Images with MDGRU</b>	<b>139</b>
<b>9</b>	<b>Discussion and Conclusion</b>	<b>153</b>
	<b>Bibliography</b>	<b>159</b>
	<b>Curriculum Vitae</b>	<b>175</b>

# Acknowledgments

First of all, I would like to express my profound gratitude to Professor Dr. Philippe Cattin, head of the Center for Medical Image Analysis and Navigation (CIAN) and head of the Department of Biomedical Engineering at the University of Basel, for allowing me the opportunity to do my Ph.D. under his excellent supervision. Second, I deeply thank my advisor Dr. Simon Pezold, for his enduring constructive guidance. The high expertise of and the regular meetings with both Philippe and Simon made this work possible. I would like to thank PD Dr. Katrin Parmar, senior physician at the Department of Neurology of the University Hospital Basel, for her excellent leadership and coordination of this work within the scope of a larger SNSF project. I am very grateful to thank Professor Dr. Jean-Philippe Thiran from the Signal Processing Institute of the Swiss Federal Institute of Technology Lausanne (EPFL) for joining the Ph.D. committee as the external expert. Furthermore, I deeply appreciate and thank the collaboration from all our project partners at the University Hospital Basel, in particular, Dr. Charidimos Tsagkas, PD Dr. Katrin Parmar, Dr. Matthias Weigel, Dr. Michael Amann, Tanja Haas, and Professor Dr. Oliver Bieri.

I am also very grateful to Dr. Simon Andermatt, who helped me and the whole group with his great knowledge about supervised deep learning to adapt to the rapidly emerging new state of the art in solving computer vision problems. Great thanks to all my colleagues at the Department of Biomedical Engineering at the University of Basel, it was great working and sharing knowledge with Dr. Adrian Schneider, Alina Giger, Dr. Beat Fasel, Bruno Sempéré, Carlo Seppi, Dr. Christoph Jud, Corinne Eymann-Baier, Eva Schnider, Esther Zoller, Florentin Bieder, Dr. Frank Preiswerk, Iris Schmidt, Dr. Jörg Wagner, Julia Wolleb, Dr. Ketut Fundana, Dr. Kyung-won Baek, Lina Beltran, Lorenzo Iafolla, Manuela Eugster, Marek Zelechowski, Dr. Nadia Möri, Dr. Natalia Chicheroova, Peter von Niederhäusern, Robin Sandkühler, Samaneh Manavi Roodsari, Dr. Simon Andermatt, Dr. Simon Pezold, Dr. Stephan Wyder, Tiziano Ronchetti, and Dr. Uri Nahum. It was also a great pleasure to see their enthusiasm for combining computer science, sports, and academic research.

Finally, I would like to thank several people not directly involved in this scientific work: For the body and mind balance after the many working hours, I would like to thank my yoga teacher Lhundup Droyültshang, the whole university sports team, and the gymnastics club CALIBAS. Thanks to all my friends for realizing my work-life balance and for regularly pointing me to other fascinating aspects of life aside from science and sports. A special thanks goes to my partner Claudia Huck. The combination of my love for her and the focus on this work gave me emotional stability and the best motivation I could have. Finally, I wish to express my gratitude to my parents Monika Horváth-Lüthy and György Horváth for too many things to mention here. They have *always* been there for me.



# Foreword

I started this work with a mathematical background and therefore tried to analyze and understand image segmentation from a mathematical point of view. Soon, I realized that real, non-academic problems, such as segmentation of medical images, are difficult to accurately reformulate with computable abstract mathematical objects.

While it would have been interesting to experiment even more on the many different segmentation approaches and also to dive deeper into their theoretical understanding, the collaboration within the Ph.D. project urged me to find the first solution. Following the footsteps of the preliminary work of Pezold et al. at segmenting the whole spinal cord, we successfully continued applying similar variational frameworks for segmenting the inner structures of the spinal cord. Meanwhile, supervised deep learning became more and more popular and reached impressive performance in image segmentation, which led me to jump on this accelerating train. Having adapted the variational model for many, many months by hand, it was stunning to see a neural network outperform the hand-crafted model within only a few hours of training time.

Besides the project, this evolution in practice gave me reasons to describe some basic concepts from my own point of view to emphasize their ambiguity and how they are used and denoted in the different related fields. This thesis could have been written in a considerably shorter version; however, at some point during the writing, I aimed to describe things without presuming many terminologies to meet the different backgrounds of potential readers. Therefore, concise terms sometimes are paraphrased concerning different viewpoints to increase the likelihood of understanding.



# Summary

This thesis focuses on finding ways to differentiate the gray matter (GM) and white matter (WM) in magnetic resonance (MR) images of the human spinal cord (SC). The aim of this project is to quantify tissue loss in these compartments to study their implications on the progression of multiple sclerosis (MS). To this end, we propose segmentation algorithms that we evaluated on MR images of healthy volunteers.

Segmentation of GM and WM in MR images can be done manually by human experts, but manual segmentation is tedious and prone to intra- and inter-rater variability. Therefore, a deterministic automation of this task is necessary. On axial 2D images acquired with a recently proposed MR sequence, called AMIRA, we experiment with various automatic segmentation algorithms. We first use variational model-based segmentation approaches combined with appearance models and later directly apply supervised deep learning to train segmentation networks. Evaluation of the proposed methods shows accurate and precise results, which are on par with manual segmentations. We test the developed deep learning approach on images of conventional MR sequences in the context of a GM segmentation challenge, resulting in superior performance compared to the other competing methods. To further assess the quality of the AMIRA sequence, we apply an already published GM segmentation algorithm to our data, yielding higher accuracy than the same algorithm achieves on images of conventional MR sequences.

On a different topic, but related to segmentation, we develop a high-order slice interpolation method to address the large slice distances of images acquired with the AMIRA protocol at different vertebral levels, enabling us to resample our data to intermediate slice positions.

From the methodical point of view, this work provides an introduction to computer vision, a mathematically focused perspective on variational segmentation approaches and supervised deep learning, as well as a brief overview of the underlying project's anatomical and medical background.





# 1 Introduction

Capturing medical images is a standard procedure for medical diagnoses and research. For imaging, many different modalities exist, such as ultrasound, radiography (e.g. CT), and magnetic resonance (MR). After the acquisition, such images usually are analyzed by an expert, for example, by a radiologist. To extract higher-level information from the images, such as areas or volumes of specific regions, the relevant objects need to be identified and separated into different regions [1]. This process of separation is called image segmentation and is a basic and necessary step for further analysis.

In this project, we analyze axial MR images of the human spinal cord (SC) by segmenting the SC's gray matter (GM) and white matter (WM). We aim to improve our understanding of multiple sclerosis (MS), a chronic inflammatory and disabling disorder of the central nervous system, which includes the brain and the SC. It is one of the most common neurological disorders and is affecting more than 2 million people worldwide [2]. Typical symptoms can be visual, sensory, and motor impairments. Later in the disease progress, cognitive thinking may be affected as well. Usually, the first symptoms already show between the second and fourth decade of life, and unfortunately, the disease is currently incurable. Today, it is unclear, which factors act as a cause or origin of MS, but both genetic and environmental factors seem to play a principal role. Chief environmental risk factors are found to be vitamin D deficiency, cigarette smoking, and obesity [3], whereas the human leukocyte antigen gene cluster *DRB1\*15:01* is the most strongly associated genetic risk factor [4].

MS literally means *multiple hardening*, referring to CNS regions where harder, non-functional plaques, called lesions, replace the nerve cells and the supporting glial cells, following a focal inflammatory event in that region. Another well-known manifestation of MS is atrophy, which represents neural tissue loss leading to shrinking CNS structures. Atrophy is a long term effect, difficult to observe, and is thought to be both the aftermath of the inflammatory lesions and an independent diffuse neurodegenerative process [5]. It most likely represents neuronal, axonal, and myelin loss [5]. It has been shown that the progression of the disease correlates with neural atrophy as well as with the appearance of lesions in the spinal cord [6, 7, 8, 9, 10]. The presence of hypo- and hyper-intense spots representing such lesions, however, do not correlate well with the patient status regarding symptoms, which is known as the clinicoradiological paradox [11].

MR imaging is convenient for visualizing in a non-invasive way a human's SC in vivo. Compared to other conventional medical imaging modalities in radiology, such as CT, MR imaging uses non-ionizing radiation and is considered harmless, following the MR safety protocols [12]. MR is suitable for capturing the contrasts between

## 1 Introduction

different soft tissues, including GM, WM, and the SC's surrounding and protecting cerebrospinal fluid (CSF) [13]. However, one of the main problems in imaging the SC GM and WM are the similar MR characteristics of the two tissues. Thus designing new MR imaging protocols with optimal contrast is still part of ongoing research. Besides, the SC is small and bent in various degrees, and therefore a fine imaging resolution is needed to capture its delicate inner structures. Nowadays, with standard MR devices (with 1.5–3 Tesla magnetic flux density) and applicable acquisition duration the resulting SC images regarding GM-WM contrast are still noisy and blurry, see for example Figure 3.5. As a consequence, post-processing of such images, in particular, GM-WM segmentation, is challenging – as is demonstrated, for example, in the SC GM segmentation challenge [14].

This study aims at comparing axial MR images of the same patient at different time points during the disease to enable longitudinal studies. Comparing images acquired in different sessions is difficult, because of varying positioning of the subject in the scanner, and because each imaging attempt captures slightly different noise levels. The differences of the acquisitions and the subsequent tissue segmentations therein ultimately should reflect only anatomical or pathological changes. Therefore, segmentation errors due to imaging noise and the variability in patient positioning should be minimized. While the quality of such images probably is not enough for single-case studies, we first aim for large-scale studies, where segmentation errors should cancel out in a statistical sense for a large number of samples. If such conditions are met, it is possible to extract average atrophy rates of the GM and WM compartments for different populations and MS subtypes, as was already demonstrated for the SC [15, 16].

Many different segmentation approaches exist and new methods are still proposed. Roughly speaking, two classes of segmentation algorithms exist: the shallow model-based approaches and the deep learning-based approaches. In this thesis, we first start developing classic model-based methods where we tried fitting mathematical models with only a few parameters to the task of segmenting GM and WM on axial SC images. During these attempts, however, simultaneously with the increasing performance of graphics processing units (GPU), supervised deep learning approaches got more and more popular. Deep learning approaches are easily applicable and outperform many classic segmentation approaches, thus setting a new state of the art in image segmentation. We, therefore, move from the classic segmentation approaches to the supervised deep learning frameworks. We start investigating whether the proposed automatic methods can assist or even replace human expert segmentation. As a first and necessary step, the goal of this work is to validate the developed segmentation methods on MR images of healthy volunteers.

### 1.1 Motivation

It is already known, that spinal cord cross-sectional areas correlate strongly with the disability progression of MS [6]. SC atrophy has been observed throughout all stages

of the disease, from early phases to chronic progressive MS, and can quantify the pathology of the SC [17]. Thus atrophy has important implications for the treatment of all MS patients. Furthermore, there has already been some research on the segmentation of the spinal cord as a whole, for example the Spinal Cord Toolbox (SCT) project<sup>1</sup> [18] and this project's preceding Ph.D. project of Pezold et al. [19] with the cord image analyzer *cordial*<sup>2</sup>.

On the topic of segmenting the SC's inner GM and WM structures, however, not much research has been conducted, since only recently image acquisition methods have been developed that offer a high enough resolution and contrast of the GM and WM for internal segmentation [13]. Nevertheless, separating GM and WM in MR images is of great interest in research to enable studying their implications of the disease in vivo. A postmortem study shows that SC atrophy in MS primarily manifests in the WM of the upper cervical cord levels [20]. Thus it could be hypothesized that WM pathology would better reflect clinical disability than whole-SC measurements. On the other hand, another study shows, that SC GM atrophy correlates more strongly with the clinical disability than any other MRI metric, including WM atrophy [8].

State-of-the-art in-vivo SC images currently are acquired with 3 T MR machines, although machines with 7 T already exist which acquire images with higher quality [21]. Therefore, it is of utmost interest to analyze the vast amount of data being produced with current 3 T machines and to help in improving MR imaging sequences for such devices.

This project is part of a larger SNSF project<sup>3</sup> that aimed at finding appropriate image acquisition techniques and post-processing pipelines to assess GM and WM atrophy as well as detecting lesions in images of MS patients for longitudinal clinical and pharmaceutical studies. To quantify atrophy, the respective tissues have to be segmented in each scan in an accurate and precise manner. Manual segmentations of human expert raters, however, are prone to intra- and inter-rater variability, and need a lot of time to be performed for large datasets. Deterministic algorithms, in contrast, have zero intra-rater variability and can process large datasets with one and the same formula. Therefore, automatic segmentation algorithms play a crucial role in the assessment of medical images, which motivates the search for appropriate GM-WM segmentation algorithms.

## 1.2 Contribution

In this work, we analyze and use axial SC images acquired with the AMIRA sequence of healthy volunteers in multiple ways: we develop and validate automatic GM-WM segmentation algorithms; we develop a manual segmentation app to segment and visualize the different *AMIRA channels* comfortably (we refer to the AMIRA inversion images and their average projections as the *AMIRA channels*); we analyze the AMIRA

<sup>1</sup><https://github.com/neuropoly/spinalcordtoolbox> (last accessed on Sept. 10, 2019)

<sup>2</sup><https://github.com/spezold/cordial> (last accessed on Sept. 10, 2019)

<sup>3</sup>SNSF grant number: SNF 320030-156860/1

## 1 Introduction

sequence’s average images and propose another set of averages with optimal contrasts; and furthermore, we propose an interpolation technique to resample intermediate AMIRA slices.

We propose adaptations of established image segmentation frameworks to specifically segment GM and WM on axial images of the cervical SC, through (a) manually engineering variational segmentation models and (b) by using supervised deep learning.

- a) We use continuous cuts extended with appearance models [22] and propose to solve such models with the augmented Lagrangian multiplier-based maximal-flow algorithm [23]. For segmentation robustness, we include distance map-based non-terminal capacity functions, multi-channel capacity functions to include the individual AMIRA channels and additional image features, ellipsoidal rotational vector fields for anisotropic total variation, a slice similarity prior for 3D flows between the stack of AMIRA slices, and posterior appearance reconstructions based on trusted iterated pixelwise label predictions.
- b) We use a recurrent neural network with multidimensional, multi-directional, and convolutional gated recurrent units [24] and propose to combine the model’s cross-entropy loss with a generalized Dice loss. We train deep GM-WM segmentation networks that directly learn to leverage the AMIRA sequence’s tissue-specific MR relaxation curves. To test the proposed network architecture, we participated in an earlier SC GM segmentation challenge [14], which includes axial cervical slices acquired with conventional MR sequences. Official online evaluations yielded superior performance compared to the results of all the competing methods and other subsequently reported methods, setting a new state of the art in SC GM segmentation.

We evaluate the two developed segmentation algorithms on AMIRA images that are acquired in a scan-rescan scheme for intra- and inter-session comparisons, with and without repositioning of the subject in the scanner. On a cohort of 24 healthy subjects, our analysis yields accuracy, precision, and reproducibility scores, as well as additional data on SC GM-WM cross-sectional areas located at several positions of the cervix, which to date are rarely reported in the literature.

To simplify the generation of manual segmentations, we developed an application for manual segmentation in web browsers, which we implemented in JavaScript. In this application, drawing by mouse or with a pen on a touchscreen enables comfortable creation of manual segmentations. Completed segmentations are quickly and directly sent to a database through a client-server architecture, bypassing conventional data transfer means such as transport via USB drive. We specifically designed this application to enable the user to browse the different channels of the AMIRA sequence while simultaneously drawing segmentations. With this application, we also enable views together with the protocol’s additionally provided T<sub>1</sub>- and T<sub>2</sub>-weighted 3D sequences.

With a principled, compressed sensing inspired approach, we verify that the naive uniform averages of the AMIRA inversion images proposed in [25] are already close-to-optimal.

We propose to modify the registration-based slice interpolation technique [26], by including third-order Hermitian spline interpolation to estimate pixelwise location trajectories along the stack of slices with subsequent intensity interpolation along these trajectories. This method enables the creation of SC cross-sectional views at arbitrary locations.

## 1.3 Outline

In Chapter 2, we introduce the reader to the medical and technical background of this thesis, and in Chapter 3, we provide a more detailed overview of the SNSF project that this thesis is part of. Chapters 4 to 8 comprise our publications that originate from our work so far, with Chapter 4 describing a slice interpolation technique, Chapter 5 analyzing the AMIRA average images, Chapter 6 presenting our initial continuous cut model, Chapter 7 introducing adaptations to the initial model and demonstrating results on the CGM dataset, and finally Chapter 8 describing our contribution with supervised deep learning. We complete the thesis with a discussion and conclusion in Chapter 9.



## 2 Background

In this chapter, we first provide a brief introduction to the medical aspects of this work, and second, we give an extended view on several technical topics we encountered during this Ph.D. thesis. While the medical background is essential for understanding the direct application of this thesis, with the technical background, we clarify mathematical notations, provide an overview of variational segmentation models, introduce supervised deep learning, and address philosophical aspects on automation of image processing.

### 2.1 Medical and Clinical Background

In this section, we first describe the spinal cord's anatomy and then briefly summarize the neurological disease called multiple sclerosis without going into biochemical details.

#### 2.1.1 Anatomy

The spinal cord (SC) is the longitudinal rostrocaudal structure that connects the brain with most parts of its periphery, where the SC and the brain together build the central nervous system. The SC is surrounded and protected by the vertebral column, which consists of multiple connected vertebrae. On each vertebral level, the SC reaches through the vertebral foramen which lies in the center of the six basic vertebral components: the ventral vertebral body, the two lateral vertebral arches and their lateral transverse processes, and the dorsal spinous process. Inside this bony cage, the SC, like the brain, is immersed in cerebrospinal fluid (CSF), which builds an additional mechanical and also immunological protective measure. The CSF circulates in the so-called subarachnoid cavity between the outer arachnoid mater and the inner pia mater that tightly covers the SC surface. Furthermore, between the vertebral arches of two subsequent vertebrae, spinal nerves are exiting and entering through the intervertebral foramen. Each intervertebral opening is associated with specific regions, called dermatomes and myotomes, to which sensory and motoric axons are connected.

**Location** To visualize the SC's location, axial slices as well as a mid-sagittal slice of the neck, acquired with magnetic resonance imaging (MRI), are shown in Figure 2.1. In the neck region, also called the cervical region, the vertebrae are labeled as C1 to C7, where the first two are called atlas and axis. Atlas and axis together form an

## 2 Background

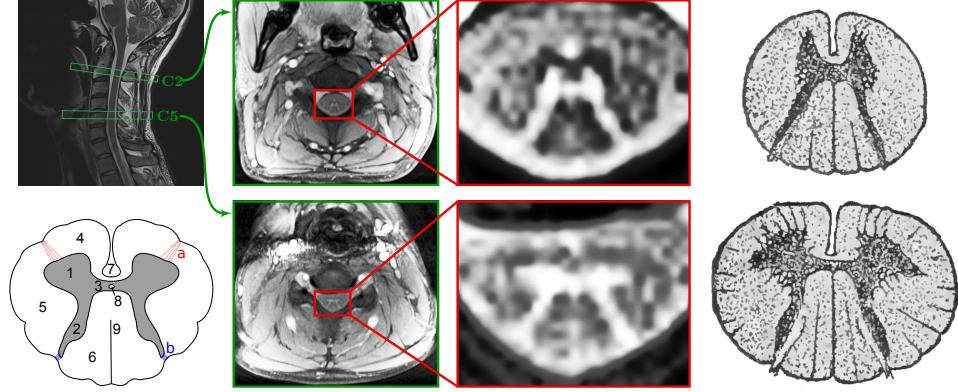


Figure 2.1: *Top left*: median slice of a T<sub>2</sub>-weighted turbo spin echo sequence scan of the neck with slice locations of the mid-column images in green; *mid-columns*: axial slices acquired with the AMIRA sequence at C2 and C5 level, respectively; *right column*: histological drawings of SC cross-sections at C1 and C5 level, respectively (reproductions of [27, p. 754], public domain); *bottom left*: schematic cross-section (reproduction of [28]) displaying GM in anterior horn (1), posterior horn (2), commissure (3), and WM in anterior funiculus (4), lateral funiculus (5), posterior funiculus (6), anterior root (a), posterior root (b), and parts filled with CSF in anterior median fissure (7), central canal (8), posterior median sulcus (9).

elongated structure, visible in the mid-sagittal slice on the top left image in Figure 2.1 near the upper green rectangle. Therefore, starting from the top, the first vertebral body after the first intervertebral disc, below the upper green rectangle, is already labeled as C3.

**Inner Structure** The SC’s ellipsoidal cross-section, as well as its inner structure, gradually varies rostral to caudal, yet its axial slices morphologically share a common structure. There exist two visual enlargements in the cervical and the lumbar regions of the SC, with the cervical enlargement starting on C5 level and ending at the beginning of the thoracic part. While the transversal and the anteroposterior diameters at C2 level are approximately 11 mm and 8 mm, respectively, the cervical enlargement causes corresponding diameters at C5 level of around 13 mm and 7 mm [29, Table 5]. The SC mainly is divided into white matter (WM) and the butterfly-shaped gray matter (GM), as depicted and labeled in more detail in a cross-sectional schematic in Figure 2.1. To underline the gradually varying SC structure, we included two axial MR images as well as corresponding histological drawings in Figure 2.1 that represent the slightly thinner anterior and posterior horns at C2 level and the enlarged anterior horns at C5 level.

Aside from glial cells, synapses and blood capillaries, GM mainly consists of neu-



ronal cell bodies with dendrites and axons. In a 3D view, the GM is separated into three (respectively five) rostrocaudal columns; the two anterior and the two posterior columns, and the lateral column which together form the axially symmetric shape of the letter "H" or of a butterfly. In the anterior column, mainly motor neurons are interconnected through axons that are traveling down from the top which are responsible for the movement of the muscles, whereas in the posterior column sensory neurons are gathering signals from the periphery and are transmitting such information upwards through the SC. The third lateral column, or also called the gray commissure, wraps around the CSF filled central canal and bidirectionally connects the left and right side. An axial cross-sectional view of the anterior and the posterior column motivates their second names, which are *ventral* and *dorsal horn*, respectively.

In contrast to GM, WM is mainly made up of myelinated axons, which render the histological white color through the myelin's high lipid content. Freshly cut WM, however, appears pinkish because of the presence of capillaries; and even more abstract, the intensity differences between WM and GM on images acquired with MRI can vary with respect to different sequences. In the exemplary axial MR images in Figure 2.1, GM appears brighter than WM. Myelinated axons are axons covered by functioning oligodendrocytes or, more peripheral, by Schwann cells that form a myelin sheath and support nerve regeneration. Through saltatory conduction, the myelin sheath increases the velocity and the amplitude of the electric nervous signal, where compared to an unmyelinated axon, the velocity is up to 10 times faster.

**Cerebrospinal fluid** The CSF in the subarachnoid space is mainly composed of water. It is performing a pulsating fluid motion that corresponds to the pressure waves generated in the blood vessels through the systolic pulses of the heart [30]. This CSF movement is called CSF pulsation and influences MR imaging with so-called CSF pulsation artifacts, where CSF that is not pulsating may reveal higher signal than pulsatile CSF [31]. Therefore, when imaging regions where CSF is involved, MR sequences need to be carefully designed to reduce CSF pulsation effects.

The high water content of CSF can be exploited for MR imaging to produce images with well visible CSF contrast, which is particularly useful for SC segmentation. However, because of the freely moving SC inside the bony cage, CSF may be pressed aside. This looseness of the SC may result in axial MR images without a ring-shaped CSF structure, see Chapter 6, Fig. 2, which is caused for example when a subject is positioned in the MR device with a strongly extended (backward bent) neck such that the SC touches the posterior part of the vertebral foramen. Such images may have impaired contrast between the SC and its background at locations where the SC touches the foramen. Such cases need to be considered in developing SC segmentation algorithms and may reduce the segmentation accuracy.

### 2.1.2 Multiple Sclerosis

Multiple sclerosis (MS) is a chronic inflammatory and disabling disorder of the central nervous system. It is one of the most common neurological disorders that has not been

## 2 Background

caused by trauma and is affecting more than 2 million people worldwide [2]. Because MS typically involves injury of nerve tissue in the brain, the SC, as well as the optic nerves, typical symptoms are impaired sensory, motor, and cognitive skills, such as monocular visual loss, double vision, coordinative disorders, locomotion problems, bladder dysfunction, equilibrium impairment, limb weakness, lack of attention [32], and deficient reasoning. Usually, the first symptoms already show between the second and fourth decades of life, and unfortunately, the disease is currently incurable. MS may have a high impact on the patient's quality of life because of the mentioned severe physical and neuro-psychological disadvantages. From a socio-economic point of view, MS causes high health-care costs which rise with increasing disability.

**Statistics of Pathogenesis** Today, it is unclear, which factors act as a cause or origin of MS and whether triggers are intrinsic or extrinsic to the central nervous system, but both genetic and environmental factors seem to play a principal role [3]. Main environmental risk factors are found to be vitamin D deficiency, cigarette smoking, and obesity; genetic risk factors include an increased probability in relatives of MS patients, female sex with a ratio between women and men higher than 2:1, and human leukocyte antigen (HLA) gene clusters with the most influential associated genetic risk factor in *HLA-DRB1\*15:01* [4]; virus infections with an increased probability for carriers of for example the Epstein-Barr virus; and geographic latitudes with a higher incidence in more temperate climates, which however may reflect in the region's varying vitamin D levels, or different pathogens and genetic distributions prevalent in these regions. Moreover, male sex, older age, and higher disability state at disease onset are associated with a worse prognosis.

**Diagnosis** Several updates in the diagnostic criteria for clinically definite MS have been proposed to improve true positive and true negative rates, as by Poser et al. in 1983 [33], McDonald et al. with the so-called McDonald criteria in 2001 [34], and Thompson et al. with the currently newest update of the McDonald criteria in 2017 [35]. The revised McDonald criteria rely on a combination of data, including patient history, clinical examination, paraclinical findings such as CSF analysis, and more. The key principle of the criteria is "dissemination of lesions in space and time", which means that characteristic MS-lesions need to be identified in various parts of the CNS together with lesions of different age. Over the years revisions of the diagnostic criteria have also come to increasingly involve the value of MRI findings to assist the space-time analysis of characteristic MS-lesions disseminated in the CNS. To quantify the severity of the neurologic impairment in MS, physicians use the expanded disability status scale (EDSS) [36] with a rating from 0 ("normal neurological exam"), through 5 ("severe to maximal impairment"), to 10 ("death due to MS").

**Pathology, Symptoms and Disease Course** The main features of MS are axonal and neuronal loss through degeneration, demyelination, and astrocytic gliosis occurring in the CNS, which manifests in characteristic lesions that are distributed over multiple

## 2.1 Medical and Clinical Background

regions in the CNS [3]. While the loss of nervous tissue, called atrophy, is a slow process with yearly atrophy rates of roughly estimated 1% volume loss [37, 15, 16], the disease embodies in the visually more prominent lesions that give the disease multiple sclerosis its name.

MS lesions occur at characteristic locations in the CNS and are spots with varying degrees of demyelination, inflammation, and glial reaction, which also includes blood-brain barrier anomalies. In the SC, the lesions mainly arise in the SC circumference, the dorsal WM tracts, the lateral corticospinal tracts, and do not respect the GM-WM borders [38]. Demyelination is assumed to be an immune attack on the myelin-oligodendrocyte complexes along the axons, and this immune reaction once activated may provoke further massive inflammatory destruction. As has been observed but poorly understood, surviving axons in a lesion may be remyelinated by a thin sheath, may be partially healed by other regenerative processes, may remain chronically inactive without remyelination, or may turn into a so-called smoldering lesion with persisting slow inflammation and demyelination [2]. As mentioned in Section 2.1.1, missing myelin nodes along axons are weakening and slowing down the electric signal conduction. Axonal disruption or axonal loss as a result of an acute focal inflammatory event or a slow degenerative process leads to complete stop of signal conduction.

MS patients show symptomatic phases with periods of stability, recovery, and relapses and based on the disease course, three clinical subtypes of MS have been proposed in 1996 [39]: the relapsing-remitting phase of MS (RRMS) with alternating and variable periods of stability, relapses, and remissions; secondary progressive MS with a final, gradually increasing worsening of the disease, often 10 to 20 years after the onset of the RRMS phase; and primary progressive MS (PPMS) with direct progressive worsening after disease onset without distinctive remissions or relapses with around 15% of all patients [2]. However, as suggested in [40, 41], the clinical subtypes do not reflect the actual disease heterogeneity, and therefore potential other subtypes have to be construed to develop more diversified and better matching patient-specific treatments.

Compared to the lesion development, also not well understood is the mentioned, slow and diffuse process of neuronal atrophy or degeneration, which is thought of as the aftermath of MS lesions, but might also occur independently of inflammation [42]. Estimated atrophy rates for the SC as well as the brain were found to be significantly higher for MS patients as compared to healthy controls, but among the clinical subtypes, these rates do not allow differentiation [43, 37, 15], though progressive forms show slightly accelerated atrophy rates. Increased rates with mean annual atrophy of around 2% were found in a small cohort of PPMS patients [16], and interestingly, the SC volume loss of this cohort progressed independently of total brain volume and brain lesion measurements, suggesting that lesions and atrophy may occur independently.

## 2 Background

**Treatment** Disease-modifying medications are available to reduce the frequency and amplitudes of the MS episodes, such as the antiviral interferon beta and glatiramer acetate; the monoclonal antibodies natalizumab, alemtuzumab, daclizumab (not used anymore), and ocrelizumab [44]; the chemotherapeutic therapy with mitoxantrone (rarely used nowadays); as well as the small-molecule agents fingolimod, dimethyl fumarate, and teriflunomide [2]. However, no medication entirely prevents progressive neurologic deterioration in all patients or reverses the already acquired neuronal injury, yet. The classification of MS currently is too coarse to design patient-specific, causal treatments.

As a general conception, regular physical and cognitive training may lead to improved long-term health by positively affecting inflammation and neurodegeneration, and by establishing new neuronal pathways and preserving the CNS's plasticity. As examined in [45], training, therefore, has meaningful consequences in MS patients, yet is not the only treatment. Recent longitudinal pharmaceutical studies show significant effects in delaying clinical progression [44, 46]. However, whether the clinical progression of MS can be delayed or even reversed is quite a controversial field of the disease [2, 46].

**MRI in MS** As suggested in the McDonald criteria for the diagnosis of MS, MRI is used to assist decisions based on the patient history by visualizing and monitoring the CNS in-vivo and in space and time. A pioneering work [43] tried to find correlations between MR images and the patient's EDSS values of the different MS subtypes, but revealed considerable heterogeneity, pointing to the so-called clinoradiological paradox [11]. In some individual cases, newly appearing lesions did not necessarily imply a clinical worsening of the disease and in the opposite logical direction, no changes between the baseline and follow-up scans did not imply an unchanged status of the disease.

With the improving quality in MRI and therefore also improving quality in detection of anatomic structures as well as pathologic tissue, however, more detailed and specific studies can be conducted to find reliable MRI biomarkers. So far, marking locations with hyper- and hypo-intense MRI intensity changes reflecting lesional tissue did not serve as a reliable biomarker. Since lesion segmentation is not as reproducible as the segmentation of anatomic structures, however, technical limitations [47] cannot be excluded.

Brain MRI is well-established and brain atrophy can be accurately and reproducibly measured [37]. It has turned out, however, that also brain atrophy does not reliably correlate with the patient's disability measured as EDSS [48]. Among the different markers in the brain, GM atrophy has shown the best correlations with the disease progression [48]. In contrast to brain MRI, SC MRI has shown to provide better biomarkers for the disability progression of MS [38], although imaging the SC is not as reproducible as imaging the brain. SC cross-sectional areas provide strong correlations to the disability progression of MS [6], and SC atrophy indeed is starting from the early phases of the disease. Thus, it has important implications for the

treatment of all MS patients [17]. Furthermore, in 2012, Yiannakas et al. [13] showed the feasibility for internal GM-WM segmentation of the SC on MR images, allowing for even more detailed studies. Although SC atrophy primarily manifests in the WM compartments of the upper cervical cord levels, as shown in a postmortem study [20], later in 2014, Schlaeger et al. [8] discovered, that cerebral SC GM atrophy correlates more strongly with EDSS than WM atrophy with – similar to the finding in [48], where GM atrophy provided best correlations in the brain.

## 2.2 Technical Background

In this section, we describe some of the basic concepts and frameworks used during this project. Since this text is written from a mathematical viewpoint, different mathematical concepts are thought of as prerequisites to understand this text. Nevertheless, the most important terms like a set and a function and how they are denoted in this thesis are carried out in the following section.

After giving the mathematical foundation, we start with how the human visual recognition could be modeled via energy functions, continue with variational principles, give an overview of variational segmentation algorithms, and finally introduce supervised deep learning.

### 2.2.1 Mathematical Terminology and Notations

In this section, we superficially describe some important concepts, putting a focus on the notations. It is not self-contained and we refer the reader to standard literature such as [49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59] for more complete information. The aim is to introduce the terms to non-technicians while trying to remain exact and mentioning relations to more advanced topics that would be necessary to analyze the approaches presented in this work numerically. This introduction also endows the reader with an overview to better understand related topics from the literature. However, as already mentioned, the primary goal is to familiarize the reader with the notations. Throughout this work, representing objects in multiple similar ways should not confuse but help the reader by offering different perspectives. In the literature of the applied fields, mathematical terms are overloaded and sometimes only vaguely defined in a way that the local community understands but outsiders first have to get a feeling for to understand.

On the one hand, terms can be defined *explicitly* through construction, or *implicitly* by defining their properties. On the other hand, there are two extreme forms in understanding: the non-interpreting form, where one doesn't understand if the term is not well-defined, or a syntactic or semantic error is present; and the *interpreting* form, where one tries to assume the most probable meaning while keeping in mind the whole set of possible fits. Failing communication, therefore, could be due to explicit constructions being too abstract, implicit properties being not well-defined or too vague, or the interpreter not using or not having the needed information.

## 2 Background

**Sets** Without mentioning modern axioms, a set is a collection of distinct elements. To gather the distinct elements, e.g.  $a, b, c$ , to a set, say  $S$ , we denote

$$S = \{a, b, c\},$$

using the curly brackets  $\{\cdot\}$ . Sometimes all the properties of an element are encoded by using a specified symbol, for example, as is often done by using  $x$  as a real number. In doing so, the symbol lets one automatically assume the intrinsic properties of that object. Using this idea, if  $m$  is the specified symbol for elements in a set  $M$ , sometimes the set of all the elements in  $M$  is denoted in a short and handy notation

$$M = \{m\},$$

which could easily be mistaken with a singleton set containing only one element instead of all possible elements of that type.

When referencing elements of a set but choosing only a part of it, we often define subsets. For example, a subset  $T$  of  $S$  can be denoted as  $T \subset S$ , and to specify which elements do belong to  $T$ , we can write

$$T = \{e \in S \mid e \text{ fulfills a specified condition}\},$$

which reads as " $T$  is the set of all elements  $e$  in  $S$ , such that  $e$  fulfills a specified condition". Moreover, a subset  $T \subset S$  does not need to be a proper subset. It could also be equal to  $S$ .

**Sequences** By adding an order to the elements of a set, we end up with a sequence, which we denote with round brackets  $(\cdot)$ . For example, in the sequence

$$S = (a, b, c),$$

$b$  follows  $a$ , and  $c$  follows  $b$ . Using 0-indexing, i.e. 0 as a starting index, one could write the latter example as  $S_0 = a$ ,  $S_1 = b$ , and  $S_2 = c$ . We can also use an index set, e.g.  $\Omega$ , and denote a sequence as

$$S = (S_i)_{i \in \Omega} = (S_i)_i,$$

where the outer index indicates, that one has to iterate over all possible indices  $i$  in  $\Omega$ . Putting a second subscript  $i$  outside the brackets,  $(S_i)_i$ , is a clever way to handle the ambiguity whether  $S_i$  refers to the  $i$ -th element in  $S$  or to the sequence  $S$  itself. To this end, identifying  $S_i$  as the whole sequence  $S$  could be realized by interpreting  $i$  as a symbolic variable placeholder which accounts for all possible indices. This ambiguity is comparable to identifying  $f(x)$  as the value of  $f$  at  $x$  or as the whole function, see next paragraph of this section.

Furthermore, the sequence notation in round brackets may also be used for arrays such as vectors and matrices by indexing each dimension.

**Functions** A mathematical function  $f$  can be denoted as

$$\begin{array}{ccc} f : & D & \longrightarrow & C \\ & x & \longmapsto & f(x), \end{array} \quad (2.1)$$

where  $f$  maps elements  $x$  from its domain  $D$  to elements  $f(x)$  in its codomain  $C$  [60]. The element  $f(x)$  is the value of  $f$  at position  $x$ . In the above notation,  $x$  appears below  $D$ , which means that  $x$  is in  $D$ , in short:  $x \in D$ . One may interpret  $x$  as an arbitrary but specific element in  $D$ , or as a variable/placeholder, and then  $f(x)$  is either an element in  $C$  or  $f(x)$  represents the whole function, respectively. On the right side of the arrows,  $f(x)$  appears below  $C$ , which means  $f(x) \in C$ . The symbol  $\mapsto$  specifies how an element is mapped to another one, and the symbol  $\rightarrow$  indicates a mapping between two sets.

In the applied field, often only the function expression " $f(x) = \dots$ " is denoted, leaving out the top part in (2.1), specifying the domain as well as the codomain only implicitly. In many cases, declaring the domains helps to understand, to calculate derivatives, and to analyze convergence properties of the mapping. Thus the domain and codomain should be specified if contextually not clear. Moreover, sometimes the position  $x$  in  $f(x)$  is being suppressed by only writing  $f$ , but meaning the value  $f(x)$ . In other words, sometimes  $f$  is used as a variable value, without indication or knowledge of its dependency on other variables.

Finally, a function actually is a set of assignments  $x \mapsto y$ , such that each element  $x \in D$  exactly has only one element  $y \in C$  with  $f(x) = y$ . This does *not* mean that different  $x_1 \neq x_2 \in D$  cannot map to the same  $y \in C$  with  $f(x_1) = f(x_2) = y$ . It only means that for  $x \in D$ ,  $f(x)$  is in  $C$  and that there do not exist different  $y_1 \neq y_2 \in C$  with  $f(x) = y_1$  and  $f(x) = y_2$ . In other words, this means that for every  $x$ ,  $f(x)$  is well-defined.

This definition leads to the following effect that the set of all inverted assignments – which would be a candidate for a function inverse – is not a function anymore if the set of assignments is not injective and surjective.  $f$  being injective means, that different  $x_1 \neq x_2$  cannot map to the same value  $f(x_1) = f(x_2)$ , whereas  $f$  being surjective means that all values in the codomain can be reached by elements from the domain. Furthermore, a function which is injective and surjective is called bijective or invertible.

Figure 2.2 illustrates a typical 1D function and also introduces two mathematical terms that help us to better understand functions in context: the image of a function and the graph of a function. The image, also called the range, of a function  $f$  is the set of all reached values, i.e.

$$\text{Image}(f) = \{f(x) \mid x \in D\} \subset C,$$

and is a subset of the codomain  $C$ . On the other hand, the graph of  $f$  is a hyperspace in the Cartesian product space  $D \times C$  of the domain and the codomain and consists of all position-value tuples, i.e.

$$\text{Graph}(f) = \{(x, f(x)) \mid x \in D\} \subset D \times C.$$

## 2 Background

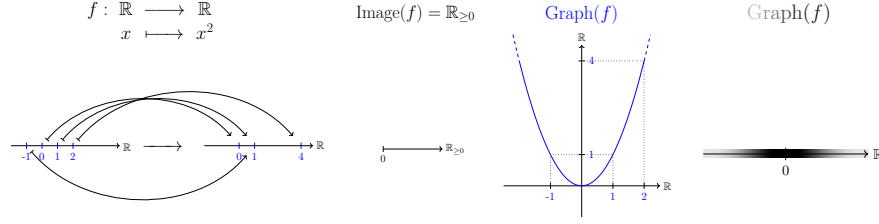


Figure 2.2: Illustrations of  $f : \mathbb{R} \rightarrow \mathbb{R}$  with  $f(x) = x^2$ .

Also often used, the preimage or inverse image  $f^{-1}(M)$  of a subset  $M$  in  $C$  is defined as

$$f^{-1}(M) = \{x \in D \mid f(x) \in M\}.$$

The preimage of  $M$  consists of all elements in the domain that map to  $M$ . It even exists if the element-wise inverse function  $f^{-1}(y)$  does not exist. The inverse image has many notations, such as the short notation  $[f \in M]$  used for example for probability measures  $P[f \in M]$  or  $\{f > 0.5\}$  for super-level sets. In such cases, they all declare a set  $\{x \mid f(x) \in M\}$ .

Furthermore, the mathematical image of  $f$  is a subset of the codomain (2<sup>nd</sup> image from the left in Figure 2.2) and should not be confused with a digital image. In a digital image, as described in Section 2.2.3, the assignment of each element in the domain is essential. In other words, the digital image is a pixel-intensity function itself or, more precisely, a representation of its graph. A graph could for example be represented using the canonical visualization of the Cartesian product space  $D \times C$  (3<sup>rd</sup> image in Figure 2.2), or the codomain could also be shown using a color space (4<sup>th</sup> image in Figure 2.2). There the colored graph of  $f$  illustrates a 1D image.

However, to give reason for the mathematical term "image", actually also a digital image can be interpreted as an  $\text{Image}(f)$ , if for example the function  $f$  is the acquisition process that maps the voxels from the imaging scene to the pixels in the image.

Last but not least, two functions can be concatenated or composed if the codomain of the first applied function is a subset of the domain of the second applied function. For example, composing the two functions  $f_1 : D \rightarrow M$  and  $f_2 : M \rightarrow C$  results in  $f = f_2 \circ f_1 : D \rightarrow C$  and can be denoted as

$$\begin{array}{ccccc} D & \xrightarrow{f_1} & M & \xrightarrow{f_2} & C \\ x & \longmapsto & f_1(x) & \longmapsto & f_2(f_1(x)) = (f_2 \circ f_1)(x). \end{array}$$

The other way round, one can also try to decompose the function  $f : D \rightarrow C$  into parts  $f_1 : D \rightarrow M$  and  $f_2 : M \rightarrow C$ . The concept of decomposition is particularly important because partitioning of a mapping process is necessary for understanding and creating algorithms that can mimic a certain process – compare with Section 2.2.10.



**Discrete vs Continuous** The set of the natural numbers  $\mathbb{N} = \{0, 1, 2, \dots\}$  is discrete and the set of the real numbers  $\mathbb{R}$  is continuous (with respect to their standard topology). Intuitively speaking, a discrete set has a minimal distance between its elements, and in a continuous set every point can be approximated arbitrarily well by other points from the set.

In the mathematical field of topology, a set is *discrete* if every point is isolated, i.e. if every point has an open set that contains itself but none of the others. Going the other direction, a set could be called *continuous* if there are no isolated points. For such a definition, one first has to define what an open set is. Openness actually depends on the chosen topology, which is simply the set of all open sets (and some other properties for this system of sets) – if a set is part of the chosen topology, then it is open, otherwise, it is not open [61]. In this abstract but not necessarily meaningful concept, every set can be “made” discrete by choosing the discrete topology, where all sets containing only one element, also called singletons, are elements of the topology. In many applications, however, we usually deal with sets that are equipped with a distance metric, a norm, or a scalar product, where these metrics already induce a topology in a more meaningful way [57].

For example, the real numbers  $\mathbb{R}$  together with the Euclidean distance  $|\cdot - \cdot|$  build the standard topology  $\mathcal{T}_{\mathbb{R}}$ . There, all intervals  $]x_0 - \epsilon, x_0 + \epsilon[ = \{x \in \mathbb{R} \mid |x - x_0| < \epsilon\}$  with  $x_0$  in  $\mathbb{R}$  and  $\epsilon > 0$  in  $\mathbb{R}$  are open, and actually they build a topological basis for  $\mathcal{T}_{\mathbb{R}}$ . It follows that singletons  $\{x_0\}$  are not open since it is not possible to draw an open set around a point  $x_0$  without including other points from the line, and thus  $\mathbb{R}$  with the standard topology is continuous.

The set of the natural numbers  $\mathbb{N}$  as a subset of  $\mathbb{R}$  together with the subspace topology  $\mathcal{T}_{\mathbb{N}} = \{S \cap \mathbb{N} \mid S \in \mathcal{T}_{\mathbb{R}}\}$  of the real standard topology  $\mathcal{T}_{\mathbb{R}}$  is discrete, since for example with  $\epsilon = 1/2$ , we realize, that all singletons  $\{n\} = ]n - 1/2, n + 1/2[ \cap \mathbb{N}$  are actually open because they are by construction in  $\mathcal{T}_{\mathbb{N}}$ .

Further examples are digital images that might be discretized during their acquisition process. A digital image’s domain is discrete because such an image typically is defined on a countable grid (which is bijective to a subset of the natural numbers  $\mathbb{N}$ ).

In the mathematical field of analysis, continuity is a property of a function and usually is not defined as a property of a set. Therefore, in contrast to our above definition of a continuous set, in the applied field, one probably assumes even more conditions than just the absence of isolated points. One might also add closedness, i.e. that the limits of all converging sequences inside the set again belong to the set. A set with absence of isolated points together with closedness then would be called a perfect set; however, such definitions would be given ad hoc for a specific problem. With simple words, a continuous set shall be similar to a real-valued vector space (equipped with a standard distance, norm or scalar product).

Nevertheless, the differentiation between discrete and continuous is essential because the literature on variational approaches uses both languages: discrete formulations are closer to implementable algorithms, and with continuous formulations, many infinitesimal concepts of the continuous analysis are imported [50, 49]. Modeling a discrete problem in a continuous domain comes along with additional problems

## 2 Background

though, where one has to interpolate the information between the discrete data points, which can lead to further discretization errors. Problems that arise in the infinitesimal calculus are not problems a priori present in the discretized world. Nevertheless, inferring a continuous model from discrete data that *originates from an actually continuous problem* can obviously be advantageous since it models the original problem more closely. In the following, we list a few basic differences between discrete and continuous math that appear in practice, cf. [50, Sec. I.1.5]:

1. Sums are used in the discrete setting and integrals for continuous spaces.
2. Finite differences are encountered in the discrete case and differential quotients in the continuous analysis. For a given resolution, finite differences can also be calculated for a continuously non-differentiable function.
3. The non-differentiability of the absolute function  $|\cdot|$  at the origin can cause problems in optimization and thus usually is smoothed out.
4. The jump discontinuity of the sign function, as the variational derivative of the absolute function  $|\cdot|$ , is often replaced by a smooth approximation.
5. The delta Dirac distribution, as the distributional derivative of the sign function, can be described precisely in the continuous domain but is problematic to discretize for arbitrary resolutions.
6. The probability for a continuous random variable to realize a single value is arbitrarily small, i.e. 0, even if its density is larger than 0. This cannot occur in a discrete setting.
7. And finally, from a topological viewpoint, a function on a discrete domain is always continuous, but without assumptions, it cannot be determined whether its analytic continuation or its continuous source is smooth or not. For example, a function defined on the natural numbers could be extended non-continuously as a step function onto the real numbers, or it could also be interpolated as a smooth continuation on the real numbers.

Among other occasions, these aspects were considered in Section 2.2.8 and led to the choice of the continuous max-flow min-cut framework (2.74).

**Differentials** In the mathematical field of analysis, a differential of a function describes the slope of a function at each position.

If you find yourself at a position in a graph and ask yourself how the situation changes when you move a tiny step away from your current position, then you can use the differential of the situation. The differential defines the ratio of how much your situation changes with respect to the step size of your tiny step. In abstract terms, for a smooth function  $f : D \subset \mathbb{R} \rightarrow \mathbb{R}$  on an open set  $D$ , if  $\Delta x$  is the step size from the current position  $x$  to the new position  $x + \Delta x$ , and  $\Delta y$  is the change of

## 2.2 Technical Background

the function's values from  $f(x)$  to  $f(x + \Delta x)$ , then one can write "by expanding by  $1 = \Delta x / \Delta x$ "

$$f(x + \Delta x) = f(x) + \Delta y = f(x) + \underbrace{\frac{\Delta y}{\Delta x}}_{\approx f'(x)} \Delta x. \quad (2.2)$$

Note that these ratios may vary at different positions  $x$  and for different step sizes  $\Delta x$ , and thus we should define the derivative  $f'$  of  $f$  at position  $x$  as

$$f'(x) = \frac{df}{dx}(x) = \frac{\partial f}{\partial x}(x) = \partial_x f(x) := \lim_{\Delta x \rightarrow 0} \frac{f(x + \Delta x) - f(x)}{\Delta x}, \quad (2.3)$$

using the important concept of a limit  $\lim_{\Delta x \rightarrow 0}$  to construct infinitely small step sizes [51, 53]. Note that in the literature, the term  $f'$  (spoken " $f$  prime") may have different notations such as  $\frac{df}{dx}$ ,  $\frac{\partial f}{\partial x}$ , or simply  $\partial_x f$  (spoken: "del  $x$  of  $f$ "). In this thesis, we mainly use the latter, which indicates the differential variable as a subscript of the short notation  $\partial$ . For the definition above, *smooth* or *smooth enough* usually means that the limits and thus the involved derivatives exist. From a geometrical point of view, the derivative exists, whenever you can uniquely place a tangent (or a tangential hyperspace) to its graph at the specific location. The slope of the tangent then is the same as the local derivative. If a function has a jump from one location to its next location or if there is a kink, it is not well defined how to place such a tangent.

The notation for a derivative with  $f'$ , as proposed above, is mainly used if the variable  $x$  is one-dimensional. If, however, the function's domain is multidimensional, then the concept of directional derivatives becomes essential. The derivative of  $f$  at position  $x$  can be taken with respect to a certain direction  $v$  through the following construction

$$\partial_v f(x) = \lim_{t \rightarrow 0} \frac{f(x + tv) - f(x)}{t}, \quad (2.4)$$

where  $v$  is an element of the tangent space of  $f$  at location  $x$ . Since in a multidimensional domain there are infinitely many directions, one usually collects the directional derivatives towards the linearly independent directions in the following manner: For a finite,  $n$ -dimensional domain  $D$  and a smooth function  $f : D \rightarrow \mathbb{R}^m$ ,  $n$  directional derivatives towards  $n$  linearly independent directions can be calculated and represented in the following matrix-valued function  $Df : D \rightarrow \mathbb{R}^{m \times n}$ :

$$Df(x) = (\partial_{x_1} f(x), \dots, \partial_{x_n} f(x)) = \begin{pmatrix} \partial_{x_1} f_1(x) & \cdots & \partial_{x_n} f_1(x) \\ \vdots & \ddots & \vdots \\ \partial_{x_1} f_m(x) & \cdots & \partial_{x_n} f_m(x) \end{pmatrix} \in \mathbb{R}^{m \times n}, \quad (2.5)$$

using the tangent space's canonical basis directions  $x_i = (\delta_{ij})_{j=1, \dots, n}$ , with the Kronecker delta symbol  $\delta_{ij} = 1$  if  $i = j$  and 0 otherwise. The derivative  $Df(x)$  in (2.5) is also known as the Jacobian, denoted as  $Jf(x)$ . The capital letter version  $Df(x)$  is used together with the presence of a basis for the tangent space, i.e. the codomain of the derivative, which allows representing the derivative as a matrix. In the absence

## 2 Background

of such a basis, we usually use the small letter version  $df(x)$  to symbolize a more abstract differential between domain and tangential space. This is for example the case for infinite-dimensional domains ( $n = \infty$ ), where we refer to Section 2.2.5.

Derivatives can also be applied multiple times, and as an example, we shed light on the second derivative  $D(Df)$  of a smooth real-valued function  $f : D \subset \mathbb{R}^n \rightarrow \mathbb{R}$ : This object  $Hf = D(Df)$  is called the Hessian and has the following structure:

$$Hf(x) = \begin{pmatrix} \partial_{x_1} \partial_{x_1} f(x) & \cdots & \partial_{x_1} \partial_{x_n} f(x) \\ \vdots & \ddots & \vdots \\ \partial_{x_n} \partial_{x_1} f(x) & \cdots & \partial_{x_n} \partial_{x_n} f(x) \end{pmatrix} \in \mathbb{R}^{n \times n}. \quad (2.6)$$

It is symmetric, i.e.  $\partial_{x_i} \partial_{x_j} = \partial_{x_j} \partial_{x_i}$ , if the second derivative of  $f$  is continuous. Furthermore, the matrix  $Hf(x)$  at position  $x$  gives important information on the local curvature – cf. the constraints in Equation (2.18) – and can be analyzed through spectral analysis by calculating eigenvalues and eigenvectors. The benefits of using Hessian information are for example to achieve algorithms with a convergence rate of second-order, or to make sure that a critical point  $x^*$  with  $Df(x^*) = 0$  is not a saddle point, cf. Figure 2.6.

Going back to first derivatives, if we can define a scalar product or a duality between derivatives and normalized displacements  $\Delta x$  with  $\|\Delta x\| = 1$ , we find the following important connection:

$$\partial_{\Delta x} f(x) = df(x) \Delta x. \quad (2.7)$$

The derivatives  $df(x)$  actually are cotangents, meaning that they are (linear) functions from the tangent spaces (where  $\Delta x$  live in) to the value space of  $f$ , as seen in (2.7): starting with the derivative  $df(x)$  of a smooth function  $f : D \rightarrow C$  at location  $x \in D$ , one needs to evaluate the (linear) function  $df(x)$  at location  $\Delta x$  to get  $\Delta y = df(x)(\Delta x)$ . However, there are different concepts of applying derivatives and by using, for example, duality brackets, scalar products or matrix products, or the representation with the *nabla* operator  $\nabla f = Df^T$ , one finds some of the following different notations:

$$\begin{aligned} \Delta y &= \partial_{\Delta x} f(x) = df(x)(\Delta x) \\ &= df(x) \Delta x = \langle df(x), \Delta x \rangle = df(x) \cdot \Delta x = \nabla f(x)^T \Delta x = Df(x) \Delta x. \end{aligned} \quad (2.8)$$

Furthermore, scalar products may be represented as sums or integrals (cf. next paragraph)

$$df(x) \Delta x = \sum_{i \in D} (df(x))_i \Delta x_i = \int_D df(x)(i) \Delta x(i) di, \quad (2.9)$$

using the sum for finite or countably infinite domains, and the integral notation for uncountably infinite domains.

Going back to our initial situation described in (2.2), one can linearly approximate the change towards a step  $\Delta x$  with

$$f(x + \Delta x) = f(x) + df(x) \Delta x + o(|\Delta x|) \approx f(x) + df(x) \Delta x. \quad (2.10)$$

This linear description of the tangent is also called first-order Taylor expansion at position  $x$ .

At positions where the function of interest has jumps or kinks, such Taylor expansions fail because it is not clear how to place the tangent space or how to calculate the derivative. However, for a non-differentiable point surrounded by a smooth setting, as is, for example, the case at the origin of the absolute function  $|\cdot|$ , it is still desirable to define such a derivative. A straightforward construction would be to take the mean value of the left and the right-sided differential quotients. It turns out that this idea, together with the following idea for a so-called weak derivative is very useful: inside an integral over a continuous domain, changes at one location (or countably many locations) of a function do not change the integral's value. Thus one considers defining derivatives inside integrals. These ideas lead to the definitions of weak or distributional derivatives  $\delta f$  of functions or distributions  $f$ , where the property

$$\langle \delta f, h \rangle = -\langle f, dh \rangle \quad (2.11)$$

needs to hold for every smooth and compactly supported test function  $h$  [62]. The equality in (2.11) is basically integration by parts, where the boundary part  $\int_{\partial \text{Domain}(h)} \delta f dh dx$  vanishes since  $h$  is compactly supported. This construction enables the calculation of weak derivatives of functions with kinks and distributional derivatives of functions with jumps by leveraging integration by parts. It also introduces the set of distributions, which enlarges the set of functions with, for example, the famous Dirac  $\delta$  distribution as the distributional derivative of the Heaviside step function.

Finally, Sobolev spaces are used to analyze weakly differentiable functions concerning different kinds of smoothness and integrability degrees [63, Ch. 4]. The concepts established for Sobolev spaces are fundamental to examine mathematical problems in image processing, such as to calculate estimations, and to design proofs for robustness [62]. However, it is beyond the scope of this work to integrate its analysis, wherein this work it was more important to show the empirical and statistical performance of the different approaches. Furthermore, it is debatable, into which Sobolev space a distribution of real-world data such as medical MR images can be embedded.

Derivatives need not be calculated via the limits in their definitions since many applicational rules exist. For elementary functions, the derivatives are already known, and the derivatives of combinations of known functions can be calculated using the linearity, the product rule, and the chain rule. Application of these rules leads to the term of *automatic differentiation*, where the derivative of a process can be calculated exactly (except for numerical errors) using the knowledge of the process's elementary functions. Also widely used are *finite differences*, where a finite  $\Delta x$  or  $t$  is used to approximate the derivative.

As a direct connection to image processing, the derivative of an image can be used to highlight edges, since edges are regions with high derivative [50, Sec. III.12.1]. To this end, one usually convolves the image with derivatives of a Gaussian to get slightly smoother values that depend less on noise. This works, because convolutions and derivatives commute.

## 2 Background

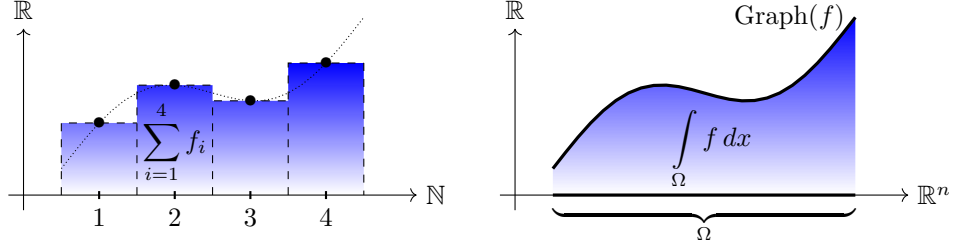


Figure 2.3: Integral as the area under the graph on a discrete and a continuous domain.

Derivatives are also used to optimize energies, as described in Section 2.2.4, since they locally point towards the steepest ascent and thus the negative gradient points towards the steepest descent. Therefore, when locally following the derivatives, one walks along a trajectory towards a local maximum.

**Integrals** The summation is not only a binary operation as in the term  $a + b$ , but one can also add three or more summands together. A summation over  $n$  summands  $(a_i)_{i=1,\dots,n} \subset A$  is denoted as

$$a_1 + \dots + a_n = \sum_{i=1}^n a_i. \quad (2.12)$$

Here  $(a_i)_i$  is a sequence, and  $a$  a function between the indices  $i$  and the sequence's values. Visualizing the sequence  $(a_i)_i$  in the product space between the indices and the values, one can interpret the above sum as the area under the function's graph as in Figure 2.3.

What happens if we add even more summands together, say infinitely many? Whether the sum

$$a_1 + a_2 + a_3 + \dots = \sum_{i=1}^{\infty} a_i \quad (2.13)$$

exists or not, depends on the summands and the set  $A$  in which these summands live. The sum on the left side of (2.13) can be interpreted as the limit of partial sums

$$a_1 + a_2 + a_3 + \dots = \sum_{i=1}^{\infty} a_i = \lim_{n \rightarrow \infty} \sum_{i=1}^n a_i. \quad (2.14)$$

This limit has the following possibilities: it may converge to an existing element in  $A$ , it can diverge to multiple accumulation points, or it can also diverge to elements outside of  $A$  [53]. For applications, we usually take care, that summations converge.

The cardinality of the index domain in the above construction may have different kinds of infinities. The discrete sequence  $(a_i)_i$  above was supposed to have countably infinite summands. What happens if we choose a continuous, uncountably infinite domain  $\Omega$ ? For  $i$  in  $\Omega$ , a summation like  $a_1 + a_2 + a_3 + \dots$  would not make sense since  $\Omega$  is uncountable and thus there might be missing values  $a_i$  between  $a_1$  and  $a_2$ . Therefore, let us choose a slightly different notation for the continuous case. We use  $x$  in  $\Omega$  instead of  $i$ , and a function  $f : \Omega \rightarrow A$  with the bracket argument notation  $f(x)$  instead of the subscript notation  $f_x$ . Now recall, that in  $\Omega$  we do not have a minimal distance between elements as compared to the natural index set. Thus in analogy to the discrete summation, where for example the distance between  $i$  and  $i + 1$  is 1, to denote a summation over all values  $f(x)$ , we introduce the arbitrarily small element  $dx$  and change symbol  $\sum$  to the symbol  $\int$ :

$$\sum_{i=1}^{\infty} a_i 1 \quad \text{in analogy to} \quad \int_{\Omega} f(x) dx. \quad (2.15)$$

The sum  $\int_{\Omega} f(x) dx$  of the values  $f(x)$  over the continuous domain  $\Omega$  is called an integral. The integral is the *entire* area under the graph of the integrand  $f$ , as visualized in Figure 2.3. Ultimately, to define such integrals as well as the arbitrarily small element  $dx$  through mathematical constructions, many different concepts have been introduced by Riemann, Stieltjes, Lebesgue, and others [55, 64]. The concepts of the above-mentioned mathematicians deal with the existence of the integral as a limit, and interpret  $dx$  as a symbolic necessity, as a differential form, or as a measure, respectively.

Indeed, the integral is a generalization of the sum: discrete sums can be seen as integrals by interpreting them as a Lebesgue integral using  $dx$  as the counting measure or by merely realizing the discrete sequence  $(a_i)_i$  as a piecewise-constant continuation with the discrete index set embedded in a continuous domain.

Integrals are often reformulated to different representations such that they offer simple calculations. For this purpose, a huge calculus has been developed over centuries, such as (a) integration by substitution, and (b) integration by parts using Stokes' theorem [55, pp. 100–124]:

- a) Some domains have a very complex structure. To integrate over complex domains, one either has to embed the domain into a flat real dimensional vector space or locally parametrize it with a part of a real dimensional vector space. For this purpose, the terminology of a manifold has been developed [65]. Similar as on a sphere's surface, a manifold is a set where one can draw a local flat map everywhere. In other words, if we can parametrize a part of the domain via a function whose derivatives are injective (e.g. full rank Jacobian matrix everywhere), then we can reformulate the integral of the domain as an integral over the parameters via integration by substitution.
- b) When trying to get rid of unwanted derivatives in an integrand, one can, if

## 2 Background

possible, use Stokes' theorem,

$$\int_{\Omega} d\omega = \int_{\partial\Omega} \omega, \quad (2.16)$$

to "push the derivative to the boundary". Here only formally mentioned,  $\Omega$  is a manifold,  $\partial\Omega$  the boundary or surface of  $\Omega$ ,  $\omega$  a special kind of a differential form, and  $d$  the exterior derivative [65]. As a special case of (2.16), we state Gauss's theorem or the divergence theorem as follows:

$$\int_{\Omega} \operatorname{div}(u p) dx = \int_{\Omega} \nabla \cdot (u p) dx = \int_{\Omega} u \nabla \cdot p + \nabla u \cdot p dx = \int_{\partial\Omega} u p \cdot n(s) ds, \quad (2.17)$$

where  $u : \Omega \rightarrow \mathbb{R}$  is a scalar valued function,  $p : \Omega \rightarrow \mathbb{R}^{\dim \Omega}$  is a vector field of the same dimension as the manifold  $\Omega$ , and  $n(s)$  is the normal vector of the surface  $\partial\Omega$  at the surface location  $s$ . Equation (2.17) includes the practical product rule for divergence,

$$\operatorname{div}(u p) = u \operatorname{div}(p) + \operatorname{grad}(u) \cdot p,$$

where  $\operatorname{div}(p) = \nabla \cdot p = \operatorname{trace}(Dp)$  is the divergence operator,  $\nabla$  is the nabla operator, and " $\cdot$ " is the scalar product in  $\mathbb{R}^{\dim \Omega}$ .

Finally, for many digital applications integrals on continuous domains – if not solvable analytically – at some point will boil down to a finite summation, since infinitely many values of  $f(x)$  cannot be stored in digital memory. To this end, continuous domains have to be discretized, and functions have to be sampled on a discretization.

**Properties of Functions** In this paragraph, we mention important analytical properties of functions such as continuity, differentiability, smoothness, integrability, and convexity. These properties, and also many others, are important for analyzing and describing optimization problems.

- A function  $f$  has a local minimum in  $x^*$ , if  $f(x^*)$  is the lowest value in an open/local neighborhood of  $x^*$ . For smooth function  $f$  we find the following sufficient and necessary conditions for a local minimum  $x^*$ :

$$\left. \begin{array}{l} Df(x^*) = 0 \\ Hf(x^*) \text{ is PD} \end{array} \right\} \xRightarrow{\neq, x^4} x^* \text{ is a local minimum of } f \xRightarrow{\neq, x^3} \left\{ \begin{array}{l} Df(x^*) = 0 \\ Hf(x^*) \text{ is PSD} \end{array} \right., \quad (2.18)$$

where  $Df$  is the first, and  $Hf$  the second derivative, PD means positive definite, i.e.  $\Delta x^T Hf(x^*) \Delta x > 0 \forall \Delta x \neq 0$ , and PSD positive semi-definite, i.e.  $\Delta x^T Hf(x^*) \Delta x \geq 0 \forall \Delta x \neq 0$ . Note that the left implication is sufficient only, and the right implication is necessary only, since  $f(x) = x^4$  is a counterexample with  $x^* = 0$  as a minimum with curvature 0 and thus not PD, and  $f(x) = x^3$  is a counterexample with a PSD curved saddle point in  $x^* = 0$ .



- In topological terms, a function is *continuous* if the inverse images of all open sets of the codomain's topology again are open in the domains' topology, and in other words, a function is *continuous* if small changes of the parameters  $x$  only cause small changes in the values  $f(x)$  (but this depends on how "small changes" are defined).

In contrast to the above topological definition, from an applicational point of view, where data is usually stored in a discretized form, a *continuous* function simply refers to the data's possible analytic function expression, ignoring whether its source actually is continuous or non-continuous – cf. Page 18 Item 7.

- A function is called lower semi-continuous at a point  $x_0$  if all function value accumulation points of sequences that converge to  $x_0$  are larger than or equal to  $f(x_0)$ , i.e. if

$$\liminf_{x \rightarrow x_0} f(x) \geq f(x_0), \quad (2.19)$$

which means, that all accumulation points of  $(f(x_n))_n$  of all sequences  $(x_n)_n$  with  $\lim_{n \rightarrow \infty} x_n = x_0$  are larger than or equal to  $f(x_0)$ . This setting corresponds to, for example, functions with a jump in  $x_0$  and the value  $f(x_0)$  being on the lower end. Similarly also upper semi-continuity is defined.

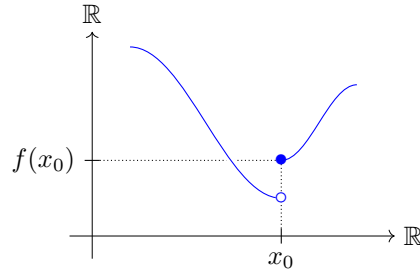


Figure 2.4: An upper semi-continuous function with a jump at  $x_0$ . The minimum does not exist.

Lower semi-continuity is a property involved in guaranteeing the existence of minimal values [49, Sec. 2.1.2]: if for example  $f$  is not lower semi-continuous at  $x_0$ , and  $x_0$  indeed would be the only candidate for an argument minimum as in Figure 2.4, then the minimum does not exist because the infimal value  $\liminf_{x \rightarrow x_0} f(x)$  is not accepted by any  $x$ , since infinitely close to  $x_0$  there are function values lower than  $f(x_0)$ , but all higher than  $\liminf_{x \rightarrow x_0} f(x)$ . In the discrete case, where  $x_0$  would be an isolated point, this scenario, of course, could not take place.

- A function is *differentiable* in the strong/classical sense, if for every point in the interior of the domain a (linear) map exists that represents the tangent

## 2 Background

space (Fréchet/Gateaux differential) [51]. If a function is differentiable, then all the partial derivatives (2.4) exist for every point in the domain. For some applications, for example, when modeling an energy functional with an integral, the derivatives of the involved functions under the integral do not necessarily need to exist for every point. Therefore less strong conditions such as the *weak* or *distributional differentiability* (2.11) exist [62]. Furthermore, if a putative weak differential cannot be expressed with a locally integrable  $L^1$  function one generalizes with a distribution, as is the case for example with the Dirac delta distribution as the distributional derivative of the Heaviside function.

- From continuous to differentiable to many times differentiable, a function is called *smooth* if it is arbitrarily many times differentiable. For formal calculations, one can also lazily assume a function to be *smooth enough* to ensure that all involved limits exist.
- A function is called *integrable* if its integral over the whole domain exists as a real number. The function is called locally integrable if it is integrable on every compact subset of the domain. Many theorems help to determine whether integrals exist or not, such as the monotone convergence or the dominated convergence theorems [55, pp. 54–69] or [59, Appendix E]. In the applied field, however, we often formally integrate without thinking of convergence. In many cases, however, we simply are summing bounded functions on a finite set, which always leads to a finite sum.
- Intuitively speaking, a function is called *convex* if it looks similar to a positively curved parabola, cf. with graphs on the right side of Figure 2.6 on Page 33. More precisely, a real-valued function  $f : \Omega \rightarrow \mathbb{R}$  is convex if the straight line between any two different points on the graph of  $f$  does not fall below the graph between those points [49, Sec. 2.1.2], i.e. if

$$\forall x_1, x_2 \in \Omega, \quad \forall \lambda \in ]0, 1[, \quad \lambda f(x_1) + (1 - \lambda) f(x_2) \geq f(\lambda x_1 + (1 - \lambda) x_2). \quad (2.20)$$

In other words, the function is convex, if everything that is above the graph, called the epigraph or supergraph, is convex as a set; where a set is said to be convex, if straight lines between any two points in the set are contained in the set. Furthermore, a function is called *strictly convex*, if the straight line between two points, excluding the points themselves, is above the graph, i.e. if

$$\forall x_1, x_2 \in \Omega, \quad \forall \lambda \in ]0, 1[, \quad \lambda f(x_1) + (1 - \lambda) f(x_2) > f(\lambda x_1 + (1 - \lambda) x_2). \quad (2.21)$$

Moreover, for a differentiable function  $f$ , convexity and strict convexity can very elegantly be expressed using the second derivative: if the second derivative (the Hessian matrix) is non-negative definite everywhere, then the function is convex, and if the second derivative is positive definite everywhere, then  $f$  is strictly convex.

For minimization, convexity is a very important property, since a convex function guarantees a unique minimal value [49, Sec. 2.1.2], also called a unique global minimum. If also the function is strictly convex, then the unique minimal value is also realized by a unique location, also called the unique global argument minimum. If the objective function is not convex, it may have different local minima, which would complicate the search for a global minimum. (In analogy, the same holds for maximization with concave functions.)

- A coercive function is a function that reaches arbitrarily high values towards any direction of the boundary of the domain [49, Sec. 2.1.2], i.e.

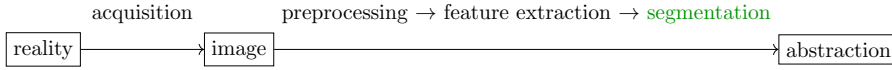
$$\forall (x_n)_n \text{ with } \lim_{n \rightarrow \infty} x_n = +\infty \quad \text{we have } \lim_{n \rightarrow \infty} f(x_n) = +\infty. \quad (2.22)$$

This simply corresponds to functions like the top right curve in Figure 2.6 on Page 33. It is a natural condition we often use when designing a functional with a centrally arranged argument minimum, as is the case when using, for example, the Euclidean norm on  $\mathbb{R}^n$ .

### 2.2.2 Computer Vision

To mimic human visual understanding, many models exist in computer vision [66]. Computer vision categorizes and studies the process of abstracting or projecting information out of a signal or an image [67, Sec. 1]. Roughly said, it is an inverse process of painting, wherein painting some (higher-level) information is instantiated by an exemplary image.

In the following graph, we see a typical pipeline of the computer vision task of image segmentation:



After the acquisition of an image, usually a preprocessing step is applied. Preprocessing steps could be, for example, re-sampling to a higher or lower resolution, noise reduction, contrast enhancement, or edge enhancement. Next, to clarify an actual segmentation, in the feature extraction phase, the characteristics of the different objects or regions are being identified. Then, with the identified different characteristics, the image domain can be segmented into the different regions. And finally, using the segmented regions, further steps can be conducted to formulate a statement. One could, for example, give the regions their names, calculate their areas, or make any kind of conclusion. The output of this process is an abstraction or a projection of the image, meaning that the output may be better understood but only includes parts of the information of the initial data. For us, most importantly, "at some point in the processing, a decision is made about which image points are relevant for further processing" [68]. Moreover, the processing steps from image to abstraction, which here

## 2 Background

are displayed in a serially distinct manner, actually often are mixed or are applied recursively.

**Segmentation** The goal of segmentation is to identify regions, called segments, where the pixels share a specific uniformity criterion. Depending on the task, the different classes are preferred to have as few segments as possible. This way the information of the image is represented such that it is easier to use for further analysis.

In computer vision, there is also the notion of classification, which refers to object classification or object recognition, giving the whole image or parts of the image a label. Note that segmentation in this sense is a pixelwise classification.

The word segmentation is ambiguous and stands for the process as well as its result. However, this is the same ambiguity as to whether with a mathematical function  $f(x)$  we mean the function itself or the value of the function at location  $x$ . In context, it is usually semantically clear to which meaning we refer.

In Sections 2.2.6 and 2.2.7 we describe in more detail how such segmentations can be represented and generated on computers, and in Section 2.2.8 we give examples for variational-based segmentation algorithms.

**Modeling** In modeling such algorithms, we can identify a sequence of three inclusions of implicitly defined philosophical domains, depicted in Figure 2.5. In these domains, similar research problems may be formulated differently, and the arrows reflect their connection or development steps. Note that in practice, a step from

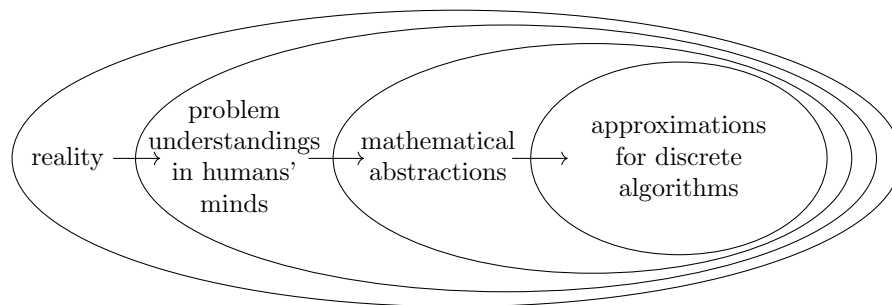


Figure 2.5: Philosophical domains of modeling a problem.

left to right involves approximations, and the approximated problem probably does not necessarily describe the same as the initial problem anymore. Using reality as a container for everything, the modeling process could be seen as a sequence of inclusions or projections into subspaces. Now, while mathematical abstractions in this conception are seen as objects from the reality, a computer engineer typically makes several iterations with the pipeline in Figure 2.5 to realize and implement a computer vision algorithm. Implementing an algorithm mainly focuses on the last step between

mathematical abstractions and approximations for discrete algorithms. The conduct for the last inclusion in Figure 2.5 could be denoted as: "Math looks nice, but for real-world applications, a lot of engineering hacks are needed." We hypothesize that often the most significant errors occur between reality and its first models. It probably won't make an already incorrect model catastrophically more incorrect by making computable assumptions.

**Examples in Modeling** The ultimate aim of medical research, to better understand and treat pain, is a difficult task: while a patient may complain about a specific feeling (which could be seen as ground truth), it can be very difficult to understand its origin. To this end, already inside the second bubble of Figure 2.5 countless many theories may exist for observations such as the signal of pain.

As an example, we refer to the clinicoradiological paradox [11]. In case of this paradox in MS, it may seem reasonable that patients with MR imaging abnormalities are worse off than patients without, with e.g. hypo- and hyper-intense spots indicating the presence of lesions at certain locations. However, MR imaging abnormalities and the patient status regarding symptoms do not correlate accordingly. Moreover, a recent study [41] described cases, where such spots were apparent in MR images, which, however, were even not identifiable on the same histological slices. Associations between clinical and radiological findings, therefore, generally are poorly understood.

As another example, we are interested in area and volume measurements of specific structures in the SC. While the correct morphology is hidden in the reality, we derive methods to abstractly represent images of the SC in-vivo and develop post-processing tools to analyze such images. In each step, approximations are involved, and thus we can never be entirely sure that projective calculations such as areas or volumes of these structures correspond to reality.

Therefore, although it lies in our nature to make simplifications to understand observations better, we should not rely on simple abstractions but should take advantage of the full information, as was also clarified by the authors in [11]. And indeed, computer vision is going towards the direction of making full use of given information. This tendency is demonstrated, for example, when comparing segmentation algorithms that only use pixelwise thresholding to segmentation networks such as MDGRU [24] that use the whole image content for their receptive field. Furthermore, systems that combine several task-specific networks may further advance towards this direction of taking full advantage of the given information.

### 2.2.3 Digital Images

A digital image  $I$  is a finite and ordered set, sequence, or array of picture elements [49, Sec. 1.2], called pixels, and can be denoted as  $(I_x)_{x \in \Omega}$  (spoken: the series of all intensities  $I$  at location  $x$  such that  $x$  is in the domain  $\Omega$ ), cf. Section 2.2.1. Here  $\Omega$  is a set of pixel locations and is called the image domain.

The image  $I$  can also be represented or denoted as a mathematical function  $I : \Omega \rightarrow \mathcal{I}$ , where  $\Omega$  is a finite subset of the  $d$ -dimensional Euclidean space  $\mathbb{R}^d$ , and  $\mathcal{I}$  is

## 2 Background

a set of possible intensity values. For gray-valued images,  $\mathcal{I}$  could be a subset in  $\mathbb{R}$ , and for colored images, we need to choose a color model in  $\mathcal{I} \subset \mathbb{R}^3$ .

However, the two definitions above are equivalent, as the pixel elements with the index notation  $I_x$  and bracket notation  $I(x)$  can be identified.

Furthermore, a digital image  $I : \Omega \rightarrow \mathcal{I}$  should not be confused with the mathematical image of a function  $I$ , which would be the unordered set of all occurring intensities. To clarify, the image of  $I$ ,

$$\text{Image}(I) = I(\Omega) = \{I(x) \mid x \in \Omega\},$$

is a subset of the codomain  $\mathcal{I}$ , and a (painted) image  $I$  actually is the graph of  $I$ ,

$$\text{Graph}(I) = \{(x, I(x)) \mid x \in \Omega\},$$

consisting of tuples  $(x, I(x))$ , where the intensities  $I(x)$  or  $I_x$  at locations  $x$  are usually visualized as gray or color values. As demonstrated in Figure 2.2, graphs can be represented differently as with displaying 1D intensities as heights in the canonical product space, or by displaying 1D intensities as gray values. Similarly, 2D or 3D intensities could be represented as typical vector fields or as colors using a color model, depending on whether colors or vectors are more meaningful.

Digital images  $I$  are discrete, which means, that their domains and hence also their mathematical images  $\text{Image}(I)$  are discrete. Therefore, creating a digital image deals with sampling and quantization. Sampling is concerned with an adequate discretization of the domain, and quantization with the discretization of the codomain. In conjunction with Fourier analysis, the Nyquist–Shannon sampling theorem becomes essential. This theorem states that if a function  $I$  contains no frequencies higher than  $B$ , it is completely determined by representing its values on a domain with a pixel spacing smaller than  $1/(2B)$  [69], otherwise aliasing effects might occur.

**Basic Image Manipulations** An image  $I : \Omega \rightarrow \mathcal{I}$  can be manipulated in the following ways: purely by deforming its domain with a deformation  $d : \hat{\Omega} \rightarrow \Omega$  by concatenating  $d$  before  $I$  as

$$I \circ d : \begin{array}{ccccc} \hat{\Omega} & \xrightarrow{d} & \Omega & \xrightarrow{I} & \mathcal{I} \\ x & \longmapsto & d(x) & \longmapsto & (I \circ d)(x) = I(d(x)); \end{array} \quad (2.23)$$

purely by adjusting the intensities with a function  $f : \mathcal{I} \rightarrow \hat{\mathcal{I}}$  by concatenating  $f$  after  $I$  as

$$f \circ I : \begin{array}{ccccc} \Omega & \xrightarrow{I} & \mathcal{I} & \xrightarrow{f} & \hat{\mathcal{I}} \\ x & \longmapsto & I(x) & \longmapsto & (f \circ I)(x) = f(I(x)), \end{array} \quad (2.24)$$

or actually through any kind of process  $I \mapsto \hat{I}(x, I, \lambda)$  that depends on any kind of parameter  $\lambda$ . Here  $\hat{\Omega}$  and  $\hat{\mathcal{I}}$  represent some image domain and intensity range, that could also be the same as  $\Omega$  and  $\mathcal{I}$ , respectively.

A typical example where image deformations (2.23) are used, is image registration between two images. During registration of two images, their domains  $\Omega$  are transformed such that corresponding image landmarks of both images coincide [70]. Furthermore, changing the domain to  $\hat{\Omega}$  can also be useful to represent the image on another domain: for example, to embed a plane  $\hat{\Omega}$  onto the surface of some object  $\Omega$ . On the other side, pure intensity manipulations (2.24) are mainly used, for example, to adjust the contrast with gamma correction, thresholding, or more generally with a lookup table.

For adding or removing image noise [71], or applying image segmentation, additional knowledge needs to be included in mapping. To this end, simple concatenations as (2.23) or (2.24) might not be enough to describe image manipulations like  $I \mapsto \hat{I}(x, I, \lambda)$  that depend on the whole image  $I$  and additional information  $\lambda$ .

### 2.2.4 Variational Principle

A powerful tool to model a problem or an algorithm is the *variational principle*, where a solution is found as an extremum, e.g. maximum or minimum, of an objective function using the calculus of variation [72, 49] [59, Sec. 8.1]. To this end, we define a function space  $V$  containing the admissible solutions, and a functional  $E : V \rightarrow \mathbb{R}$  that scores every admissible solution with a real-valued number. Here, the term functional indicates a real-valued function with a vector space or function space as the domain. If we look for a solution with a minimal score, this functional  $E$  is often called energy, cost, loss, or objective functional. Conversely, it could be called a profit, fitness, or again energy function, if it is being maximized. We denote such optimization problems as

$$\operatorname{argmin}_{v \in V} E(v) \quad \text{or} \quad \operatorname{argmax}_{v \in V} E(v). \quad (2.25)$$

For a correctly modeled problem, the desired solution(s) should be among the argument(s) of the minimum  $\operatorname{argmin}$  (or the argument(s) of the maximum  $\operatorname{argmax}$ ), where the energy is minimal (or maximal), i.e. where the energy cannot be decreased (or increased) anymore. Mathematically, there is no big difference between maximization and minimization, apart from a sign, since every maximization problem can be reformulated to a minimization problem via

$$\max_{v \in V} E(v) = -\min_{v \in V} (-E(v)). \quad (2.26)$$

This principle is motivated from physics, where the extrema often describe a stable state into which systems converge after a settling time. The same principle can be used to derive image processing algorithms, as demonstrated in Section 2.2.8.

**Solving an Optimization Problem** In simple cases, minima can be calculated analytically (this means, with algebraic transformations), but in more complex cases, often only some of the minima can be found using numerical methods. For this purpose, many numerical methods have been proposed: with global optimization methods such

## 2 Background

as the naive approach *exhaustive search* or *DIRECT* [73]; and with local optimization methods such as line search descent methods in various discretizations [49, 50, 52]. Choosing a suitable method depends on the properties of the energy functional  $E$ . However, this work does not focus on applying the different optimization methods, and for simplicity, we stick to gradient descent in this section. Moreover, there is nothing wrong with gradient descent in combination with proper initialization and mainly convex energy functionals, as explained in the next paragraph.

From an abstract point of view, the gradient descent method applied to the energy  $E$  is a numerical approximation of the *gradient descent differential equation*

$$\partial_t u = -\frac{\delta E}{\delta u}, \quad (2.27)$$

where  $\frac{\delta E}{\delta u}$  represents the variational derivative of  $E$ , defined in Section 2.2.5. The derivatives, or also called the gradients of the energy, locally are pointing towards the steepest ascent, and conversely, the negative gradients are pointing towards the steepest descent. Therefore, the gradients can be used to stepwise improve the currently guessed solutions until all directional derivatives at that position are zero. Using the concept of iterations, we initialize the sought solution with a first guess, denoting it in the super-script notation  $u^0$ , and then use the algorithm's formula  $\mathcal{A}$  to derive the next iteration  $u^{k+1} = \mathcal{A}(u^k)$  for any  $k \geq 0$ . As an example for  $\mathcal{A}$ , we can discretize the gradient descent equation (2.27) with the explicit Euler method to derive

$$u^{k+1} = \mathcal{A}(u^k) = u^k - \Delta t \frac{\delta E}{\delta u}(u^k), \quad (2.28)$$

where  $\Delta t$  is the discretization of the artificially introduced time parameter  $t$ , cf. [50, 52].

**Designing an Energy Function** Defining the function space  $V$  and the functional  $E$  is, of course, problem-specific; however, certain regularizations have to be considered to simplify the optimization process. For this purpose, we introduce basic terms regarding the well-posedness of a problem and therefore depict in Figure 2.6 different sorts of extrema. As already mentioned with Equation (2.18) for the smooth case, extrema might appear where the derivatives are 0, *but* they can also be located on the boundary of  $V$ . Therefore, to apply gradient descent methods more easily and robustly, energies should be designed as smooth and as convex as possible, and the points of interest should not lie towards the boundary of  $V$ . A captious but important fact is that for an interior point the extremality condition on the gradient,

$$\frac{\delta E}{\delta u}(u^k) = 0, \quad (2.29)$$

is only a necessary but not a sufficient condition to provide maxima or minima, cf. (2.18). A point satisfying the condition in (2.29) is called a *critical* or *stationary point*, which includes among maxima and minima also *saddle points*, as depicted in



## 2.2 Technical Background

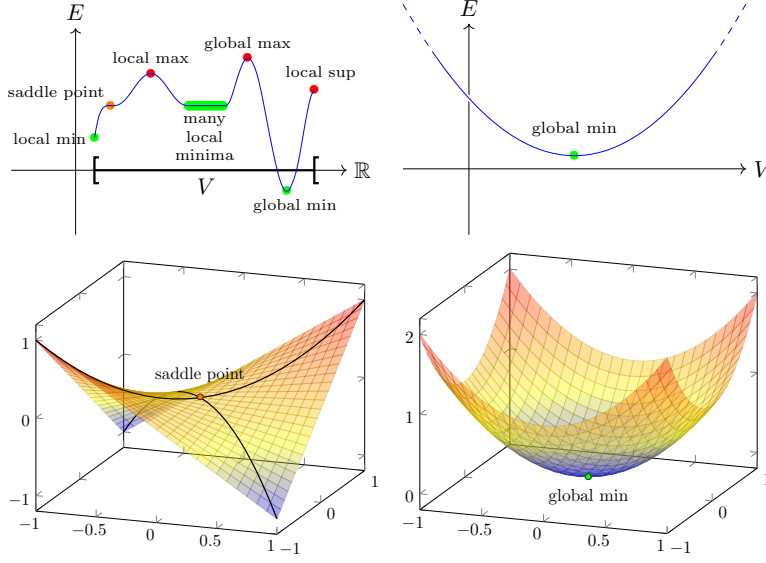


Figure 2.6: Examples of energies' extrema.

Figure 2.6. A saddle point has a zero gradient, but the curvature may be positive towards one direction and negative towards another, and such a point thus would not describe a correct solution for a minimization problem. The existence of saddle points poses an additional hurdle for gradient descent algorithms as their goal is to find critical points only. Nevertheless, in practice gradient descents almost always converge to minima [74]. As exemplified in the bottom left illustration in Figure 2.6, the probability to randomly initialize a starting point for a trajectory that converges to a saddle point is zero, since the paths of bad initial points (drawn in **black**) are zero-measure sets with respect to the area measure. In terms of gradient fields, saddle points are often attracting only along a few trajectories but are otherwise unstable.

In an ideal case, a variational formulation would be formulated such that

1. a solution exists,
2. the solution is unique, and
3. the solution continuously depends on the given data and parameters.

If these three conditions are met, the given problem is said to be *well-posed*, otherwise *ill-posed* [49, Sec. 2.5.6]. To this end, the following sufficient conditions are useful and are often considered when designing variational formulations:

- If  $E$  is continuous on a compact domain  $V$  then a minimum of  $E$  exists.

## 2 Background

- If  $E$  is lower semi-continuous (2.19) and coercive (2.22) then a minimum of  $E$  exists [49, Sec. 2.1.2].
- If  $E$  is convex and has a minimum, then the minimum (the minimum value of  $E$ ) is unique.
- If  $E$  is strictly convex and has a minimum, then the argument  $v \in V$  of the minimum value is unique.

Unfortunately, problems such as segmentation and registration of medical images naturally are ill-posed problems [75]: in many cases, it is unclear whether an optimal solution exists; some cases have many different optimal solutions; and sometimes already small changes in the data lead to unpredictably large differences in the result [76].

Therefore, the following **fundamental idea** is essential in designing a variational model: energies are composed of two terms, the data term and the regularization term. The data term is some distance between the input data and the admissible solutions, and the regularization term includes a prior assumption that is not already present in the space of the admissible solutions. Using only the data term often leads to highly ill-posed formulations, and therefore further conditions are included to regularize the formulation. From another point of view, regularization terms are conditions from the space of the admissible solutions included into the energy in the form of a Lagrangian.

In our case, the input data are medical images, and the space of admissible solutions would be a model space, for example, the space of segmentations. The distance between an image and a segmentation could be, for example, the sum of squared differences (SSD). A typical regularization of the solution would include, for example, the Euclidean norm of latent variables to avoid numerically diverging values, smoothness constraints such as total variation (TV)-regularization to deal with image noise, or distances to a statistical shape model for staying close to realistic representations.

As a thought experiment, we could model the human visual understanding as a variational problem: the input images are the visual information coming from the optic nerves, the function space is the set of all possible visual 3D interpretations in one human's mind, the data term is some intensity-based distance between the optical input and its interpretation, and the regularization term includes prior knowledge from the cerebral cortex as a weighting function that favors more plausible admissible solutions. With such a model, optical illusions, for example, would be explained with the regularization term misleading us to an illusory visual interpretation from the brain. Dreams, for example, could be modeled with noisy close-to-zero inputs from the eye's photoreceptor cells that are regularized to the scenery that reflects the current state of mind. This setup, however, is not meant to describe the human visual understanding accurately, but to provide an example of how variational formulations may be used for computer vision tasks, as is addressed with so called human visual system models [66].

**Connection to Bayesian Inference** From a mathematical point of view, variational formulations are direct applications of the mathematical field of functional analysis and show via the Euler-Lagrange equations (2.27) and (2.40) an interesting connection to differential equations, as depicted on the right side of Figure 2.7. On the other

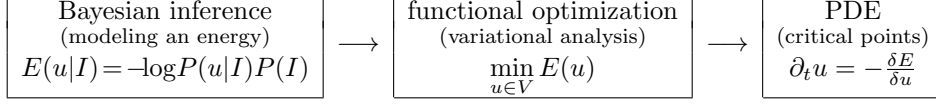


Figure 2.7: Pipeline of the variational modeling framework.

hand, objective functions can be derived using Bayesian inference models, as depicted on the left side of Figure 2.7. Bayesian inference links statistics to optimization [77], and we therefore briefly introduce in this paragraph the construction of energies via negative logarithms of posterior probability maps, following [78].

Let  $\hat{A} : \Omega_1 \rightarrow S_1$  and  $\hat{B} : \Omega_2 \rightarrow S_1$  be two random variables of two probability spaces  $(\Omega_1, \mathcal{F}_1, P_1)$  and  $(\Omega_2, \mathcal{F}_2, P_2)$  from the reality into two probability spaces  $(S_1, \mathcal{A}_1, P_{S_1})$  and  $(S_2, \mathcal{A}_2, P_{S_2})$  from the measurement space. Exemplarily here for  $\hat{A}$ , we define the probability that the random variable  $\hat{A}$  takes on an event  $A$  of the  $\sigma$ -algebra  $\mathcal{A}$  by

$$P_1(A) := P_1(\hat{A} \in A) := P_1\left(\left\{x \in \Omega_1 \mid \hat{A}(x) \in A\right\}\right). \quad (2.30)$$

Furthermore, conditional probabilities of events  $A \in \mathcal{A}$  and  $B \in \mathcal{B}$  are defined as

$$P(B|A) := \frac{P(B \cap A)}{P(A)}, \quad (2.31)$$

using the joint probability distribution  $P$  of  $P_1$  and  $P_2$  and the intersection  $A \cap B$  in the joint probability space  $S$ . The use of the joint probability also allows dropping the indices of  $P_1(A)$  and  $P_2(B)$  to  $P(A)$  and  $P(B)$ , respectively, in interpreting them as marginals. The relation between  $P(A|B)$  and  $P(B|A)$  is given by Bayes' formula

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}. \quad (2.32)$$

These notations make sense for discrete random variables, however for continuous random variables, singleton events most probably have the probability 0, and therefore *probability densities* are used: for singletons  $A = \{a\}$ , we identify the notations

$$P(\hat{A} \in A) = P(A) = P(a) = P(\hat{A} = a), \quad (2.33)$$

though, for continuous random variables, the probability  $P(a)$  instead has to be interpreted as a probability density (oversimplified, to not introduce further variables).

## 2 Background

Table 2.1: Overview of the prior, likelihood, and posterior densities; MLE, and MAP.

$P(I), P(u)$	prior probability density functions
$P(I u)$	<u>likelihood</u> density function
$P(u I) = \frac{P(I u)P(u)}{P(I)}$	<u>posterior</u> probability density function
$u_{\text{ML}} = \underset{u}{\operatorname{argmax}} P(I u)$	Maximum Likelihood Estimation (MLE)
$u_{\text{MAP}} = \underset{u}{\operatorname{argmax}} P(u I)$	Maximum A Posteriori Probability Estimation (MAP)

Let us now assume we want to infer the most likely segmentation  $u$  given an image  $I$ . We want to infer information  $u \in S_1$  out of some measurement  $I \in S_2$ . Finding the most probable inference means that we are abstractly searching for the most likely real 3D scene in  $\hat{A}^{-1}(\{I\}) \subset \Omega_1$  and the most likely true anatomical differentiation in the 3D scene in  $\hat{B}^{-1}(\{u\}) \subset \Omega_2$ . However, since we cannot perform this in the spaces  $\Omega_1$  and  $\Omega_2$  from the reality, we, therefore, use the imaging process  $\hat{A}$  and the segmentation process  $\hat{B}$  to represent the real objects as  $u$  and  $I$ .

We introduce further important density functions and their maximum estimations in Table 2.1. The maximum a posteriori probability estimation  $u_{\text{MAP}}$  can be reformulated as follows:

$$u_{\text{MAP}} = \underset{u}{\operatorname{argmax}} P(u|I) = \underset{u}{\operatorname{argmax}} \frac{\overset{\text{a prior enters}}{\overbrace{P(I|u) P(u)}}}{\underbrace{P(I)}_{\text{constant w.r.t. } u}} = \underset{u}{\operatorname{argmin}} -\log P(I|u)P(u), \quad (2.34)$$

where the prior  $P(I)$  in the denominator can be discarded, because it is a constant with respect to the segmentation  $u$  and thus inside the maximum, the posterior can be represented as the likelihood times the prior of the segmentation. Furthermore, using the monotonously increasing property of the logarithm and the min-to-max conversion (2.26) we end up with the right side of (2.34). Because of the linearity of the logarithm, the negative log posterior can be rewritten as

$$-\log P(u|I)P(I|u)P(u) = -\log P(I|u) - \log P(u), \quad (2.35)$$

and indeed, the right hand side of (2.35) has the form of

$$E = \text{likelihood} + \text{prior}, \quad (2.36)$$

as suggested before in the fundamental design on Page 34.

Being aware of the involvement of Bayesian inference in modeling an energy, prior knowledge can indeed simply be *added* to the energy function. Conversely, we often indirectly use Bayesian inference when adding a regularization term on  $u$  to the energy. Moreover, if prior information is available *then one should use it* because assuming a uniformly distributed prior  $P(u)$  results in the equality between MLE and MAP,

yet uniform distributions are rarely the case. Finally, note that many densities are designed using an exponential function such that the negative logarithm of the density directly extracts the negative exponent, as demonstrated in Section 2.2.8. Such models are called *log-linear models*.

### 2.2.5 Variational Derivative

René Gâteaux generalized the directional derivatives for a wide range of vector spaces also in the infinite-dimensional case, and these derivatives were used to formalize the variational derivatives. We only formally present definitions for variational derivatives, without going into detail whether the limits exist.

The *functional* or *variational derivative* of  $E$  in the direction  $h \in V$  is the Gâteaux derivative

$$\begin{aligned} \delta E(u)(h) &= \lim_{\epsilon \rightarrow 0} \frac{E(u + \epsilon h) - E(u)}{\epsilon} = \left. \frac{d}{d\epsilon} E(u + \epsilon h) \right|_{\epsilon=0} \\ &= \left\langle \frac{\delta E(u)}{\delta u}, h \right\rangle = \langle \delta_u E(u), h \rangle = \langle \delta E(u), h \rangle \\ &= \int_{\Omega} \frac{\delta E(u)}{\delta u}(x) h(x) dx = \int_{\Omega} \delta E(u) h(x) dx \\ &= \text{projection of the gradient } \frac{\delta E(u)}{\delta u} \text{ on the dimension in direction } h. \end{aligned} \tag{2.37}$$

We think of the operator  $\delta E(u) = \frac{\delta E(u)}{\delta u}$  as the gradient of  $E$  at the point  $u$ , and of  $\langle \frac{\delta E(u)}{\delta u}, h \rangle$  as the directional derivative.

With the fundamental lemma of the calculus of variations and density arguments of compactly supported smooth test functions in  $V$ , the extremality principle for the directional derivative and the gradient itself become equivalent:

$$\left\langle \frac{\delta E(u)}{\delta u}, h \right\rangle = 0 \quad \forall h \in V \quad \Leftrightarrow \quad \frac{\delta E(u)}{\delta u} = 0. \tag{2.38}$$

Energies are often formulated using integrals which results in the variational derivative having a particular structure described by the Euler-Lagrange equations [59, Sec. 8.1]. Let us, therefore, assume an energy of the following form:

$$E(u) = \int_{\Omega} F(x, u(x), du(x)) dx. \tag{2.39}$$

For image processing,  $\Omega$  typically is the image domain,  $F$  in this case a pixelwise smooth objective function,  $u$  an admissible solution that one is searching for, and  $d$  a differential operator, for example the nabla operator  $\nabla$ . With the concept of distributional derivatives, i.e. with integration by parts and a boundary condition

## 2 Background

(BC)  $\int_{\partial\Omega} \frac{\partial F}{\partial du} h \, d\sigma = 0$ , the *Euler-Lagrange equation* (in green) looks as follows:

$$\begin{aligned}
\delta E(u)(h) &= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int F(x, u(x) + \epsilon h(x), du(x) + \epsilon dh(x)) - F(x, u(x), du(x)) \, dx \\
&= \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \int F(x, u, du) + \epsilon \frac{\partial F}{\partial u} \cdot h + \epsilon \frac{\partial F}{\partial du} \cdot dh + o(\epsilon) - F(x, u, du) \, dx \\
&= \int \frac{\partial F}{\partial u} \cdot h + \frac{\partial F}{\partial du} \cdot dh \, dx \\
&= \int_{\Omega} \frac{\partial F}{\partial u} \cdot h + d^* \frac{\partial F}{\partial du} \cdot h \, dx - \int_{\partial\Omega} \frac{\partial F}{\partial du} h \, d\sigma \\
&= \int \underbrace{\left( \frac{\partial F}{\partial u} + d^* \frac{\partial F}{\partial du} \right)}_{=\delta E(u) \stackrel{!}{=} 0} \cdot h \, dx,
\end{aligned} \tag{2.40}$$

where  $d^*$  is the dual operator of  $d$ , found using integration by parts, e.g.  $\nabla^* = -\text{div}$  (2.17). Finally, the extremality condition on the functional derivative is equivalent to the following differential equation

$$\delta E(u)(h) = 0 \quad \Leftrightarrow \quad \begin{cases} \frac{\delta E}{\delta u} = \frac{\partial F}{\partial u} + d^* \frac{\partial F}{\partial du} = 0 & \text{in } \mathring{\Omega} \\ \frac{\partial F}{\partial du} h = 0 & \text{on } \partial\Omega. \end{cases} \tag{2.41}$$

Two extreme cases of the BC  $\frac{\partial F}{\partial du} h = 0$  in  $\partial\Omega$  are the

$$\begin{aligned}
&\text{homogeneous Dirichlet BC} & \frac{\partial F}{\partial du} = 0 & \text{on } \partial\Omega & \text{ and the} \\
&\text{homogeneous Neumann BC} & h = 0 & \text{on } \partial\Omega.
\end{aligned} \tag{2.42}$$

In an image such conditions hold, if for example the image has a uniform boundary (e.g. air in MR images) or is all around zero-padded.

### 2.2.6 Segmentation Representations

Given a digital image  $I : \Omega \rightarrow \mathcal{I}$ , a segmentation assigns a label to every pixel  $x \in \Omega$ . Let us denote the set of labels or classes with  $\mathcal{L}$ . We differentiate between binary segmentation, where the number of labels  $|\mathcal{L}|$  is 2, and multi-label segmentation, where  $|\mathcal{L}|$  is larger than 2.

Focusing on binary segmentation, where we differentiate between the object of interest and the background, we denote the set of object pixel locations as  $\mathcal{O}$  and the background as  $\mathcal{B} = \Omega \setminus \mathcal{O}$ . For implementation purposes, we usually use  $\mathcal{L} = \{0, 1\}$  with 0 corresponding to the background pixels and 1 to the object pixels. If, for example, the objective were to segment GM, 1 would map to GM and 0 to background, which here would be WM and everything else that is not GM.

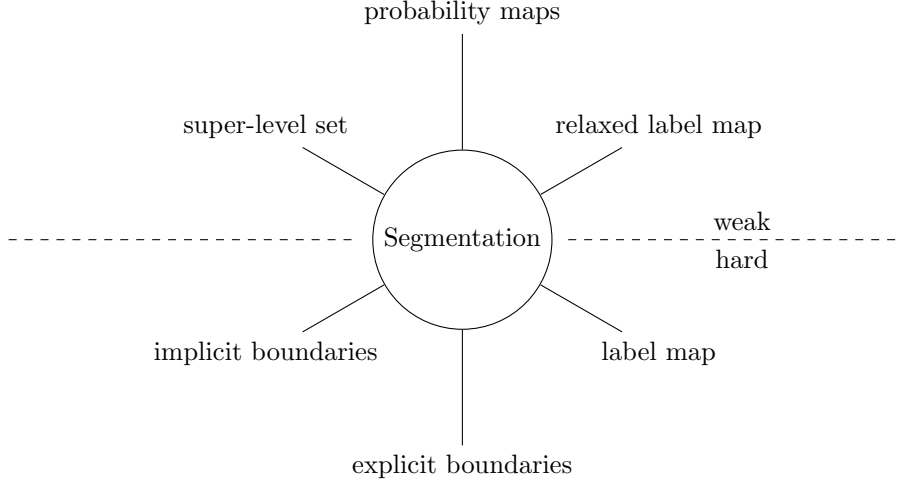


Figure 2.8: Different representations of segmentations.

Depending on the task and the imaging contrasts, a multi-label segmentation would be for example to assign 3 to all GM pixels, 2 to WM pixels, 1 to CSF pixels, and 0 to the rest.

In the following, we describe different ways to represent a segmentation, which we visualize in Figure 2.8. Representing a segmentation as a *label map*, as *explicit boundaries*, or as *implicit boundaries*, creates hard, discrete, and final boundaries that leave no room for uncertainties and further thresholding. Weak representations with *relaxed label maps*, *super-level sets*, and *probability maps*, in contrast, allow for uncertainties.

**Label Map** A label map is an elegant and straightforward way to represent a segmentation. It is simply a function  $u : \Omega \rightarrow \mathcal{L}$ . In this representation, the axioms of the mathematical function make sure, that every pixel is labeled, and that every pixel only has one label. Therefore, a segmentation represented by a label map is complete (all pixels labeled), and its regions are non-intersecting.

Notice that we can also choose  $\mathcal{L} = \mathcal{I}$ , which would make the image itself a segmentation. Given a medical MR image, however, taking the image itself as a segmentation would not be meaningful if the task were to segment anatomical structures.

**Explicit Boundary** Another way of representing a segmentation is to explicitly parametrize the boundaries  $\Gamma_\ell$  between the different segments as curves  $\psi_\ell : [0, 1] \rightarrow \Omega$  with  $\psi_\ell([0, 1]) = \Gamma_\ell$ , where  $\ell$  corresponds to the classes in  $\mathcal{L}$  [79]. Objects that are fully contained in the image domain have closed boundaries, whereas objects leaving

## 2 Background

the image domain may disjoin the parametrizations. Thus, for image segmentations using this representation, the objects are preferred to be fully contained in the domain, to not being concerned with joining and disjoining of curves during optimization. For the same reason, it is also preferred that one class region does not have multiple connected components.

In the case of binary segmentation only one curve  $\psi : [0, 1] \rightarrow \Omega$  is necessary to parametrize the object boundary  $\psi([0, 1]) = \partial\mathcal{O}$ . If the object only has one connected component that is fully contained in the interior of the image domain, the boundary is a closed, non-self-crossing, continuous curve, which is topologically equivalent to a circle.

**Implicit Boundary** A popular way of representing a segmentation is to implicitly define the boundaries  $\Gamma_\ell$  as equations [50]. By introducing signed distance functions  $\phi_\ell : \Omega \rightarrow \mathbb{R}$  for each class, the boundary of an object then is the kernel of  $\phi_\ell$ , or also called the 0-level set or just level set of  $\phi_\ell$ ,

$$\Gamma_\ell = \text{kernel}(\phi) = \phi_\ell^{-1}(\{0\}) = \{x \in \Omega \mid \phi_\ell(x) = 0\}. \quad (2.43)$$

A signed distance function  $\phi_\ell$

1. has positive values outside the object  $\mathcal{O}_\ell$ ,
2. has negative values inside the object  $\mathcal{O}_\ell$ ,
3. has, as already described, 0-values along the boundary  $\Gamma_\ell$ , and
4. since it shall be a distance, the magnitude of its derivative has to be 1 almost everywhere except on the boundary itself.

*Almost everywhere* means, everywhere except to a 0-measure subset with respect to the measure used in  $\Omega$ . The measure in  $\Omega$  would refer, for example, to the area measure in 2D or the volume measure in 3D.

These conditions for a signed distance function often cause problems in segmentation algorithms that use an implicit representation, because update steps for  $\phi_\ell$  usually do not preserve the necessary conditions. Therefore, so-called reinitialization steps have to be conducted regularly during optimization [50, Sec. II.7.2]. Without reinitializations, different slopes arise in the graph of  $\phi_\ell$ , and the update's force vector fields are unequal along the boundary, which can lead to divergence. Several kinds of resorts to this problem have been proposed [80], among them a clever way that completely replaces the gradient descent equation with a simpler update rule [81], see also (2.83).

Compared to the explicit boundary parametrizations, level sets can easily change the object's topology during optimization. The noninvariance in topology can be beneficial if the number of connected components of the object is a priori unclear. At the same time, it can also be a curse because an intentionally single connected component could split up into several segments during optimization.



A combinatorial way of using signed distance functions may save memory and computation time for problems with more than 3 labels [82]. For example, with only 2 signed distance functions  $\phi_1$  and  $\phi_2$ , already 4 label-regions can be reproduced using a mapping similar to  $(\ell_1, \ell_2, \ell_3, \ell_4) \leftrightarrow (-, +, -, +)$ , where  $+-$  symbolizes  $\phi_1(x)$  being positive and  $\phi_2(x)$  being negative. This way,  $L$  signed distance functions can already represent  $2^L$  classes.

**Relaxed Label Map** Suppose the set of class labels is finite and suppose we update a label map  $u$  with an update  $\Delta u$ , then  $u + \Delta u$  again should be a valid label map. This setup would imply that  $\Delta u$  can only make integer label changes, which presents a conflict with real-valued update steps, according to (2.28). As described in Section 2.2.4, convexity is a much desired property of an optimization algorithm and does not only demand the convexity of the functional, but also of the object's representation, as we just realized above: if  $u_1$  and  $u_2$  are two segmentations with valid representations, then for every  $\lambda \in [0, 1]$  also every linear combination  $\lambda u_1 + (1 - \lambda) u_2$  between  $u_1$  and  $u_2$  has to be a segmentation with a valid representation. In variational algorithms, usually the updates are real-valued, and therefore label maps need to be relaxed, i.e. convexified.

A label map  $u : \Omega \rightarrow \{0, 1, \dots, L-1\}$  could be relaxed, for example, to a relaxed label map  $u : \Omega \rightarrow [0, L]$ , where the set of  $L$  discrete labels  $\{0, \dots, L-1\}$  is convexified to the interval of all real numbers between 0 and  $L$ , including 0 and  $L-1$ . This completely makes sense for binary segmentations, where  $\{0, 1\}$  simply is extended to  $[0, 1]$ . For multilabel problems, however, the label map  $u : \Omega \rightarrow \{0, 1, \dots, L-1\}$  is often split into several relaxed one-vs-all binary label maps  $u_\ell : \Omega \rightarrow [0, 1]$  [83], which is called one-hot encoding. Furthermore, the same combinatorial trick as explained before for the signed distance functions can be utilized to reduce the  $L$  binary one-vs-all label maps to only  $\lceil \log_2 L \rceil$  binary label maps [83].

Moreover, many functionals fulfill a certain thresholding theorem, which guarantees that globally optimal relaxed segmentations can (again) be thresholded to globally optimal discrete segmentations, see Figure 2.10 on Page 54.

**Super-Level Set** Another relaxation technique is realized by functional lifting, using the set of all super-level sets, as is demonstrated by Pock et al. [84, 85], who lifted functionals similar to the piecewise-constant Mumford-Shah model (pcMS) and proposed a thresholding theorem for such representation. Nosrati et al. [86] used the same technique directly on the pcMS, and managed to incorporate even geometrical constraints to the segmentation in a globally optimal way. This technique, however, comes at a slightly higher computational cost, since individual super-level sets have to be calculated, and since it introduces another parameter that regulates how many super-level sets are used.

A  $\theta$ -super-level of  $u$  is the set of all points  $x \in \Omega$  that have a value  $u(x)$  higher than  $\theta$ , i.e.

$$[u > \theta] = \{x \in \Omega \mid u(x) > \theta\}, \quad (2.44)$$

## 2 Background

and can be represented using the characteristic set function

$$\phi(x, \theta) = \mathbb{1}_{[u > \theta]}(x) = \begin{cases} 1 & \text{if } u(x) > \theta, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \quad (2.45)$$

The "super-level set" then is the set of all super-levels

$$\{[u > \theta] \mid \theta \in [0, L]\}, \quad (2.46)$$

and can be represented as

$$(\phi(x, \theta))_{\theta \in [0, L]}. \quad (2.47)$$

The functional's integral then is not only integrated over the image domain  $\Omega$ , but also over the lifted domain part  $[0, L]$ . Furthermore, the super-level set representations  $(\phi(x, \theta))_{\theta \in [0, L]}$  need to be regularized, such that the super-levels  $[u > \theta]$  are growing with larger  $\theta$ , i.e.  $[u > \theta_1] \subset [u > \theta_2]$  for arbitrary  $\theta_1 > \theta_2$  [86].

**Probability Map** Finally, weak label maps can also be read with probabilistic interpretation as categorical distributions. For each class label  $\ell \in \mathcal{L}$ , probability maps  $p_\ell : \Omega \rightarrow [0, 1]$  with

$$p_\ell(x) = P[x \text{ belongs to class } \ell] \quad (2.48)$$

can be estimated. Such representations especially are used as an output in neural networks and are normally calculated with the softmax [87]

$$p_\ell(x) = \text{softmax}(z_\ell)_x = \sigma(z_\ell)_x = \frac{e^{z_{\ell,x}}}{\sum_{l \in \mathcal{L}} e^{z_{l,x}}}, \quad (2.49)$$

where the function  $\sigma : \mathbb{R}^{|\Omega|} \rightarrow \mathbb{R}^{|\Omega|}$  is a symbol for the softmax, and  $z \in \mathbb{R}^{|\Omega|}$  is the output state of the net before the softmax calculation. Here,  $|\Omega|$  is the number of data points in the finitely discretized domain  $\Omega$ .

Probability maps again can be thresholded in the end for binary segmentations. A straightforward way to generate a label map  $u : \Omega \rightarrow \mathcal{L}$ , given the  $L = |\mathcal{L}|$  probability maps  $(p_\ell)_{\ell=1, \dots, L}$ , is calculated through the argmax construction,

$$u(x) = \underset{\ell \in \mathcal{L}}{\operatorname{argmax}} p_\ell(x). \quad (2.50)$$

Probability maps have the same range as relaxed binary label maps, but are used in a slightly different context: while the probability maps' values can be interpreted as probabilities, it is not directly clear how the relaxed binary label maps' values between 0 and 1 correspond to uncertainty – or whether they do so in the first place – if used in combination with continuous cuts (2.86). To achieve binary segmentations with high accuracy, in both cases, however, optimal thresholds are problem-specific.

### 2.2.7 Segmentation Algorithms

How to get to a segmentation and what the optimal segmentation differentiates is formulated within the segmentation process.

A segmentation can be manually drawn by hand with a pen directly onto a printed image, for example, by explicitly delineating the boundaries or by filling out the regions. Manual segmentations are very complex and are defined within the "neural networks" of the experts' brains that include a lot of prior knowledge. Furthermore, manual segmentations depend on the experts' motor controls and their end effectors.

To teach this process to a computer, a set of operations have to be implemented, that transform a digital image  $I$  to a segmentation by using one of the representations described in Section 2.2.6. There are many image segmentation techniques in the literature [88, 89, 90, 91] and a complete review is beyond the scope of this thesis. They vaguely can be separated into two important classes: *shallow model-based algorithms* and the *supervised deep learning algorithms*. Other separation attempts differentiate between automatic and semi-automatic algorithms or try to differentiate based on the used energy function or the segmentation's representation.

**Automatic or Semi-Automatic** An automatic algorithm simply needs to be executed, whereas a semi-automatic algorithm needs more manual intervention such as selecting a seed point at a specific location in the data or manually choosing a field of view. Typical examples for semi-automatic segmentation algorithms are interactive graph cuts [92, 93], where one iteratively marks down a representative part of the foreground. In [92], for example, Gaussian mixture models of the supervised foreground and background are iteratively re-estimated, and graph cuts (2.84) are used to calculate the foreground-background segmentations.

**Energy-based Differentiation** This differentiation focuses on variational segmentation algorithms and divides energy terms into pixel-based, edge-based, region-based, and texture-based formulations [77]. These terms do not distinctively distinguish segmentations, because energy terms can be based on a combination of them. Pixel-based algorithms, such as pixelwise thresholding (2.67), use the information of each pixel separately without including neighbor information; edge-based algorithms mainly use first- or second-order derivative information of the image, such as edge-based active contours [79], cf. equations (2.76), (2.77), and (2.79); and region- and texture-based energies incorporate different intensity distributions of the segments, as is carried out with the Chan-Vese model [94], cf. Equation (2.82) with  $C \equiv 1$ . The Chan-Vese model (with arbitrary  $C$ ) as well as graph cuts [93], cf. left side of the equations in (2.95), already represent combinations of above classes.

**Representation-based Differentiation** The representation used for the segmentation, cf. Section 2.2.6, builds a classification from an implementational point of view. The chosen representation strongly influences the segmentation algorithm's advantages and drawbacks: in the ability to change the topology, in numerical robustness,

## 2 Background

computational parallelizability, memory usage, and implementational simplicity. To mention a few: active contours with explicit boundaries preserve their initialized topology [79], cf. Equation (2.76), at least with the naive implementation without enabling joining and disjoining of curve elements; active contours with explicit or implicit boundaries have critical instable stopping criteria [79, 94]; implicit boundaries that use signed distance functions need reinitializations and need to approximate absolute and sign functions as well as delta distributions [50, 80], cf. equations (2.77) and (2.79); large neighborhood connections with graph cuts may need more memory than continuous versions [86], cf. Equation (2.96) and the paragraph about metrization errors on Page 53; continuous max-flow min-cut algorithms only need a few lines of code with simple operations and are pleasingly parallelizable [23], cf. Equation (2.103); variational formulations with convex regularization through functional lifting introduce an additional discretization parameter [85, 84, 86]. Moreover, this differentiation also shows beautiful connections between segmentation algorithms and their variational formulations, as presented in Section 2.2.8 on Page 55 for variations of the piecewise-constant Mumford-Shah model.

**Shallow or Deep** The boundary between the two classes *shallow model-based algorithms* and *supervised deep learning algorithms* is blurry, because the model-based algorithms may include supervised learning, and the supervised learning algorithms themselves are also modeled. Therefore the differentiation lies in the "deepness" of the algorithms, which depends on how "deep" a process is being decomposed into its constituent subprocesses. The deepness reflects the algorithm's complexity and approximation power and could be measured, for example, in how many real-dimensional linearly independent, effective parameters are used [95]. In terms of neural networks, the deepness also depends on how many abstract layers between the input and the output layers are used [95], see also Section 2.2.10.

Without giving a specific threshold between *shallow* and *deep* models, among shallow segmentation algorithms we have the classic segmentation algorithms such as simple intensity thresholding (2.67), split and merge algorithms [96], atlas-based segmentation approaches [97, 98, 99], active contours [50], graph cuts [93], and continuous cuts [23, 100]. Further variational segmentations use formulations similar to energies used in active contours, graph cuts, and continuous cuts with first- and second-order clique potentials but additionally include statistical information such as appearance and shape models [101, 90, 102], see Section 2.2.9. Including statistical information through addition of an additional energy term makes a variational segmentation somewhat deeper; however, this addition probably does not reach the same connectedness between the data term and the statistical information as deep neural networks can learn during training, see Section 2.2.10. In contrast to shallow models, the many parameters of deep models are hard to tweak by hand, and therefore training in the form of a supervised learning framework is necessary, cf. Section 2.2.10.

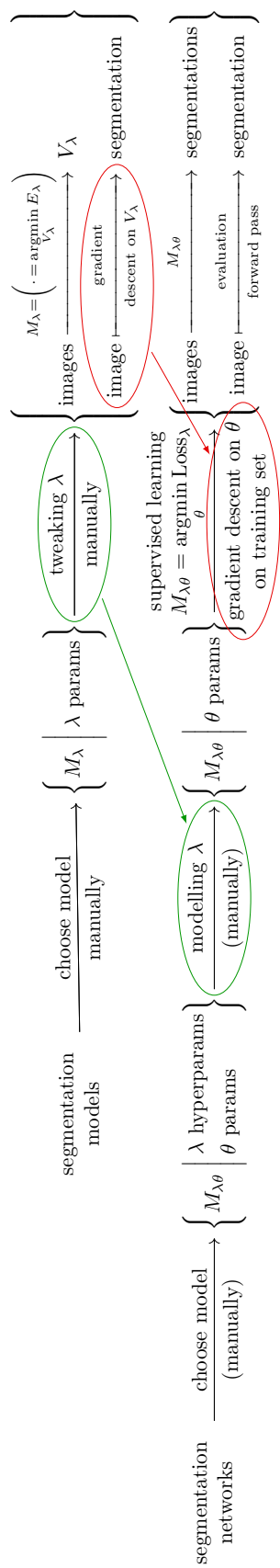


Figure 2.9: Variational segmentation approach (upper pipeline) vs. supervised deep learning (lower pipeline).

## 2 Background

In Figure 2.9 on Page 45, we show abstractions of the two pipelines: *manual modeling for a variational formulation* and *supervised learning for deep neural networks*. For manual modeling, we select a segmentation framework  $\{F_\lambda \mid \lambda \text{ parameters}\}$ , where  $F_\lambda$  represents a segmentation algorithm, and  $\lambda$  a set of parameters. Typically, the number of parameters increases during modeling, when adding new subprocesses, and is in the range of 10 to 1000. These parameters involve the modeling of the energy, the algorithm, and the stopping criteria for the descent method. Tweaking the parameters is conducted during the manual modeling part on selected images from a training/validation dataset before the actual segmentation of the target image. Though, some of the parameters are also automatically adapted during the segmentation process to adjust to the target image. This adaptation, however, could be seen as part of the segmentation algorithm itself. Depending on how much and how often the training/validation data are being used to optimize the parameters, we can say that pure model-based approaches are unsupervised. Finally, the daily business comprises testing the designed model on unseen test data to check if it generalizes well.

The big difference, computation-wise, between shallow model-based variational segmentation and supervised deep learning lies in where we use energy optimization, as is clearly visualized and marked with the red ellipses in Figure 2.9. To calculate a segmentation of a target image, the classic approach uses a descent method on an energy that depends on the target image, whereas a neural network simply applies a forward pass. The variational optimization phase for the neural network, in contrast, is applied during the training phase over a whole set of training data and moves the manual modeling part to the next level, as indicated with the green ellipses. Moving the gradient descent has the following advantage: whereas variational segmentations struggle with the stopping criteria for each target image anew, the critical task of applying the stopping criteria during supervised training has to be executed only once. To this end, we usually take a small validation dataset to check the quality of the network iterations and can use this validation to design stopping criteria.

The neural network’s parameter set  $\theta$  has a cardinality in the range of millions which typically is much larger than the parameter set used for classic approaches. However, as indicated with the green arrow, there exists a similar modeling experience for supervised deep learning, where one has to set up a loss function for training and several other parameters for the network’s architecture. The resulting parameters of this modeling phase are gathered with so-called hyperparameters  $\lambda$  and can be manually tweaked or themselves optimized by another network. To this end, ideas for recursively learning to learn have already been proposed [103]. For such ”deepness”, however, the need for computational power increases even more.

**Philosophical Discussion** In this paragraph, we accumulated some inexact philosophical thoughts, ideas, and hypotheses on the contention of variational segmentation approaches and deep learning neural networks.

It took several years for people to build confidence in applying deep neural networks for real-world problems [104]. We believe the prevailing notion was, that trained neu-

ral networks are difficult to understand, because of their cryptic way of approximating a problem, and therefore neural networks are not trustworthy. At that time, when the hardware was not ready for practical neural network approaches, referring to the AI winter [105], people preferred classic model-based approaches which at that time probably also seemed more principled. Regarding the confidence on using neural networks, however, we simply have to spend time dealing with an abstract object and we automatically build up a confident perception – quoting John von Neumann<sup>1</sup>: “Young man, in mathematics you don’t understand things. You just get used to them.”.

With variational segmentation approaches, the goal is to model functionals with their global minima being “globally optimal solutions”. At that point, we have to keep in mind that a segmentation being an optimal solution does not mean it is the correct optimal solution of the real problem. It merely is an optimal solution to the formulated problem and whether this problem accurately represents the real problem is another question. Here, with a real problem, we understand problems that involve processes of nature (and the human population) which are not necessarily already accurately formulated within abstract human languages. Such problems typically are very complex, and it is questionable whether they can be accurately simulated. Nevertheless, it lies in our nature to try to approximate and understand such problems with simpler models, and this is why model-based approaches seem more principled.

As a vantage point for deep learning, one could hypothesize that the simpler a formulation gets (e.g. being convex), the rougher and the more unsubtle it is in representing the complex real problem. A possible interpretation is that approximations of processes with neural networks are not necessarily mathematically *unprincipled* – since obviously, the network is a mathematical construction – however, the network’s trained weights may be as puzzling as the real problem itself. If one would interpret that “the inability of simple models exactly representing a complex model” is an invariant between *complex natural phenomena* and *abstractions from our minds*, then neural networks that accurately do their job but are challenging to understand potentially are closer to the real problem than models designed with simpler mathematical formulas. However, the possibility to discover a simple and computable model that describes a complex problem is still present. Finally, depending on the point of view, trained neural networks anyhow can be seen as relatively simple models.

### 2.2.8 The Mumford-Shah Functional

Some of the variational methods we introduced in Section 2.2.7 are described in more detailed in this section. Not all of these methods are part of this thesis’s publications; however, they were commonly used in medical image analysis during the last two decades and are therefore worth mentioning here. In this section, we show connections to the Mumford-Shah functional, which as an envelope closely connects many variational algorithms: where the algorithm’s implementations might be completely

---

<sup>1</sup>Reply, according to Dr. Felix Smith of Stanford Research Institute, to a physicist friend who had said “I’m afraid I don’t understand the method of characteristics”.

## 2 Background

different because of different used representations for their segmentations, their variational formulations indeed are similar to each other.

In particular, we motivate the choice of the continuous max-flow min-cut algorithm of Yuan et al. [23] in the following. In short: with this algorithm, the piecewise-constant Mumford-Shah functional is put to an optimization algorithm in a straightforward way; and the duality between the min-cut and the max-flow shows a familiar connection to the well-established graph cut methods.

**Definitions** The Mumford-Shah functional [106] was introduced in 1989 and is an energy formulation to simultaneously construct an optimal piecewise smooth approximation of an image together with an optimal discontinuity set of the image and its approximation. Given an image  $I : \Omega \subset \mathbb{R}^2 \rightarrow \mathcal{I} \subset \mathbb{R}$  on a bounded domain  $\Omega$ , we simultaneously optimize for a piecewise smooth approximation  $u : \Omega \rightarrow \mathcal{I}$  of  $I$  and a discontinuity set  $K$ :

$$\min_{u, K} \int_{\Omega} (u(x) - I(x))^2 dx + \int_{\Omega \setminus K} \|\nabla u(x)\|^2 dx + \int_K d\sigma, \quad (2.51)$$

where  $\int_K d\sigma$  is the length of  $K$  and  $d\sigma$  is its appropriate Hausdorff measure. In the following, we suppress the dependence of  $u$  and  $I$  on  $x$  to shorten the notation. Also, for the sake of notational simplicity, we left out the weighting parameters of the three additive terms. As explained in [106], the first term,  $\int_{\Omega} (u - I)^2 dx$ , is a coupling term and asks that  $u$  is an approximation of  $I$ ; the second term,  $\int_{\Omega \setminus K} \|\nabla u\|^2 dx$ , asks that  $u$  does not vary much on each connected component on  $\Omega \setminus K$ , and in combination with the first term, the second term ensures also that the intensities of  $I$  are similar for each connected component; and the third term,  $\int_K d\sigma$ , asks that the boundary of the discontinuity set  $K$  is as short as possible, yielding an optimal number of connected components with piecewise smooth boundaries. Note that in the second term, the classic derivatives  $\nabla u$  need not be calculated on the discontinuity set  $K$ , which makes sense because  $u$  has discontinuous changes across  $K$ .

This functional has been studied extensively, and a survey is given in [49, pp. 153–173]: different analytic approximations were suggested for which existence and uniqueness were analyzed; and also the behavior of the discontinuity set has been construed. In a special analytic setting where  $K$  is a finite union of smooth curves, it has been shown, for example, that triple-junctions have to meet in a  $60^\circ$  angle, and quadruple junctions would need to split into two triple-junctions. Such analytic observations are interesting, however for applications on discrete images they are not directly relevant, especially for images with only one object of interest that consists of only one connected component. Furthermore, we focus on a reduced special case of the Mumford-Shah functional where the discontinuity set is directly given by the approximator  $u$  – namely the piecewise-constant Mumford-Shah model (pcMS):

$$\min_{u, K \neq \emptyset} \int_{\Omega} (u - I)^2 dx + \int_K d\sigma, \quad (2.52)$$



where we additionally force that the function  $u$  is piecewise-constant, i.e.  $\nabla u = 0$  on  $\Omega \setminus K$ , and  $K \subset \Omega$  is a non-empty set  $K \neq \emptyset$ .

A fundamental observation is that  $u$ 's piecewise-constant parts take on the mean values of the connected components. That is: if  $R \subset \Omega$  is a connected component, surrounded by  $K$ , then

$$u(x) \equiv \text{mean}((I(y))_{y \in R}) \quad \forall x \in R. \quad (2.53)$$

Therefore,  $u$  can be encoded by  $K$ . Conversely,  $K$  can also be encoded by  $u$ , by simply extracting the discontinuities of the piecewise constant  $u$ , and thus the pcMS (2.52) becomes an optimization problem solely on  $u$  or on  $K$ .

Already in the full Mumford-Shah model (2.51),  $u$  and  $K$  are somewhat coupled, however, there is more freedom than in the piecewise-constant version. In the initial version, a discontinuity line may be added to  $K$  or may be mollified by  $u$  because of allowed smooth transitions in  $u$ . Imposing, however, that  $u$  needs to be piecewise-constant rather than just piecewise smooth, simplifies optimization, as mentioned above, through the mutual coupling between  $K$  and  $u$ . In practice, however, the mentioned ambiguity between  $u$  and  $K$  in the piecewise smooth version is not taken away entirely in the piecewise-constant version, as  $u$  could be optimized with a jump across a smooth transition in the image, or the smooth transition could be left as one region, resulting in the adjustment of the respective constant of  $u$  to the mean value of the combined connected component. Moreover, in the piecewise-constant version, the different values of  $u$  can be directly used to represent different labels.

**Normalization of the Segmentation Model** The approximation  $u$  of the pcMS can further be normalized such that  $u$ 's piecewise-constant mean values realize integer values  $\{0, \dots, L-1\}$ . For the sake of visual simplicity, we demonstrate this transformation in the binary case  $L = 2$ . As explained in (2.53), a solution  $u^*$  of the pcMS (2.52) realizes the two constants  $m_{\mathcal{O}} = \text{mean}((I(x))_{x \in \mathcal{O}})$  and  $m_{\mathcal{B}} = \text{mean}((I(x))_{x \in \mathcal{B}})$ , where  $\mathcal{O}$  is the object surrounded by  $K$ , and  $\mathcal{B} = \Omega \setminus \mathcal{O}$  denotes the background. To use the mapping

$$\begin{aligned} m_{\mathcal{B}} &\mapsto 0, \\ m_{\mathcal{O}} &\mapsto 1, \end{aligned} \quad (2.54)$$

we reformulate (2.52) to

$$\min_{u: \Omega \rightarrow \{0,1\}} \int_{\Omega} u(x) \underbrace{(I(x) - m_{\mathcal{O}})^2}_{=C_{\mathcal{O}}} + (1 - u(x)) \underbrace{(I(x) - m_{\mathcal{B}})^2}_{=C_{\mathcal{B}}} dx + \int_{\partial \mathcal{O}} d\sigma, \quad (2.55)$$

where the mean values,  $m_{\mathcal{O}}$  and  $m_{\mathcal{B}}$ , and the object  $\mathcal{O}$  itself depend on the segmentation  $u$  with  $\mathcal{O} = [u = 1] = \{x \in \Omega \mid u(x) = 1\}$ . Furthermore, instead of the mean intensity differences  $(I(x) - \text{mean}(I(\text{region})))^2$ , one can generalize the pcMS by choosing any kind of model functions  $C_{\mathcal{O}}$  and  $C_{\mathcal{B}}$ .

Another way to denote the segmentation model (2.55) is to introduce the mapping

$$I_{\text{model}} = m_{\mathcal{O}} u + m_{\mathcal{B}}(1 - u), \quad (2.56)$$

## 2 Background

as done in [22], with which the pcMS (2.52) intuitively reads as

$$\min_{u: \Omega \rightarrow \{0,1\}} \int_{\Omega} (I(x) - I_{\text{model}}(x))^2 dx + \int_{\partial \mathcal{O}} d\sigma. \quad (2.57)$$

**Connection to the ROF model** The model of Rudin, Osher, and Fatemi (ROF), which was introduced in 1992 [107], describes an image denoising model based on total variation (TV) minimization, which is similar to the pcMS (2.52):

$$\min_{u: \Omega \rightarrow \mathcal{I}} \int_{\Omega} (u - I)^2 dx + \int_{\Omega} \|\nabla u\| dx. \quad (2.58)$$

For the sake of visual simplicity, we left out the weighting parameters between the two integral terms. Comparing this model to (2.52), only the first integral term  $\int_{\Omega} (u - I)^2 dx$  is the same, whereas the second term,  $\int_{\Omega} \|\nabla u\| dx$ , called TV for smooth functions (e.g. in  $W^{1,1}$ -Sobolev spaces), does not appear in (2.52) nor in (2.51). More importantly,  $u$  is simply a denoised approximation of  $I$  and is not directly a segmentation, as the ROF model does not force  $u$  to be piecewise-constant.

Using the ROF model but imposing  $u$  to be piecewise-constant, and assuming that w.l.o.g.  $u$  realizes the mean values 0 and 1, we actually reproduce the pcMS (2.52) for the binary case. To show the similarity, the coarea formula for bounded variation functions [63, Theorem 5.9],

$$\int_{\Omega} \|\nabla u\| dx = \int_0^1 \int_{\partial[u \geq \theta]} d\sigma d\theta, \quad (2.59)$$

shows that the boundary integral term  $\int_{\partial \mathcal{O}} d\sigma = \int_K d\sigma$  and the TV term  $\int_{\Omega} \|\nabla u\| dx$  actually are the same if  $u$  is regular enough [108], e.g.  $u \in W^{1,1}$ <sup>2</sup>. This is because the discontinuity set  $K$  coincides with the jump-set  $\partial[u \geq \theta]$  for almost all  $\theta \in [0, 1]$  and thus the perimeter of the foreground object  $[u = 1]$  fulfills (using the mean value theorem for integrals)

$$\int_K d\sigma = \int_{\partial[u \geq 0.5]} d\sigma = (1 - 0) \int_{\partial[u \geq 0.5]} d\sigma = \int_0^1 \int_{\partial[u \geq \theta]} d\sigma, \quad (2.60)$$

and therefore we have

$$\int_K d\sigma = \int_{\Omega} \|\nabla u\| dx. \quad (2.61)$$

Indeed, the TV regularization which is used for denoising with the ROF model has a similar effect as denoising with the Mumford-Shah model.

<sup>2</sup>  $W^{1,1}$  is used to express the total variation  $\int_{\Omega} |Du|$  [49, p. 45] by means of the weak derivative  $\nabla u$ ;  $|Du|$  is the distributional derivative measure [63, Sec. 5.1].

## 2.2 Technical Background

**Bayesian Inference** In Figure 2.7, we mentioned that energy functionals could be generated through Bayesian inference, and indeed, the pcMS can be derived as a log-linear model. We demonstrate its construction in the discrete and in the continuous setting, following [109].

Let  $I = (I_1, \dots, I_m)$  be the pixel intensities of the given image  $I$  and  $u = (u_1, \dots, u_m)$  be the desired piecewise-constant image  $u$ , and let us furthermore assume the following two properties:

1. all measurements  $I_1, \dots, I_m$  are mutually independent (m.i.); and
2. we can use the first-order Markov approximations (MA) for the segmentation prior  $P(u)$ :

$$P(u) = P(u_1, \dots, u_m) = P(u_m) \prod_{i=1}^{m-1} P(u_i|u_{i+1}) \propto \prod_{i=1}^{m-1} P(u_i|u_{i+1}).$$

Now, modeling the likelihood  $P(I|u)$  and the prior  $P(u)$  with the following functions

$$\begin{aligned} P(I|u) &\stackrel{\text{m.i.}}{=} \prod_i P(I_i|u_i) \propto \prod_i \exp\left(-\frac{(u_i - I_i)^2}{2\sigma^2}\right) = \exp\left(-\sum_i \frac{(u_i - I_i)^2}{2\sigma^2}\right), \\ P(u) &\stackrel{\text{MA}}{=} P(u_m) \prod_{i=1}^{m-1} P(u_i|u_{i+1}) \propto \prod_{i=1}^{m-1} P(u_i|u_{i+1}) \\ &\propto \prod_{i=1}^{m-1} \exp(-\lambda |u_i - u_{i+1}|) \end{aligned} \quad (2.62)$$

results in the negative log maximum likelihood estimation (MLE) being the discrete Mumford-Shah energy:

$$E = -\log(P(I|u)P(u)) = \sum_{i=1}^m \frac{(u_i - I_i)^2}{2\sigma^2} + \underbrace{\log(\sqrt{2\pi}\sigma)}_{\text{irrelevant for min}} + \lambda \sum_{i=1}^{m-1} |u_i - u_{i+1}|. \quad (2.63)$$

The energy  $E$  coincides with the Ising model [110] if we rescale  $E$ , by multiplying  $2\sigma^2$  into  $\lambda$ .

With similar steps but utilizing a product integral, one can directly generate the pcMS in the continuous case [77]. For notational simplicity, we again choose the binary-labeling case; and to formulate a slightly more general framework, we use an arbitrary regional intensity likelihood  $P(I|u)$ . To this end, let the regional likelihood and segmentation prior be modeled with

$$\begin{aligned} P(I|u) &= \prod_{x \in \Omega} P(I(x)|u(x)) = \left( \prod_{x \in \mathcal{O}} P_{\mathcal{O}}(I(x)|u(x)) \right) \left( \prod_{x \in \mathcal{B}} P_{\mathcal{B}}(I(x)|u(x)) \right), \\ P(u) &= \exp\left(-\lambda \int_{\Omega} \|\nabla u\| \, dx\right). \end{aligned} \quad (2.64)$$

## 2 Background

Furthermore, let the product integral for a suitable function  $f$  be defined as

$$\prod_{\Omega} f(x)^{dx} := \exp \left( \int_{\Omega} \log f(x) dx \right), \quad (2.65)$$

then we can find the negative log MLE to be

$$\begin{aligned} E(u) &= -\log(P(I|u)P(u)) = -\log \left( \prod_{\Omega} P(I(x)|u(x))^{dx} \right) - \log(P(u)) \\ &= \int_{\mathcal{O}} \log P_{\mathcal{O}}(I(x)|u(x)) dx + \int_{\mathcal{B}} \log P_{\mathcal{B}}(I(x)|u(x)) dx + \int_{\Omega} \lambda \|\nabla u\| dx. \end{aligned} \quad (2.66)$$

If the model functions  $P_{\mathcal{O}}$  and  $P_{\mathcal{B}}$  are unknown, they can be estimated using some kind of density estimation technique on local or more global regions, e.g. histograms on selected regions, Gaussian mixture models, or kernel density estimation techniques [93, 80].

**History** As explained in [106, p. 581], the pcMS in a discretized two-label case is closely related to the Ising model [110]. Therefore, in a sense, Ising proposed a segmentation model as early as 1925 by modeling the natural clustering of magnetization spins under the influence of an external magnetic field, arranged on a discrete lattice. There, the spins align with the magnetic field, where neighboring spins regularize each others' direction.

In 1943, Ashkin and Teller reformulated the Ising model for four configurations, and later, in 1951, Potts proposed in his dissertation to generalize the model to arbitrary many configurations. A complete review of the physics-related development of the Potts model is given in [111]. Comparing our field with statistical physics, digital images take on the place of the external magnetic fields in the Ising model. Furthermore, from a more theoretical, statistical point of view, Markov Random Fields can be used to represent discrete segmentation models like the Ising or the Potts model [88, 93, 102].

Important to mention is the fundamental difference in complexity between the Ising model, which differentiates between two classes, and the Potts model, which differentiates between more than two classes. The binary-label problem can be efficiently solved with, for example, the polynomial-time algorithm of Ford and Fulkerson [112], which uses the max-flow min-cut Duality. The multi-label problem, in contrast, is known to be NP-hard, and therefore only near-optimal solutions can be found in polynomial time, e.g. [85].

Finally, Markov Random Fields with interactions between immediate neighbors, like in the Ising model, can be represented in a continuous setting with the pcMS (2.52). This analogy between discrete and continuous models introduces many more solvers. While discrete models are solved with graph-based discrete optimization algorithms such as the Ford-Fulkerson algorithm, continuous models can also be solved via partial differential equations and their numerical discretizations, following Figure 2.7.

**Metrication Errors** As mentioned in [23], one major drawback of discrete methods is the grid bias which leads to discretization errors, or sometimes called metrication errors: curved boundaries that should be smooth appear with staircase artifacts. While such staircase artifacts naturally appear in visualizations of binary segmentations, the discrete interaction regularization in the Ising model reinforces such artifacts during segmentation if the interaction potentials between neighboring pixels penalize some spatial directions more than others.

This discretization error can be reduced by including larger discrete neighborhoods and thus more complex interactions between non-direct neighbors, as shown for example in [86, Fig. 5]. However, solving such models comes with a higher computational cost and memory usage.

Another way to reduce the visual staircases is to relax the binary segmentation by including gray values through relaxed label maps, as is typically done for continuous models, see Page 55. In the end, when we need a binary segmentation, we have to threshold the relaxed label map again, and this way the natural staircase artifacts reappear, but without the artifacts caused through the discrete interaction regularization, explained above. An example is shown in [86, Fig. 5], where the thresholded relaxed segmentations have a similar visual appearance as binary segmentations that are calculated with large interaction neighborhoods.

**Relaxation and Thresholding Theorems** As mentioned in the previous paragraph, relaxing the segmentation's binary values to gray values allows dealing with metrication errors. However, if we want to reproduce a binarized segmentation, we have to threshold the relaxed segmentation again.

Reducing metrication errors is not the only motivation for relaxation. More importantly, often the whole energy formulation needs to be relaxed to create a convex problem, which guarantees that globally optimal solutions are found. This reformulation involves the relaxation of the energy functional  $E$  to a convex energy  $\tilde{E}$ , the relaxation of the domain of admissible solutions  $V$  to  $\tilde{V}$ , and therefore also the relaxation of the range of the admissible solutions  $u$ . The latter also guarantees that an admissible solution stays an admissible solution after a descent step, as explained on Page 41. Relaxation to a convex problem is sometimes also called convexification.

In the following, let us denote the threshold of a function  $u : \Omega \rightarrow [0, 1]$  at level  $\theta \in [0, 1]$  with

$$\mathbb{1}_{u \geq \theta}(x) = \begin{cases} 1 & \text{if } u(x) \geq \theta, \text{ and} \\ 0 & \text{otherwise.} \end{cases} \quad (2.67)$$

Note, the thresholded function  $\mathbb{1}_{u \geq \theta} : \Omega \rightarrow \{0, 1\}$  is basically the same as the  $\theta$ -super-level set function in (2.45).

Ultimately, the question arises, whether an optimal solution of the convexified problem is an optimal solution of the initial problem after thresholding at level  $\theta$  – as depicted in Figure 2.10. Surprisingly, there exist such thresholding theorems which guarantee that almost all threshold levels lead to an optimal solution, at least for the

## 2 Background

$$\boxed{\min_{u \in V} E(u) \xrightarrow[\text{and the range of } u]{\text{convexify } V \text{ and } E} u^* = \operatorname{argmin}_{u \in \tilde{V}} \tilde{E}(u) \xrightarrow[u^* \rightarrow \mathbb{1}_{u^* \geq \theta}]{\text{thresholding}} \tilde{E}(\mathbb{1}_{u^* \geq \theta}) \stackrel{= ?}{\geq} \tilde{E}(u^*)}$$

Figure 2.10: Relaxing a problem and thresholding an optimal relaxed solution.

convexified functional  $\tilde{E}$ . This way, continuous cuts can be seen as complicated, filtered thresholding, where the TV regularization avoids the segmentation of noise. As an advantage, TV regularization allows sharp discontinuities in the solutions, leading to near-binary results [108].

To show such theorems for a given energy  $\tilde{E}$ , one typically uses the layer-cake formula

$$u(x) = \int_0^1 \mathbb{1}_{u \geq \theta}(x) d\theta \quad (2.68)$$

and a coarea formula (such as (2.59)) to devise a property as follows: (cf. (2.73))

$$\tilde{E}(u) = \int_0^1 \tilde{E}(\mathbb{1}_{u \geq \theta}(x)) d\theta \quad \forall u \text{ in a neighborhood of } u^*. \quad (2.69)$$

Using this property (2.69), the optimality of  $u^*$ , and the continuity of  $\int_0^1 \tilde{E}(\mathbb{1}_{u^* \geq \theta}(x)) d\theta$  in  $\theta$  one can show via contradiction that

$$\tilde{E}(u^*) = \tilde{E}(\mathbb{1}_{u^* \geq \theta}) \quad \forall \theta \in (0, 1). \quad (2.70)$$

We now demonstrate convexification of the pcMS for the use of binary segmentation. With equations (2.59) and (2.60), the reduced Mumford-Shah version for binary segmentation (2.55) can be rewritten to

$$\min_{u: \Omega \rightarrow \{0,1\}} \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + \|\nabla u\| dx. \quad (2.71)$$

Let us assume fixed  $C_{\mathcal{O}}$  and  $C_{\mathcal{B}}$  during optimization that are independent of any iterations of  $u$ . This assumption is generally not the case during optimization; however, we assume that we can model  $C_{\mathcal{O}}$  and  $C_{\mathcal{B}}$  such that they do not change much at some point during convergence, referring to  $\Gamma$ -convergence [49, p. 40]. Hence, in a simplified view, at some point during convergence, the functional in (2.71) is convex because the point-wise additions and multiplications, as well as the  $\nabla$  operator, are linear and the Euclidean norm  $\|\cdot\|$  is convex. However, incorporating the optimization of  $C_{\mathcal{O}}$  and  $C_{\mathcal{B}}$  to the problem, as in the pcMS, generally leads to non-convex problems.

The space of admissible solutions  $V = \{u : \Omega \rightarrow \{0, 1\}\}$  is not convex, since two admissible solutions in  $V$  cannot be convexly combined, because for some  $\lambda \in ]0, 1[$ , the combination  $\lambda u + (1 - \lambda) v$  between  $u$  and  $v$  is not in  $V$  anymore. This combination

## 2.2 Technical Background

generally is not in  $V$ , because for example for some  $x$ , where  $u(x) = 1$  but  $v(x) = 0$ , the value of the combined function at that position is  $\lambda u(x) + (1 - \lambda) v(x) = \lambda$ , which is not in  $\{0, 1\}$ . Therefore  $V$  needs to be relaxed to  $\tilde{V} = \{u : \Omega \rightarrow [0, 1]\}$ , and we end up with the relaxed segmentation model

$$\min_{u: \Omega \rightarrow [0, 1]} \underbrace{\int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + \|\nabla u\| \, dx}_{=E(u)}. \quad (2.72)$$

Finally, we can indeed derive the sufficient condition (2.69) as follows:

$$\begin{aligned} E(u) &\stackrel{(2.68)}{=} E\left(\int_0^1 \mathbb{1}_{u \geq \theta} \, d\theta\right) \\ &= \int_{\Omega} \left(\int_0^1 \mathbb{1}_{u \geq \theta} \, d\theta\right) C_{\mathcal{O}} + \left(1 - \int_0^1 \mathbb{1}_{u \geq \theta} \, d\theta\right) C_{\mathcal{B}} + \overbrace{\int_0^1 \int_{\partial[u \geq \theta]} d\sigma \, d\theta}^{(2.59)} \\ &= \underbrace{\int_0^1 \int_{\Omega} \mathbb{1}_{u \geq \theta} C_{\mathcal{O}} + (1 - \mathbb{1}_{u \geq \theta}) C_{\mathcal{B}} \, dx \, d\theta}_{\text{Fubini}} + \int_0^1 \underbrace{\int_{\Omega} \nabla \mathbb{1}_{u \geq \theta} \, dx}_{(2.59) \& (2.60) \text{ on } [u \geq \theta]} \, d\theta \quad (2.73) \\ &= \int_0^1 E(\mathbb{1}_{u \geq \theta}) \, d\theta. \end{aligned}$$

Moreover, the thresholding theorem also holds if the TV-regularization  $\int \|\nabla u(x)\| \, dx$  is modified with anisotropic total variation (ATV) to  $\int \sqrt{u(x)^T A(x) u(x)} \, dx$ . To this end, Olsson et al. [113] proved a coarea formula that works for ATV if the anisotropic matrix field  $A : \Omega \rightarrow \mathbb{R}^{\dim \Omega \times \dim \Omega}$  is strongly positive definite.

**Representations** In this paragraph, we demonstrate how to generate different algorithms based on the pcMS-inspired binary segmentation model

$$\min_u \int_{\mathcal{O}} C_{\mathcal{O}} \, dx + \int_{\Omega \setminus \mathcal{O}} C_{\mathcal{B}} \, dx + \text{Length}(\partial \mathcal{O}, C) \quad (2.74)$$

by using different representations. In this model (2.74),  $\text{Length}(\partial \mathcal{O}, C)$  calculates the weighted length of the object boundary, given a weighting function  $C$ . The term  $\text{Length}(\partial \mathcal{O}, C)$  is present, for example, as the external edge force (2.75) in active contours [79, 114, 94] or as the TV regularization in graph cuts and continuous cuts [23]. In the following, we merely focus on the form of the functionals and their variational derivatives and do not go into detail for implementations.

We start with the geodesic active contour model [114] or [49, p. 176],

$$J_2(c) = \text{Length}(\partial \mathcal{O}, C) = \int_K C \, d\sigma = \int_0^1 C(c(s)) c'(s) \, ds, \quad (2.75)$$

## 2 Background

where  $c : [0, 1] \rightarrow \Omega$  is an explicit parametrization of the object boundary  $\partial\mathcal{O} = K$  with  $K = c([0, 1]) = \text{Image}(c)$ , and  $C : \Omega \rightarrow \mathbb{R}$  weighs the contour length, such that the optimal curve  $c$  follows along the desired delineation in the image. A typical example for the model function  $C$  is to invert the image gradient magnitudes, e.g.  $C(x) = 1/(1 + \|\nabla I(x)\|)$  [50, Sec. III.12.1]. Indeed, this model represents the length term in (2.74) and is part of the Mumford-Shah functional (2.51) if  $C \equiv 1$  (which then would, however, not make sense without regional information). The following curve evolution converges to a local minimum of  $J_2$  (2.75) and is called *snakes*:

$$\partial_t c = (\kappa C - \langle \nabla C, N \rangle) N, \quad (2.76)$$

where  $\kappa$  is the curvature,  $N(s)$  is the normal vector to  $c$  at position  $c(s)$ , and  $t$  is the numerical evolution time parameter. Moreover,  $\kappa C N$  represents the so-called mean curvature motion, and  $\langle \nabla C, N \rangle N$  represents the edge force.

Instead of an explicit parametrization, Caselles et al. [115] use a signed distance function  $\phi : \Omega \rightarrow \mathbb{R}$  to reformulate the curve evolution of *snakes* (2.76):

$$\partial_t \phi = \left( \kappa C + \left\langle \nabla C, \frac{\nabla \phi}{\|\nabla \phi\|} \right\rangle \right) \|\nabla \phi\| = \kappa C \|\nabla \phi\| + \langle \nabla C, \nabla \phi \rangle. \quad (2.77)$$

With this representation, the curvature and the unit inward normal are given by  $\kappa = \text{div}(\nabla \phi / \|\nabla \phi\|)$  and  $N = -\nabla \phi / \|\nabla \phi\|$ , respectively.

The curve evolutions (2.76) and (2.77), so far, have been derived through geometric considerations. If, however, we calculate the gradient descent equation (2.27) with the variational derivative of (2.75) formulated for the level set  $H(\phi)$ ,

$$\tilde{J}_2(\phi) = \int_{\Omega} C \|\nabla(H(\phi))\| \, dx, \quad (2.78)$$

where  $H$  is the Heaviside function, we end up with the following equation,

$$\partial_t \phi = \left( \kappa C + \left\langle \nabla C, \frac{\nabla \phi}{\|\nabla \phi\|} \right\rangle \right) \delta(\phi), \quad (2.79)$$

where we recognize the mean curvature motion and the edge force but instead of  $\nabla \phi$  the Dirac delta function  $\delta$  is used. These two forces are split up through the application of the product and chain rule on the variational derivative:

$$\begin{aligned} -\frac{\delta \tilde{J}_2}{\delta \phi} &= \text{div} \partial_{\nabla \phi} (C \|\nabla(H(\phi))\|) \stackrel{\substack{\text{chain rule} \\ \text{on TV}}}{=} \text{div} \partial_{\nabla \phi} (C \delta(\phi) \|\nabla \phi\|) \quad (2.80) \\ &\stackrel{\substack{\text{chain} \\ \text{rule}}}{=} \text{div} \left( C \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|} \right) \\ &\stackrel{\substack{\text{product} \\ \text{rule}}}{=} \left\langle \nabla C, \frac{\nabla \phi}{\|\nabla \phi\|} \right\rangle \delta(\phi) + C \underbrace{\text{div} \left( \delta(\phi) \frac{\nabla \phi}{\|\nabla \phi\|} \right)}_{= \nabla \delta(\phi) \kappa + \delta(\phi) \kappa}, \end{aligned}$$



using  $\nabla\delta(\phi) = 0$ . If  $C \equiv 1$ , then the edge force vanishes. This approach, with  $C \equiv 1$ , is utilized in the Chan-Vese model [94] as an "active contour without edges", where the Chan-Vese model is the pcMS represented with level sets  $H(\phi)$ :

$$\min_{\phi, m_{\mathcal{O}}, m_{\mathcal{B}}} \int_{\Omega} H(\phi) (I - m_{\mathcal{O}})^2 + (1 - H(\phi)) (I - m_{\mathcal{B}})^2 + C \|\nabla(H(\phi))\| \, dx, \quad (2.81)$$

where the according gradient descent equation is given by

$$\partial_t \phi = \left( \kappa C + \left\langle \nabla C, \frac{\nabla \phi}{\|\nabla \phi\|} \right\rangle - (I - m_{\mathcal{O}})^2 + (I - m_{\mathcal{B}})^2 \right) \delta(\phi). \quad (2.82)$$

Actually, Chan and Vese [94] added a balloon force to (2.81) and (2.82) to accelerate convergence. This force, however, can easily mess up convergence if set too high compared to the regional terms and the edge force. In such cases, the active contours vanish completely through divergence towards the empty set or the whole image domain.

The two different curve evolutions (2.77) and (2.79) ((2.82), respectively) differ in  $\nabla\phi$  and  $\delta(\phi)$ , yet, depending on the numerical discretizations of  $\delta$ ,  $\delta(\phi)$  can be smoothed to a similar shape as  $\|\nabla\phi\|$ , cf. Page 18, [50, p. 16] and [94]. Leaving out the length regularization term  $\int_{\Omega} C \|\nabla(H(\phi))\| \, dx$  in (2.81) completely, Chan and Zhu [81] developed a very simple update scheme by replacing the gradient descent equation as follows:

$$\partial_t \phi = A \delta(\phi) \xrightarrow{\text{replacing}} \phi(x, t) = \begin{cases} 1 & \text{if } A(x, t) \geq 0, \\ -1 & \text{if } A(x, t) < 0, \end{cases} \quad (2.83)$$

where  $A$  represents the respective forces. This update rule works, because the energy without length regularization only depends on  $H(\phi)$  and not directly on  $\phi$ . The length term, however, is crucial for image segmentation, for which reason Chan and Zhu [81] chose this update rule in combination with (length) regularization through coupling with a shape model.

**Representations – Max-Flow Min-Cut Duality** Indeed, the curvature calculation of the signed distance functions and the need for their reinitializations contain numerical complications [50, Sec. II.7.2], cf. Page 18. Therefore, from our point of view, the use of label maps  $u : \Omega \rightarrow \{0, 1\}$  and their relaxed versions  $u : \Omega \rightarrow [0, 1]$  lead to the more stable segmentation algorithms: graph cuts and continuous cuts.

For graph cuts [93], the theory of flow-networks and the duality between min-cuts and max-flows from graph theory is used. In short, the pixels are represented with vertices  $V$  of a graph, and the edges  $E \subset V \times V$  between neighboring vertices are equipped with non-terminal capacities  $C : V \times V \rightarrow \mathbb{R}$ . Two additional virtual source and sink/target vertices  $s$  and  $t$  are also connected with all pixel vertices, yielding the terminal capacities  $C_s : \{s\} \times V \rightarrow \mathbb{R}$  and  $C_t : \{t\} \times V \rightarrow \mathbb{R}$ . These terminal and

## 2 Background

non-terminal capacities are the framework's model functions. The goal is then to find a cut  $K \subset \hat{E} = E \cup (\{s\} \times V) \cup (\{t\} \times V)$  such that the sum over the capacities of the cut edges (2.84) is minimal. In Figure 2.11, we provide a minimal simplification of a graph cut. In (2.84) and (2.85), we directly compare the discrete graph cut with

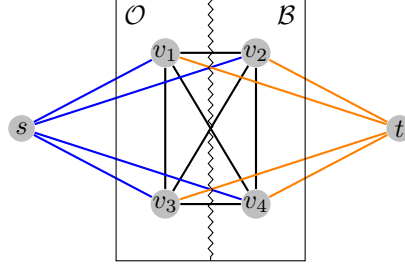


Figure 2.11: Simplified illustration of a graph cut: non-terminal capacities (black), source capacities (blue), sink capacities (orange), and the cut (zigzag line).

the continuous cut for understanding their common ground:

$$\min_{K \subset \hat{E}} \sum_{e \in K} C(e) = \min_{\mathcal{O} \subset V} \sum_{v \in \mathcal{B}} C_s(v) + \sum_{u \in \mathcal{O}} C_t(u) + \sum_{(u,v) \in \mathcal{O} \times \mathcal{B}} C(u,v), \quad (2.84)$$

$$\min_{\mathcal{O} \subset \Omega} \int_{\mathcal{B}} C_{\mathcal{O}}(x) dx + \int_{\mathcal{O}} C_{\mathcal{B}}(x) dx + \int_{\partial \mathcal{O}} C(x) d\sigma(x), \quad (2.85)$$

whereas  $C_s$ ,  $C_t$  and  $C$  are functions on edges, in the continuous case,  $C_{\mathcal{O}}$ ,  $C_{\mathcal{B}}$ , and  $C$  live on the domain  $\Omega$ .

For both graph cuts and continuous cuts, dually formulated representations exist, e.g. max-flow and continuous max-flow, respectively, which do not directly minimize the minimal cut energy (2.84) or (2.85), respectively. As already mentioned in the paragraph of the pcMS's history on Page 52, the discrete duality is established through the max-flow min-cut theorem [116], and similarly, a continuous max-flow min-cut duality can be shown using  $L^2$ -functions and the Cauchy-Schwartz inequality as follows: Starting again with the generic relaxed min-cut segmentation model (2.72) endowed with the model functions  $C$

$$\min_{u: \Omega \rightarrow [0,1]} \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + C \|\nabla u\| dx, \quad (2.86)$$

we use the Cauchy-Schwartz inequality, also known as the cosine formula, between two vectors  $p$  and  $q$  in  $\mathbb{R}^{\dim \Omega}$ , to formulate a non-terminal saturation property similar to the discrete max-flow:

$$\max_{\|p\| \leq C} p \cdot q = \max_{\|p\| \leq C} \|p\| \|q\| \cos(\angle(p, q)) = C \|q\|.$$

## 2.2 Technical Background

With this saturation property for  $q = \nabla u(x)$ , we can rewrite (2.86) to

$$\min_u \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + \left( \max_{\|p(x)\| \leq C(x)} p(x) \cdot \nabla u(x) \right) dx, \quad (2.87)$$

and moving the maximum out of the integral by interpreting  $\|p\| \leq C$  as a point-wise comparator and the involved functions as  $L^2$ -functions, we arrive at the saddle point problem

$$\max_{\|p\| \leq C} \min_u \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + p \cdot \nabla u dx. \quad (2.88)$$

With integration by parts we can pass the derivative over to  $p$ , which results in a model that can produce algorithms without the need of differentiating the desired segmentation,

$$\max_{\|p\| \leq C} \min_u \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} - u \operatorname{div} p dx. \quad (2.89)$$

The negative sign of the linear term  $-u \operatorname{div} p$  can be stashed by redefining  $p \mapsto -p$ , and thus with the linearity of the divergence we have

$$\max_{\|p\| \leq C} \min_u \int_{\Omega} C_{\mathcal{B}} + u (C_{\mathcal{O}} - C_{\mathcal{B}} + \operatorname{div} p) dx. \quad (2.90)$$

We now introduce the flow variables  $p_{\mathcal{O}}$  and  $p_{\mathcal{B}}$  and formulate the terminal saturation properties,

$$\max_{p_{\mathcal{O}} \leq C_{\mathcal{O}}} p_{\mathcal{O}} = C_{\mathcal{O}}, \quad (2.91)$$

$$\max_{p_{\mathcal{B}} \leq C_{\mathcal{B}}} p_{\mathcal{B}} = C_{\mathcal{B}}, \quad (2.92)$$

which we use to formulate the continuous max-flow min-cut formulation

$$\max_{\substack{\|p\| \leq C \\ p_{\mathcal{O}} \leq C_{\mathcal{O}} \\ p_{\mathcal{B}} \leq C_{\mathcal{B}}}} \min_u \int_{\Omega} p_{\mathcal{B}} + u (p_{\mathcal{O}} - p_{\mathcal{B}} + \operatorname{div} p) dx, \quad (2.93)$$

through moving the maxima out of the integral and interpreting  $p_{\mathcal{O}}$  and  $p_{\mathcal{B}}$  as  $L^1$ -functions.

Finally, the flow variables  $p : \Omega \rightarrow \mathbb{R}^{\dim(\Omega)}$ , and  $p_s, p_t : \Omega \rightarrow \mathbb{R}$  can be seen as Lagrangian dual variables of the continuous min-cut (2.86), and thus the continuous max-flow model is given by

$$\begin{cases} \max_{p_{\mathcal{B}}} \int_{\Omega} p_{\mathcal{B}} dx \\ \text{s.t. } p_{\mathcal{B}} \leq C_{\mathcal{B}}, p_{\mathcal{O}} \leq C_{\mathcal{O}}, \|p\| \leq C, \\ \operatorname{div} p - p_{\mathcal{B}} + p_{\mathcal{O}} = 0. \end{cases} \quad (2.94)$$

## 2 Background

Conversely, if we interpret the max-flow model (2.94) as the primal model, then the max-flow min-cut model (2.93) is its Lagrangian dual formulation and the segmentation  $u$  can be seen as the max-flow's Lagrange multiplier.

As in the comparison of (2.84) and (2.85), comparing the discrete max-flow with the continuous max-flow (2.94) again yields striking similarities:

$$\left\{ \begin{array}{l} \max_{p_s} \sum_{v \in \Omega} p_s(v) \\ \text{s.t. } 0 \leq p_s \leq C_s, 0 \leq p_t \leq C_t, \|p\| \leq C, \\ \left( \sum_{v' \in \mathcal{N}_v} p(v, v') \right) - p_s + p_t = 0, \end{array} \right. \quad \left\{ \begin{array}{l} \max_{p_{\mathcal{B}}} \int_{\Omega} p_{\mathcal{B}} dx \\ \text{s.t. } p_{\mathcal{B}} \leq C_{\mathcal{B}}, p_{\mathcal{O}} \leq C_{\mathcal{O}}, \|p\| \leq C, \\ \operatorname{div} p - p_{\mathcal{B}} + p_{\mathcal{O}} = 0. \end{array} \right. \quad (2.95)$$

Both last lines of (2.95) represent flow conservation constraints, where in the discrete case  $\mathcal{N}_v$  represents the neighborhood of the vertex  $v$ . The main differences are given in the non-terminal capacities and flows since their domains and codomains differ as follows:

$$\begin{array}{ll} \text{discrete:} & \text{continuous:} \\ C : E \subset V \times V \rightarrow \mathbb{R}_{\geq 0}, & C : \Omega \rightarrow \mathbb{R}_{\geq 0}, \\ p : E \subset V \times V \rightarrow \mathbb{R}_{\geq 0} & \rightsquigarrow \left( \begin{array}{c} p(v, v+dx_1)/dx_1 \\ \vdots \\ p(v, v+dx_{\dim \Omega})/dx_{\dim \Omega} \end{array} \right) \rightsquigarrow p : \Omega \rightarrow \mathbb{R}^{\dim \Omega}. \end{array} \quad (2.96)$$

In the discrete case, the capacities and flows are values on the graph's edges. In the continuous case, the edges become infinitely small and are represented by the spatial positions directly. This leads to the effect that the continuous flows need to be represented as point-wise multidimensional densities. This is symbolically denoted in (2.96) between the two leads-to arrows, where  $p(x)$  would need to be evaluated through directional derivatives  $\lim_{\Delta x_i \rightarrow 0} p(v, v+\Delta x_i)/\Delta x_i$  by introducing a vertex  $v+\Delta x_i$  infinitely close to  $v$  and a flow  $p(v, v+\Delta x_i)$  between  $v$  and  $v+\Delta x_i$  which could locally represent a connection towards that direction. This reformulation of a discrete flow to a continuous one seems practically irrelevant, whereas in the opposite direction the reformulation of a continuous flow density to a discrete flow is straightforward since such discretizations are naturally performed with graph cuts [93, 86].

Moreover, in the continuous case one can extend the non-terminal capacity to  $C : \Omega \rightarrow \mathbb{R}_{\geq 0}^N$  with  $1 \leq N \leq \dim \Omega$ . This way we can control the spatial dimensions separately by adapting the component-wise comparator in  $p \leq C$  accordingly – as demonstrated in [117, 118] and Equation (6.2).

Finally, super-level sets, as one of the representations from Section 2.2.6, unused in this section so far, lead to another representation of the pcMS which mainly is a lifted version of (2.93) and is illustrated in the publication of Nosrati et al. [86]. This version introduces an additional critical parameter for the discretization of the

## 2.2 Technical Background

”lifting dimension” which generates higher computational costs, as described in Section 2.2.6. We did not further utilize this convexification technique for our purposes, and therefore, continue with the unlifted version (2.93) in the following.

As mentioned earlier on Page 55, we can augment TV to ATV by introducing a strongly positive definite [113] Riemannian metric matrix field  $A : \Omega \rightarrow \mathbb{R}^{\dim \Omega \times \dim \Omega}$  with which the generic segmentation model (2.86) is modified to

$$\min_{u: \Omega \rightarrow [0,1]} \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + C \sqrt{\nabla u^T A \nabla u} dx. \quad (2.97)$$

The matrix field  $A$  being a metric requires  $A(x)$  to be symmetric and positive definite in every point  $x \in \Omega$  and therefore we can uniquely decompose  $A(x) = S(x) S(x)^T$ , which allows us to write  $\sqrt{\nabla u(x)^T A(x) \nabla u(x)} = \|S(x)^T \nabla u(x)\|$ . Following [19, Appendix 8.A], we also find the non-terminal saturation property (2.87) for the anisotropic case and thus we can reformulate the max-flow min-cut formulation (2.93) to

$$\max_{\substack{\|p\| \leq C \\ p_{\mathcal{O}} \leq C_{\mathcal{O}} \\ p_{\mathcal{B}} \leq C_{\mathcal{B}}}} \min_u \int_{\Omega} p_{\mathcal{B}} + u (\operatorname{div} S p - p_{\mathcal{B}} + p_{\mathcal{O}}) dx. \quad (2.98)$$

**Optimizer** Even if the analytic models are formulated in the continuous setting, the optimization algorithms finally need to be discretized, for example, by following (2.28). To this end, different approaches to optimize the min-cut problem have been proposed, which also necessitate slightly different dualities, e.g. [23, 119, 108].

In the remainder of this section, we motivate the optimizer we have chosen for the work in Chapter 6 and 7, which is based on the continuous max-flow algorithm of Yuan et al. [23]. Furthermore, instead of the isotropic TV, we enhance the regularization to the anisotropic case of ATV and derive an algorithm as in [120]. In the following, we directly refer to the max-flow min-cut with ATV (2.98), where the algorithm proposed in [23] is a special case with  $S$  being the identity matrix.

As already mentioned above, the min-cut and max-flow formulations can both be seen as primal and dual. The authors in [23] see the max-flow (2.94) as primal, the max-flow min-cut (2.93) as primal-dual, and the min-cut (2.85) as dual. Seeing the max-flow min-cut formulation as primal-dual makes sense because it is a mixture of both min-cut and max-flow formulations, where both min and max operators are present, respectively. However, this labeling, ”primal-dual”, could be misleading because of the well-known primal-dual algorithms [108, 121], which are based on the Legendre-Fenchel-Rockafellar duality. The max-flow min-cut algorithm proposed by Yuan et al. [23], however, is based on Lagrange duality, where the primal max-flow is

## 2 Background

modified with an augmented Lagrangian:

$$\max_{\substack{\|p\| \leq C \\ p_B \leq C_B \\ p_O \leq C_O}} \min_u \underbrace{\int_{\Omega} p_B + u (\operatorname{div}(S p) - p_O + p_B) - \frac{c}{2} \|\operatorname{div}(S p) - p_O + p_B\|^2 dx}_{=: L_c}. \quad (2.99)$$

They (correctly) refer to the proposed algorithm as the augmented Lagrangian multiplier-based continuous max-flow algorithm. Using the categories discussed in the work of Esser et al. [121], this augmented Lagrangian algorithm is a "splitting algorithm applied to a dual formulation", whereas the primal-dual algorithm proposed in [108] is similar to the primal-dual hybrid gradient algorithm [122] and is a generalization of the Arrow-Hurwicz algorithm [123].

Indeed, the duality described by Yuan et al. [23] is not referring to the Lagrangian or Fenchel duality; however, they exhibit the duality between max-flow and min-cut to establish a connection to the well-known graph cut models by reformulating saturation for the continuous case. Meanwhile, using Fenchel's duality, Chambolle and Pock [108] were looking for algorithms with nice convergence properties that solve non-smooth convex optimization problems with known saddle-point structure, such as the present TV-regularized min-cut problem.

The multiplier-based max-flow algorithm [23] alternately maximizes the three flow variables  $p$ ,  $p_B$ , and  $p_O$  and minimizes the multiplier  $u$ , which in this case is the labeling map:

$$\begin{aligned} p^{k+1} &= \operatorname{argmax}_{\|p\| \leq C} L_c(p_B^k, p_O^k, p, u^k), \\ p_B^{k+1} &= \operatorname{argmax}_{p_B \leq C_s} L_c(p_B, p_O^k, p^{k+1}, u^k), \\ p_O^{k+1} &= \operatorname{argmax}_{p_O \leq C_t} L_c(p_B^{k+1}, p_O, p^{k+1}, u^k), \\ u^{k+1} &= u^k - c (\operatorname{div}(S p^{k+1}) - p_B^{k+1} + p_O^{k+1}). \end{aligned} \quad (2.100)$$

For each iteration  $k \rightarrow k+1$ , each variable  $p$ ,  $p_B$ ,  $p_O$ , and  $u$  is being updated by fixing the others. While each update looks like a complicated optimization on its own, they actually can be solved explicitly by setting the following partial variational derivatives to 0 where possible:

$$\begin{aligned} \frac{\delta L_c}{\delta p} &= -S^T \nabla (u - c (\operatorname{div}(S p) - p_B + p_O)) \stackrel{!}{=} 0 && \rightsquigarrow \text{gradient ascent } \frac{\partial p}{\partial t} = \frac{\delta E}{\delta p}, \\ \frac{\delta L_c}{\delta p_B} &= 1 - u - c (\operatorname{div}(S p) - p_B + p_O)(-1) \stackrel{!}{=} 0 && \rightsquigarrow \text{solve for } p_B, \\ \frac{\delta L_c}{\delta p_O} &= u - c (\operatorname{div}(S p) - p_B + p_O) \stackrel{!}{=} 0 && \rightsquigarrow \text{solve for } p_O, \\ \frac{\delta L_c}{\delta u} &= \operatorname{div}(S p) - p_B + p_O \stackrel{!}{=} 0 && \rightsquigarrow \text{gradient descent } \frac{\partial u}{\partial t} = -\frac{\delta E}{\delta u}. \end{aligned} \quad (2.101)$$

## 2.2 Technical Background

The slightly more complicated derivation of the variational derivative  $\frac{\delta L_c}{\delta p}$  with the involvement of  $S$  can be reproduced by realizing that the dual of the differential operator  $d = \text{div} S$  is  $d^* = -S^T \nabla$ , i.e.  $\frac{\delta L_c}{\delta p} = \partial_p L_c - S^T \nabla \partial_{\text{div} S p} L_c$ . The derivation is shown in more detail in [19, Appendix 8.B].

Solving the extremality equation  $\frac{\delta E}{\delta p} \stackrel{!}{=} 0$  for  $p^{k+1}$  for each step is inefficient since it is a second-order partial differential equation. We, therefore, follow Yuan et al. [124] and choose to update  $p$  by gradient ascent. The updates for  $p_{\mathcal{B}}$  and  $p_{\mathcal{O}}$  can be found by solving for the root, which is possible because of the introduced augmented Lagrangian. Without the augmentation, their derivatives would be  $\frac{\delta E}{\delta p_{\mathcal{B}}} = 1 - u$  and  $\frac{\delta E}{\delta p_{\mathcal{O}}} = u$ , where the required variables are missing. The derivative towards direction  $u$  still does not depend on  $u$  itself, even with augmentation, and therefore we choose to find the minimum by gradient descent. However, we only perform one gradient descent step, which is one iteration with the explicit Euler method (2.28), and this also coincides with the augmented Lagrangian's multiplier update in (2.100), proposed in [23]. Thus, the last line in (2.100) could be written as  $u^{k+1} = \text{argmin}_u L_c(p_{\mathcal{B}}^{k+1}, p_{\mathcal{O}}^{k+1}, p^{k+1}, u)$ . The considerations in (2.101) lead to the following update scheme,

$$\begin{aligned} p^{k+1} &= p^k + \Delta t_1 \frac{\delta E}{\delta p}(p^k), \\ p_{\mathcal{B}}^{k+1} &= \text{div} S p^k + p_{\mathcal{O}}^k + (1-u^k)/c, \\ p_{\mathcal{O}}^{k+1} &= -\text{div} S p^k + p_{\mathcal{B}}^{k+1} + u^k/c, \\ u^{k+1} &= u^k - \Delta t_2 \frac{\delta E}{\delta u}(u^k), \end{aligned} \tag{2.102}$$

though, unfortunately  $p^{k+1}$ ,  $p_{\mathcal{B}}^{k+1}$ , and  $p_{\mathcal{O}}^{k+1}$  do not necessarily respect the constraints  $p_{\mathcal{B}} \leq C_s$ ,  $p_{\mathcal{O}} \leq C_t$ , and  $\|p\| \leq C$ . To this end, after each update step, the mentioned variables have to be projected onto the constraints' domain boundary:

$$\begin{aligned} p^{k+1} &= \mathcal{P}_p \left( p^k + \Delta t_1 c S^T \nabla \left( \text{div} S p^k - p_{\mathcal{B}}^k + p_{\mathcal{O}}^k - \frac{u^k}{c} \right) \right), \\ p_{\mathcal{B}}^{k+1} &= \mathcal{P}_{p_{\mathcal{B}}} \left( \text{div} S p^k + p_{\mathcal{O}}^k + \frac{1-u^k}{c} \right), \\ p_{\mathcal{O}}^{k+1} &= \mathcal{P}_{p_{\mathcal{O}}} \left( -\text{div} S p^k + p_{\mathcal{B}}^{k+1} + \frac{u^k}{c} \right), \\ u^{k+1} &= \mathcal{P}_u \left( u^k - \Delta t_2 (\text{div} S p^{k+1} - p_{\mathcal{B}}^{k+1} + p_{\mathcal{O}}^{k+1}) \right), \end{aligned} \tag{2.103}$$

where  $\mathcal{P}$  represents the corresponding projections. Interestingly, in our experiments, the proposed updates for  $u$  kept the values of  $u$  in  $[0, 1]$  without any projections directly on  $u$  necessary.

Regarding the choice of this algorithm of Yuan et al. [23], we refer to a comparison between the augmented Lagrangian method and the primal-dual approaches, as conducted in [125]. Indeed, primal-dual methods show the fastest convergence rates [108],

## 2 Background

however, as tested in [125] on their chosen binary-labeling problem, their primal-dual implementation was not much faster. We, therefore, chose to use the augmented Lagrangian method [23] because of its implementational simplicity and the mentioned duality between min-cut and max-flow. Furthermore, at the beginning of this Ph.D., computational efficiency was not as relevant as designing a well-generalizing model. Provided a model’s critical points represent the right solutions for the given problem, the choice of an optimizer, in this case, was not the crucial component in terms of accuracy. The same max-flow min-cut models could be solved with other, faster-converging algorithms, and it would have been interesting to try out the primal-dual algorithms listed in the well-organized work of Esser et al. [121]. It is, however, not part of this thesis, since we abandoned the approach of variational segmentation and solved the problem with a deep learning framework later, see Chapter 8.

### 2.2.9 Including Appearance Models

Before we jump to the topic of supervised deep learning, first, as an approach towards that direction, we talk about including shape and appearance models to the variational segmentation approaches.

The admissible solutions of the Mumford-Shah functional, so far, do not know what they are representing. Given a deteriorated image of a SC, for example, with motion artifacts, an optimal solution of the pcMS may represent an utterly unrealistic object. To counteract, we need to include prior knowledge of the population of possible realistic segmentations. Provided, for example, a set of manual reference segmentations, we can estimate probability maps to evaluate the likelihood of the appearance of a segmentation.

For proper evaluation of such a likelihood, typically the target image and the admissible segmentation need to be co-registered to the likelihood’s model space [90]. Likewise, all reference segmentations first need to be pairwise co-registered to estimate a probability map that excludes the variability of different positioning of the object of interest. On the topic of image registration we refer to Page 31 and Equation (4.1).

We differentiate between shape models and appearance models, where shape models represent object boundaries and appearance models also include the texture of the objects [90]. Both models can be learned from a set of examples, which usually involves co-registration of the samples to a model space. This can be done through rigid co-registration to align differently positioned objects, or additionally also through pixel-to-pixel freeform co-registration to more precisely differentiate the shape variability. Sampling from a model created with pixelwise co-registration is done within the space of displacement fields and generates sharp and more realistic instances, whereas sampling from a model with rigidly aligned samples merely superimposes objects from the population, yielding blurry or probabilistic instances, see also [126, Sec. 2].

For our case of 2D label maps, we decided to create appearance models using rigid co-registration only. Superimposing binary label maps from the population yields gray-valued maps that can be combined with relaxed label maps from the current



subject the following way. Given a relaxed label map as an iteration of a continuous cut during optimization, the projection of such an iteration into the appearance model space regularizes the segmentation towards a more realistic representation. Including pixelwise freeform co-registration between appearance/shape models and the segmentations may improve the regularization, however, at a higher computational cost in each iteration of the segmentation process. Furthermore, pixelwise co-registration introduces further errors and most certainly fails to generalize for deteriorated images.

In the following, we list approaches that combine the pcMS with appearance models. As mentioned in [81], the segmentation functional and the shape/appearance functional mostly are combined linearly, where the shape/appearance regularization mainly is modeled by a mean squared difference between the appearance  $\psi$  and the iterated segmentations  $u$  as follows:

$$\operatorname{argmin}_{u, \psi} E_{\text{segmentation}}(u, I) + \gamma \|u - \psi\|^2, \quad (2.104)$$

where  $\gamma$  controls the influence of the shape/appearance regularization. Generally, such algorithms then are designed, such that  $\psi$  converges together with  $u$ , in the sense of  $\Gamma$ -convergence [49, p. 40].

Among others, Chan and Zhu [81] represented shape models as signed distance functions and included them into Chan and Vese's segmentation model (2.81) [94]. To find correspondence between the shape model and the segmentation, they proposed to rigidly register the shape model's signed distance function to the segmentation iterations during optimization, where, however, the shape model itself is static.

Later, Cremers et al. [126] proposed to use probability maps to represent shapes and combined this representation with a pcMS similar to (2.66) but without boundary length regularization. Similar to [81], the segmentation is only being regularized by the shape model energy. In addition to the rigid registration as in [81], they introduced dynamic Gaussian shape models, which do not only change their positioning but also adapt their shapes to the iterated segmentations through coupling with a Mahalanobis distance. Moreover, they showed that the set of relaxed shape probability maps is convex, and the dynamic shape models can be globally optimized.

The work of Overgaard et al. [22] showed that the additive combination of the Chan-Vese energy and the shape regularization (2.104) can be combined to a pcMS energy (2.57) to apply any continuous cut optimization. To this end, they introduced the image model  $I_{\text{model}}$ , the prior model  $I_{\text{prior}}$ , and the effective image  $I_{\text{eff}}$  which depend on the weak segmentation  $u : \Omega \rightarrow [0, 1]$  and the weak shape representation  $f : \Omega \rightarrow [0, 1]$  as follows:

$$\begin{aligned} I_{\text{model}} &= c_{\mathcal{O}} u + c_{\mathcal{B}}(1 - u), \\ I_{\text{prior}} &= b_{\mathcal{O}} f + b_{\mathcal{B}}(1 - f), \\ I_{\text{eff}} &= \lambda I + \mu I_{\text{prior}}, \end{aligned} \quad (2.105)$$

where  $c_{\mathcal{O}}$ ,  $c_{\mathcal{B}}$ ,  $b_{\mathcal{O}}$ ,  $b_{\mathcal{B}}$  are the mean image intensities of the respective regions, and  $\lambda$ ,  $\mu$  with  $\lambda + \mu = 1$  are positive weights to balance the influence of the shape model. The

## 2 Background

model image  $I_{\text{model}}$ , in this case, is the relaxed analog of the pcMS approximation, cf. Equation (2.56). They show that the energy

$$\int_{\Omega} \|\nabla u\| + \lambda (I - I_{\text{model}}(u))^2 + \mu (I_{\text{model}}(u) - I_{\text{prior}}(f))^2 dx \quad (2.106)$$

corresponds to the energy (2.104), if  $c_{\mathcal{O}} \approx b_{\mathcal{O}}$  and  $c_{\mathcal{B}} \approx b_{\mathcal{B}}$ , and finally, the continuous cut's model functions are found to be

$$C_{\mathcal{O}}(x, u, f) = (I_{\text{eff}}(x, f) - c_{\mathcal{O}}(u))^2 \quad \text{and} \quad C_{\mathcal{B}}(x, u, f) = (I_{\text{eff}}(x, f) - c_{\mathcal{B}}(u))^2. \quad (2.107)$$

Interestingly, this reformulation of the coupling between segmentation and shape energy (2.104) translates into the effective image  $I_{\text{eff}}$ , where the prior model  $I_{\text{prior}}$  is being superimposed onto the image  $I$ , resulting in a highlighting effect of the object of interest. Furthermore, the authors in [22] incorporated a new gradient descent procedure for the rigid registration of the statistical model by adapting the numerical step sizes to the curvature of *the prior and image dependent rigid transformation group*.

Indeed, segmentation models that include such statistical knowledge do not necessarily have a shallow architecture. However, the constructions presented above, where we simply combine prior knowledge through addition in the energy term, never reach the complexity of the non-linear approximation capability of deep neural networks.

Moreover, above segmentation models are closely related to atlas-based segmentation [98, 99, 127, 97]. Whereas appearance models consist of reference segmentations only, atlases are composed of tuples of MR images and their reference segmentations. Atlas-based algorithms register a collection of MR images onto the target image with pixelwise freeform registration, choose the best fitting reference segmentations as candidate segmentations, and subsequently combine the candidates with a label fusion technique such as the simultaneous truth and performance level estimation (STAPLE) [128]. In this sense, atlas-based algorithms are regularized by the collection of images in the used atlas. With the method (2.107) presented above, in contrast, the appearance model is not primarily co-registered with the target image but with the iterated segmentations, influencing the segmentation through a highlighting effect [22]. This way, the method does not directly use the appearance model's label maps as candidate solutions but lets the segmentation algorithm optimize for the best solution with a strong dependency on the original image information. Depending on the parameter  $\gamma$  in (2.104) or the balance between  $\lambda$  and  $\mu$  in (2.105) and (2.106), the involvement of an appearance model is thought of as a weak regularization.

### 2.2.10 Supervised Deep Learning

At the beginning of this decade, supervised deep learning induced an abrupt change in state-of-the-art segmentation methods. Classic segmentation frameworks such as morphological segmentation approaches, clustering algorithms, and variational approaches introduced in Section 2.2.7 were commonly used, but with the increasing

computing performance, and the better availability of datasets, supervised deep learning became the new state of the art. Many challenges, such as the SC GM segmentation challenge [14], were dominated by the performance of supervised deep learning approaches.

In this section, we briefly introduce concepts of supervised deep learning without going into details of designing a network. The main principle in applying supervised deep learning is as follows: first, we design or choose a network that is complex enough, i.e. capable of approximating a given problem; and second, we gather a large set of training data, consisting of images and manual segmentations; to third, optimize the network's weights in a training phase.

**Network Architectures** As a short literature review, convolutional neural networks revolutionized and simplified the modeling part of neural networks, where the U-net [129] introduced a new concept for segmentation networks. With the U-net's U-shaped architecture, low- as well as high-level features are gathered to produce an output of the same resolution as the input image. An alternative approach was to use recurrent neural networks (RNN) for image segmentation, which use recurrent iterations in a mode similar to variational segmentation approaches. In the work of [130], an RNN with long short-term memory (LSTM) was used in a multi-directional scanning scheme over the data points. On the basis of this work, Andermatt et al. developed a convolutional, recurrent neural network that uses simplified LSTM units, called gated recurrent units (GRU), in a convolutional, multidimensional, and multi-directional way, resulting in an architecture called multidimensional gated recurrent units (MDGRU) [24]. The work of Andermatt et al. outperformed other deep neural networks including U-Net approaches in several medical image segmentation challenges over extended periods of time [131].

The goals of this paragraph are to describe deep networks as non-linear approximators that can learn to divide complex processes into their constituent subprocesses and to clarify the notions "shallow" and "deep". To this end, we follow and borrow notations from the theoretical work of Poggio et al. [95].

The terms present for artificial neural networks are obviously influenced by the biological counterpart; however, artificial neural networks do not directly model a biological neural network. However, the nodes represent neurons, and the nodes' units represent connections like the axons between nerve cells. The units' activation functions mimic the saturation of the signal a nerve cell receives, and a node linearly combines all such signals. A linear combination of such activation functions is typically called a ridge function,

$$\sum_{i=1}^Q c_i \sigma(\langle x, w_i \rangle + b_i). \quad (2.108)$$

Some articles compute the node's output as  $\sigma(\sum_{i=1}^Q \langle x, w_i \rangle + b_i)$ , allowing to omit the coefficients  $c_i$  in (2.108), however, if these summations are applied recursively, conceptually the results are the same. Having a closer look at the activation function:

## 2 Background

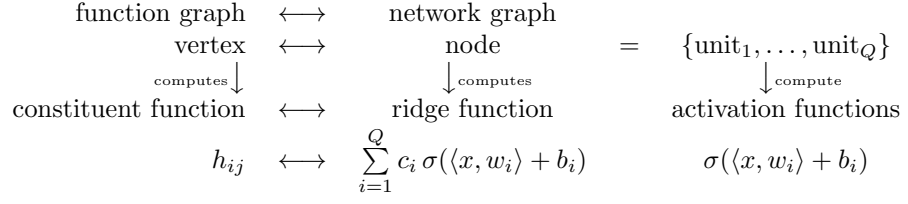


Figure 2.12: Hierarchically local compositional functions and neural networks.

$w_i$  is its weight;  $b_i$  its bias;  $x$  are the input data which could be a subdomain of the initial data or also a collection of outputs of other nodes; and  $\sigma$  the activation function itself, which could for example be the rectified linear unit  $\max(x, 0)$ . Ridge functions need to be efficient to compute, and therefore many other activation functions  $\sigma$  exist.

Besides the connection between the artificial and biological concepts of neural networks, in Figure 2.12 we illustrate the connection between the network's architecture and the approximated functions, where the ridge functions correspond to constituent functions, and the nodes correspond to evaluation vertices of sub-functions. To this end, the authors in [95] introduce the set of hierarchically local, compositional functions. Such functions are recursively composed from a set of constituent functions, as for example a binary-hierarchical function  $f : \mathbb{R}^8 \rightarrow \mathbb{R}$  is composed from a set of binary functions  $h_i : \mathbb{R}^2 \rightarrow \mathbb{R}$ :

$$f(x) = h_3 \left[ h_{21} \left( h_{11}(x_1, x_2), h_{12}(x_3, x_4) \right), h_{22} \left( h_{13}(x_5, x_6), h_{14}(x_7, x_8) \right) \right]. \quad (2.109)$$

Here, *local compositionality* refers to the sparse connectivity of the constituent functions, i.e. the dimensionality of the constituent functions' input arguments. The *connectivity order* of the net then could be defined as the maximal local connectivity. In the above function  $f$ , the connectivity order is 2 because all constituent functions depend on two variables. On the other hand, the *hierarchicality* refers to the graph of the hierarchical compositionality. To exemplify the plethora of neural networks, Poggio et al. (Figure 2.13) use the binary-hierarchical, locally composed function  $f$  in (2.109) as well as two other extreme cases: a shallow, non-locally composed function and a recursive function with higher connectivity. These functions and their associated counterparts as neural networks, according to Figure 2.12, are illustrated in the bottom part of Figure 2.13.

As pointed out in [95], binary-hierarchical networks represent an idealized version of the many deep convolutional neural networks in the literature, since the convolution's kernel sizes define the number of connections. In the binary-hierarchical network in the middle of Figure 2.13, the feature maps of all convolutions' outputs of one layer could be represented by the outputs of all nodes of one layer (here, with a seamless and non-overlapping stride). In this sense, a convolution can sparsely and compactly represent the behavior of multiple nodes of one layer, which simplifies the modeling of the network, as mentioned before.

## 2.2 Technical Background

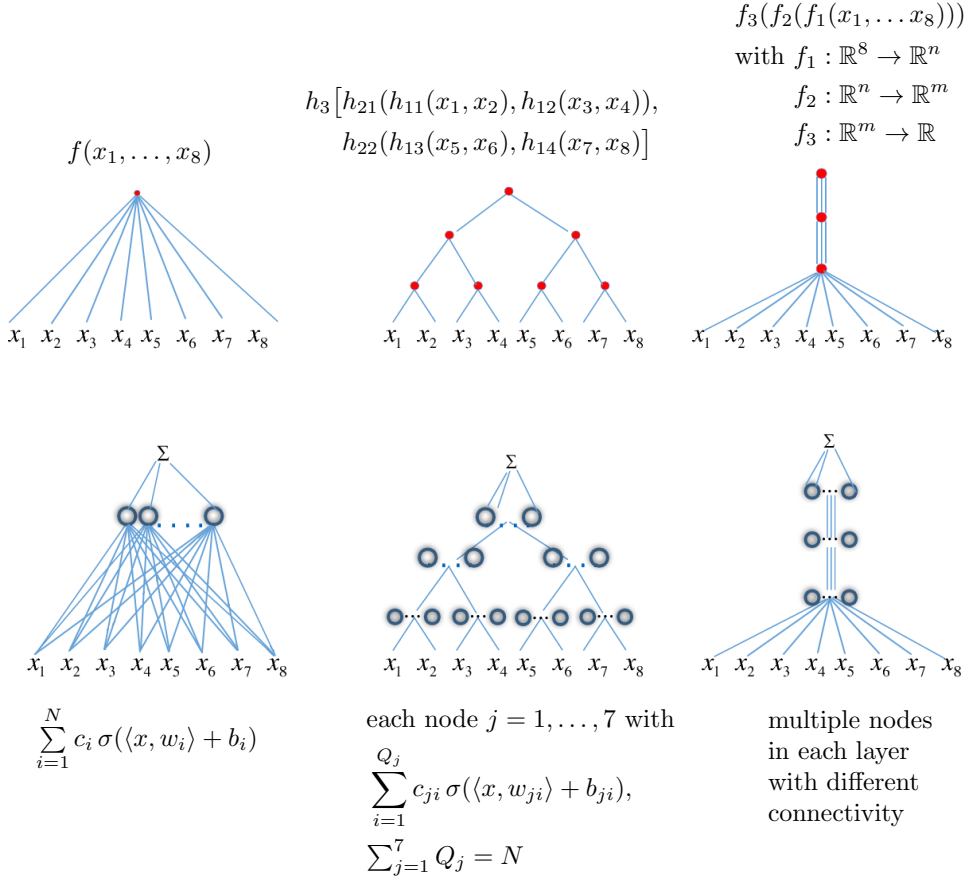


Figure 2.13: *Top row*: functions, where the small red dots represent evaluation vertices; *bottom row*: the top row's associated networks according to Figure 2.12, where the grayish circles represent the neural network's units, and the local collections of units represent the nodes.

*Left*: shallow network with one hidden layer and high connectivity; *middle*: deep binary-hierarchical network with low connectivity and high compositionality; *right*: deep hierarchical network with high connectivity. For visual simplicity, only one node per layer is drawn in the graphs on the right.

This figure is taken from [95, Fig. 1] with the Creative Commons license<sup>a</sup>. We included blue lines to indicate the networks' full connectedness, where necessary.

<sup>a</sup>Creative Commons Attribution 4.0 International License

## 2 Background

We define a network to be deep, if it has more than one hidden layer, a connectivity order of at least two, and if each weight of all units represents a degree of freedom. Ignoring the third condition that the units' weights should be freely optimizable, we could perceive the proposed continuous cut models in Chapters 6 and 7 as deep models. Continuous cuts could be seen as networks of the type visualized on the right side in Figure 2.13, with many hidden layers and without pooling in the end to a single value. Each hidden layer would represent a gradient descent iteration, where most weights would be modeled as repeated constants from layer to layer, driving the layers' outputs to a steady-state. Now, although many hidden layers are present, continuous cuts are not as deep as convolutional neural networks, because deep networks adapt their free weights in a more globally connected manner during the supervised learning phase.

Finally, both shallow and deep networks enjoy the universal approximation theorem, which states that any continuous function can be arbitrarily well approximated, provided enough degrees of freedom are given in the network – which, on the other hand, also describes the curse of dimensionality. To this regard, it does not matter whether higher connectivity (though, at least connectivity order of 2) or higher local compositionality is present. However, deep networks can deal with the curse of dimensionality better than shallow networks, as has been shown in [95], by imposing smoothness constraints on the constituent functions: for special classes of smooth, local compositional functions, exponentially fewer parameters are needed for deep networks to reach the same accuracy as shallow networks, if the deep network adopts an architecture that approximately matches the approximated function's compositionality. For accurate approximations, therefore, a network has to have high connectivity or a problem-specific hierarchical structure with at least a local connectivity of 2.

The authors in [95] hypothesize that the wiring in the cortex, or more precisely the limited number of local connections between the neurons in the brain, impose the hierarchical, local compositionality onto the way of thinking, which reflects in our language. If so, deep neural networks are even closer to the human way of thinking than shallow models with very high connectivity, as argued in the philosophical discussion on Page 46. Such compositionality, however, is not necessarily part of all natural problems, as there might still be problems that can be approximated better by shallow high-order networks.

**Supervised Learning** As already described in Section 2.2.7, a training set needs to be available to learn the optimal weights of a network. A training set consists of tuples of input data and their results. In the case of image segmentation, the input data are images, and the results are (manual) reference segmentations. As shown in Figure 2.9, supervised learning is a gradient descent process on the weights  $c_i$ ,  $b_i$ , and  $w_i$  (2.108) of all nodes over the training set. Due to the huge size of the training set, in practice, not all the training data are used for each gradient step, but random batches from the image domains of randomly chosen training samples are used. Such an optimization process is usually realized with some variant of stochastic gradient

descent (SGD). As a loss function for the descent process one chooses a distance between the network’s result and the reference result of a current random batch. An example of a loss function is the cross-entropy loss. Applying SGD on a neural network’s loss function mainly involves two steps: first, the forward pass, where the network evaluates a given subsample and calculates the result’s loss; and second, the backward pass, where the difference of the network’s result and the reference result is backpropagated through the network. During backpropagation, the gradients for the individual weights  $c_i$ ,  $b_i$ , and  $w_i$  (2.108) of all nodes are calculated by applying the chain rule to the loss function, which is called automatic differentiation, cf. Page 21.

As pointed out in the previous paragraph, a network approximates a given problem best, if the problem’s constituent architecture is contained in this network’s architecture. Since, however, the architecture of the problem is not always known, a general strategy is to design the network as complex as possible but with a reasonable amount of parameters, such that the weights are learnable. Such an overparameterization is not directly a devastating problem since SGD is particularly efficient in optimizing overparametrized neural networks [132, 133], since with smaller subsamples the estimated, calculated gradients have less noise.

Finally, similar to classical model-based segmentation, the goal of deep learning is to generate a model which generalizes to unseen data. Inherent to overparametrized neural networks is the problem of overfitting to the training data which leads to the inability to generalize on other data. Countermeasures against overfitting are, for example, regularization through dropout [134] or dropconnect [135]. Common practice is to divide the available data into three mutually exclusive datasets with similar distributions: the training, the validation, and the test datasets. The validation dataset is used to determine when to stop the learning process, and independently of this stopping criteria, the test dataset is used to evaluate the final model.

### 2.2.11 Validation and Comparison of Segmentation Models

Segmentations can be validated through comparison with other segmentations. For that purpose, throughout this work, we specifically compare the segmentation’s label maps with the Dice similarity coefficient (DSC) and the Hausdorff distance (HD), and the regions’ areas via the relative standard deviation (RSD), which is also called the coefficient of variation (CV).

For binary segmentation, the DSC between two segmentations refers to the DSC of their objects. The DSC is defined as the area of the intersection of the two objects divided by the mean value of their individual areas. For multi-label segmentations, the same is applied for each object, utilizing the so-called one-hot encoding, cf. Page 41. The DSC yields a value of 1 if the prediction and the reference are identical, and a value of 0, if they do not have any pixels in common. This score is not very sensitive to noise, because if most pixels are identical, except for e.g. a few distant pixels, the DSC is still close to 1. To capture such cases, we utilize the HD measure, which rates noisy missegmentations with higher values. The HD is the largest value of the point-wise minimal distances between the predicted and the reference object. Finally, the CV

## 2 Background

directly measures the reproducibility of the areas of the predicted and the reference object, without consideration of the object's shapes. Two disjoint objects with the same area, for example, yield a perfect reproducibility although they probably are not comparable. Such cases, however, are already captured and penalized by scores of DSC and HD. Therefore, it is usually advisable to consider multiple such measures together, rather than evaluating methods only on the basis of one of them.

Keeping in mind that for in-vivo measurements, all segmentations – including the manual reference standards – are predictions, we nevertheless refer to comparisons between a method's prediction and the reference standard as *accuracy* scores. But inevitably, segmentations that are closer to the (unknown) truth than the manual reference standard would theoretically yield worse accuracy than segmentations that are closer to the (probably not entirely correct) manual reference standard. In addition to the accuracy score, we refer to comparisons between several segmentations of the same scan as *precision* scores. While we measure precision in our publications in Chapters 7 and 8 with all three scores DSC, HD, and CV, for accuracy, we chose to use only DSC and HD. We discarded CV in conjunction with accuracy since different approaches naturally lead to (consistent) under- or over-segmentation, for which the CV is not the appropriate measure because of its absolute values.

Assuming well-segmented images, the most relevant precision value for our task of area quantification is the CV. This work aims at estimating tissue atrophy of cross-sectional areas, and therefore, the CV of the areas is well suited because of its close relation to atrophy. We demonstrate this relationship in the following. Let  $A_1$  and  $A_2$  be two cross-sectional areas of a specific tissue, and let  $A_2$  be the result from the yearly follow-up scan with  $A_1 > A_2$ ; then the percentage of atrophy is given by

$$\text{relative atrophy} = \frac{A_1 - A_2}{A_1} = \frac{|A_1 - A_2|}{A_1}. \quad (2.110)$$

Taking a closer look at the CV of the two values  $A_1$  and  $A_2$ , by using the corrected sample variation  $\text{Var}$  and defining  $\bar{A} = \text{mean}(A_1, A_2) = (A_1 + A_2)/2$ , then

$$\begin{aligned} \text{CV}(A_1, A_2) &= \frac{\sqrt{\text{Var}(A_1, A_2)}}{\bar{A}} = \frac{\sqrt{(A_1 - \bar{A})^2 + (A_2 - \bar{A})^2}}{\bar{A}} = \frac{\sqrt{2} (A_1 - \bar{A})}{\bar{A}} \\ &= \sqrt{2} \frac{|A_1 - \bar{A}|}{\bar{A}} = \frac{\sqrt{2}}{2} \frac{|A_1 - A_2|}{\bar{A}} = \sqrt{2} \frac{|A_1 - A_2|}{A_1 + A_2} \end{aligned} \quad (2.111)$$

reveals that the CV is approximately a factor  $\sqrt{2}/2 = 1/\sqrt{2} \approx 0.7$  smaller than the relative atrophy, assuming  $\bar{A}$  is similar to  $A_1$ . In the case of follow-up scans without the presence of atrophy, the CV should be around 0 but yields positive values because of acquisition and segmentation errors. With the above reasoning, the scan-rescan statistics' CV values multiplied by a factor  $\sqrt{2} \approx 1.4$  can be seen as a lower bound for the detection of atrophy, at least in single-patient studies. Statistical effects caused through larger sample sizes may reduce this lower bound for detecting atrophy, as leveraged for example in [16].



A reference standard for segmentation is often created via a consensus of manual segmentations, calculated, for example, via pixelwise majority voting or STAPLE [136]. In Chapters 7 and 8, we however used the segmentations of only one expert rater as a manual reference standard because additional manual segmentations were unavailable. In turn, the strength of our work is rooted in data acquired in a scan-rescan fashion, which enables us to calculate precision scores even for deterministic computer-based methods.

To estimate accuracy errors caused during the processes of acquisition and segmentation more precisely, we experimented with acquiring high-resolution images of an ex-vivo SC specimen in formaldehyde using nanotomography. We, however, did not pursue this approach because of missing GM-WM contrast in the acquired images. Scanning whole dead bodies (without extraction of the SC) with AMIRA sequences of different resolutions, as listed on Page 85, would further enable to estimate errors caused through partial volume effects.

Another form of validation is prosecuted via the comparison of one and the same approach’s performance on different kinds of data, which is often applied to test an algorithm’s generalization capability. However, comparisons of one approach on different data and, likewise, comparisons between different approaches on the same data, are difficult to carry out. This difficulty arises from the fact that algorithms that are specifically designed to work on a specific type of data would most probably need to be adapted to work on other data, ultimately leading to a different algorithm. Therefore, not only because of the time-consuming task of adaptation but also because reasoning based on unintentionally different algorithms, such comparisons pose additional complications. However, provided small enough adaptations, such cross-comparisons reveal further insights on which variabilities stem from acquisition or segmentation errors. These circumstances made it particularly difficult to compare segmentation algorithms developed for the AMIRA sequence to others that had been developed for common MR sequences, such as [97, 137, 138, 14]. Nevertheless, we managed to test the deep learning approach with multidimensional gated recurrent units (MDGRU) on the data of the SC GM segmentation challenge, and to test the iterative non-local STAPLE (iNLS) approach [97] on AMIRA data, see Chapters 7 and 8.



## 3 The Underlying Project

This work is part of an SNSF project<sup>1</sup> that includes three work packages (WP) for three parties. In this section, we shortly describe our interactions with the groups of the other two WPs and summarize their results that reflect this project’s development.

1. The group of the first WP is searching for MR sequences with optimal GM-WM contrast.
2. Our part aims at computer-based analyses of the images acquired with the MR sequences from the first WP, regarding GM-WM differentiability.
3. The group of the third WP is conducting the clinical research, which includes recruiting volunteers and evaluating the results of the first two WPs. Satisfied with the quality of the acquisition and post-processing pipeline, the automatic segmentation approach finally can be used for statistical analyses to better understand the implications of pathological GM-WM changes in connection with the progression of MS.

The timelines of these WPs are not strictly sequenced after each other but have undergone several iterations.

### 3.1 The Search for an Optimal MR Sequence

For this section, basic terminologies for magnetic resonance (MR) imaging are pre-requisites. An introduction to MR can be found in [139].

Imaging the SC in-vivo is challenging due to 1) the SC’s fine inner structure, 2) low contrast between GM and WM, 3) variable bending of the SC, reinforced through different positioning of the head and neck, 4) the gradually varying GM-WM structure along the SC, rostral to caudal, 5) possible CSF pulsation effects [31], and 6) breathing and swallowing motion in the neck region. For several years already, different MR sequences such as 2D and 3D  $T_2^*$ -weighted gradient echo sequences and phase-sensitive inversion recovery (PSIR) sequences have been used to acquire SC images with GM-WM contrast [140, 14, 141, 47]. According to [25], the PSIR sequence was one of the only MR imaging methods with a sufficient SC GM-WM contrast useful for assessing GM-WM atrophy quantitatively; yet, PSIR sequences typically have long acquisition times for fine resolutions. Therefore, the need for fast and accurate GM-WM MR sequences is still present.

---

<sup>1</sup>SNSF grant number: SNF 320030-156860/1

### 3 The Underlying Project

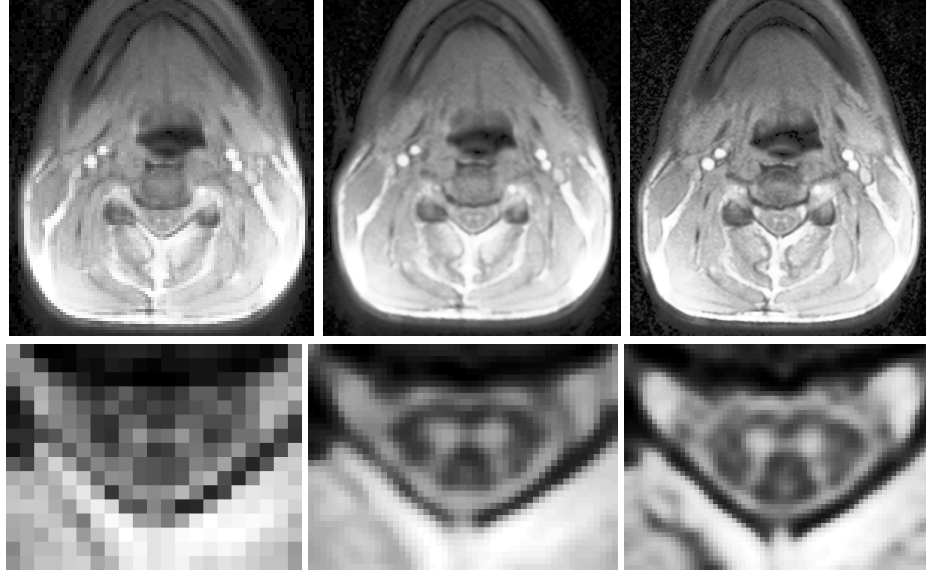


Figure 3.1: Three acquisitions of one axial slice on C3 level with a 3D-PSIR sequence of different resolutions. Slice thickness 4 mm. Histogram-equalized images. *Top row*: full view, *bottom row*: zoom on SC. *Left to right*: in-plane resolutions of 1.33 mm  $\times$  1.33 mm, 0.53 mm  $\times$  0.53 mm, and 0.40 mm  $\times$  0.40 mm, acquisition times of 1:10 min, 1:10 min, 2:20 min.

At the beginning of this project, the first images indeed were acquired by a 3D PSIR sequence, visualized in Figure 3.1, and those images were used for first segmentation attempts. Later, an implementation of the Modified Look Locker Inversion Recovery sequence (MOLLI) [142] was adapted for GM-WM contrast [143], with which we acquired images with remarkable tissue contrast, depicted in Figures 3.2 and 3.7. Scans acquired with the MOLLI sequence were used for the publication in Chapter 6. As an improvement of the MOLLI sequence concerning acquisition time, the Averaged Magnetization Inversion Recovery Acquisitions (AMIRA) sequence was developed [25], which produces images similar to the MOLLI sequence, as shown in Figures 3.5 to 3.7. The AMIRA sequence was used to acquire the CGM dataset, described on Page 88, which in turn was used for the publications in Chapters 7 and 8.

We would like to point out that from visual examination of axial SC GM-WM MR images, the butterfly-shaped GM is easily visible in the unzoomed full view images; when zooming in, however, the visual appearance gets worse, especially when looking at the raw data (using only nearest-neighbor interpolation), as can be seen in Figures 3.1 and 3.7. For the human vision, upsampling of the images helps to improve the visual appearance and furthermore it also helps to minimize numerical errors for the calculation of similarity scores such as the Dice coefficient and the Hausdorff

### 3.1 The Search for an Optimal MR Sequence

distance. Throughout this work, we usually upsampled the images by interpolating pixels with MATLAB's Lanczos3 kernel.

Feedback of the groups of the three WPs influenced the development of the MR sequences in terms of

- *quality* with signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), in-plane resolution, slice distance, slice thickness, slice alignment, and sharpness,
- *acquisition time* to reduce motion artifacts induced through breathing and swallowing, and
- *reproducibility*, which involves the number of axial slices, the slice distance, slice alignment, and practical simplicity in preparing and conducting acquisitions.

The final MR acquisition protocol is a carefully balanced compromise between all of these contradicting aspects and thus cannot accept the individual factors' optima, because, for example, a finer resolution with the same CNR requires a longer acquisition time. An in-plane resolution of  $0.24\text{ mm} \times 0.24\text{ mm}$  with 8 mm slice thickness needs around 7 min to acquire one slice with the AMIRA sequence, which already leads to high probabilities for motion artifacts, especially when patients are involved. As a rule of thumb, reducing the voxel volume by a factor of 2 leads to a fourfold increase in acquisition time. Finer resolutions, however, play an essential role in reducing partial volume effects which are strongly present for the delicate SC structures, as can be seen in the exemplary images in Figures 3.1, 3.9 and 3.10.

Besides the trade-off between resolution and acquisition time, another crucial requirement for short acquisition times lies in the reproducibility. To accurately reproduce SC GM-WM volumes in the neck region, a high number of axial slices with small slice distances are important, because a few distant slices alone could not represent the whole cervical SC. The SC's variability is further increased by the patients' variable poses on the MR scanner bed resulting in different bendings of the SC. If the orientation of the acquired slice is not perpendicular to the SC centerline, larger slice thicknesses result in more partial volume effects of the thin GM structure, as can be seen in Figure 3.9 on Page 84. Two straightforward solutions to that problem are smaller voxel sizes as in almost isotropic 3D scans or manual alignment of the slices perpendicular to the SC centerline. Now, the benefits of having single slices compared to 3D scans are shorter acquisition times per scan, resulting in a lower chance for motion artifacts. Given the gradually varying course of the SC, therefore, proper alignment of the slices perpendicular to the spinal cord allows for thicker slices without a drastic rise in partial volume effects, which even leads to higher SNR, cf. Figure 3.9.

In the remainder of this section, we describe the two sequences MOLLI and AMIRA that we used for our publications in Chapters 6 to 8. In addition, we describe two experiments to assess the impact on the image quality with respect to the resolution, the slice thickness, and slice angulation. These experiments, however, are based on a few samples only, and evaluations were mainly conducted by visual appearance. Nevertheless, we believe that the findings are interesting and worth mentioning.

### 3 The Underlying Project

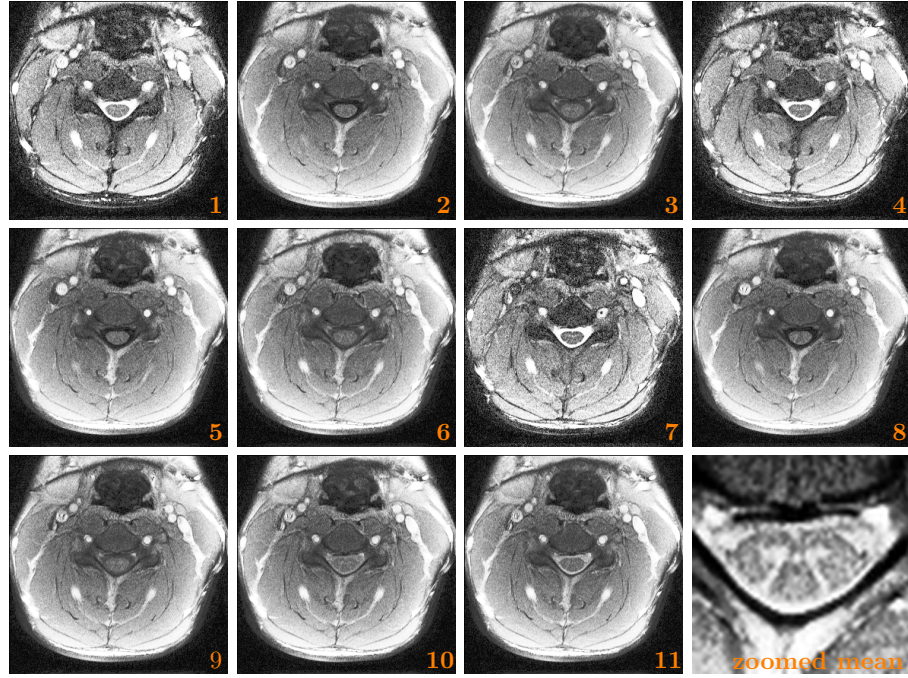


Figure 3.2: MOLLI sequence, an IR sequence with 11 inversion images. One axial slice on C4 level, slice thickness 8 mm, in-plane resolution of  $0.67 \text{ mm} \times 0.67 \text{ mm}$ . *Last image*: Zoomed average of all acquired inversion images. Histogram-equalization was used for better visualization.

**MOLLI Sequence** The adapted Modified Look Locker Inversion Recovery sequence (MOLLI) [143], is based on the original work of Messroghli et al. [142], which was developed for cardiac  $T_1$  relaxometry. The MOLLI sequence acquires inversion images during 3 consecutive inversion recovery (IR) experiments, where the inversion images are timed to capture the state during end-diastoles, controlled through cardiac gating. At the beginning of each period, a  $180^\circ$  inversion pulse is applied, and only a few milliseconds after, several inversion images are acquired at different effective inversion times. During the first and the second inversion periods, 3 inversion images are acquired, followed by another 5 in the third period, resulting in a total of 11 inversion images. From inversion pulse to inversion pulse, relatively long repetition times (TR) on the order of seconds are needed to allow a (partial to full) recovery of tissue magnetization towards the thermal equilibrium. To sample the inversion images, instead of spoiling the magnetization, Messroghli et al. used a balanced steady-state free precession (bSSFP) readout, which, according to [25], offers the highest SNR

### 3.1 The Search for an Optimal MR Sequence

per unit time of all MRI sequences. Availability of code for the MOLLI sequence led to testing this combination of an IR sequence with bSSFP sampling for SC GM-WM imaging. Moreover, the approach with bSSFP became a valuable tool because of robustness in motion and flow, and therefore several magnetization preparation techniques with application-specific contrasts have recently been proposed [144]. No CSF pulsation artifacts were observed with the MOLLI sequence.

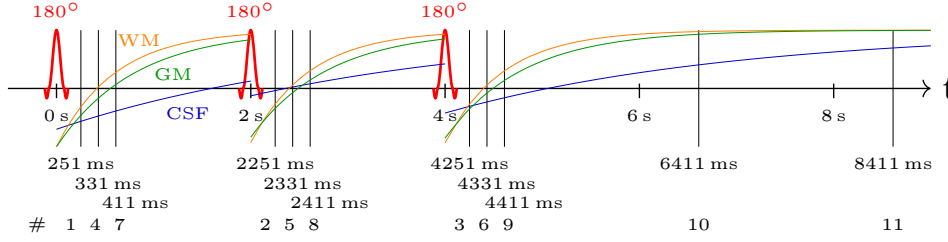


Figure 3.3: The MOLLI sequence’s schematic inversion times and tissue-specific longitudinal relaxation curves. Inversion image numbers at the bottom.

The work of Weigel and Bieri [143] adapts the original MOLLI sequence [142] as follows: three consecutive inversion pulses are applied with a fixed repetition time of 2 s, where the effective inversion times are (251, 331, 411) ms during the first two inversion periods and (251, 331, 411, 2411, 4411) ms during the third period, as shown in Figure 3.3. This sequence was used to acquire images with a field of view of  $128 \text{ mm} \times 128 \text{ mm}$ , a sample grid of  $288 \times 288$ , an in-plane resolution of  $0.4 \text{ mm} \times 0.4 \text{ mm}$ , and a slice thickness of 8 mm.

In Figure 3.3, we additionally illustrate tissue-specific longitudinal magnetization relaxation curves. These curves are only schematic and should give an idea of how the different tissue contrasts in the 11 inversion images are generated – compare with Figure 3.2.

Inversion images 1, 4, and 7 have high CSF-WM contrasts and low noise levels, whereas the other images contain information for GM-WM differentiation and at the same time have slightly lower SNR, which is due to the signal decay at larger inversion times after each inversion pulse. Interestingly, the average of all inversion images together produces images with remarkable GM-WM contrasts, cf. top row of Figure 3.7. We used such averaged images in our first publication in Chapter 6. Unfortunately, with a total acquisition time of 2:14 min for only one slice, the proposed sequence is not optimal for clinical use.

**AMIRA Sequence** The AMIRA sequence [25] is a mixed  $T_1$ - and  $T_2$ -weighted inversion recovery (IR) sequence that uses balanced steady-state free precessing (bSSFP) [144] sampling and exploits and improves the setup of the MOLLI sequence [143], to make it applicable for clinical studies. Using a multi-shot IR-prepared sequence with



### 3 The Underlying Project

a time-limited cine bSSFP readout, Weigel and Bieri improved the sample efficiency, resulting in shorter acquisition times. The AMIRA sequence captures 8 consecutive images in time directly after inversion, whereas the MOLLI sequence acquires 11 images unevenly distributed over three inversion periods. As a consequence, the two approaches produce different tissue contrasts for the acquired inversion images, as visualized in Figure 3.5. Four different averages of the 8 inversion images are shown in Figure 3.6 and a direct comparison to the MOLLI sequence on exemplary images is provided in Figure 3.7.

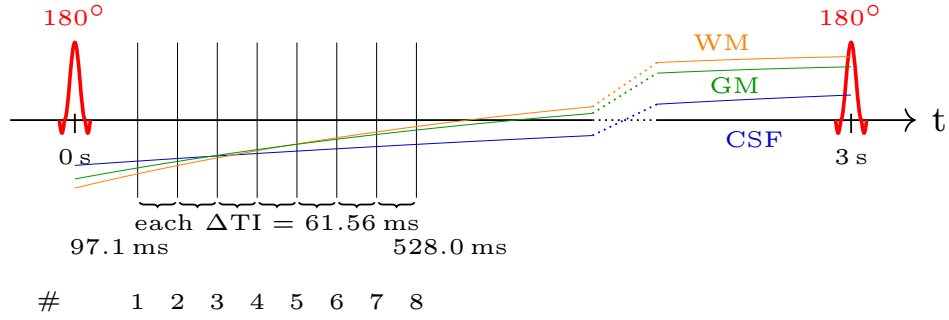


Figure 3.4: The AMIRA sequence's schematic inversion times and approximated tissue-specific relaxation curves after steady-state from TR to TR was reached.

The AMIRA sequence has a longer inversion pulse repetition time of 3 s as compared to 2 s for MOLLI, but only needs 17 such repetitions to fully encode k-space for a sample grid of  $192 \times 192$ , and directly is acquiring 8 instead of only 3 images per preparation. At effective inversion times of (97.1, 158.7, 220.2, 281.8, 343.3, 404.9, 466.5, 528.0) ms of each repetition, a part of the k-space for each inversion image is retrieved, visualized in Figure 3.4. For a field of view of  $128\text{ mm} \times 128\text{ mm}$ , the sample grid of  $192 \times 192$  yields an in-plane resolution of  $0.67\text{ mm} \times 0.67\text{ mm}$ , where 17 repetitions of 3 s each result in an acquisition time of only 51 s.

The presented steady-state relaxation curves in Figure 3.4 are approximations of mixed effects of bSSFP sampling during the first 600 ms and free recovery afterward, which is analyzed in more detail in [25, Fig. 2]. Nevertheless, the curves in Figure 3.4 approximately reflect the course of the different tissue contrasts of the inversion images, shown in the mid-row of Figure 3.5. As already shown for the MOLLI sequence's longitudinal relaxation curves in Figure 3.3, similarly, the magnetizations in the AMIRA sequence do not fully recover in each IR episode due to short repetition times. Moreover, during the acquisition of the 8 inversion images in the first 600 ms, the relaxation curves of CSF, WM, and GM are all negative and increase towards zero. These relaxation curves lead to a decreasing signal for the inversion images acquired at higher effective inversion times, which is visible in the first row



### 3.1 The Search for an Optimal MR Sequence

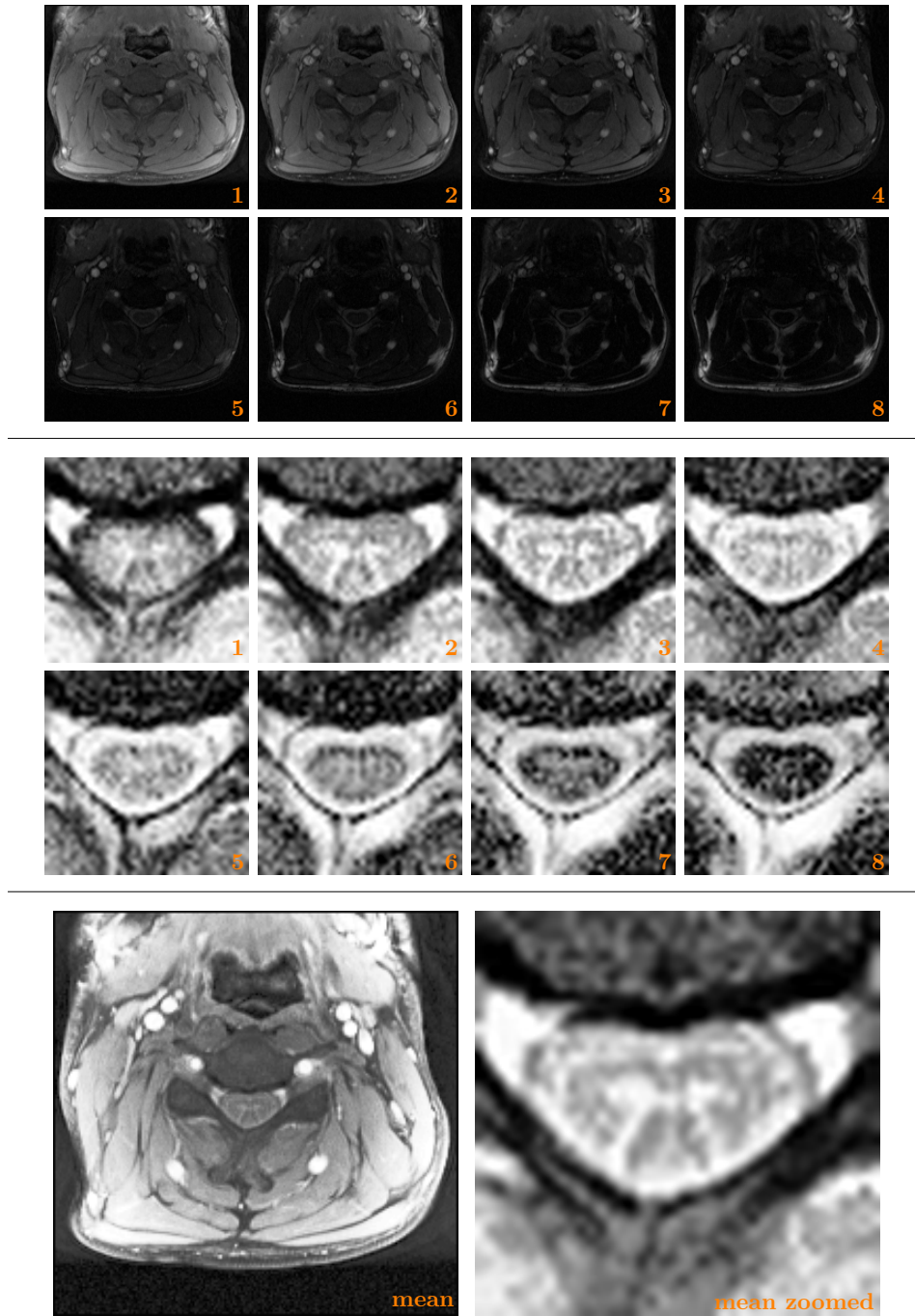


Figure 3.5: AMIRA, IR sequence with 8 inversion images. One axial slice, C4 level, slice thickness 8 mm, in-plane resolution  $0.67 \text{ mm} \times 0.67 \text{ mm}$ . *Top*: raw images, *middle*: zoomed, histogram-equalized, upsampled, *bottom*: average of the first 5 acquired inversion images, histogram-equalized.

### 3 The Underlying Project

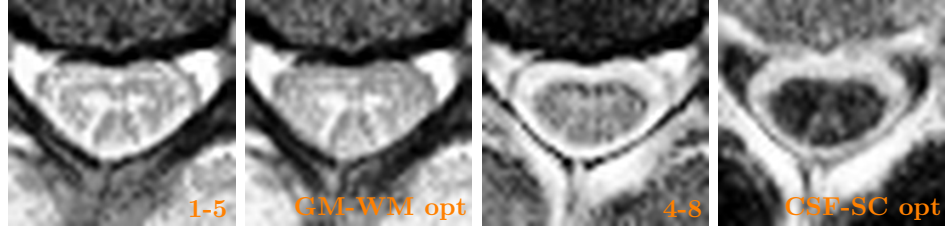


Figure 3.6: AMIRA averages, cf. Chapter 5. *Left to right*: average of the first 5 inversion images; optimized GM-WM contrast (observe that CSF-WM contrast is not preserved); average of the last 5 inversion images; optimized CSF-SC contrast.

of Figure 3.5. Through normalization, however, tissue contrasts are still observable – especially for CSF – at the cost of a slightly lower SNR.

While the MOLLI sequence achieves high CSF-SC contrasts in the inversion images 1, 4, and 7 of the first inversion pulse and GM-WM contrast in the other images, cf. Figures 3.2 and 3.3, AMIRA has the highest GM-WM contrast in the first two inversion images and increasing CSF-SC contrast in the last three images. As explained in [25], averaging the inversion images leads to an expected reduction of noise by a factor of  $1/N$ , where  $N$  is the number of involved images. Therefore, averaging the first and the last few images leads to improved CNR between GM and WM and good CNR between CSF and SC, respectively, as also analyzed in Chapter 5 and visualized in Figure 3.6.

**The Angulation Experiment** The slices’ orientation and thickness strongly influence the image quality via partial volume effects. To estimate the impact, we acquired 20 images of one subject with different slice thicknesses of 8 mm, 6 mm, and 5 mm and different slice orientations that are tilted towards the anteroposterior direction of slices orthogonal to the SC centerline with angles of  $-20^\circ$ ,  $-15^\circ$ ,  $-10^\circ$ ,  $-5^\circ$ ,  $0^\circ$ ,  $+5^\circ$ ,  $+10^\circ$ ,  $+15^\circ$ , and  $+20^\circ$ , as depicted in Figure 3.8. Among those, 15 images were acquired for all combinations of the mentioned thicknesses and negative angles located at the axial level of the spinal disc C2-C3. For the positive angles only five slices were acquired with a thickness of 8 mm, located on the level of the spinal disc C4-C5. Slices with a thickness of 5 mm were averaged over two subsequent acquisitions to improve SNR. The results are shown in Figure 3.9.

To estimate the impact of different angles on GM segmentation, all 20 images were manually segmented by one experienced rater and were also automatically segmented with an algorithm similar to the one described in Section 6.A (but with the GM appearance model used in Chapter 6 and without slice-similarity). These segmentations, together with their Dice similarity coefficients (DSC), are shown in Figure 3.9. Please note that the calculated DSC values have a sample size of 1 for each combination, and therefore, no significant conclusion can be drawn; yet, perfectly aligned orthogonal

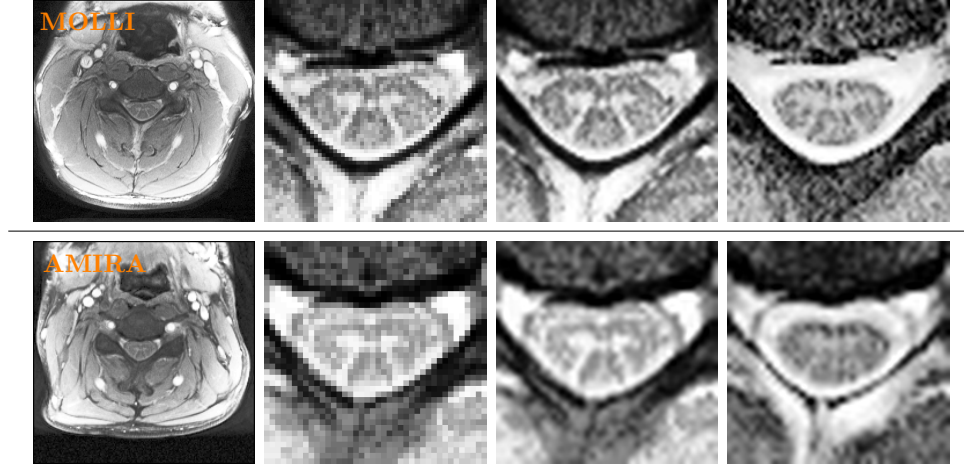


Figure 3.7: MOLLI and AMIRA averages of a healthy subject acquired at C4 level in different sessions. *Left to right*: full view; zoom on SC (raw; 5-fold upsampled; with CSF contrast). First three images are averages of all 11 MOLLI images (*top*), and averages of the first 5 AMIRA inversion images (*bottom*); last image with high CSF contrast is the average of the MOLLI images 1, 4, 7 (*top*), and the average of the last 5 AMIRA inversion images (*bottom*).

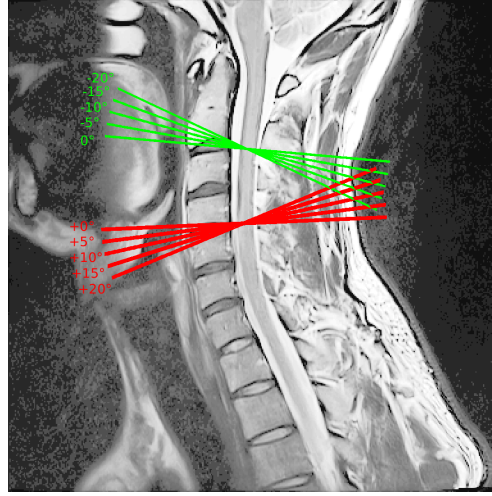


Figure 3.8: Localization of the slices acquired with different orientations used in Figure 3.9. In *red*, the slices with positive angles wrt. the plane orthogonal to the SC centerline, located at C4-C5 level, and in *green*, the slices with negative angles on C2-C3 level. Shown are the mid-positions of the slices.

### 3 The Underlying Project

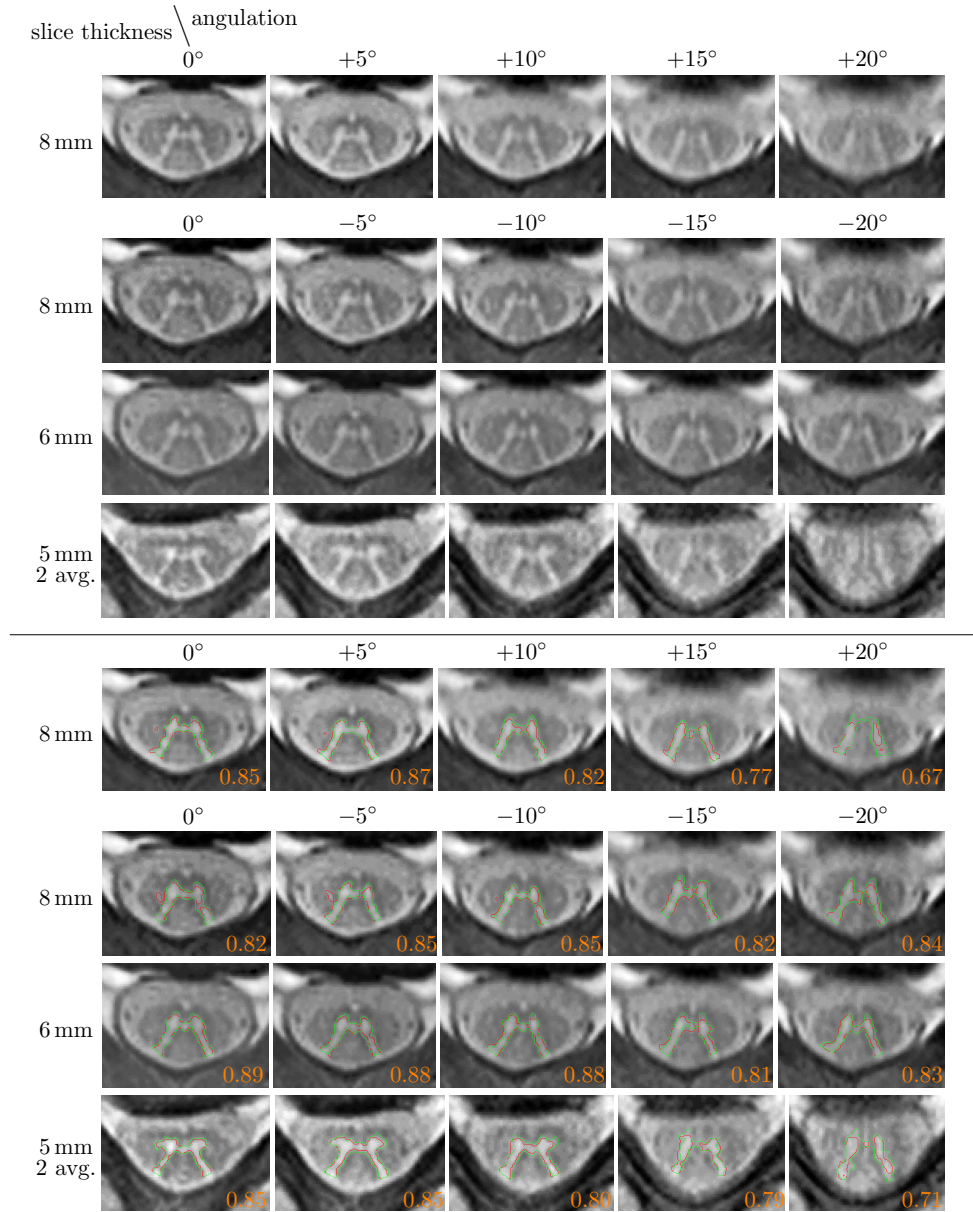


Figure 3.9: AMIRA slices on C2-C3 level (negative angles), and C4-C5 level (positive angles), as visualized in Figure 3.8. Manual segmentations in *green*, automated segmentations in *red*, and Dice similarity coefficients in *orange*.

### 3.1 The Search for an Optimal MR Sequence

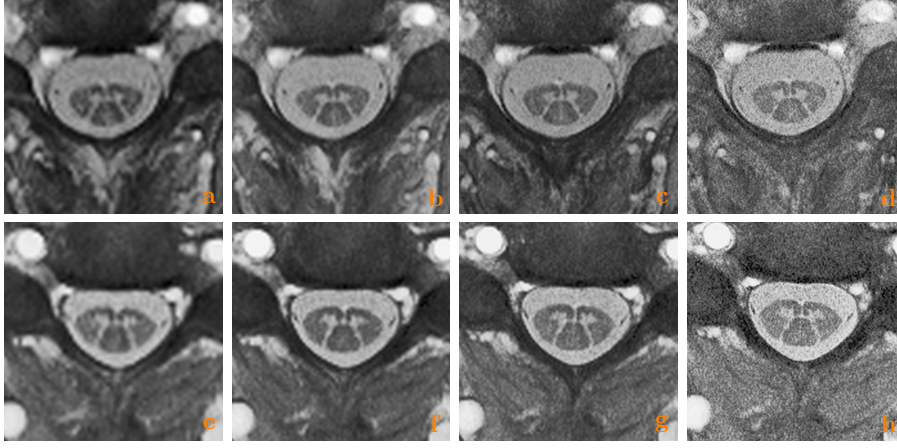


Figure 3.10: High resolution images located at the spinal disc C2-C3 of two different healthy subjects (*top* and *bottom* row). *Left to right*: in-plane resolutions of  $0.67\text{ mm} \times 0.67\text{ mm}$ ,  $0.5\text{ mm} \times 0.5\text{ mm}$ ,  $0.38\text{ mm} \times 0.38\text{ mm}$ ,  $0.24\text{ mm} \times 0.24\text{ mm}$ ; acquisition times of 51 s, 2:27 min, 3:27 min, 6:45 min; and slice thickness 8 mm, 5 mm, 8 mm, 8 mm. For better SNR, the images with resolution finer than  $0.67\text{ mm}$  were averaged over two acquisitions. For visualization of each image, the 8 AMIRA inversion images were averaged and subsequently the average was upsampled to a common interpolated resolution of  $0.067\text{ mm}$ .

slices tend to have less variability between the two segmentation approaches. A similar trend can be visually confirmed with orthogonally aligned slices having less partial volume effect. Conversely, thicker slices induce a higher impact on partial volume effects for tilted orientation. In these images, slices with a thickness of 8 mm seem to be a suitable compromise between partial volume effects and SNR. This investigation shows the importance of properly angulated slices.

**The Low-Resolution High-Resolution Experiment** To assess the impact of different voxel sizes on partial volume effects, different settings for high-resolution AMIRA images were tested on healthy volunteers. Exemplary axial images located at the spinal disc C2-C3 are shown in Figure 3.10.

The acquired images of this experiment showed that averaging two subsequent acquisitions improves SNR and the visual appearance for resolutions finer than  $0.67\text{ mm}$ ; averaging over more than two acquisitions, however, yielded blurry images because of small motion artifacts. The acquisition times without double-acquisitions are 51 s, 1:14 min, 1:44 min, 3:23 min and increase more than linearly with respect to the in-plane resolutions of  $0.67\text{ mm} \times 0.67\text{ mm}$ ,  $0.5\text{ mm} \times 0.5\text{ mm}$ ,  $0.38\text{ mm} \times 0.38\text{ mm}$ , and  $0.24\text{ mm} \times 0.24\text{ mm}$  and 8 mm slice thickness.



### 3 The Underlying Project

The in-plane resolution of  $0.38 \text{ mm} \times 0.38 \text{ mm}$  yields impressive images where the ventral horns begin to reveal their fine structure. On image [c](#) of Figure [3.10](#), even the posterior median sulcus is visible which on most images with a coarser resolution manifests as a blurry spot because of partial volume effects.

Blurred representations of the posterior median sulcus are indeed an obstacle for GM segmentation algorithms, as experienced with the variational approaches as well as the supervised deep learning approaches, since these spots consist of similar intensities as GM. Ongoing experiments also revealed a possibility for misclassification between MS lesions and blurred posterior median sulci. Moreover, partial volume effects of the ventral horns in these experiments have shown to be problematic for the distinction between GM and lesions. Therefore, images acquired with higher resolutions could potentially increase the accuracy and precision of segmentation algorithms; however, with increasing acquisition times, the probability for motion artifacts increases, too.

**The AMIRA Protocol** Finally, although finer resolutions more accurately represent the SC anatomy and pathological changes, above considerations with strong weight on the time factor led us to choose the AMIRA sequence with a resolution of  $0.67 \text{ mm} \times 0.67 \text{ mm}$ , a slice thickness of 8 mm, and an acquisition time of 51 s.

Already slightly higher resolutions with similar SNR markedly increase the acquisition time, where for instance, an in-plane resolution of  $0.5 \text{ mm} \times 0.5 \text{ mm}$  already requires an acquisition time of 2:27 min. Such long acquisition times, as just mentioned, massively increase the probability of motion artifacts, especially for neurological patients with significant motor impairment or spasticity. Furthermore, shorter acquisition times allow acquiring more images over a more extended segment of the SC within a comparable session time.

While a slice thickness of 8 mm anatomically seems to be very coarse, we need to clarify that in the MR acquisition process, information is not accumulated uniformly across the axial slices in inferior-superior direction but rather in a Gaussian-like weighting. This thickness, finally, was chosen because of the higher SNR compared to thinner slices.

Addressing the reproducibility: to acquire SC segments on similar axial levels each time a subject is positioned in the scanner, the first slice is aligned perpendicular to the SC such that the lower end of the slice matches the upper end of the intervertebral disc C2-C3. From this position on, slices are shifted caudally with a slice distance of 4 mm and are successively realigned perpendicular to the SC, to acquire a total of 12 slices that cover approximately 5.2 cm of the SC from mid C2 to the beginning of C5. Because of the varying bending of the SC, the slice numbers cannot be associated with exact vertebral levels, yet the correspondences are roughly given by  $((1, 2), (3-6), (7-10), (11, 12)) \leftrightarrow (\text{C2}, \text{C3}, \text{C4}, \text{C5})$ .

With these settings, acquiring each slice needs 51 s, which results in approximately 10 min for all the slices. Since it was not initially apparent during our planning that the inversion images 5 to 8 produce an acceptable CSF-SC contrast, each slice

is directly followed by an additional axial CSF scan that mimics the first inversion images of the MOLLI sequence. Such images offer additional information for CSF-SC segmentation, but are not perfectly aligned with the AMIRA images. Furthermore, this protocol is accompanied by two sagittal 3D scans: a  $T_1$ -weighted MPRAGE scan with 120 sagittal slices, a resolution of  $1\text{ mm} \times 1\text{ mm} \times 1\text{ mm}$ , TR 1580 ms, TE 3.52 ms, TI 900 ms, flip angle  $8^\circ$ , and an acquisition time of 3:41 min; and a  $T_2$ -weighted TSE scan with 15 sagittal slices, averaged over 2 acquisitions each, with a resolution of  $0.6\text{ mm} \times 0.6\text{ mm} \times 3\text{ mm}$ , TR 3500 ms, TE 306 ms, flip angle  $160^\circ$ , and an acquisition time of 2:46 min. This way, information gathered by the AMIRA scans can be enriched by the sagittal 3D scans, as is utilized in the Manual Segmentation App, see next section and Figure 3.11.

## 3.2 Clinical Validation

The main research question from our medical collaborators addresses the correlation of GM-WM atrophy regarding the MS disease progression. We propose using the AMIRA protocol, described on Page 86, to acquire in-vivo SC GM-WM images of MS patients. However, before the ongoing experiments with images of MS patients, an essential step is to measure the reliability and performance of the AMIRA protocol in combination with our developed post-processing tools. To this end, we applied two pipelines on healthy subjects where no presence of radiological findings and GM-WM atrophy is assumed. The results are published in [145, 146], and presented in Chapters 7 and 8.

Because there is no ground truth available for our in-vivo experiments, we estimate the accuracy and precision of the proposed pipelines based on the variability resulting from scan-rescan and "segment-resegment" experiments. In "segment-resegment" experiments, we had multiple neuro-radiological experts segment the same images (multiple times). The agreement of the manual segmentations varies between the experts and also varies if one expert segments the same image twice, to which we refer as inter- and intra-rater variability, respectively. Because the task of manual segmentation is time-consuming, large datasets would need to be split into multiple parts, where a different expert rater segments each part. This splitting causes additional variability and biases, which are undesired for further statistical evaluations. Deterministic algorithms, in contrast, have zero variability in segmenting one image multiple times and can unrestingly process images of a huge dataset with one and the same algorithm. Nonetheless, manual segmentations are essential for training and validation of automated methods. Segmenting the AMIRA sequence by hand, however, is not straightforward since this sequence produces 8 relatively noisy inversion images that are difficult to keep in mind during segmentation simultaneously. We believe that a human expert needs less noisy one-channel projections of the 8 inversion images for comfortable and reliable delineation and therefore we proposed different average images, as shown in Figure 3.6. The collection of all inversion images, however, probably contains more information than a single one-channel projection, and

### 3 The Underlying Project

therefore we designed a browser-based application called Manual Segmentation App (MSA) that allows for simultaneous drawing and channel browsing. With MSA we provide a split-screen view of the AMIRA channels (we refer to the AMIRA inversion images and their average projections as the AMIRA channels) together with the  $T_1$ - and  $T_2$ -weighted sagittal 3D scans from the AMIRA protocol, which entail even more information. Finally, the client-server architecture within MSA simplifies data exchange.

In the remaining two paragraphs of this section, we describe our dataset consisting of AMIRA images of health subjects, called the CGM dataset, and briefly introduce the developed application MSA.

**CGM Dataset** The Cervical Gray Matter (CGM) dataset consists of a cohort of 24 healthy subjects (14 female, 10 male, age  $40 \pm 11$ ) and a dataset of 72 MS patients (44 female, 29 male, age  $49 \pm 13$ ). Since the MS patients currently still are being scanned and manually segmented, this thesis focuses on the dataset of the healthy subjects.

Each of the 24 healthy controls was scanned 3 times with the AMIRA protocol, described on Page 86. The first 3 scans were acquired in a scan-rescan fashion: the 1<sup>st</sup> and 2<sup>nd</sup> scans were acquired shortly after each other without repositioning of the subject in the scanner, describing the intra-session variability; for the 3<sup>rd</sup> scan, the subject left the scanner, walked around, and laid back into the scanner to simulate varying positioning and bending of the SC in inter-session comparisons, which is inevitably given in yearly follow-up scans. In total, with all 3 scans from the scan-rescan sessions, 875 slices were acquired, 10 slices of which were acquired twice because of motion artifacts; and one scan recorded a 13<sup>th</sup> slice on an even more caudal vertebral level. From the remaining regular  $24 \cdot 3 \cdot 12 = 864$  slices, we had to remove another 9 slices because of severe imaging artifacts, resulting in 5 scans with less than 12 slices: we discarded the last three caudal slices of one scan, the last two slices of two scans, and the last slice of another two scans.

Before segmentation, we upsampled all images 10-fold with MATLAB's Lanczos3 kernel. Two experienced raters, both with more than 4 years of experience in SC neuroimaging, manually segmented parts of the CGM dataset by manually delineating the background-CSF and the CSF-SC boundaries on one of the AMIRA average images with CSF contrast, shown in Figure 3.6, and the GM-WM boundary on one of the averages with GM-WM contrast:

- Rater 1 segmented all first three scans for each of the 24 healthy subjects and additionally segmented 60 randomly chosen slices with respect to slice location, scan session, and subject, without knowledge of such information, which we used to estimate the intra-rater variability;
- Rater 2 segmented all 1<sup>st</sup> scans of the 24 healthy subjects, which we used to estimate inter-rater variability.

All manual segmentations were performed with MSA (see next paragraph) except for the segmentation of the scan-rescan dataset delineated by Rater 1. Rater 1 first used



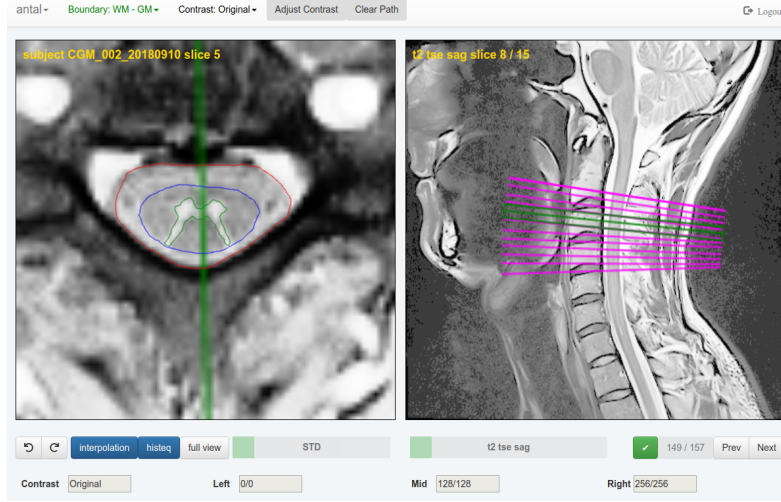


Figure 3.11: Screenshot of the Manual Segmentation App.

ITK-SNAP<sup>2</sup> for delineation, transferring the CSF segmentation that was obtained from the CSF contrast image onto the GM-WM contrast image to continue with the delineation of GM inside the provided SC mask. This sequential arrangement of tasks was tedious and was simplified with the use of MSA.

**Manual Segmentation App** With the JavaScript-based browser application called Manual Segmentation App<sup>3</sup> (MSA), we simplify the simultaneous visualization and segmentation of the data provided by the AMIRA protocol and speed up the data transfer of the raw image data and the segmentation results. On the user side, this application directly runs in a web browser and works without installation. The application connects to a server from where it receives the MR data and to where it saves the user's progress. The mode of operation is to slice-wise segment the respective tissues by manually drawing with a mouse or a stylus and clicking the "next" button to submit the drawings and receive new data. A screenshot is provided in Figure 3.11. To stay independent of the browser's window resolution, the positions of the manual delineations are captured in image coordinates using floating point values such that they could be resampled to any reasonable resolution afterward.

In the following, we go through MSA's features which are explained in more detail in MSA's documentation. The AMIRA images are visualized on the left side, while sagittal, axial, or coronal views of the T<sub>1</sub>- and T<sub>2</sub>-weighted 3D sequences are shown on the right. With the top-left drop-down list, the user can select specific slices, and next to it, the "Boundary" drop-down list enables the user to select the drawing

<sup>2</sup><http://www.itksnap.org> (last accessed on Sept. 10, 2019)

<sup>3</sup><https://github.com/neonroehre/MSA> (last accessed on Sept. 10, 2019)

### 3 The Underlying Project

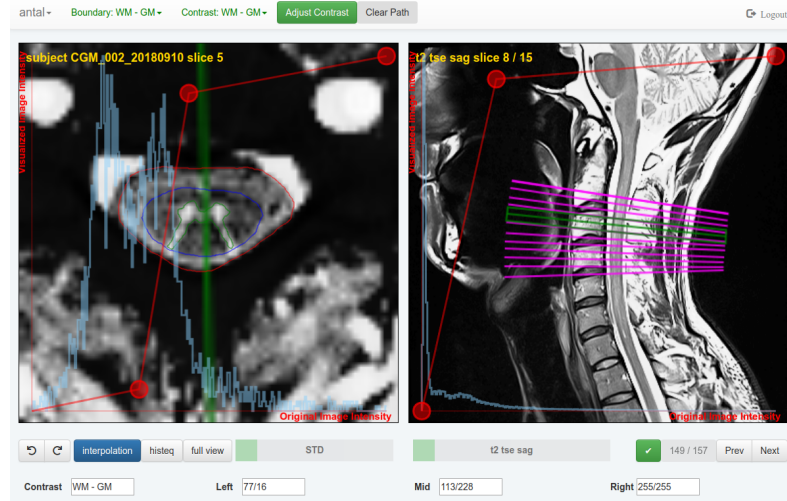


Figure 3.12: Screenshot of the MSA when adjusting the contrast.

color for the current object boundary. Then there is the "Contrast" drop-down list to add or to choose between manually defined image contrasts, which the users can design by themselves while having activated the "Adjust Contrast" button, as shown in Figure 3.12. Next there is a series of buttons that allow clearing the current path; boundary-specific undoing and redoing of drawing path segments; switching the visualization of the active boundary curve between "on" and "off"; toggling histogram equalization; and toggling between full and zoom view. With the scrollbar below the left image, we enable the navigation of the AMIRA channels, which is also possible via mouse scrolling or touch gestures on the left image frame; and with the scrollbar below the right image, we enable choosing between sagittal, axial, and coronal views of the T<sub>1</sub>- and T<sub>2</sub>-weighted 3D sequences. Finally, "previous" and "next" buttons enable the user to navigate through the AMIRA slices.

With activated "Adjust Contrast" button, see Figure 3.12, intensity histograms and intensity manipulation curves are shown in blue and red, respectively. With three red control points, the user can modify the visualized intensities with instant visual feedback via drag and drop. The coordinates of the control points, as well as a name for the currently active contrast, can be specified in the text boxes in the bottom row of MSA, allowing better reproducibility among users.

## 4 High-Order Slice Interpolation for Medical Images

The publication presented in this chapter addresses the large slice distance between the images acquired with the AMIRA protocol, described on page Page 86 and visualized in Figure 3.11. The goal of this work is to interpolate the missing information between the given slices to enable accurate resampling of the given data to any desired resolution. The presented algorithm finds application, for example, when visualizing a stack of AMIRA slices in 3D space, assisting volume quantification algorithms, or calculating intermediate slices for data augmentation during supervised deep learning.

Intensity-based interpolation algorithms linearly combine the intensities of neighboring pixels, but in cases where neighboring pixels are located on different slices, we need to find pixelwise correspondence between subsequent slices to combine the correct information. Since this Ph.D. thesis mainly focuses on *image segmentation*, we did not yet define the process of finding pixel-correspondence maps, which is called *image registration*, cf. Page 31. In short: given two digital 2D images  $I_1$  and  $I_2 : \Omega \rightarrow \mathcal{I}$  with the common intrinsic image domain  $\Omega \subset \mathbb{R}^{m \times n}$ , one tries to find a displacement field  $v : \mathbb{R}^{m \times n} \rightarrow \mathbb{R}^2$  by optimizing

$$\min_v \mathcal{D}(I_1 \circ (\text{Id} + v), I_2) + \mathcal{R}(v), \quad (4.1)$$

with an appropriate image distance  $\mathcal{D}$ , the identity function  $\text{Id}(x) = x$ , and a displacement field regularization  $\mathcal{R}$ .

We follow the idea of slice interpolation of Baghaie et al. [26], who propose to find displacement fields by mutually warping two subsequent slices towards each other. We improved their algorithm in three aspects:

1. we do not only interpolate one mid-slice between one pair of slices but variationally optimize a continuous model onto a whole stack of slices;
2. we do not only perform linear interpolation along one displacement field but include higher-order displacement field terms to interpolate structural changes along with the slices smoothly; and
3. we also adjust the linear combination of corresponding pixel intensities to a smooth intensity interpolation along the smoothly interpolated correspondence trajectories.

This approach describes a framework for arbitrary registration schemes that allows choosing a suitable scheme for a given problem. As in the work of Baghaie et al., we

#### 4 High-Order Slice Interpolation for Medical Images

chose to use the sum of squared differences (SSD) for  $\mathcal{D}$  in combination with curvature regularization for  $\mathcal{R}$  in (4.1).

With leave-one-slice-out experiments on phantom images as well as on MR images, we demonstrate superior performance with the proposed slice interpolation algorithm as compared to a re-implementation of the algorithm proposed by Baghaie et al. in [26].

The algorithm, however, does not directly consider different orientations of the slices. To compensate for that, one could embed the interpolated slices into the patient coordinate system given the slices' interpolated locations and orientations.

**Publication** The proposed method was presented at the workshop *Simulation and Synthesis in Medical Imaging* (SASHIMI 2017) in conjunction with the *20<sup>th</sup> International Conference on Medical Image Computing and Computer Assisted Intervention* (MICCAI) on the 10<sup>th</sup> of September 2017 in Quebec City, Canada. It was published<sup>1</sup> as part of the workshop proceedings [147]. MATLAB code<sup>2</sup> is made publicly available.

---

<sup>1</sup>[http://dx.doi.org/10.1007/978-3-319-68127-6\\_8](http://dx.doi.org/10.1007/978-3-319-68127-6_8) (last accessed on Sept. 10, 2019)

<sup>2</sup><https://mathworks.com/matlabcentral/fileexchange/63907> (last accessed on Sept. 10, 2019)

# High Order Slice Interpolation for Medical Images

Antal Horváth<sup>1</sup>() , Simon Pezold<sup>1</sup>, Matthias Weigel<sup>1,2</sup>, Katrin Parmar<sup>3</sup>,  
and Philippe Cattin<sup>1</sup>

<sup>1</sup> Department of Biomedical Engineering,  
University of Basel, Allschwil, Switzerland  
`antal.horvath@unibas.ch`

<sup>2</sup> Radiological Physics, Clinics of Radiology,  
University Hospital Basel, Basel, Switzerland

<sup>3</sup> Department of Neurology,  
University Hospital Basel, Basel, Switzerland

**Abstract.** In this paper we introduce a high order object- and intensity-based method for slice interpolation. Similar structures along the slices are registered using a symmetric similarity measure to calculate displacement fields between neighboring slices. For the intensity-based and curvature-regularized registration no manual landmarks are needed but the structures between two subsequent slices have to be similar. The set of displacement fields is used to calculate a natural spline interpolation for structural motion that avoids kinks. Along every correspondence point trajectory, again high order intensity interpolating splines are calculated for gray values. We test our method on an artificial scenario and on real MR images. Leave-one-slice-out evaluations show that the proposed method improves the slice estimation compared to piecewise linear registration-based slice interpolation and cubic interpolation.

**Keywords:** Slice interpolation · Image registration · Splines

## 1 Introduction

Medical images often have anisotropic resolution. For example, in magnetic resonance (MR) images the in-plane resolution is often higher than the through-plane resolution. Reslicing and upsampling are standard preprocessing steps when dealing with such data. This motivates the search for slice interpolating methods to increase the resolution between the slices. Standard intensity interpolations such as *nearest neighbor*, *linear*, or *cubic* interpolations are well established. By using such intensity interpolations between two slices with different structure we would calculate linear combinations of intensities of points that do not belong together, see Fig. 2 row 3 in the red box. Therefore it is better to interpolate the object structure and position, which involves registration techniques to find correspondences. Morphing images is often done using manual landmarks, but since putting landmarks manually would be tedious, an automatic registration is

70 A. Horváth et al.

advantageous. Grevera et al. [4] present a first comparison of slice interpolation methods that use correspondence points. In general, to guarantee meaningful slice interpolations, the structures of two subsequent slices have to be similar and the registration approach has to find an appropriate displacement field. Registration-based slice interpolation is an important field of research. Penney et al. [7] use a nonrigid registration algorithm with a spatial B-spline basis [8] but linear interpolation along the displacement field. Frakes et al. [3] use a modified version of control grid interpolation and a cubic interpolator for the displacement fields. Leng et al. [6] use a multi-resolution registration method and linear intensity interpolation along Catmull-Rom spline interpolated displacement fields.

Baghaie et al. [1] introduce a method with a symmetric similarity measure and curvature regularization

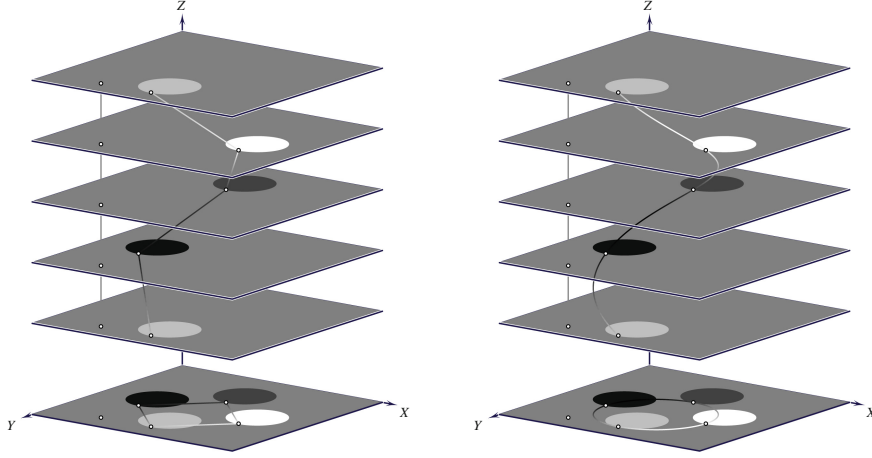
$$\operatorname{argmin}_{v: \Omega \rightarrow \mathbb{R}^2} \frac{1}{2} \int_{\Omega} [I_1(\mathbf{x} - v(\mathbf{x})/2) - I_2(\mathbf{x} + v(\mathbf{x})/2)]^2 + \lambda ((\Delta v_x(\mathbf{x}))^2 + (\Delta v_y(\mathbf{x}))^2) d\mathbf{x}, \quad (1)$$

where they look for a displacement field  $v$  by minimizing the intensity differences between the simultaneously displaced images and the bending of the resulting displacement field. The two images  $I_1$  and  $I_2$  share the common intrinsic image domain  $\Omega \subset \mathbb{R}^2$  and  $v_x$  and  $v_y$  are the displacement field components in  $x$  and  $y$  direction. Using a symmetric similarity has the advantage that both the “reference” and “target” can be warped in a more symmetric way. After minimizing energy (1), they can interpolate an image  $1/2 I_1(\mathbf{x} - v(\mathbf{x})/2) + 1/2 I_2(\mathbf{x} + v(\mathbf{x})/2)$  in the middle of  $I_1$  and  $I_2$  through linear intensity interpolation at the automatically registered and transformed correspondence points.

As an analytical property the zero level set of the curvature regularization contains harmonic functions and among them affine transformations [2]. This means that during minimization processes affine transformations are preferred as long as possible, as can be nicely seen in an example in [2, Fig. 2]. Moreover, the gradient descent steps with the curvature regularization can be iterated efficiently with a stable, implicit finite difference scheme.

Stacking together these linear interpolations between many neighboring slices results in a piecewise linear interpolation. At the stitching points, kinks may appear, which could be smoothed out using a higher order interpolation; see Fig. 1, left.

In this paper we derive a method that interpolates a whole stack of images through both object and intensity interpolation. Given the point correspondences between the slices, it calculates spline trajectories for all the correspondences and along these trajectories spline interpolations for the gray values; see Fig. 1, right. To find the correspondences between the slices we use the slice registration proposed in [1], but the proposed method can easily be adapted to other distance measures. Our contribution lies in solving the problem of combining higher order interpolations of structure motion and intensity. We describe the approach and the algorithm in Sect. 2, we test the proposed method on a test scenario and on real 3D images in Sect. 3, and we conclude in Sect. 4.



**Fig. 1.** Registration-based slice interpolation schemes: piecewise linear (*left*) vs. smooth interpolation (*right*), with their projections on the  $x$ - $y$ -planes (*bottom*). In both interpolations on the left and right, two exemplary correspondence curves are shown: correspondences of points in a flat region and correspondences of a boundary pixel of the circling ellipse. To get a full interpolation, for all pixels in the slices such correspondence curves are established. The proposed method optimizes over the whole  $z$  range for a smooth interpolation, similar as in the right scheme, avoiding kinks on the correspondence curves at the given slice positions.

## 2 Method

Let  $(I_k)_{k=1,\dots,P}$  be an ordered stack of  $P$  similar 2D images in  $\mathbb{R}^{M \times N}$ , i.e. the slices of a volumetric image or the frames of a movie. Assume slices  $I_k$  lie parallel to the  $x$ - $y$ -plane and their positions  $z_k \in \mathbb{R}$  along the  $z$ -axis are given with  $z_k < z_{k+1}$ . Our task is to interpolate a new slice at any distance between two subsequent slices or to refine the slice distances  $h_k = z_{k+1} - z_k$  by a factor  $R \in \mathbb{N}$ .

Given an image distance measure  $\mathcal{D}$  and a displacement field regularization  $\mathcal{R}$ , we minimize the summed up registration energies of all the neighboring image pairs at specific registration evaluation points  $\mathcal{S}$  along the  $z$ -axis:

$$\underset{\mathbf{v} = \{v_k: \Omega \rightarrow \mathbb{R}^2\}_{k=1}^{P-1}}{\operatorname{argmin}} \sum_{k=1}^{P-1} \sum_{s \in \mathcal{S}} \mathcal{D}[I_k \circ \vec{p}_k(s), I_{k+1} \circ \overleftarrow{p}_k(1-s)] + \lambda \mathcal{R}(\mathbf{v}_k). \quad (2)$$

Similar to [1], we use sum of squared distances for  $\mathcal{D}$  and curvature regularization for  $\mathcal{R}$  [2] but add the evaluation points  $\mathcal{S} = \{0, 1/2, 1\}$ . To clarify the notation, the intensity of the transformed images  $I \circ p(z)$  at point  $\mathbf{x}$  in the intrinsic image domain  $\Omega$  can be read out by  $I(p(z, \mathbf{x}))$ , whereas the transformations  $\vec{p}_k(s, \mathbf{x})$  and  $\overleftarrow{p}_k(1-s, \mathbf{x})$  are defined as follows: For each point  $\mathbf{x}$  we interpolate a trajectory  $p(z, \mathbf{x})$  for  $z \in [z_1, z_P]$  along the slices. With the displacement fields  $v_k$ , which give correspondences for subsequent slices, we construct a natural spline with the

third order Hermite interpolation polynomials  $p_k(z, \mathbf{x}) = p(z, \mathbf{x})|_{[z_k, z_{k+1}]}$  defined on the intervals  $[z_k, z_{k+1}]$  [5, p. 48]. Because of the well known inverse effect of transforming image domains, out of  $p_k$  we define two special transformations  $\vec{p}_k$  and  $\overleftarrow{p}_k$  with different parameter origins such that they are both formulated for the common intrinsic image domain  $\Omega$  of all slices:  $\vec{p}_k$  transforms the image in positive direction along the  $z$ -axis and  $\overleftarrow{p}_k$  in negative direction. Furthermore we reparametrize the transformations. For every  $k$  let  $z = z_k + s h_k$  with  $s \in [0, 1]$  and

$$\begin{aligned} \vec{p}_k(s, \mathbf{x}) &= \mathbf{x} - s v_k(\mathbf{x}) - s(s-1) \left[ s (a_{k+1}(\mathbf{x}) h_k - v_k(\mathbf{x})) + \right. \\ &\quad \left. + (s-1) (a_k(\mathbf{x}) h_k - v_k(\mathbf{x})) \right], \\ \overleftarrow{p}_k(1-s, \mathbf{x}) &= \mathbf{x} + (1-s) v_k(\mathbf{x}) - s(s-1) \left[ (1-s) (a_{k+1}(\mathbf{x}) h_k - v_k(\mathbf{x})) + \right. \\ &\quad \left. - s (a_k(\mathbf{x}) h_k - v_k(\mathbf{x})) \right]. \end{aligned}$$

For using piecewise polynomial Lagrange interpolation of degree three, four supporting points, or more abstractly, four degrees of freedom have to be set on each interval  $[z_k, z_{k+1}]$ . Between the obvious supporting points  $p_k(z_k, \mathbf{x}) = \mathbf{x}$  and  $p_k(z_{k+1}, \mathbf{x}) = \mathbf{x} + v_k(\mathbf{x})$  the remaining two are placed exactly at the same spots  $z_k$  and  $z_{k+1}$  [5]. At the positions where two supporting points are now on top of each other, Hermite interpolation can be used to calculate the derivatives  $a_k = p'(z_k)$  and  $a_{k+1} = p'(z_{k+1})$ . They define the other two degrees of freedom. The slopes  $a_k$  can be calculated through a nicely conditioned tridiagonal linear equation system of size  $P \times P$  by incorporating smoothness conditions of the trajectory interpolation  $p$ . For the third order spline to be two times differentiable, the first and second derivatives at the stitching positions of the neighboring interpolation polynomials are set to be equivalent:  $p'_k(z_{k+1}) = p'_{k+1}(z_{k+1})$ ,  $p''_k(z_{k+1}) = p''_{k+1}(z_{k+1})$ . This way, along each correspondence point trajectory, the spline interpolator minimizes the bending energy  $\int_{z_1}^{z_P} (p''(z))^2 dz$  among all two times differentiable interpolators [5]. Adding the natural spline condition  $p''(z_1) = p''(z_P) = 0$  defines all the degrees of freedom for the splines. For every  $\mathbf{x} \in \Omega$ , the optimal first derivatives  $a(\mathbf{x}) = (a_k(\mathbf{x}))_{k=1, \dots, P} = ((a_{kx}(\mathbf{x}), a_{ky}(\mathbf{x})))_{k=1, \dots, P}$  in  $x$  and  $y$  direction now can be calculated through solving the  $2 \cdot M \cdot N$  linear systems  $A a_{\cdot x}(\mathbf{x}) = d_{\cdot x}(\mathbf{x})$  and  $A a_{\cdot y}(\mathbf{x}) = d_{\cdot y}(\mathbf{x})$  of size  $P \times P$ , where

$$A = \begin{pmatrix} 2 & 1 & 0 & \dots & 0 \\ \frac{1}{h_1} 2 \left( \frac{1}{h_1} + \frac{1}{h_2} \right) & \frac{1}{h_2} & 0 & \dots & 0 \\ 0 & \ddots & & & \vdots \\ \vdots & & & \frac{1}{h_{P-2}} 2 \left( \frac{1}{h_{P-2}} + \frac{1}{h_{P-1}} \right) & \frac{1}{h_{P-1}} \\ 0 & \dots & 0 & 1 & 2 \end{pmatrix}, \quad \begin{aligned} d_1(\mathbf{x}) &= 3 \frac{v_1(\mathbf{x})}{h_1}, \\ d_k(\mathbf{x}) &= 3 \left( \frac{v_{k-1}(\mathbf{x})}{h_{k-1}} + \frac{v_k(\mathbf{x})}{h_k} \right), \\ d_P(\mathbf{x}) &= 3 \frac{v_{P-1}(\mathbf{x})}{h_{P-1}}. \end{aligned} \quad (3)$$

These systems can efficiently be solved by exploiting the structure of the invertible tridiagonal matrix  $A$ . After a Cholesky decomposition  $A = LL^T$ , the systems  $Aa = d$  can be rewritten by  $Le = d$  and  $L^T a = e$ . The lower triangular matrix



**Algorithm 1.** Proposed Slice Interpolation Algorithm

---

**Data:** slices  $I \in \mathbb{R}^{M \times N \times P}$ , slice distances  $h \in \mathbb{R}^{P-1}$ , refinement factor  $R$

**Result:** interpolated slices  $I^{\text{interp}} \in \mathbb{R}^{M \times N \times ((P-1)R+1)}$

Set up the spline matrix  $A \in \mathbb{R}^{P \times P}$  and calculate Cholesky decomp  $A = LL^T$ .  
 Calculate  $D \in \mathbb{R}^{M \times N}$ ,  $D_{ij} = 1 + \tau \lambda (-4 + 2 \cos((i-1)\pi/M) + 2 \cos((j-1)\pi/N))^2$  [2].  
 Initialize  $v = (v_x, v_y) = 0 \in \mathbb{R}^{M \times N \times (P-1) \times 2}$ .

**while**  $\|v^{\text{new}} - v^{\text{old}}\| \geq TOL$  **do**

% calculate displacement fields between slices

**for**  $k = 1, \dots, P-1$  **do**

$F_x = 0; F_y = 0;$  %  $\odot$  and  $\oslash$ : pointwise mult. and div.

**for**  $s \in \mathcal{S}$  **do**

$\Delta I = I_{k+1} \circ \overleftarrow{p_k}(s) - I_k \circ \overrightarrow{p_k}(1-s);$

$F_x = F_x + \Delta I \odot (s \partial_x I_k \circ \overrightarrow{p_k}(s) + (1-s) \partial_x I_{k+1} \circ \overleftarrow{p_k}(1-s));$

$F_y = F_y + \Delta I \odot (s \partial_y I_k \circ \overrightarrow{p_k}(s) + (1-s) \partial_y I_{k+1} \circ \overleftarrow{p_k}(1-s));$

**end**

$v_{kx}^{\text{new}} = \text{IDCT}(\text{DCT}(v_{kx}^{\text{old}} - \tau F_x) \oslash D);$  %  $v_{kx}, v_{ky} \in \mathbb{R}^{M \times N}$

$v_{ky}^{\text{new}} = \text{IDCT}(\text{DCT}(v_{ky}^{\text{old}} - \tau F_y) \oslash D);$

**end**

% calculate spline coefficients for object interpolations  $\overleftarrow{p_k}, \overrightarrow{p_k}$

Set up  $d \in \mathbb{R}^{m \times n \times p \times 2}$  as in (3) with  $v^{\text{new}}$  and solve  $Aa = d$  with (4).

**end**

% calculate spline coefficients for intensity interpolation

Set up  $d_I \in \mathbb{R}^{m \times n \times p}$  as in (6) and solve the systems  $Aa_I = d_I$  with (7).

% interpolate slices

$I_1^{\text{interp}} = I_1; l = 1;$

**for**  $k = 1, \dots, P-1$  **do**

**for**  $r = 1, \dots, R-1$  **do**

$s = r/R;$

% spline intensity interpolation of spline morphed slices

$\Delta I = I_{k+1} \circ \overleftarrow{p_k}(1-s) - I_k \circ \overrightarrow{p_k}(s);$

$I_l^{\text{interp}} = I_k \circ \overrightarrow{p_k}(s) + s \Delta I +$

$+ s(s-1) [s(h_k a_{I_{k+1}} \circ \overleftarrow{p_k}(1-s) - \Delta I) + (s-1)(h_k a_{I_k} \circ \overrightarrow{p_k}(s) - \Delta I)];$

$l = l + 1;$

**end**

$I_l^{\text{interp}} = I_{k+1}; l = l + 1;$

**end**

---

$L$  is invertible and only has one secondary diagonal on the first off-diagonal, and thus the spline coefficients  $a$  can be calculated by forward and backward substitution as follows:

$$\begin{cases} e_1 = d_1/L_{11} \\ e_k = (d_k - L_{k,k-1} d_{k-1})/L_{kk}, k = 2, \dots, P, \end{cases} \quad \begin{cases} a_P = e_P/L_{PP} \\ a_k = (e_k - L_{k,k+1} e_{k+1})/L_{kk}, k = P-1, \dots, 1. \end{cases} \quad (4)$$

74 A. Horváth et al.

For minimizing (2), we use gradient descent: alternately we calculate a new set of displacement fields through a descent step and calculate their spline interpolation – compare with the while loop of the pseudocode (Algorithm 1).

We could stop here and interpolate slices between the images  $I_k$  and  $I_{k+1}$  at any  $z = z_k + s h_k$  by linearly combining the intensities of the warped images:

$$I^{\text{interp}} = (1 - s) I_k \circ \vec{p}_k(s) + s I_{k+1} \circ \overleftarrow{p}_k(1 - s). \quad (5)$$

This would, however, result in kinks of the intensities along the trajectories and would result in a piecewise linear approximation.

We now construct a spline interpolation of the intensities along these trajectories. The correspondence interpolation is correct only locally along the interpolating axis, and registration errors between the corresponding points are summed up over several slices. Thus intensity information of slices further away should not have strong influence in the calculation of the interpolation. Because of the bounded supports  $[z_k, z_{k+1}]$  of the piecewise interpolating polynomials  $p_k$  and the local dependencies between each other, encoded by the matrix  $A$ , we use the same approach of spline interpolation as before to smooth out the kinks of the intensities at the stitching positions. We compare the pure intensity differences at the stitching positions  $z_k$  and calculate the intensity spline coefficients  $a_I$  through solving the linear system  $A a_I(\mathbf{x}) = d_I(\mathbf{x})$  with the following inhomogeneities:

$$\begin{aligned} d_{I_1} &= 3 \left( \frac{I_2 \circ \vec{p}_1(1) - I_1}{h_1} \right), \quad d_{I_P} = 3 \left( \frac{I_P - I_{P-1} \circ \overleftarrow{p}_{P-1}(1)}{h_{P-1}} \right), \\ d_{I_k} &= 3 \left( \frac{I_k - I_{k-1} \circ \overleftarrow{p}_{k-1}(1)}{h_{k-1}} + \frac{I_{k+1} \circ \vec{p}_k(1) - I_k}{h_k} \right). \end{aligned} \quad (6)$$

To make sure we can compare the images in (6) we added the registration evaluation points 0 and 1 to  $\mathcal{S}$ , recalling that Baghaie et al. [1] only registered them at  $s = 1/2$ . Preliminary experiments showed that forcing  $I_k \circ \vec{p}_k(1/2)$  and  $I_{k+1} \circ \overleftarrow{p}_k(1/2)$  to be similar does not guarantee that  $I_k$  and  $I_{k+1} \circ \overleftarrow{p}_k(1)$  or  $I_k \circ \vec{p}_k(1)$  and  $I_{k+1}$  are similar. The registration point  $s = 1/2$  is still needed to optimize the third order polynomials. Other registration points can also be realized, in fact, every reslicing point could be used as a registration point. While solving  $A a_I(\mathbf{x}) = d_I(\mathbf{x})$ , we combine  $d_{I_k}(\mathbf{x})$  at different locations  $z_k$ . In order to properly register them, we again utilize the tridiagonal structure of the matrix  $A$ : The lower triangular matrix  $L$  of the Cholesky decomposition  $A = LL^T$  only has one secondary diagonal on the first off-diagonal. Therefore we can elegantly solve  $Le = d_I$  and  $L^T a_I = e$  with forward and backward substitution where for each subtraction the involved variables are warped to the mutual  $z$ -position:

$$\begin{cases} e_1 = d_{I_1}/L_{11} & (k=2, \dots, P) \\ e_k = (d_{I_k} - L_{k,k-1} (d_{I_{k-1}} \circ \overleftarrow{p}_{k-1}(1))) / L_{kk}, \end{cases} \begin{cases} a_{I_P} = e_P / L_{PP} & (k=P-1, \dots, 1) \\ a_{I_k} = (e_k - L_{k,k+1} (e_{k+1} \circ \vec{p}_k(1))) / L_{kk}. \end{cases} \quad (7)$$

Now we can interpolate slices between the images  $I_k$  and  $I_{k+1}$  at any  $z = z_k + s h_k$ , by replacing the linear combination (5) with (compare lower part

of Algorithm 1)

$$\begin{aligned}\Delta I &= I_{k+1} \circ \overleftarrow{p}_k(1-s) - I_k \circ \overrightarrow{p}_k(s) \\ I^{\text{interp}} &= I_k \circ \overrightarrow{p}_k(s) + s \Delta I + s(s-1) [s(h_k a_{I_{k+1}} \circ \overleftarrow{p}_k(1-s) - \Delta I) + \\ &\quad + (s-1)(h_k a_{I_k} \circ \overrightarrow{p}_k(s) - \Delta I)].\end{aligned}\quad (8)$$

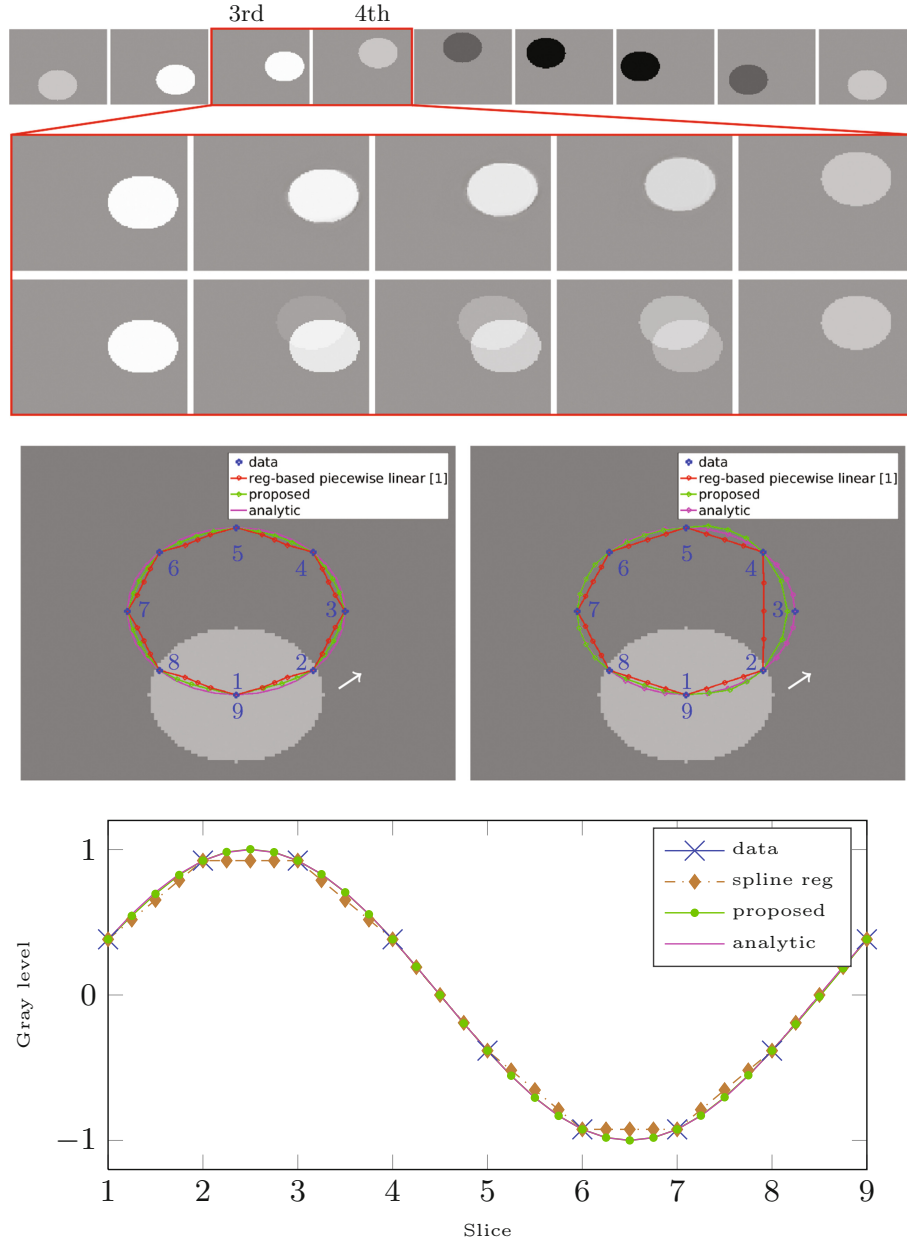
### 3 Experiments and Results

We implemented<sup>1</sup> the proposed algorithm in MATLAB. In all our experiments we linearly rescaled the images to values between 0 and 1, chose the parameters  $\lambda = 10$  and  $\tau = 10$  and stopped the optimization process when the mean square distance of the update is less than 0.1% for 10 consecutive iterations. For the in-slice image transformations  $\circ$  we used bilinear interpolation during the registration phase and bicubic interpolation for (6)–(8). The optimization time of the proposed method is comparable to the time of the linear registration-based method. The complexity of one registration iteration is  $\mathcal{O}(PMN \log(MN))$ , where DCT with  $\mathcal{O}(MN \log(MN))$  is the main contributor. Solving for the spline coefficients  $a$  with (4) is of order  $\mathcal{O}(PMN)$ .

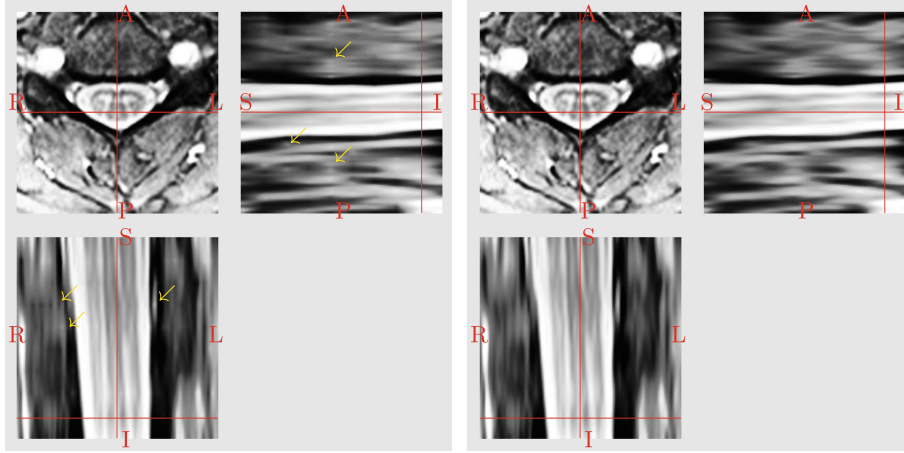
The artificial scenario involves two tests: shape interpolation and intensity interpolation along the correspondence point trajectories. *Shape interpolation* involved non-linear, ellipsoidal movement of a 2D ellipse, see Fig. 2. We sampled 9 slices counterclockwise every eighth from 6 o'clock to 6 o'clock. Comparing the results of the proposed algorithm in the 2nd and the 3rd row in Fig. 2, we clearly see the benefit of the proposed against an intensity interpolation without calculating displacement fields. The second advantage of the proposed method is the non-linear movement estimation: In the 4th row we see the motion trajectory of the center points of the interpolated ellipses. The algorithm in [1] estimates the object motion piecewise linearly while the proposed approach calculates a spline interpolated motion field, which results in a better approximation of the true solution. In the 4th row on the right in Fig. 2 we clearly see the advantage of the proposed method with a leave-one-slice-out test. The calculated center point of the ellipse in the third slice is close to the analytic solution. To test the *intensity interpolation* we colored the ellipses along the slices with a sinusoidal, as shown in Fig. 2. The proposed method performs better than the proposed structure interpolation with only linear intensity (5) (spline reg).

For a second scenario, we use 42 datasets of the human spinal cord along the neck captured with a slice-wise inversion recovery MR sequence (in-slice resolution  $0.67 \text{ mm} \times 0.67 \text{ mm}$ , slice distance 4 mm, slice thickness 8 mm) which we cropped for a centered view to a size of  $120 \times 120 \times 10$  voxels; see Fig. 3. With a leave-one-slice-out interpolation we quantitatively evaluated how well the left out slices can be interpolated. In particular, we compared linear and cubic interpolation without registration, a reimplement of the linear registration-based method of [1], the proposed interpolation with piecewise linear intensity

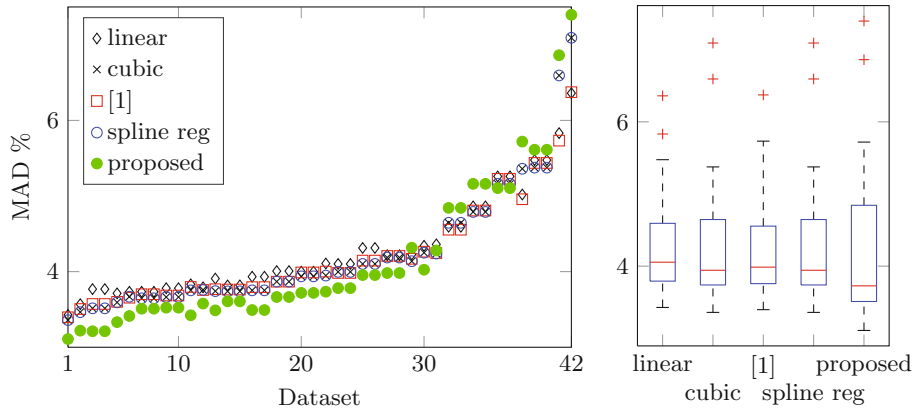
<sup>1</sup> <https://mathworks.com/matlabcentral/fileexchange/63907>.



**Fig. 2.** 1st row: 9 slices from left to right of a counterclockwise circling ellipse. 2nd row: proposed interpolation between the 3rd and 4th slice. Images are shown in larger size than in row 1 for visualization without upsampling. 3rd row: linear intensity interpolation. 4th row: movement line of the center points of the interpolated ellipses of the proposed method compared against the slice interpolation of [1]. 4th row, right: center points, when leaving the third slice out. 5th row: gray values of the center points of the proposed solution are nicely interpolated compared to the analytic solution.



**Fig. 3.** Exemplary validation dataset. Slice interpolated with the method of Baghaie et al. [1] (*left*) and the proposed method (*right*). Upsampled and histogram equalized for better visualization. Transverse cut (*upper left*), sagittal cut (*upper right*), coronal cut (*below*). In the sagittal and coronal cuts on the left kinks along the stitching positions are visible (yellow arrows). The proposed interpolation on the right is smooth. (Color figure online)



**Fig. 4.** Leave-one-slice-out interpolation of MR data. *Left*: MAD in percent of the interpolated slices compared to the corresponding left out slices of all datasets; datasets sorted for better visualization. *Right*: medians from left to right: 4.1, 3.9, 4.0, 3.9, 3.7.

changes (5) (spline reg), and the proposed method; see Fig. 4. As an evaluation metric we chose the mean absolute difference (MAD), comparing all the  $(P-2)$  interpolated slices of the  $(P-2)$  leave-one-slice-out interpolations to the left out slices of one dataset. In most datasets the registration process provided acceptable correspondences. In Fig. 4 we can see, that the proposed interpolation

reaches a higher accuracy as long as the correspondences are accurate. The proposed method interpolated the datasets with 0.5% less error in intensity than [1]. Errors in the correspondences can increase the spline intensity interpolation error which also explains the slightly higher variance. Nevertheless, the proposed method improves on the slice interpolation capability of the pure intensity-based methods and of the registration-based linear interpolation methods.

## 4 Discussion and Conclusion

In this paper we derived a new method for registration-based slice interpolation. The structural motions along the interpolating axis are spline interpolated, and along these motion trajectories the intensities are also spline interpolated. We presented a way to solve the problem of combining motion and intensity interpolation. We used piecewise polynomial interpolators of degree three between the slices and the additional free degrees of freedom to even out the kinks at the stitching positions. The method produces two times differentiable structure- and intensity interpolations. Provided by accurate point correspondences between the slices, the smooth interpolation can be a better approximation than the ones from linear registration-based interpolations and from intensity-based cubic interpolations. To better guarantee point correspondences, we would like to point out that the proposed approach can be used with other, more sophisticated image distances and regularizations. The proposed slice interpolation framework is flexible and can be extended in several aspects. For example, the polynomial interpolators can be transformed to include bending of the interpolating axis, in case the slices are not parallel to each other.

## References

1. Baghaie, A., Yu, Z.: An optimization method for slice interpolation of medical images [cs], February 2014. [arXiv:1402.0936](https://arxiv.org/abs/1402.0936)
2. Fischer, B., Modersitzki, J.: A unified approach to fast image registration and a new curvature based registration technique. *Linear Algebra Appl.* **380**, 107–124 (2004)
3. Frakes, D.H., Dasi, L.P., Pekkan, K., Kitajima, H.D., Sundareswaran, K., Yoganathan, A.P., Smith, M.J.T.: A new method for registration-based medical image interpolation. *IEEE Trans. Med. Imaging* **27**(3), 370–377 (2008)
4. Grevera, G.J., Udupa, J.K.: An objective comparison of 3-D image interpolation methods. *IEEE Trans. Med. Imaging* **17**(4), 642–652 (1998)
5. Hairer, E., Wanner, G.: *Introduction à l'analyse numérique* (2005) [Accessed: 28 July 2017]. <http://www.unige.ch/~hairer/poly/poly.pdf>
6. Leng, J., Xu, G., Zhang, Y.: Medical image interpolation based on multi-resolution registration. *Comput. Math. Appl.* **66**(1), 1–18 (2013)
7. Penney, G.P., Schnabel, J.A., Rueckert, D., Viergever, M.A., Niessen, W.J.: Registration-based interpolation. *IEEE Trans. Med. Imaging* **23**(7), 922–926 (2004)
8. Rueckert, D., Sonoda, L.I., Hayes, C., Hill, D.L., Leach, M.O., Hawkes, D.J.: Non-rigid registration using free-form deformations: application to breast MR images. *IEEE Trans. Med. Imaging* **18**(8), 712–721 (1999)

## 5 A Principled Approach to Combining Inversion Recovery Images

The published abstract presented in this chapter focuses on combining the 8 inversion images of the AMIRA sequence, described on page 79. As pointed out in [25, Tab. 3], averaging the first five inversion images reveals an enhanced contrast-to-noise-ratio (CNR) between GM and WM, while averaging the last five inversion images leads to a good CNR between CSF and WM. Averaging, however, linearly combines the involved images with equal weights only.

In this work, we modeled a compressed sensing-inspired energy functional to optimize non-uniform weights in the 8-dimensional space for optimal CNR 1) between GM and WM, and 2) between CSF and SC (with minimal variations inside the SC between GM and WM). To calculate the CNR values appearing in the energy, we used manual segmentations of 68 slices of 4 different healthy subjects. We point out that this optimization procedure is, however, a chicken and egg problem, since the manual segmentations were created on the uniform averages, which probably slightly influences the results.

The optimized weights are similar to the uniform weights and only slightly improve CNR. Therefore the uniform averages proposed in [25] can be interpreted as close-to-optimal. To assess the robustness of the presented approach, we performed "leave-all-other-subjects-out" cross-validations that show consistent results, independent of the subject for which the coefficients were optimized.

We would like to mention that one task of this approach was to optimize for optimal *CSF-SC* contrast, which is not the same as optimal *CSF-WM* contrast. Opposed to this goal, the original work [25] described a uniform averaging of the last 5 images which leads to a close-to-optimal *CSF-WM* contrast. However, averages with optimal *CSF-SC* contrast achieve lower variations inside the SC, which contributes to more robust CSF segmentation with the variational approach presented in Section 6.A. Experiments revealed that averages of the last 3 images have a significantly better *CSF-SC* contrast as compared to the averages of the last 5 inversion images. This observation is explained through the inversion images 4 and 5 providing more GM-WM contrast than images 6 to 8, cf. Fig. 2 in the abstract below. These reasons led us to compare the optimized *CSF-SC* contrast to the uniform averaging technique of the last 3 inversion images, instead of the last 5.

**Publication** The following abstract was presented as a poster at the 27<sup>th</sup> annual meeting of the International Society for Magnetic Resonance in Medicine (ISMRM) on the 18<sup>th</sup> of June 2018 in Paris, France. It was published as part of the conference proceedings [148]. The following text was reformatted for appearance in this thesis.

# A Principled Approach to Combining Inversion Recovery Images

Antal Horváth<sup>1</sup>, Matthias Weigel<sup>1,2</sup>, Christoph Jud<sup>1</sup>, Simon Pezold<sup>1</sup>,  
Charidimos Tsagkas<sup>3</sup>, Katrin Parmar<sup>3</sup>, Oliver Bieri<sup>1,2</sup>, and Philippe Cattin<sup>1</sup>

<sup>1</sup>Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland

<sup>2</sup>Division of Radiological Physics, Department of Radiology, University Hospital  
Basel, Basel, Switzerland

<sup>3</sup>Department of Neurology, University Hospital Basel, Basel, Switzerland

## 1 Synopsis

The averaged magnetization inversion recovery acquisitions (AMIRA) spinal cord imaging sequence acquires images of different inversion contrasts [2]. Despite the different contrasts the images can be combined to even enhance tissue contrast. We give a principled justification for such averaging. Using energy optimization, we describe how to automatically optimize the contrast-to-noise ratio between different tissues using a compressed sensing inspired approach [1]. We show that the uniform weights in the recently proposed AMIRA sequence are close to the optimum but can nevertheless still be improved. As an example we optimize the contrast-to-noise ratio between different compartments in the spinal cord.

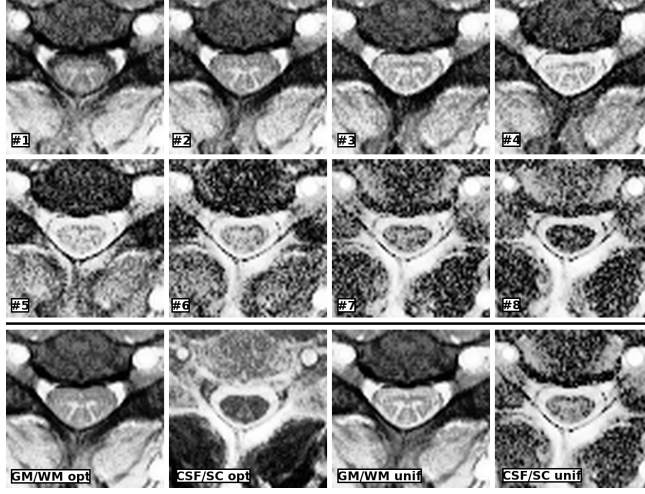
## 2 Purpose

Recently, a fast high-resolution spinal cord imaging sequence with averaged magnetization inversion recovery acquisitions (AMIRA) has been presented [2]. This sequence acquires 8 images at different inversion times, whose averages give better contrast-to-noise ratio. The original work proposes to average the first 5 images for optimal spinal cord gray matter (GM) and white matter (WM) contrast and to average the last 5 images for optimal WM and cerebrospinal fluid (CSF) contrast. Direct averaging of some of the acquired images is a promising but still simple concept. In the following we analyze whether a non-uniform averaging could improve the contrast. Using energy minimization, we calculate coefficients for optimal linear combinations of the 8 inversion images with best contrast-to-noise ratios.

## 3 Method

We present an energy functional, which is based on the idea of compressed sensing [1]. Given some images  $I_1, \dots, I_n$  with values between 0 and 1, we search





**Fig. 1.** *Top and middle row:* series of the 8 acquired inversion images of the AMIRA sequence in chronological acquisition order from top left to bottom right. *Bottom row:* CSF/SC and GM/WM averages of the AMIRA images combined with the optimized mean coefficient values of Figure 2 and the uniform 1:5 and 6:8-averages, respectively. All images zoomed in, upscaled, and histogram equalized.

for coefficients  $c = (c_1, \dots, c_n)$  where the linear combination  $I(c) := c_1 \cdot I_1 + \dots + c_n \cdot I_n$  yields an optimal contrast-to-noise ratio. Suppose we have manual segmentations of tissue  $A$  and  $B$  (subsets of the image domain) and suppose tissue  $A$  has bright intensities close to 1 and tissue  $B$  has dark intensities close to 0 (or at least darker than tissue  $A$ ). We define the energy

$$\begin{aligned}
 \mathcal{E}(c) := & \lambda_1 \frac{\sum_{x \in A} (I(c)(x) - 1)^2}{|A|} + \lambda_2 \frac{\sum_{x \in B} (I(c)(x))^2}{|B|} + \\
 & - \lambda_3 \left| E[I(c)(A)] - E[I(c)(B)] \right| + \lambda_4 V[I(c)(A)] + \lambda_5 V[I(c)(B)] + \\
 & - \lambda_6 \sum_{k=1}^n c_k \left| E[I_k(A)] - E[I_k(B)] \right| + \\
 & + \lambda_7 \sum_{k=1}^n c_k V[I_k(A)] + \lambda_8 \sum_{k=1}^n c_k V[I_k(B)] + \lambda_9 \mathcal{R}(c),
 \end{aligned} \tag{1}$$

where  $|A|$  is the cardinality of  $A$ ,  $I(A)$  is the set of all intensities of tissue  $A$  on an image  $I$ , and  $E[I(A)]$  and  $V[I(A)]$  are the mean and variance of the intensities of tissue  $A$  on an image  $I$ , respectively. The first two summands in  $\mathcal{E}$  encourage the linear combination  $I(c)$  to be as close to the segmentation as possible, the third negative term maximizes the contrast, and the fourth and fifth terms minimize noise. The terms 6, 7 and 8 in the third line also maximize

contrast and minimize noise, however on level of the individual inversion images. With the regularization

$$\mathcal{R}(c) := \left(1 - \sum_{k=1}^n |c_k|\right)^2, \quad (2)$$

we weakly constrain the coefficients' absolute values to sum up to 1. This way we omit arbitrarily large coefficients and we enable meaningful negative coefficients. A negative coefficient  $c_k$  flips the contrast of an inversion image  $I_k$  between tissue  $A$  and  $B$ .  $\mathcal{E}$  is minimized using BFGS.

We optimized the coefficients for a total of 68 slices on 4 different subjects at different vertebral heights (C2-C5) acquired with the AMIRA sequence. For each slice, we calculated linear coefficients  $c_{\text{CSF/SC}}$  for optimal CSF-SC contrast and another set of linear coefficients  $c_{\text{GM/WM}}$  for optimal GM-WM contrast, see Figure 2.

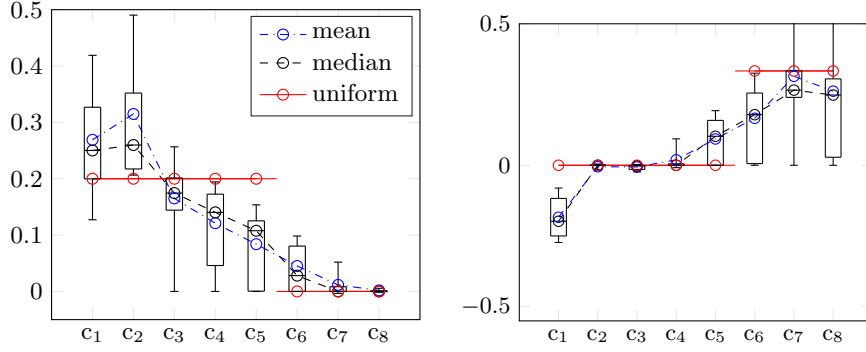
## 4 Results and Discussion

Figure 1 shows a representative series of the 8 images acquired by the AMIRA sequence and the suggested optimized CSF-SC and GM-WM contrast averages of an exemplary single axial slice located at level C2-C3. The optimized coefficients and their statistics are shown in Figure 2. The calculated optimal coefficients for GM-WM and for CSF-SC contrast are similar to uniform averaging, but they also address the weaker signal of inversion images at later inversion times, e.g. the drop of  $c_8$  in the right side of Figure 2. For the calculated CSF-SC coefficients, we may now include the first inversion image, which has darker CSF and brighter WM, by allowing negative coefficients  $c_k$ . A flipped image  $c_k \cdot I_k$  can be combined with the inversion images  $I_5$  to  $I_8$ , where the CSF is brighter than WM. We empirically have chosen  $\lambda_1, \dots, \lambda_9$  to be 100, 100, 1, 1, 1, 10, 10, 10, and 1000, respectively.

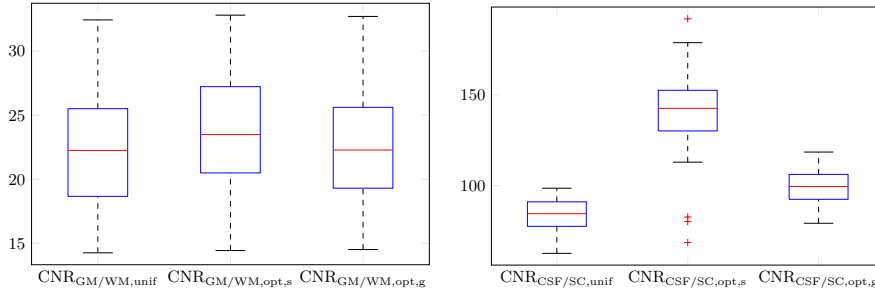
For quantitative comparison we calculated contrast-to-noise ratios  $\text{CNR}_{A/B} = \text{SNR}_A - \text{SNR}_B$  and signal to noise ratios  $\text{SNR}_A = E[I(A)]/SD[I(C)]$ . We estimated the standard deviation of noise  $SD[I(C)]$  on a homogeneous part  $C$  of the background. Figure 3 shows a quantitative comparison of CNR between uniform and proposed averaging. Leave-one-subject-out results are shown in Table 1.

## 5 Conclusion

With the proposed method we analyze the uniform averaging technique of the inversion images of the AMIRA sequence. The found calculated coefficients are close to the uniform coefficients and the contrast-to-noise ratios can slightly be improved. For CSF-SC contrast a notable improvement was possible. With the developed approach, we could justify the decisions made in [2] on a quantitative basis. The optimization coefficients can be chosen for the needs, e.g. prioritizing less noise or better contrast.



**Fig. 2.** Optimized linear coefficients  $c_{\text{GM}/\text{WM}}$  for optimal GM-WM contrast (left), and  $c_{\text{CSF}/\text{SC}}$  for optimal CSF-SC contrast (right), in comparison to the uniform 1:5-averaging (left) and 6:8-averaging (right). In solid red, the uniform averaging; in black box plots of the optimized coefficients of all 68 slices with median, lower and upper quartile box, and 10/90-th percentile whiskers; and in dash-dotted blue, the mean values are shown.



**Fig. 3.** Contrast-to-noise ratios, comparing the uniform averaging (unif) and the optimized coefficients (opt). The optimized coefficients are evaluated for two cases: the optimal case (s), where the optimal coefficients  $c$  of each slice were used to calculate the linear combination  $I(c)$  for each slice; and the global case (g), where the blue mean value coefficients of the optimal  $c_{\text{GM}/\text{WM}}$  and  $c_{\text{CSF}/\text{SC}}$  in Figure 2, respectively, were used for all slices. *Left:* CNR comparison for CSF-SC contrast enhancement; *right:* comparison for GM-WM contrast enhancement.

Mean values from left to right: 83, 139, 97, 22, 24, 23.

Left out subject	$\text{CNR}_{\text{CSF}/\text{SC}, \text{unif}}$	$\text{CNR}_{\text{CSF}/\text{SC}, \text{opt}}$	$\Delta \text{CNR}_{\text{CSF}/\text{SC}, \text{opt-unif}}$	$\text{CNR}_{\text{GM}/\text{WM}, \text{unif}}$	$\text{CNR}_{\text{GM}/\text{WM}, \text{opt}}$	$\Delta \text{CNR}_{\text{GM}/\text{WM}, \text{opt-unif}}$
1	81.98	91.87	9.88	24.99	25.36	0.37
2	90.26	101.57	11.31	24.66	25.64	0.98
3	87.37	102.44	15.07	22.71	23.65	0.94
4	72.64	88.15	15.51	17.37	18.07	0.70

**Table 1.** Leave-one-subject-out cross-validation: CNR mean values over all slices of the subjects that were not left out are shown for uniform averaging and optimized averaging. Each cross-validation uses the coefficients  $c_{\text{GM}/\text{WM}}$  and  $c_{\text{CSF}/\text{SC}}$  averaged over the optimal coefficients of the slices of the left out subject. Positive improvements with the proposed method are possible for both contrasts.

## 6 Acknowledgements

This work was supported by the Swiss National Science Foundation (SNF grant No. SNF 320030-156860/1).

## References

1. Donoho, D.L.: Compressed sensing. *IEEE Transactions on Information Theory* **52**(4), 1289–1306 (Apr 2006). <https://doi.org/10.1109/TIT.2006.871582>
2. Weigel, M., Bieri, O.: Spinal cord imaging using averaged magnetization inversion recovery acquisitions. *Magnetic Resonance in Medicine* **79**(4), 1870–1881 (Apr 2018)

## 6 Variational Segmentation of the White and Gray Matter in the Spinal Cord Using a Shape Prior

The publication presented in this chapter describes an automatic segmentation method for axial MR images of the SC on C3 level acquired with the MOLLI sequence, described on Page 78. After reviewing and experimenting with the various segmentation frameworks presented in Section 2.2.8, we chose the continuous cut model (2.97) adapted to the use of appearance models according to the idea of Overgaard et al. [22], see Equation (2.107). We solved the model with the multiplier-based max-flow algorithm of Yuan et al., cf. Equation (2.103).

We separated the algorithm into two segmentation parts: in a first part, the method locates and segments the ring-shaped CSF and subsequently extracts the SC from it; in a second part, the algorithm differentiates between GM and WM using the SC as a mask. Both parts essentially use the continuous cut framework (2.107) but with individually adapted parameters. During segmentation, the method automatically estimates the involved objects' intensity distributions, which are used for the terminal capacities.

The goal of this paper was to describe a first working experimental setup as complete and as reproducible as possible. While we manually tweaked the involved parameters by observing the algorithm's iterations using only a few validation images, we tested the proposed algorithm on 16 axial images at C3 level acquired with the MOLLI sequence. We achieved robust segmentation results that are on par with the manual segmentations.

We, however, did not provide an ablation analysis on the impact of the different mathematical ingredients. We believe that the algorithm's performance was mainly limited by the low resolution of the raw data ( $0.4\text{ mm} \times 0.4\text{ mm}$ ), which results in a SC diameter of approximately 20 pixels. We did not resample the data to higher resolutions before automatic segmentation. Therefore, the calculated similarity metrics are prone to numerical imprecision and thus not sensitive enough for the true performance. As a consequence, we considerably improved the proposed algorithm after this publication by adapting the parameters for use with higher resolutions and by harmonizing the two segmentation parts. These adaptations are described in this chapter's Appendix 6.A and led to the publication in Chapter 7,

We point out that the argument of "non-differentiability of  $\|\cdot\|$  at the origin" on page 29 of this publication, which supported the use of the divergence on the flow-variable instead of the gradient on the primal variable, holds for the  $L^1$ -norm but does not hold


when using the Euclidean  $L^2$ -norm. Nevertheless, the second argument of avoiding derivative calculations along jump-parts is still valid.

**Publication** The following paper was presented at the workshop *Computational Methods and Clinical Applications for Spine Imaging* (CSI 2016) in conjunction with the *19<sup>th</sup> International Conference on Medical Image Computing and Computer Assisted Intervention* (MICCAI) on the 17<sup>th</sup> of October 2016 in Athens, Greece. It was published<sup>1</sup> as part of the workshop proceedings [149].

---

<sup>1</sup>[http://dx.doi.org/10.1007/978-3-319-55050-3\\_3](http://dx.doi.org/10.1007/978-3-319-55050-3_3) (last accessed on Sept. 10, 2019)

# Variational Segmentation of the White and Gray Matter in the Spinal Cord Using a Shape Prior

Antal Horváth<sup>1</sup>() , Simon Pezold<sup>1</sup>, Matthias Weigel<sup>1,2</sup>, Katrin Parmar<sup>3</sup>,  
Oliver Bieri<sup>1,2</sup>, and Philippe Cattin<sup>1</sup>

<sup>1</sup> Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland  
[antal.horvath@unibas.ch](mailto:antal.horvath@unibas.ch)

<sup>2</sup> Radiological Physics, Clinics of Radiology,  
University Hospital Basel, Basel, Switzerland

<sup>3</sup> Department of Neurology, University Hospital Basel, Basel, Switzerland

**Abstract.** Segmenting the inner structure of the spinal cord on magnetic resonance (MR) images is difficult because of poor contrast between white and gray matter (WM/GM). We present a variational formulation to automatically detect cerebrospinal fluid and WM/GM. The segmentation results are obtained by continuous cuts combined with a shape prior. Intensity-based segmentation guarantees high accuracy while the shape prior aims at precision. We tested the algorithm on a set of MR images with visual WM/GM contrast and evaluated it w.r.t. manual GM segmentations. The automated GM segmentations are on a par with the manual results.

## 1 Introduction

Numerous neurological diseases manifest not only in the brain but also in the spinal cord (SC). Accurate SC segmentation recently gained increasing attention. Measuring the SC cross-sectional area on magnetic resonance (MR) images has shown to be a good quantitative measure to study diseases of the central nervous system like multiple sclerosis (MS). MS shows strong influence on the SC, which manifests e.g. in atrophy and lesion formation [7]. Spinal cord atrophy is occurring early in the disease progress and was shown to correlate very well with the clinical evaluation (EDSS) of the patient i.e. with MS disability progression [7]. Especially the SC gray matter area was shown to correlate strongly with MS disability [11].

Delineating white (WM) and gray matter (GM) and measuring their areas or volumes in-vivo is challenging because of their fine structure, poor WM/GM imaging contrast, limited practical MR resolution, and inter- and inpatient variability of the captured images and of the SCs themselves. On top of that, high intra- and interobserver variability in manual segmentations make further statistical evaluations difficult.

To overcome these challenges and to further deepen the knowledge about the GM/WM atrophy in the SC an automatic quantitative segmentation method is required that has high accuracy as well as precision.

Given an MR image with optimal contrast and signal-to-noise ratio, an intensity-based segmentation method is perfectly suited for that job. The more noise is involved the more we need to regularize the method. To handle partial volume effects of captured fine structures, even shape priors may be necessary.

Yiannakas et al. [16] show the feasibility of segmenting WM/GM in the SC. Tang et al. [12] use a Bayesian three-class classifier, and Asman et al. [1] use groupwise multi-atlas segmentation to discriminate WM/GM slice-wise. Taso et al. [13] construct atlases for cerebrospinal fluid (CSF), WM/GM and propose atlas based 3D segmentation and classification methods. De Leener et al. [5] provide a more comprehensive review of the available GM segmentation methods. However, no standard method has yet been established, which motivates the search for alternative approaches.

In this paper, we propose an automatic variational segmentation method that can locate CSF and segment the WM/GM of the spinal cord. We describe our data in Sect. 2, introduce our generic models in Sect. 3, show the results in Sect. 4, and discuss them in Sect. 5.

## 2 Data

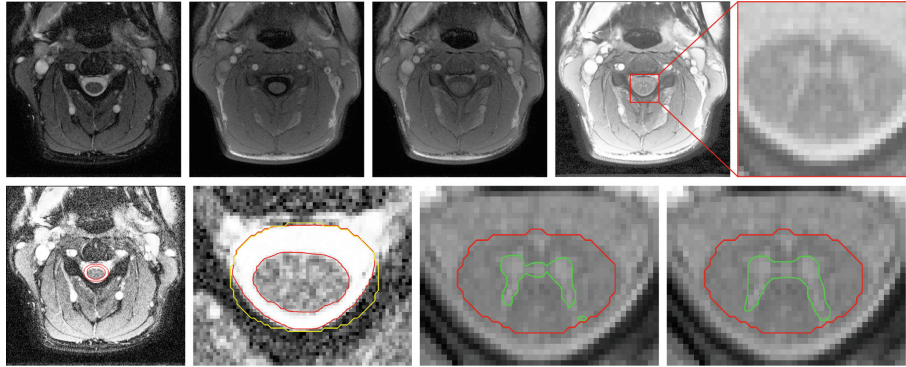
For this paper 10 volunteers (6 male, 4 female) were scanned with an experimental MR sequence (approved by the local ethics review board). In total we acquired 16 axial cross-sectional sets of images on C3 level with acceptable WM/GM contrast using a 2D-MOLLI sequence [15] with  $0.4 \times 0.4 \text{ mm}^2$  in-plane resolution and 8 mm slice thickness. The MOLLI sequence acquires altogether 11 aligned images in each image set with different inversion times TI per slice [15]. The first image of each image set has good CSF contrast and the average of each set has good WM/GM contrast. In Fig. 1 we see the first three images and the mean image of one set. Figure 3 shows more mean images of different contrast quality and of different subjects.

Two experienced raters segmented the 16 images manually for GM. Rater 1 rated two times at different days with different techniques. In one technique images in the original resolution were segmented pixel-wise, see Fig. 3. In the other, 10-fold upsampled cubic interpolated images were segmented whose masks were downsampled again afterwards. The second technique resulted in a grayscale segmentation. Rater 2 used the same up- and downsampling technique.

## 3 Method

In this paper we present a variational approach that segments CSF, WM, GM, and background given a set of MR images of the same slice. Because binary segmentation algorithms are more robust than multi-labeling algorithms and because of the special situation of the MOLLI sequence, where the first image has good CSF contrast, we split our fully automatic approach into two steps: a CSF segmentation step and a WM/GM segmentation step. The second step





**Fig. 1.** *Upper row:* first three images of the MOLLI sequence of a cross-sectional neck scan on C3 level; histogram-equalized mean image and its zoom depict CSF, WM, GM. *Lower row, segmentation steps:* CSF contrast image with ellipsoidal prior; zoomed view on the CSF segmentation (red) and mask (yellow); GM segmentation (green) and the boundary towards CSF (red) before and after shape regularization. (Color figure online)

makes use of the previously labeled CSF, where we extract the interior of the ring-shaped CSF and use it as a mask for the WM/GM discrimination, see Fig. 1.

In recent years a tendency towards relaxed convex variational formulations can be observed because their solutions all enjoy to have the optimal score. The motivation of using continuous cuts [17] lies in the mathematical beauty and in the simple algorithmic implementation. The model is convex and finds intended segmentations robustly, independent of any specific algorithm initialization. Its dual formulation, also called continuous max flow, has analogies to graph cut [3], but enables subpixel accuracy and has less metrication errors. We added different additional features to the continuous cut formulation: anisotropic total variation (ATV) [8], pose invariant shape priors [4, 9], an additive Bhattacharyya coefficient (BC) [14], and prior boundary curvature dependent capacities.

### 3.1 Mathematical Ingredients

Before we describe the CSF and WM/GM segmentation steps, we first introduce continuous cuts [17] and additional energies, which we use in both steps.

The most basic segmentation method is intensity thresholding, where the pixels are divided into two categories: those with intensity values lower and respectively higher than a certain threshold. Because of the presence of noise in high resolution MR images, before thresholding, first an approximation of the image has to be calculated where the noise is reduced. In the literature this problem can be modeled with the Mumford-Shah functional [2]. We make use of a generalized special case of the piecewise-constant Mumford-Shah problem

$$\operatorname{argmin}_{\mathcal{O} \subset \Omega} \int_{\mathcal{O}} C_t(x) dx + \int_{\Omega \setminus \mathcal{O}} C_s(x) dx + \operatorname{TV}_C(\partial \mathcal{O}), \quad (1)$$

which is also a generalized version of the Chan-Vese model and can be seen as a generic segmentation model [17]: Inside the image domain  $\Omega$  the object  $\mathcal{O}$  and the background  $\mathcal{B} := \Omega \setminus \mathcal{O}$  shall be found, where  $C_t, C_s$  and  $C$  are the model parameters, here called capacity functions. The name *capacity* and the indices of  $C_s$  and  $C_t$  find their roots in graph cuts and stand for source and target.  $C_t$  has to be low on the object  $\mathcal{O}$ , and  $C_s$  has to be low in the background.

For an associated algorithm to be automatic, it is necessary to estimate proper capacity functions automatically. In this task, mean intensity differences turned out to be well-suited for the terminal capacities  $C_s$  and  $C_t$ . For non-terminal capacities  $C$  we chose negative exponential image gradients. The weighted total variation  $\text{TV}_C$  as an object boundary length regularizer can be adapted anisotropically to the image structures when introducing a Riemannian metric tensor  $A$  [8], replacing  $\text{TV}_C$  by  $\text{ATV}_C$ .

To find a convex version of the above functional (1) one introduces a relaxed labeling  $u : \Omega \rightarrow [0, 1]$ , such that  $x$  is in  $\mathcal{O}$  if  $u(x)$  is close to 1 and  $x$  is in  $\mathcal{B}$  if  $u(x)$  is close to 0. This way we can write down the continuous cut with ATV

$$\operatorname{argmin}_{u: \Omega \rightarrow [0,1]} \int_{\Omega} C_t(x)u(x) + C_s(x)(1 - u(x)) + C(x) \|\nabla u(x)\|_A dx, \quad (2)$$

where  $\|\nabla u(x)\|_A = \sqrt{\nabla u(x)^T A(x) \nabla u(x)} = \|S(x)^T \nabla u(x)\|_2$ , and  $A = SS^T$  is a strongly positive definite, matrix valued function [8]. Olsson et al. [8] proofed that a coarea formula for ATV holds and thus a thresholding theorem exists for solutions of (2). This means that for a minimizer  $u^*$  of (2) and any threshold  $\theta$  in  $]0, 1]$  the thresholded superlevel set  $\mathbb{1}_{u^* \geq \theta}$  is again a minimizer of (2).

Algorithms that minimize (2) itself may struggle with the non-differentiability of  $\|\nabla u(x)\|_2$  twofold: the non-differentiability of  $\|\cdot\|_2$  at the origin and the calculation of  $\nabla u$  along jump-parts. A nice work-around is provided by the primal-dual formulation of (2). As proposed in [10], adding an augmented Lagrangian and calculating the variational derivatives results in a very short algorithm: We iterate a valid initialization  $(p_s^0, p_t^0, p^0, u^0)$  with

$$\begin{aligned} p_s^{k+1} &= \min \left( (1-u^k)/c + \operatorname{div} S p^k + p_t^k, C_s \right), \\ p_t^{k+1} &= \min \left( u^k/c - \operatorname{div} S p^k + p_s^{k+1}, C_t \right), \\ p^{k+1} &= \mathcal{P} \left( p^k + \gamma S^T \nabla (\operatorname{div} S p^k - p_s^{k+1} + p_t^{k+1} - u^k/c) \right), \\ u^{k+1} &= u^k - c \left( \operatorname{div} S p^{k+1} - p_s^{k+1} + p_t^{k+1} \right), \end{aligned} \quad (3)$$

where  $\mathcal{P}(p(x)) = \operatorname{sign}(p(x)) \min(|p(x)|, C(x))$ . We use  $c = 0.3$  and  $\gamma = 0.16$ .

Up to here the model is convex and fulfills a thresholding theorem. Now we vary the capacity functions and make them dependent on  $u$ , thus in general we lose the mathematical global optimality property. In practice, as long as the capacity functions do not change too fast, they converge with  $u$ . We lose convexity anyway, as we include additional energy terms like BC [14], and a mean squared difference to a shape prior [4, 9], see next sections. Of course, the mathematical properties of the continuous cut in practice still help to balance

out the local properties of the additional energies, as long as they are weighted appropriately. In turn, BC helps out the segmentation process to stick to the local intensity structure while the shape prior term includes prior knowledge.

**Pose Invariant Shape Prior.** Given a grayscale image  $I : \Omega \rightarrow [0, 1]$ , a shape prior  $f_0 : \Omega \rightarrow [0, 1]$ , and a relaxed labeling  $u$ , we introduce the effective image

$$I_{\text{eff}} = \lambda I_{\text{model}} + \mu I_{\text{prior}}, \quad I_{\text{model}} = c_0 (1 - u) + c_1 u, \quad I_{\text{prior}} = b_0 (1 - f) + b_1 f, \quad (4)$$

with  $c_1$  and  $c_0$  the mean intensities of the background and foreground,  $f = f_0 \circ T_p : \Omega \rightarrow [0, 1]$  a rigidly transformed version of  $f_0$ , and  $b_1$  and  $b_0$  the mean model intensities on the rigidly transformed shape prior area and its complement [9]. The idea is to minimize

$$\int_{\Omega} \lambda (I - I_{\text{model}})^2 + \mu (I_{\text{model}} - I_{\text{prior}})^2 + C \|\nabla u\|_A \, dx, \quad (5)$$

where we segment and force the segmented result to be close to the prior by minimizing the mean squared distance to the relaxed piecewise constant approximation  $I_{\text{model}}$ . By factoring out with remainder, we see that minimizing the latter is equivalent to minimizing

$$E(u, c, f, b) = \int_{\Omega} (I_{\text{eff}} - c_1)^2 u + (I_{\text{eff}} - c_0)^2 (1 - u) + C \|\nabla u\|_A \, dx, \quad (6)$$

which is in the form of a continuous cut, thus can be optimized by (3). In practice, we replace  $(I_{\text{eff}}(x) - c_1)^2$  and  $(I_{\text{eff}}(x) - c_0)^2$  with functions  $C_t(x)$  and  $C_s(x)$  that stay close to the idea of (4): The square function is replaced with the absolute function and  $c_0$  and  $c_1$  are varied slightly, see below.

Following [9], the rigid coordinates  $p = (a, b, \theta, \exp \sigma)$  of the prior  $f$  can be iterated by a gradient descent through

$$\begin{aligned} a^{n+1} &= a^n - \iota \frac{\langle f - u, -\partial_{x_1} f \rangle}{\|f\|^2}, & \theta^{n+1} &= \theta^n - \iota \frac{\langle f - u, -\nabla f^T J(x - (a, b)^{n+1}) \rangle}{\|x - (a, b) \|\nabla f\|^2}, \\ b^{n+1} &= b^n - \iota \frac{\langle f - u, -\partial_{x_2} f \rangle}{\|f\|^2}, & \sigma^{n+1} &= \sigma^n - \iota \frac{\langle f - u, -\nabla f^T (x - (a, b)^{n+1}) \rangle}{\|x - (a, b) \|\nabla f\|^2}, \end{aligned} \quad (7)$$

where  $J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ , and  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  denote the  $L^2$  scalar product and norm. The denominators, obtained through the metric on the Lie group of the transformed priors, can be seen as automatic step size controllers. We use the stepsize  $\iota = 1$ .

**Bhattacharyya Coefficient.** BC is a measure of how different two densities are. The goal of intensity-based segmentation can be described as finding regions with maximally distinct histograms. Given the two histogram densities  $f_{\mathcal{O}}$  and  $f_{\mathcal{B}}$  of the object and the background with values in  $Z$ , their BC is given by

$$BC(f_{\mathcal{O}}, f_{\mathcal{B}}) = \int_Z \sqrt{f_{\mathcal{O}}(z) f_{\mathcal{B}}(z)} \, dz. \quad (8)$$

Following Wang et al. [14] we calculate the variational derivative  $\frac{\delta}{\delta u}$  of BC:

$$\begin{aligned} \frac{\delta BC}{\delta u}(x) = & \frac{1}{2} \delta(u(x) - \theta) \int_Z \sqrt{f_O(z) f_B(z)} \left( \frac{1}{A_B} - \frac{1}{A_O} \right) + \\ & + \delta(I(x) - z) \left( \frac{\sqrt{f_O(z) f_B(z)}}{f_B(z)} \frac{1}{A_O} - \frac{\sqrt{f_O(z) f_B(z)}}{f_O(z)} \frac{1}{A_B} \right) dz, \end{aligned} \quad (9)$$

where the histograms and the areas  $A_O$  and  $A_B$  are represented through thresholded  $u$  segmentations, and  $\delta(\cdot)$  stands for the Dirac-delta distribution. A more global variant of the BC gradient could be calculated by plugging in  $u$  directly, omitting the factor  $\delta(u(x) - \theta)$  in (9). But since we only want to influence the continuous cut close to the boundary of the thresholded solution, we chose this version. In the algorithm we set  $\theta$  to 0.5 and use a standard arctan approximation with  $\epsilon = 1$  for the  $\delta$  function outside the integral. Given a discrete image, we already calculate discrete histogram densities which also discretizes  $\int_Z dz$  to  $\sum_{z \in Z}$  and  $\delta(I(x) - z)$  to the Kronecker delta  $\delta_{I(x), z}$ .

Wang et al. [14] combined the level set representation of the Chan-Vese energy with BC. We now combine the continuous cut energy for shape priors with BC, weighted with a factor  $\nu$ :

$$\min_{\Omega} \int_{\Omega} C_t(u, f, I) u + C_s(u, f, I) (1 - u) + C(f, I) \|\nabla u\|_A dx + \nu BC(u). \quad (10)$$

With this external energy, the update rule for  $u$  in (3) changes to

$$u^{k+1} = \mathcal{P}_u \left( u^k - c \left( \operatorname{div} S p^{k+1} - p_s^{k+1} + p_t^{k+1} + \nu \frac{\delta BC}{\delta u} \right) \right). \quad (11)$$

The local behavior of the variational derivative of BC is guided by the convex property of the continuous cut. BC influences twofold: For the terminal capacities, aside mean intensity differences, also object and background histogram densities can be used. This choice, in practice, turned out unsuited for automatic segmentation when the object is not already well initialized. Using BC we can incorporate the histogram densities simultaneously with mean intensity differences. The second benefit is: While continuous cut (6) is computing on the effective image  $I_{\text{eff}}$ , BC is dealing with the original image intensities. BC makes the choice of the capacities less sensitive and balances the influence of the prior in the segmentation process.

**Statistical Appearance Model.** We do not only want the shape prior  $f_0$  to be a statistical mean shape, but we also want the shape prior to adapt to the actual segmentation. Cremers et al. [4] use an appearance model for this task, because projection and backprojection are low cost compared to a displacement field model, where non-rigid registration is involved. They also showed that optimizing the model parameter inside the set of meaningful parameters is convex.

Let us model the shape priors by a Gaussian distribution with mean  $f_0$  and covariance  $C$  and let  $C = V\Lambda^2V^T$  be its eigendecomposition. Following [4], we use our continuous cut energy (10) for an appearance model realization optimization together with an appearance prior regularizer  $\xi \alpha^T \Lambda \alpha$ :

$$\min_{\alpha} \int_{\Omega} C_t f + C_s(1 - f) + C \|\nabla f\|_A dx + \nu BC(u) + \xi \alpha^T \Lambda \alpha, \quad (12)$$

with the backprojection  $f = f(\alpha, p) = \min(\max(f_0(p) + V(p)\Lambda\alpha, 0), 1)$ , where  $f_0(p)$  and  $V(p)$  are the rigidly transformed mean GM and eigenshapes at the rigid coordinates  $p$ .

In (12) compared to (10)  $f(\alpha, p)$  takes the role of  $u$  and since they have to be similar, we initialize the model parameter  $\alpha$  with the projection of  $u^{k_0}$ , where  $u^{k_0}$  is an acceptable solution of the model (10). We iterate the model parameter simultaneously with projected  $u$ -updates from (11):

$$\begin{aligned} \alpha^0 &= \Lambda^{-1}V(p)^T(u^{k_0} - f_0(p)), \\ \alpha^{n+1} &= \alpha^n - \Delta t_{\alpha} \Lambda^{-1}V(p)^T \left( \operatorname{div} Sp^{k+1} - p_s^{k+1} + p_t^{k+1} + \nu \frac{\delta BC}{\delta u} \right) - 2\xi \Lambda \alpha^n. \end{aligned} \quad (13)$$

The idea behind this algorithm is: projecting  $u$  in every iteration into the model space and using its approximation  $f$  as the prior. But since we can choose the stepsize  $\Delta t_{\alpha}$ , the prior and the segmented structure are, up to the point where  $f$  is initialized, uncoupled from each other.

**Anisotropic Total Variation.** We use the structure tensor of the image to design the Riemannian metric matrix field  $A : \Omega \rightarrow \mathbb{R}^{2 \times 2}$  [6]. In  $\int_{\Omega} \|\nabla u\|_A dx$  we want the integrand to be small at high image gradients. Thus, for each  $x$  in  $\Omega$  the eigenvalue of  $A(x)$  in image gradient direction, which is parallel to  $\nabla u$ , has to be small. Let  $B(x)\hat{A}(x)B(x)^T$  be the eigendecomposition of the symmetric structure tensor at point  $x$  with decreasing order of eigenvalues and let  $\lambda(x)$  denote the larger eigenvalue. We use the gamma-transformed normalized negative transform of all image gradient magnitudes  $(1 - N(\lambda)(x))^2$ , to scale the image gradient direction, such that  $\nabla u$  gets less penalized when the image gradient is stronger. We define  $S(x) = B(x) \operatorname{diag}((1 - N(\lambda)(x))^2 + \epsilon, 1)$ , where we add  $\epsilon > 0$  to guarantee strong positiveness of the matrix field. When  $\lambda$  gets small,  $(1 - N(\lambda)(x))^2$  tends quadratically to 1, thus  $A$  approaches the isotropic identity matrix in regions with weak edges.

**Notations and Definitions.** For super- and lowerlevel sets we set  $\{u \sim \theta\} = \{x \in \Omega \mid u(x) \sim \theta\}$ , where  $\sim$  stands for a relation ( $=, >, <, \dots$ ). We denote the normalization of a real-valued function  $v : \Omega \rightarrow \mathbb{R}$  as the function  $N(v) : \Omega \rightarrow [0, 1]$  with  $N(v)(x) = (v(x) - \min_s v(s)) / (\max_s v(s) - \min_s v(s))$ , and the normalized gamma transform  $\Gamma(v) : \Omega \times \mathbb{R}_+ \rightarrow [0, 1]$  with  $\Gamma(v)(x, \gamma) = (N(v)(x))^{\gamma}$ . We denote the

shortest Euclidean distance between  $x$  and the boundary of  $\{f = 1\}$  by  $d(x, f)$ .  $\mathcal{F}_{\text{median}}$  is a median filter of size  $3 \times 3$ . For all the terms that need to be regularized we add a constant  $\epsilon = 10^{-5}$ . Tolerances have been chosen, such that the  $L^1$  norm of an update is smaller than 1 pixel for 10 consecutive iterations.

### 3.2 CSF Segmentation

Given a histogram-equalized MR image  $I : \Omega \subset \mathbb{R}^2 \rightarrow [0, 1]$  with good CSF contrast we introduce an ellipsoidal shape prior  $f_0 : \Omega \rightarrow \{0, 1\}$ , shown in Fig. 2. We use the segmentation model (10) with  $\lambda = \mu = 0.5$ ,  $\nu = 0$ , and

A while  $f_0$  is being registered:

$$\begin{aligned} C_s(x) &= \mathcal{F}_{\text{median}}(\Gamma(|I_{\text{eff}}(x) - \text{mean}(I(\{f > 0.5\}))|, 3/2)), \\ C_t(x) &= \mathcal{F}_{\text{median}}(\Gamma(|I_{\text{eff}}(x) - \text{mean}(I(\{f \leq 0.5\}))|, 2/3)), \\ C(x) &= 1/10 \log(1 + d(x, f)) + \epsilon, \end{aligned}$$

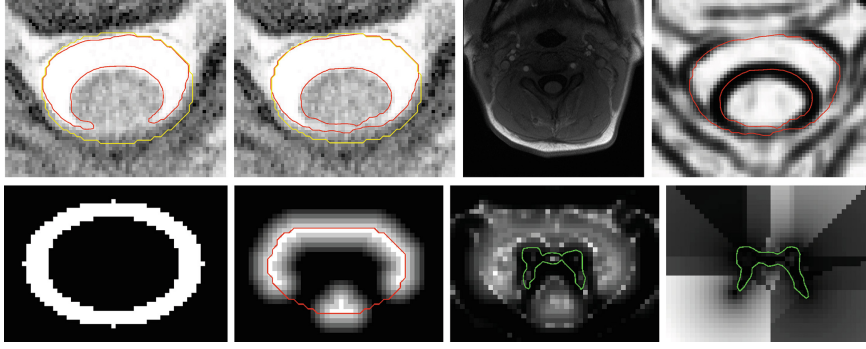
B after  $f_0$  has been registered:

$$\begin{aligned} C_s(x) &= \mathcal{F}_{\text{median}}\left(\frac{\max(\log(f_{\mathcal{O}} + \epsilon) - \log(f_{\mathcal{B}} + \epsilon), 0)}{\log(1 + \epsilon) - \log \epsilon}\right), \\ C_t(x) &= \mathcal{F}_{\text{median}}\left(\frac{\max(\log(f_{\mathcal{B}} + \epsilon) - \log(f_{\mathcal{O}} + \epsilon), 0)}{\log(1 + \epsilon) - \log \epsilon}\right), \\ C(x) &= \gamma_1 \exp(-\gamma_2 \|\nabla I_{\text{assist}}(x)\|). \end{aligned}$$

In A, the translational initial point of  $f$  is the center of mass of the neck, calculated through Otsu's thresholding, which is then iterated through (7). The terminal capacity functions are gamma-transformed with experimentally chosen exponents  $3/2$  and  $2/3$  to enhance the contrast between object and background. In B, when  $f$  has been registered to the CSF position, we calculate a mask and use it as the new image domain – marked yellow in Fig. 2. We use the convex hull of the ellipsoid and dilate it with a box-shaped structuring element of size 4, to make sure that the segmentation does not leak into surrounding bright intensity areas, when we switch to other capacity functions. The non-terminal capacity  $C$  is then changed to a negative exponential of gradient magnitudes of an assisting image  $I_{\text{assist}}$ . The assisting image at a bigger TI relaxation time, where the CSF is black and the spinal tract is brighter, helps in case the captured CSF of the subject is not ring-shaped. CSF can be pressed away though the gravity of the spinal tract when lying on the back, compare Fig. 2.

### 3.3 WM/GM Segmentation

Given an image  $I : \Omega \cap M \subset \mathbb{R}^2 \rightarrow [0, 1]$  with good WM/GM contrast, the inner part of the segmented CSF as a mask  $M$  and a statistical model for GM, we use (10) and (12) with  $\lambda = \mu = 0.5$ ,  $\nu = 5$ ,  $\xi = 0$ ,  $\Delta t_{\alpha} = 0.2$  (13), and



**Fig. 2.** *Upper row:* CSF segmentation without adapted non-terminal capacity; CSF segmentation with adapted non-terminal capacity; assisting image  $I_{\text{assist}}$ ; adapted non-terminal capacity for CSF – low values along the posterior CSF result in weak length regularization. *Lower row:* Ellipsoidal-shaped CSF prior (major and minor axis roughly 10 mm and 7 mm); special mask  $M_{\text{special}}$  with high values where GM cannot be; negative exponential of gradient magnitudes combined with  $M_{\text{special}}$ ; negative transformed GM curvature multiplied by distance map. (Color figure online)

**A** during initial segmentation and rigid registration:

$$\begin{aligned} C_s(x) &= \mathcal{F}_{\text{median}}(\Gamma(|I(x) - \text{mean}(I(M))|, 3/2)), \\ C_t(x) &= \mathcal{F}_{\text{median}}(\Gamma(|I_{\text{eff}}(x) - 0|, 2/3)), \\ C(x) &= g(x) + \frac{\text{median}(g) + \max(g)}{2} \cdot M_{\text{special}}, \quad g(x) = \gamma_1 \exp(-\gamma_2 \|\nabla I(x)\|), \end{aligned}$$

**B** during appearance model registration:

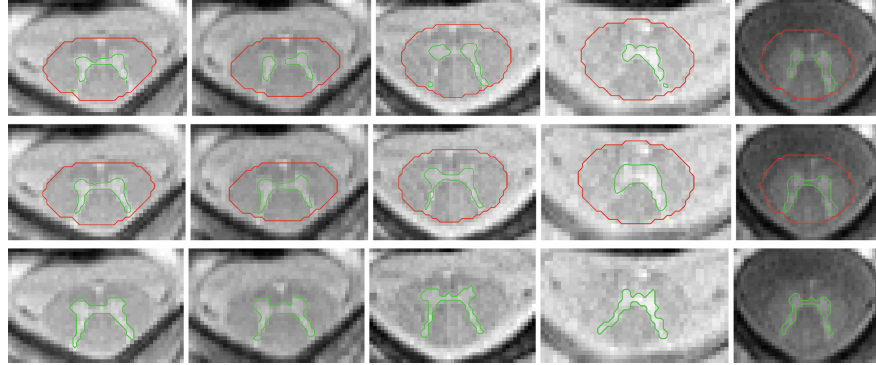
$$\begin{aligned} C_s(x) &= \mathcal{F}_{\text{median}}(\Gamma(|I_{\text{eff}}(x) - \text{mean}(I(\{f > 0.5\}))|, 4)), \\ C_t(x) &= \mathcal{F}_{\text{median}}(\Gamma(|I_{\text{eff}}(x) - \text{mean}(I(\{f \leq 0.5\}))|, 1/4)), \\ C(x) &= 1/10 \log(1 + d(x, f)) \cdot \zeta(x, f) + \epsilon. \end{aligned}$$

The time-domain **A** starts with an initial segmentation according to (10) without a prior. Once the segmentation updates reach the tolerance, the rigid registration part starts. The initial translational coordinate is set to the center of mass of the WM/GM mask  $M$ . For  $C$  we construct a special mask  $M_{\text{special}}$ , where capacities are forced to be high in regions where GM cannot be, see Fig. 2.

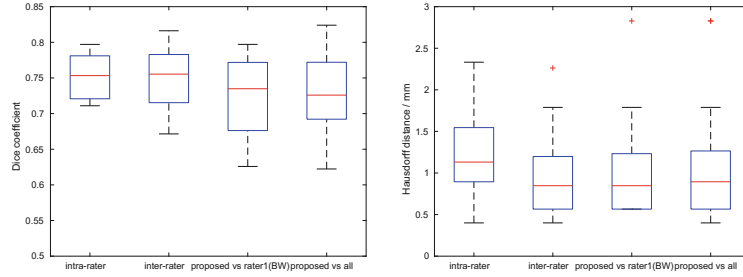
When the rigid registration and segmentation updates reach the tolerance, time-domain **B** starts. Here the model appearance  $f$  is being registered according to (12). We observed that, as long as  $f$  is registered appropriately, calculating mean intensities of object and background depending on the prior  $f$  seems more stable than looking at the actual areas defined through  $u$ . For  $C_s$  and  $C_t$ ,  $I_{\text{eff}} = (I_{\text{model}} + I_{\text{prior}})/2$  is replaced with  $(I + I_{\text{prior}})/2$ , compare (4). This way we combine again unblurred original image intensities with prior information.

With the curvature  $\kappa$  on the boundary of the prior, we define for any point in  $\Omega$  the normalized negative transform of the curvature  $\zeta(x, f) = 1 - N(\kappa)(\hat{x})$





**Fig. 3.** *Upper row:* Solutions of (10) show disconnected GM regions and depend on the image quality. *Middle row:* Proposed solutions show anatomically consistent results. *Lower row:* GM contours of pixel-wise manual segmentations of rater 1.



**Fig. 4.** GM segmentation results. *Dices:* intra-rater: binary vs grayscale; inter-rater: rater 1 vs 2 (grayscale); proposed vs rater 1 (binary); proposed (grayscale) vs rater 1 and 2 (binary and grayscale). *Hausdorff:* thresholded grayscale segmentations.

of the nearest neighbor  $\hat{x}$ . For the capacity function  $C$  we multiply the prior distance map with  $\zeta$ . This way the boundary has more freedom in regions with high curvature, see Fig. 2.

It is an advantage to choose  $\Delta t_\alpha$  smaller than 1 (we chose 0.2) because then the prior does not change too fast, in case the Chan-Vese solution at the initialization time has good quality and would be affected badly by the prior, which, in turn, would change the prior again.

## 4 Results

We implemented our models and algorithms in MATLAB and tested the GM results against the manual segmentation. For each dataset we created a statistical appearance model for (12) from the aligned manual segmentations of rater 1 and 2 of the remaining datasets, using principal component analysis. The CSF localization, spinal tract extraction and GM segmentation worked robustly and



fully automatic on all 16 images. Figure 3 shows exemplary GM results. We notice that the CSF segmentation is not perfect, because the anterior median fissure is not seen in the first image, where the CSF has bright intensity and overrules the thin structure because of partial volume effects.

Figure 4 shows mean Dice coefficients (0.75) and average Hausdorff distances (1 mm) of the proposed GM results. The Hausdorff distances have the tendency to be large, because the thin-structured posterior horns are not always delineated with the same length. Taso et al. [13] reach mean GM Dice coefficients of 0.83. Since our image resolution and signal-to-noise ratio are low, our manual segmentation only reaches a mean intra- and inter-rater Dice of 0.75. Therefore we expect Dices at best in the latter range, when comparing our method to the manual raters. Hand-segmented results are thresholded with 0.6, because of the downsampling technique, and the automated results with a standard 0.5. The error for thresholding is in the range of non-thresholded inter-rater variance. The proposed automated solution therefore can be seen as a third rater.

## 5 Discussion and Conclusion

We developed a variational approach to segment GM inside the spinal cord on C3 level. The algorithm works robustly on the given data sets and achieves similar Dice and Hausdorff measures as hand-segmented results. The solutions of the algorithm depend more on the intensity values gathered through the MR sequence and less on prior knowledge. Prior knowledge is included but does not overrule the information given in the MR image.

The quality of the results depends strongly on the imaging quality because of the fine structures present in the GM. We expect better quality in new images and thus better segmentation results. The MR sequence is still experimental and further adaptations are necessary for use in clinical practice.

We consider to implement a spatially adaptive weighting factor between the information of a given MR image and of the prior knowledge. The simple ellipsoidal-shaped prior was only used to locate the CSF, and in the future CSF and WM appearance models will be included. Since variational algorithms can easily be extended to additional dimensions, our model can be adapted for 3D MR images with good WM/GM contrast.

## References

1. Asman, A.J., Bryan, F.W., Smith, S.A., Reich, D.S., Landman, B.A.: Groupwise multi-atlas segmentation of the spinal cord’s internal structure. *Med. Image Anal.* **18**(3), 460–471 (2014)
2. Aubert, G., Kornprobst, P.: The segmentation problem. *Mathematical Problems in Image Processing. Applied Mathematical Sciences*, vol. 147, pp. 153–187. Springer, New York (2002)
3. Boykov, Y.Y., Jolly, M.P.: Interactive graph cuts for optimal boundary & region segmentation of objects in N-D images. In: *ICCV 2001*, vol. 1, pp. 105–112 (2001)

4. Cremers, D., Schmidt, F.R., Barthel, F.: Shape priors in variational image segmentation: Convexity, Lipschitz continuity and globally optimal solutions. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–6 (2008)
5. De Leener, B., Taso, M., Cohen-Adad, J., Callot, V.: Segmentation of the human spinal cord. *Magn. Reson. Mater. Phys.* **29**(2), 125–153 (2016)
6. Grasmair, M., Lenzen, F.: Anisotropic total variation filtering. *Appl. Math. Optim.* **62**(3), 323–339 (2010)
7. Losseff, N.A., et al.: Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain J. Neurol.* **119**(Pt 3), 701–708 (1996)
8. Olsson, C., Byröd, M., Overgaard, N.C., Kahl, F.: Extending continuous cuts: anisotropic metrics and expansion moves. In: ICCV 2009, pp. 405–412 (2009)
9. Overgaard, N.C., Fundana, K., Heyden, A.: Pose invariant shape prior segmentation using continuous cuts and gradient descent on lie groups. In: Tai, X.-C., Mørken, K., Lysaker, M., Lie, K.-A. (eds.) *SSVM 2009. LNCS*, vol. 5567, pp. 684–695. Springer, Heidelberg (2009). doi:[10.1007/978-3-642-02256-2\\_57](https://doi.org/10.1007/978-3-642-02256-2_57)
10. Pezold, S., Horváth, A., Fundana, K., Tsagkas, C., Andělová, M., Weier, K., Amann, M., Cattin, P.C.: Automatic, robust, and globally optimal segmentation of tubular structures. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) *MICCAI 2016. LNCS*, vol. 9902, pp. 362–370. Springer, Cham (2016). doi:[10.1007/978-3-319-46726-9\\_42](https://doi.org/10.1007/978-3-319-46726-9_42)
11. Schlaeger, R., et al.: Spinal cord gray matter atrophy correlates with multiple sclerosis disability. *Ann. Neurol.* **76**(4), 568–580 (2014)
12. Tang, L., Wen, Y., Zhou, Z., von Deneen, K.M., Huang, D., Ma, L.: Reduced field-of-view DTI segmentation of cervical spine tissue. *Magn. Reson. Imaging* **31**(9), 1507–1514 (2013)
13. Taso, M., et al.: A reliable spatially normalized template of the human spinal cord – applications to automated white matter/gray matter segmentation and tensor-based morphometry (TBM) mapping of gray matter alterations occurring with age. *NeuroImage* **117**, 20–28 (2015)
14. Wang, X.F., Min, H., Zou, L., Zhang, Y.G.: A novel level set method for image segmentation by incorporating local statistical analysis and global similarity measurement. *Pattern Recogn.* **48**(1), 189–204 (2015)
15. Weigel, M., Bieri, O.: A simple and fast approach for spinal cord imaging at 3T with high in-plane resolution and good contrast. In: *Proceedings of the 24th Annual Meeting of ISMRM*, Singapore. p. 4408, May 2016
16. Yiannakas, M., Kearney, H., Samson, R., Chard, D., Ciccarelli, O., Miller, D., Wheeler-Kingshott, C.: Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: a pilot study with application to magnetisation transfer measurements. *NeuroImage* **63**(3), 1054–1059 (2012)
17. Yuan, J., Bae, E., Tai, X.C.: A study on continuous max-flow and min-cut approaches. In: *CVPR 2010*, pp. 2217–2224 (2010)

## 6.A Adaptation of the Continuous Cut Model to AMIRA

This section describes the improvements to the original algorithm proposed in this chapter's publication [149]. The MOLLI sequence (cf. Page 78) was replaced by the AMIRA sequence (cf. Page 86) because of the AMIRA sequence's shorter acquisition times. To this end, the proposed algorithm needed to be adapted to the new sequence, which also includes acquired axial slices of lower cervical levels. The modified version was used as "the automatic segmentation algorithm" in the publication of Chapter 7. MATLAB code<sup>2</sup> is made publicly available.

Similar to the original method of this chapter [149], we again split the segmentation process into two parts: CSF segmentation with subsequent SC extraction, and GM-WM segmentation on the provided SC mask. We harmonized both parts in having three phases, which are depicted in Chapter 7, Fig. 2: in the first phase, the continuous cut iterations provide a first estimate of the segmentation; in the second phase, the mean appearance of a collection of translationally aligned manual segmentations is being rigidly aligned and superimposed onto the target image; and in the final third phase, a statistical appearance model of the same collection of manual segmentations is being adapted to the actual continuous cut iterations. We point out that during all three phases, the continuous cut iterations deliver new estimates for the final segmentation. During phases two and three, the appearance model is not only adapting to the current iterated segmentation but is also influencing the iterations by altering the terminal and non-terminal capacities according to Equation (2.107). We improved this setup in adding four major modifications: 1) support for multi-channel input, 2) a slice similarity prior that allows extending the 2D model to a 3D continuous cut model, 3) posterior appearance models based on trusted prior pixels, and 4) an ellipsoidal vector field for ATV during CSF segmentation.

1. We augmented the continuous cut model (2.97) for input images of  $M$  channels by linearly combining the channels' individual capacity functions  $C_{\cdot,i}$  for the terminal and the non-terminal capacities  $C_O$ ,  $C_B$ , and  $C$  using the following scheme:

$$C((I_i)_{i=1}^M, x) = \sum_{i=1}^M \lambda_{\cdot,i} C_{\cdot,i}(I_i, x), \quad (6.1)$$

where  $I_i$  are the individual input channels and  $\lambda_{\cdot,i}$  describe the channels' weighted contributions to the final capacity functions  $C_{\cdot}$ . The indexed dots symbolize the use for the terminal capacities  $C_O$ ,  $C_B$ , or the non-terminal capacity  $C$ .

In contrast to this modification, the original model of this chapter [149] used averaged one-channel projections of the MOLLI sequence's 11 inversion images. With this modification we directly feed the AMIRA sequence's individual inversion images and may also feed additional feature channels, such as the additional

---

<sup>2</sup><https://github.com/neonroehre/AJNR2019> (last accessed on Sept. 10, 2019)

CSF contrast image provided with the AMIRA protocol (cf. Page 86) and filtered or histogram-equalized projections of the IR sequences. Similar to the original model [149], we use mean intensity differences on the individual channels' terminal capacities during phases 1 and 2, because they can be efficiently calculated; during phase 3, we use the histogram-based terminal capacities of Boykov et al. [93] applied to the individual channels. Finally, we discarded the additional energy term with the Bhattacharyya distance in [149, Eq. 10], since histogram information is already utilized with the histogram-based terminal capacities during phase 3.

2. We added a slice similarity prior  $|\partial_{x_3} u(x)|$  to the anisotropic total variation (ATV) term  $\|\nabla u(x)\|_A = \sqrt{\nabla_{x_1, x_2} u^T A \nabla_{x_1, x_2} u}$  in (2.97),

$$\min_{u: \Omega \rightarrow [0,1]} \int_{\Omega} u C_{\mathcal{O}} + (1 - u) C_{\mathcal{B}} + C \|\nabla_{x_1, x_2} u(x)\|_A + \beta |\partial_{x_3} u(x)| dx, \quad (6.2)$$

weighted by a factor  $\beta$ . To this end, we collected all slices acquired with the AMIRA protocol (cf. Page 86) and co-registered them in a successive slice-wise fashion from top to bottom using integer-valued translations only to prevent interpolation. Slice-wise registration is performed two times, once before CSF segmentation and a second time before GM-WM segmentation using the estimated SC label maps.

3. Inspired by the probabilistic posterior models for partial information [150], we involved the appearance model more closely in the segmentation during phase 3 [149, Eq. 13] with the calculation of such posterior probabilities. In each iteration, we reconstruct the most likely posterior appearance given well-segmented pixels only.

Let the probabilistic appearance model of segmentations be modeled as follows:

$$f \sim \mathcal{N}(\mu, U(D^2 + \sigma^2 I)U^T), \quad (6.3)$$

where  $\mu \in \mathbb{R}^N$  is the mean appearance of a population of aligned and vectorized manual segmentations with  $N$  denoting the number of pixels,  $U \in \mathbb{R}^{N \times M_{\text{pca}}}$  is a collection of the populations' highest-ranking principal components with  $M_{\text{pca}}$  denoting the number of used principal components,  $D^2 \in \mathbb{R}^{M_{\text{pca}} \times M_{\text{pca}}}$  is a diagonal matrix with the principal components' eigenvalues,  $\sigma^2 \in \mathbb{R}$  is the variance of the added probabilistic Gaussian noise, and  $I \in \mathbb{R}^{M_{\text{pca}} \times M_{\text{pca}}}$  is the identity matrix.

Assuming an iterated segmentation  $u^k$  rigidly aligned to the probabilistic appearance model, we can calculate an energy of  $u^k(x)$  in the model space pixelwise through the negative log-likelihood of the pixelwise marginal of (6.3):

$$E(u^k)(x) = (u^k(x) - \mu(x)) \sum_{j=1}^{M_{\text{pca}}} ((UD_{\text{pca}})_{x,j})^2, \quad (6.4)$$

## 6.A Adaptation of the Continuous Cut Model to AMIRA

leaving out any proportional factors and additive terms, where the diagonal matrix  $D_{\text{ppca}} = \sqrt{D^2 + \sigma^2 \bar{I}}$  consists of the probabilistic singular values calculated via element-wise square-roots. Using the pixelwise energies  $E(u^k)(x)$  we trust those pixels that have an energy smaller than a selected threshold of trust. In the following, we denote with index  $b$  in  $u_b^k$ ,  $\mu_b$ , and  $W_b$  the respective entities that live on the trusted pixels only. The matrix  $W_b$ , in this case, is the collection of the weighted principal components  $UD$  with deleted rows of those pixel dimensions that are not trusted. Based on these trusted pixels of the iterated segmentation  $u^k$ , we reconstruct the most likely posterior appearance  $f_{\text{MLE}}$  by first calculating the most likely latent variable

$$\alpha_{\text{MLE}} = \frac{1}{\sigma^2} \left( I + \frac{1}{\sigma^2} W_b^T W_b \right)^{-1} (W_b^T (u_b^k - \mu_b)) \quad (6.5)$$

and then finally projecting the latent variable back into the appearance space through

$$f_{\text{MLE}} = \mu + UD \alpha_{\text{MLE}}. \quad (6.6)$$

Such posterior appearances  $f_{\text{MLE}}$  are calculated for each slice by generating a model that consists of only those manual segmentations located on similar vertebral levels as the target slice and by excluding those samples of the same subject. The slicewise posterior appearances are then superimposed onto the target image's channels according to (2.107). Whereas the original model [149] uses adaptive appearance models only for the GM segmentation part, we harmonized the two parts and included such models also for the CSF segmentation part.

4. To the third phase of the CSF segmentation, we added a Riemannian metric matrix field  $A$  with strong ellipsoidal rotational components and weak center normal components positioned at the center of the SC mask. With this metric field  $A$ , ATV favors ellipsoidal segmentations and punishes label changes perpendicular to the rotational streamlines. This setup helps inpainting boundaries in cases where the ring-shaped CSF has an appearing gap in the image data (cf. Chapter 6, Fig. 2) which occurs if, for example, the SC is touching the posterior part of the vertebral column.

With these modifications, many more parameters were introduced that enable a better capability of generalization as compared to the original model [149]. We optimized all parameters groupwise through exhaustive search on discretized parameter grids on a small validation set. Still, an ablation analysis of the proposed method in terms of accuracy was difficult to perform since all these modifications greatly influenced the method's robustness: Including the additional CSF contrast image led to 4% fewer CSF segmentation failures; varying the channels' contribution to the capacities via Equation (6.1) greatly influenced the algorithm's performance in terms of CPU time, robustness, and accuracy; the slice similarity prior improved segmentation stability as compared to pure slicewise segmentation; the posterior shape models

## 6 Variational Segmentation of SC GM-WM

with the probabilistic noise factor and the threshold of trust introduced another two parameters that have a fair influence on robustness and accuracy, which we observed during the parameter search; and the ellipsoidal ATV for CSF segmentation significantly increased the chance for closing gaps in the ring-shaped CSF.

## 7 Novel Reliable Automatic Pipeline for Spinal Cord Gray Matter Quantification

In the publication presented in this chapter, we assess the reproducibility and accuracy of measurements from images acquired with the AMIRA protocol (cf. Page 86) by segmenting the GM-WM structures of the acquired images in the CGM dataset (described on Page 88) manually and with the modified continuous cut algorithm described in Section 6.A.

Results show high reproducibility for both the manual and the automatic segmentations, where the automatic segmentations showed more reproducible results in intra- and inter-session comparisons. The similarity between the automatic segmentations and the manual reference standard, interpreted as accuracy, was high with a mean DSC of 0.86 for GM and 0.90 for WM.

Keeping in mind that the quantification of the SC's inner structure depends on both the imaging and the post-processing part, the focus of this publication lies less on the analysis of the modified continuous cut algorithm, but rather on the quality of the AMIRA images on healthy subjects. Considering the challenges of SC MRI (cf. Page 75) and considering that with the proposed continuous cut approach a shallow algorithm is capable of providing state-of-the-art segmentations in 88% of all acquired slices, the high quality of the acquisitions with the AMIRA protocol is substantiated.

To further analyze the quality of the AMIRA images, we adapted the iterative non-local STAPLE (iNLS) algorithm for GM segmentation proposed by Asman et al. [97], which is also the underlying algorithm of the atlas-based GM segmentation algorithm in the spinal cord toolbox (SCT) [137]. Applying the method of Asman et al. [97] on AMIRA images yields superior results to the same method being applied on  $T_2^*$ -weighted 3D gradient echo images, as reported in [97]. In a second attempt, we wanted to test the SCT's implementation of iNLS [137] together with the SCT's provided MNI-Poly-AMU atlas [127]. Unfortunately, this did not perform as expected on the AMIRA images and would have required the construction of a new atlas, based on the AMIRA images in the data format of the SCT. However, since according to the SCT developers, this code is not maintained anymore, we were not able not to generate such an atlas even after several weeks of trial. Hence, this attempt is not part of the publication.

Moreover, the proposed continuous cut algorithm was not designed to differentiate between healthy and inflamed nervous tissue and therefore fails in segmenting AMIRA images of MS patients with presence of lesions, as presented in this chapter's Fig. 6.

This limitation, however, does not devalue the results of this publication regarding reproducibility of images without lesions acquired with the AMIRA protocol.

**Publication** The publication was written with joint first authorship by Charidimos Tsagkas and Antal Horváth, with A. Horváth developing and applying the post-processing pipeline and conducting the statistical analysis, and C. Tsagkas focusing on the clinical setup and evaluation. It was accepted by the American Journal of Neuroradiology (AJNR) in June 2018 and was published<sup>1</sup> in August 2019 [145]. MATLAB code<sup>2</sup> is made publicly available.

---

<sup>1</sup><https://doi.org/10.3174/ajnr.A6157> (last accessed on Sept. 19, 2019)

<sup>2</sup><https://github.com/neonroehre/AJNR2019> (last accessed on Sept. 10, 2019)



# Automatic Spinal Cord Gray Matter Quantification: A Novel Approach

C. Tsagkas, A. Horvath, A. Altermatt, S. Pezold, M. Weigel, T. Haas, M. Amann, L. Kappos, T. Sprenger, O. Bieri, P. Cattin, and K. Parmar



## ABSTRACT

**BACKGROUND AND PURPOSE:** Currently, accurate and reproducible spinal cord GM segmentation remains challenging and a noninvasive broadly accepted reference standard for spinal cord GM measurements is still a matter of ongoing discussion. Our aim was to assess the reproducibility and accuracy of cervical spinal cord GM and WM cross-sectional area measurements using averaged magnetization inversion recovery acquisitions images and a fully-automatic postprocessing segmentation algorithm.

**MATERIALS AND METHODS:** The cervical spinal cord of 24 healthy subjects (14 women; mean age,  $40 \pm 11$  years) was scanned in a test-retest fashion on a 3T MR imaging system. Twelve axial averaged magnetization inversion recovery acquisitions slices were acquired over a 48-mm cord segment. GM and WM were both manually segmented by 2 experienced readers and compared with an automatic variational segmentation algorithm with a shape prior modified for 3D data with a slice similarity prior. Precision and accuracy of the automatic method were evaluated using coefficients of variation and Dice similarity coefficients.

**RESULTS:** The mean GM area was  $17.20 \pm 2.28 \text{ mm}^2$  and the mean WM area was  $72.71 \pm 7.55 \text{ mm}^2$  using the automatic method. Reproducibility was high for both methods, while being better for the automatic approach (all mean automatic coefficients of variation,  $\leq 4.77\%$ ; all differences,  $P < .001$ ). The accuracy of the automatic method compared with the manual reference standard was excellent (mean Dice similarity coefficients:  $0.86 \pm 0.04$  for GM and  $0.90 \pm 0.03$  for WM). The automatic approach demonstrated similar coefficients of variation between intra- and intersession reproducibility as well as among all acquired spinal cord slices.

**CONCLUSIONS:** Our novel approach including the averaged magnetization inversion recovery acquisitions sequence and a fully-automated postprocessing segmentation algorithm demonstrated an accurate and reproducible spinal cord GM and WM segmentation. This pipeline is promising for both the exploration of longitudinal structural GM changes and application in clinical settings in disorders affecting the spinal cord.

**ABBREVIATIONS:** AMIRA = averaged magnetization inversion recovery acquisitions; CV = coefficient of variation; DSC = Dice similarity coefficient; HD = Hausdorff distance; SC = spinal cord

The human spinal cord (SC) can be affected by numerous neurologic disorders of variable pathophysiology (eg, genetic, inflammatory, demyelinating, degenerative, and so forth),<sup>1,2</sup> and

MR imaging is a valuable part of the diagnostic work-up in patients with suspected intramedullary pathology.<sup>3,4</sup> SC gray matter and white matter can be involved to a various extent not only among different SC disorders but also among patients with the same disease (eg, multiple sclerosis, amyotrophic lateral sclerosis).<sup>5,6</sup> Hence, quantification of SC compartments may add to our understanding of SC pathology<sup>5,6</sup> and hopefully help in the management of individual patients in the future.

However, the SC presents additional challenges for MR imaging.

Received November 26, 2018; accepted after revision June 25, 2019.

From the Neurologic Clinic and Policlinic (C.T., M.A., L.K., T.S., K.P.), Department of Medicine and Biomedical Engineering; Translational Imaging in Neurology Basel (C.T., A.A., M.A., M.W., L.K., K.P.), Department of Medicine and Biomedical Engineering; Division of Radiological Physics (M.W., T.H., O.B.), Department of Radiology; and Division of Diagnostic and Interventional Neuroradiology (M.A.), Department of Radiology, University Hospital Basel, University of Basel, Basel, Switzerland; Medical Image Analysis Center (C.T., A.A., M.A.), Basel, Switzerland; Department of Biomedical Engineering (A.H., A.A., S.P., M.W., O.B., P.C.), University of Basel, Allschwil, Switzerland; and Department of Neurology (T.S.), DKD HELIOS Klinik, Wiesbaden, Germany.

C. Tsagkas and A. Horvath contributed equally as first authors.

This work was supported by the Swiss National Science Foundation (grant No. 320030\_156860) and the Foundation for Sponsorship of Gastroenterological and General Clinical Research as well as of Medical Imaging (application ID 02/2015).

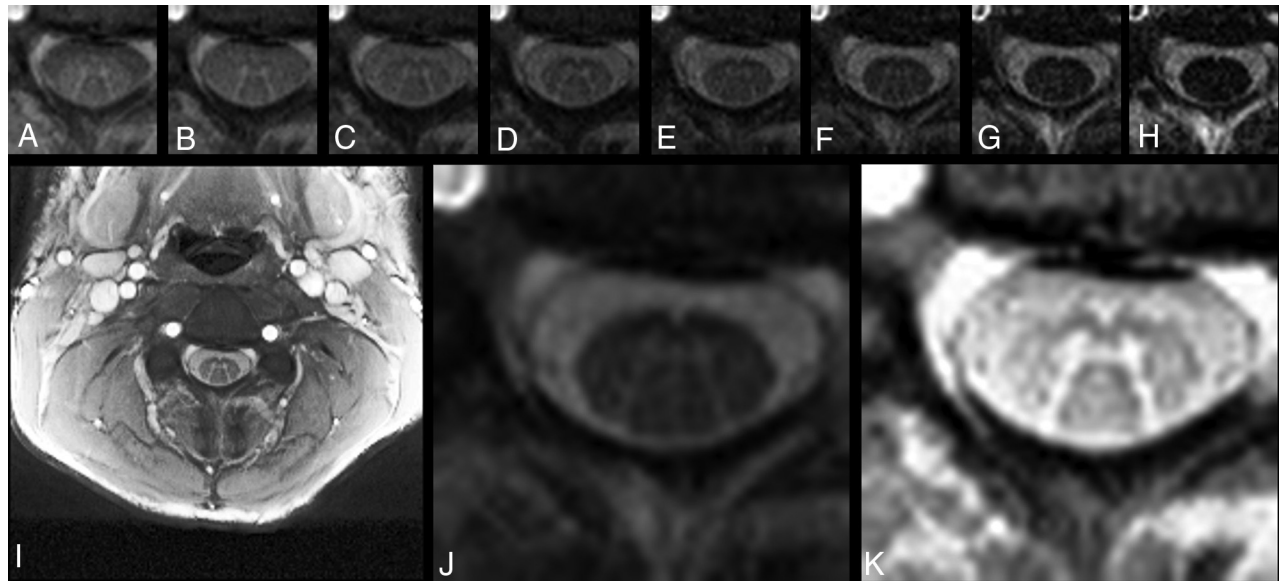
Please address correspondence to Katrin Parmar, MD, Department of Neurology, University Hospital Basel, Petersgraben 4, CH-4031 Basel, Switzerland; e-mail: katrin.parmar@usb.ch

Indicates open access to non-subscribers at [www.ajnr.org](http://www.ajnr.org)

Indicates article with supplemental on-line table.

Indicates article with supplemental on-line photo.

<http://dx.doi.org/10.3174/ajnr.A6157>



**FIG 1.** Exemplary axial AMIRA slice of 1 representative volunteer at the C4 level. A–H, Eight images of different tissue contrast acquired by the AMIRA sequence, shown in chronologic order from lowest-to-highest TI. I, Average image from A to E in full view, which delivers a high contrast-to-noise-ratio for GM/WM. J, Average image from F to H, which delivers a high contrast-to-noise ratio for SC/CSF. K, Same average image as in I but histogram-equalized and zoomed.

ing. The SC is surrounded by a number of different tissue types, including CSF, bone, and air. These create considerable signal inhomogeneities along this thin, elongated structure.<sup>3,4,7,8</sup> As a result, conventional SC MR imaging was—until recently—not able to differentiate sufficiently among SC GM, WM, and CSF. In the past, there were first attempts toward this differentiation using a series of acquisition approaches.<sup>9–12</sup> More recently an averaged magnetization inversion recovery acquisitions (AMIRA) sequence was proposed, delivering a notable SC GM/WM contrast while maintaining short acquisition times at the same time.<sup>13</sup> The latter is especially important for imaging small-sized structures (like the SC GM/WM) in patients with disabilities having a short time window in which they can lie still.

Moreover, accurate SC GM segmentation remains challenging. First, manual approaches demonstrated the feasibility of distinguishing between WM and GM.<sup>9</sup> However, manual approaches require a considerable amount of time, are prone to error, and demonstrate significant interobserver and intraobserver variability. As a result of improvements in image quality and postprocessing techniques, the first fully automatic SC GM segmentation methods were established in the past few years.<sup>14–17</sup> These methods have deployed atlas-based GM segmentation algorithms, which may, however, lead to misestimations or segmentation errors, especially in case of pathology, image artifacts, or large between-individual anatomic variations.<sup>18,19</sup> A noninvasive broadly accepted reference standard for accurate and reproducible SC GM measurements is still a matter of ongoing discussion.

In this study, we validate a fully automatic method for SC GM and WM segmentation in terms of its reproducibility and accuracy in segmenting the cervical SC of healthy controls against a manual segmentation. The proposed approach used a variational segmentation algorithm with a shape prior,<sup>20</sup> modified for 3D data with a slice similarity prior on AMIRA images.

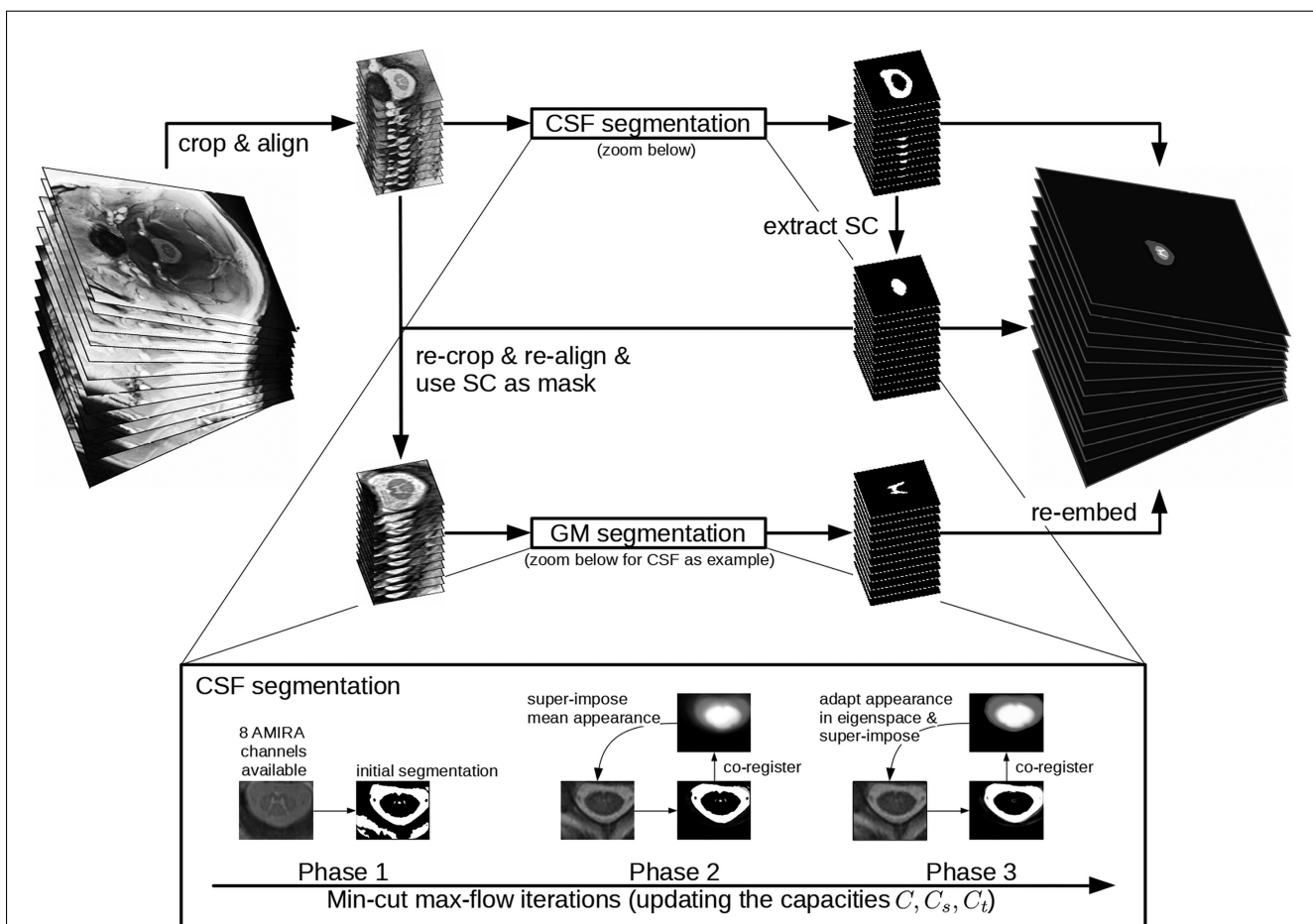
## MATERIALS AND METHODS

### Subjects and MR Imaging Acquisition

Twenty-four healthy subjects (14 women; mean age,  $40 \pm 11$  years) were scanned in a test-retest fashion on a 3T whole-body MR imaging system (Magnetom Prisma; Siemens, Erlangen, Germany). All subjects provided written consent. Experimental procedures conformed to the Declaration of Helsinki, and the study protocol was approved by the local ethics committee. We acquired 12 axial AMIRA images<sup>13</sup> (FOV =  $128 \times 128$  mm<sup>2</sup>, slice thickness = 8 mm, slice overlap = 4 mm, in-plane resolution =  $0.67 \times 0.67$  mm<sup>2</sup>, TE<sub>balanced steady-state free precession</sub> = 2.14 ms, TR<sub>balanced steady-state free precession</sub> = 5.13 ms, signal averaging = 1, acquisition time = 51 seconds per slice) over a 48-mm cervical SC segment, extending approximately from the C2–C5 vertebral level.<sup>13</sup> The most rostrally acquired slice was placed with its lower surface adjacent to the most rostral surface of the C2/C3 intervertebral disc. For precise positioning of each individual slice and its orthogonal angulation to the course of the SC, a strongly T2-weighted TSE with high contrast between CSF and SC was used as a reference. For each slice, the AMIRA approach acquired 8 images of considerably different tissue contrast among GM, WM, and CSF with effective TI = 97.1, 158.7, 220.2, 281.8, 343.3, 404.9, 466.5, 528.0 ms. Averaging the first 5 images enhances the GM/WM contrast-to-noise ratio, whereas averaging the last 3 images clearly improves the WM/CSF contrast-to-noise ratio (Fig 1). For more details on the AMIRA sequence, please see Weigel and Bieri, 2018.<sup>13</sup>

Each subject was scanned 3 times in 1 MR imaging session. The first 2 scans were performed in a back-to-back fashion without repositioning to allow intrasession comparisons. The third scan was obtained after patient repositioning to allow intersession comparisons.

All scans underwent basic preprocessing including 2D and 3D correction for field inhomogeneities using the scanner soft-



**FIG 2.** Flow chart of the automatic segmentation pipeline. As a first step of the algorithm to align the 12 slices, the images are center-cropped and slice-wise successively coregistered rostral to caudal using translations in pixel-size steps to prevent further interpolation. Then, the algorithm automatically locates and delineates the ring-shaped CSF from its surroundings and extracts the cross-sectional SC surface. Finally, it uses the previously segmented SC surface as a mask for GM/WM differentiation. The iterative steps of CSF segmentation are shown as a zoomed-in view. GM segmentation uses essentially the same steps and is thus not shown in detail.

ware before segmentation. To minimize numeric errors of the validation metrics, we performed a 5-fold in-slice upsampling of the slices using the Lanczos-3 interpolation kernel.

### SC Segmentation

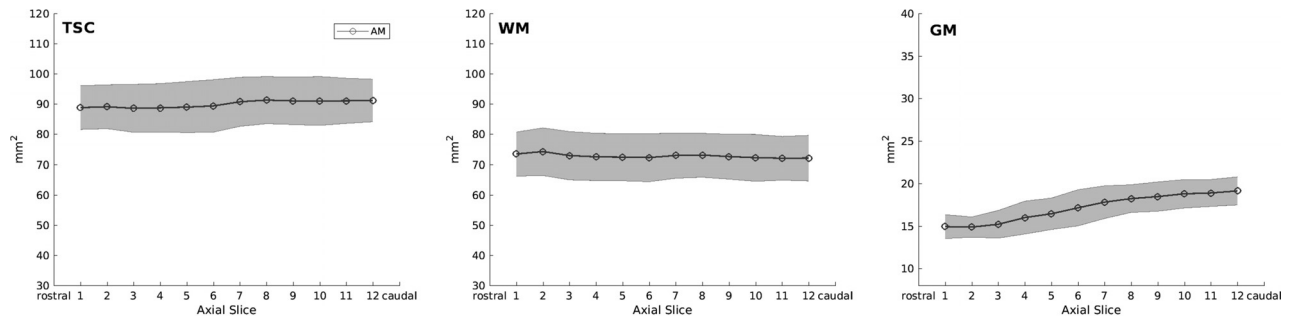
As proposed in a previous study,<sup>20</sup> a variational segmentation approach based on the continuous min-cut max-flow framework was used, which includes total variation regularization to segment WM and GM. The min-cut max-flow capacity functions are modeled using edge, region, and prior information as well as an appearance model built from manual segmentations. Aiming for high accuracy, the proposed approach prefers intensities of the actual image and tries to include prior information as little as possible, which regularizes for higher precision. Compared with the previous study,<sup>20</sup> we added a slice similarity prior,<sup>21</sup> included all inversion images of one slice (Fig 1) into the calculation of the max-flow capacity functions, and improved the adaptation of the appearance model with posterior models that reconstruct the most likely appearance only based on well segmented pixels.<sup>22</sup>

As a first step of the algorithm to align the 12 slices, the images are center-cropped and slice-wise successively coregistered rostral to caudal using translations in pixel-size steps to prevent further interpolation. Then, the algorithm automatically locates and delineates the ring-shaped CSF from its surroundings and extracts

the cross-sectional SC surface. Finally, it uses the previously segmented SC surface as a mask for GM/WM differentiation. An illustration of the algorithm is shown in Fig 2. Segmentations were achieved in a leave-one-subject-out cross-validation—that is, with the currently segmented subject being left out in the appearance model used.

The segmentation algorithm was implemented in Matlab (MathWorks, Natick, Massachusetts). Processing time on a Xeon CPU E5-2620 v3 @ 2.40GHz (Intel, Santa Clara, California) is around 1 minute for each segmentation step (CSF-SC and WM-GM segmentation), and fewer than 8 GB of RAM is used to segment a stack of 12 slices. Code is available on <https://github.com/neonroehre/AJNR2019>.

Two experienced raters (C.T. and A.A.) were involved in the manual segmentations. Both raters had >4 years' experience in neuroimaging research, including SC volumetric studies. In a first step, segmentations were conducted on the average of the last 3 AMIRA images for the total SC cross-sectional area. Using the already delineated total SC masks, we then performed manual segmentations of the GM and WM cross-sectional areas on the average of the first 5 AMIRA images (On-line Figure). C.T. segmented all images once. These results were further applied as a "manual reference standard." C.T. also conducted a second



**FIG 3.** Cross-sectional areas of total spinal cord, white matter, and gray matter per axial slice as measured by automatic segmentations. Notice the slight increase of total spinal cord (TSC) and the marked GM cross-sectional area increase caudally, which corresponds to the cervical SC enlargement. The light gray area depicts the limits of  $\pm 1$  SD.

“run” of 60 randomly selected slices to assess intrarater comparisons. This second run was conducted with slightly different contrast adjustments than the first to evaluate the robustness of intrarater manual segmentation. A.A. segmented all images of the first scan of all 24 healthy controls to allow interrater comparisons.

To evaluate the performance of our method on SC slices, in which the fully automatic approach failed (in total 12% of acquired slices, see also Results), we applied a semiautomatic approach as follows: The SC/CSF boundaries were segmented manually (manual reference standard) and segmentation of the GM and WM was then performed using the fully-automatic approach described above, given the manual total SC masks.

To compare our automatic method with currently available algorithms, we tested the iterative nonlocal STAPLE algorithm<sup>23</sup> on our AMIRA images using the algorithm in the SCFusion\_Demo package ([https://www.nitrc.org/frs/download.php/7666/scfusion\\_demo.zip](https://www.nitrc.org/frs/download.php/7666/scfusion_demo.zip)). Asman et al<sup>23</sup> used atlases consisting of SC-GM-WM contrast images and SC-GM-WM manual reference standard segmentations, which are rigidly registered to the target slice and fused together with the most fitting manual reference standard segmentation as an estimation of the targeted segmentation. We built our own atlases and tested the iterative nonlocal STAPLE in a leave-one-subject-out fashion.

### Statistical Analysis

Intra- and intersession and intra- and interrater reproducibility of the 2 approaches were evaluated using coefficients of variation (CVs), Dice similarity coefficients (DSCs), and Hausdorff distances (HDs). The accuracy of the automatic method compared with the manual reference standard was evaluated using the DSC and HD. CVs between the 2 masks *A* and *B* were calculated with the following formula:

$$CV = \sqrt{2} \times \frac{|A - B|}{|A + B|} \%$$

DSCs were calculated as follows:

$$DSC = 2 \times \frac{|MRS \cap AM|}{|MRS| + |AM|}$$

HDs were calculated as follows:

$$d(X \rightarrow Y) = \max(d_i^{X \rightarrow Y}, i = 1 \dots N_x)$$

$$HD = \max[d(MRS \rightarrow AM), d(AM \rightarrow MRS)],$$

where *d* is the Euclidean distance between voxels *x* and *y*, MRS is the manual reference standard and AM the automatic method.

Because of non-normally distributed data, we performed a square root transformation of the CV, a cubic transformation of the DSC, and a logarithmic transformation of the HD before conducting all *t* test analyses and MANOVA. Two-paired *t* tests were performed for the following comparisons after controlling for normal data distribution: 1) manual segmentation versus automatic method reproducibility, and 2) manual segmentation versus automatic method total SC, WM, and GM cross-sectional area. For the automatic method, differences in measures of reproducibility and accuracy between intrasession and intersession; among GM, WM, and total SC; and among the axial slice levels (1–12) were investigated using MANOVA. Additional Tukey post hoc tests were conducted when applicable.

### RESULTS

In total, 864 slices were acquired from 24 volunteers with 12 slices per scan, and each scan was performed a total of 3 times for each subject. Of 864 acquired axial SC slices, 9 were excluded from further analysis because of severe imaging artifacts. The automatic method successfully segmented 88% (752 slices) of all remaining slices. Because of imaging artifacts, localization problems, or posterior gaps of the CSF, 8% of all slice-wise SC segmentations and 4% of GM segmentations would have needed further manual interventions and thus were excluded from the reproducibility analysis.

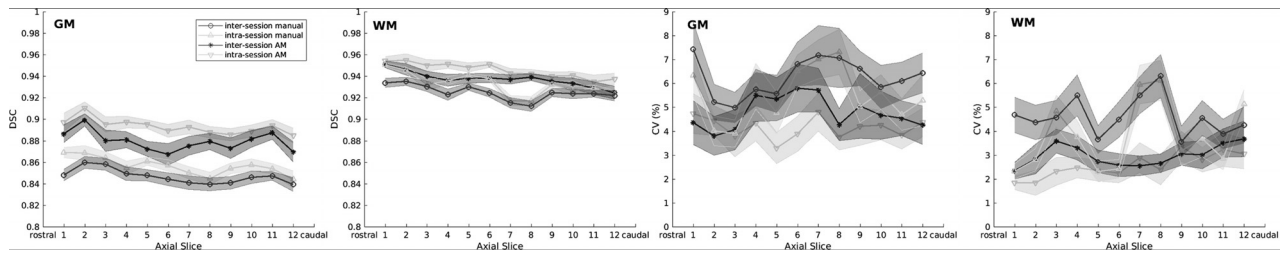
#### Cross-Sectional SC Measurements

The mean total SC area was  $89.98 \pm 7.88 \text{ mm}^2$ , the mean WM area was  $72.71 \pm 7.55 \text{ mm}^2$ , and the mean GM area was  $17.20 \pm 2.28 \text{ mm}^2$  as measured by the automatic method. Compared with the manual reference standard, the automatic method delivered significantly higher total SC and WM area as well as significantly lower GM area ( $86.88 \pm 11.87$ ,  $69.18 \pm 10.16$ , and  $17.77 \pm 3.05 \text{ mm}^2$ , respectively; all,  $P < .001$ ). Cross-sectional areas per slice of the automatic method are shown in Fig 3.

#### Reproducibility

Measurements of intra- and intersession and intra- and interrater reproducibility are shown in the On-line Table. Reproducibility





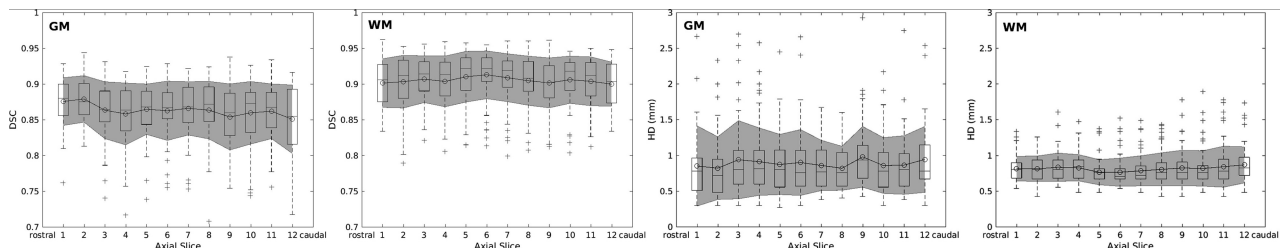
**FIG 4.** Comparison between the reproducibility of manual and automatic measurements (AM) of spinal cord gray matter and white matter per axial slice. Intrasession and inter-session reproducibility is assessed in terms of Dice coefficients (graphics on the left) and coefficients of variation (graphics on the right). Manual and automatic inter-session reproducibility is shown in dark gray, whereas manual and automatic intrasession reproducibility is shown in light gray. Error bars display mean values  $\pm$  0.2 SDs.

#### SC GM, WM, and TSC accuracy of automatic and semiautomatic segmentations against the manual reference standard segmentations<sup>a</sup>

	Parameter	GM	WM	TSC
AM (on 88% of acquired slices, nondiscarded from initial analysis)	DSC	0.86 $\pm$ 0.04 (0.87)	0.90 $\pm$ 0.03 (0.91)	0.95 $\pm$ 0.03 (0.95)
	HD (mm)	0.90 $\pm$ 0.44 (0.72)	0.82 $\pm$ 0.22 (0.75)	0.64 $\pm$ 0.27 (0.67)
SAM (on 12% of acquired slices, discarded from initial analysis)	DSC	0.83 $\pm$ 0.04 (0.84)	0.96 $\pm$ 0.01 (0.96)	
	HD (mm)	1.11 $\pm$ 0.55 (0.93)	0.64 $\pm$ 0.15 (0.67)	
AM (on nondiscarded samples) and SAM (on discarded slices), mixed (100% of acquired slices)	DSC	0.86 $\pm$ 0.04 (0.86)	0.91 $\pm$ 0.04 (0.92)	0.96 $\pm$ 0.03 (0.96)
	HD (mm)	0.91 $\pm$ 0.46 (0.81)	0.80 $\pm$ 0.22 (0.75)	0.60 $\pm$ 0.29 (0.55)

**Note:**—SAM indicates semiautomatic segmentation; TSC, total spinal cord; AM, automatic segmentations; DSC, Dice coefficient; HD, Hausdorff distances.

<sup>a</sup> All values are shown as mean  $\pm$  SD (median).



**FIG 5.** Accuracy measurements in terms of Dice coefficients (graphics on the left) and Hausdorff distances (graphics on the right) of white matter and gray matter per slice. Overlaid boxplots display median values as well as 25th and 75th percentile values. Gray areas depict the mean standard error values  $\pm$  1 SD.

of SC GM and WM is also depicted per slice in Fig 4. All mean CVs of the automatic method were  $\leq 4.77\%$ , and the mean DSC was  $\geq 0.88$  between scans and raters. The latter was significantly better than for the manual segmentation (all,  $P < .001$ ).

By means of MANOVA with DSC, HD, and CV as multivariate outcomes, a significant difference between intra- and inter-session reproducibility for total SC, WM, and GM using the automatic method was shown ( $P < .001$  for all 3 models). However, CVs differed only for WM and total SC ( $P < .05$  and  $P < .001$ , respectively), but not for GM. In our automatic method, intra- and inter-session reproducibility was significantly decreased in the order total SC  $\rightarrow$  WM  $\rightarrow$  GM (all  $P < .001$ ), as shown by MANOVA and post hoc tests. No difference was found in intra- and inter-session reproducibility among slices for GM, but a significant decrease was found for WM (both  $P < .001$ ) and total SC ( $P < .05$  and  $P < .001$ , respectively) in caudal slices, as shown by MANOVA. However, CVs were similar for all SC metrics among all slices.

#### Accuracy

Measurements of accuracy of our fully automatic method compared with a manual reference standard are shown in detail in the Table and are also shown per slice in Fig 5. The automatic method showed a mean DSC of  $\geq 0.86$  in all SC metrics. Accuracy was

significantly decreased in the order total SC  $\rightarrow$  WM  $\rightarrow$  GM (all  $P < .001$ ), as shown by MANOVA and post hoc tests. In MANOVA, accuracy was lower for GM ( $P < .05$ ) and total SC ( $P < .001$ ) in caudal slices, but not for WM. However, the DSC was similar among acquired slices for total SC.

Measurements of accuracy of the initially discarded SC slices (12% of all acquired AMIRA slices) analyzed in a semiautomatic fashion are also shown in the Table. The semiautomatic approach showed a mean DSC of  $\geq 0.83$  in both GM and WM. When we compared it with the fully automatic approach on the initially nondiscarded SC slices, a statistically significant accuracy decrease was observed in the semiautomatic approach (both  $P < .001$ ).

#### Comparison with the Iterative Nonlocal STAPLE Algorithm

In comparison with the original study<sup>23</sup> performed on T2\* MR images, the application of the iterative nonlocal STAPLE algorithm in our AMIRA images showed a higher accuracy. Mean DSC and HD for the total SC, GM, and WM were as follows—total SC: mean DSC =  $0.93 \pm 0.03$  (median, 0.94), mean HD =  $0.96 \pm 0.39$  mm (median, 0.84 mm); GM: mean DSC =  $0.80 \pm 0.06$  (median, 0.82), mean HD =  $1.09 \pm 0.42$  mm (median, 1.04 mm); WM: mean DSC =  $0.87 \pm 0.04$  (median, 0.88), mean HD =

$0.98 \pm 0.37$  mm (median, 0.89 mm).<sup>23</sup> Moreover, our proposed automatic method had higher accuracy for all total SC, WM, and GM compared with the iterative nonlocal STAPLE algorithm in our AMIRA images (all,  $P < .001$ ).

## DISCUSSION

Visualization of the SC GM in MR imaging has been hampered by technical difficulties until recently.<sup>8</sup> Despite technologic advancements, segmentation of SC compartments remains a challenge.<sup>24</sup> In this work, we successfully deployed the novel MR imaging approach AMIRA and a fully automatic variational segmentation algorithm with a shape prior modified for 3D data with a slice similarity before demonstrating a fully automated approach for segmentation of SC, GM, and WM.

In contrast to brain MR imaging, the environment of the SC presents additional challenges for MR imaging methods and inherently for SC segmentation. The greatest challenges include magnetic field inhomogeneities across the SC, cord curvature, shape and size, contact of the SC and the osseous canal, osteophytes causing focal changes in CSF flow dynamics within the canal, motion artifacts, Gibbs artifacts, partial volume effects, and  $B_1$  inhomogeneity.<sup>25-27</sup> The AMIRA sequence<sup>13</sup> used is based on a 2D approach that is, generally, less motion-sensitive than a 3D sequence. It uses a relatively short acquisition time of 51 seconds per slice, which leads to a reduction of motion artifacts and is especially suitable for disabled patients with limited ability to lie still (eg, due to spasticity). The AMIRA approach makes use of a balanced steady-state free precession readout, which is inherently of low-flow sensitivity or inherently flow compensated.<sup>13</sup> The inversion recovery preparation is global and nonselective; hence, it does not pose an issue for CSF flow sensitivity either.

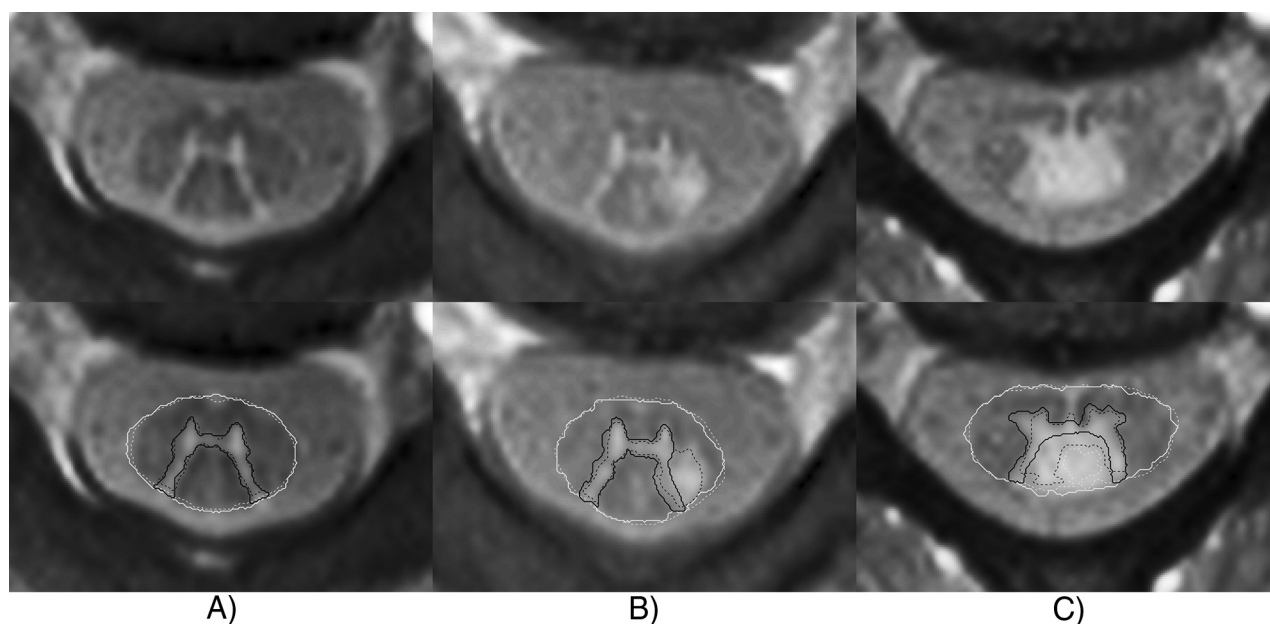
Furthermore, because the SC has a small cross-section of roughly  $1.3 \times 0.7$  cm and our slices were located close to the isocenter, effects of  $B_1$  inhomogeneity do not play a significant role for the present AMIRA images. The even smaller size of the SC GM presents additional difficulties for MR imaging methods, requiring submillimeter in-plane resolutions, especially for morphometry. Visualization and segmentation of the SC GM and WM are hampered by the similar relaxation times of the 2 SC compartments, limiting the use of conventional SC MR imaging for that purpose. Finally, the complex butterfly shape of the SC GM makes the segmentation of the structure a rather difficult task for computer-based segmentation methods. The AMIRA approach was able to produce SC images with a high GM/WM contrast in all participants. This was achieved in clinically feasible acquisition times (10.2 minutes for a 48-mm cervical SC segment).

Of 864 slices, only 9 were excluded due to image artifacts, mainly resulting from magnetic field inhomogeneities produced by bone structures (clavicles, scapulae, humeri, ribs, and so forth) as well as due to aliasing and motion artifacts. Although, these artifacts occurred in a rather small percentage of the acquired images (1%), they should be taken into account in future applications of the AMIRA approach. A further argument in favor of the use of AMIRA for SC GM and WM quantification is that our pipeline was able to deliver not only higher accuracy measures compared with a previous study<sup>24</sup> demonstrating results from

various MR imaging sequences and segmentation algorithms, but also better accuracy performance of an established algorithm on AMIRA compared with T2\* MR images (see also below). This result may be an indirect indication of the superior quality of AMIRA compared with other sequences used so far for spinal cord GM and WM quantification. Nevertheless, due to AMIRA having a nonisotropic resolution, our MR imaging acquisition may have been more prone to partial volume effects, despite our slices being angulated individually in an orthogonal way to the course of the SC.

The proposed automatic segmentation method showed excellent precision in terms of inter- and intrasession reproducibility and was superior to the manual segmentation performed by experienced raters for all SC metrics, as measured by both CV and DSC. Our automatic method was also superior in terms of the HD for total SC, though it did not differ with regard to SC WM and GM. At the same time, the accuracy of the automatic method was high for total SC, GM, and WM, as measured by both the DSC and HD. Comparing the present data with results of the SC GM segmentation challenge,<sup>24</sup> we achieved a superior mean GM DSC of 0.86 versus 0.80 performed by the deepSeg (<https://pypi.org/project/deepSeg/>) algorithm in the SC GM segmentation challenge dataset. This achievement could be potentially explained by the high quality of the AMIRA images and/or the use of a multicenter dataset within the challenge with results from various MR imaging sequences and segmentation algorithms. Application of the previously published iterative nonlocal STAPLE algorithm on our AMIRA images showed higher accuracy than the original work of Asman et al<sup>23</sup> (SC GM: median DSC of 0.82 versus 0.75, median HD of 1.04 versus 2.5 mm), which was performed on T2\*-weighted 3D gradient-echo images. While our atlases were constructed from a pool of around 800 samples, Asman et al had around 2000 available slices. Thus, the better accuracy seen here can be explained by a possible higher image quality in AMIRA images compared with T2\*-weighted 3D gradient-echo images; however, a direct comparison of MR images within the same subjects was not performed. The shallow architecture of the proposed algorithm with only a few parameters may make it less prone to overfitting to the training set compared with a state-of-the-art deep neural network. However, a direct comparison of our method with deepSeg<sup>28</sup> was not possible in this study.

Precision and accuracy of our automatic method was decreased in the order of total SC→WM→GM. This decrease may be caused by the accordingly decreasing size of WM and GM compared with total SC because small differences may be translated into a larger variance. Moreover, the more complex geometry of the GM and WM compared with total SC may be more prone to misclassification errors. Finally, despite the good image quality, signal contrast was stronger for SC/CSF compared with GM/WM, which, in turn, could have partly contributed to differences in total SC and GM segmentation. Moreover, a slightly lower reproducibility and accuracy of our measurements in more caudally acquired slices could also be identified, which may reflect a decrease in contrast intensity and a “noise” increase in AMIRA images acquired closer to the lungs and surrounded by overall greater body mass (thorax, shoulders, and arms) compared with the more rostral cervical SC.



**FIG 6.** Examples of segmentations of representative patients with MS. The thick continuous line indicates automatic segmentation; the dashed line, manual reference standard. A, A 54-year-old female patient with MS. Rostral cervical SC slices of the C1/C2 level without focal lesions. Automatic segmentation highly corresponds to the manual reference standard. B, A 32-year-old male patient with MS. Rostral cervical SC slice of the C2 level with a focal posterolateral lesion fused with the left posterior gray matter horn. Automatic segmentation misclassifies the focal lesion as SC GM. C, A 33-year-old female patient with MS. A cervical SC slice of the C3/C4 level with a focal posterior lesion fusing with the posterior SC GM horns and the central SC GM commissure. Automatic segmentation misclassifies the focal lesion as SC GM and CSF.

Our automatic method also showed significantly lower intra-session than inter-session variability for all SC metrics. However, GM intra- and inter-session CVs were similar, with mean values ranging between 4.10% and 4.77%. Accordingly, our method demonstrated similar mean intra- and inter-session CVs between 2.54% and 2.95% for WM. We, therefore, conclude that patient repositioning only slightly influences GM and WM area measurements; this conclusion provides evidence for the suitability of our automatic segmentation method in longitudinal settings.

In our work, minimal contrast adjustment differences in our manual segmentation led to a marked decrease of reproducibility, especially in GM area quantifications, as shown in the manual intra- and interrater measurements (mean CV up to 19.18%). Because the proposed method is fully automatic and requires no user-software interaction, it is devoid of additional variation produced by intra- and interrater variability. Therefore, our method provides significant advantages in large datasets or multicenter studies and, as mentioned above, may also be valuable in the longitudinal evaluation of individual patients (eg, patients with MS).

Compared with the manual reference standard, the automatic method slightly overestimated total SC and WM area, while underestimating the GM area. This result might be due to different intensity-thresholding in the manual segmentation compared with the automatic method. The caudal GM area increase shown in Fig 3 can be explained by the increased volume of motor cells of the cervical SC enlargement in the GM ventral horns, which innervate the upper limb muscles.

Although a fully automatic segmentation was not feasible on 12% of acquired SC slices, a semiautomatic approach with manual total SC segmentation and fully automatic GM and WM segmentation could be performed on those slices. This approach also

showed high-accuracy measurements with mean DSC of  $\geq 0.83$  in both GM and WM. However, compared with the fully automatic method on the initially nondiscarded slices, a slight accuracy decrease was observed, which could be interpreted in terms of a lower image quality of those AMIRA images. Nevertheless, these results demonstrate a relative robustness of our automatic approach even in MR images of suboptimal quality, which are a rather common phenomenon in clinical routine.

The present work focused on SC GM and WM segmentation using AMIRA images of healthy controls. Nevertheless, the motivation of our research is to deploy this method in patient data (eg, patients with MS) for the development of a potential widely applied MR imaging biomarker. Exemplary segmentation of data of patients with MS (not shown in detail here) showed that lesion appearance was similar to that of GM and therefore challenged the algorithm where lesions did not respect the GM boundaries (Fig 6). In future work, we intend to adjust the current method to address its current limitations. As an alternative approach, we plan to apply a deep learning-based segmentation approach on pathologic images as already performed on the data of healthy subjects.<sup>28,29</sup>

## CONCLUSIONS

The AMIRA sequence is presented as a time-efficient and reproducible MR imaging approach within the cervical cord. Our fully automatic segmentation method for SC GM and WM demonstrated further high reproducibility and accuracy. We were able to show that a shallow algorithm produces state-of-the-art GM-WM segmentation results on the AMIRA data. It is therefore suitable in large longitudinal studies investigating upper cervical SC volumes. Reproducibility measures of this work could be further



used for effect size calculations of SC compartment metrics for studies using the same processing approach. In future work, we will address the use of deep learning approaches, as demonstrated in recent studies.<sup>28</sup>

Disclosures: Charidimos Tsagkas—RELATED: Grant: Swiss National Science Foundation (grant No. 320030\_156860) and the Foundation for Sponsorship of Gastroenterological and General Clinical Research as well as of Medical Imaging (application ID 02/2015); UNRELATED: Grants/Grants Pending: University of Basel, Forschungsfonds für exzellente Nachwuchsforschende, pending. Antal Horvath—RELATED: Support for Travel to Meetings for the Study or Other Purposes: Swiss National Science Foundation.\* Simon Pezold—UNRELATED: Grants/Grants Pending: Novartis Research Foundation.\* Matthias Weigel—RELATED: Grant: Swiss National Science Foundation, Comments: grant No. 320030\_156860, Method Development for MRI of Spinal Cord\*; UNRELATED: Employment: Department of Biomedical Engineering, University of Basel, Switzerland, Comments: postdoctorate; Patents (Planned, Pending or Issued): University Hospital Freiburg, Freiburg, Germany, Comments: patent for Optical Motion Correction, neither relevant nor used in the present work; Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: International Society for Magnetic Resonance in Medicine, European Society for Magnetic Resonance in Medicine and Biology, MR Compact, Comments: travel reimbursement and accommodation for invited, scientific talks about MRI physics. Ludwig Kappos—UNRELATED: Board Membership: Actelion, Almirall, Bayer AG, Biogen, Celgene/Receptos, df-mp, Excemed, Genzyme, Merck, Novartis, Pfizer, Roche, Sanofi-Aventis, Tanabe Pharma, Novartis, Roche, Sanofi-Aventis, Santhera, Teva Pharmaceutical Industries, Vianex\*; Consultancy: Actelion, Almirall, Bayer AG, Biogen, Celgene/Receptos, df-mp, Excemed, Genzyme, Japan Tobacco, Merck, Minoryx, Mitsubishi Tanabe Pharma, Novartis, Roche, Sanofi-Aventis, Santhera, Teva Pharmaceutical Industries, Vianex\*; Grants/Grants Pending: Bayer AG, Biogen, Novartis, Roche, the Swiss MS Society, the Swiss National Research Foundation\*; Payment for Lectures Including Service on Speakers Bureaus: Allergan, Almirall, Baxalta, Bayer AG, Biogen, CSL Behring, Desitin, Excemed, Genzyme, Merck, Novartis, Pfizer, Roche, Sanofi-Aventis, Teva Pharmaceutical Industries\*; Payment for Development of Educational Presentations: Allergan, Almirall, Baxalta, Bayer AG, Biogen, CSL Behring, Desitin, Excemed, Genzyme, Merck, Novartis, Pfizer, Roche, Sanofi-Aventis, Teva Pharmaceutical Industries.\* Till Sprenger—RELATED: Grant: Swiss National Science Foundation\*; UNRELATED: Consultancy: Actelion, ATI, Biogen, Desitin, electroCore, Sanofi Genzyme, Novartis, and Mitsubishi Tanabe Pharma Europe\*; Grants/Grants Pending: EFIC-Grüenthal, Novartis Switzerland, the Swiss Multiple Sclerosis Society\*; Payment for Lectures Including Service on Speakers Bureaus: Sanofi Genzyme, Biogen, Novartis, Teva Pharmaceutical Industries.\* Oliver Bieri—RELATED: Grant: Swiss National Science Foundation\*; UNRELATED: Grants/Grants Pending: Swiss National Science Foundation.\* Katrin Parmar—RELATED: Grant: Swiss National Science Foundation\*; UNRELATED: Grants/Grants Pending: Baasch Medicus Foundation; Payment for Lectures Including Service on Speakers Bureaus: Excemed, Novartis Switzerland\*; Travel/Accommodations/Meeting Expenses Unrelated to Activities Listed: Novartis Switzerland.\* Money paid to the institution.

## REFERENCES

- Ginsberg L. Disorders of the spinal cord and roots. *Pract Neurol* 2011;11:259–67 CrossRef Medline
- Cook SA. Spinal disease: neoplastic, degenerative, and infective spinal cord diseases and spinal cord compression. April 10, 2015. <https://clinicalgate.com/spinal-disease-neoplastic-degenerative-and-infective-spinal-cord-diseases-and-spinal-cord-compression/>. Accessed April 4, 2015
- Wheeler-Kingshott CA, Stroman PW, Schwab JM, et al. The current state-of-the-art of spinal cord imaging: applications. *Neuroimage* 2014;84:1082–93 CrossRef Medline
- Kearney H, Miller DH, Ciccarelli O. Spinal cord MRI in multiple sclerosis: diagnostic, prognostic and clinical value. *Nat Rev Neurol* 2015;11:327–38 CrossRef Medline
- Schlaeger R, Papinutto N, Panara V, et al. Spinal cord gray matter atrophy correlates with multiple sclerosis disability. *Ann Neurol* 2014;76:568–80 CrossRef Medline
- Paquin MÊ, Mendili MM, Gros C, et al. Spinal cord gray matter atrophy in amyotrophic lateral sclerosis. *AJNR Am J Neuroradiol* 2018;39:184–92 CrossRef Medline
- Stroman PW, Bosma RL. Spinal cord imaging. In: Weidner N, Rupp R; Tansey K, eds. *Neurological Aspects of Spinal Cord Injury*. Cham: Springer International Publishing; 2017:237–57
- Stroman PW, Wheeler-Kingshott C, Bacon M, et al. The current state-of-the-art of spinal cord imaging: methods. *Neuroimage* 2014;84:1070–81 CrossRef Medline
- Yiannakas MC, Kearney H, Samson RS, et al. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: a pilot study with application to magnetisation transfer measurements. *Neuroimage* 2012;63:1054–59 CrossRef Medline
- Held P, Dorenbeck U, Seitz J, et al. MRI of the abnormal cervical spinal cord using 2D spoiled gradient echo multiecho sequence (MEDIC) with magnetization transfer saturation pulse: a T2\* weighted feasibility study. *J Neuroradiol* 2003;30:83–90 Medline
- Papinutto N, Schlaeger R, Panara V, et al. 2D phase-sensitive inversion recovery imaging to measure in vivo spinal cord gray and white matter areas in clinically feasible acquisition times. *J Magn Reson Imaging* 2015;42:698–708 CrossRef Medline
- Weigel M, Bieri O. A simple and fast approach for spinal cord imaging at 3T with high in-plane resolution and good contrast. In: *Proceedings of the Annual Meeting of International Society of Magnetic Resonance in Medicine*, Singapore; May 7–13, 2016:4408
- Weigel M, Bieri O. Spinal cord imaging using averaged magnetization inversion recovery acquisitions. *Magn Reson Med* 2018;79:1870–81 CrossRef Medline
- Prados F, Cardoso MJ, Yiannakas MC, et al. Fully automated grey and white matter spinal cord segmentation. *Sci Rep* 2016;6:36151 CrossRef Medline
- Datta E, Papinutto N, Schlaeger R, et al. Gray matter segmentation of the spinal cord with active contours in MR images. *Neuroimage* 2017;147:788–99 CrossRef Medline
- Dupont SM, De Leener B, Taso M, et al. Fully-integrated framework for the segmentation and registration of the spinal cord white and gray matter. *Neuroimage* 2017;150:358–72 CrossRef Medline
- De Leener B, Lévy S, Dupont SM, et al. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *Neuroimage* 2017;145:24–43 CrossRef Medline
- Bach Cuadra M, Duay V, Thiran JP. Atlas-based segmentation. In: Paragios N, Duncan J, Ayache N, eds. *Handbook of Biomedical Imaging: Methodologies and Clinical Research*. Boston: Springer US; 2015:221–44
- Wachinger C, Golland P. Atlas-based under-segmentation. *Med Image Comput Assist Interv* 2014;17:315–22 Medline
- Horváth A, Pezold S, Weigel M, et al. Variational segmentation of the white and gray matter in the spinal cord using a shape prior. In: Yao J, Vrtovc T, Zheng G, et al, eds. *Computational Methods and Clinical Applications for Spine Imaging*. Cham: Springer International Publishing; 2016:26–37
- Pezold S, Fundana K, Amann M, et al. Automatic segmentation of the spinal cord using continuous max flow with cross-sectional similarity prior and tubularity features. In: Yao J, Glocker B, Klinder T, et al, eds. *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*. Cham: Springer; 2015:107–18
- Lüthi M, Albrecht T, Vetter T. Probabilistic modeling and visualization of the flexibility in morphable models. In: Hancock ER, Martin RR, Sabin MA, eds. *Mathematics of Surfaces XIII*. Berlin: Springer-Verlag; 2009:251–64
- Asman AJ, Smith SA, Reich DS, et al. Robust GM/WM segmentation of the spinal cord with iterative non-local statistical fusion. *Med Image Comput Assist Inter* 2013;16(Pt 1):759–67 Medline
- Prados F, Ashburner J, Blaiotta C, et al. Spinal cord grey matter segmentation challenge. *Neuroimage* 2017;152:312–29 CrossRef Medline
- Bhadelia RA, Bogdan AR, Kaplan RF, et al. Cerebrospinal fluid pulsation amplitude and its quantitative relationship to cerebral blood flow pulsations: a phase-contrast MR flow imaging study. *Neuroradiology* 1997;39:258–64 CrossRef Medline
- Figley CR, Stroman PW. Investigation of human cervical and upper thoracic spinal cord motion: implications for imaging spinal cord



**structure and function.** *Magn Reson Med* 2007;58:185–89 CrossRef Medline

27. Piché M, Cohen-Adad J, Nejad MK, et al. **Characterization of cardiac-related noise in fMRI of the cervical spinal cord.** *Magn Reson Imaging* 2009;27:300–10 CrossRef Medline
28. Perone CS, Calabrese E, Cohen-Adad J. **Spinal cord gray matter seg-**

**mentation using deep dilated convolutions.** *Sci Rep* 2018;8:5966 CrossRef Medline

29. Horvath A, Tsagkas C, Andermatt S, et al. **Spinal cord gray matter-white matter segmentation on magnetic resonance AMIRA images with MD-GRU.** <https://arxiv.org/abs/1808.02408>. Accessed August 7, 2018



## 8 Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA Images with MDGRU

The approach presented in this chapter describes our first attempt in applying a supervised deep-learning framework to the CGM dataset of healthy subjects (cf. Page 88). In particular, we chose a recurrent neural network with multidimensional gated recurrent units (MDGRU) [24], which was designed for the segmentation of biomedical data and proved excellent performance in GM-WM segmentation. We applied MDGRU to the 2D images of the CGM dataset slicewise, directly feeding the 8 AMIRA inversion images. Moreover, for the training phase, we modified MDGRU’s cross-entropy loss (CEL) by proposing and including a generalized Dice loss (GDL) with automated class weighting [151].

To demonstrate the architecture’s capability of generalization, we applied the proposed approach to the spinal cord gray matter segmentation (SCGM) challenge data [14]. We set a new state of the art in this challenge with a significant margin between the mean GM DSC of 0.90 achieved by the proposed approach compared to 0.85 of the predecessor, see Table 2 of this chapter’s publication. The proposed method furthermore yields mean GM Dice coefficients of  $0.92 \pm 0.03$  on the AMIRA images as compared to  $0.90 \pm 0.03$  on the challenge data. This difference may point towards AMIRA’s favorable image quality compared to the quality of the common MR sequences gathered for the SCGM challenge.

The use of a Dice loss was inspired by SCT’s ”deepseg” network of Perone et al. [138]. They showed that the implementation of a Dice loss in the network’s learning loss not only improves the testing scores in terms of DSC but also leads to sharper predicted probability maps calculated through softmax, cf. Equation (2.49). Sharper probability maps have less uncertainty in the labels’ boundary regions, as depicted in Fig. 2 of this chapter’s publication. The publication focuses on giving an ablation analysis on the combination with GDL: first, the combination of CEL and GDL together performs better than the versions with either only CEL or only GDL; second, the automated class weighting establishes a comfortable way of dealing with class imbalance; third, the convex linear combination of CEL and GDL is quite stable because of GDL’s linear scale and CEL’s logarithmic scale; fourth, there is no noticeable difference between the multi-label Dice loss and its generalized version, except

in the implementation of the generalized version having fewer regularization terms to prevent divisions by zero.

We, however, did not include an ablation analysis on the effect of feeding the network with the 8 inversion images directly rather than providing only a projected mean image. Experiments revealed inferior performance when feeding only one mean image. We believe that directly feeding the 8 AMIRA inversion images enables the network to learn the tissue-specific MR relaxation curves shown in Figure 3.4.

Finally but importantly, the estimated SC, WM, and GM cross-sectional areas of the manual segmentations as well as those of the proposed automatic approach correspond to similar values reported in the literature, such as [152], [153, Tab. 2], [154, Fig. 8], [13, Tab. 1], and [21, Tab. 4].


**Publication** The publication was written with joint first authorship by Charidimos Tsagkas and Antal Horváth, with A. Horváth developing and applying the proposed method and performing the statistical analysis, and C. Tsagkas focusing on the clinical setup and providing manual segmentations. The paper was presented at the workshop *Computational Methods and Clinical Applications for Spine Imaging* (CSI 2018) in conjunction with the *21<sup>st</sup> International Conference on Medical Image Computing and Computer Assisted Intervention* (MICCAI) on the 16<sup>th</sup> of September 2018 in Granada, Spain. It was published<sup>1</sup> as part of the workshop proceedings [146].

---

<sup>1</sup>[https://link.springer.com/chapter/10.1007/978-3-030-13736-6\\_1](https://link.springer.com/chapter/10.1007/978-3-030-13736-6_1) (last accessed on Sept. 10, 2019)



# Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA Images with MD-GRU

Antal Horváth<sup>1</sup> , Charidimos Tsagkas<sup>2</sup>, Simon Andermatt<sup>1</sup>, Simon Pezold<sup>1</sup>,  
Katrín Parmar<sup>2</sup>, and Philippe Cattin<sup>1</sup>

<sup>1</sup> Department of Biomedical Engineering, University of Basel, Allschwil, Switzerland  
[antal.horvath@unibas.ch](mailto:antal.horvath@unibas.ch)

<sup>2</sup> Department of Neurology, University Hospital Basel, Basel, Switzerland

**Abstract.** The small butterfly shaped structure of spinal cord (SC) gray matter (GM) is challenging to image and to delineate from its surrounding white matter (WM). Segmenting GM is up to a point a trade-off between accuracy and precision. We propose a new pipeline for GM-WM magnetic resonance (MR) image acquisition and segmentation. We report superior results as compared to the ones recently reported in the SC GM segmentation challenge and show even better results using the averaged magnetization inversion recovery acquisitions (AMIRA) sequence. Scan-rescan experiments with the AMIRA sequence show high reproducibility in terms of Dice coefficient, Hausdorff distance and relative standard deviation. We use a recurrent neural network (RNN) with multi-dimensional gated recurrent units (MD-GRU) to train segmentation models on the AMIRA dataset of 855 slices. We added a generalized dice loss to the cross entropy loss that MD-GRU uses and were able to improve the results.

**Keywords:** Segmentation · Spinal cord · Gray matter · White matter · Deep learning · RNN · MD-GRU

## 1 Introduction

Cervical spinal cord (SC) segmentation in magnetic resonance (MR) images is a viable means for quantitatively assessing the neurodegenerative effects of diseases in the central nervous system. While conventional MR sequences only allowed differentiation of the boundary between SC and cerebrospinal fluid (CSF), more recent sequences can be used to distinguish the SC's inner gray matter (GM) and white matter (WM) compartments. The latter task, however, remains challenging as state-of-the-art MR sequences only achieve an in-slice resolution of around 0.5 mm while maintaining a good signal-to-noise ratio (SNR) and an

---

A. Horváth and C. Tsagkas—These two authors contributed equally.

4 A. Horváth et al.

acceptable acquisition time. This resolution is barely enough to visualize the SC’s butterfly-shaped GM structure.

The 2016 spinal cord gray matter segmentation (SCGM) challenge [7] reported mean Dice similarity coefficients (DSC) of 0.8 in comparison to a manual consensus ground truth for the best SC GM segmentation approaches at that time. Porisky et al. [6] experimented with 3D convolutional encoder networks but did not improve the challenge’s results. Perone et al.’s U-Net approach [5] later managed to push the DSC value to 0.85. More recently, Datta et al. [3] reported mean DSC of 0.88 on images of various MR sequences with a morphological geodesic active contour model.

Still, this means that a high number of subjects would be necessary to get reliable findings from clinical trials. Hence, despite recent developments, there is a need for improvement of the reproducibility of SC GM and WM measurements. An accurate and precise segmentation of the SC’s inner structures in MR images under the mentioned limiting trade-off between resolution, SNR, and time therefore remains a challenge, especially when focusing on the GM.

In this work, we present a new robust and fully automatic pipeline for the acquisition and segmentation of GM and WM in MR images of the SC. On the segmentation side, we propose the use of multi-dimensional gated recurrent units (MD-GRU), which already proved fit for a number of medical segmentation tasks [1], to gain accurate and precise SC GM and WM segmentations. To this end, we adapt MD-GRU’s original cross-entropy loss by integrating a generalized Dice loss (GDL) [8] and show improved segmentation performance compared to the original. Using the proposed setup, we manage to set a new state of the art on the SCGM challenge data with a mean DSC of 0.9. On the imaging side, we propose to use the AMIRA MR sequence [9] for gaining improved GM-WM and WM-CSF contrast in axial cross-sectional slices of the SC. Using the proposed MD-GRU approach in combination with this new imaging sequence, we manage to gain an even higher accuracy of DSC 0.91 wrt. a manual ground truth, as we demonstrate in experiments on scan-rescan images of healthy subjects, for both SC GM and WM.

The remaining paper is structured as follows: in Sect. 2, we present our segmentation method; in Sect. 3, we briefly describe the AMIRA MR sequence and the two datasets (SCGM challenge, AMIRA images) that we use for the experiments of Sect. 4, before we conclude in Sect. 5.

## 2 Method

The Multi-Dimensional Gated Recurrent Unit (MD-GRU) [1] is a generalization of a bi-directional recurrent neural network (RNN), which is able to process images. It achieves this task by treating each direction along each of the spatial dimensions independently as a temporal direction. The MD-GRU processes the image using two convolutional GRUs (C-GRUs) for each image dimension, one in forward and one in backward direction, and combines the results of all individual C-GRUs. The gated recurrent unit (GRU), compared to the more popular and

established long short-term memory (LSTM), uses a simpler gating structure and combines its state and output. The GRU has been shown to produce comparable results while consuming less memory than its LSTM counterpart when applied to image segmentation and hence allows for larger images to be processed [1].

We directly feed the 2D version of MD-GRU the 8-channel AMIRA images (cf. Sect. 3.1) to train AMIRA segmentation models, but only use the single channel images of the SCGM dataset (cf. Sect. 3.2) for the challenge models. To address the high class imbalance between background, WM and GM, similar to [5] we added a GM Dice loss (DL), but also included DLs for all the other label classes using the generalized Dice loss (GDL) formulation of Sudre et al. [8].

## 2.1 Dice Loss

A straightforward approximation of a DL for a multi-labelling problem is

$$L_D = - \frac{1}{\sum_{\ell \in \mathcal{L}} \omega_\ell} \sum_{\ell \in \mathcal{L}} \omega_\ell \frac{2 \sum_{x \in X} p_{\ell x} r_{\ell x}}{\sum_{x \in X} p_{\ell x} + r_{\ell x}}, \quad (1)$$

with the image domain  $X$ , labels  $\mathcal{L}$ , predictions  $p$ , raters  $r$ , and class weights  $\omega$ . Sudre et al. [8] described a Generalized Dice Loss (GDL)  $L_{GD}$  where they divide the weighted sum of the intersections of all labels by the weighted sum of all predictions and targets of all labels, instead of just linearly combining the individual Dice coefficients:

$$L_{GD} = - \frac{2 \sum_{\ell \in \mathcal{L}} \omega_\ell \sum_{x \in X} p_{\ell x} r_{\ell x}}{\sum_{\ell \in \mathcal{L}} \omega_\ell \sum_{x \in X} p_{\ell x} + r_{\ell x}}. \quad (2)$$

As stated in [2], compared to the DL (1), the GDL (2) allows all labels to contribute equally to the overall overlap (denominator in (2)).

The (squared) inverse volume weighting

$$\omega_\ell = \frac{1}{\left(\sum_{x \in X} r_{\ell x}\right)^2}, \quad (3)$$

as proposed in [2], deals with the class imbalance problem: large regions only contribute very little to  $L_D$  or  $L_{GD}$ , whereas small regions are weighted more and thus are more important in the optimization process.

To avoid division by zero in  $\omega_\ell$  for image samples with absence of label  $\ell$ , we regularize the denominator of (3) and formulate the weighting we used:

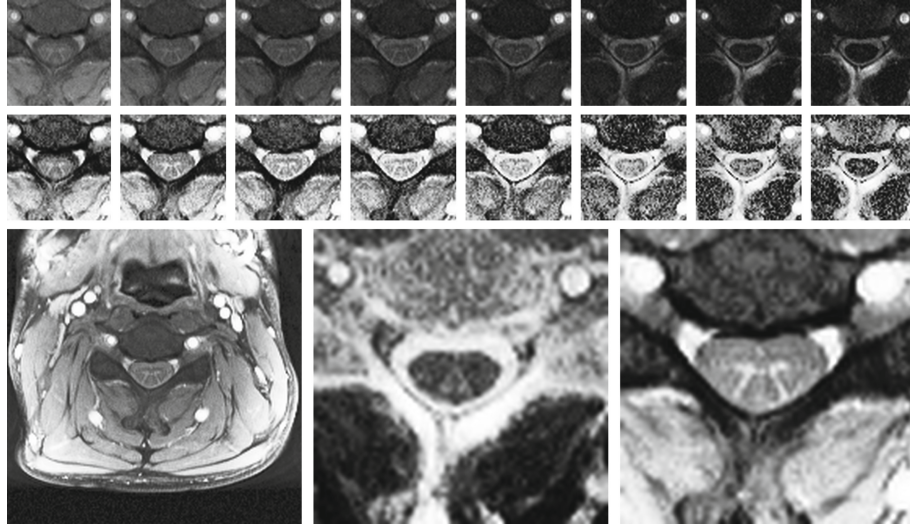
$$\omega_\ell = \frac{1}{1 + \left(\sum_{x \in X} r_{\ell x}\right)^2}. \quad (4)$$

The weighting (4) compared to (3) only slightly decreases its value as long as the object of interest has enough pixels. Note, that during training of a network, it is possible, that not all labels occur in a random subsample with random location.

Finally, we combine DL or GDL with the cross entropy loss  $L_C$  (CEL) with a factor  $\lambda \in [0, 1]$ :

$$L = \lambda L_{D \text{ or } GD} + (1 - \lambda) L_C.$$

6 A. Horváth et al.



**Fig. 1.** AMIRA sequence of an exemplary slice on C4 level. All images 10-fold upsampled. *Top and middle row:* Inversion images with increasing inversion times from left to right. Original cropped images (*top*), and histogram equalized (*middle*). *Bottom row:* Histogram equalized sum of the first 5 inversion images in full view (*left*), weighted average with optimal CSF-WM contrast (*middle*), and optimal GM-WM contrast (*right*).

### 3 Data

In the following subsections, we describe the images used for the experiments: healthy subjects scan-rescan AMIRA dataset (own), which we call the AMIRA dataset, and the SCGM challenge dataset<sup>1</sup> [7], which we refer to as SCGM dataset.

#### 3.1 AMIRA Dataset

The first dataset used in this paper consists of 24 healthy subjects (14 female, 10 male, age  $40 \pm 11$  years). Each subject was scanned 3 times, remaining in the scanner between the first and second scan, and leaving the scanner and being repositioned between the second and third scan. Each scan contains 12 axial cross-sectional slices of the neck acquired with the AMIRA sequence [9] that were manually aligned at acquisition time perpendicular to the SC's centerline with an average slice distance of 4 mm starting from vertebra C3 level in caudal direction.

<sup>1</sup> <http://cmictig.cs.ucl.ac.uk/niftyweb/program.php?p=CHALLENGE> last accessed: September 13, 2018.



Because of severe imaging artifacts some slices had to be discarded. For one scan the last three caudal slices, for two scans the last two slices and for another two scans the last slice, in total 9 out of the 864 slices were discarded.

The AMIRA sequence consists of 8 inversion images of the same anatomical slice captured at different inversion times after  $180^\circ$  MR pulses that have an in-slice resolution of  $0.67 \text{ mm} \times 0.67 \text{ mm}$ . Exemplary inversion images and different averages of an exemplary slice on vertebra C4 level are shown in Fig. 1. For human raters, to manually segment the AMIRA images, different single channel projections of the 8 channel images are necessary. Weighted averages of the inversion images with e.g. optimal CSF-WM or GM-WM contrast, see Fig. 1, were calculated with an approach that maximizes between-class intensity mean values and minimizes within-class intensity variances [4].

In order to reduce the numerical errors for the calculated measures, we 10-fold upsampled all slices with Lanczos interpolation. Since all images were manually centered at the SC, we consequently trimmed one third of the image size on each side and thus cropped out the inner ninth to a size of  $650 \times 650$  pixels for faster processing.

One experienced rater segmented all 855 images manually for WM and GM and segmented again 60 randomly chosen slices over all subjects, scans and slices, without knowledge of their origin, to enable an intra-rater comparison.

### 3.2 SCGM Dataset

The SCGM segmentation challenge data [7] consists of 40 training datasets and 40 test datasets acquired at 4 different sites. Both training and test datasets each have 10 samples of each site. The 4 sites have different imaging protocols with different field of view, size and resolution. Each dataset was manually segmented by 4 experts and to assess rater performance, with majority voting (more than 2 positive votes) a consensus segmentation of the 4 raters was calculated.

For training and testing of our MD-GRU models, we resampled all axial slices of all the datasets to the common finest resolution of  $0.25 \text{ mm} \times 0.25 \text{ mm}$  and center cropped or padded all datasets to a common size of  $640 \times 640$  pixels. Before submitting the testing results for evaluation, we padded and resampled all slices to their original sizes and resolutions.

## 4 Experiments and Results

In the following subsections, we describe our experiments, the chosen MD-GRU options, and show their results.

### 4.1 AMIRA Segmentation Model

We split the 24 subjects into 3 groups of 8 subjects each for 3 cross-validations: training on two groups and testing on a third group. To handle over-fitting, of each training set we excluded one subject and used it for validation.

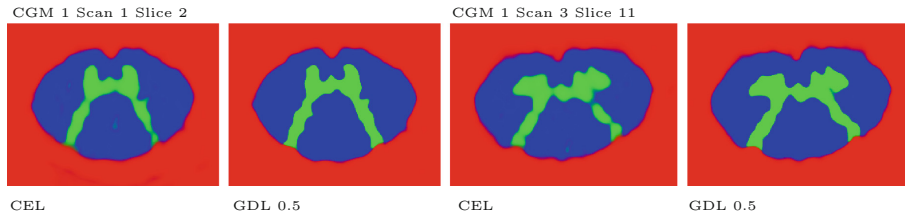
8 A. Horváth et al.

We used the standard MD-GRU<sup>2</sup> model with default settings and residual learning, dropout rate 0.5, and dropconnect on state. We chose the following problem specific parameters: Gaussian high pass filtering with variance 10, batch size 1, and window size  $500 \times 500$  pixels. In each iteration of the training stage, for data augmentation, a subsample of the training data with random deformation field at a random location was selected. Random deformations included an interpolated deformation field on 4 supporting points with randomly generated deformations of standard deviation of 15, random scaling of a factor between  $4/5$  and  $5/4$ , random rotation of  $\pm 10^\circ$ , and random mirroring along the anatomical median plane. To prevent zero padding of the subsamples, we only allowed random sampling within a safe distance of 45 pixels from the image boundary and truncated the random deformation magnitudes to 45 pixels, which is 3 times the chosen standard deviation.

We trained the networks with Adadelta with a learning rate of 1 for 30'000 iterations, where one iteration approximately took 10s on an NVIDIA GeForce GTX Titan X. Cross entropy and DSC on the evaluation set already reached their upper bounds after around 20'000 iterations, and dropconnect on state prevented from overfitting as we can see in Fig. 3.

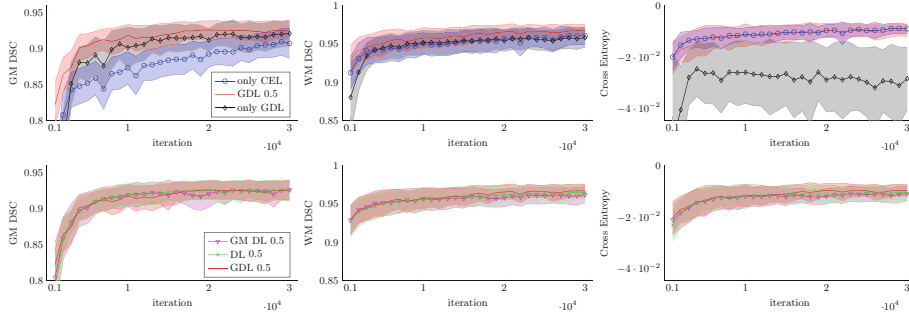
The time for segmenting a slice with the trained network approximately took 7 s.

Prior to the final model generation, we experimented in adding only a GM DL to the CEL with weightings  $\lambda = 0, 0.25, 0.5, 0.75, 1$  and figured that 0.5 produced the best results. DL produces values close to  $-1$  whereas CEL tends to have small values close to 0. Moreover, CEL holds the information of all labels, since it is calculated over all labels. Now, when adding only GM DL, because of the imbalance of the loss values, higher values of  $\lambda$  strongly weaken the information for WM and background that in this setup is carried only within CEL. The best weighting  $\lambda$  depends on the cross entropy and thus depends on the class imbalance and label uncertainty of each specific segmentation task.



**Fig. 2.** Exemplary prediction probability maps of the three labeling maps background (red), GM (green) and WM (blue) of MD-GRU with CEL and with GDL in RGB colors. (Color figure online)

<sup>2</sup> <https://github.com/zubata88/mdgru> last accessed: September 13, 2018.



**Fig. 3.** GM DSC, WM DSC and cross entropy over the training iterations of the validation set of group 1 in the AMIRA dataset in the format mean  $\pm$  one standard deviation. *Top row:* models with  $\lambda = 0$  (only CEL),  $\lambda = 1$  (only GDL), and combined with  $\lambda = 0.5$  (GDL 0.5). *Bottom row:* GM DL 0.5, DL 0.5, and GDL 0.5 show similar performance.

We observed that the auxiliary DL produces sharper probability maps at the boundaries as compared to only using CEL, see Fig. 2, and that DL helps to delineate weak contrasts e.g. between GM and WM.

Further experiments showed, that the proposed automatic weightings  $\omega_\ell$  (4) for the DLs between all label classes is a good strategy to simplify the selection of  $\lambda$ . In our case, the evaluation scores did not show big differences for  $\lambda$  in a range from 0.25 to 0.75, when using the class weights  $\omega_\ell$  according to (4) for both DL and GDL. MD-GRU with the trivial linear combinations  $\lambda = 0$  (only CEL) and  $\lambda = 1$  (only GDL) did not perform as good as true combinations between the two losses. We show the improvement in the scores of GDL with  $\lambda = 0.5$  in Fig. 3 and Table 1.

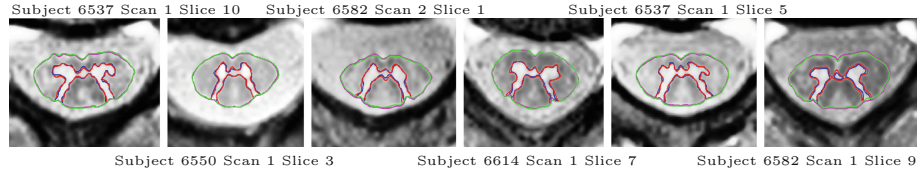
**Table 1.** Improvement between native MD-GRU with CEL and the proposed MD-GRU with GDL together with the manual segmentation’s precision and intra-rater accuracy values. Intra-rater accuracy of the human expert was calculated for the 60 randomly chosen slices.

GM	Accuracy		Intra-session			Inter-session		
	DSC	HD(mm)	DSC	HD(mm)	RSD(%)	DSC	HD(mm)	RSD(%)
MD-GRU CEL	0.90 $\pm$ 0.04	0.68 $\pm$ 0.43	0.89 $\pm$ 0.03	0.71 $\pm$ 0.46	3.22 $\pm$ 2.87	0.88 $\pm$ 0.04	0.70 $\pm$ 0.43	3.65 $\pm$ 3.97
MD-GRU GDL 0.5	0.91 $\pm$ 0.03	0.56 $\pm$ 0.33	0.88 $\pm$ 0.03	0.58 $\pm$ 0.32	2.93 $\pm$ 2.63	0.88 $\pm$ 0.03	0.61 $\pm$ 0.35	3.86 $\pm$ 3.49
Manual			0.86 $\pm$ 0.03	0.67 $\pm$ 0.24	5.55 $\pm$ 4.11	0.85 $\pm$ 0.03	0.71 $\pm$ 0.27	6.27 $\pm$ 4.70
Intra-rater	0.85 $\pm$ 0.07	0.62 $\pm$ 0.30						
WM								
	DSC	HD(mm)	DSC	HD(mm)	RSD(%)	DSC	HD(mm)	RSD(%)
MD-GRU CEL	0.94 $\pm$ 0.03	0.47 $\pm$ 0.26	0.94 $\pm$ 0.02	0.51 $\pm$ 0.25	2.07 $\pm$ 2.16	0.94 $\pm$ 0.02	0.52 $\pm$ 0.22	2.40 $\pm$ 2.22
MD-GRU GDL 0.5	0.95 $\pm$ 0.02	0.43 $\pm$ 0.22	0.94 $\pm$ 0.02	0.51 $\pm$ 0.22	2.14 $\pm$ 2.35	0.94 $\pm$ 0.02	0.53 $\pm$ 0.23	2.69 $\pm$ 2.54
Manual			0.93 $\pm$ 0.02	0.54 $\pm$ 0.13	3.78 $\pm$ 3.32	0.92 $\pm$ 0.02	0.58 $\pm$ 0.15	4.59 $\pm$ 3.77
Intra-rater	0.96 $\pm$ 0.02	0.44 $\pm$ 0.15						

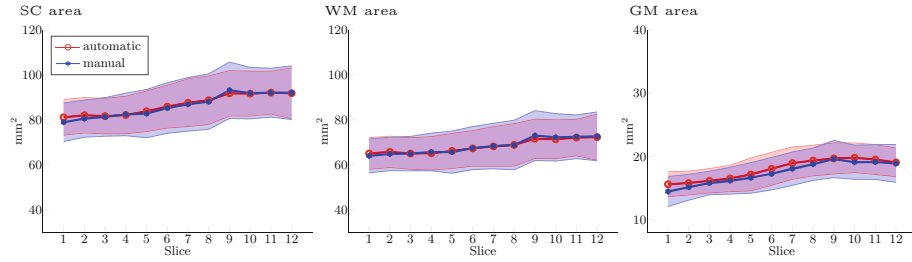
10 A. Horváth et al.

Finally, comparisons between GM DL 0.5, auto-weighted DL 0.5 and GDL 0.5, all with  $\lambda = 0.5$ , are shown in Fig. 3 on the bottom row. As can be expected, the similarity of the terms DL (1) and GDL (2) is reflected in their almost identical segmentation performance.

GM DL 0.5 shows comparable WM segmentation performance to the losses that have WM DL included. This can be explained, because the GM boundary is part of the WM boundaries and thus influences the WM scores, and furthermore the outer WM boundary is already well delineated even without any DL through the good CSF-WM contrast. Choosing a DL as a surrogate for GM DSC only, as proposed in [5], is thus justifiable.

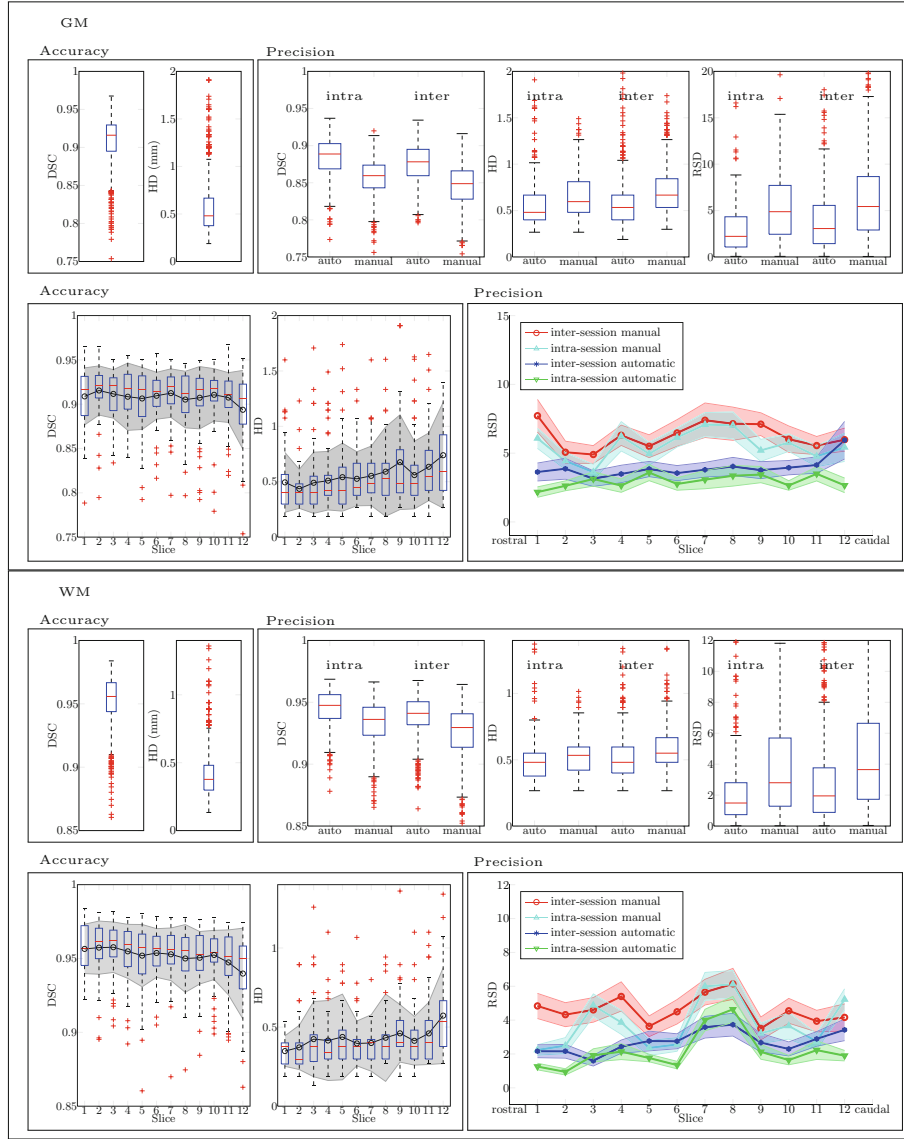


**Fig. 4.** Exemplary slices of the AMIRA dataset with automatic GM (*red*) and CSF-WM (*green*) boundaries, and manual GM (*blue*) and CSF-WM (*magenta*) boundaries. (Color figure online)



**Fig. 5.** SC, WM, and GM areas of GDL 0.5 (automatic) and manual segmentations wrt. the anatomical slice positions in mean  $\pm$  one standard deviation.

While the SCGM challenge results only provide GM segmentation accuracy, for the AMIRA dataset we additionally also provide WM segmentation results. For the statistics, we gathered all slice-wise test results of all cross-validations for the proposed method GDL 0.5 and compare it with those of CEL. Pairwise two-tailed Hotelling's T-tests for GM accuracy in DSC and labelmap Hausdorff distance (HD) show, that the test results of the MD-GRU models trained on the different groups are not significantly different from each other ( $p > 0.3$  for both GDL and CEL).



**Fig. 6.** GM and WM accuracy and precision plots of the AMIRA dataset. For both boxes GM and WM: *Top row:* Accuracy (*left*) in DSC and HD of all the 855 slices of the proposed method; intra-session (intra) and inter-session (inter) precision (*right*) of the proposed method (auto) and the manual segmentations in DSC, HD, and area RSD. *Bottom row:* Accuracy box plots (*left*) in DSC and HD wrt. the slice positions with overlaid error bars in the format mean  $\pm$  one standard deviation; precision error bars (*right*) for area RSD wrt. the slice positions, for better visualization shown with 0.2 standard deviations. HD is measured in millimeters, and RSD in percents.

In Fig. 6 and in Table 1 we show GM and WM accuracy and precision of all gathered slice results in DSC, HD and relative standard deviation of the areas (RSD), also known as coefficient of variation. With intra- and inter-session precision we compare segmentations of the same slice for different scans with and without repositioning, respectively. The proposed automatic segmentations shows better reproducibility as the manual segmentations. Additionally, we show the anatomical GM and WM areas wrt. the slice positions in Fig. 5 and show randomly chosen results in Fig. 4. Training multiple networks with data from multiple human raters as ground truth data, as we did with the SCGM data, cf. Subsect. 4.2, might further improve the performance.

#### 4.2 SCGM Challenge Model

To enable comparison with other methods, we tested MD-GRU on the SCGM dataset [7]. We trained four MD-GRU models, one for each expert rater's ground truth, and in the end performed majority voting on the individual test results to mimic the challenge's consensus segmentation.

We used the same MD-GRU setup but with a window size of  $200 \times 200$  pixels for a similar anatomical field of view as the AMIRA models. Random subsamples in each training iteration were drawn with a distance of 200 pixels from the image boundary. We trained the networks for 100'000 iterations and observed, that the scores reached their upper bounds after around 60'000 iterations. One training iteration took around 4s and segmentation of one slice took less than 1s.

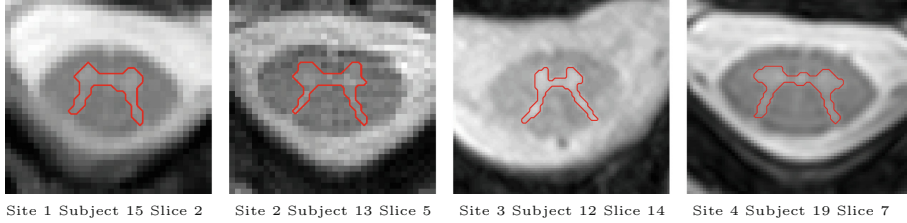
In Table 2, the proposed model shows a new state-of-the-art in almost all metrics. This comparison shows MD-GRU's strong performance in learning the GM segmentation problem. In Table 3, we additionally show the improvement for the auto-weighted GDL, compared to the native MD-GRU approach with only CEL. Figure 7 shows randomly chosen results of the proposed model.

**Table 2.** Results of the SCGM challenge competitors including the results of Porisky et al. [6], Perone et al. [5] and ours. The metrics are Dice coefficient (DSC), mean surface distance (MD), Hausdorff surface distance (HD), skeletonized Hausdorff distance (SHD), skeletonized median distance (SMD), true positive rate (TPR), true negative rate (TNR), precision (P), Jaccard index (J), and conformity (C). Best results on each metric are highlighted in bold font. Distances are measured in millimeters.

	JCSGS	DEEPPSEG	MGAC	GSBME	SCT	VBEM	[6]	[5]	Proposed
DSC	0.79 $\pm$ 0.04	0.80 $\pm$ 0.06	0.75 $\pm$ 0.07	0.76 $\pm$ 0.06	0.69 $\pm$ 0.07	0.61 $\pm$ 0.13	0.80 $\pm$ 0.06	0.85 $\pm$ 0.04	<b>0.90</b> $\pm$ 0.03
MD	0.39 $\pm$ 0.44	0.46 $\pm$ 0.48	0.70 $\pm$ 0.79	0.62 $\pm$ 0.64	0.69 $\pm$ 0.76	1.04 $\pm$ 1.14	0.53 $\pm$ 0.57	0.36 $\pm$ 0.34	<b>0.21</b> $\pm$ 0.20
HD	2.65 $\pm$ 3.40	4.07 $\pm$ 3.27	3.56 $\pm$ 1.34	4.92 $\pm$ 3.30	3.26 $\pm$ 1.35	5.34 $\pm$ 15.35	3.69 $\pm$ 3.93	2.61 $\pm$ 2.15	<b>1.85</b> $\pm$ 1.16
SHD	1.00 $\pm$ 0.35	1.26 $\pm$ 0.65	1.07 $\pm$ 0.37	1.86 $\pm$ 0.85	1.12 $\pm$ 0.41	2.77 $\pm$ 8.10	1.22 $\pm$ 0.51	0.85 $\pm$ 0.32	<b>0.71</b> $\pm$ 0.28
SMD	0.37 $\pm$ 0.18	0.45 $\pm$ 0.20	0.39 $\pm$ 0.17	0.61 $\pm$ 0.35	0.39 $\pm$ 0.16	0.54 $\pm$ 0.25	0.44 $\pm$ 0.19	<b>0.36</b> $\pm$ 0.17	0.37 $\pm$ 0.17
TPR	77.98 $\pm$ 4.88	78.89 $\pm$ 10.33	87.51 $\pm$ 6.65	75.69 $\pm$ 8.08	70.29 $\pm$ 6.76	65.66 $\pm$ 14.39	79.65 $\pm$ 9.56	94.97 $\pm$ 3.50	<b>96.22</b> $\pm$ 2.69
TNR	<b>99.98</b> $\pm$ 0.03	99.97 $\pm$ 0.04	99.94 $\pm$ 0.08	99.97 $\pm$ 0.05	99.95 $\pm$ 0.06	99.93 $\pm$ 0.09	99.97 $\pm$ 0.04	99.95 $\pm$ 0.06	<b>99.98</b> $\pm$ 0.03
P	81.06 $\pm$ 5.97	82.78 $\pm$ 5.19	65.60 $\pm$ 9.01	76.26 $\pm$ 7.41	67.87 $\pm$ 8.62	59.07 $\pm$ 13.69	81.29 $\pm$ 5.30	77.29 $\pm$ 6.46	<b>85.46</b> $\pm$ 4.96
J	0.66 $\pm$ 0.05	0.68 $\pm$ 0.08	0.60 $\pm$ 0.08	0.61 $\pm$ 0.08	0.53 $\pm$ 0.08	0.45 $\pm$ 0.13	0.67 $\pm$ 0.07	0.74 $\pm$ 0.06	<b>0.82</b> $\pm$ 0.05
C	47.17 $\pm$ 11.87	49.52 $\pm$ 20.29	29.36 $\pm$ 29.53	33.69 $\pm$ 24.23	6.46 $\pm$ 30.59	44.25 $\pm$ 90.61	48.79 $\pm$ 18.09	64.24 $\pm$ 10.83	<b>77.46</b> $\pm$ 7.31

**Table 3.** SCGM challenge results of the native MD-GRU with only CEL in comparison to the proposed GDL 0.5. Abbreviations of the metrics taken from Table 2.

	DSC	MD	HD	SHD	SMD	TPR	TNR	P	J	C
MD-GRU CEL	$0.87 \pm 0.03$	$0.30 \pm 0.31$	$2.14 \pm 1.20$	$0.85 \pm 0.36$	$0.40 \pm 0.20$	$93.93 \pm 3.85$	$99.98 \pm 0.03$	$82.04 \pm 5.42$	$0.78 \pm 0.05$	$70.90 \pm 9.06$
MD-GRU GDL 0.5	$0.90 \pm 0.03$	$0.21 \pm 0.20$	$1.85 \pm 1.16$	$0.71 \pm 0.28$	$0.37 \pm 0.17$	$96.22 \pm 2.69$	$99.98 \pm 0.03$	$85.46 \pm 4.96$	$0.82 \pm 0.05$	$77.46 \pm 7.31$



**Fig. 7.** For each site of the SCGM dataset, one randomly chosen result of the proposed model in cropped view.

## 5 Conclusion

We presented a new pipeline of acquisition and automatic segmentation of SC GM and WM. The AMIRA sequence produces 8 channel images for different inversion times which the proposed deep learning approach with MD-GRU used for segmentation. Using the 8 channels, tissue specific relaxation curves can be learned and used for GM-WM segmentation.

Comparing our segmentation results to the results of the ex-vivo high-resolution dataset of Perone et al. [5], we show comparable accuracy for in-vivo data. The acquired AMIRA dataset in scan-rescan fashion, with and without repositioning in the scanner, shows high reproducibility in terms of GM area RSD. Thus we believe that the presented pipeline is a candidate for longitudinal clinical studies. Further tests with patient data have to be conducted.

We added a generalized multi-label Dice loss to the cross entropy loss that MD-GRU uses. We observed, that the segmentation performance was stable for a larger region of the weighting  $\lambda$  between the two losses. In a future work, we will study the effects of small  $\lambda$ s that correspond well with the logarithmical magnitudes of CEL. Our proposed segmentation model outperforms the methods from the SC GM segmentation challenge. Training the MD-GRU models directly on the 3D data might further improve the performance compared to slice-wise segmentation.

Given the small and fine structure of the GM, we like to point out, that the achieved results of the metrics are near optimal. Higher resolutions of the imaging sequence will improve the accuracy more easily.

**Acknowledgments.** We thank Dr. Matthias Weigel, Prof. Dr. Oliver Bieri and Tanja Haas for the MR acquisitions with the AMIRA sequence.

14 A. Horváth et al.

## References

1. Andermatt, S., Pezold, S., Cattin, P.: Multi-dimensional gated recurrent units for the segmentation of biomedical 3D-data. In: Carneiro, G., et al. (eds.) LABELS/DLMIA-2016. LNCS, vol. 10008, pp. 142–151. Springer, Cham (2016). [https://doi.org/10.1007/978-3-319-46976-8\\_15](https://doi.org/10.1007/978-3-319-46976-8_15)
2. Crum, W.R., Camara, O., Hill, D.L.G.: Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Trans. Med. Imaging* **25**(11), 1451–1461 (2006)
3. Datta, E., Papinutto, N., Schlaeger, R., Zhu, A., Carballido-Gamio, J., Henry, R.G.: Gray matter segmentation of the spinal cord with active contours in MR images. *NeuroImage* **147**, 788–799 (2017)
4. Horváth, A., et al.: A principled approach to combining inversion recovery images. In: Proceedings of the 26th Annual Meeting of ISMRM, Paris, France, June 2018
5. Perone, C.S., Calabrese, E., Cohen-Adad, J.: Spinal cord gray matter segmentation using deep dilated convolutions. *Sci. Rep.* **8**(1), 5966 (2018)
6. Porisky, A., et al.: Grey matter segmentation in spinal cord MRIs via 3D convolutional encoder networks with shortcut connections. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS-2017. LNCS, vol. 10553, pp. 330–337. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-67558-9\\_38](https://doi.org/10.1007/978-3-319-67558-9_38)
7. Prados, F., et al.: Spinal cord grey matter segmentation challenge. *NeuroImage* **152**, 312–329 (2017)
8. Sudre, C.H., Li, W., Vercauteren, T., Ourselin, S., Jorge Cardoso, M.: Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In: Cardoso, M.J., et al. (eds.) DLMIA/ML-CDS-2017. LNCS, vol. 10553, pp. 240–248. Springer, Cham (2017). [https://doi.org/10.1007/978-3-319-67558-9\\_28](https://doi.org/10.1007/978-3-319-67558-9_28)
9. Weigel, M., Bieri, O.: Spinal cord imaging using averaged magnetization inversion recovery acquisitions. *Magn. Reson. Med.* **79**(4), 1870–1881 (2018)



## 9 Discussion and Conclusion

The goal of this Ph.D. thesis was to develop a segmentation algorithm that enables automated quantification of gray matter (GM) and white matter (WM) inside the cervical spinal cord (SC) on magnetic resonance (MR) images. Segmentation of the SC's inner structures on MR images is particularly challenging because of low contrast-to-noise ratio (CNR) between GM and WM. As part of a larger project supported by the Swiss National Science Foundation (SNSF), we collaborated with magnetic resonance physicists in the development of an MR sequence appropriate for GM-WM segmentation. On the other hand, we were working together with neurological experts, who served us with their medical expertise, recruited healthy volunteers and multiple sclerosis (MS) patients for MR imaging (MRI) of the SC, and provided us with the manual GM-WM segmentations of such images. This joint work led to the development of a web application, called manual segmentation app (MSA), for visualization and manual segmentation of images acquired with the AMIRA protocol. This application makes use of the different representations of the AMIRA sequence's inversion images and simplifies the data transfer. The focus of this work, however, was the development and implementation of an automatic SC GM-WM segmentation algorithm that uses statistical knowledge to assist the pixelwise intensity-based decisions for getting a robust, accurate, and precise segmentation outcome, which, in turn, enables reliable SC GM and WM quantifications.

With the absence of training data at the beginning of this thesis, we first started experimenting with variational segmentation methods, while towards the end of this thesis, more and more data had been manually segmented, enabling us to apply deep learning. A literature review on variational segmentation approaches led us to the choice of continuous cut models equipped with statistical appearance models which can be elegantly optimized through the multiplier-based max-flow min-cut algorithm of Yuan et al. [23]. We published a prototype of such a continuous cut approach capable of segmenting SC GM-WM structures on axial 2D images acquired with the MOLLI sequence at the cervical vertebral level C3, see Chapter 6. We improved the model by extending it to a 3D model applicable also for other cervical levels and adapting it to the AMIRA sequence, which led to the publication in Chapter 7. In this publication, we analyzed the accuracy on a larger dataset, called the CGM dataset. We included intra- and inter-session comparisons to estimate the precision of the whole pipeline of imaging and post-processing. With supervised deep learning setting a new state of the art in image segmentation, we experimented with adapting and applying a deep-learning framework to the CGM dataset. In particular, we proposed to modify the training loss of the multidimensional gated recurrent units (MDGRU) [24] with a generalized Dice loss [151]. This modification enabled us robust and

## 9 Discussion and Conclusion

automatic segmentation of the inner SC structure with surprisingly high accuracy and reproducibility, which we published in Chapter 8. Moreover, this approach set a new state of the art in the SC GM segmentation (SCGM) challenge [14]. Both segmentation methods, the trained MDGRU network as well as the variational model are on par with the manual segmentations of the CGM dataset.

Parallel to the development of automated segmentation algorithms, we were investigating two other aspects of the AMIRA protocol. First, we analyzed the AMIRA sequence’s inversion images for the best one-channel projections in terms of CSF-SC contrast and GM-WM contrast, which are necessary for manual and variational segmentation. The resulting principled approach for combining AMIRA images led to the publication presented in Chapter 5. Second, we investigated dealing with the large slice distances between the 12 acquired axial AMIRA slices. For this we developed and implemented a slice interpolation algorithm capable of resampling the provided slices at any intermediate slice position, which led to the publication in Chapter 4.

**Discussion** This work was performed at the verge of the era of deep learning and was part of the sudden change from manual modeling techniques to supervised deep learning. With variational segmentation frameworks, we tried splitting the segmentation problem into several sub-problems using the mathematical language to reformulate each sub-problem to a computable version. Manual modeling in this way involved observing the algorithm’s performance and identifying possible improvements for many hours. It is, therefore, open to question whether this effortful task of manual modeling justifies the variational automation of the tedious task of manual segmentation – especially with supervised deep learning at hand. The automation of this problem tends to be more cost-effective when conducting the cumbersome task of manual segmentation on a set of images with which a deep neural network may automatically figure out how to solve the problem. From a practical point of view, it does not matter whether an algorithm’s model is hand-crafted or computer-generated, or, in other words, whether a problem is described by a simple, mathematically principled model or automatically optimized through deep learning: for complex ill-posed problems, reasoning with the statistical performance in terms of robustness, accuracy, and precision is more crucial than reasoning with mathematical principles.

With manual variational modeling as well as with deep learning, we generate segmentation models designed to work for a specific kind of data that both cannot be applied directly to other data. When trying to segment slightly different data, for example, images acquired with a different MR sequence, with deep learning, we can retrain the same architecture more conveniently compared to a manual approach. Efficiently designed forward passes of neural networks and applicable deep learning frameworks make retraining and ablation analyses more straightforward. With the variational pipeline described in Section 6.A, in contrast, ablation analyses involved individual, parameter-specific processing errors. Proper ablation analyses, however, are important to identify the parameters’ or the model features’ impact on the performance. The main problem in manual modeling lies in carefully handling the many

different cases where new problems eventually arise with unseen data. This aspect is more pleasing with neural networks, where forward passes normally do not throw syntactic or semantic errors if a network fails to generalize on unseen data. Supervised deep learning with neural networks, therefore, is desirable when it comes to comparisons of one approach applied to different datasets. The continuous cut approach in Section 6.A which was designed to work on cervical, axial AMIRA images, therefore, was not applied to the data of the SCGM challenge, whereas we applied the approach with MDGRU in Chapter 8 to the CGM dataset as well as the challenge dataset. In this sense, deep learning is scalable. The downside of supervised deep learning, however, is that it needs a lot of already annotated data to train on, to which it might be biased in addition. Since a shallow variational model has fewer parameters and regularizes more strongly, it might be "trained" on fewer samples.

Regarding the approximation accuracy, if a deep neural network's architecture is rich enough, then the network has the freedom of decomposing an actual task into its constituent functions, providing they exist in the first place. In contrast, a shallow model's architecture might be too hard-coded to represent the problem's complexity. For our task of segmenting GM and WM on MR images, however, both approaches eventually encountered samples where they failed; no matter how well they had been trained. The deep learning approach with MDGRU in Chapter 8 generalized more robustly and led to superior performance with less hand-crafted regularization compared to the variational segmentation approach in Section 6.A. The overregularization in the variational segmentation model, mainly caused by the regularization with slice similarity and appearance models, led to the slightly overestimated SC areas in the upper cervical regions, shown in Chapter 7, Fig. 3<sup>1</sup>. Moreover, about 12% of all slices in the CGM dataset needed to be excluded for the continuous cut approach because of failing subprocesses. On the other hand, experiments with the SCGM challenge have shown that a deep neural network trained on annotated data of one individual rater can mimic the rater's peculiarities. In Chapter 8, we showed that combining the outputs of several such individually trained networks generalizes the expert raters' consensus segmentation.

The AMIRA sequence produces 8 inversion images with different tissue contrasts which computer-based segmentation methods may process directly, but human raters first need one-channel projections with better CNR to conduct manual segmentation. The analysis in Chapter 5 shows that uniform averaging of the first five or the last three inversion images produces images with close-to-optimal CNR between GM and WM or CSF and SC, respectively. This analysis also yielded two other average images with even slightly better CNR regarding GM-WM and CSF-SC differentiability. All four average projections and the 8 inversion images, as well as the additional CSF contrast image from the AMIRA protocol (see Page 86), are presented within the Manual Segmentation App (MSA), where a human rater may scroll through all images while segmenting simultaneously. The shallow variational approaches from the publications in Chapter 6 and Chapter 7 mainly focused on using the MOLLI and the AMIRA

---

<sup>1</sup>The SC areas of the manual segmentations, in contrast, are visualized in Chapter 8, Fig. 5.

## 9 Discussion and Conclusion

sequences' less noisy average images, too. Experiments in Section 6.A showed, that the inclusion of the AMIRA sequence's individual inversion images in Equation (6.1) did not influence the segmentation performance significantly when choosing the  $\lambda_i$  in (6.1) to be the optimal weights proposed in Chapter 5 for optimal CNR. A parameter search on the optimal channel weights  $\lambda_i$  in (6.1) for optimal segmentation performance, however, has not been conducted. The additional CSF contrast image from the AMIRA protocol led to 4% fewer CSF segmentation failure cases with the continuous cut approach. Finally, the concept of using all individual inversion images of the AMIRA sequence is directly carried out with the MDGRU approach in Chapter 8, where the CSF segmentation robustly worked without the need of average images or the additional CSF contrast image. This additional CSF image has been neglected on purpose since MDGRU segments the CSF-SC boundary excellently without it, and co-registration errors with the AMIRA sequence might worsen the results.

The AMIRA protocol's large slice distances, in relation to the fine axial in-plane resolution, lead to an anisotropic 3D resolution. To address any needs for resampling of the anisotropic data, we developed a high-order slice interpolation method, proposed in Chapter 4. The improvements of this method compared to conventional interpolation techniques are only small in terms of mean absolute intensity differences. Accuracy measurements with manually placed landmarks could have revealed further insights for the method's performance. The main contribution, however, is the avoidance of kinks at the slice's stitching positions, which should provide better segmentation performance on resampled data. For now, we did not yet further follow this path, since the analyses of the individual slices already entail all the available information.

The magnetic field strength of an MR device has a strong influence on the quality of the acquired images in terms of the achievable CNR and resolution. The experiments in this work involved an MR device with a field strength of 3 T, since such machines represent the current state of the art in medical imaging and are well-spread around the world besides the 1.5 T machines. With 7 T machines being the next step in medical applications, better images with higher resolution and better CNR are expected [21], which probably would allow segmentation methods to achieve higher accuracy. However, with images of better quality probably also the segmentation tasks' demands increase towards the edge of the possible and thus the segmentation problems' difficulty increases as well. Whatever quality in MRI the future brings, segmentation algorithms should nevertheless be capable of dealing with today's devices for the foreseeable future, because of (a) the slow spread of 7 T scanners and more importantly because of (b) older data in longitudinal trials that were acquired with older technology. This work addresses the current limitations on quantifying SC GM-WM cross-sectional areas in-vivo with state-of-the-art MR imaging and post-processing pipelines. To provide our clinical partners with SC GM and WM atrophy rates, quantification of these limitations is necessary to assess the reliability of such rates.

Considering the large variability in assessing the SC GM and WM cross-sectional areas with respect to different methods [21, Tab. 4], our reported findings of cross-

sectional areas in healthy subjects, presented in Chapter 8, Fig. 5, are well in line with the literature.

**Future Work** Although the solution with the deep learning approach proposed in Chapter 8 is promising for longitudinal clinical studies, its performance may be further improved with newer network architectures which may also address the automatic estimation of the segmentation’s reliability and measures of uncertainty. With the healthy subjects’ follow-up data still being acquired, it will be interesting to analyze the yearly variability in addition to the already published intra- and inter-session variabilities of the GM, WM, and SC areas. The heavy presence of MS lesions in the SC, observed in AMIRA images of patients from the CGM dataset, complicates the morphometric analysis, mainly because lesion sites occlude the separability of GM and WM. Ongoing promising experiments include SC lesion segmentation as well as GM-WM segmentation on AMIRA images of MS patients. To analyze the variability of the GM-WM separability inside lesions, more manual segmentations are currently being produced that offer more data for the deep neural networks to train on. These mentioned steps are necessary for further clinical analyses, such as the estimation of reliable MS subtype-specific GM-WM atrophy rates, which is one of the clinical objectives of this research project.

So far, we aimed for automatic approaches to eliminate the variability induced through the interaction of a human rater. However, to enable accurate subject-specific analyses, instead of just aiming for large-scale medical studies, we plan to include the possibility to correct the computer-based results. This idea introduces exciting applications in combination with deep reinforcement learning to improve the segmentation models based on the amendments made by the user [155]. Furthermore, a potential future direction might also involve weakly supervised learning, where algorithms automatically learn the segmentations without providing pixelwise manual reference segmentations but manual annotations that are less time-consuming to acquire [131].

**Conclusion** We are living in an exciting time with fast-paced new discoveries made almost daily. We started with segmentation models from the first decade of the 2000s and ended with modern neural network architectures for deep learning. Supervised deep learning needs many training samples that include manual annotations but at the same time drastically minimizes the time needed to generate a segmentation algorithm. However, with both approaches, the manual modeling attempt and the supervised deep learning attempt, we created segmentation algorithms that conveniently and automatically segment the GM-WM compartments of the cervical SC on AMIRA images without hypo- or hyper-intense spots. Both methods, compared to manual segmentation, reduce the inter- and intra-session variabilities without losing a significant share of accuracy. Both methods are on par with manual segmentations. Thus, we are confident that our work enables further large scale, longitudinal clinical and pharmaceutical studies.



# Bibliography

- [1] Nigel W. John. Segmentation of Radiological Images. In Emanuele Neri, Davide Caramella, and Carlo Bartolozzi, editors, *Image Processing in Radiology: Current Applications*, Medical Radiology, pages 45–54. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. [1](#)
- [2] Daniel S. Reich, Claudia F. Lucchinetti, and Peter A. Calabresi. Multiple Sclerosis. *New England Journal of Medicine*, 378(2):169–180, January 2018. [1](#), [10](#), [11](#), [12](#)
- [3] Alan J. Thompson, Sergio E. Baranzini, Jeroen Geurts, Bernhard Hemmer, and Olga Ciccarelli. Multiple sclerosis. *The Lancet*, 391(10130):1622–1636, April 2018. [1](#), [10](#), [11](#)
- [4] Jill A. Hollenbach and Jorge R. Oksenberg. The Immunogenetics of Multiple Sclerosis: A Comprehensive Review. *Journal of autoimmunity*, 64:13–25, November 2015. [1](#), [10](#)
- [5] Natalia Petrova, Daniele Carassiti, Daniel R. Altmann, David Baker, and Klaus Schmierer. Axonal loss in the multiple sclerosis spinal cord revisited. *Brain Pathology*, 28(3):334–348, 2018. [1](#)
- [6] N. A. Losseff, S. L. Webb, J. I. O’Riordan, R. Page, L. Wang, G. J. Barker, P. S. Tofts, W. I. McDonald, D. H. Miller, and A. J. Thompson. Spinal cord atrophy and disability in multiple sclerosis. A new reproducible and sensitive MRI method with potential to monitor disease progression. *Brain: A Journal of Neurology*, 119 (Pt 3):701–708, June 1996. [1](#), [2](#), [12](#)
- [7] Adam B. Cohen, Mohit Neema, Ashish Arora, Elisa Dell’Oglio, Ralph H. B. Benedict, Shahamat Tauhid, Daniel Goldberg-Zimring, Christian Chavarro-Nieto, Antonella Ceccarelli, Joshua P. Klein, James M. Stankiewicz, Maria K. Houtchens, Guy J. Buckle, David C. Alsop, Charles R. G. Guttmann, and Rohit Bakshi. The Relationships among MRI-Defined Spinal Cord Involvement, Brain Involvement, and Disability in Multiple Sclerosis. *Journal of Neuroimaging*, 22(2):122–128, 2012. [1](#)
- [8] Regina Schlaeger, Nico Papinutto, Valentina Panara, Carolyn Bevan, Iryna V. Lobach, Monica Bucci, Eduardo Caverzasi, Jeffrey M. Gelfand, Ari J. Green, Kesshi M. Jordan, William A. Stern, H.-Christian von Büdingen, Emmanuelle Waubant, Alyssa H. Zhu, Douglas S. Goodin, Bruce A. C. Cree, Stephen L.

## BIBLIOGRAPHY

- Hauser, and Roland G. Henry. Spinal cord gray matter atrophy correlates with multiple sclerosis disability. *Annals of Neurology*, 76(4):568–580, October 2014. [1](#), [3](#), [13](#)
- [9] Wei Qiu, Sonja Raven, Ian James, Yuebei Luo, Jingshan Wu, Alison Castley, Frank T. Christiansen, William M. Carroll, Frank L. Mastaglia, and Allan G. Kermode. Spinal cord involvement in multiple sclerosis: a correlative MRI and high-resolution HLA-DRB1 genotyping study. *Journal of the Neurological Sciences*, 300(1-2):114–119, January 2011. [1](#)
- [10] A. Papadopoulos, S. Gatzonis, A. Gouliamos, S. Trakadas, A. Kalovidouris, P. Sgouropoulos, L. Vlachos, and C. Papavasiliou. Correlation between spinal cord MRI and clinical features in patients with demyelinating disease. *Neuroradiology*, 36(2):130–133, February 1994. [1](#)
- [11] Frederik Barkhof. The clinico-radiological paradox in multiple sclerosis revisited. *Current Opinion in Neurology*, 15(3):239–245, June 2002. [1](#), [12](#), [29](#)
- [12] Mary F Dempsey, Barrie Condon, and Donald M Hadley. MRI safety review. *Seminars in Ultrasound, CT and MRI*, 23(5):392–401, October 2002. [1](#)
- [13] M. C. Yiannakas, H. Kearney, R. S. Samson, D. T. Chard, O. Ciccarelli, D. H. Miller, and C. A. M. Wheeler-Kingshott. Feasibility of grey matter and white matter segmentation of the upper cervical cord in vivo: A pilot study with application to magnetisation transfer measurements. *NeuroImage*, 63(3):1054–1059, November 2012. [2](#), [3](#), [13](#), [140](#)
- [14] Ferran Prados, John Ashburner, Claudia Blaiotta, Tom Brosch, Julio Carballido-Gamio, Manuel Jorge Cardoso, Benjamin N. Conrad, Esha Datta, Gergely Dávid, Benjamin De Leener, Sara M. Dupont, Patrick Freund, Claudia A. M. Gandini Wheeler-Kingshott, Francesco Grussu, Roland Henry, Bennett A. Landman, Emil Ljungberg, Bailey Lyttle, Sebastien Ourselin, Nico Papinutto, Salvatore Saporito, Regina Schlaeger, Seth A. Smith, Paul Summers, Roger Tam, Marios C. Yiannakas, Alyssa Zhu, and Julien Cohen-Adad. Spinal cord grey matter segmentation challenge. *NeuroImage*, 152:312–329, May 2017. [2](#), [4](#), [67](#), [73](#), [75](#), [139](#), [154](#)
- [15] Charidimos Tsagkas, Stefano Magon, Laura Gaetano, Simon Pezold, Yvonne Naegelin, Michael Amann, Christoph Stippich, Philippe Cattin, Jens Wuerfel, Oliver Bieri, Till Sprenger, Ludwig Kappos, and Katrin Parmar. Spinal cord volume loss: A marker of disease progression in multiple sclerosis. *Neurology*, 91(4):e349–e358, July 2018. [2](#), [11](#)
- [16] Charidimos Tsagkas, Stefano Magon, Laura Gaetano, Simon Pezold, Yvonne Naegelin, Michael Amann, Christoph Stippich, Philippe Cattin, Jens Wuerfel, Oliver Bieri, Till Sprenger, Ludwig Kappos, and Katrin Parmar. Preferential



## BIBLIOGRAPHY

- spinal cord volume loss in primary progressive multiple sclerosis. *Multiple Sclerosis (Houndmills, Basingstoke, England)*, 25(7):947–957, June 2019. [2](#), [11](#), [72](#)
- [17] Xia Lin, Christopher R. Tench, Nikos Evangelou, Timothy Jaspan, and Cris S. Constantinescu. Measurement of spinal cord atrophy in multiple sclerosis. *Journal of Neuroimaging: Official Journal of the American Society of Neuroimaging*, 14(3 Suppl):20S–26S, July 2004. [3](#), [13](#)
  - [18] Benjamin De Leener, Simon Lévy, Sara M. Dupont, Vladimir S. Fonov, Nikola Stikov, D. Louis Collins, Virginie Callot, and Julien Cohen-Adad. SCT: Spinal Cord Toolbox, an open-source software for processing spinal cord MRI data. *NeuroImage*, 145, Part A:24 – 43, 2017. [3](#)
  - [19] Simon Pezold. *Quantification of spinal cord atrophy in magnetic resonance images*. Ph.D. thesis, University of Basel, 2016. [3](#), [61](#), [63](#)
  - [20] Christopher P. Gilmore, Gabriele C. DeLuca, Lars Bö, Trudy Owens, James Lowe, Margaret M. Esiri, and Nikos Evangelou. Spinal cord atrophy in multiple sclerosis caused by white matter volume loss. *Archives of Neurology*, 62(12):1859–1862, December 2005. [3](#), [13](#)
  - [21] E. E. Sigmund, G. A. Suero, C. Hu, K. McGorty, D. K. Sodickson, G. C. Wiggins, and J. A. Helpert. High-resolution human cervical spinal cord imaging at 7 T. *NMR in Biomedicine*, 25(7):891–899, 2012. [3](#), [140](#), [156](#)
  - [22] Niels Christian Overgaard, Ketut Fundana, and Anders Heyden. Pose Invariant Shape Prior Segmentation Using Continuous Cuts and Gradient Descent on Lie Groups. In Xue-Cheng Tai, Knut Mørken, Marius Lysaker, and Knut-Andreas Lie, editors, *Scale Space and Variational Methods in Computer Vision*, number 5567 in Lecture Notes in Computer Science, pages 684–695. Springer Berlin Heidelberg, June 2009. [4](#), [50](#), [65](#), [66](#), [109](#)
  - [23] J. Yuan, E. Bae, and X. C. Tai. A study on continuous max-flow and min-cut approaches. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2217–2224, June 2010. [4](#), [44](#), [48](#), [53](#), [55](#), [61](#), [62](#), [63](#), [64](#), [153](#)
  - [24] Simon Andermatt, Simon Pezold, and Philippe Cattin. Multi-dimensional Gated Recurrent Units for the Segmentation of Biomedical 3d-Data. In Gustavo Carneiro, Diana Mateus, Loïc Peter, Andrew Bradley, João Manuel R. S. Tavares, Vasileios Belagiannis, João Paulo Papa, Jacinto C. Nascimento, Marco Loog, Zhi Lu, Jaime S. Cardoso, and Julien Cornebise, editors, *Deep Learning and Data Labeling for Medical Applications*, Lecture Notes in Computer Science, pages 142–151. Springer International Publishing, 2016. [4](#), [29](#), [67](#), [139](#), [153](#)

## BIBLIOGRAPHY

- [25] Matthias Weigel and Oliver Bieri. Spinal cord imaging using averaged magnetization inversion recovery acquisitions. *Magnetic Resonance in Medicine*, 79(4):1870–1881, April 2018. 5, 75, 76, 78, 79, 80, 82, 103
- [26] Ahmadreza Baghaie and Zeyun Yu. Curvature-Based Registration for Slice Interpolation of Medical Images. In David Hutchison, Takeo Kanade, Josef Kittler, Jon M. Kleinberg, Alfred Kobsa, Friedemann Mattern, John C. Mitchell, Moni Naor, Oscar Nierstrasz, C. Pandu Rangan, Bernhard Steffen, Demetri Terzopoulos, Doug Tygar, Gerhard Weikum, Yongjie Jessica Zhang, and João Manuel R. S. Tavares, editors, *Computational Modeling of Objects Presented in Images. Fundamentals, Methods, and Applications*, volume 8641, pages 69–80. Springer International Publishing, Cham, 2014. 5, 91, 92
- [27] Henry Gray and Henry Vandyke Carter. *Anatomy of the human body*. Lea & Febiger, Philadelphia, 20th ed. / thoroughly rev. and re-edited by Warren H. Lewis edition, 1918. Open Library ID: OL24786057M. 8
- [28] Wikipedia, The Free Encyclopedia. Gray matter. [https://en.wikipedia.org/wiki/Grey\\_matter#/media/File:Medulla\\_spinalis\\_-\\_Section\\_-\\_English.svg](https://en.wikipedia.org/wiki/Grey_matter#/media/File:Medulla_spinalis_-_Section_-_English.svg), 2019. [Online; accessed 27-August-2019]. 8
- [29] Arvid Frostell, Ramil Hakim, Eric Peter Thelin, Per Mattsson, and Mikael Svensson. A Review of the Segmental Diameter of the Healthy Human Spinal Cord. *Frontiers in Neurology*, 7, December 2016. 8
- [30] L. Sakka, G. Coll, and J. Chazal. Anatomy and physiology of cerebrospinal fluid. *European Annals of Otorhinolaryngology, Head and Neck Diseases*, 128(6):309–316, December 2011. 9
- [31] D. R. Enzmann, J. B. Rubin, R. DeLaPaz, and A. Wright. Cerebrospinal fluid pulsation: benefits and pitfalls in MR imaging. *Radiology*, 161(3):773–778, December 1986. 9, 75
- [32] L. M. J De Sonnevile, J. B Boringa, I. E. W Reuling, R. H. C Lazeron, H. J Adèr, and C. H Polman. Information processing characteristics in subtypes of multiple sclerosis. *Neuropsychologia*, 40(11):1751–1765, January 2002. 10
- [33] Charles M. Poser, Donald W. Paty, Labe Scheinberg, W. Ian McDonald, Floyd A. Davis, George C. Ebers, Kenneth P. Johnson, William A. Sibley, Donald H. Silberberg, and Wallace W. Tourtellotte. New diagnostic criteria for multiple sclerosis: Guidelines for research protocols. *Annals of Neurology*, 13(3):227–231, 1983. 10
- [34] W. Ian McDonald, Alistair Compston, Gilles Edan, Donald Goodkin, Hans-Peter Hartung, Fred D. Lublin, Henry F. McFarland, Donald W. Paty, Chris H. Polman, Stephen C. Reingold, Magnhild Sandberg-Wollheim, William Sibley, Alan Thompson, Stanley Van Den Noort, Brian Y. Weinshenker, and Jerry S.

## BIBLIOGRAPHY

- Wolinsky. Recommended diagnostic criteria for multiple sclerosis: Guidelines from the international panel on the diagnosis of multiple sclerosis. *Annals of Neurology*, 50(1):121–127, 2001. [10](#)
- [35] Alan J Thompson, Brenda L Banwell, Frederik Barkhof, William M Carroll, Timothy Coetzee, Giancarlo Comi, Jorge Correale, Franz Fazekas, Massimo Filippi, Mark S Freedman, Kazuo Fujihara, Steven L Galetta, Hans Peter Hartung, Ludwig Kappos, Fred D Lublin, Ruth Ann Marrie, Aaron E Miller, David H Miller, Xavier Montalban, Ellen M Mowry, Per Soelberg Sorensen, Mar Tintoré, Anthony L Traboulsee, Maria Trojano, Bernard M J Uitdehaag, Sandra Vukusic, Emmanuelle Waubant, Brian G Weinshenker, Stephen C Reingold, and Jeffrey A Cohen. Diagnosis of multiple sclerosis: 2017 revisions of the McDonald criteria. *The Lancet Neurology*, 17(2):162–173, February 2018. [10](#)
- [36] John F. Kurtzke. Rating neurologic impairment in multiple sclerosis. *Neurology*, 33(11):1444–1444, 1983. [10](#)
- [37] Nicola De Stefano, Marco Battaglini, and Stephen M. Smith. Measuring Brain Atrophy in Multiple Sclerosis. *Journal of Neuroimaging*, 17(s1):10S–15S, 2007. [11](#), [12](#)
- [38] Hugh Kearney, David H. Miller, and Olga Ciccarelli. Spinal cord MRI in multiple sclerosis—diagnostic, prognostic and clinical value. *Nature Reviews Neurology*, 11(6):327–338, June 2015. [11](#), [12](#)
- [39] Fred D. Lublin and Stephen C. Reingold. Defining the clinical course of multiple sclerosis. *Neurology*, 46(4):907–911, 1996. [11](#)
- [40] Andreas Bitsch and Wolfgang Brück. Differentiation of Multiple Sclerosis Subtypes. *CNS Drugs*, 16(6):405–418, June 2002. [11](#)
- [41] Bruce D. Trapp, Megan Vignos, Jessica Dudman, Ansi Chang, Elizabeth Fisher, Susan M. Staugaitis, Harsha Battapady, Sverre Mork, Daniel Ontaneda, Stephen E. Jones, Robert J. Fox, Jacqueline Chen, Kunio Nakamura, and Richard A. Rudick. Cortical neuronal densities and cerebral white matter demyelination in multiple sclerosis: a retrospective study. *The Lancet Neurology*, 17(10):870–884, October 2018. [11](#), [29](#)
- [42] Hans Lassmann. What drives disease in multiple sclerosis: Inflammation or neurodegeneration? *Clinical and Experimental Neuroimmunology*, 1(1):2–11, January 2010. [11](#)
- [43] G. J. Nijeholt, M. A. van Walderveen, J. A. Castelijns, J. H. van Waesberghe, C. Polman, P. Scheltens, P. F. Rosier, P. J. Jongen, and F. Barkhof. Brain and spinal cord abnormalities in multiple sclerosis. Correlation between MRI parameters, clinical subtypes and symptoms. *Brain: A Journal of Neurology*, 121 ( Pt 4):687–697, April 1998. [11](#), [12](#)

## BIBLIOGRAPHY

- [44] Xavier Montalban, Stephen L. Hauser, Ludwig Kappos, Douglas L. Arnold, Amit Bar-Or, Giancarlo Comi, Jérôme de Seze, Gavin Giovannoni, Hans-Peter Hartung, Bernhard Hemmer, Fred Lublin, Kottil W. Rammohan, Krzysztof Selmaj, Anthony Traboulsee, Annette Sauter, Donna Masterman, Paulo Fontoura, Shibeshih Belachew, Hideki Garren, Nicole Mairon, Peter Chin, Jerry S. Wolinsky, and ORATORIO Clinical Investigators. Ocrelizumab versus Placebo in Primary Progressive Multiple Sclerosis. *The New England Journal of Medicine*, 376(3):209–220, 2017. [12](#)
- [45] Robert W. Motl and Lara A. Pilutti. The benefits of exercise training in multiple sclerosis. *Nature Reviews Neurology*, 8(9):487–497, September 2012. [12](#)
- [46] Timothy Coetzee and Alan J. Thompson. Unified understanding of MS course is required for drug development. *Nature Reviews Neurology*, 14:191–192, January 2018. [12](#)
- [47] Charley Gros, Benjamin De Leener, Atef Badji, Josefina Maranzano, Dominique Eden, Sara M. Dupont, Jason Talbott, Ren Zhuoquiong, Yaou Liu, Tobias Granberg, Russell Ouellette, Yasuhiko Tachibana, Masaaki Hori, Kouhei Kamiya, Lydia Chougar, Leszek Stawiarz, Jan Hillert, Elise Bannier, Anne Kerbrat, Gilles Edan, Pierre Labauge, Virginie Callot, Jean Pelletier, Bertrand Audoin, Henitsoa Rasoanandrianina, Jean-Christophe Brisset, Paola Valsasina, Maria A. Rocca, Massimo Filippi, Rohit Bakshi, Shahamat Tauhid, Ferran Prados, Marios Yiannakas, Hugh Kearney, Olga Ciccarelli, Seth Smith, Constantina Andrada Treaba, Caterina Mainero, Jennifer Lefeuvre, Daniel S. Reich, Govind Nair, Vincent Auclair, Donald G. McLaren, Allan R. Martin, Michael G. Fehlings, Shahabeddin Vahdat, Ali Khatibi, Julien Doyon, Timothy Shepherd, Erik Charlson, Sridar Narayanan, and Julien Cohen-Adad. Automatic segmentation of the spinal cord and intramedullary multiple sclerosis lesions with convolutional neural networks. *NeuroImage*, 184:901–915, January 2019. [12](#), [75](#)
- [48] Blandine Grassiot, Béatrice Desgranges, Francis Eustache, and Gilles Defer. Quantification and clinical relevance of brain atrophy in multiple sclerosis: a review. *Journal of Neurology*, 256(9):1397, April 2009. [12](#), [13](#)
- [49] Gilles Aubert and Pierre Kornprobst. *Mathematical Problems in Image Processing: Partial Differential Equations and the Calculus of Variations*. Applied Mathematical Sciences. Springer-Verlag, New York, 2 edition, 2006. [13](#), [17](#), [25](#), [26](#), [27](#), [29](#), [31](#), [32](#), [33](#), [34](#), [48](#), [50](#), [54](#), [55](#), [65](#)
- [50] Stanley Osher and Ronald Fedkiw. *Level Set Methods and Dynamic Implicit Surfaces*. Applied Mathematical Sciences. Springer-Verlag, New York, 2003. [13](#), [17](#), [18](#), [21](#), [32](#), [40](#), [44](#), [56](#), [57](#)
- [51] T. M. Flett. *Differential Analysis: Differentiation, Differential Equations and Differential Inequalities*. Cambridge University Press, Cambridge, reissue edition, November 2008. [13](#), [19](#), [26](#)

## BIBLIOGRAPHY

- [52] Alfio Quarteroni and Silvia Quarteroni. *Numerical models for differential problems*, volume 2. Springer, 2009. [13](#), [32](#)
- [53] Otto Forster. *Analysis 1: Differential- und Integralrechnung einer Veränderlichen*. Grunkurs Mathematik. Springer Spektrum, 12 edition, 2016. [13](#), [19](#), [22](#)
- [54] Otto Forster. *Analysis 2: Differentialrechnung im  $\mathbb{R}^n$ , gewöhnliche Differentialgleichungen*. Grunkurs Mathematik. Vieweg+Teubner Verlag, 9 edition, 2010. [13](#)
- [55] Otto Forster. *Analysis 3: Maß- und Integrationstheorie, Integralsätze im  $\mathbb{R}^n$  und Anwendungen*. Aufbaukurs Mathematik. Vieweg+Teubner Verlag, 7 edition, 2012. [13](#), [23](#), [26](#)
- [56] Michael Artin. *Algebra*. Pearson, 2 edition, 2010. [13](#)
- [57] Hans W. Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer-Verlag, 3 edition, 1999. [13](#), [17](#)
- [58] R. Tyrrell Rockafellar. *Convex Analysis*. Princeton University Press, 1970. Google-Books-ID: 1TiOka9bx3sC. [13](#)
- [59] Lawrence C. Evans. *Partial Differential Equations*. American Mathematical Society, Providence, R.I, 2 edition, 2010. [13](#), [26](#), [31](#), [37](#)
- [60] F William Lawvere and Stephen H Schanuel. *Conceptual mathematics: a first introduction to categories*. Cambridge University Press, 2009. [15](#)
- [61] James Munkres. *Topology*. Pearson, Upper Saddle River, NJ, 2 edition edition, January 2000. [17](#)
- [62] William P. Ziemer. *Weakly Differentiable Functions: Sobolev Spaces and Functions of Bounded Variation*. Graduate Texts in Mathematics. Springer-Verlag, New York, 1989. [21](#), [26](#)
- [63] Lawrence Craig Evans and Ronald F. Gariepy. *Measure Theory and Fine Properties of Functions*. CRC Press, December 1991. Google-Books-ID: HZONacVPGIMC. [21](#), [50](#)
- [64] Measure Theory and Fine Properties of Functions, Revised Edition. [23](#)
- [65] Shigeyuki Morita. *Geometry of differential forms. Translations of Mathematical Monographs*. Number 201. American Mathematical Society, 2001. [23](#), [24](#)
- [66] Marcus J Nadenau, Stefan Winkler, David Alleysson, and Murat Kunt. Human vision models for perceptually optimized image processing—a review. *Proceedings of the IEEE*, 32, 2000. [27](#), [34](#)

## BIBLIOGRAPHY

- [67] Richard Szeliski. Computer Vision: Algorithms and Applications. *Draft*, page 979, 2010. [27](#)
- [68] Wikipedia, The Free Encyclopedia. Computer vision. [http://en.wikipedia.org/w/index.php?title=Computer%20vision&oldid=896283892#System\\_methods](http://en.wikipedia.org/w/index.php?title=Computer%20vision&oldid=896283892#System_methods), 2019. [Online; accessed 29-May-2019]. [27](#)
- [69] C. E. Shannon. Communication in the Presence of Noise. *Proceedings of the IRE*, 37(1):10–21, January 1949. [30](#)
- [70] Francisco P. M. Oliveira and João Manuel R. S. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, January 2014. [31](#)
- [71] Ajay Kumar Boyat and Brijendra Kumar Joshi. A Review Paper: Noise Models in Digital Image Processing. *arXiv:1505.03489 [cs]*, May 2015. *arXiv*:1505.03489. [31](#)
- [72] F Guichard, L Moisan, and J-M Morel. A review of pde models in image processing and image analysis. In *Journal de Physique IV (Proceedings)*, volume 12, pages 137–154. EDP sciences, 2002. [31](#)
- [73] D. R. Jones, C. D. Perttunen, and B. E. Stuckman. Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications*, 79(1):157–181, October 1993. [32](#)
- [74] Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient Descent Converges to Minimizers. *arXiv:1602.04915 [cs, math, stat]*, February 2016. [33](#)
- [75] Bernd Fischer and Jan Modersitzki. Ill-posed medicine—an introduction to image registration. *Inverse Problems*, 24(3):034008, 2008. [34](#)
- [76] Mark R Banham and Aggelos K Katsaggelos. Digital image restoration. *IEEE signal processing magazine*, 14(2):24–41, 1997. [34](#)
- [77] Daniel Cremers, Mikael Rousson, and Rachid Deriche. A Review of Statistical Approaches to Level Set Segmentation: Integrating Color, Texture, Motion and Shape. *International Journal of Computer Vision*, 72(2):195–215, April 2007. [35](#), [43](#), [51](#)
- [78] Christopher Bishop. *Pattern Recognition and Machine Learning*. Information Science and Statistics. Springer-Verlag, New York, 2006. [35](#)
- [79] Michael Kass, Andrew Witkin, and Demetri Terzopoulos. Snakes: Active contour models. *International Journal of Computer Vision*, 1(4):321–331, January 1988. [39](#), [43](#), [44](#), [55](#)

- [80] Xiao-Feng Wang, Hai Min, Le Zou, and Yi-Gang Zhang. A novel level set method for image segmentation by incorporating local statistical analysis and global similarity measurement. *Pattern Recognition*, 48(1):189–204, January 2015. [40](#), [44](#), [52](#)
- [81] T. Chan and Wei Zhu. Level set based shape prior segmentation. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 1164–1170 vol. 2, June 2005. [40](#), [57](#), [65](#)
- [82] Luminita A Vese and Tony F Chan. A multiphase level set framework for image segmentation using the mumford and shah model. *International journal of computer vision*, 50(3):271–293, 2002. [41](#)
- [83] Egil Bae, Jing Yuan, Xue-Cheng Tai, and Yuri Boykov. A Fast Continuous Max-Flow Approach to Non-convex Multi-labeling Problems. In Andrés Bruhn, Thomas Pock, and Xue-Cheng Tai, editors, *Efficient Algorithms for Global Optimization Methods in Computer Vision*, Lecture Notes in Computer Science, pages 134–154. Springer Berlin Heidelberg, 2014. [41](#)
- [84] Thomas Pock, Daniel Cremers, Horst Bischof, and Antonin Chambolle. Global Solutions of Variational Models with Convex Regularization. *SIAM J. Imaging Sciences*, 3:1122–1145, 2010. [41](#), [44](#)
- [85] Thomas Pock, Thomas Schoenemann, Gottfried Graber, Horst Bischof, and Daniel Cremers. A Convex Formulation of Continuous Multi-label Problems. In David Forsyth, Philip Torr, and Andrew Zisserman, editors, *European Conference on Computer Vision (ECCV)*, number 5304 in Lecture Notes in Computer Science, pages 792–805. Springer Berlin Heidelberg, October 2008. [41](#), [44](#), [52](#)
- [86] Masoud S Nosrati, Shawn Andrews, and Ghassan Hamarneh. Bounded labeling function for global segmentation of multi-part objects with geometric constraints. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 2032–2039, 2013. [41](#), [42](#), [44](#), [53](#), [60](#)
- [87] Neena Aloysius and M Geetha. A review on deep convolutional neural networks. In *2017 International Conference on Communication and Signal Processing (ICCSP)*, pages 0588–0592. IEEE, 2017. [42](#)
- [88] Nikhil R Pal and Sankar K Pal. A review on image segmentation techniques. *Pattern Recognition*, 26(9):1277–1294, September 1993. [43](#), [52](#)
- [89] Dzung L. Pham, Chenyang Xu, and Jerry L. Prince. Current Methods in Medical Image Segmentation. *Annual Review of Biomedical Engineering*, 2(1):315–337, 2000. [43](#)
- [90] Tobias Heimann and Hans-Peter Meinzer. Statistical shape models for 3d medical image segmentation: A review. *Medical Image Analysis*, 13(4):543–563, August 2009. [43](#), [44](#), [64](#)



## BIBLIOGRAPHY

- [91] Geethu Mohan and M. Monica Subashini. Chapter 4 - Medical Imaging With Intelligent Systems: A Review. In Arun Kumar Sangaiah, editor, *Deep Learning and Parallel Computing Environment for Bioengineering Systems*, pages 53–73. Academic Press, January 2019. [43](#)
- [92] Carsten Rother, Vladimir Kolmogorov, and Andrew Blake. Grabcut -interactive foreground extraction using iterated graph cuts. *ACM Transactions on Graphics (SIGGRAPH)*, August 2004. [43](#)
- [93] Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. In *Eighth IEEE International Conference on Computer Vision (ICCV). Proceedings*, volume 1, pages 105–112 vol.1, 2001. [43](#), [44](#), [52](#), [57](#), [60](#), [124](#)
- [94] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Transactions on Image Processing*, 10(2):266–277, February 2001. [43](#), [44](#), [55](#), [57](#), [65](#)
- [95] Tomaso Poggio, Hrushikesh Mhaskar, Lorenzo Rosasco, Brando Miranda, and Qianli Liao. Why and when can deep-but not shallow-networks avoid the curse of dimensionality: A review. *International Journal of Automation and Computing*, 14(5):503–519, October 2017. [44](#), [67](#), [68](#), [69](#), [70](#)
- [96] Richard Szeliski. Segmentation. In Richard Szeliski, editor, *Computer Vision: Algorithms and Applications*, Texts in Computer Science, pages 235–271. Springer London, London, 2011. [44](#)
- [97] Andrew J. Asman, Seth A. Smith, Daniel S. Reich, and Bennett A. Landman. Robust GM/WM Segmentation of the Spinal Cord with Iterative Non-local Statistical Fusion. In Kensaku Mori, Ichiro Sakuma, Yoshinobu Sato, Christian Barillot, and Nassir Navab, editors, *Medical Image Computing and Computer-Assisted Intervention MICCAI*, Lecture Notes in Computer Science, pages 759–767. Springer Berlin Heidelberg, 2013. [44](#), [66](#), [73](#), [127](#)
- [98] M. Bach Cuadra, V. Duay, and J.-Ph. Thiran. Atlas-based Segmentation. In Nikos Paragios, James Duncan, and Nicholas Ayache, editors, *Handbook of Biomedical Imaging: Methodologies and Clinical Research*, pages 221–244. Springer US, Boston, MA, 2015. [44](#), [66](#)
- [99] Mariano Cabezas, Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Meritxell Bach Cuadra. A review of atlas-based segmentation for magnetic resonance brain images. *Computer Methods and Programs in Biomedicine*, 104(3):e158–e177, December 2011. [44](#), [66](#)
- [100] Jing Yuan, Egil Bae, Xue-Cheng Tai, and Yuri Boykov. A Continuous Max-Flow Approach to Potts Model. In *European Conference on Computer Vision (ECCV)*, Lecture Notes in Computer Science, pages 379–392. Springer, Berlin, Heidelberg, September 2010. [44](#)



## BIBLIOGRAPHY

- [101] D. Freedman and Tao Zhang. Interactive graph cut based segmentation with shape priors. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 755–762 vol. 1, June 2005. [44](#)
- [102] Tahir Majeed. *Facial soft tissue segmentation*. Ph.D. thesis, University of Basel, 2014. [44](#), [52](#)
- [103] Marcin Andrychowicz, Misha Denil, Sergio Gómez, Matthew W Hoffman, David Pfau, Tom Schaul, Brendan Shillingford, and Nando de Freitas. Learning to learn by gradient descent by gradient descent. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 3981–3989. Curran Associates, Inc., 2016. [46](#)
- [104] Phil Kim. Deep learning. In *MATLAB Deep Learning*, pages 103–120. Springer, 2017. [46](#)
- [105] Wikipedia, The Free Encyclopedia. Ai winter. [https://en.wikipedia.org/wiki/AI\\_winter](https://en.wikipedia.org/wiki/AI_winter), 2019. [Online; accessed 02-September-2019]. [47](#)
- [106] David Mumford and Jayant Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5):577–685, 1989. [48](#), [52](#)
- [107] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, November 1992. [50](#)
- [108] Antonin Chambolle and Thomas Pock. A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, May 2011. [50](#), [54](#), [61](#), [62](#), [63](#)
- [109] Thomas Brox and Daniel Cremers. On the statistical interpretation of the piecewise smooth mumford-shah functional. In *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 203–213. Springer, 2007. [51](#)
- [110] Ernst Ising. Beitrag zur Theorie des Ferromagnetismus. *Zeitschrift für Physik*, 31(1):253–258, February 1925. [51](#), [52](#)
- [111] F. Y. Wu. The Potts model. *Reviews of Modern Physics*, 54(1):235–268, January 1982. [52](#)
- [112] L. R. Ford and D. R. Fulkerson. Maximal Flow Through a Network. *Canadian Journal of Mathematics*, 8:399–404, 1956. [52](#)

## BIBLIOGRAPHY

- [113] C. Olsson, M. Byröd, N. C. Overgaard, and F. Kahl. Extending continuous cuts: Anisotropic metrics and expansion moves. In *IEEE 12th International Conference on Computer Vision (ICCV)*, pages 405–412, September 2009. [55](#), [61](#)
- [114] Vicent Caselles, Ron Kimmel, and Guillermo Sapiro. Geodesic Active Contours. *International Journal of Computer Vision*, 22(1):61–79, February 1997. [55](#)
- [115] Vicent Caselles, Francine Catté, Tomeu Coll, and Françoise Dibos. A geometric model for active contours in image processing. *Numerische Mathematik*, 66(1):1–31, December 1993. [56](#)
- [116] George Bernard Dantzig and Richard Cottle. *The Basic George B. Dantzig*. Stanford University Press, 2003. Google-Books-ID: ZpYca36h464C. [58](#)
- [117] W. Qiu, J. Yuan, E. Ukwatta, Y. Sun, M. Rajchl, and A. Fenster. Prostate Segmentation: An Efficient Convex Optimization Approach With Axial Symmetry Using 3-D TRUS and MR Images. *IEEE Transactions on Medical Imaging*, 33(4):947–960, April 2014. [60](#)
- [118] Simon Pezold, Ketut Fundana, Michael Amann, Michaela Andelova, Armanda Pfister, Till Sprenger, and Philippe C. Cattin. Automatic Segmentation of the Spinal Cord Using Continuous Max Flow with Cross-sectional Similarity Prior and Tubularity Features. In *Recent Advances in Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computational Vision and Biomechanics, pages 107–118. Springer, Cham, 2015. [60](#)
- [119] Ke Wei, Xue-Cheng Tai, Tony F Chan, and Shingyu Leung. Primal-dual method for continuous max-flow approaches. In *Computational Vision and Medical Image Processing V-Proceedings of 5th ECCOMAS Thematic Conference on Computational Vision and Medical Image Processing, VipIMAGE*, volume 2015, pages 17–24, 2016. [61](#)
- [120] Simon Pezold, Antal Horváth, Ketut Fundana, Charidimos Tsagkas, Michaela Andělová, Katrin Weier, Michael Amann, and Philippe C Cattin. Automatic, robust, and globally optimal segmentation of tubular structures. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 362–370. Springer, 2016. [61](#)
- [121] John Ernest Esser. *Primal dual algorithms for convex models and applications to image restoration, registration and nonlocal inpainting*. PhD thesis, University of California, 2010. OCLC: 609714479. [61](#), [62](#), [64](#)
- [122] Mingqiang Zhu and Tony Chan. An efficient primal-dual hybrid gradient algorithm for total variation image restoration. *UCLA Cam Report*, 08-34, 2008. [62](#)

- [123] Kenneth J Arrow, Leonid Hurwicz, and Hirofumi Uzawa. Studies in linear and non-linear programming. 1958. [62](#)
- [124] Jing Yuan, Egil Bae, Xue-Cheng Tai, and Yuri Boykov. A study on continuous max-flow and min-cut approaches. *UCLA Cam Report*, 10-61, 2010. [63](#)
- [125] Ke Wei. Primal-dual method for continuous max-flow approaches, 2015. [63](#), [64](#)
- [126] D. Cremers, F. R. Schmidt, and F. Barthel. Shape priors in variational image segmentation: Convexity, Lipschitz continuity and globally optimal solutions. In *IEEE Conference on Computer Vision and Pattern Recognition, 2008. CVPR*, pages 1–6, June 2008. [64](#), [65](#)
- [127] Manuel Taso, Arnaud Le Troter, Michael Sdika, Julien Cohen-Adad, Pierre-Jean Arnoux, Maxime Guye, Jean-Philippe Ranjeva, and Virginie Callot. A reliable spatially normalized template of the human spinal cord – Applications to automated white matter/gray matter segmentation and tensor-based morphometry (TBM) mapping of gray matter alterations occurring with age. *NeuroImage*, 117:20–28, August 2015. [66](#), [127](#)
- [128] Simon K. Warfield, Kelly H. Zou, and William M. Wells. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE transactions on medical imaging*, 23(7):903–921, July 2004. [66](#)
- [129] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI*, Lecture Notes in Computer Science, pages 234–241. Springer, Cham, 2015. [67](#)
- [130] Marijn F. Stollenga, Wonmin Byeon, Marcus Liwicki, and Juergen Schmidhuber. Parallel Multi-dimensional LSTM, with Application to Fast Biomedical Volumetric Image Segmentation. In *Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 2*, NIPS, pages 2998–3006, Cambridge, MA, USA, 2015. MIT Press. event-place: Montreal, Canada. [67](#)
- [131] Simon Andermatt. *Automated brain lesion segmentation in magnetic resonance images*. Ph.D. thesis, University of Basel, 2018. [67](#), [157](#)
- [132] Nitish Shirish Keskar, Dheevatsa Mudigere, Jorge Nocedal, Mikhail Smelyanskiy, and Ping Tak Peter Tang. On Large-Batch Training for Deep Learning: Generalization Gap and Sharp Minima. *arXiv:1609.04836 [cs, math]*, September 2016. arXiv: 1609.04836. [71](#)
- [133] Chiyuan Zhang, Qianli Liao, Alexander Rakhlin, Brando Miranda, Noah Golowich, and Tomaso Poggio. Theory of Deep Learning IIb: Optimization

## BIBLIOGRAPHY

- Properties of SGD. *arXiv:1801.02254 [cs]*, January 2018. arXiv: 1801.02254. [71](#)
- [134] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A Simple Way to Prevent Neural Networks from Overfitting. *Journal of Machine Learning Research*, 15:1929–1958, 2014. [71](#)
- [135] Li Wan, Matthew Zeiler, Sixin Zhang, Yann Le Cun, and Rob Fergus. Regularization of Neural Networks using DropConnect. In *International Conference on Machine Learning*, pages 1058–1066, February 2013. [71](#)
- [136] Alain Jungo, Raphael Meier, Ekin Ermis, Marcela Blatti-Moreno, Evelyn Herrmann, Roland Wiest, and Mauricio Reyes. On the Effect of Inter-observer Variability for a Reliable Estimation of Uncertainty of Medical Image Segmentation. In Alejandro F. Frangi, Julia A. Schnabel, Christos Davatzikos, Carlos Alberola-López, and Gabor Fichtinger, editors, *Medical Image Computing and Computer Assisted Intervention – MICCAI*, Lecture Notes in Computer Science, pages 682–690. Springer International Publishing, 2018. [73](#)
- [137] Sara M. Dupont, Benjamin De Leener, Manuel Taso, Arnaud Le Troter, Sylvie Nadeau, Nikola Stikov, Virginie Callot, and Julien Cohen-Adad. Fully-integrated framework for the segmentation and registration of the spinal cord white and gray matter. *NeuroImage*, 150:358–372, 2017. [73](#), [127](#)
- [138] Christian S. Perone, Evan Calabrese, and Julien Cohen-Adad. Spinal cord gray matter segmentation using deep dilated convolutions. *Scientific Reports*, 8(1):5966, April 2018. [73](#), [139](#)
- [139] Dominik Weishaupt, Victor D. Köchli, and Borut Marincek. *How does MRI work?: An Introduction to the Physics and Function of Magnetic Resonance Imaging*. Springer-Verlag, Berlin Heidelberg, 2 edition, 2006. [75](#)
- [140] Seth A. Smith, Richard A.E. Edden, Jonathan A.D. Farrell, Peter B. Barker, and Peter C.M. Van Zijl. Measurement of T1 and T2 in the Cervical Spinal Cord at 3 Tesla. *Magnetic resonance in medicine : official journal of the Society of Magnetic Resonance in Medicine / Society of Magnetic Resonance in Medicine*, 60(1):213–219, July 2008. [75](#)
- [141] Esha Datta, Nico Papinutto, Regina Schlaeger, Alyssa Zhu, Julio Carballido-Gamio, and Roland G. Henry. Gray matter segmentation of the spinal cord with active contours in MR images. *NeuroImage*, 147:788–799, February 2017. [75](#)
- [142] Daniel R. Messroghli, Aleksandra Radjenovic, Sebastian Kozerke, David M. Higgins, Mohan U. Sivananthan, and John P. Ridgway. Modified Look-Locker inversion recovery (MOLLI) for high-resolution T1 mapping of the heart. *Magnetic Resonance in Medicine*, 52(1):141–146, 2004. [76](#), [78](#), [79](#)

## BIBLIOGRAPHY

- [143] M. Weigel and O. Bieri. A Simple and Fast Approach for Spinal Cord Imaging at 3T with High In-Plane Resolution and Good Contrast. In *Proceedings of the 24th Annual Meeting of ISMRM, Singapore*, page 4408, May 2016. [76](#), [78](#), [79](#)
- [144] Oliver Bieri and Klaus Scheffler. Fundamentals of balanced steady state free precession MRI. *Journal of Magnetic Resonance Imaging*, 38(1):2–11, July 2013. [79](#)
- [145] C. Tsagkas, A. Horvath, A. Altermatt, S. Pezold, M. Weigel, T. Haas, M. Amann, L. Kappos, T. Sprenger, O. Bieri, P. Cattin, and K. Parmar. Automatic Spinal Cord Gray Matter Quantification: A Novel Approach. *American Journal of Neuroradiology*, August 2019. [87](#), [128](#)
- [146] Antal Horváth, Charidimos Tsagkas, Simon Andermatt, Simon Pezold, Katrin Parmar, and Philippe Cattin. Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA Images with MD-GRU. In Guoyan Zheng, Daniel Belavy, Yunliang Cai, and Shuo Li, editors, *Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computer Science, pages 3–14. Springer International Publishing, 2019. [87](#), [140](#)
- [147] Antal Horváth, Simon Pezold, Matthias Weigel, Katrin Parmar, and Philippe Cattin. High order slice interpolation for medical images. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 69–78. Springer, 2017. [92](#)
- [148] Antal Horváth, Christoph Jud, Simon Pezold, Matthias Weigel, Charidimos Tsagkas, Katrin Parmar, Oliver Bieri, and Philippe Cattin. A Principled Approach to Combining Inversion Recovery Images. In *Proceedings of the 26th Annual Meeting of ISMRM, Paris, France*, June 2018. [103](#)
- [149] Antal Horváth, Simon Pezold, Matthias Weigel, Katrin Parmar, Oliver Bieri, and Philippe Cattin. Variational segmentation of the white and gray matter in the spinal cord using a shape prior. In *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pages 26–37. Springer, 2016. [110](#), [123](#), [124](#), [125](#)
- [150] M. Lüthi, T. Albrecht, and T. Vetter. Probabilistic Modeling and Visualization of the Flexibility in Morphable Models. In *Mathematics of Surfaces XIII*, Lecture Notes in Computer Science, pages 251–264. Springer, Berlin, Heidelberg, September 2009. [124](#)
- [151] Carole H. Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M. Jorge Cardoso. Generalised Dice Overlap as a Deep Learning Loss Function for Highly Unbalanced Segmentations. In M. Jorge Cardoso, Tal Arbel, Gustavo Carneiro, Tanveer Syeda-Mahmood, João Manuel R.S. Tavares, Mehdi Moradi, Andrew Bradley, Hayit Greenspan, João Paulo Papa, Anant Madabhushi, Jacinto C. Nascimento, Jaime S. Cardoso, Vasileios Belagiannis, and Zhi

## BIBLIOGRAPHY

- Lu, editors, *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, Lecture Notes in Computer Science, pages 240–248. Springer International Publishing, 2017. [139](#), [153](#)
- [152] Christopher R. Tench, Paul S. Morgan, and Cris S. Constantinescu. Measurement of cervical spinal cord cross-sectional area by MRI using edge detection and partial volume correction. *Journal of Magnetic Resonance Imaging*, 21(3):197–203, March 2005. [140](#)
- [153] Adam Cadotte, David W. Cadotte, Micha Livne, Julien Cohen-Adad, David Fleet, David Mikulis, and Michael G. Fehlings. Spinal Cord Segmentation by One Dimensional Normalized Template Matching: A Novel, Quantitative Technique to Analyze Advanced Magnetic Resonance Imaging Data. *PLOS ONE*, 10(10):e0139323, October 2015. [140](#)
- [154] V. S. Fonov, A. Le Troter, M. Taso, B. De Leener, G. L  v  que, M. Benhamou, M. Sdika, H. Benali, P. F. Pradat, D. L. Collins, V. Callot, and J. Cohen-Adad. Framework for integrated MRI average of the spinal cord white and gray matter: The MNI-Poly-AMU template. *NeuroImage*, 102:817–827, November 2014. [140](#)
- [155] Chaitanya Baweja, Ben Glocker, and Konstantinos Kamnitsas. Towards continual learning in medical imaging. *arXiv:1811.02496 [cs]*, 2018. [157](#)

# Antal Horváth

---

## Curriculum Vitae

---

### Education

- 2008–2011 **Bachelor of Science in Mathematics**, *Department of Mathematics and Computer Science*, University of Basel.
- 2011–2014 **Master of Science in Mathematics**, *Department of Mathematics and Computer Science*, University of Basel.  
Major subjects in Numerical Analysis and Commutative Algebra.  
Master thesis in Explicit Low-dispersion Low-dissipation Runge-Kutta Schemes for the Wave Equation.
- 2015–2019 **Ph.D. in Biomedical Engineering**, *Department of Biomedical Engineering*, University of Basel.  
Topic: Spinal Cord Gray Matter White Matter Segmentation  
Supervisor: Prof. Dr. Philippe Cattin.

---

### References

#### Prof. Dr. Philippe C. Cattin

Department of Biomedical Engineering  
University of Basel  
Gewerbestrasse 14, 4123 Allschwil (Switzerland)  
✉ [philippe.cattin@unibas.ch](mailto:philippe.cattin@unibas.ch)  
☎ +41 61 207 54 00

#### Prof. Dr. Marcus J. Grote

Department of Mathematics and Computer Science  
University of Basel  
Spiegelgasse 1, 4051 Basel (Switzerland)  
✉ [marcus.grote@unibas.ch](mailto:marcus.grote@unibas.ch)  
☎ +41 61 207 39 96

---

### Publications

**Antal Horváth**, Simon Pezold, Matthias Weigel, Katrin Parmar, Oliver Bieri, and Philippe Cattin. Variational segmentation of the white and gray matter in the spinal cord using a shape prior. In *International Workshop on Computational Methods and Clinical Applications for Spine Imaging*, pages 26–37. Springer, 2016.

Simon Pezold, **Antal Horváth**, Ketut Fundana, Charidimos Tsagkas, Michaela Andělová, Katrin Weier, Michael Amann, and Philippe C Cattin. Automatic, robust, and globally optimal segmentation of tubular structures. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 362–370. Springer, 2016.

**Antal Horváth**, Simon Pezold, Matthias Weigel, Katrin Parmar, and Philippe Cattin. High order slice interpolation for medical images. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 69–78. Springer, 2017.

**Antal Horváth**, Christoph Jud, Simon Pezold, Matthias Weigel, Charidimos Tsagkas, Katrin Parmar, Oliver Bieri, and Philippe Cattin. A Principled Approach to Combining Inversion Recovery Images. In *Proceedings of the 26th Annual Meeting of ISMRM, Paris, France*, June 2018.

Charidimos Tsagkas, **Antal Horváth**, Matthias Weigel, Tanja Haas, Ludwig Kappos, Till Sprenger, Oliver Bieri, Philippe Cattin, and Katrin Parmar. Reliability Of Automatic Spinal Cord Gray Matter Segmentation Using Averaged Magnetization Inversion Recovery Acquisitions. In *Proceedings of the 13th Annual ARSEP MRI Workshop, Paris, France*, pages 28–29, 2018.

**Antal Horváth\***, Charidimos Tsagkas\*, Simon Andermatt, Simon Pezold, Katrin Parmar, and Philippe Cattin. Spinal cord gray matter-white matter segmentation on magnetic resonance AMIRA images with MD-GRU. In *Computational Methods and Clinical Applications for Spine Imaging*, Lecture Notes in Computer Science, pages 3–14. Springer, 2019.

Simon Andermatt, **Antal Horváth**, Simon Pezold, and Philippe Cattin. Pathology Segmentation Using Distributional Differences to Images of Healthy Origin. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, Lecture Notes in Computer Science, pages 228–238. Springer International Publishing, 2019.

Charidimos Tsagkas\*, **Antal Horváth\***, Anna Altermatt, Simon Pezold, Matthias Weigel, Tanja Haas, Michael Amann, Ludwig Kappos, Till Sprenger, Oliver Bieri, Philippe Cattin,

---

Asterisks (\*) denote equally contributing first authors.



and Katrin Parmar. Automatic Spinal Cord Gray Matter Quantification: A Novel Approach. *American Journal of Neuroradiology*, August 2019.

## Posters

- 2016 Antal Horváth, Simon Pezold, Matthias Weigel, Katrin Parmar, Oliver Bieri, and Philippe Cattin. Variational Segmentation of the White and Gray Matter in the Spinal Cord using a Shape Prior, *CSI Satellite Event MICCAI 2016*.
- 2017 Antal Horváth, Simon Pezold, Matthias Weigel, Katrin Parmar, and Philippe Cattin. High-Order Slice Interpolation for Medical Images, *SASHIMI Satellite Event MICCAI 2017*.
- 2018 Antal Horváth, Christoph Jud, Simon Pezold, Matthias Weigel, Charidimos Tsagkas, Katrin Parmar, Oliver Bieri, and Philippe Cattin. A Principled Approach to Combining Inversion Recovery Images, *ISMRM 2018*.
- 2018 Antal Horváth, Charidimos Tsagkas, Simon Andermatt, Simon Pezold, Katrin Parmar, and Philippe Cattin. Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA images with MD-GRU, *DBE Research Day 2018*.

## Talks

- 2015 **Segmentation and Classification of Inner Structures of the Spinal Cord on MR Images with Continuous Cuts and Principal Component Analysis**, *Bernoulli's Tafelrunde*, Department of Mathematics and Computer Science, University of Basel, March 26 2015.
- 2016 **Variational Segmentation of the White and Gray Matter in the Spinal Cord**, *Special Interest Group*, Allschwil, Switzerland, March 10 2016.
- 2016 **Variational Segmentation of the White and Gray Matter in the Spinal Cord using a Shape Prior**, *CSI Satellite Event MICCAI*, Athens, Greece, October 17 2016.
- 2017 **High-Order Slice Interpolation for Medical Images**, *SASHIMI Satellite Event MICCAI*, Quebec City, Canada, September 10 2017.
- 2018 **Spinal Cord Gray Matter-White Matter Segmentation on Magnetic Resonance AMIRA images with MD-GRU**, *CSI Satellite Event MICCAI*, Granada, Spain, September 16 2018.