Computational Modeling of Medical Images of Brain Tumor Patients for Optimized Radiation Therapy Planning

Agn, Mikael; Van Leemput, Koen; Larsen, Rasmus

Publication date:
2017

Document Version
Publisher's PDF, also known as Version of record

Computational Modeling of Medical Images of Brain Tumor Patients for Optimized Radiation Therapy Planning

Mikael Agn

Kongens Lyngby 2016
PHD-2016-442

Technical University of Denmark
Department of Applied Mathematics and Computer Science
Richard Petersens Plads, building 324,
2800 Kongens Lyngby, Denmark
Phone +45 4525 3031
compute@compute.dtu.dk
www.compute.dtu.dk

PHD-2016-442. ISSN: 0909-3192

Summary (English)

In brain tumor radiation therapy, the aim is to maximize the delivered radiation dose to the targeted tumor and at the same time minimize the dose to sensitive healthy structures – so-called organs-at-risk (OARs). When planning a radiation therapy session, the tumor and the OARs therefore need to be delineated on medical images of the patient's head, to be able to optimize a radiation dose plan. In clinical practice, the delineation is performed manually with limited assistance from automatic procedures, which is both time-consuming and typically suffers from poor reproducibility. There is, therefore, a need for automated methods that can segment both brain tumors and OARs. However, there is a noticeable lack in the literature of methods that simultaneously segment both types of structures.

To automatically segment medical images of brain tumor patients is difficult because brain tumors vary greatly in size, shape, appearance and location within the brain. Furthermore, healthy structures surrounding a tumor are pushed and deformed by the so-called mass effect of the tumor. Moreover, medical imaging techniques often result in imaging artifacts and varying intensity across imaging centers.

The goal of this PhD-project was to develop automated segmentation methods that can handle both brain tumors and OARs. In the first part of the project, we developed a model for tumor shape and used it to develop a fully automated generative method specifically for brain tumor segmentation. This method performed favorably compared to other state-of-the-art methods. In the second part of the project, we used a probabilistic atlas-based model capable of detailed modeling of the spatial organization in a healthy brain, and extended it

to handle various OARs. We incorporated this model into the previously used modeling framework. In experiments, we showed that the resulting model was capable of simultaneous segmentation of brain tumors and OARs, while also being capable of adapting to varying image sequences and images from different imaging centers.

Summary (Danish)

I strålebehandling af hjernetumorer er målet at maksimere leveret strålingsdosis til den målrettede tumor og samtidig minimere dosis til følsomme sunde strukturer - såkaldte organs-at-risk (OARs). Når man planlægger en stråleterapi session skal derfor tumoren og de sunde strukturerne afgrænses på medicinske billeder af patientens hoved, for at kunne optimere en strålingsdosisplan. I klinisk praksis udføres afgrænsningen manuelt med begrænset hjælp fra automatiske procedurer, hvilket er både tidskrævende og typisk fører til dårlig reproducerbarhed. Der er derfor behov for automatiserede metoder, der kan segmentere både hjernetumorer og OARs. Der er imidlertid en mærkbar mangel i litteraturen af metoder, der samtidig kan segmentere begge typer af strukturer.

At automatisk segmentere medicinske billeder af hjernetumorpatienter er vanskeligt, fordi hjernetumorer varierer meget i størrelse, form, udseende og placering i hjernen. Endvidere skubbes og deformeres sunde strukturer af tumoren. Desuden resulterer medicinske scanningsteknikker ofte i artefakter i billederne og varierende intensitet på tværs af billeddiagnostiske centre.

Målet med dette PhD-projekt var at udvikle automatiserede segmenteringsmetoder, der kan håndtere både hjernetumorer og OARs. I den første del af projektet udviklede vi en model der beskriver udformningen af tumorer, og brugte den til at udvikle en fuldautomatisk generativ metode til segmentering af hjernetumorer. Denne metode præsterede fordelagtigt i forhold til andre state-of-the-art metoder. I den anden del af projektet brugte vi en probabilistisk atlasbaseret model, som er i stand til detaljeret modellering af den rumlige organisation i en sund hjerne, og udvidede den til at håndtere forskellige OARs. Vi inkorporerede denne model i den tidligere anvendte modelleringsramme. Vi viste i eks-

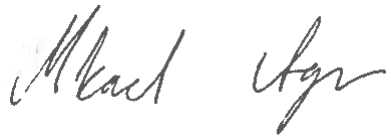
perimenter, at den resulterende model var i stand til samtidig segmentering af hjernetumorer og OARs. Den var også i stand til at tilpasse sig til varierende billedsekvenser og billeder fra forskellige billeddiagnostiske centre.

Preface

This thesis was prepared at the *Department of Applied Mathematics and Computer Science* at the *Technical University of Denmark* in partial fulfillment of the PhD degree requirements. From this department, associate professor Koen Van Leemput acted as main supervisor and professor Rasmus Larsen acted as co-supervisor. The PhD project was performed in collaboration with the *Department of Oncology* and the *Department of Clinical Physiology, Nuclear Medicine and PET* at *Rigshospitalet* in Denmark. From Rigshospitalet, Per Munck af Rosenschöld and Ian Law acted as co-supervisors. Per Munck af Rosenschöld is the head of research at the *Department of Oncology* and associate professor at the *Niels Bohr Institute* at *University of Copenhagen*. Ian Law is the chief physician at the *Department of Clinical Physiology, Nuclear Medicine and PET* and professor at the *Faculty of Health and Medical Sciences* at *University of Copenhagen*. The thesis was made possible with funding from the *Lundbeck fundation*.

The thesis deals with automated segmentation methods of medical images of brain tumor patients for use in radiation therapy planning.

Lyngby, 30-November-2016



Mikael Agn

Acknowledgements

Firstly, I would like to thank my main supervisor Koen Van Leemput for his patience, guidance and vast knowledge in the field of medical imaging and generative modeling. Without his extensive insight into the field, the project would not have progressed as far as it did. Secondly, I would like to thank my co-supervisors at the hospital, Per Munck af Rosenschöld and Ian Law, for their invaluable guidance in all matters relating to clinical practice, medical images and treatment of brain tumors. I would also like to thank my co-supervisor Rasmus Larsen for guidance in more practical matters at the university section. Thanks also to the Lundbeck foundation for financially supporting the project.

Many thanks goes to John Ashburner for inviting me to UCL in London for my external stay. We had many valuable discussions relating to generative modeling from a slightly different perspective than with my main supervisor. I would also like to thank his PhD students Claudia Blaiotta and Mikael Brudfors. Many thanks also to Jorge Cardoso – whom John introduced me to – for inspiring discussions, valuable input and for further introducing me to Laura Mancini and Sotirios Bisdas. Laura and Sotirios showed great interest in my work and provided me with further clinical data to test my methods on, so many thanks go to them as well.

I am grateful towards my colleagues for creating a motivating working environment and especially for many inspiring discussions and valuable input. Special thanks go to Oula Puonti, Daniel Andreasen, Christian Thode Larsen and Mark Lyksborg at the university section; and Michael Lundemann Jensen at the hospital, who should also be thanked for his great assistance when collecting clinical data at the hospital.

Above all, I wish to thank my family and boyfriend for their immense support throughout the project.

Scientific Contributions

Papers included in this thesis

Paper A: Agn, M., Puonti, O., Munck af Rosenschöld, P., Law, I., and Van Leemput, K. (2016). Brain tumor segmentation using a generative model with an RBM prior on tumor shape. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, Lecture Notes in Computer Science, volume 9556, pages 168–180. Springer.

This work was presented in the 2015 Multimodal Brain Tumor Segmentation Challenge – at the Brainles workshop held in conjunction with the conference MICCAI 2015 – where it ranked the 3rd overall best tumor segmentation method and the best among fully automated methods. We were therefore invited to submit this peer-reviewed article to the post-proceedings of the workshop.

Paper B: Agn, M., Law, I., Munck af Rosenschöld, P., and Van Leemput, K. (2016). A generative model for segmentation of tumor and organs-at-risk for radiation therapy planning of glioblastoma patients. In *Proceedings of conference SPIE Medical Imaging 2016: Image Processing*, volume 9784, pages 97841D–97841D-9.

Paper C: Agn, M., Munck af Rosenschöld, P., Mancini, L., Ashburner, J., Law, I., and Van Leemput, K. (2016). Simultaneous segmentation of gliomas and organs-at-risk by a sequence-adaptive generative method for radiotherapy planning. *Journal manuscript.*

Paper D: Lyksborg, M., Puonti, O., Agn, M., and Larsen, R. (2015). An ensemble of 2D convolutional neural networks for tumor segmentation. In *Proceedings of the 19th Scandinavian Conference on Image Analysis, SCIA 2015*, Lecture Notes in Computer Science, volume 9127, pages 201-211. Springer.

Papers excluded from the thesis

- Agn, M., Svarer, C., Frokjaer, V. G., Greve, D. N., Knudsen, G. M., and Van Leemput, K. (2014). Improved resolution and reliability in dynamic PET using Bayesian regularization of MRTM2. In *Proceedings of the 2014 IEEE 11th International Symposium on Biomedical Imaging (ISBI)*, pages 955–958.
- Frokjaer, V. G., Pinborg, A., Holst, K. K., Overgaard, A., Henningsson, S., Heede, M., Larsen, E. C., Jensen, P. S., Agn, M., Nielsen, A. P., and Stenbæk, D. S. (2015). Role of serotonin transporter changes in depressive responses to sex-steroid hormone manipulation: a positron emission tomography study. *Biological psychiatry*, 78(8), pages 534-543.

Contents

Chapter 1

Introduction

The goal of this PhD project was to develop segmentation methods that can be of use when planning a radiation therapy session to treat patients with brain tumors. Although radiation therapy can be used to treat many types of brain tumors, we focused mainly on primary glioblastomas, which are the most common and severe type of tumors originating within the brain [Preusser et al., 2011]. These high-grade tumors are typically treated with a combination of surgical removal of tumor tissue, chemotherapy and radiation therapy [Chinot et al., 2014]. In a radiation therapy session, the patient's head is subjected to radiation beams from different directions, with the aim of maximizing the delivered radiation dose to the targeted tumor while minimizing the dose to sensitive healthy structures – so-called organs-at-risk [Shaffer et al., 2010].

In the planning of a radiation therapy session, a radiation dose plan need to be optimized. To be able to optimize this plan, the location of the tumor target and the organs-at-risk need to be known. Therefore, a number of medical images are acquired of the patient's head and the structures of interest are subsequently delineated on these images. In clinical practice, the delineation is performed manually by a radiologist with limited assistance from automatic procedures. This manual procedure is time consuming for the radiologist and it is known that – even for radiologists – the inter-rater variability is high [Dolz et al., 2015b, Deeley et al., 2011, Menze et al., 2015]. There is therefore a need for automated methods that are capable of segmenting both the tumor

target and the organs-at-risk. However, there is a notable lack in the literature of methods for simultaneous brain tumor and organ-at-risk segmentation.

To automatically segment medical images of brain tumor patients is difficult for two main reasons. The first reason is that tumors in themselves are difficult to model. Brain tumors vary greatly in size, shape and location within the brain. Furthermore, tumors might have been partly removed by surgery in a previous treatment stage. The appearance of tumor tissue in medical images is also complicated due to varying biological changes both between tumors and within one tumor. The second reason is that healthy structures surrounding a tumor are pushed and deformed by the growth of the tumor, due to the so-called mass effect. Many brain tumor patients are also older with co-occurring age-related abnormalities in the brain. Moreover, the medical imaging techniques often result in imaging artifacts and varying intensity across imaging centers. For these reasons, computational modeling of medical images for brain tumor patients is both a challenging and interesting task.

1.1 Contributions

For paper A, we developed a generative method specifically focused on brain tumor segmentation. We participated with the method in the Brain Tumor Segmentation (BRATS) challenge at the Brainles workshop held in conjunction with the 2015 MICCAI conference. In that challenge, our method was among the top-performing methods. On tumor core – which is the region of interest in radiation therapy – our method performed particularly well, obtaining the highest average Dice score among all methods.

In paper B, we proposed an extension of the previous method aimed to handle organs-at-risk. To be able to also model organs-at-risk, we incorporated a whole-brain probabilistic atlas into the modeling framework. This atlas has previously been thoroughly validated for detailed whole-brain segmentation on healthy subjects. We showed the feasibility of our approach on a small data set of glioblastoma patients by evaluating the segmentation of tumor core and two organs-at-risk already included in the atlas – namely hippocampus and brainstem.

In paper C, we presented the full generative method for simultaneous brain tumor and organs-at-risk segmentation. To increase the relevance for radiation therapy planning, we included several other important organs-at-risk in the probabilistic atlas model. Primarily, we extended the whole-brain atlas to model non-brain structures in the eye socket region. In this paper, we described the

full modeling framework in more detail, showed the flexibility of the method in including new medical images and validated the segmentation performance on a larger data set and for additional organs-at-risk. To the best of our knowledge, a simultaneous brain tumor and organs-at-risk segmentation method has not been presented before.

In Paper D, an approach based on convolutional neural networks (CNNs) was presented for brain tumor segmentation. Although 3D CNNs have been used in a few approaches for brain tumor segmentation, CNNs are typically trained in 2D to reduce the complexity of the filters that need to be learned, and to lower the computational burden. However, when a CNN is trained only on 2D patches from e.g., transversal slices, the connectivity between the slices is lost. To improve on the connectivity between slices, the approach in this paper uses an ensemble of 2D CNNs. The segmentation is performed in three steps. For the first step, three CNNs are trained for segmenting the whole tumor region – one for each image plane, i.e., the transverse, sagittal and coronal plane. A segmentation is then obtained by performing voxel-wise majority voting on the outputs of these three CNNs. In the second step, the segmentation is refined using a cellular automaton-based seed growing method called growcut. Finally, in the third step, tumor core and enhanced core are segmented within the whole tumor region by an additional ensemble of 2D CNNs. The method was shown to give competitive performance on the test data for the 2014 BRATS challenge. In this work, I helped mainly with formulating the method on a general level, writing parts of the introduction and also proof-reading the paper.

1.2 Overview of thesis

This thesis consist an overview of the research conducted during the PhD project and the four papers presented in the previous section. The overview is designed to act as a complement to the papers. Here, we will present the developed methods on a higher level of abstraction and further discuss the background, research and some aspects of the developed models not included in the papers. The overview is divided into the following chapters:

- Chapter 2 presents an overview of the radiation therapy treatment procedure and the typical medical images acquired during the planning of a radiation therapy session.
- Chapter 3 presents an overview of current methods related to the segmentation of medical images for brain tumor patients.

- Chapter 4 describes the first part of the PhD project, where we focused exclusively on modeling and segmenting brain tumors. This chapter mainly describes the research related to paper A.
- Chapter 5 describes the second part of the PhD project, where we focused on the modeling of organs-at-risk and the joint segmentation of brain tumors and organs-at-risk. This chapter describes the research related to papers B and C.
- Chapter 6 provides the conclusion of the thesis, where we summarize the conducted research and the developed methods. Here, we also discuss possible ways forward to further develop the methods presented in the thesis.

Chapter 2

Radiation therapy treatment of brain tumors

This chapter describes the radiation therapy treatment procedure and the medical images acquired when planning a radiation therapy session. The chapter is divided into the following sections:

- The first section focuses on the medical images acquired to visualize brains with tumors, which are needed when planning a radiation therapy session.
- The second section briefly describes the planning and realization of a radiation therapy session.

2.1 Medical images of brain tumor patients

When planning a radiation therapy session various medical images are acquired of the patient's head. These images are acquired with various imaging techniques or *modalities*, such as X-ray computed tomography (CT), magnetic resonance imaging (MRI) and positron emission tomography (PET). An example of such a set of images can be seen in figure 2.1. A CT scan is crucial for computing the optimal radiation dose distribution, as it can be used to calculate the attenuation

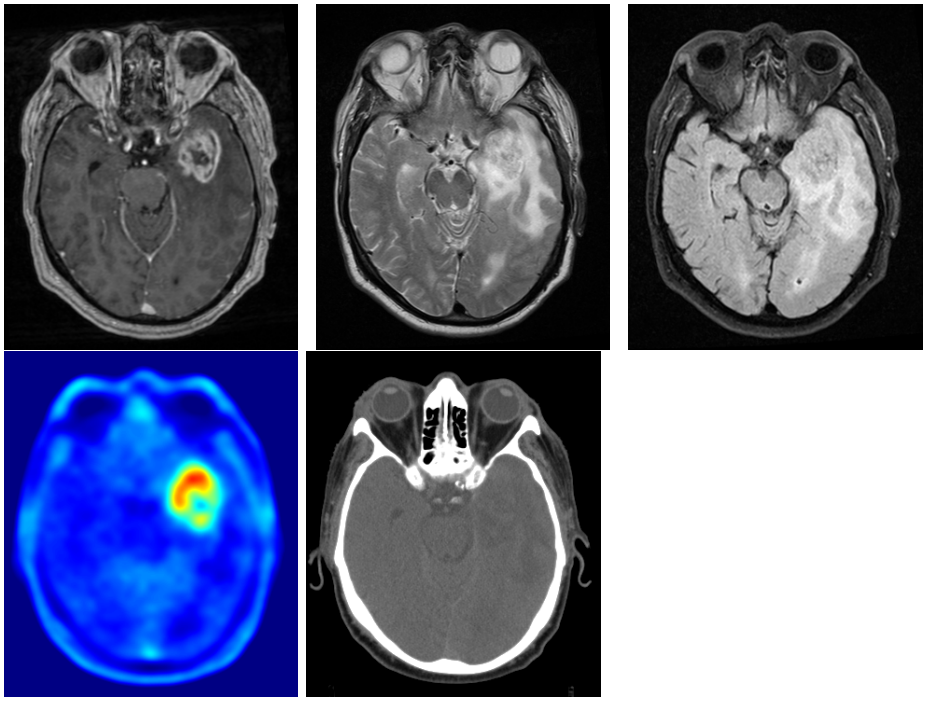


Figure 2.1: Set of medical images often acquired in radiation therapy planning of glioblastomas. Upper row from left to right: MR-sequences T1c, T2 and Flair. Lower row: FET-PET on the left, CT on the right.

and scatter of high-energy photon radiation passing through tissues in the head. However, CT scans have rather poor soft tissue contrast and are thus not optimal for segmenting brain structures. Instead, Magnetic resonance (MR) scans are widely used in order to segment the tumor target and the organs-at-risk, as this imaging modality has excellent soft tissue contrast. Additionally, other imaging modalities can be used for visualizing certain aspects of tumor tissue, such as PET that can visualize metabolic information.

MRI uses the fact that hydrogen atoms in an external magnetic field can absorb and emit radio-frequency radiation. Soft tissue in the human body contains large amounts of hydrogen atoms, which behave differently depending on their composition in a tissue. Therefore, MRI can be used to acquire images with a good contrast between different types of soft tissues. By varying acquisition parameters in a so-called *pulse sequence*, soft tissue contrasts highlighting certain tissue differences can be obtained. To visualize brains with tumors, a series of sequences are typically acquired, detailed in an imaging protocol. This series

of acquired images is also referred to as a multi-sequence image. The standard sequences are detailed in table 2.1.

Table 2.1: Sequence types used in standard MRI protocols for gliomas.

Sequence short name	
T1	T1-weighted sequences are the most commonly used for segmenting healthy brain tissue, since they result in images with a good contrast between gray and white matter structures in the brain. Especially the 3D-acquired sequence MPRAGE results in high-quality images with a good resolution.
T1c	In a contrast-enhanced T1-weighted image, a contrast-enhancing agent is injected into the patient's blood stream prior to scanning. The regions where the agent accumulates will then have an enhanced contrast in the image. Gadolinium is commonly used for patients with high-grade gliomas. This agent cannot cross the intact blood-brain barrier. Brain regions with a damaged blood-brain barrier will thus be contrast enhanced, which is a characteristic of high-grade gliomas.
T2	In comparison to T1-weighted sequences, T2-weighted sequences obtain a superior visualization of many abnormalities – such as edema and tumor tissue. However, the bright signal intensity of abnormalities in a T2 image is similar to that of free-flowing fluid such as cerebrospinal fluid (CSF).
FLAIR	A T2-weighted fluid attenuated inversion recovery (*FLAIR*) sequence suppresses the signal of free-flowing fluid, which results in a clear visualization of abnormalities in the brain. As the sequence is fairly time consuming, a scan is usually acquired in 3-5 mm thick 2D slices. Optimized 3D-acquisition FLAIR sequences do exist, but are not yet common in clinical brain tumor imaging protocols.

The specific acquisition parameters used for each type of sequence vary between imaging centers and scanners. Even when using the same acquisition parameters, the intensity of acquired MR images can vary significantly. Furthermore, MR images are affected by many artifacts, such as motion and flow artifacts due to the often long acquisition time. The most significant type of artifact is perhaps so-called *bias fields*, which appear as smoothly varying changes in intensity over the image. Another issue is the partial volume effect (PVE) when

different types of tissue are mixed in one image voxel.

Various MR sequences are being researched for brain tumor visualization, e.g., T2-weighted double inversion recovery (DIR) sequences that suppress signals of both fluid and white matter, which further increases the visualization of abnormalities. More advanced MR sequences – such as diffusion-weighted MRI, perfusion-weighted MRI and magnetic resonance spectroscopic imaging – are also being researched for this purpose [Sauwen et al., 2016].

PET is becoming more common in radiation therapy planning. Using PET, physiological images can be obtained by injecting a radioactive compound in the bloodstream of a patient. The compound is designed to get attached to certain molecules of interest and send out radiation which is subsequently recorded. As the recorded signal is fairly weak, PET images have a low resolution. ^{18}F-FET is an often-used tracer that provides information on tumor metabolism [Poulsen et al., 2016]. An example slice of FET-PET is shown in figure 2.1.

2.2 Radiation therapy treatment planning

Radiation therapy is an important treatment modality of glioblastomas, which is used in combination with chemotherapy and surgery. In a radiation therapy session, the tumor is radiated with high-energy photon radiation in order to damage tumorous cells. However, the radiation will also damage healthy tissue. The goal of the treatment is therefore to deliver a high radiation dose to the tumor target, while keeping the radiation dose low for surrounding healthy tissue. Ideally, one would like to deliver a large amount of radiation to all tumorous cells but no radiation at all to healthy cells. There are two main reasons why this is not possible. The first reason is that currently used medical images are not able to visualize the location of all tumorous cells. The second reason is that it is not possible to deliver such a precise dose using conventional external beam linear accelerator-based photon radiation therapy. Advanced radiation therapy equipment can deliver a dose at millimeter-precision by delivering radiation beams from varying directions with different intensity profiles, e.g., intensity-modulated arc therapy [Cedric and Tang, 2011, Munck af Rosenschöld et al., 2011] for which an example is shown in figure 2.2.

The planning of a treatment session aims to find a balance between irradiating tumor tissue while sparing important healthy structures from radiation. These important healthy structures are called organs-at-risk. In the head, there are many important and sensitive structures that can be considered organs-at-risk [Munck af Rosenschöld et al., 2015]. The optic system – which includes the

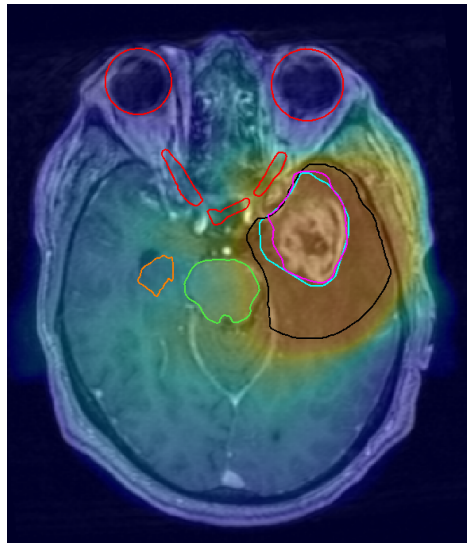


Figure 2.2: Radiation dose plan overlayed on a T1c image. High radiation in red, low radiation in blue/violet. Lines show edges of important structures: optic system in red, hippocampus in orange, brainstem in green, MR-defined gross tumor volume in magenta, PET-defined gross tumor volume in cyan and clinical tumor volume in black.

eyes, optic nerves, optic chiasm and optic tracts – and brainstem are important organs-at-risk. Sometimes, hippocampi are also included. Figure 2.3 illustrates these structures in three-dimensions from different angles. In order to deliver the optimal radiation dose, these organs-at-risk and the tumor target need to be delineated on the acquired medical images.

In clinical practice, the delineation is done manually with limited assistance from automatic procedures. However, manual delineation is time consuming and typically suffers from poor reproducibility [Dolz et al., 2015b, Deeley et al., 2011, Menze et al., 2015]. The typical mean time for a radiologist to analyze and delineate brain tumor images has been reported to 86 minutes [Dolz et al., 2015b]. Furthermore, the resulting delineation can vary considerably from one manual rater to another [Dolz et al., 2015b]. This could affect the outcome of the radiation therapy treatment and complicate studies in comparing the outcome of different radiation therapy approaches. There is therefore a need for automated segmentation methods that can segment both the tumor target and organs-at-risk – to save valuable time for the human experts involved and to obtain a more consistent delineation result. Furthermore, new approaches to radiation

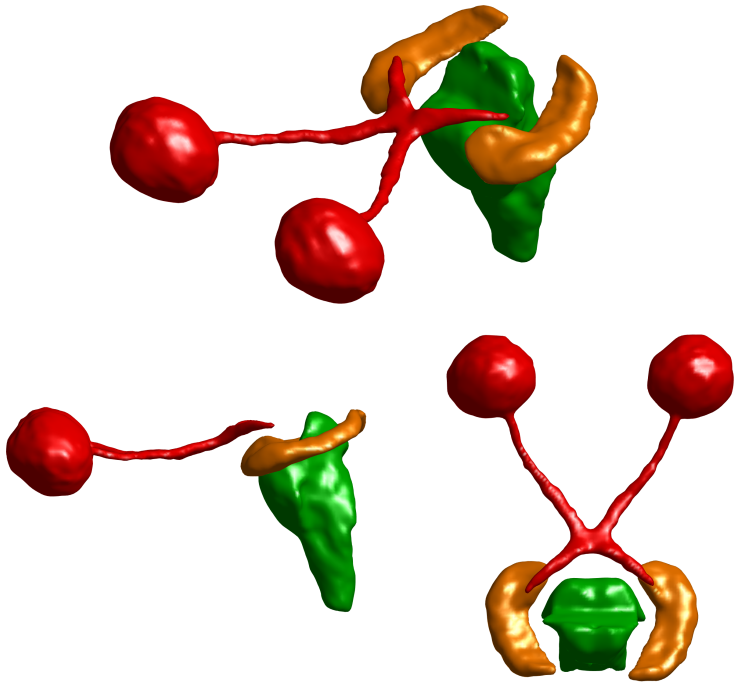


Figure 2.3: Surface plots of important organs-at-risk from three different viewing angles. Optic system in red including eyes, optic nerve, optic chiasm and optic tract. Hippocampi in orange. Brainstem in green.

therapy are being researched where MRI images are provided at the same time as the treatment to enable more effective tumor targeting and organs-at-risk sparing [Lagendijk et al., 2016]. In such an approach, automated segmentation would be invaluable.

Segmentation of tumor. As described in the previous section, the tumor is typically visualized with a number of MR sequences, but also PET is sometimes used. The different types of images visualize different aspects of the tumor, but none of them visualizes the full extent of tumor cells. The imaged tumor-affected tissue can be divided into tumor core and peritumoral edema. *Peritumoral edema* is tissue surrounding the tumor core that has been affected by the tumor due to e.g., the pushing of healthy tissue that occurs when a tumor grows. However, edema does not necessarily contain tumor cells and the visualization of edema in e.g., a FLAIR sequence is not an indicator of the

presence of tumor cells. The *tumor core* is the region of interest when segmenting a tumor from medical images. The tumor core is called gross tumor volume (GTV) in clinical practice [Burnet et al., 2004]. The GTV is the tumor that can be visualized by medical images. The border of the GTV may differ between MRI and PET, so they are typically segmented separately and then merged, as can be seen in figure 2.2 [Munck af Rosenschöld et al., 2015]. Because the GTV might not include all tumor cells, it is extended with an up to 20 mm large margin within the brain to form the clinical tumor volume (CTV) [Burnet et al., 2004, Munck af Rosenschöld et al., 2015], which can be seen in figure 2.2. The CTV is then extended with a small safety margin designed to allow for uncertainties in planning and treatment delivery to form the planning tumor volume (PTV) – which is then used when planning the radiation dose.

In this PhD project, the focus is on segmenting the MR-defined GTV, which we will call the tumor core. The definition of the GTV for glioblastomas is focused on the enhanced tumor tissue in T1c, i.e., where the blood-brain barrier is damaged. It also includes other regions in direct connection to the enhanced tumor: necrotic tumor tissue (seen with dark intensity in figure 2.2), cavities from surgically removed tumor, and unenhanced tumor tissue directly connected to the enhanced tumor tissue.

Chapter 3

Overview of segmentation methods for medical images

This chapter gives an overview of current automatic and semi-automatic approaches for segmenting medical images of human heads, i.e., the process of dividing an image of the head into multiple meaningful segments. We only list a subset of available approaches with the purpose of relating the research in this thesis to the research field. Therefore, we will here focus mainly on segmentation of brain tumors and organs-at-risk, but also discuss more general whole-brain segmentation methods. As there is a noticeable lack in the literature of methods for simultaneous segmentation of brain tumors and organs-at-risk, we will discuss the two types of structures separately. We will start by describing segmentation methods for brain images on a general level (section 3.1). We will then describe methods for brain tumor segmentation (section 3.2). Finally, we will describe methods for healthy tissue segmentation and specifically for organs-at-risk in radiation therapy (section 3.3).

3.1 General description of segmentation methods for brain images

On a general level, segmentation methods are often explained in dichotomies based on their underlying models. In this thesis, we will use the dichotomies: *discriminative vs. generative*, *supervised vs. unsupervised*, and *parametric vs. nonparametric*.

Say we have a multi-sequence MR image denoted by $\mathbf{D}$. We want to find an optimal segmentation $\mathbf{l}$ given the image data in $\mathbf{D}$, with all image voxels divided into a number of anatomically relevant segments. Here, $\mathbf{l}$ will contain one segment label for each image voxel. From a probabilistic perspective, a *discriminative* approach aims to find a model that directly predicts $\mathbf{l}$ from $\mathbf{D}$, e.g., by modeling the conditional probability distribution $p(\mathbf{l}|\mathbf{D})$. The name comes from the fact that the model will learn directly how to discriminate between different segmentation classes given the image intensity information. A limitation of discriminative models is that they typically need a large amount of manually segmented training data to train the model. Here, training the model refers to estimating proper values for the parameters of the model. Furthermore, as these models explicitly use the intensity information in the training data, they are restricted to segmenting images with the same image contrast. In MRI the intensity information can vary significantly even within one type of MRI sequence. Therefore, even within one type of sequence, careful preprocessing is crucial – such as intensity normalization [Nyúl et al., 2000, Roy et al., 2013], bias field correction [Tustison et al., 2010, Larsen et al., 2014] and skull-stripping [Ségonne et al., 2004, Fennema-Notestine et al., 2006]. To extend a discriminative method to handle other imaging sequences or modalities, new manually segmented training data are often needed. Alternatively, contrast synthesis [Roy et al., 2013, Iglesias et al., 2013a, Andreasen et al., 2015] could be applied, where an image is not directly segmented but rather used to synthesize a new image with the desired image contrast. However, this still typically requires new data containing both image types to learn a proper synthesis process.

Generative approaches, on the other hand, aim to build a cohesive model of the formation process of the image to be segmented, modeling the joint probability distribution $p(\mathbf{l}, \mathbf{D})$, typically parameterized as $p(\mathbf{l}, \mathbf{D}) = p(\mathbf{D}|\mathbf{l})p(\mathbf{l})$. Because a generative model is a full probabilistic model of both the labels $\mathbf{l}$ and the image data $\mathbf{D}$, it can be used to generate any of them. An image can be segmented by applying Bayes' rule to obtain the conditional distribution $p(\mathbf{l}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{l})p(\mathbf{l})$. The likelihood function $p(\mathbf{D}|\mathbf{l})$ models the image intensities given the labels, usually in the form of Gaussian mixture models (GMMs) where each label is

connected to some Gaussian distributions. Because the distribution of image intensities is explicitly modeled, a generative model can often be made to adapt to the intensities in the image to be segmented. Thus, a generative model often does not need to be fully retrained to handle new image sequences or images from different scanners. The prior probability distribution $p(\mathbf{l})$ models prior knowledge about the structures to be segmented such as the shape, size and relative spatial organization of structures. The label prior could for example be a model based on a physical process or a structural model trained with manual segmentations from training data. A prior model for healthy brain structures can often be in the form of a probabilistic atlas that encodes the spatial organization of healthy structures in a brain. The resulting method is then called an atlas-based method. The atlas can either be stationary – where it is just affinely co-register with the target subject – or deformable, meaning that it is capable of deforming during the segmentation process to fit the target subject. An inconvenience with generative models is the difficulty to transform human prior knowledge into appropriate probabilistic models.

In *supervised* learning labeled data is used to train a model, which can subsequently be used to label new data. Discriminative approaches are typically supervised. In *unsupervised* learning, on the other hand, a model is used to find structure in unlabeled data. The different parts that a generative model is comprised of are often largely unsupervised by themselves – e.g., a GMM clusters unlabeled data and a prior model can find structure in segmentations. Lastly, *nonparametric* methods – in contrast to *parametric* methods – do not summarize training data into a set of parameters but instead use the training data directly to segment an image. This can for example be done in a so-called multi-atlas approach, where segmented training images are first registered to the target image. The corresponding segmentations are then warped using the same transformations and then fused to obtain a final segmentation of the target subject. Many well-performing brain segmentation methods are non-parametric, which will be described in the last section of this chapter.

3.2 Methods for segmenting brain tumors

Many methods exist with a singular focus on brain tumor segmentation – with some achieving a good although not very robust performance. In particular, the annual Brain Tumor Segmentation (BRATS) challenge held the first time in conjunction with the 2012 MICCAI conference has benchmarked several well-performing methods [Menze et al., 2015]. The methods are best divided into discriminative and generative methods.

By far most brain tumor segmentation methods are discriminative. Therefore, they typically rely heavily on the intensity information of annotated training data, which is used to directly learn how to discriminate between the appearance of tumor and that of other tissue. To incorporate spatial context, different strategies are employed. Many methods rely on user-engineered image features that are fed into a classifier, such as random forests, e.g., [Tustison et al., 2015, Islam et al., 2013, Zikic et al., 2012, Maier et al., 2016] or support vector machines [Bauer et al., 2011]. Recently, convolutional neural networks (CNNs) have successfully been employed for brain tumor segmentation, e.g., [Pereira et al., 2016, Havaei et al., 2016, Kamnitsas et al., 2016]. The key to their success is their capability of automatically learning image features from training data. CNNs learn these features in a deep hierarchy that can capture complex interactions in the data. In paper D of this thesis, a CNN method was developed for brain tumor segmentation [Lyksborg et al., 2015]. All of these methods are parametric. A few non-parametric methods have also been used for brain tumor segmentation, e.g. [Cordier et al., 2016], in a so-called patch-based approach. Patch-based methods make use of annotated image patches from training data that are compared directly with the images to be segmented.

Generative methods for brain tumor segmentation have generally been atlas-based, wherein they typically rely on a stationary probabilistic atlas to model healthy tissues surrounding brain tumors. The difficulty is then to incorporate prior information about tumor tissue into the model. Tumor tissue has been modeled as outliers in the intensity model, e.g., in [Menze et al., 2010], or with more involved models on tumor growth that deforms the atlas according to the tumor mass-effect, e.g., in [Kwon et al., 2014, Gooya et al., 2012]. To improve performance, these methods have also been used with a subsequent discriminative step in [Menze et al., 2016, Bakas et al., 2016]. In [Sanjuán et al., 2013], an extra class prior is added to the atlas designed to pick up abnormal image voxels. The method proceeds to segment the subject, and then subsequently uses a fuzzy clustering outlier detection procedure. This procedure compares the resulting probability maps for healthy tissue to healthy controls. The output of this procedure is then used to refine the extra class prior, and the algorithm is run a second time with the updated prior.

3.3 Methods for segmenting healthy brain tissue

Various methods exist for detailed whole-brain segmentation with excellent performance on reasonably healthy brains. In detailed whole-brain segmentation, numerous neuroanatomical structures are segmented within the brain – where several structures have an identical intensity distribution and can only be sepa-

rated by their specific location within the brain. For example, hippocampi have the same intensity distribution as several other gray matter structures. Generally, these methods include segmentation of both hippocampi and brainstem. However, they are typically not designed to handle significant abnormalities in the brain. Many of the best-performing methods are non-parametric, in the form of multi-atlas methods, such as [Sabuncu et al., 2010, Wang et al., 2013, Coupé et al., 2011, Asman and Landman, 2013]. Many parametric whole-brain segmentation methods are based on generative atlas-based models, such as [Fischl, 2012, Puonti et al., 2016, Pohl et al., 2006], where a deformable atlas is used to encode the detailed spatial organization of brain structures.

As described in the previous section, generative atlas-based methods have been used for the single purpose of brain tumor segmentation, while still also incorporating some information about healthy tissue. For the purpose of healthy tissue segmentation, some studies have investigated how to take tumors into account when deforming atlases [Bauer et al., 2013, Cuadra et al., 2004, Conson et al., 2014]. However, this has not yet been specifically explored for typical organs-at-risk. An atlas-based method for segmenting organs-at-risk – but not tumors – has been evaluated on a limited data set of brain tumor patients [Isambert et al., 2008], with good performance on brainstem and eyes but lower performance on optic nerve and chiasm. A few discriminative methods specifically focused on brainstem segmentation have been evaluated on brains with tumors, showing good performance on a limited data set, by using support vector machines [Dolz et al., 2016b] and denoising autoencoders [Dolz et al., 2016a].

Segmentation of the optic system has generally received little attention because optic nerve and eye balls are outside the brain, but some methods have been developed for use in radiation therapy planning. Some focus solely on optic nerve segmentation by using e.g., support vector machines (SVMs) [Dolz et al., 2015a] or non-parametric multi-atlas approaches [Panda et al., 2014]. Geometric-driven methods have been successful in capturing the specific nature of the optic system, such as in [Noble and Dawant, 2011], and also including eyeballs in [Deeley et al., 2011, Bekes et al., 2008]. In these methods, the geometrical structure is explicitly modeled, i.e., that the optic nerves have a tubular structure with a crossing at the chiasm. CNN-based methods have also been developed for segmentation of the optic system, such as [Ibragimov and Xing, 2016]. The head and neck auto-segmentation challenge, held in conjunction with the 2015 MICCAI conference, examined the segmentation on CT images of organs-at-risk in radiation therapy of head and neck tumors. The challenge test dataset included 15 subjects with manual segmentations. Several included organs-at-risk are not relevant for brain tumors. However, this challenge included segmentation of brainstem, optic nerves and optic chiasm. Among other methods, a multi-atlas approach [Chen and Dawant, 2015] and a patch-based approach [Orbes Arteaga et al., 2015] were tried.

Segmentation of brain tumors and organs-at-risk have been studied separately. However, to the best of our knowledge, simultaneous brain tumor and organs-at-risk segmentation has not been fully explored before in the literature. Therefore, the aim of this project was to fill this gap in the literature.

Chapter 4

Modeling brain tumors with generative neural networks

In this chapter, we focus on the first part of the PhD project, which deals exclusively with the automatic segmentation of brain tumors. The chapter mainly describes the research related to paper A, and is divided in the following sections:

- The first section introduces and describes the generative neural network we use to model the shape of brain tumors: convolutional restricted Boltzmann machines. We will describe the model, the estimation of its parameters of the model and how it can be used to generate tumor shapes.

- The second section describes how the tumor shape model is used as a prior in a novel generative method for tumor segmentation in medical images, specifically multi-sequence MR images.

- In the third chapter, we discuss the method's performance on segmenting brain tumors compared to the state of the art in the field.

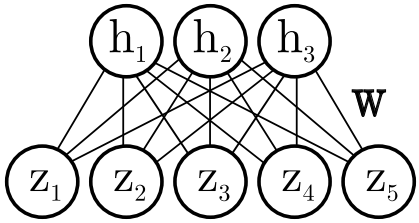


Figure 4.1: 1D example of an RBM. Each of the visible units in **z** are connected to all hidden units in **h** through weighted connections, with weights in **W**.

4.1 Modeling brain tumors with convolutional restricted Boltzmann machines

We will start by introducing the standard restricted Boltzmann machine (RBM) model. The model will be presented for a 1D case to avoid cluttered equations, but it directly generalizes to 3D images.

An RBM is a type of generative neural network that consist of stochastic units in two layers: a visible layer and a hidden layer [Smolensky, 1986, Fischer and Igel, 2014, Hinton, 2012]. The two layers are connected through weighted connections. Figure 4.1 shows a small 1D example of the network structure, with the visible layer denoted by **z** and the hidden layer denoted by **h**. In our case, we will use binary visible and hidden units.

As with other neural networks, the weight parameters in a restricted Boltzmann machine can be learned automatically by subjecting the network to training data. The weights act as filters and the hidden units as feature detectors capable of detecting typical features in the data. Therefore, there is no need for user-engineered filters (such as e.g. maximum response filters, seen on the right in figure 4.2), which have been and is still commonplace to use when developing segmentation algorithms. On the left in figure 4.2, RBM filters automatically learned on MNIST (a binary data set of hand-written digits) are shown as an example.

As the RBM model is a generative model, it learns the distribution of the training data. Once the parameters of an RBM have been learned, the model can thus be used to generate new data from the same distribution. RBMs have been shown to effectively learn the distribution of the binary MNIST data, and are able to generate new numbers with good quality [Hinton, 2002]. In this PhD project, we use RBMs to learn a generative model of binary tumor maps.

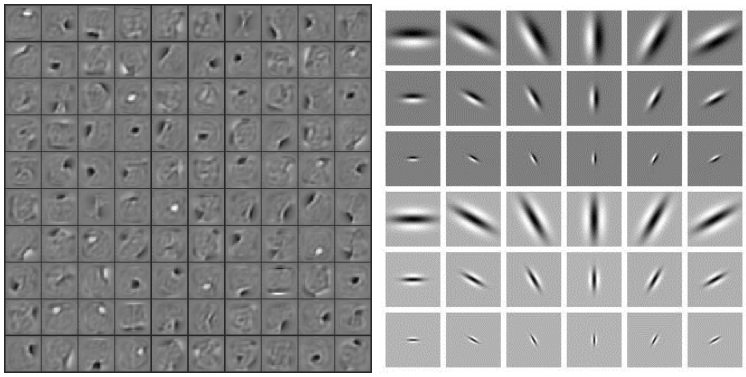


Figure 4.2: Filters. On the right: automatically learned RBM filters on MNIST. On the left: user-engineered maximum response filters.

From another perspective, the RBM model can also be interpreted as a higher-order Markov random field (MRF) [Fischer and Igel, 2014]. The model can be contrasted to a traditionally used first-order MRFs, in which only pairwise interactions between neighboring voxels are modeled. First-order MRFs are often used as priors in generative models to encode the prior knowledge that neighboring image voxels often belong to the same label, i.e., that there is a spatial consistency in an image. In contrast, a higher-order MRF is capable of taking larger neighborhood structures into account which leads to a more expressive model.

In the RBM model the probability of a set of visible units is defined as

$$p(\mathbf{z}) = \sum_{\mathbf{h}} p(\mathbf{z}, \mathbf{h}),$$

where the binary visible units are collected in $\mathbf{z} = (z_1, ..., z_I)$ and the binary hidden units are collected in $\mathbf{h} = (h_1, ..., h_J)$. The joint probability over both the visible units $\mathbf{z}$ and the hidden units $\mathbf{z}$ is

$$p(\mathbf{z}, \mathbf{h}) = \frac{1}{\mathcal{Z}} \mathrm{e}^{-E_{\mathrm{RBM}}(\mathbf{z}, \mathbf{h})}$$

with the energy

$$E_{\mathrm{RBM}}(\mathbf{z}, \mathbf{h}) = -\mathbf{z}^T \mathbf{W} \mathbf{h} - \mathbf{b}^T \mathbf{h} - \mathbf{z}^T \mathbf{a}, \tag{4.1}$$

where $\mathbf{W}$ is the weight matrix that encodes the weighted connections, $\mathbf{b}$ is the hidden bias vector and $\mathbf{a}$ is the visible bias vector - jointly denoted by $\boldsymbol{\lambda}$. The purpose of the bias vectors is to encourage units to be enabled or disabled. The

normalizing constant $\mathcal{Z}$ is given by

$$\mathcal{Z} = \sum_{\mathbf{z}} \sum_{\mathbf{h}} \mathrm{e}^{-E_{\mathrm{RBM}}(\mathbf{z},\mathbf{h})}.$$

Because there are no connections between units in the same layer in an RBM, it is straightforward to sample from the model. The conditional distributions can be derived as

$$p(h_j = 1|\mathbf{z}) = \sigma(\sum_i w_{ji} z_i + b_j) \tag{4.2}$$

and

$$p(z_i = 1|\mathbf{h}) = \sigma(\sum_j w_{ji} h_j + a_i), \tag{4.3}$$

where the sigmoid function $\sigma(t) = 1/(1 + e^{-t})$. The sampling can therefore be performed with block-Gibbs sampling, which is a Markov chain Monte Carlo (MCMC) method. Here, all hidden units can be sampled at the same time from equation 4.2 given the visible units, and all visible units can be sampled at the same time from equation 4.3 given the hidden units.

RBM training. Appropriate values for the RBM parameters $\boldsymbol{\lambda}$ are learned through unsupervised learning on training data (in our case binary tumor maps). The learning is based on maximum-likelihood estimation, which finds the parameters that maximize the likelihood of the model given the training data. As these parameters cannot be found analytically for the RBM model, gradient ascent on the log-likelihood is typically used. In gradient ascent, the parameters are iteratively updated based on the gradient of the log-likelihood. For an RBM, this gradient can be written as a sum of two terms. The first term is the expectation of the energy gradient under the conditional distribution of the hidden units given training examples of $\mathbf{z}$ and the second term is the expectation of the energy gradient under the model distribution [Fischer and Igel, 2014]. This can be expressed as

$$\nabla \boldsymbol{\lambda} = \frac{\partial \langle \log p(\mathbf{z}|\boldsymbol{\lambda}) \rangle_d}{\partial \boldsymbol{\lambda}} = -\Big\langle \frac{\partial E_{\mathrm{RBM}}(\mathbf{z},\mathbf{h})}{\partial \boldsymbol{\lambda}} \Big\rangle_d + \Big\langle \frac{\partial E_{\mathrm{RBM}}(\mathbf{z},\mathbf{h})}{\partial \boldsymbol{\lambda}} \Big\rangle_m,$$

where $\langle \cdot \rangle_d$ denotes the expectation under $p(\mathbf{h}|\mathbf{z}) p_e(\mathbf{z})$ (with p_e denoting the empirical data distribution) and $\langle \cdot \rangle_m$ denotes the expectation under the model distribution $p(\mathbf{z}, \mathbf{h})$ [Melchior et al., 2013]. The gradient in respect to each type of parameter then becomes

$$\begin{aligned} \nabla \mathbf{W} &= \langle \mathbf{z}\mathbf{h}^T \rangle_d - \langle \mathbf{z}\mathbf{h}^T \rangle_m, \\ \nabla \mathbf{a} &= \langle \mathbf{z} \rangle_d - \langle \mathbf{z} \rangle_m, \\ \nabla \mathbf{b} &= \langle \mathbf{h} \rangle_d - \langle \mathbf{h} \rangle_m. \end{aligned}$$

In the gradient ascent algorithm, the parameters are updated in each iteration by taking a step in the gradient direction, i.e.,

$$\boldsymbol{\lambda} = \boldsymbol{\lambda} + \omega \nabla \boldsymbol{\lambda},$$

where ω is a user-specified learning rate. The first expectation in the gradient is easily obtained directly from the observed training examples. The main issue is to compute the expectation under the model distribution $p(\mathbf{z}, \mathbf{h})$, which is intractable to compute exactly for any RBM of a usable size. We can approximate this term by sampling from the model by Gibbs sampling, using the conditional distributions of both $\mathbf{h}$ and $\mathbf{z}$ in equations 4.2 and 4.3 respectively. However, it takes many samples to obtain an unbiased estimate, which would make the training time prohibitively long. In a standard setting, the so-called contrastive divergence (CD) approximation is used instead. This approximation is obtained by initializing the Gibbs sampler with training examples and running it for just a few steps. Often, only one step is used, in which case the approximation is called CD-1 [Hinton, 2002]. In practice – due to computational time considerations – stochastic descent is typically used, where the gradient is approximated over a subset of the training examples in each iteration.

There are difficulties associated with training an RBM in practice, in part because of the approximate nature of the gradient computations. The training is sensitive to the selected training parameters, such as the learning rate. On small RBMs, where the log-likelihood can be evaluated fairly easily, it has been shown that the parameters may start to diverge from an optimal solution [Fischer and Igel, 2014]. Besides this, the larger size of most used RBMs make the evaluation of the log-likelihood impractical during training. Furthermore, the training is not invariant to the representation of the data, e.g., an RBM trained on the MNIST data set will have a better performance than an RBM trained on the same data set, but where every binary unit has been flipped to the opposite value (1-MNIST) [Melchior et al., 2013]. This is not due to the model's capacity but because of the invariance properties of the gradient. In fact, an RBM trained on MNIST can be transformed to model 1-MNIST while retaining the same performance [Melchior et al., 2013]. An approach to solve this invariance problem is the so-called enhanced gradient, which has been shown to be invariant of the data representation [Cho et al., 2013]. This gradient has been derived as a weighted sum of the gradients obtained for any combination of flips on the binary units, and is given by

$$\begin{aligned}
\nabla_e \mathbf{W} &= \Big\langle \big(\mathbf{z} - \langle \mathbf{z} \rangle_d\big) \, \big(\mathbf{h} - \langle \mathbf{h} \rangle_d\big)^T \Big\rangle_d - \Big\langle \big(\mathbf{z} - \langle \mathbf{z} \rangle_m\big) \, \big(\mathbf{h} - \langle \mathbf{h} \rangle_m\big)^T \Big\rangle_m, \\
\nabla_e \mathbf{a} &= \langle \mathbf{z} \rangle_d - \langle \mathbf{z} \rangle_m - \nabla_e \mathbf{W} \frac{1}{2} \Big(\langle \mathbf{h} \rangle_d + \langle \mathbf{h} \rangle_m \Big), \\
\nabla_e \mathbf{b} &= \langle \mathbf{h} \rangle_d - \langle \mathbf{h} \rangle_m - \nabla_e \mathbf{W}^T \frac{1}{2} \Big(\langle \mathbf{x} \rangle_d + \langle \mathbf{x} \rangle_m \Big).
\end{aligned}$$

There are also other training strategies apart from the standard CD-1, e.g., persistent contrastive divergence (PCD) and parallel tempering (PT) [Fischer and Igel, 2014, Melchior et al., 2013], which have been shown to yield superior performance for MNIST. Despite this, in early experiments comparing standard CD-1 with PCD on our data set – with varying learning rates – we found that CD-1 performed better for our purpose. However, an exhaustive comparison using the enhanced gradient was not performed, which would probably be appropriate for future work.

Convolutional RBM. The standard RBM works well on small images with a pre-defined size, such as the MNIST data set with a set image size of 28×28 pixels. However, medical images are typically far larger, three-dimensional and without a set size. Therefore, we instead use a convolutional variant of the RBM model (cRBM) [Lee et al., 2011] which can cope with this type of images. In this variant, the connection weights are shared among all locations in the form of convolutional filters, which only have a fraction of the size of a full image. Figure 4.3 shows a small 1D example of the cRBM model. The energy in the cRBM is defined as

$$E_{\text{cRBM}}(\mathbf{z}, \mathbf{H}) = -\sum_k \mathbf{h}_k \bullet (\mathbf{w}_k * \mathbf{z}) - \sum_k b_k \sum_j h_{kj} - a \sum_i z_i,$$

where $\bullet$ denotes element-wise product followed by summation and $*$ denotes convolution. Each hidden group $\mathbf{h}_k \in \mathbf{H}$ is connected to the visible units in $\mathbf{z}$ with a convolutional filter $\mathbf{w}_k$. There are two main differences from the standard energy (equation 4.1): (1) the hidden units detect features locally rather than over the whole image and (2) the visible bias term is a scalar instead of a vector and acts over the whole image rather than for each image unit. The conditional distributions of the hidden and visible units can be derived as

$$p(h_{kj} = 1|\mathbf{z}) = \sigma\Big((\mathbf{w}_k * \mathbf{z})_j + b_k\Big) \tag{4.4}$$

and

$$p(z_i = 1|\mathbf{H}) = \sigma\Big(\sum_k (\tilde{\mathbf{w}}_k * \mathbf{h}_k)_i + a\Big),$$

where $\tilde{\mathbf{w}}$ denotes a mirror-reversed version of the filter $\mathbf{w}$. The parameters of the cRBM model can be learned in exactly the same way as for the standard RBM, due to their similar structure.

Tumor shape model. The visible tumor in MRI consist of two parts: the tumor core, which is the target for radiation therapy, and the peritumoral edema surrounding the tumor core. To model both structures, we construct a model

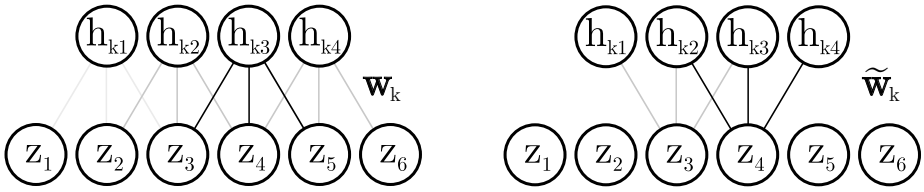


Figure 4.3: 1D example of a cRBM. Visible units in $\mathbf{z}$ are connected to hidden units in a hidden group $\mathbf{h}_k$ through a convolutional filter $\mathbf{w}_k$ of size 3. The first illustration shows the model from the hidden layer's perspective. The second shows the model from the visible layer's perspective, where $\tilde{\mathbf{w}}_k$ is a mirror-reversed version of the filter. Note that boundary units in the visible layer are set to 0.

combining two cRBMs. The first cRBM models the "whole" tumor region, which include both the tumor core and the edema. We denote this region by $\mathbf{z}$. The second cRBM models only the tumor core, which we denote by $\mathbf{y}$. The two cRBM models are trained separately and then combined to form the full model, which we define as

$$p(\mathbf{z}, \mathbf{y}) = \sum_{\mathbf{H},\mathbf{G}} p(\mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}) \tag{4.5}$$

with

$$p(\mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}) \propto e^{-E_{\text{comb}}(\mathbf{z},\mathbf{y},\mathbf{H},\mathbf{G})}$$

and the combined energy term

$$E_{\text{comb}}(\mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}) = E_z(\mathbf{z}, \mathbf{H}) + E_y(\mathbf{y}, \mathbf{G}) + \sum_i f(z_i, y_i), \tag{4.6}$$

where $\mathbf{G}$ denotes the hidden units connected to the tumor core map $\mathbf{y}$. The function $f(z, y)$ restricts tumor core to only be allowed within the whole tumor region and is constructed as follows: $f(z, y) = \infty$ if $z = 1$ and $y = 0$, and otherwise $f(z, y) = 0$. Another way to model both labels would be to train them together as a multinomial distribution, as e.g., in [Salakhutdinov et al., 2007]. However, the construction of a multinomial RBM increases the complexity of the modeled distribution and would force us to learn filters also for the background class, which increases the amount of parameters and the computational time in the segmentation algorithm.

To train the tumor shape model. When constructing a cRBM model, there are two key aspects to consider: the number of filters k and the size of the filters. In choosing the number of filters, we have to consider the complexity of the distribution but we are also limited by the fact that more filters increase

the computational time. The size of the filters will not affect the computational time significantly, but this aspect is perhaps more troublesome. A large filter size could potentially capture long-range features better than a small filter size, but in a 3D setting the amount of parameters to tune per filter increases fast with an increasing filter size. In our experiments, we found that a filter size larger than $7 \times 7 \times 7$ voxels decreased the performance of the model. This might be due to overfitting to the training data, which is common when many parameters are estimated on a small training set. It could also be that the training algorithm did not converge to the maximum-likelihood estimate, which is common for RBMs and is difficult to measure. To be able to better capture long-range features, we therefore composed an approach where we tie the filter weights in $(2 \times 2 \times 2)$-blocks, effectively treating each block as one parameter. Figure 4.4 illustrates this approach in 1D for one single hidden unit. For the purpose of tumor segmentation, this approach with a filter size of $14 \times 14 \times 14$ (with $7 \times 7 \times 7$ parameters) gave the best performance in initial experiments.

Figure 4.5 shows filters trained for tumor core in three different settings. In all settings, the same learning rate and training data set was used. The training set consisted of 30 binary three-dimensional maps augmented to 240 by flipping the maps in 8 different direction. In the first row, filters learned with standard CD-1 are shown. In the second row, filters learned with CD-1 and enhanced gradient are shown. In the third row, filters learned with CD-1 and enhanced gradient in a multinomial cRBM are shown. Although it might not always be clear from the filters how well the model performs, the filters do show the kind of features they capture in the data. The filters trained with standard CD-1 do not exhibit the same variety of features as the filters trained with enhanced gradient and CD-1. Furthermore, a significant amount of CD-1 filters do not seem to encode any specific feature. The filters trained with enhanced gradient are all distinct. The filters for the multinomial cRBM illustrates the complexity of this distribution. CD-1 learning with enhanced gradient on binary units was found to give a superior performance, and is thus the approach we use in this thesis.

Generative performance of the cRBM model. To illustrate the generative performance of the cRBM model, we show a small inpainting experiment in figure 4.6. In this experiment, a cRBM model trained to model whole tumor is used to fill in missing data in two whole tumor maps. A three dimensional checkerboard patterned mask with 3 cm^3 large cubes is used to set some of the voxels to zero in the original maps. The other voxels are kept fixed to their original values throughout the experiment. Gibbs sampling is then performed with 600 sampling steps to test the capability of the cRBM model to generate new data where data is missing. As seen in figure 4.6, the model is capable of

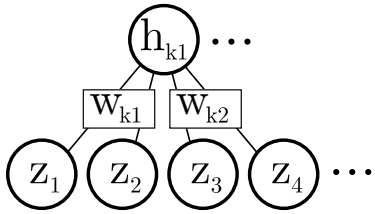


Figure 4.4: Illustration of the sharing of weights in a 1D example. Adjacent filter weights are tied to the same parameter, so that a filter of length 4 has only 2 parameters.

generating new data that captures the general shape of whole tumor well. The generated data is smoother than in the original maps, which is not surprising considering that only 40 filters are used to model the highly variable shape of tumors. Still, we can conclude that the model can connect isolated tumor regions fairly far apart, which is seen clearly in the first case, and fill in large holes, which is seen clearly in the second case.

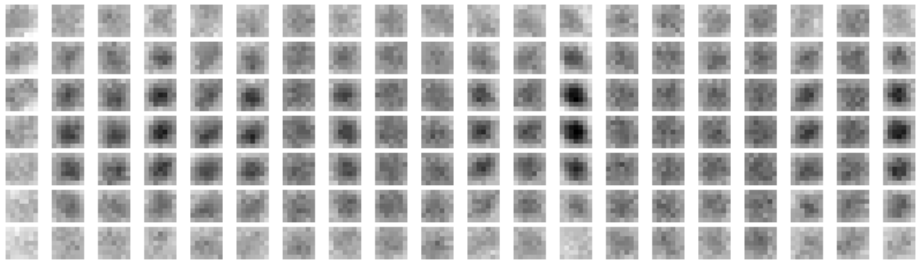
4.2 Segmenting brain tumors with a generative method

In order to use the tumor shape model to segment brain tumors in medical images, we need to assemble a framework to inform the model about the structure of the surrounding brain and the properties of medical images. To accomplish this, we incorporate the tumor shape model as a prior in a generative healthy tissue segmentation model. This type of generative model has been used extensively for segmenting MR images of the brain, e.g., in [Ashburner and Friston, 1997, Van Leemput et al., 1999b, Van Leemput et al., 1999a, Van Leemput et al., 2001, Menze et al., 2010]

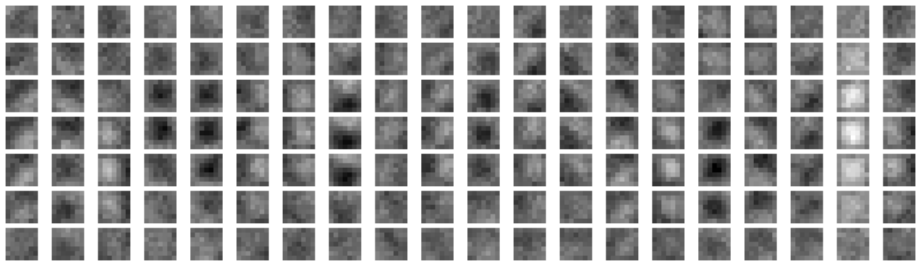
4.2.1 Basic generative model for healthy tissue segmentation

The basic model consist of two parts: a *likelihood function* modeling image intensities given segmentation labels and a *prior probability* modeling the spatial organization of the labeled brain structures. Let $\mathbf{D} = (\mathbf{d}_1, ..., \mathbf{d}_I)$ denote the multi-sequence MR image of a subject, where I is the number of image voxels

Standard CD-1, 9600 gradient steps, tumor core



CD-1 with enhanced gradient, 9600 gradient steps, tumor core



Multinomial cRBM, CD-1 with enhanced gradient, 11280 gradient steps, tumor core

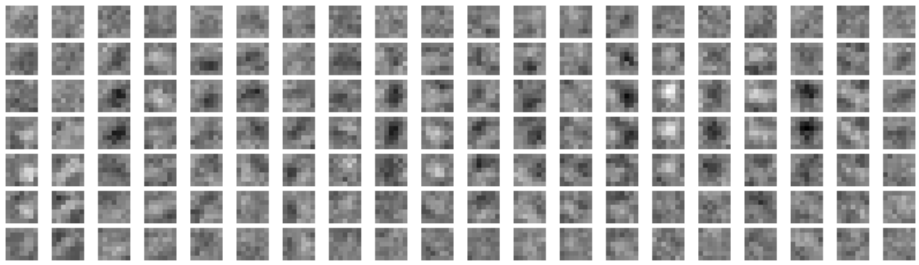


Figure 4.5: Comparison of filters relating to tumor core. Each row shows 20 filters out of 40 filters. *First row* shows filters learned with standard CD-1. *Second row* shows filters learned with CD-1 with enhanced gradient. *Third row* shows filters learned for a multinomial cRBM with a total of 3 labels: tumor core (shown filters), edema and background.

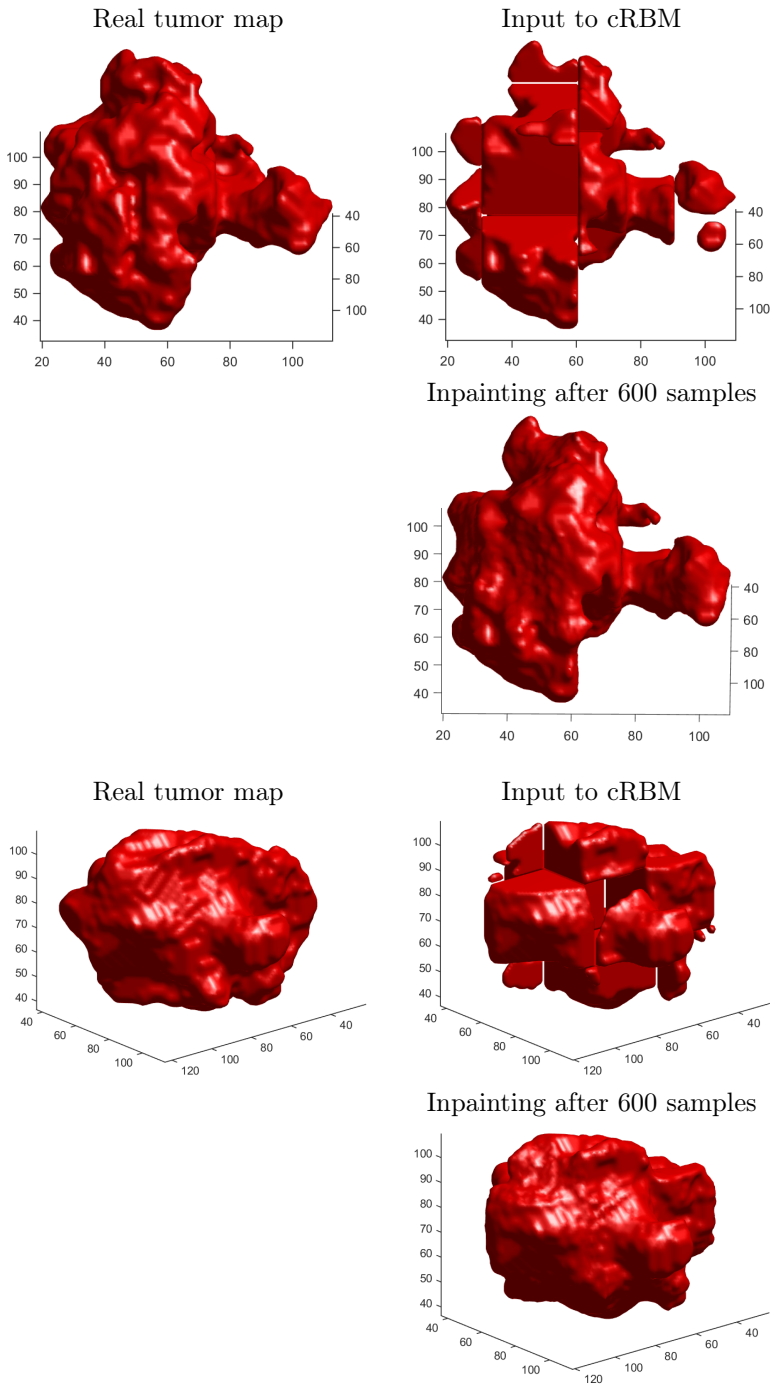


Figure 4.6: Inpainting of whole tumor by sampling from cRBM model.

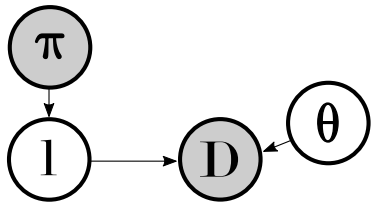


Figure 4.7: Graphical representation of basic model for healthy brain segmentation. The atlas-based prior $\boldsymbol{\pi}$ models healthy tissue labels in $\mathbf{l}$. The labels in $\mathbf{l}$ predict the multi-sequence data $\mathbf{D}$ according to the model parameters $\boldsymbol{\theta}$. Shading indicates observed variables.

and $\mathbf{d}_i$ contains the image intensities at voxel i. Furthermore, let $\mathbf{l} = (l_1, ..., l_I)$ denote the healthy labels, with $l_i \in \{1, .., K\}$ for K labels. The likelihood is defined as

$$p(\mathbf{D}|\mathbf{l}) = \int_{\boldsymbol{\theta}} p(\mathbf{D}|\mathbf{l}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) d\boldsymbol{\theta}, \tag{4.7}$$

where $\boldsymbol{\theta}$ contains free model parameters. The healthy tissue prior $p(\mathbf{l})$ is in the form of a probabilistic atlas that contains a probability vector for each voxel encoding the probability of each label to occur, defined as $p(\mathbf{l}) = \prod_i \pi^i_{l_i}$. The model is illustrated in figure 4.7. A segmentation of a brain image can then be obtained by evaluating the posterior

$$p(\mathbf{l}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{l}) p(\mathbf{l}). \tag{4.8}$$

Likelihood. The likelihood function models the image intensities by Gaussian mixture models as illustrated in figure 4.8, one connected to each label. By log-transforming the MR intensities, the bias field artifacts in MR images can be modeled as a linear combination of spatially smooth basis functions added to the log-transformed intensities. Let $\boldsymbol{\theta}_l$ denote the Gaussian parameters connected to label l and $\mathbf{C}$ denote the parameters of the bias field model. A Gaussian mixture model can then be defined at voxel i as

$$p_i(\mathbf{d}_i|\boldsymbol{\theta}_l, \mathbf{C}) = \sum_{g=1}^{G_l} \gamma_{lg} \mathcal{N}(\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i | \boldsymbol{\mu}_{lg}, \boldsymbol{\Sigma}_{lg}), \tag{4.9}$$

where G_l is the number of Gaussian components connected to label l; and γ_{lg}, $\boldsymbol{\mu}_{lg}$ and $\boldsymbol{\Sigma}_{lg}$ are the weight, mean and covariance matrix of component g. Furthermore, the vector $\boldsymbol{\phi}_i \in \mathbb{R}^P$ evaluates the P basis functions at voxel i and $\mathbf{C} = (\mathbf{c}_1, ..., \mathbf{c}_N)^T$, where $\mathbf{c}_n \in \mathbb{R}^P$ denotes the parameters of the bias field model

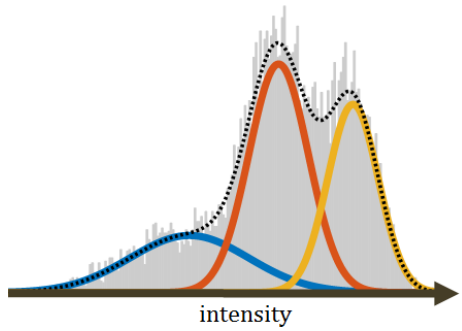


Figure 4.8: A Gaussian mixture model with three Gaussian components models the intensity distribution in an image.

for MR sequence n. All voxel probabilities are collected to form the likelihood

$$p(\mathbf{D}|\mathbf{l}, \boldsymbol{\theta}) = \prod_i p_i(\mathbf{d}_i|\boldsymbol{\theta}_{l_i}, \mathbf{C}), \tag{4.10}$$

where $\boldsymbol{\theta}$ contains all parameters connected to the Gaussian mixture models and the bias field model. Finally, a prior $p(\boldsymbol{\theta})$ is added over these intensity parameters. For healthy tissue segmentation, it is often enough to just use uninformative priors.

Modeling of bias field artifacts in MR images. Figure 4.9 shows white matter segmentations using this basic model on different sagittal slices of one subject. The figure illustrates the effect of including the bias field correction in the model. As can be seen, bias field modeling is crucial to obtain accurate segmentations. In many segmentation methods – as almost all discriminative brain tumor segmentation methods and some generative methods such as [Kwon et al., 2014] – bias field correction is regarded as a pre-processing step. Here, we explicitly include the bias field correction by modeling the bias field as spatially smooth basis functions added to the (log-transformed) MR images, as in [Van Leemput et al., 1999a]. Other approaches exist, as in [Ashburner and Friston, 2005] where the same basis function model is used directly as a multiplicative effect. Several types of basis functions, denoted by $\boldsymbol{\phi}$ in equation 4.9, can be appropriate to model bias fields. Throughout this thesis, we use cosine basis functions. Specifically, if we have P number of basis functions per dimension of an image, we use the P lowest frequencies of the discrete cosine transform (DCT). Other works have used e.g, polynomial basis functions [Van Leemput et al., 1999a] or B-splines [Larsen et al., 2014].

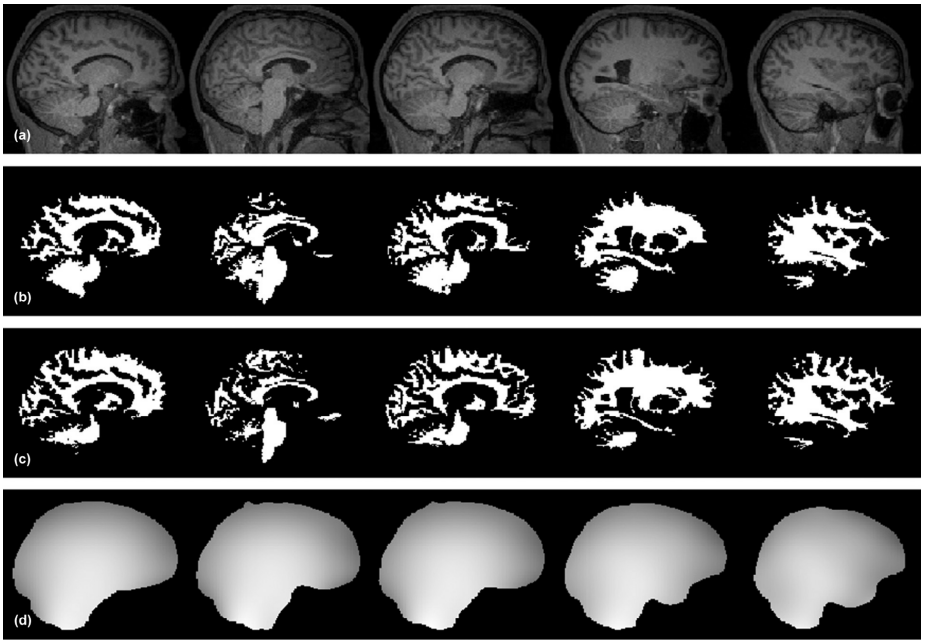


Figure 4.9: Effect on segmentation performance from bias field correction. The white matter in T1-weighted images (a) is segmented without bias field correction (b) and with bias field correction (c). The estimated bias field is shown in (d). The image is taken from [Van Leemput et al., 1999a].

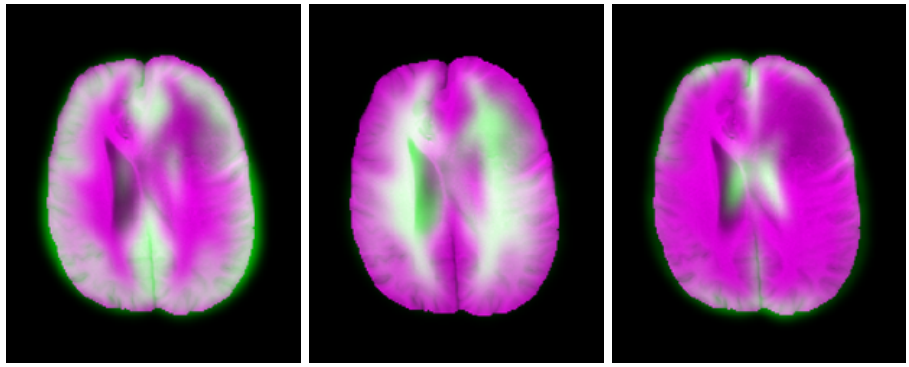


Figure 4.10: Tissue probabilities in the affine atlas $\boldsymbol{\pi}$ overlayed on a T1-weighted image after affine registration. From left to right: gray matter, white matter and CSF.

Prior on healthy labels. The probabilistic atlas in $p(\mathbf{l})$ is assumed to be known and remains constant during the segmentation process, after an initial affine co-registration with the subject to be segmented. The atlas used in this thesis is shown in figure 4.10 overlayed on a T1 image. It has been computed from manually segmented subjects that were affinely registered to each other and include smooth probability maps of gray matter, white matter and CSF. Further details on this atlas can be found in [Ashburner and Friston, 1997].

4.2.2 Incorporating the tumor shape model into the basic generative model

To extend the basic generative model for healthy brain segmentation to also handle tumor segmentation, we introduce into the model the binary map $\mathbf{z}$ to indicate voxels that are affected by tumor and the binary map $\mathbf{y}$ to indicate voxels that are inside the tumor core. A voxel i with label l_i will change its status to tumor when $z_i = 1$. If $z_i = 1$ and $y_i = 0$ the voxel will be part of edema and if $z_i = 1$ and $y_i = 1$ the voxel will be part of tumor core. Furthermore, we introduce two new Gaussian mixture models: one to model the data under the edema label and one to model the data under the tumor core label. We can then include the tumor shape model in equation 4.5 as a prior in the segmentation model so that the posterior becomes

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y})p(\mathbf{l})p(\mathbf{z}, \mathbf{y}). \tag{4.11}$$

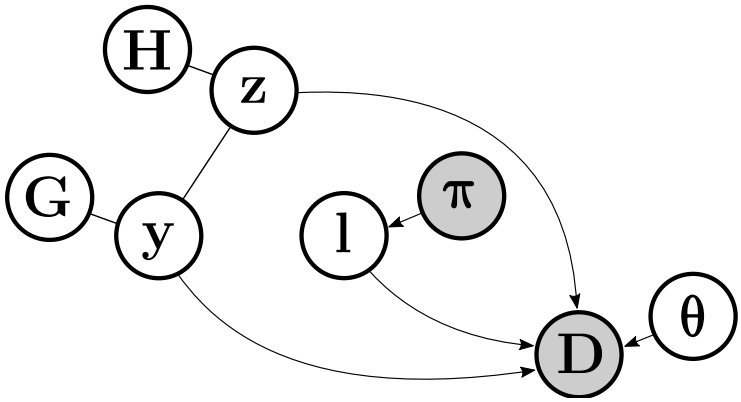


Figure 4.11: Graphical representation of model focused on brain tumor segmentation. The atlas-based prior $\boldsymbol{\pi}$ models healthy tissue labels in $\mathbf{l}$. The complete tumor label $\mathbf{z}$ and core label $\mathbf{y}$ are connected to the hidden units of their cRBM models $\mathbf{H}$ and $\mathbf{G}$, respectively. The labels $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ jointly predict the multi-sequence data $\mathbf{D}$ according to the model parameters $\boldsymbol{\theta}$. Shading indicates observed variables.

To make clear how the two introduced Gaussian mixture models relate to the tumor prior, the likelihood in equation 4.10 will change to

$$p(\mathbf{D}|\mathbf{l},\mathbf{z},\mathbf{y},\boldsymbol{\theta}) = \prod_i \begin{cases} p_i(\mathbf{d}_i|\boldsymbol{\theta}_{l_i},\mathbf{C}) & \text{if } z_i = 0 \text{ and } y_i = 0, \text{ (healthy)} \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_e,\mathbf{C}) & \text{if } z_i = 1 \text{ and } y_i = 0, \text{ (edema)} \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_c,\mathbf{C}) & \text{if } z_i = 1 \text{ and } y_i = 1, \text{ (core)} \\ (y_i = 1 \text{ and } z_i = 0 \text{ prohibited by prior, eq. 4.5}) \end{cases},$$

where $\boldsymbol{\theta}_e$ contains the Gaussian parameters related to edema and $\boldsymbol{\theta}_c$ contains the Gaussian parameters related to tumor core. Figure 4.11 shows a graphical representation of the new extended model.

Robustness of the intensity model. A key difference between the atlas-based healthy tissue prior and the cRBM-based tumor shape prior is that the tumor shape prior does not include any prior information about location of tumors within the brain. Expected location of tumors could have been modeled by the visible bias term in the cRBM model, as in e.g., [Puonti and Van Leemput, 2016] for multiple sclerosis lesions. However, to model this accurately for tumors a lot more training examples would be needed than we had access to. Even if we would have access to a multitude of training examples, this aspect would probably not make a lot of sense to model for brain tumors, as they can appear

virtually anywhere within the brain. Nevertheless, this lack of location information has some impact on the robustness of the segmentation model. The healthy tissue segmentation model described in section 4.2.1 is usually used with uninformative priors on the Gaussian parameters, because the location information in the atlas is enough to estimate appropriate values. For tumor segmentation, the need arises to include some informative prior on these parameters.

As described in section 2.1, the intensity distribution of MR images obtained with the same MR sequence tends to vary significantly. Therefore, it is difficult to include prior information related to intensity parameters without loosing the adaptability that generative models are known for. However, we can assume e.g., that tumor tissue – especially edema – will be significantly brighter than any healthy tissue in a FLAIR image and that some parts of the tumor core in a high-grade glioma will be significantly brighter than healthy white matter and gray matter in a T1c image. To encode this prior knowledge, we use an informative prior on the mean vectors related to tumor. This prior introduces linear constraints on mean values, e.g.,

$$p(\boldsymbol{\mu}_e, \boldsymbol{\mu}_{WM}) = \begin{cases} \propto 1 & \text{if } \mu_{e,\text{FLAIR}} \geq \alpha\, \mu_{\text{WM,FLAIR}} \\ 0 & \text{otherwise} \end{cases}.$$

Here, the scalar α defines the limit of how close $\mu_{e,\text{FLAIR}}$ can be to $\mu_{\text{WM,FLAIR}}$. Appropriate values of these constraints can be estimated on training data which optimally include images from several imaging centers. We have found this type of prior to work well in making the segmentation method more robust. However, it does limit the model's flexibility in the sense that some reconsideration is required if e.g., FLAIR is not present for a subject.

A related issue is the number of Gaussian components for the Gaussian mixture models related to healthy structures. If one compares the number of components used for healthy structures in a healthy tissue segmentation method – e.g., [Puonti et al., 2016] – with the number of components in Paper A or C, one can see that we use less components to model each structure. This is also due to the lack of tumor location information. If we use a large number of healthy Gaussian components, we run into the risk that one healthy component incorrectly starts to model tumor tissue. However, the small number of healthy components can result in the opposite situation, that the Gaussian components related to tumor tissue starts to incorrectly model healthy tissue. Therefore, we regularize the values of the covariance matrices by introducing informative conjugate priors. In Paper A and B, we imposed priors specifically on the covariance matrices related to tumor, which we used in a nonstandard way. However, we later found that a more general proper prior on all covariance matrices achieves a similar performance. This prior was used in Paper C.

4.2.3 Inference

To obtain a segmentation of an image, we need to evaluate the posterior in equation 4.11. However, there are two complications with our model that hinder us from directly inferring the optimal segmentation: (1) the tumor shape model does not factorize over voxels and (2) the likelihood model contains a marginalization over the intensity parameters in $\boldsymbol{\theta}$, which results in an intractable integration. If we for the moment disregard the tumor shape model, we have the healthy tissue model defined in equation 4.8, where the second complication is still present. This type of model is commonly solved with the expectation-maximization (EM) algorithm [Dempster et al., 1977]. The EM algorithm focus on the posterior of the intensity parameters given the data, instead of the posterior of the labels in 4.8. Using Bayes' rule, the posterior of the intensity parameters is given by

$$\begin{aligned} p(\boldsymbol{\theta}|\mathbf{D}) &\propto p(\mathbf{D}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \\ &= \left(\sum_{\mathbf{l}} p(\mathbf{D}|\mathbf{l},\boldsymbol{\theta})p(\mathbf{l})\right) p(\boldsymbol{\theta}). \end{aligned}$$

The EM algorithm is designed to find the parameter values that maximize this function. It iterates between two steps: the E-step (expectation step) and M-step (maximization step). In the E-step, the algorithm builds a lower bound to the objective function which can be optimized analytically. This lower bound is a local approximation that touches the objective function at the current parameter estimates. In the M-step, the current parameter estimates are then updated to the values that maximize the lower bound, which are given in analytical form. The algorithm is guaranteed to increase the objective function at each iterative step, and will thus find at least a local maximum of the parameter posterior.

However, the bias field model we use complicates the M-step slightly. Analytical expressions of the update rules for the intensity parameters can be found only if the bias field parameters are kept fixed and vice versa. Alternating between updating the two parameter sets will eventually give the values that maximize the lower bound. In practice, however, it is enough to update the two parameter sets once, which will not maximize the lower bound but still guarantee an increase of the objective function at each iterative step. This is then called a generalized expectation-maximization (GEM) algorithm [Van Leemput et al., 1999a].

Once the maximum a-posteriori (MAP) estimates of the intensity parameters have been obtained, a segmentation can be obtained by using the empirical Bayes approximation, where the parameters are fixed to these estimates. The

approximate posterior of the labels then becomes

$$p(\mathbf{l}|\mathbf{D}) \approx p(\mathbf{D}|\mathbf{l}, \hat{\boldsymbol{\theta}})p(\mathbf{l}).$$

The maximum a-posteriori (MAP) estimates of the labels can then easily be obtained, as the label posterior now factorizes over the voxels. The MAP estimate at one voxel will simply be the label with the highest posterior probability at that voxel.

As the tumor shape model does not factorize over the voxels, it is not possible to directly use the GEM algorithm to evaluate the tumor segmentation posterior defined in equation 4.11. One approach to evaluate this posterior is to sample from the model using block-Gibbs sampling. In practice, we would sample from the joint distribution $p(\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{G}, \mathbf{H}, \boldsymbol{\theta}|\mathbf{D})$ by iteratively sampling from the conditional distribution of each set of variables given the other variables. All of these conditional distributions are easy to sample from, because each of them factorizes over its components. We then simply collect samples of $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$; and perform a voxel-wise majority voting across the collected samples to obtain the final segmentation. However, if the values used to initialize the sampling chain are far away from appropriate values and the state space is high-dimensional, a Gibbs sampler may need an impractically large number of sampling steps to reach a stationary distribution and may become trapped in one isolated region of the state space. Therefore, for an efficient Gibbs sampler for our model, it is important to have reasonable initial values.

We therefore developed an inference algorithm with two steps: in the first step we use GEM and in the second step we use Gibbs sampling. To be able to use GEM, we temporarily replace the energy term in the tumor shape model (defined in equation 4.6) with a simple factorizable energy term. This temporary energy term essentially defines uniform spatial prior probabilities for $\mathbf{z}$ and $\mathbf{y}$. The effect of the temporary energy term is that the tumor Gaussian distributions will model tissue that is not well explained by the healthy Gaussian distributions. We then use the MAP estimates of $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ to initialize a Gibbs sampler, where we reintroduce the proper energy term of the tumor shape model. Note that we did not sample from the conditional distribution of the bias field parameters in paper A, but instead kept them fixed to their MAP estimates obtained from GEM. However, later in the PhD project, we also implemented this.

A potential improvement of the inference could be to incorporate the tumor shape model into the GEM procedure. In [Van Leemput et al., 1999a], a first-order MRF – which also does not factorize over voxels – was used as a prior in the basic healthy tissue model defined in section 4.2.1. By using a mean field approximation, the resulting model could be evaluated with GEM. For the tumor shape model however, this approach both becomes computationally

demanding and seems to result in a poor approximation of the complicated true distribution, judging from experiments we carried out. However, this could be studied further.

4.3 Results and discussion

In paper A, we describe the participation of the above method in the brain tumor segmentation (BRATS) challenge at the BrainLes workshop, which was held in conjunction with the 2015 MICCAI conference. In that challenge, the segmentation performance of 12 brain tumor segmentation methods was compared on a test data set of 54 subjects – including both high- and low-grade gliomas. The data set included the MR-sequences FLAIR, T1c, T2 and T1. All images were skull-stripped by the challenge organizers, re-sampled to 1 mm isotropic resolution and registered to a common space. Each participating team had two days to segment the images and subsequently uploaded the resulting segmentations to an online platform [Kistler et al., 2013] for comparison to manual ground truth segmentations. In the challenge, our method was one of the top-performing methods. Figure 4.12 shows box plots of the reported scores for our method (method 1) and three other top-performing methods. Method 2 is based on a semi-automated generative method which incorporates a tumor growth model and a discriminative post-processing step designed to improve performance [Bakas et al., 2016]. Method 3 and 4 are discriminative CNN-based methods [Pereira et al., 2016, Havaei et al., 2016]. Scores were reported for three regions: enhanced core, tumor core (which includes enhanced core) and whole tumor (which includes tumor core and edema). The tumor core region corresponds to $\mathbf{y}$ in our model and the whole tumor region corresponds to $\mathbf{z}$ in our model. As the region of interest in radiation therapy is tumor core, we show scores for this region, but also for whole tumor to show the full performance of our tumor shape model. Two types of scores were reported: (1) Dice score which measure the overlap between a segmentation and the ground truth, where a score of 1 means perfect overlap and 0 means no overlap; (2) robust Hausdorff distance which is a measure of the distance between the segmentation borders.

As can be seen from the figure, our method performed particularly well on tumor core, judging from the Dice score. However, the average Hausdorff distance is somewhat lower than at least two of the other methods, which is probably explained by the post-processing used by the other methods. In general, none of these state-of-the-art methods can be categorized as robust, considering the wide range of scores for each method. The test data set has been designed to be particularly difficult, with many complicated high-grade tumors and low-grade

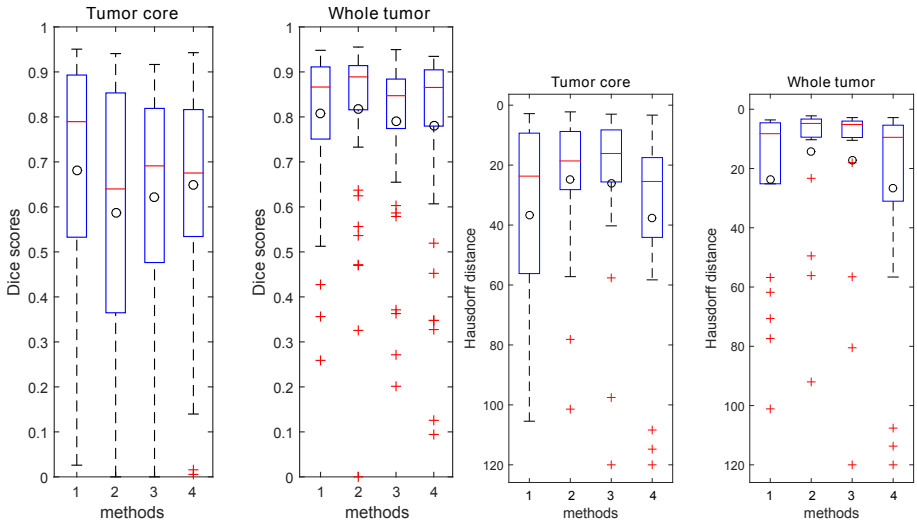


Figure 4.12: Comparison between top performing methods in BRATS 2015 on Dice scores and Hausdorff distances for tumor core and whole tumor. Method 1 is the method presented in this section, method 2 is a generative method with a discriminative post-processing step to enhance performance [Bakas et al., 2016] and method 3-4 are CNN-based methods [Pereira et al., 2016, Havaei et al., 2016]. Circles show mean values, red central lines show median values, red crosses show outliers and whiskers show full range of values excluding outliers.

tumors included and with a wide range of tumor sizes. Nevertheless, these results illustrate the general robustness of current state-of-the-art methods well and shows the high complexity in modeling brain tumors. Figure 4.13 shows a representative subject together with the automatic segmentation by our method. In the image slices, we can see the complicated appearance of particularly tumor core. The T2 image exhibit a ringing artifact, which is a fairly common MRI artifact. Furthermore, our method has incorrectly segmented some areas around the ventricles as edema, because of their bright appearance in FLAIR, which result in a large Hausdorff distance.

Figure 4.14 shows two subjects where our method segmented the tumor poorly. In both these subjects, some inner parts of the core exhibit a similar appearance to healthy white matter. Here, we can clearly see a limitation of the tumor shape model. The learnt features in the cRBMs are mainly focused on the shape of the tumor border, which results in a model that is strong at the borders but fairly weak when it comes to modeling the inner parts of a tumor. The model

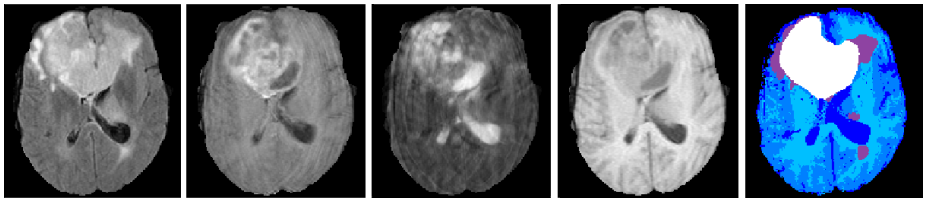


Figure 4.13: Example slice of tumor segmentation. From left to right: MR sequences FLAIR, T1c, T2, T1; and final segmentation. Shades of blue: healthy tissues, white: tumor core, lilac: edema.

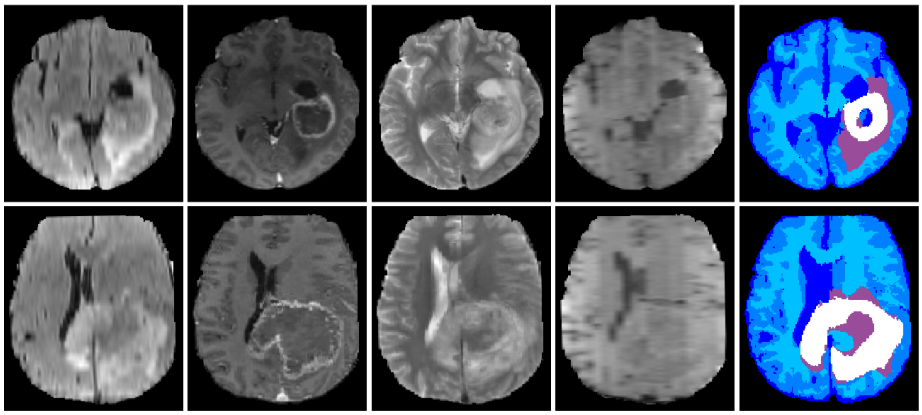


Figure 4.14: Examples where tumor prior fail to segment inner part of tumor. From left to right: MR sequences FLAIR, T1c, T2, T1; and final segmentation. Shades of blue: healthy tissues, white: tumor core, lilac: edema.

still works well for most tumors because the appearance of inner parts of a tumor typically deviates sufficiently from the appearance of healthy structures. However, in a few cases – such as the ones in the figure – this is not true, which results in a poor segmentation. A deeper architecture in the cRBM models with more than just one hidden layer could potentially result in a more global tumor shape model that could handle these cases better. A convolutional deep belief network (cDBN) is an example of a generative deep neural network that uses cRBMs as building blocks [Lee et al., 2011]. However, a cDBN is even harder to train than a cRBM and would increase the computational time considerably. More recent generative neural networks, such as variational autoencoders could potentially be considered [Kingma and Welling, 2013, Bengio et al., 2013].

Chapter 5

Combined segmentation of brain tumors and organs-at-risk

In this chapter, we focus on the second part of the PhD project, which deals with the simultaneous automatic segmentation of brain tumors and organs-at-risk (healthy structures that are important to spare in radiation therapy). The chapter describes the research related to papers B and C, and is divided into the following sections:

- The first section introduces and describes the probabilistic deformable atlas-based model that we use to model the healthy brain in more detail than the atlas used in the previous chapter. This allows us to segment many more structures than just white matter, gray matter and cerebrospinal fluid. This section also describes how we extend this brain atlas to handle non-brain organs-at-risk.
- The second section describes how we incorporate this deformable atlas into the modeling framework for brain tumor segmentation presented in the previous chapter.
- In the third chapter, we discuss the method's performance on segmenting organs-at-risk and brain tumors simultaneously.

5.1 Building a mesh-based atlas for modeling healthy brain structures

In the previous chapter, healthy brain tissues were modeled by a affinely-registered atlas that remained constant during the inference over the model. This type of atlas is easy to include in the modeling framework and seems to be sufficient to differentiate the appearance of tumor from that of healthy tissue. In this chapter, however, we turn our attention to the segmentation of specific anatomical structures in the brain – specifically to the detailed segmentation of several organs-at-risk. For this purpose, the previously used atlas lacks the necessary complexity. Although the general structure of healthy human brains is the same, individual shape differences are fairly large (which is the reason for the smoothness of the atlas in figure 4.10). In addition, large tumors often deform healthy structures, and brains with tumors frequently include other co-occurring pathologies. Therefore, we instead include a more complex framework that can model both the tissue probabilities at certain locations and the deformation of those locations.

In contrast to voxel-based approaches such as [Ashburner and Friston, 2005, Fischl et al., 2002], we use a mesh-based approach. This mesh-based atlas is in the form of a tetrahedral mesh, which consists of a number of vertices. The positions of the vertices are collected in $\boldsymbol{\eta}$. Each of these vertices is associated with a probability vector, where each element of the vector specifies the probability of a label to occur around that vertex. During the atlas building process, the number of vertices is locally adapted, which results in a mesh that is sparse in areas with less information and dense in areas with more information. This allows for a compact representation of the structure of the brain. An example of this type of atlas is shown in figure 5.1. The model was first introduced in [Van Leemput, 2009] and further described and validated for detailed healthy brain segmentation in [Puonti et al., 2016]. It has also been used successfully for detailed segmentation of hippocampus and brainstem subregions [Iglesias et al., 2015a, Iglesias et al., 2015c].

Fundamentally, we want to be able to predict the probability of seeing a segmentation given a number of manual training segmentations. In this thesis, we used the expert segmentations from a data set of 39 subjects scanned with a T1-weighted MR sequence – including 28 healthy subjects and 11 subjects with varying severity of Alzheimer's disease, with an age range from under 30 years to over 60 years old [Sabuncu et al., 2010]. The segmentations were performed by a validated semi-automated segmentation protocol [Caviness et al., 1989, Caviness et al., 1996, Kennedy et al., 1989]. The segmentations consist of 39 brain structures – including three important organs-at-risk: brainstem, hip-

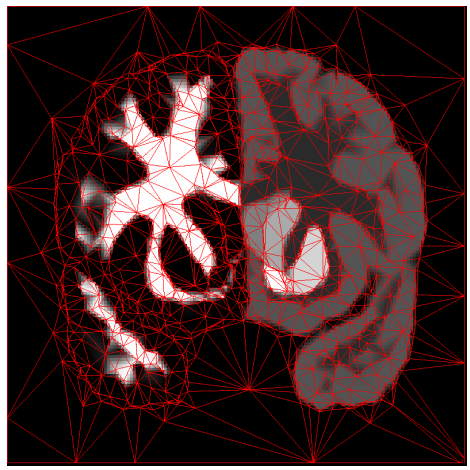


Figure 5.1: An example of a mesh-based brain atlas. Lines show connections between vertices and smooth color variation show probability of different brain structures. Image is taken from [Van Leemput, 2009].

pocampi and optic chiasm. This data set has also been used to build the atlases in the well-known publicly available software package FreeSurfer [Fischl, 2012] and in [Puonti et al., 2016], both for healthy whole-brain segmentation. An example slice of one of the segmented subjects is shown in figure 5.2.

As it was shown in [Puonti et al., 2016] that the segmentation performance does not increase substantially by using more than 10 training subjects for whole-brain segmentation, we selected 10 representative subjects – including both healthy subjects and subjects with varying progression of Alzheimer's disease. As we are interested only in organs-at-risk, we merged most of the 39 labels to the following labels: the organs-at-risk labels hippocampi (HC), brainstem (BS) and optic chiasm (CH); as well as remaining white matter (WM), remaining gray matter (GM), cerebrospinal fluid (CSF) and background.

Some important organs-at-risk are not included in the available manual segmentations. Structures outside of the brain are clearly not included i.e., eyes, eye lenses and optic nerves. Also, the optic tracts are not included. Therefore, we manually segmented these on the T1 images of the 10 selected subjects. In simplified terms, an eye includes fluid within the eye, a solid outer layer and the lens. Fluid has a different intensity profile in MRI than solid tissue. Thus, we segmented it separately. In order to model a structure well, it is important to give some context around the structure. Consequently, we decided to segment the whole eye socket, which – apart from eye and optic nerve – also includes

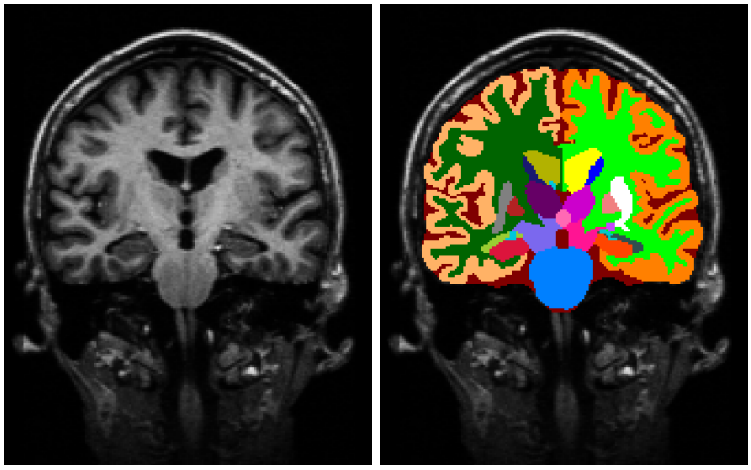


Figure 5.2: A slice of a subject used in building the atlas. From left to right: T1 image of the subject's head and corresponding expert segmentation into 39 neuroanatomical structures. Figure from [Puonti et al., 2016].

muscles and fatty tissue.

To be able to predict the probability of seeing a segmentation $\mathbf{l}$ given the 10 training segmentations, we first build a generative model of how segmentations are formed. We define a reference mesh over the image domain with vertex positions $\boldsymbol{\eta}_r$ and a label probability vector associated with each vertex, with all probability vectors collected in $\boldsymbol{\alpha}$. We define a prior on the deformation of a mesh with vertex positions $\boldsymbol{\eta}$ from the reference mesh as

$$p(\boldsymbol{\eta}|\boldsymbol{\eta}_r, \beta) \propto \exp\left(-\beta \sum_t U_t(\boldsymbol{\eta}, \boldsymbol{\eta}_r)\right),$$

where t denotes a tetrahedron in the mesh, $U_t(\cdot)$ is a penalty for deforming tetrahedron t from its reference position to its actual position, and β is a user-defined constant that controls the stiffness of the mesh. The deformation penalty is the same as proposed in [Ashburner et al., 2000]. This penalty prevents the mesh from tearing or folding onto itself.

We can now build an atlas by refining the reference mesh and learning appropriate values for $\boldsymbol{\alpha}$ given the manual training segmentations. The algorithm is initialized with a reference mesh of high resolution, and by defining a mesh for each segmentation. The algorithm then alternates between two steps. In the *first step*, the probabilities in $\boldsymbol{\alpha}$ and the vertex positions for each training

segmentation mesh are updated iteratively until convergence. While keeping the mesh vertex positions fixed, $\boldsymbol{\alpha}$ is updated using an EM algorithm. The mesh vertex positions are then updated by performing a group-wise non-rigid registration, while keeping $\boldsymbol{\alpha}$ fixed. In the *second step* of the algorithm, the topology of the reference mesh is optimized given the MAP estimates of $\boldsymbol{\alpha}$ and the vertex positions in the training segmentation meshes. For further detail of this atlas building process, see [Van Leemput, 2009].

Our atlas model is then defined as

$$p(\mathbf{l}|\hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\eta}}_r) = \int_{\boldsymbol{\eta}} p(\mathbf{l}|\boldsymbol{\eta}, \hat{\boldsymbol{\alpha}}) p(\boldsymbol{\eta}|\hat{\boldsymbol{\eta}}_r, \beta) \mathrm{d}\boldsymbol{\eta},$$

where the hat variables have been learned in the atlas building process. For the rest of the thesis, we drop the dependency on these learned variables and simply write

$$p(\mathbf{l}) = \int_{\boldsymbol{\eta}} p(\mathbf{l}|\boldsymbol{\eta}) p(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta} \tag{5.1}$$

where we define

$$p(\mathbf{l}|\boldsymbol{\eta}) = \prod_i p(l_i|\boldsymbol{\eta}).$$

Note that we use interpolation to go from vertex positions in the mesh to image voxels. Figure 5.3 shows a slice of the resulting atlas.

5.2 Incorporating mesh-based atlas in the generative brain tumor segmentation model

We now want to incorporate this new deformable atlas model – defined in equation 5.1 – into the segmentation model we used in the previous chapter. We do this by replacing the fixed affine atlas with this new more advanced atlas. However, as the new atlas include the eye socket region, there are some complications involved in this change of atlas model. As there is no barrier such as the blood-brain barrier in parts of the eye socket these parts are enhanced in T1c images, which can result in a similar appearance to enhanced tumor core. Furthermore, the fine organization of tissues with varying image appearance – such as muscles, fatty tissue, blood vessels and nerves – the issue with partial volume effects is more severe. Therefore, we found it necessary to model a restriction for tumor tissue to only appear within the brain. Although tumors can appear virtually anywhere in the body, high-grade gliomas that we focus on in this thesis are confined to the brain. In rare cases, high-grade gliomas can metastasize and spread to other parts of the body and primary gliomas can

Figure 5.3: The resulting atlas including the eye socket. Eye fluid is in yellow. Lenses and the solid outer layers of the eyes are in light orange. Optic nerves are in red. Hippocampi are in dark orange. Brainstem is in green. Other healthy structures included in the atlas are shown in varying shades of blue.

appear in e.g., the optic nerve, but these cases would entail a different treatment protocol and are thus outside of the scope of this thesis.

To model the restriction for tumors to only exist within the brain, we form a joint prior on both the healthy labels $\mathbf{l}$ and the tumor maps $\mathbf{z}$ and $\mathbf{y}$. Using the atlas model defined in equation 5.1 and the tumor shape model defined in equation 4.5, we can form a joint prior defined as

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}) = \int_{\boldsymbol{\eta}} p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\boldsymbol{\eta}) p(\boldsymbol{\eta}) d\boldsymbol{\eta}$$

with

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\boldsymbol{\eta}) = \sum_{\mathbf{H},\mathbf{G}} p(\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}|\boldsymbol{\eta}),$$

where

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}|\boldsymbol{\eta}) \propto \mathrm{e}^{-E_{\mathrm{comb}}(\mathbf{z},\mathbf{y},\mathbf{H},\mathbf{G}) + \sum_i \log p(l_i|\boldsymbol{\eta}) - s(z_i, l_i)}.$$

Here, the added restriction function $s(z, l)$ is constructed as follows: $s(z, l) = \infty$ if $z = 1$ and $l \neq B$, where B denotes healthy brain labels. For all other

combination of values $s(z, l) = 0$. Note that without this function, the prior would devolve to simply $p(\mathbf{l})p(\mathbf{z}, \mathbf{y})$. The updated segmentation posterior is now

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y})p(\mathbf{l}, \mathbf{z}, \mathbf{y}).$$

The full model is illustrated in figure 5.4.

The posterior can be evaluated with the same two-step procedure used in the previous chapter, with just two differences. The most important difference relates to the positions of the mesh vertices $\boldsymbol{\eta}$ that did not exist in the model of the previous chapter. These parameters also need to be taken into account during the inference. Another difference is of course that the simpler energy term that temporarily replaces the combined energy term in the tumor shape model – in the first step of the algorithm – also needs to have the restriction added that tumors can only exist inside the brain.

In the first step of the inference – similar to the previous chapter – we focus on evaluating the posterior of the model parameters, here including both the intensity parameters and the mesh vertex positions. Using Bayes' rule, this posterior is given by

$$\begin{aligned} p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{D}) &\propto p(\mathbf{D}|\boldsymbol{\theta}, \boldsymbol{\eta})p(\boldsymbol{\theta})p(\boldsymbol{\eta}) \\ &= \left(\sum_{\mathbf{l},\mathbf{z},\mathbf{y}} p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\eta})\ p_{\text{tmp}}(\mathbf{l}, \mathbf{z}, \mathbf{y}|\boldsymbol{\eta}) \right) p(\boldsymbol{\theta})p(\boldsymbol{\eta}), \end{aligned}$$

where $p_{\text{tmp}}(\mathbf{l}, \mathbf{z}, \mathbf{y}|\boldsymbol{\eta})$ denotes the prior probability where the cRBM energy term have been replaced with the temporary factorizable energy term, similar to section 4.2.3.

We already know that we can find MAP estimates of the intensity parameters with the GEM algorithm. However, MAP estimates of the vertex positions cannot be found with GEM. Therefore, the posterior is evaluated with a coordinate ascent algorithm. In coordinate ascent, the objective function is maximized iteratively by maximizing in one coordinate direction in each iteration. Here, the posterior is maximized with respect to the intensity parameters by GEM, while keeping the vertex positions fixed. Then, the posterior is maximized with respect to the vertex positions while keeping the intensity parameters fixed. For this, we use a standard conjugate gradient optimizer [Shewchuk, 1994], where the vertex positions are updated iteratively by using the gradient. The algorithm alternate between updating each set of parameters while keeping the other fixed until convergence. A detailed description of this approach can be found in [Van Leemput, 2009].

Similar to the previous chapter, we reintroduce the tumor shape model in the

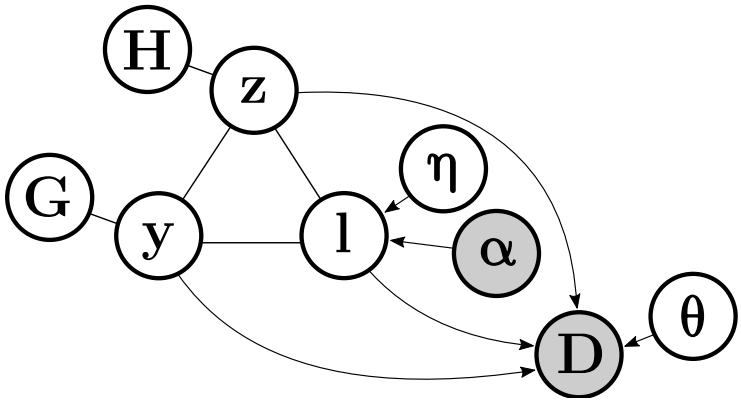


Figure 5.4: Graphical representation of full segmentation model. The mesh-based healthy tissue atlas is defined by node locations in $\boldsymbol{\eta}$ and probabilities in $\boldsymbol{\alpha}$. The complete tumor label $\mathbf{z}$ and core label $\mathbf{y}$ are connected to the hidden units of their cRBM models $\mathbf{H}$ and $\mathbf{G}$, respectively. The labels $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ jointly predict the multi-sequence data $\mathbf{D}$ according to the model parameters $\boldsymbol{\theta}$. Shading indicates observed variables.

second step of the algorithm and initialize a Gibbs sampler with the MAP estimates of the labels – obtained from step 1. We keep the mesh vertex positions fixed throughout the sampling procedure. We could sample from the vertex positions as well, as in e.g., [Iglesias et al., 2013b], which could potentially improve performance. However, as this is time consuming, we chose to keep them fixed. Another problem is that the tumor segmentation is still somewhat sensitive to the initialization of the algorithm. A more robust approach could potentially be to include the sampling in an iterative scheme together with the first inference step, although this would require some further development of the inference algorithm. A further development could perhaps be inspired by the tumor segmentation approach in [Sanjuán et al., 2013]. In this approach, the algorithm is run two times, where the results from the first run is used to refine the segmentation in the second run.

Illustrating the inference algorithm. Figure 5.5 illustrates the different steps of the inference on a representative subject. The first row in the figure shows the included MR sequences for this subject, the second row shows the posterior at different steps of the algorithm overlayed on the T1c image and the third row shows the deformation of the atlas for the first step of the algorithm. In the *initialization* of the algorithm, the atlas is affinely registered to the subject

and the intensity model is initialized based on the atlas (except for the two tumor Gaussians, which are initialized in relation to the average image intensities). Clearly, the atlas does not represent the subject well. In the end of *step 1*, the atlas has adapted to the subject, which is most clearly seen for the eyes and hippocampi that have moved to the right locations. Note that the left optic nerve is slightly off, which is due to the noticeable inconsistencies between the images – which could be due to eye movements, small registration errors between the images or partial volume effects. Furthermore, in the end of step 1, the tumor Gaussians are modeling edema and enhanced core fairly well. In *step 2* of the algorithm, after 200 sampling steps, the tumor shape prior has cleaned up the tumor segmentation considerably and the cavity from surgically removed tumor tissue has been included in the tumor core segmentation. Note that the atlas does not deform during the second step of the algorithm.

5.3 Results and discussion

In Paper C, we validated the presented segmentation method on a data set of 42 glioblastoma patients that underwent radiation therapy treatment at Rigshospitalet in Denmark around 2012. The patients were scanned with three MR sequences: FLAIR, T1c and T2. Also, a CT scan and a FET-PET scan were acquired. As part of the treatment planning, the following structures have been manually segmented: tumor core, brainstem, hippocampi, eyes, optic nerves, optic chiasm, optic tracts, lens and whole brain. The manual segmentations were performed on the MR images resampled to fit the CT image (with a slice thickness of around 3 mm). This is the standard work-up for radiation therapy of glioblastoma patients at Rigshospitalet. The full planning procedure is detailed in [Munck af Rosenschöld et al., 2011, Munck af Rosenschöld et al., 2015]. In the validation, we chose to focus on the most important structures: tumor core, brainstem, hippocampi, eyes, optic nerve and optic chiasm.

As in the previous chapter, we evaluated the automatic segmentations by comparing to the corresponding manual segmentations using Dice scores and Hausdorff distances, i.e., we treat the manual segmentations as the ground truth. The Dice score measures overlap; a perfect overlap results in a score of 1 and no overlap results in a score of 0. Hausdorff distance measures distance between the segmentation borders. Figure 5.6 shows box plots of the results. The scores for tumor core are fairly consistent with the scores on the BRATS data set in figure 4.12. The scores are somewhat better here, because the BRATS data included many low-grade tumors – for which tumor core is less defined – and also some quite challenging high-grade tumors.

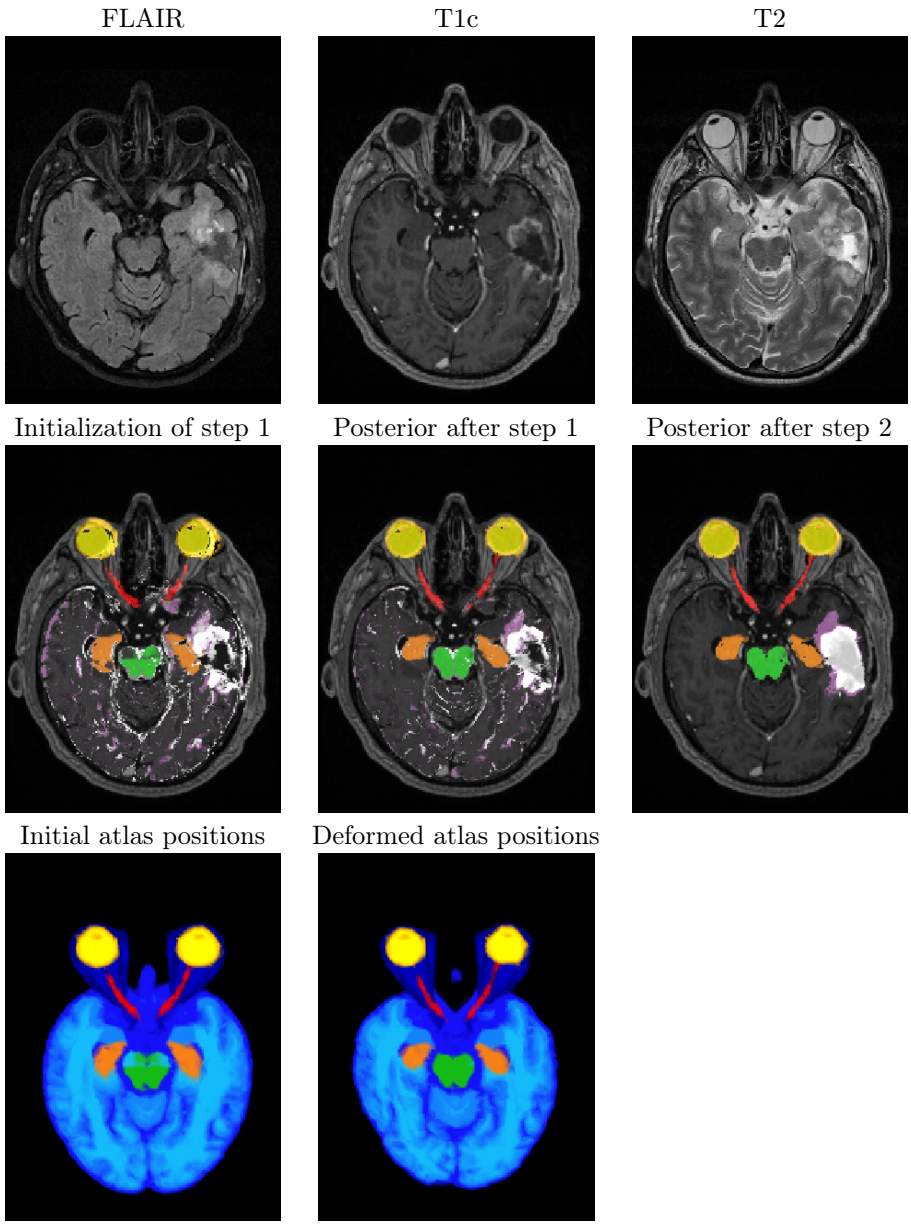


Figure 5.5: Illustration of the two steps of the algorithm for one subject. *First row:* MR sequences of the subject. *Second row*, from left to right: prior of organs-at-risk and tumor after initial affine coregistration of the atlas with the subject, posterior after step 1, posterior after 200 sampling steps in step 2. *Third row*, from left to right: initial position of atlas after affine registration, deformed atlas after step 1. Note that atlas positions are kept fixed in step 2.

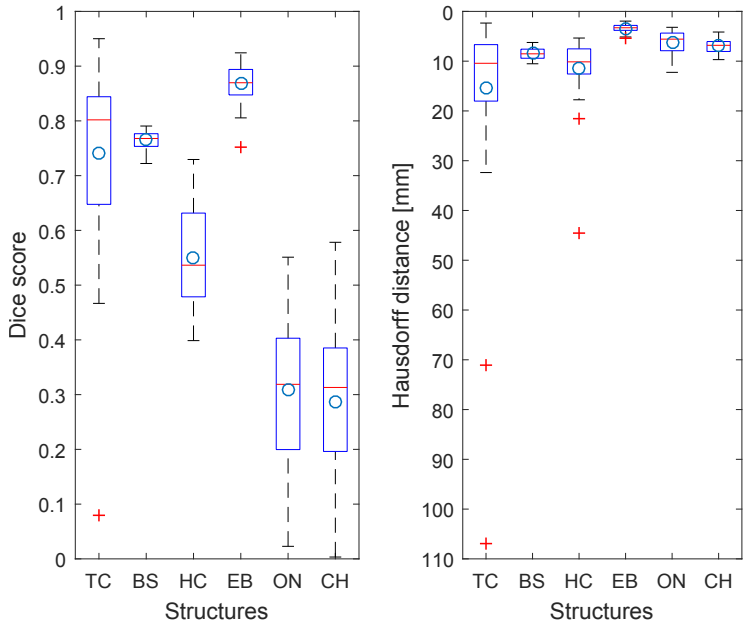


Figure 5.6: Boxplot of Dice scores and Hausdorff distances for the following structures: tumor core (TC), brainstem (BS), hippocampi (HC), eyes (EB), optic nerves (ON), chiasm (CH). Middle red line indicates median value and cyan circle indicates mean value.

The scores for brainstem are consistent over the subjects, but the average Dice score is not very high. Figure 5.7 shows a comparison of the automatic and manual segmentation for one representative subject, to illustrate the differences. In general, both the automatic and manual segmentations capture the brainstem well – although the automatic segmentations seem to be slightly undersegmented. However, the segmentation protocols differ for the training segmentations used to build the atlas and the segmentations from the clinic. Most importantly, a large part of the midbrain is not included in brainstem in the protocol used for the training segmentations.

At first sight, the numerical scores for hippocampi seems to suggest a fairly low performance. However, a careful inspection of the automatic and manual segmentations revealed that the quality problem is rather in the manual segmentations. Although they do capture the central parts of the hippocampus region, they tend to undersegment the full region in a rather inconsistent manner. The hippocampus segmentation of a representative subject is shown in figure 5.8. In this figure, we can see that the automatic segmentation appears to be superior. However, our method tends to slightly oversegment some parts of the

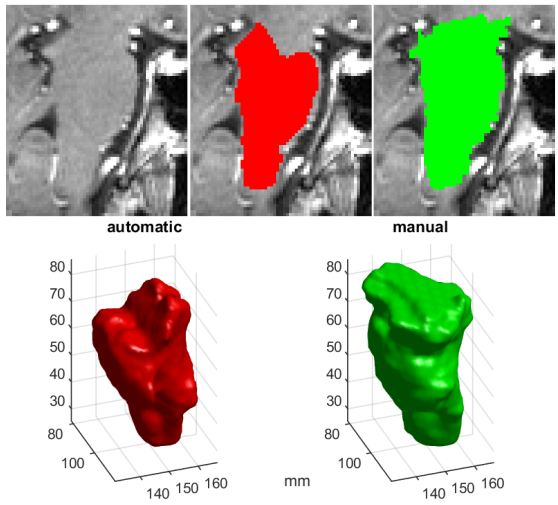


Figure 5.7: Representative example of brainstem segmentation, where the automatic segmentation is shown in red and the manual ground truth segmentation in green. First row show segmentations overlayed on a sagittal slice of the T1c image. Second row show surface plots of the segmentations.

hippocampus, which can be seen in the closest hippocampus in the surface plot.

The scores for eyes show a fairly good and consistent performance. However, the scores for other parts of the optic system, i.e., optic nerves and chiasm, are not optimal. The Hausdorff distances seem somewhat high, but most importantly the Dice scores are low. In part, this is because the Dice score measures overlap and is thus more sensitive for small and thin structures than large structures. This can be seen for a representative subject in figure 5.9. Although the optic nerve segmentations are reasonably close, the Dice score is only 0.33. To further illustrate this point, we can compare to the inter-rater variability shown in [Deeley et al., 2011], which reported an average inter-rater Dice score as low as 0.50 for optic nerves and 0.39 for optic chiasm. Still, our average Dice scores are lower than these results.

To demonstrate our method's flexibility in handling varying MR sequences as input, we conducted a small exploratory experiment on a data set of 7 tumor patients. In contrast to the other data sets we have used in the project, this data set includes a double-inversion recovery (DIR) image but no T1 image. Another difference is that the images were acquired in 3D rather than in 2D slices. The only two changes we did in the algorithm was to (1) affinely register the atlas to

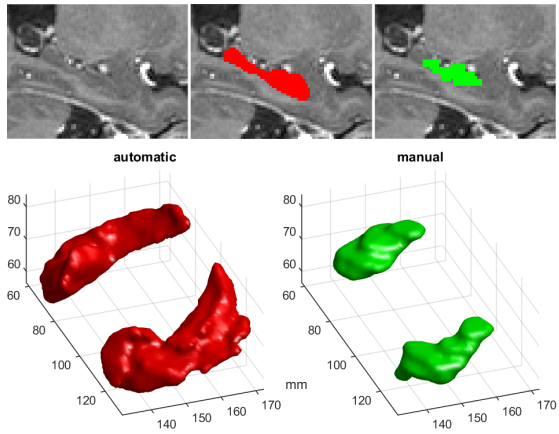


Figure 5.8: Representative example of hippocampus segmentation, where the automatic segmentation is shown in red and the manual ground truth segmentation in green. First row show segmentations overlayed on a sagittal slice of the T1c image. Second row show surface plots of the segmentations.

T2 instead of T1c in the initialization, and (2) initialize the tumor mean values in DIR the same way as for FLAIR. Unfortunately, this data set did not include any manual segmentations to compare against, but a visual inspection revealed that the method performed similar as for the other data sets. A representative subject is shown in figure 5.10.

For radiation therapy, FET-PET is an important modality to include in the segmentation process, because it visualizes metabolic activity of tumors. Theoretically, a PET image could easily be included in the same way as DIR. However, there are two main issues when considering PET: PET images are substantially smoother in comparison to MR images and, more importantly, the boundary of tumor core often differs between PET and MR. Therefore, in practice, the inclusion of PET might not be as straight forward and is in this project left for future work.

In general, the results demonstrate the feasibility of the method for simultaneous segmentation of tumors and organs-at-risk for application in radiation therapy. However, to further validate the automatic segmentations, it would be relevant to study the effect they would have on the radiation dose plan computations. Furthermore, it seems clear that a ground truth with higher quality should be obtained to better validate the model. As the inter-rater variability is usually high [Deeley et al., 2011], a ground truth obtained by fusing segmentations from several experts would be ideal. Another issue with using segmen-

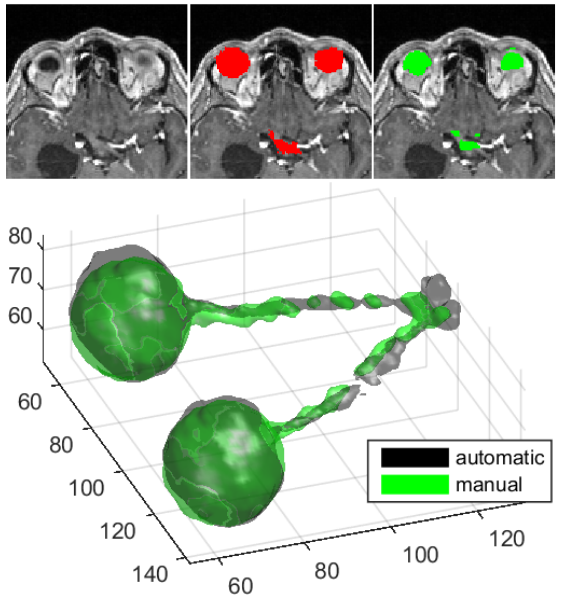


Figure 5.9: Representative example of optic system segmentation (including eyes, optic nerve and chiasm), where the automatic segmentation is shown in red and the manual ground truth segmentation in green. First row show segmentations overlayed on a sagittal slice of the T1c image. Second row show a surface plot of the segmentations, where automatic segmentation is in black and manual segmentation in green.

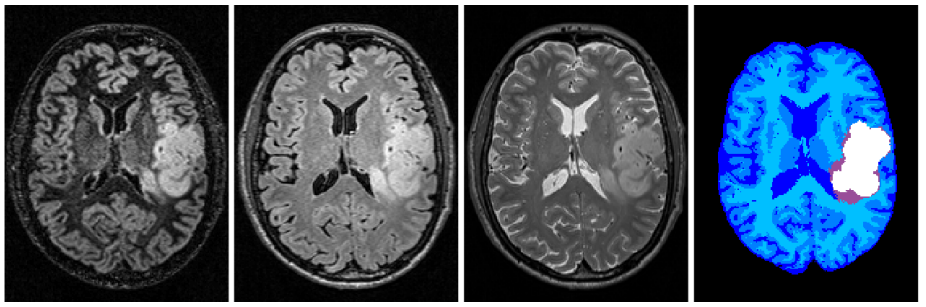


Figure 5.10: Segmentation of a subject with MR DIR sequence. From left to right: MR sequences DIR, FLAIR, T2; and automatic segmentation by the proposed method.

tations from radiation therapy planning is that organs-at-risk far away from the tumor target might not be carefully segmented by the radiologist, as they would not be expected to have a significant effect in the dose plan computation anyway. Furthermore, the atlas should be updated to the same segmentation protocol for brainstem as in the clinic. As we only need 10 manual segmentations to build the atlas, it would be quite feasible to simply adjust each of these segmentations. Another approach could be to fuse the atlas with an atlas with the appropriate segmentation protocol. A framework for this has been presented in [Iglesias et al., 2015b] and investigated for detailed hippocampus segmentation in [Iglesias et al., 2015a] and detailed brainstem segmentation in [Iglesias et al., 2015c].

The segmentation performance on optic nerves and chiasm could probably be improved by incorporating prior knowledge about their specific geometrical structure, e.g., we know that the optic nerves will extend all the way from the eyes to the chiasm without any breaks. Another related issue is that there can be inter-sequence inconsistencies in the appearance of healthy structures, due to small registration errors and imaging artifacts. For most structures, this does not usually pose a problem, but for the thin optic nerve it does. In addition, the optic nerve might move during scanning due to eye movements, which is harder to control than movement of the head. We could consider to weigh the intensity model for optic nerve more heavily on certain sequences, such as T1c – which often have the highest resolution and clearest visualization of this structure.

Chapter 6

Conclusions and future work

In this PhD project, we have developed generative models capable of modeling the shape and structure of brain tumors and several organs-at-risk with application in radiation therapy. We have used the models to incorporate prior knowledge of these structures in generative segmentation methods of medical images, specifically for multi-sequence magnetic resonance images. We have shown that these generative methods have a good segmentation performance for brain tumors with varying grade of severity and several organs-at-risk. Furthermore, the generative framework allows the methods to adapt to changes in the intensity distribution of the input images. Although the methods rely on certain MR sequences to be present, previously unseen MR sequences can also easily be included.

In the first part of the PhD project, we concentrated on the modeling of brain tumors – with a specific interest in high-grade tumors. We developed a tumor shape model based on convolutional restricted Boltzmann machines (cRBMs), which are generative networks capable of automatically learning features from training data. We used cRBMs to learn tumor shape features from binary tumor maps. From a fairly small amount of training data, we were able to automatically learn prior information about the shape of tumors that resulted in a well performing brain tumor segmentation method. In comparison to other state-of-the-art methods, the method performed especially well on tumor core, which is the tumor target in radiation therapy. However, as with other brain

tumor segmentation methods, further development is needed to obtain a truly robust segmentation performance.

In the second part of the PhD project, we turned our focus to modeling organs-at-risk. For this purpose, we used a previously validated probabilistic atlas-based model capable of detailed modeling of the spatial organization in a healthy brain. The parameters of the model are learned from training data in the form of manual brain segmentations. As a small amount of training data is needed to reach an optimal segmentation performance, the atlas can fairly easily be extended to handle new structures or differing segmentation protocols. In this project, we extended the atlas to handle also non-brain organs-at-risk. We incorporated the atlas into the modeling framework that we used in the first part of the project. In experiments, we showed that the updated method was capable of segmenting both the tumor and the organs-at-risk simultaneously. Although more validation is needed with a ground truth of higher quality, the final fully automated segmentation method seems feasible for use in clinical radiation therapy planning. It would work well as an aid in manual segmentation to decrease inter-rater variability and save time for the radiologist. However, further development is needed to be able to use the method in a fully automated radiation therapy planning pipeline.

6.1 Future work

In this section, we will briefly discuss some possible ways forward to further develop the segmentation method developed in this project.

A further validation of the segmentation performance on organs-at-risk is needed. In hindsight, the manual segmentations used in clinical radiation therapy planning are not a good choice as ground truth segmentations. This is both due to the fairly high inter-rater variability but also because some structures far away from the tumor target might not be carefully segmented. At least a careful review of the segmentations would be needed, but fused segmentations from several raters would be ideal. Furthermore, the effect on the radiation dose plan would be appropriate to measure. The brainstem segmentation protocol used for the atlas should also be updated to better adhere to the protocol used in the clinic. In addition, also other organs-at-risk should be validated, such as the lens of the eye and the optic tracts, which were actually included in the atlas.

Although PET is not routinely used in radiation therapy planning at this moment, the use will most probably increase. Therefore, further research on how the segmentation method could handle this modality would be reasonable. How-

ever, PET images are substantially smoother than MR images – which increase the problem with partial volume effects – and more importantly differ compared to MR in the visualization of the tumor border. Therefore, it is probably not as straight forward to include PET as it is with a new MR sequence. Possibly, the PET image could be treated separately by the segmentation method, which would require a more involved development of the modeling framework. As an example, the generative method in [Menze et al., 2010] treats all MR sequences separately to segment different parts of the tumor.

The modeling of the optic chiasm and optic nerves could be further developed. As these structures are both considerably smaller and thinner in comparison to other validated organs-at-risk, the segmentation of these is more sensitive to imaging artifacts, partial volume effects and small registration errors between the images. In addition, the optic nerve can move depending on eye position. One approach could be to focus on one of the images for the segmentation of these structures. Another approach would be to incorporate further prior information on the specific shape of these structures, e.g., that a healthy optic nerve always will extend from the eye to the chiasm without any gaps.

The tumor shape model works well in the sense that it results in a segmentation method with highly competitive segmentation performance. However, it is still fairly limited to local modeling of tumor shape. A model with a more global capability could potentially be obtained with a deeper architecture with more than one hidden layer. The cRBM model has been used in such an architecture, i.e., in deep belief networks. However, these models are even harder to train than a cRBM and would increase the computational time considerably. Alternative generative networks, such as variational autoencoders could potentially be considered.

Improvements in the inference of the model could be considered, to obtain a more robust segmentation performance. A limitation in the current inference algorithm is that the tumor shape model does not inform the deformation of the healthy atlas, as it is introduced in the second step of the algorithm. Another problem is that the tumor segmentation is still somewhat sensitive to the initialization of the algorithm. A more robust inference could perhaps be obtained by a more iterative approach, such as in [Sanjuán et al., 2013], where the algorithm is run two times and the results from the first run is used to refine the segmentation in the second run. Finally, the Gibbs sampler is a fairly basic sampler which does not necessarily explore the full space of possible parameter combinations efficiently. Thus, a more advanced sampler could possibly improve robustness.

Chapter 7

Paper A

Brain Tumor Segmentation Using a Generative Model with an RBM Prior on Tumor Shape

Mikael Agn[1( )], Oula Puonti[1], Per Munck af Rosenschöld[2], Ian Law[3], and Koen Van Leemput[1,4]

[1] Department of Applied Mathematics and Computer Science, Technical University of Denmark, Kongens Lyngby, Denmark
miag@dtu.dk
[2] Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[3] Department of Clinical Physiology, Nuclear Medicine and PET, Rigshospitalet, Copenhagen University Hospital, Copenhagen, Denmark
[4] Martinos Center for Biomedical Imaging, MGH, Harvard Medical School, Boston, USA

Abstract. In this paper, we present a fully automated generative method for brain tumor segmentation in multi-modal magnetic resonance images. The method is based on the type of generative model often used for segmenting healthy brain tissues, where tissues are modeled by Gaussian mixture models combined with a spatial atlas-based tissue prior. We extend this basic model with a tumor prior, which uses convolutional restricted Boltzmann machines (cRBMs) to model the shape of both tumor core and complete tumor, which includes edema and core. The cRBMs are trained on expert segmentations of training images, without the use of the *intensity information* in the training images. Experiments on public benchmark data of patients suffering from low- and high-grade gliomas show that the method performs well compared to current state-of-the-art methods, while not being tied to any specific imaging protocol.

1 Introduction

Brain tumor segmentation from multi-modal magnetic resonance (MR) images is of high value in radiotherapy planning. Automatic tumor segmentation is challenging since tumor location, shape and appearance vary greatly across patients. Moreover, brain tumor images often exhibit significant intensity inhomogeneity as well as large intensity variations between subjects, particularly when they are acquired with different scanners or at different imaging facilities.

Most current state-of-the-art methods in brain tumor segmentation use a *discriminative* approach, which exploits the specific intensity contrast information of annotated training images, e.g., [1–3]. This may hinder their applicability to images acquired at different centers, because the intensity contrast depends

A. Crimi et al. (Eds.): BrainLes 2015, LNCS 9556, pp. 168–180, 2016.
DOI: 10.1007/978-3-319-30858-6_15

on the scanner and the imaging protocol that has been used. Many discriminative methods have been based on the random forest (RF) classification scheme, which predicts segmentation labels from user-engineered image features. One such a method is the winner of the 2013 brain tumor segmentation (BRATS) challenge [4], developed by Tustison et al. [1]. Another large group of discriminative methods are based on deep convolutional neural networks (CNNs) that are capable of automatically learning features from image intensity information. CNNs have recently proved successful in many segmentation tasks. At the 2015 BRATS challenge, two such methods achieved a high segmentation accuracy: Havaei et al. [2] developed a two-way architecture of CNNs that captures both local details and larger contexts; whereas Pereira et al. [3] trained one CNN for high-grade gliomas and another for low-grade gliomas, which proved useful because of the differences between these two types of tumors. However, the latter method requires the user to manually select one of the CNNs.

In contrast to these discriminative methods, Kwon et al. [5] developed a successful semi-automatic *generative* method, which does not use intensity information from training images. This method, however, requires the user to manually assign tumor seed points and radii to initialize the tumor growth model used in the method. For the 2015 BRATS challenge, the same group extended a version of this generative method with a discriminative post-processing step using a gradient boosting multi-class classification scheme followed by a patient-wise refinement step, which increased the segmentation accuracy [6]. Some fully automated generative approaches have previously been proposed, such as [7,8], but with generally lower segmentation accuracy.

In this paper we propose a fully automated generative method that achieves segmentation accuracy comparable to state-of-the-art discriminative methods while being contrast-adaptive. To achieve this, we incorporate a prior on tumor shape into an atlas-based probabilistic model for healthy tissue segmentation. The prior models tumor shape by convolutional restricted Boltzmann machines (cRBMs), which are higher-order Markov random fields (MRFs) capable of modeling more complex interactions than traditionally used first-order MRFs. The features of the cRBMs are learned automatically from expert segmentations of training data without the use of the *intensity information* corresponding to these segmentations. This allows the model to adapt to varying intensity contrasts during the segmentation phase. Experiments on the test data sets of the 2013 and 2015 BRATS challenges show that the method compares well to the current state-of-the-art.

2 Generative Modeling Framework

Let $\mathbf{D} = (\mathbf{d}_1, ..., \mathbf{d}_I)$ denote the multi-contrast MR data of a subject, where I is the number of voxels and $\mathbf{d}_i$ contains the (log-transformed) intensities at voxel i. We aim to segment each voxel i into either one of K healthy tissue labels $l_i \in \{1, ..., K\}$ or tumor tissue $z_i \in \{0, 1\}$, and within tumor tissue into either edema or core $y_i \in \{0, 1\}$. We also aim to segment the voxels in the

core that are enhanced in T1c (see [4] for a description of the tumor tissue labels). For this purpose we build a generative model that describes the image formation process, illustrated in Fig. 1. We then use this model to obtain a fully automated segmentation algorithm by focusing on the posterior of the segmentation variables given the data:

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D}) \propto p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y})p(\mathbf{l})p(\mathbf{z}, \mathbf{y}) \quad \text{with} \tag{1}$$

$$p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}) = \int_{\boldsymbol{\theta}} p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta})p(\boldsymbol{\theta})\mathrm{d}\boldsymbol{\theta},$$

where $\mathbf{l} = (l_1, ..., l_I)$, $\mathbf{z} = (z_1, ..., z_I)$ and $\mathbf{y} = (y_1, ..., y_I)$; and $\boldsymbol{\theta}$ contains free model parameters. The model consists of the likelihood function $p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta})$, which links labels to MR intensities; and the priors $p(\mathbf{l})$, $p(\mathbf{z}, \mathbf{y})$ and $p(\boldsymbol{\theta})$.

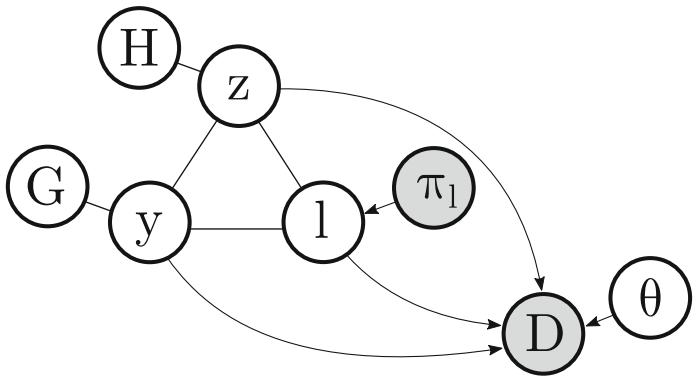


Fig. 1. Graphical representation of the model. The atlas-based prior $\boldsymbol{\pi}_l$ models healthy tissue labels $\mathbf{l}$. The complete tumor label $\mathbf{z}$ and core label $\mathbf{y}$ are connected to the hidden units of their RBM models $\mathbf{H}$ and $\mathbf{G}$, respectively. The labels $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ jointly predict the multi-contrast data $\mathbf{D}$ according to the model parameters $\boldsymbol{\theta}$. Shading indicates observed variables.

For the likelihood $p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta})$, we use Gaussian mixture models (GMMs) to model the relationships between tissue labels and image intensities. Furthermore, we model bias fields corrupting the MR scans as linear combinations of spatially smooth basis functions added to the scans [9]. Specifically, we define the likelihood as

$$p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta}) = \prod_i \begin{cases} p_i(\mathbf{d}_i|\boldsymbol{\theta}_{l_i}, \mathbf{C}) & \text{if } z_i = 0 \text{ and } y_i = 0, \text{ (healthy)} \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_e, \mathbf{C}) & \text{if } z_i = 1 \text{ and } y_i = 0, \text{ (edema)} \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_c, \mathbf{C}) & \text{if } z_i = 1 \text{ and } y_i = 1, \text{ (core)} \\ (y_i = 1 \text{ and } z_i = 0 \text{ prohibited by prior, see Eq. 11}) \end{cases}, \tag{2}$$

where $\boldsymbol{\theta}_x$ denotes the parameters of the GMM connected to tissue x and $\mathbf{C} = (\mathbf{c}_1, ..., \mathbf{c}_N)$, where $\mathbf{c}_n$ denotes the parameters of the bias field model for MR contrast n. All GMM and bias field parameters are collected in $\boldsymbol{\theta}$. We define a Gaussian mixture model, with G_x Gaussian components, as $p_i(\mathbf{d}_i|\boldsymbol{\theta}_x, \mathbf{C}) = \sum_{g=1}^{G_x} \gamma_{xg} \mathcal{N}(\mathbf{d}_i - \mathbf{C}^T \boldsymbol{\phi}^i | \boldsymbol{\mu}_{xg}, \boldsymbol{\Sigma}_{xg})$, where subscript g denotes a Gaussian component within the Gaussian mixture model; $\mathcal{N}(\cdot)$ denotes a normal distribution; and the parameters γ_{xg}, $\boldsymbol{\mu}_{xg}$ and $\boldsymbol{\Sigma}_{xg}$ are the weight, mean and covariance of the corresponding Gaussian. Furthermore, $\boldsymbol{\phi}^i$ evaluates the basis functions of the bias field model at voxel i. We assume that one of the Gaussian components of the core will correspond to the enhanced parts of the core.

For the healthy tissue prior $p(\mathbf{l})$, we use a probabilistic affine atlas computed from segmented healthy subjects [10], defined as $p(\mathbf{l}) = \prod_i \pi^i_{l_i}$. In the ith voxel, π^i_{WM}, π^i_{GM}, π^i_{CSF} and π^i_{BG} denote the prior probability for white matter (WM), gray matter (GM), cerebrospinal fluid (CSF) and background (BG) respectively. Note that the atlas does not include a vessel label, i.e., vessels are not directly handled by the model. However, they do not typically affect the final tumor segmentation due to their small size. The affine registration of the atlas is often insufficient for capturing the displacement of healthy tissues seen in many tumor patients due to the so-called mass effect of tumors. We therefore add an extra healthy label OTHER to the atlas with a constant prior probability $\pi^i_{\text{OTHER}} = 0.1$, to put some probability mass in otherwise unexpected places. We then re-normalize the probability maps to ensure that they sum to one.

For the prior $p(\boldsymbol{\theta})$ on the distribution parameters, we use uniform priors on $\mathbf{C}$ and most mean vectors. However, we found it beneficial to use a prior with a linear constraint for edema and WM. We model these two tissues with just one Gaussian component each, and define a prior over their mean vectors as

$$p(\boldsymbol{\mu}_e, \boldsymbol{\mu}_{WM}) = \begin{cases} \propto 1 & \text{if } \mu_{e,\text{FLAIR}} \geq \alpha\, \mu_{\text{WM,FLAIR}} \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

to encode our prior knowledge that edema appears brighter than WM in FLAIR. Here, the scalar α defines the limit of how close $\mu_{e,\text{FLAIR}}$ can be to $\mu_{\text{WM,FLAIR}}$.

For each GMM's mixture weights, collected in vector $\boldsymbol{\gamma}_x$, we use the conjugate prior

$$p(\boldsymbol{\gamma}_x) = \text{Dir}(\boldsymbol{\gamma}_x|\boldsymbol{\beta}), \tag{4}$$

which is Dirichlet distributed [11, Ch. 3.4]. Each element of $\boldsymbol{\beta}$ is set to 1000 to discourage the removal of Gaussian components. For each GMM's covariances, we use the conjugate prior

$$p(\boldsymbol{\Sigma}_{xg}) = \mathcal{W}^{-1}(\boldsymbol{\Sigma}_{xg} \,|\, v^0_{xg} \boldsymbol{\Sigma}_0,\, v^0_{xg}), \tag{5}$$

which is inverse-Wishart distributed [11, Ch. 4.6]. The matrix $\boldsymbol{\Sigma}_0$ is our prior belief of the covariance structure. We set off-diagonal elements in $\boldsymbol{\Sigma}_0$ to zero and diagonal elements to the variances of the intensities in the whole brain divided

by the number of Gaussians in the full model. The scalar v_{xg}^0 defines the strength of the prior. As healthy tissues are rather well-defined, we set v_{xg}^0 to zero for the healthy Gaussians to obtain uniform priors. For the tumor Gaussians, we set v_{xg}^0 to 20 % of their estimated volumes (cf. Sect. 3.1 for details).

Finally, for $p(\mathbf{z}, \mathbf{y})$ we use a convolutional RBM model, defined below.

2.1 Tumor Prior

We model tumor shape by RBMs, which are higher-order MRFs that are capable of modeling higher-order interactions. An RBM is a graphical model with a set of visible units and a set of hidden units, where connections exist between the two sets but not between the units within each set. This restriction facilitates inference with the model. To allow for more efficient inference over large images without a predefined size, we use *convolutional* RBMs (cRBMs), where the connection weights are shared among all locations [12]; see Fig. 2 for an example. In particular, for modeling tumor label $\mathbf{z}$ we use a binary cRBM of the form $p(\mathbf{z}) = \sum_{\mathbf{H}} p(\mathbf{z}, \mathbf{H})$, with $p(\mathbf{z}, \mathbf{H}) \propto \mathrm{e}^{-E(\mathbf{z},\mathbf{H})}$ and the energy term

$$E(\mathbf{z}, \mathbf{H}) = -\sum_k \mathbf{h}_k \bullet (\mathbf{w}_k * \mathbf{z}) - \sum_k b_k \sum_j h_{kj} - a \sum_i z_i, \tag{6}$$

where $\bullet$ denotes element-wise product followed by summation and $*$ denotes convolution. Here the model is defined in 1D to avoid cluttered equations; it is trivial to extend it to 3D images. Each hidden group $\mathbf{h}_k \in \mathbf{H}$ is connected to the visible units in $\mathbf{z}$ with a convolutional filter $\mathbf{w}_k$, which models interactions between the hidden and visible units, effectively detecting specific features in $\mathbf{z}$. Furthermore, each hidden group has a bias b_k and visible units have a bias a, encouraging units to be enabled or disabled.

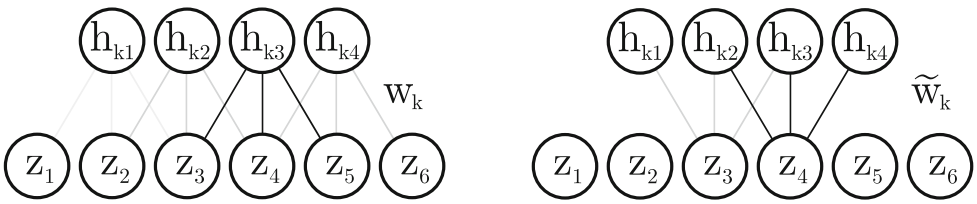


Fig. 2. A small 1D example of a cRBM. Visible units in $\mathbf{z}$ are connected to hidden units in a hidden group $\mathbf{h}_k$ through a convolutional filter $\mathbf{w}_k$ of size 3. The first illustration shows the model from the hidden layer's perspective. The second shows the model from the visible layer's perspective, where $\tilde{\mathbf{w}}_k$ is a mirror-reversed version of the filter. Note that boundary units in the visible layer are set to 0.

We train a cRBM for the complete tumor label $\mathbf{z}$, where we learn the filters and bias terms from expert segmentations of the complete tumor obtained from training data. This is done by stochastic gradient ascent with the contrastive

divergence (CD) approximation of the log-likelihood gradients with one block-Gibbs sampling step (persistent CD was also tried as an alternative to standard CD, but yielded inferior results in our experiments) [13]. We use the so-called enhanced gradient together with the CD approximation to obtain more distinct filters [14]. Due to the structure of the cRBM model, the conditional distributions needed for block-Gibbs sampling are easily obtained as $p(\mathbf{z}|\mathbf{H}) = \prod_i p(z_i|\mathbf{H})$ and $p(\mathbf{H}|\mathbf{z}) = \prod_k \prod_j p(h_{kj}|\mathbf{z})$, where

$$p(z_i = 1|\mathbf{H}) = \sigma((\sum_k \tilde{\mathbf{w}}_k * \mathbf{h}_k)_i + a) \tag{7}$$

$$\text{and} \quad p(h_{kj} = 1|\mathbf{z}) = \sigma((\mathbf{w}_k * \mathbf{z})_j + b_k). \tag{8}$$

Here, $\sigma(t) = 1/(1 + \mathrm{e}^{-t})$ and tilde denotes a mirror-reversal of the filter in each direction. Similarly, we train a cRBM for the tumor core label $\mathbf{y}$, with conditional distributions

$$p(y_i = 1|\mathbf{G}) = \sigma((\sum_k \tilde{\mathbf{u}}_k * \mathbf{g}_k)_i + c) \tag{9}$$

$$\text{and} \quad p(g_{kj} = 1|\mathbf{y}) = \sigma((\mathbf{u}_k * \mathbf{y})_j + d_k), \tag{10}$$

where $\mathbf{G}$ denotes the hidden units connected to $\mathbf{y}$; $\mathbf{u}$ the filters; and c and d the bias terms. After the training phase we combine the two cRBMs to form the joint tumor shape prior:

$$p(\mathbf{z}, \mathbf{y}) = \sum_{\mathbf{H},\mathbf{G}} p(\mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}) \tag{11}$$

$$\text{with} \quad p(\mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}) \propto \mathrm{e}^{-E(\mathbf{z},\mathbf{H})-E(\mathbf{y},\mathbf{G})-\sum_i f(z_i,y_i)},$$

which models both edema and core simultaneously. Here, $f(z_i, y_i) = \infty$ if $z_i = 0$ and $y_i = 1$, and otherwise 0, restricting tumor core to only exist within the complete tumor.

2.2 Inference

Exact inference of $p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D})$ requires an intractable integration over all possible combinations of model parameters. Moreover, even if the model parameters were known the model does not factorize over the voxels, as the cRBM introduces non-local dependencies between them. Therefore, we resort to Markov chain Monte Carlo sampling (MCMC) to generate samples of $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ from $p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D})$, and perform a voxel-wise majority voting across the collected samples to obtain the final segmentation. In particular, we generate samples of $\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}$ and $\boldsymbol{\theta}$ by block-Gibbs sampling from the distribution $p(\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}, \boldsymbol{\theta}|\mathbf{D})$, and ignore the samples of $\mathbf{H}, \mathbf{G}$ and $\boldsymbol{\theta}$ as they are of no interest to us.

Block-Gibbs sampling is straightforward to implement as each of the conditional distributions factorizes over its components: the labels $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ are

sampled simultaneously from the conditional distribution $p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D}, \mathbf{H}, \mathbf{G}, \boldsymbol{\theta})$, for each voxel independently:

$$p(l_i, z_i, y_i|\mathbf{d}_i, \mathbf{H}, \mathbf{G}, \boldsymbol{\theta}) \propto \begin{cases} p_i(\mathbf{d}_i|\boldsymbol{\theta}_{l_i}, \mathbf{C})\, p(z_i = 0|\mathbf{H})\, p(y_i = 0|\mathbf{G})\, \pi^i_{l_i} & \text{if } z_i = 0, y_i = 0 \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_e, \mathbf{C})\, p(z_i = 1|\mathbf{H})\, p(y_i = 0|\mathbf{G})\, \pi^i_{l_i} & \text{if } z_i = 1, y_i = 0 \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_c, \mathbf{C})\, p(z_i = 1|\mathbf{H})\, p(y_i = 1|\mathbf{G})\, \pi^i_{l_i} & \text{if } z_i = 1, y_i = 1 \\ 0 & \text{if } z_i = 0, y_i = 1 \end{cases}.$$

The hidden layers $\mathbf{H}$ and $\mathbf{G}$ are sampled from the conditional distributions $p(\mathbf{H}|\mathbf{z})$ and $p(\mathbf{G}|\mathbf{y})$, given by Eqs. (8) and (10). For the GMM parameters, we iteratively assign voxels to individual GMM components and sample the parameters accordingly ([11, p. 840]). We use rejection sampling to satisfy the constraint of Eq. (3). Note that we could also easily sample from the bias field model, since its conditional distribution is a multi-variate Gaussian. However, this was not implemented. Instead, we use the point estimates of the bias field model parameters $\mathbf{C}$ obtained with the initialization algorithm described below.

We initialize the MCMC sampler with the *maximum a posteriori* segmentation obtained with a generalized expectation-maximization algorithm (GEM) [9]. Since we are only interested in a good parameter initialization at this stage, we temporarily replace the combined cRBM's energy with a simple energy of the form: $-\sum_i [l_i \neq BG](z_i \log w + (1 - z_i) \log(1 - w))$, where w represents the probability of a voxel to be tumor. This reduces the model to the same form as in [9] (with the addition of $p(\boldsymbol{\mu}_e, \boldsymbol{\mu}_{WM})$). We set w to the average fraction of tumor tissue within brain tissue in the training data. At this stage we simply use uniform priors on all covariance matrices.

3 Experiments

We demonstrate the performance of our method on the data of the BRATS brain tumor segmentation challenges. The data sets include high- and low-grade gliomas and consist of four MR-sequences: FLAIR, T2, T1 and contrast-enchanced T1 (T1c). The data are publicly available at the virtual skeleton online platform [15]. Previous to the release of the data sets, all data were skull-stripped and resampled to 1 mm isotropic resolution and the four MR-sequences of each subject were co-registered.

To learn the parameters of the cRBM model, we used the expert segmentations of the BRATS 2013 training data, consisting of 20 high-grade gliomas (HGGs) and 10 low-grade gliomas (LGGs). To internally test our method, we used the 2015 training data with available ground truth segmentations. This data set contains 200 HGGs and 44 LGGs (we excluded the subset of BRATS 2013 training subjects). We then tested the method on two independent test data sets from 2013: the data set used in the 2013 on-site challenge with 10 HGGs, and the leaderboard data set with 25 subjects used for an off-site evaluation including both HGGs and LGGs [4]. Furthermore, we participated in the 2015 challenge where the test data set consisted of 53 subjects, including both HGGs

and LGGs. Note that the ground truth segmentations and tumor grades of the test data sets were not publicly available. Instead, we evaluated our method and compared to other methods by uploading segmentations to the online platform.

3.1 Implementation

We used 40 filters of size $(7 \times 7 \times 7)$ for each cRBM, corresponding to 40 hidden groups. Each cRBM was trained with 9600 gradient steps of size 0.1. A subset of 10 training examples was used to compute the gradient at each step. As the training data set is small, we augmented it by flipping the tumor segmentations in 8 different directions. Furthermore, to reduce the number of parameters to be estimated, we let each element in an cRBM filter model two neighboring elements in $\mathbf{z}$ or $\mathbf{y}$, i.e., a filter of size 7 will span over 14 visible units.

We registered the healthy tissue atlas by an affine transformation and log-transformed the MR intensities, to account for the additive bias field model [9]. The number of components in each GMM was chosen as follows: we represented the core label $\mathbf{y}$ with one Gaussian during GEM initialization, and three during MCMC: one for enhanced core and two for unenhanced core. Before starting the MCMC procedure, the unenhanced core Gaussians were initialized by randomly setting $y_i = 1$ to a fraction of the voxels with $z_i = 1$ and $y_i = 0$ in the GEM segmentation. The fraction was chosen so that the total fraction of core within the complete tumor equaled the average fraction in the training data set. All other labels were represented by one Gaussian each, except CSF and BG that were represented by two Gaussians each.

The healthy tissues' GMM parameters were initialized based on the atlas, except for the label OTHER's mean values which were initialized as the 30th percentile of the brain intensities in each MR-contrast. For tumor tissue, we used the knowledge that edema is always brighter than healthy tissue in FLAIR and T2, and additionally that enhanced core is brighter than any other label in T1c. We therefore initialized the mean values to the percentiles $\{90, 70, 50, 50\}$ and $\{90, 70, 50, 95\}$ in FLAIR, T2, T1 and T1c for edema and core respectively. When validating the method, we found that this initialization is adequate in most cases, i.e., the algorithm is able to adapt to the intensity distribution of a subject. However, the method might fail if e.g., the intensity distribution of tumor tissue is not sufficiently different from healthy tissue due to a bias field.

Due to the large size variation of tumors, we found it necessary to individualize the bias term a connected to $\mathbf{z}$ in Eq. (7) to better represent the tumor to be segmented. We therefore added $\log\left(\frac{p_{zs}(1-p_{zt})}{p_{zt}(1-p_{zs})}\right)$ to a, where p_{zs} denotes the fraction of tumor within the GEM-segmented brain and p_{zt} denotes the average tumor size in the data used to train the cRBM. We did the same for the bias term connected to $\mathbf{y}$ in Eq. (9), matching it with the average fraction of core within complete tumor in the training data set.

As discussed previously for the Gaussians modeling tumor tissue, we set the strength of the covariance priors (v^0_{xg} in Eq. 6) to 20 % of the number of voxels belonging to each Gaussian in the initial segmentation. However, due to the large

changes of tumor size during sampling, we found it necessary to re-estimate v_{xg}^0 to 20 % of the volumes in the updated segmentations during sampling. Note that this is non-standard and a more proper use of the priors is left for future work.

Finally, we estimated the limit parameter α (in Eq. 3) to 1.08 by estimating the means on the BRATS 2013 training subjects using the described GEM, but with tumor labels fixed to the ground truth, and subsequently building statistics of the FLAIR mean values in edema and WM.

All computations were done on a i7-5930K CPU and a GeForce GTX Titan Black GPU in MATLAB 2014b. The training phase of each cRBM took around 3 days on the GPU. The full segmentation algorithm takes approximately 30 min per subject, including atlas registration (CPU), GEM-initialization (CPU) and sampling (GPU). The sampling is the most time consuming part, taking 25 min on average, mainly due to the many convolutions that are involved. We generated 15 samples after a burn-in period of 200 samples and obtained the final segmentation by majority voting on these 15 samples.

3.2 Results

In Table 1, we compare our method on the three test data sets described in the beginning of this section to the five state-of-the-art methods discussed in the introduction. The evaluated labels are the complete tumor (which includes tumor core and edema), the core region of the tumor, as well as the enhancing regions within the core. Our method performed comparably well on complete tumor and core, but not as well on enhanced core. When comparing average Dice scores in the 2015 challenge, out of 13 participants we ranked 2nd for complete tumor, 1st for core and 6th for enhanced core. The lower performance on enhanced core

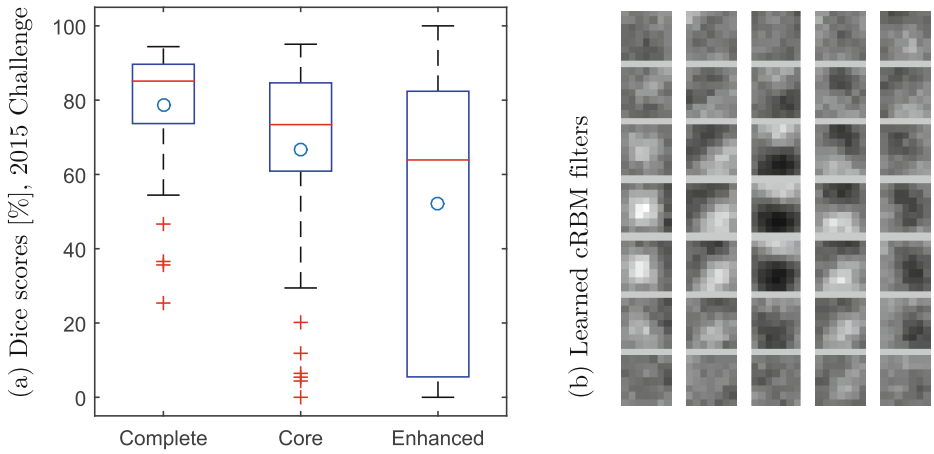


Fig. 3. (a) Box plot of Dice scores, 2015 challenge. Circles show mean values, central lines show medians, edges of boxes show the 25th and 75th percentiles, and outliers are marked with '+'. (b) 5 learned cRBM filters for complete tumor.

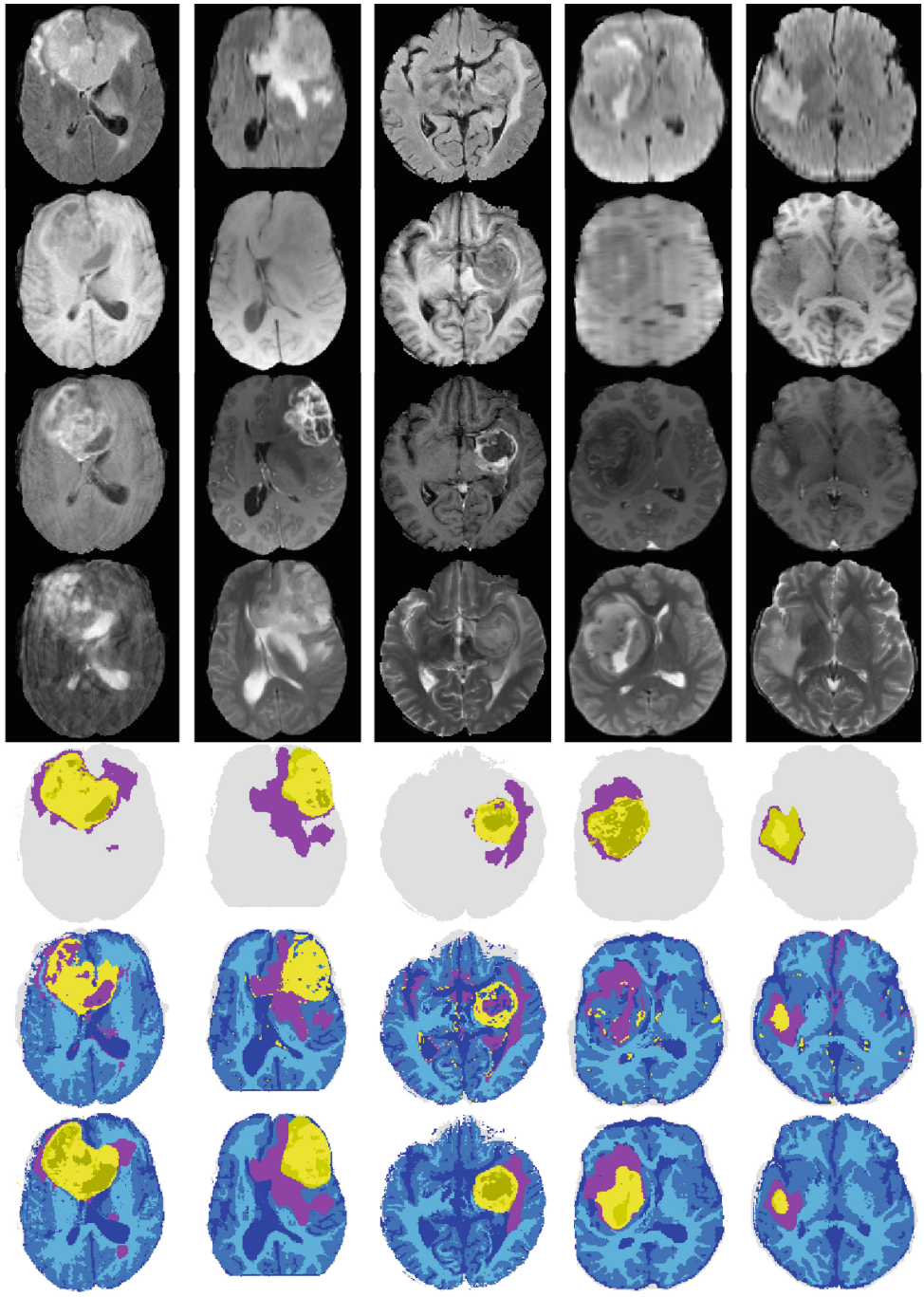


Fig. 4. Slices of five exemplary subjects, 2015 training data. The last subject has a low-grade tumor and the rest high-grade tumors. From top to bottom: MR-contrasts: FLAIR, T1, T1c and T2; ground-truth segmentation; initial GEM-segmentation; and final segmentation. Healthy labels are in blue to cyan, edema is in lilac and core is in different shades of yellow (Color figure online).

Table 1. Average Dice scores (%) for the BRATS test data sets [15].

Data set:	2013 Challenge			2013 Leaderboard			2015 Challenge		
	Comp	Core	Enh	Comp	Core	Enh	Comp	Core	Enh
Our method	87	82	70	83	71	54	81	68	65
Random forest method [1]	87	78	74	79	65	53	–	–	–
Two-way CNN method [2]	88	78	73	81	67	55	79	62	72
Grade-specific CNN method [3]	88	83	77	84	72	62	78	65	75
Generative method [5]	88	83	72	86	79	59	–	–	–
Generative-Discriminative [6]	–	–	–	–	–	–	82	59	74

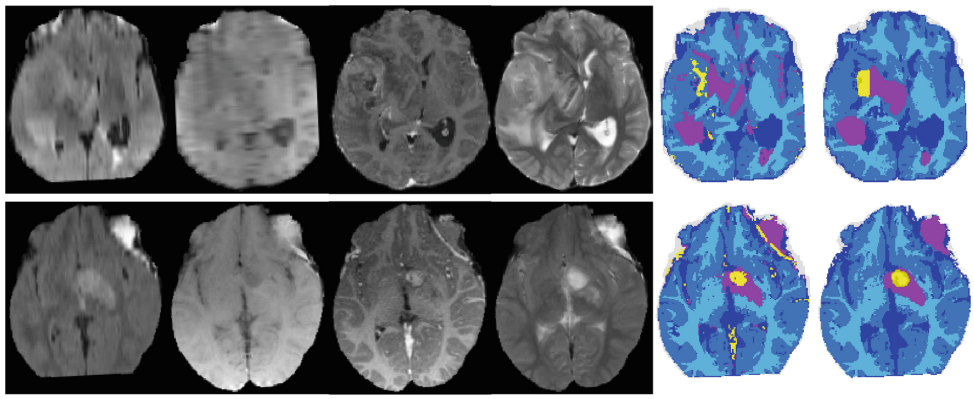


Fig. 5. Slices of two failed segmentations, 2015 test data. From left to right: MR-contrasts, initial GEM segmentation and final segmentation. Dice scores (%) of complete tumor, core and enhanced core: (54, 12, 18) and (36, 87, 0).

is not surprising, as we base this segmentation on a single Gaussian intensity distribution without any spatial prior to separate it from the rest of the core.

Figure 3a shows a box plot of the resulting Dice scores for the test data of the 2015 challenge. We can see that the method on average performs well on complete tumor and core, but with a considerable amount of more or less failed segmentations. It performs substantially worse on the enhanced core. Figure 3b shows five of the automatically learned filters of the complete tumor cRBM.

A few example segmentations by the proposed method are shown in Fig. 4, together with initial GEM-segmentations, the ground truth segmentations and the MR data. Here we can see that the method is capable of capturing varying tumor shapes, removing many false positives from the initial segmentation (e.g., vessels) and recovering when a large part of the core initially has been labeled as healthy tissue. However, the rather localized shape model does have limitations, e.g., it has difficulties to remove sizable ventricular CSF flow artifacts and it tends to oversmooth the tumor border. Furthermore, the intensity difference between edema and core is not always clear; the last subject has a typical low-grade tumor appearance, where this difference is almost non-existent.

Figure 5 shows two failed subjects. In the first subject, large parts of the core exhibit a similar intensity distribution to GM, mainly due to a bias field in FLAIR combined with low enhancement in T1c. The interaction between the strong edge detecting cRBM filters and the smooth affine healthy atlas is not ideal in this case. For the second subject, a large part of non-tumor tissue is better explained by the intensity distribution of tumor than the healthy labels.

4 Discussion

In this paper, we have proposed a fully automated generative method for brain tumor segmentation, with a tumor prior that uses convolutional restricted Boltzmann machines to model tumor shape. We have shown that the method's performance compares well to current state-of-the-art methods on public benchmark data sets. Moreover, it is not tied to any specific imaging protocol as the optimal parameters of the tumor model are estimated only from expert *segmentations* of annotated training images, without using intensity information.

Described here is a work in progress with many potential paths of improvement still to be explored. The structure of healthy tissues could probably be better explained by a deformable atlas. Furthermore, it was observed that the proposed sampling method only explores a small part of the total space of possible configurations. This is due to the Gibbs sampling framework and the fairly strong edge detecting filters obtained by training the cRBM model, which result in a slow mixing of the MCMC chain. Although the method is effective in the sense that just a few sampling steps are needed to produce competitive segmentations, it could be more efficient and less dependent on initialization when using a better sampling framework. Future work will involve further experimentation with different filter configurations and deformable atlases, exploration of more efficient sampling frameworks and simultaneous segmentation of important healthy structures for radiotherapy.

Acknowledgments. This research was supported by NIH NCRR (P41-RR14075), NIBIB (R01EB013565) and the Lundbeck Foundation (R141-2013-13117).

References

1. Tustison, N., et al.: ANTs and Árboles. In: Proceedings of the MICCAI-BRATS 2013 (2013)
2. Havaei, M., et al.: A convolutional neural network approach to brain tumor segmentation. In: Proceedings of the MICCAI-BRATS 2015 (2015)
3. Pereira, S., et al.: Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In: Proceedings of the MICCAI-BRATS 2015 (2015)
4. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015)

5. Kwon, D., Shinohara, R.T., Akbari, H., Davatzikos, C.: Combining generative models for multifocal glioma segmentation and registration. In: Golland, P., Hata, N., Barillot, C., Hornegger, J., Howe, R. (eds.) MICCAI 2014, Part I. LNCS, vol. 8673, pp. 763–770. Springer, Heidelberg (2014)
6. Bakas, S., et al.: Segmentation of gliomas in multimodal magnetic resonance imaging volumes based on a hybrid generative-discriminative framework. In: Proceedings of the MICCAI-BRATS 2015 (2015)
7. Menze, B.H., et al.: Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation. In: Proceedings of the MICCAI-BRATS 2012 (2012)
8. Haeck, T., et al.: Automated model-based segmentation of brain tumors in MR images. In: Proceedings of the MICCAI-BRATS 2015 (2015)
9. Leemput, V., et al.: Automated model-based tissue classification of MR images of the brain. IEEE Trans. Med. Imaging **18**(10), 897–908 (1999)
10. Ashburner, J., et al.: Statistical Parametric Mapping. The Wellcome Department Cognitive Neurology, University College London, London, UK. http://www.fil.ion.ucl.ac.uk/spm/
11. Murphy, K.P.: Machine learning: a probabilistic perspective. MIT Press, Cambridge (2012)
12. Lee, H., et al.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. Commun. ACM **54**(10), 95–103 (2011)
13. Fischer, A., et al.: Training restricted boltzmann machines: an introduction. Pattern Recogn. **47**(1), 25–39 (2014)
14. Melchior, J., et al.: How to center binary restricted boltzmann machines. arXiv preprint (2013). arXiv:1311.1354
15. Kistler, M., et al.: The virtual skeleton database: an open access repository for biomedical research and collaboration. J. Med. Internet Res **15**(11), e245 (2013)

Chapter 8

Paper B

A Generative Model for Segmentation of Tumor and Organs-at-Risk for Radiation Therapy Planning of Glioblastoma Patients

Mikael Agn[a], Ian Law[b], Per Munck af Rosenschöld[c], and Koen Van Leemput[a,d]

[a]Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark
[b]Department of Clinical Physiology, Nuclear Medicine and PET, Rigshospitalet, Copenhagen University Hospital, Denmark
[c]Department of Oncology, Rigshospitalet, Copenhagen University Hospital, Denmark
[d]Martinos Center for Biomedical Imaging, MGH, Harvard Medical School, USA

ABSTRACT

We present a fully automated generative method for simultaneous brain tumor and organs-at-risk segmentation in multi-modal magnetic resonance images. The method combines an existing whole-brain segmentation technique with a spatial tumor prior, which uses convolutional restricted Boltzmann machines to model tumor shape. The method is not tuned to any specific imaging protocol and can simultaneously segment the gross tumor volume, peritumoral edema and healthy tissue structures relevant for radiotherapy planning. We validate the method on a manually delineated clinical data set of glioblastoma patients by comparing segmentations of gross tumor volume, brainstem and hippocampus. The preliminary results demonstrate the feasibility of the method.

1. INTRODUCTION

When planning for radiotherapy of brain tumors, several structures need to be segmented from multi-modal magnetic resonance (MR) images. Important structures are healthy sub-cortical structures that should be spared from radiation (so-called organs-at-risk) and the MR-defined gross tumor volume (GTV). Automatic segmentation of tumor-affected brains is challenging since the location, shape and appearance of tumors, as well as the effect of tumors on surrounding healthy tissue, vary greatly across patients. Moreover, brain tumor images often exhibit significant intensity inhomogeneity as well as large intensity variations between patients, particularly when they are acquired with different scanners or at different imaging facilities.

A number of methods have been proposed for automatic brain tumor segmentation, including e.g., [2,3] and other methods evaluated within the MICCAI brain tumor segmentation (BRATS) challenges [1]. However, these methods have typically been focused solely on the segmentation of tumors and are difficult to extend to also segment healthy organs-at-risk. In contrast, there are many successful methods for healthy brain segmentation which are capable of segmenting even small structures with a fair level of accuracy, such as [4,5]. However, these methods typically have difficulties handling abnormal brain tissue such as tumors. A method capable of segmenting healthy brain structures in detail while at the same time being able to handle tumors or other types of abnormal tissue still remains an open problem. Some research with atlas-based methods for brain tumor segmentation have started to look in this direction, e.g., in [6] a semi-automatic method was presented to non-linearly register a healthy atlas to brains with tumors; and in [7] an automatic method was presented for simultaneous segmentation of tumors and sub-cortical healthy structures, but the accuracy was not validated on important organs-at-risk.

In this paper we propose a fully automated generative method to segment brain tumors and organs-at-risk simultaneously. The method is a further development of a previously presented method for brain tumor

Further author information: (Send correspondence to Mikael Agn)
E-mail: miag@dtu.dk

Medical Imaging 2016: Image Processing, edited by Martin A. Styner, Elsa D. Angelini, Proc. of SPIE Vol. 9784, 97841D · © 2016 SPIE
CCC code: 1605-7422/16/$18 · doi: 10.1117/12.2216814

segmentation that has been validated within the BRATS challenge on a publicly available data set of gliomas [8]. Similarly to this previous work, we use a spatial tumor prior consisting of convolutional restricted Boltzmann machines (cRBMs), which are a type of Markov random fields (MRFs) capable of modeling more complex interactions than traditionally used first-order MRFs. We learn the cRBM features automatically from expert segmentations without the use of the *intensity information* corresponding to these segmentations. Therefore, the model is able to adapt to varying intensity contrasts during the segmentation phase. In contrast to our previous method [8], here we combine the cRBM tumor shape prior with a whole-brain segmentation method for healthy brains [5]. The resulting probabilistic atlas-based method is capable of spatially adapting to individual brains, enabling us to obtain detailed segmentations of healthy structures, such as organs-at-risk in radiotherapy, in addition to delineations of the tumor itself.

As an initial evaluation, we test the proposed method on a manually delineated clinical data set of 20 glioblastoma patients. We report segmentation accuracy on GTV and two important organs at risk, brainstem and hippocampus, compared to the manual segmentations. These preliminary results demonstrate the feasibility of the method.

2. MODELING FRAMEWORK

Let $\mathbf{D} = (\mathbf{d}_1, ..., \mathbf{d}_I)$ denote the multi-contrast MR data of a subject after logarithmic transformation, where I is the number of voxels and $\mathbf{d}_i \in \mathbb{R}^N$ contains the log-transformed intensities at voxel i, where N is the number of MR contrasts. We aim to assign one of the following labels to each voxel i: tumor (that is to say the MR-defined GTV), peritumoral edema or one of K healthy tissue labels. For this purpose we build a generative model that describes the image formation process and use this model to obtain a fully automated segmentation algorithm.

Let $\mathbf{l} = (l_1, ..., l_I)^T$ denote the healthy segmentation, where $l_i \in \{1, ..., K\}$. We introduce a binary map $\mathbf{y} = (y_1, ..., y_I)^T$ to indicate voxels that are part of a tumor, and another binary map $\mathbf{z} = (z_1, ..., z_I)^T$ to indicate all tumor-affected voxels, i.e., voxels that are part of either tumor or peritumoral edema. This means that an edema voxel will be indicated by $z_i = 1$ and $y_i = 0$, and a tumor voxel by $z_i = 1$ and $y_i = 1$. Our generative model then consist of a likelihood function $p(\mathbf{D}|\mathbf{l}, \mathbf{y}, \mathbf{z})$ that links labels to MR intensities, and a prior distribution on labels $p(\mathbf{l}, \mathbf{y}, \mathbf{z}) = p(\mathbf{l})p(\mathbf{y}, \mathbf{z})$, where $p(\mathbf{l})$ encodes prior knowledge of healthy brain anatomy and $p(\mathbf{y}, \mathbf{z})$ models the shape of tumor and edema.

For the likelihood $p(\mathbf{D}|\mathbf{l}, \mathbf{y}, \mathbf{z})$, we use Gaussian mixture models (GMMs) to model the relationship between tissue labels and MR intensities. Furthermore, we model bias fields that typically corrupt MR scans as linear combinations of spatially smooth basis functions added to the scans. Letting $\boldsymbol{\theta}$ denote all GMM and bias field parameters with a prior $p(\boldsymbol{\theta})$, the resulting likelihood is given by

$$p(\mathbf{D}|\mathbf{l}, \mathbf{y}, \mathbf{z}) = \int_{\boldsymbol{\theta}} p(\mathbf{D}|\mathbf{l}, \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta} \qquad \text{with}$$

$$p(\mathbf{D}|\mathbf{l}, \mathbf{y}, \mathbf{z}, \boldsymbol{\theta}) = \prod_i \begin{cases} p_i(\mathbf{d}_i|\boldsymbol{\theta}_{l_i}, \mathbf{C}) & \text{if } z_i = 0 \text{ and } y_i = 0, \text{ (healthy)} \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_e, \mathbf{C}) & \text{if } z_i = 1 \text{ and } y_i = 0, \text{ (edema)} \\ p_i(\mathbf{d}_i|\boldsymbol{\theta}_t, \mathbf{C}) & \text{if } z_i = 1 \text{ and } y_i = 1, \text{ (tumor)} \\ (y_i = 1 \text{ and } z_i = 0 \text{ prohibited by prior, see Eq. (1)}) \end{cases} .$$

Here $\boldsymbol{\theta}_x$ denotes the parameters of the GMM connected to tissue x and $\mathbf{C} = (\mathbf{c}_1, ..., \mathbf{c}_N)$, where $\mathbf{c}_n$ denotes the parameters of the bias field model for MR contrast n. We define a Gaussian mixture model, with G_x Gaussian components, as $p_i(\mathbf{d}_i|\boldsymbol{\theta}_x, \mathbf{C}) = \sum_{g=1}^{G_x} \gamma_{xg} \mathcal{N}(\mathbf{d}_i - \mathbf{C}^T \boldsymbol{\phi}^i | \boldsymbol{\mu}_{xg}, \boldsymbol{\Sigma}_{xg})$, where subscript g denotes a Gaussian component within the Gaussian mixture model; $\mathcal{N}(\cdot)$ denotes a normal distribution; and the parameters γ_{xg}, $\boldsymbol{\mu}_{xg}$ and $\boldsymbol{\Sigma}_{xg}$ are the weight, mean and covariance of the corresponding Gaussian. Furthermore, $\boldsymbol{\phi}^i$ evaluates the basis functions of the bias field model at voxel i. We restrict healthy structures with similar intensity properties to have the same GMM parameters. We use the following superstructures: non-brain tissues, cerebrospinal fluid, gray matter structures (including hippocampus) and white matter structures (including brainstem). For the prior $p(\boldsymbol{\theta})$, we follow the formulations in [8]; with uniform priors on $\mathbf{C}$ and most mean vectors, conjugate

priors on weights and covariances, and a prior with a linear constraint on the mean vectors related to edema and healthy white matter structures.

For the healthy segmentation prior $p(\mathbf{l})$, we use a probabilistic atlas learned from manual annotations in 39 subjects as described in [5]. Among many other structures, the atlas includes spatial probability maps of brainstem and hippocampus. The atlas is parametrized by a sparse tetrahedral mesh with node positions $\boldsymbol{\eta}$. Assuming conditional independence of the labels between voxels given $\boldsymbol{\eta}$, the prior is given by

$$p(\mathbf{l}) = \int_{\boldsymbol{\eta}} p(\mathbf{l}|\boldsymbol{\eta}) p(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta} \qquad \text{with}$$

$$p(\mathbf{l}|\boldsymbol{\eta}) = \prod_{i=1}^{I} p_i(l_i|\boldsymbol{\eta}),$$

where $p(\boldsymbol{\eta})$ is a topology-preserving deformation prior [9].

Finally, for $p(\mathbf{y}, \mathbf{z})$ we use a cRBM model, defined below.

2.1 Spatial tumor prior using convolutional RBMs

We model the shape of tumor-affected tissue by restricted Boltzmann machines (RBMs), a type of MRFs that are capable of modeling higher-order interactions between voxels through local connections to hidden units. Connections directly between voxels or between hidden units do not exist, which facilitates inference with the model. To allow for more efficient inference over large images without a predefined size, we use *convolutional* RBMs (cRBMs), where the connection weights are shared among all locations [10]. As in [8], we combine two separate cRBMs to form the tumor prior: one that models interactions in the tumor-affected map $\mathbf{z}$ and one that models interactions in the tumor map $\mathbf{y}$.

In particular, the cRBM that models $\mathbf{z}$ is defined by $p(\mathbf{z}) = \sum_{\mathbf{H}} p(\mathbf{z}, \mathbf{H})$, with $p(\mathbf{z}, \mathbf{H}) \propto \mathrm{e}^{-E(\mathbf{z}, \mathbf{H})}$ and the energy term

$$E(\mathbf{z}, \mathbf{H}) = -\sum_k \mathbf{h}_k \bullet (\mathbf{w}_k * \mathbf{z}) - \sum_k b_k \sum_j h_{kj} - a \sum_i z_i,$$

where $\bullet$ denotes element-wise product followed by summation and $*$ denotes convolution. Here the model is defined in 1D to avoid cluttered equations; it is straightforward to extend it to 3D images. Each hidden group $\mathbf{h}_k \in \mathbf{H}$ is connected to the visible units in $\mathbf{z}$ with a convolutional filter $\mathbf{w}_k$, which models interactions between the hidden and visible units, effectively detecting specific features in $\mathbf{z}$. Furthermore, each hidden group has a bias b_k and visible units have a bias a, encouraging units to be enabled or disabled.

The two cRBMs, defined by $p(\mathbf{z})$ with hidden units $\mathbf{H}$ and $p(\mathbf{y})$ with hidden units $\mathbf{G}$, are trained separately. During the training phase we learn the filters and bias terms from expert segmentations of tumor and edema obtained from training data. This is done by stochastic gradient ascent with the contrastive divergence (CD) approximation of the log-likelihood gradients with one block-Gibbs sampling step [11]. We use the so-called enhanced gradient together with the CD approximation to obtain more distinct filters [12, 13].

After the training phase we combine the two cRBMs to form the joint tumor shape prior:

$$p(\mathbf{y}, \mathbf{z}) = \sum_{\mathbf{G}, \mathbf{H}} p(\mathbf{y}, \mathbf{z}, \mathbf{G}, \mathbf{H})$$

$$\text{with} \quad p(\mathbf{y}, \mathbf{z}, \mathbf{G}, \mathbf{H}) \propto \mathrm{e}^{-E_{\mathrm{comb}}(\mathbf{y}, \mathbf{z}, \mathbf{G}, \mathbf{H})}$$

$$\text{where} \quad E_{\mathrm{comb}}(\mathbf{y}, \mathbf{z}, \mathbf{G}, \mathbf{H}) = E(\mathbf{y}, \mathbf{G}) + E(\mathbf{z}, \mathbf{H}) + \sum_i f(y_i, z_i), \tag{1}$$

which models both the tumor and the surrounding edema simultaneously. Here, $f(z_i, y_i) = \infty$ if $z_i = 0$ and $y_i = 1$, and otherwise 0, restricting tumor to be within the tumor-affected region.

2.2 Inference

Exact inference of $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D})$ requires an intractable integration over all possible combinations of model parameters. To side-step this difficulty, we use the approximation $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}) \simeq p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$, where $\{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}\}$ are the parameter values that maximize $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{D})$. We first use the same model parameter optimization algorithm as in [5] to maximize $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{D})$. After this, we sample from $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \hat{\boldsymbol{\eta}})$ by using Markov chain Monte Carlo sampling (MCMC) and subsequently use voxel-wise majority voting to obtain the final segmentation.

To maximize $p(\boldsymbol{\theta}, \boldsymbol{\eta}|\mathbf{D})$ we can use coordinate ascent as in [5], where the atlas deformation parameters $\boldsymbol{\eta}$ are optimized with a conjugate gradient algorithm, and the intensity model parameters $\boldsymbol{\theta}$ with a generalized expectation-maximization (GEM) algorithm. The optimization is done by iteratively alternating between keeping the deformation parameters fixed while optimizing the intensity model parameters and vice versa until convergence. However, the cRBM model introduces non-local dependencies between the voxels. We therefore temporarily replace the combined cRBM's energy $E_{\text{comb}}(\mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G})$ with a simple energy of the form

$$E_{\text{tmp}}(\mathbf{z}) = -\sum_{i=1}^{I} [l_i \neq NB] \Big(z_i \big(y_i \log(wv) + (1 - y_i) \log(w(1 - v)) \big) + (1 - z_i) \log(1 - w) \Big),$$

where NB denotes non-brain tissues, and w and v are user-specified parameters that essentially define uniform spatial prior probabilities for tumor-affected tissue and tumor, respectively, to occur within the brain. This reduces the model to the same form as in [5], so we can use the same optimization method.

After the initial parameter estimation, we replace the temporary energy with the original cRBM energy and use MCMC to generate samples of $\mathbf{l}$, $\mathbf{y}$ and $\mathbf{z}$ from $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \hat{\boldsymbol{\eta}})$. Note that we keep the parameters $\boldsymbol{\eta}$ fixed to $\hat{\boldsymbol{\eta}}$ during sampling, as these atlas deformation parameters are difficult to sample from. In particular, we generate samples of $\mathbf{l}, \mathbf{y}, \mathbf{z}, \mathbf{G}, \mathbf{H}$ and $\boldsymbol{\theta}$ by block-Gibbs sampling from the distribution $p(\mathbf{l}, \mathbf{y}, \mathbf{z}, \mathbf{G}, \mathbf{H}, \boldsymbol{\theta}|\mathbf{D}, \hat{\boldsymbol{\eta}})$. After that, we discard the samples of $\mathbf{G}, \mathbf{H}$ and $\boldsymbol{\theta}$ as they are of no interest to us, and perform voxel-wise majority voting across the collected samples of $\mathbf{l}$, $\mathbf{y}$ and $\mathbf{z}$ to obtain the final segmentation. Block-Gibbs sampling is straightforward to implement as each of the conditional distributions $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \mathbf{G}, \mathbf{H}, \boldsymbol{\theta}, \hat{\boldsymbol{\eta}})$, $p(\mathbf{G}|\mathbf{y})$, $p(\mathbf{H}|\mathbf{z})$ and $p(\boldsymbol{\theta}|\mathbf{D}, \mathbf{l}, \mathbf{y}, \mathbf{z}, \hat{\boldsymbol{\eta}})$ factorizes over its components. We initialize the MCMC sampler with a *maximum a posteriori* segmentation obtained by evaluating the posterior probability $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}})$ of the initial optimization method.

3. EXPERIMENTS

We demonstrate the performance of the proposed method on a clinical data set of 20 glioblastoma patients that have undergone radiotherapy treatment at Rigshospitalet at Copenhagen University Hospital. These patients have been scanned with a CT scanner and a Siemens Magnetom Espree 1.5T MRI scanner. The data set includes three MR-sequences: FLAIR, T2 and contrast-enhanced T1 (T1c); with a voxel size of $(1 \times 1 \times 3)$, $(1 \times 1 \times 3)$ and $(0.5 \times 0.5 \times 1)$ mm respectively. The CT scans have a voxel size of $(0.5 \times 0.5 \times 3)$ mm. As part of the treatment planning, the MR-defined GTV and several organs-at-risk (such as hippocampus, brainstem, eyeballs and optic nerve) have been manually delineated on the transversal slices of the CT scan of each patient, with the MR-sequences transformed and resampled to the CT scan.

In our method, we use the three MR-sequences to automatically segment the GTV, hippocampus and brainstem. Note that our method do not currently segment eyeballs and the optic nerve, as they are not part of the healthy atlas we are using. These structures will be included in the future. As the only pre-processing step before running the algorithm, we co-register the MR scans and resample them to 1 mm isotropic resolution. As a post-processing step, we transform and resample the resulting segmentations to the CT scan, before comparing with the manual segmentations.

To learn the parameters of the cRBM model, we used the expert segmentations of the training data for the MICCAI Brain Tumor Segmentation challenge of 2013, consisting of edema and tumor segmentations of 20 high-grade gliomas and 10 low-grade gliomas [1]. As this training data set is small, we augmented it by flipping the segmentations in 8 different directions.

3.1 Implementation

For the initial model parameter optimization method, we closely follow the implementation details of the whole-brain segmentation method described in [5]. For the sampling and the training of the cRBM model, we closely follow the implementation details in [8]. We used 40 filters of size $(7 \times 7 \times 7)$ for each cRBM, corresponding to 40 hidden groups. Furthermore, to reduce the number of parameters to be estimated, we let each element in an cRBM filter model two neighboring elements in $\mathbf{y}$ or $\mathbf{z}$, i.e., a filter of size 7 will span over 14 visible units.

For the GMMs connected to the different structures, we used the following number of Gaussians: three for non-brain tissues, two for cerebrospinal fluid, one for gray matter structures (which includes hippocampus), one for white matter structures (which includes brainstem), one for tumor and one for edema. The healthy GMM parameters were initialized based on the structure probabilities given by the atlas prior model after affine registration to the subject. We initialized the mean vectors of the tumor and edema Gaussians to the percentiles {90,70,95} and {90,70,50} respectively, in FLAIR, T2 and T1c. Before starting the MCMC sampler, we added two more Gaussian for tumor. We initialized the Gaussians by randomly setting $y_i = 1$ to a fraction of the voxels with $z_i = 1$ and $y_i = 0$ in the segmentation used to initialize the sampler, as detailed in [8].

We implemented the algorithm in MATLAB 2015a, except for the atlas mesh deformation part which was implemented in C++. The full segmentation algorithm (including both the initial parameter estimation and the subsequent sampling) was performed on a Core i5-2400 3.1 GHz CPU, taking roughly 2 hours per subject. After a burn-in period of 100 samples, we generated 100 samples and obtained the final segmentation by majority voting on these 100 samples.

3.2 Results

Figure 1 shows four exemplary segmentations by the method. Although the method can segment many more structures [5], we focus here on the structures that we can validate against the manual segmentations and that are important for radiotherapy, i.e., brainstem, hippocampus and GTV. As can be seen, the atlas is able to deform well to fit the subjects. The method also captures GTV well although the segmentations are slightly underestimated compared to the manual segmentations. Furthermore, brainstem and hippocampus are captured well, although the hippocampus is slightly larger than in the manual segmentations.

Figure 2 shows a brainstem segmentation by the method from a sagittal view compared to the corresponding manual segmentation. The method captures brainstem well, but as can be seen from this view a part of the midbrain is missing in the method's segmentation. This is due to the different protocols used by the experts at the clinic and for the healthy subjects used to construct the atlas, where this part of the midbrain has been labeled as ventral diencephalon.

Figure 3 shows three hippocampus segmentation by the method from a sagittal view compared to the corresponding manual segmentations. Note that the manual segmentation appear blocky as they are delineated on the transversal slices of the CT scans. In the two first examples, the hippocampus is well captured by the method. In contrast, the manual segmentations are missing parts of the hippocampus. It is clear that the protocol used for the manual segmentations is quite conservative. In the last example however, the method slightly overestimate the hippocampus.

In figure 4, we show quantitative results evaluating the spatial overlap in the structures brainstem, hippocampus and GTV; for the 20 test subjects. In order to compare our method to the manual segmentations, we use Dice score $\mathrm{D} = \frac{\mathrm{TP}}{(\mathrm{FN+FP+2\,TP})/2}$, true positive rate $\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP+FN}}$ and positive prediction value $\mathrm{PPV} = \frac{\mathrm{TP}}{\mathrm{TP+FP}}$ as performance metrics. Here, TP, FP and FN count the true positive, false positive and false negative voxels compared to the manual segmentations. For all metrics, a perfect overlap would give a score of 1 and no overlap a score of 0. From left to right, the columns in each plot shows the scores for brainstem, hippocampus and GTV, respectively. The scores for brainstem are the most consistent, with a fairly high average Dice score. The trend of the TPR and PPV shows that brainstem is clearly underestimated compared to the manual segmentations, which is mainly due to the different protocols. Hippocampus, on the other hand, has a low Dice score as in general it is overestimated compared to the manual segmentations. This is mainly due to the conservative protocol of the manual segmenters, but the hippocampus is also sometimes overestimated by the method. It should be noted that hippocampus is a small structure, with an average volume which is 10 times less than the other two

structures in this study. This makes the scores more sensitive to segmentation differences. For most subjects, GTV gets fairly well segmented, but fails for a few subjects which drags down the average scores considerably.

Figure 5 shows two failed segmentation. In the first subject, the method has difficulties to include a resection with a similar intensity distribution as cerebrospinal fluid. This issue affected the segmentations of three subjects in the data set. In the second subject in the figure, the method captures an old infarct in another region of the brain and adapts to its intensity distribution, which has the unfortunate effect that the enhanced ring of the tumor gets labeled as peritumoral edema. In one other subject, the method segmented the GTV fairly well but also segmented a tumor that was not to be included for radiotherapy.

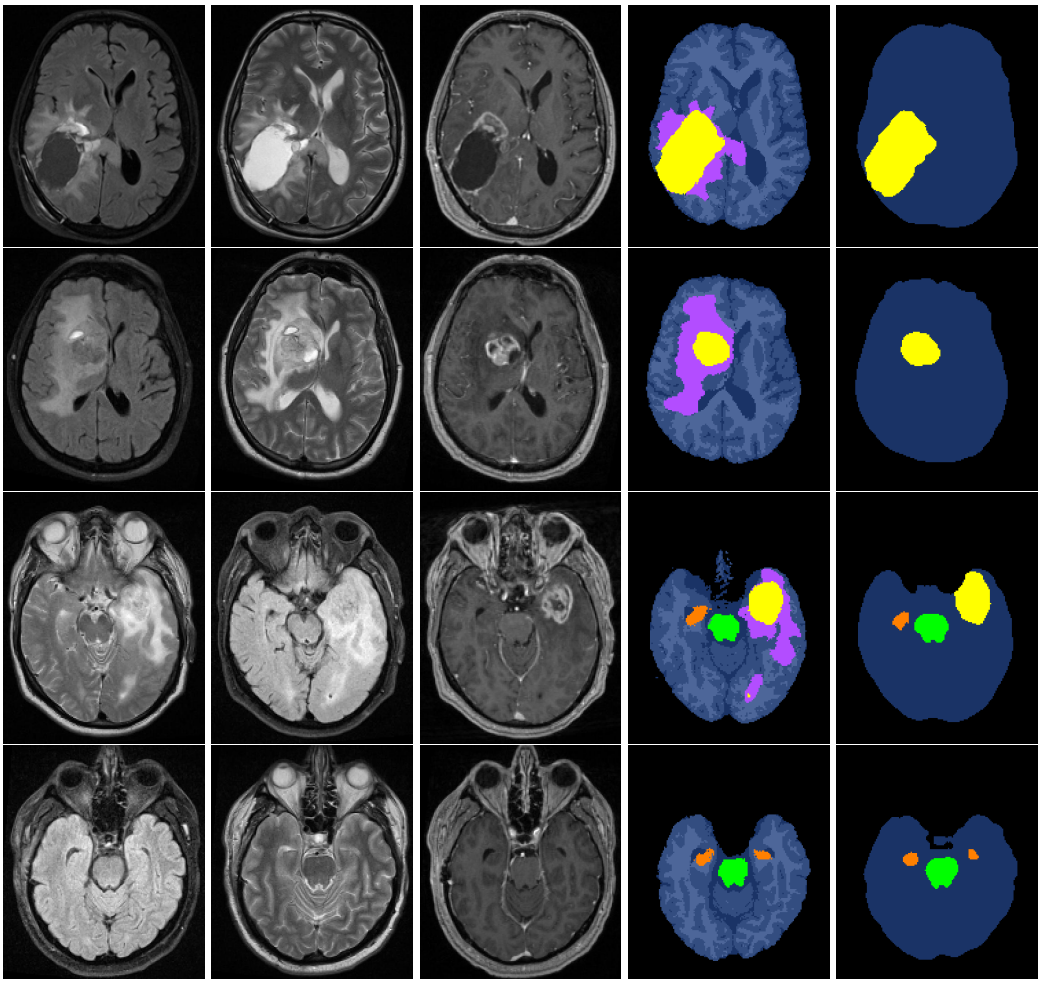


Figure 1. Slices of four exemplary subjects. From left to right: MR-contrasts: FLAIR, T2 and T1c; segmentation by method; and manual segmentation. GTV is in yellow, brainstem is in green, hippocampus is in orange, edema is in lilac, non-brain tissue is in black and healthy superstructures are in shades of blue.

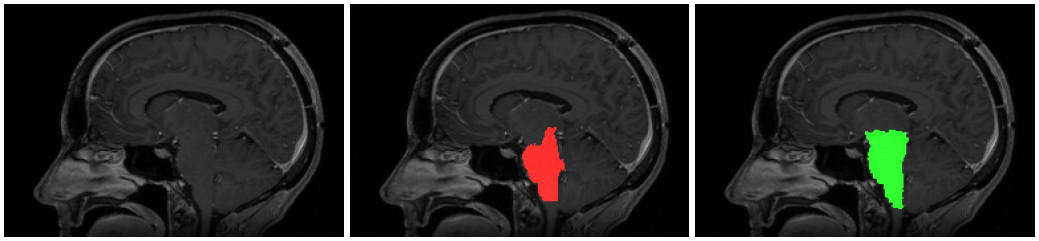


Figure 2. Sagittal view of brainstem. From left to right: T1c, segmentation by the proposed method, and manual segmentation.

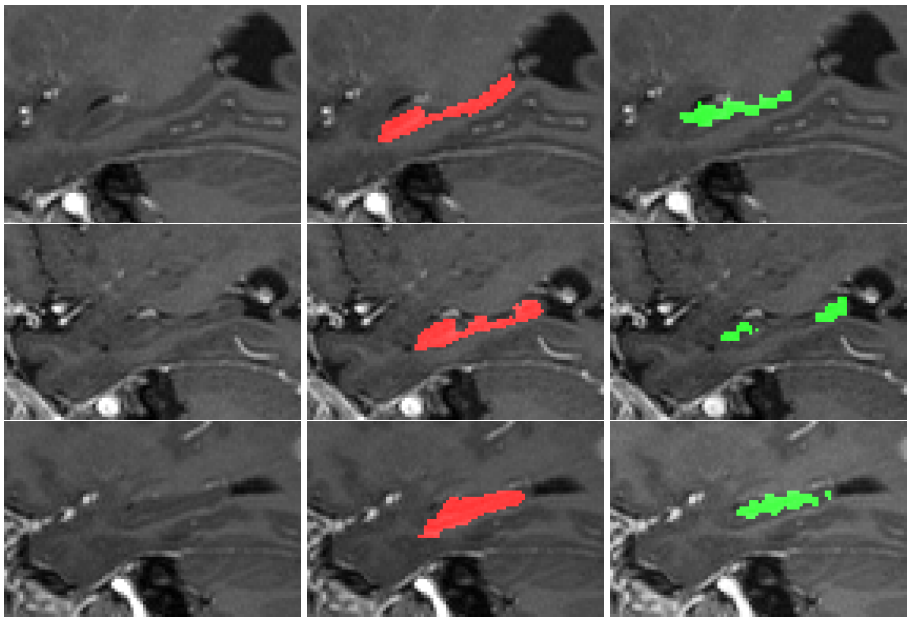


Figure 3. Sagittal view of hippocampus in three subjects. From left to right: T1c, segmentation by the proposed method, and manual segmentation.

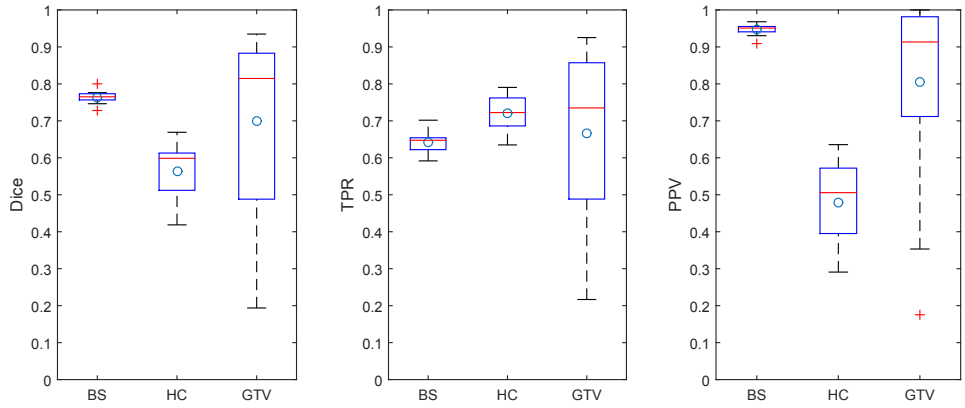


Figure 4. Box plots showing Dice scores, TPR and PPV of brainstem (BS), hippocampus (HC) and GTV. Circles show mean values, central lines show medians, edges of boxes show the 25th and 75th percentiles, and outliers are marked with '+'.

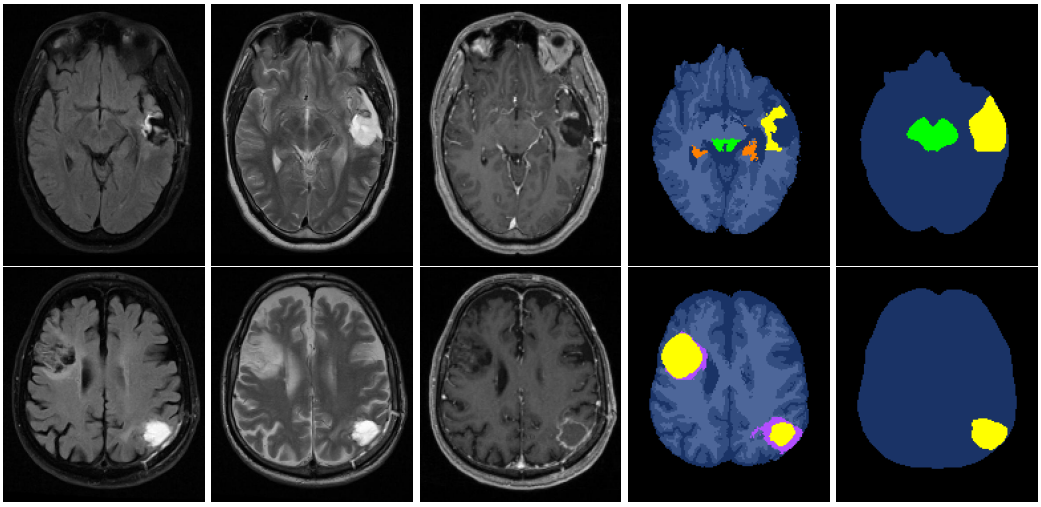


Figure 5. Slices of two failed subject. From left to right: MR-contrasts: FLAIR, T2 and T1c; segmentation by method; and manual segmentation. GTV is in yellow, brainstem is in green, hippocampus is in orange, edema is in lilac, non-brain tissue is in black and healthy superstructures are in shades of blue.

4. DISCUSSION

In this paper, we have proposed a fully automated generative method for simultaneous brain tumor and organs-at-risk segmentation for radiotherapy planning. The method combines an existing whole-brain segmentation method with a tumor prior that uses convolutional restricted Boltzmann machines to model tumor shape. It is not tied to any specific imaging protocol as the parameters of the tumor model are estimated only from expert *segmentations* of annotated training images, without using intensity information. To initially test the feasibility of the method, we have evaluated its performance on a manually delineated clinical data set of 20

glioblastoma patients. We have reported the segmentation accuracy on gross tumor volume and two organs at risk in radiotherapy, i.e., brainstem and hippocampus; with varying but promising results. The evaluation revealed some misalignment between the protocols used by the segmentation experts of the test data set and the protocols used when constructing the healthy atlas. Future work will involve further experiments to improve the tumor prior and use more informed priors on the GMM parameters; inclusion of other important organs at risk in the atlas, such as eyeballs and the optic nerve; and validation on a larger data set.

ACKNOWLEDGMENTS

This research was supported by NIH NCRR (P41-RR14075), NIBIB (R01EB013565) and the Lundbeck Foundation (R141-2013-13117).

REFERENCES

[1] Menze, B. H., et al.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2015)

[2] Tustison, N., Wintermark, M., Durst, C., and Avants, B.: ANTs and Árboles. In: Proc MICCAI-BRATS 2013 (2013)

[3] Havaei, M., et al.: Brain Tumor Segmentation with Deep Neural Networks. arXiv preprint arXiv:1505.03540 (2015)

[4] Fischl, B.: FreeSurfer. Neuroimage, 62(2), 774-781 (2012)

[5] Puonti, O., Iglesias, J. E., and Van Leemput, K.: Fast, Sequence Adaptive Parcellation of Brain MR Using Parametric Models. In: MICCAI 2013. Lecture Notes in Computer Science, vol. 8149, 727-734 (2013)

[6] Kwon, D., Shinohara, R. T., Akbari, H., and Davatzikos, C.: Combining generative models for multifocal glioma segmentation and registration. In: MICCAI 2014. Lecture Notes in Computer Science, vol. 8673, 763-770 (2014)

[7] Bauer, S., Lu, H., May, C. P., Nolte, L. P., Büchler, P., and Reyes, M.: Integrated segmentation of brain tumor images for radiotherapy and neurosurgery. International Journal of Imaging Systems and Technology, 23(1), 59-63 (2013)

[8] Agn, M., Puonti, O., Law, I., Munck af Rosenschöld, P., and Van Leemput, K.: Brain Tumor Segmentation Using a Generative Model with an RBM Prior on Tumor Shape. To appear in: Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries - Brainles workshop at MICCAI 2015, Revised Selected Papers. Lecture Notes in Computer Science, vol. 9556 (2016) (pre-print: http://people.compute.dtu.dk/miag/brainles-agn.pdf)

[9] Van Leemput, K.: Encoding Probabilistic Brain Atlases Using Bayesian Inference. IEEE Transactions on Medical Imaging 28(6), 822-837 (2009)

[10] Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y.: Unsupervised learning of hierarchical representations with convolutional deep belief networks. Communications of the ACM 54(10), 95-103 (2011)

[11] Fischer, A., and Igel, C.: Training restricted Boltzmann machines: An introduction. Pattern Recognition 47(1), 25-39 (2014)

[12] Melchior, J., Fischer, A., and Wiskott, L.: How to Center Binary Restricted Boltzmann Machines. arXiv preprint arXiv:1311.1354 (2013)

[13] Cho, K., Raiko, T., and Ilin, A.: Enhanced gradient for training restricted Boltzmann machines. Neural computation 25(3), 805-831 (2013)

Chapter 9

Paper C

Simultaneous Segmentation of Tumor and Organs-at-Risk for Radiation Therapy Planning of Glioblastomas using a Sequence-Adaptive Method

Mikael Agn*, Per Munck af Rosenschöld, Laura Mancini, Anastasia Papadaki, Steffi Thust, John Ashburner, Ian Law, Koen Van Leemput

***Abstract*—When planning a radiation therapy session to treat a brain tumor patient, the tumor target need to be segmented and also so-called organs-at-risk, which are healthy structures that should be spared from radiation. In this paper, we present a fully automated generative method that can simultaneously segment both types of structures. The method extends an existing generative whole-brain segmentation method to also handle organs-at-risk outside of the brain, and combines this method with a spatial tumor shape model. The method uses a mesh-based probabilistic atlas modeling the organization of healthy structures coupled with an intensity model, modeling the image intensity distribution together with imaging artifacts. The shape of tumors are modeled by convolutional restricted Boltzmann machines. The method is capable of adapting to varying image sequences and protocols. We validate the method on a clinical dataset with multi-contrast MR data as well as on a public benchmark dataset of glioma patients from several imaging centers. In addition, we demonstrate the method's flexibility on a small dataset with a differing set of MR contrasts. Our experiments indicate that the segmentation of brain tumor compares well to the state-of-the-art and that we achieve a good performance on brainstem, hippocampi and eyes; while the performance on optic nerves and chiasm was generally lower.**

I. Introduction

GLIOBLASTOMAS are the most common type of tumors originating within the brain [1]. Current treatment practice involves a combination of surgical resection, conmittant and adjuvant chemotherapy, and radiation therapy [2]. During radiation therapy, the patient is typically subjected to radiation beams from different directions and with different intensity profiles, with the aim of maximizing the delivered radiation dose to the targeted tumor while minimizing the dose to sensitive healthy structures [3]. When planning the radiation therapy session, many structures therefore need to be delineated on medical images of the patient's head – most importantly on multi-contrast Magnetic Resonance (MR) scans [4]. The MR-defined gross tumor volume (GTV) need to be delineated and also so-called organs-at-risk (OARs). OARs are healthy structures that need to be spared from radiation. The specific OARs used vary between clinics; in this paper we will focus on optic nerves, optic chiasm, eyes, brainstem and hippocampi [5]. In clinical practice, the delineation is performed manually with limited assistance from automatic procedures. However, manual delineation is time consuming and typically suffers from poor reproducibility [6], [7], [8]. Therefore, there is a need for automated segmentation methods that can segment both brain tumors and healthy structures in brain tumor patients.

To automatically segment medical images of brain tumor patients is difficult for two main reasons. The first reason is that tumors in themselves are difficult to capture as they vary greatly in size, shape, appearance and location within the brain; and additionally they might have been partly removed by surgery in a previous treatment stage. The second reason is that the healthy structures surrounding a tumor are pushed and deformed by the growth of the tumor (so-called mass effect). Furthermore, patients suffering from brain tumors fairly frequently have other abnormalities such as enlarged ventricles due to their age, making the segmentation task challenging. Moreover, MR scans often exhibit significant intensity inhomogeneity – so-called bias fields – as well as large intensity variations across imaging centers. In addition, the visualization of tumors through medical images is an active research area with potentially interesting imaging protocols and techniques emerging [9].

There is a notable lack in the literature of methods that simultaneously segment brain tumors and organs-at-risk [10]. In contrast, many methods exist with a singular focus on brain tumor segmentation – with some achieving a good although not very robust performance. In particular, the annual Brain Tumor Segmentation (BRATS) challenge held the first time in conjunction with the 2012 MICCAI conference have benchmarked several well-performing methods [8]. Such methods can be divided into two broad categories: discriminative and generative methods.

Discriminative methods typically rely heavily on the intensity information of annotated training data, which is used to

*M. Agn is with the Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark.

Per Munck af Rosenschöld is with the Department of Oncology, Rigshospitalet, 2100 Copenhagen, Denmark.

Laura Mancini, Anastasia Papadaki and Steffi Thust are with the Neuroradiological Academic Unit, Department of Brain Repair and Rehabilitation, UCL Institute of Neurology, University College London, WC1N 3BG London, UK, and also with the Lysholm Department of Neuroradiology, National Hospital for Neurology and Neurosurgery, UCLH NHS Foundation Trust, WC1N 3BG London, UK.

John Ashburner is with the Wellcome Trust Centre for Neuroimaging, UCL Institute of Neurology, University College London, WC1N 3BG London, UK.

Ian Law is with the Department of Clinical Physiology, Nuclear Medicine and PET, Rigshospitalet, 2100 Copenhagen, Denmark.

K. Van Leemput is with the Athinoula A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School, Charlestown, MA 02129, USA, and also with the Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Lyngby, Denmark.

directly learn how to discriminate between the appearance of tumor and that of other tissue. To incorporate spatial context, different strategies are employed. Patch-based methods make use of annotated image patches from training data that are compared directly with the images to be segmented, e.g., [11]. Many methods rely on user-engineered image features that are fed into a classifier, such as random forests, e.g., [12], [13], [14], [15], or support vector machines (SVMs) [16]. Recently, convolutional neural networks (CNNs) have successfully been employed for brain tumor segmentation, e.g., [17], [18], [19], [20]. The key to their success is their capability of automatically learning image features from training data. CNNs learn these features in a deep hierarchy that can capture complex interactions in the data. Because discriminative methods explicitly depend on training data intensity information, they are limited to the imaging protocols and modalities included in the training data. Even within one imaging protocol, careful preprocessing is crucial – such as bias field correction, skull-stripping and intensity normalization. Additionally, these type of methods often need a large amount of annotated training data. To train a discriminative method to handle a new image contrast, additional training data typically have to be acquired that include the new contrast.

Generative methods, on the other hand, aim to build a cohesive model of the formation process of the data to be segmented. Instead of directly relying on the intensity information of training data, a generative model typically incorporates detailed prior knowledge of behaviors of brain structures and imaging artifacts coupled with an adaptive model on the distribution of intensities. Therefore, these methods are more flexible in adapting to varying situations than discriminative methods, such as unseen images. An inconvenience with generative methods however is the difficulty to encode detailed human prior knowledge into appropriate probabilistic models. For brain tumor segmentation, generative methods have generally been atlas-based, wherein a probabilistic atlas encodes the spatial organization of healthy brain structures. The difficulty is then to incorporate prior information about tumor tissue into the model. Tumor tissue has been modeled as outliers in the intensity model, e.g., in [21], or with more involved models on tumor growth, e.g., in [22], [23]. To improve performance, these methods have also been used with a subsequent discriminative step in [24], [25].

Various methods exist for detailed whole-brain segmentation with excellent performance on reasonably healthy brains – such as [26], [27], [28], which all include hippocampus and brainstem segmentation. However, they are typically not designed to handle significant abnormalities in the brain. As described in the previous paragraph, generative atlas-based methods have been used for the single purpose of brain tumor segmentation, while the employed atlas also incorporates some information about healthy tissue. Some studies have investigated how to take tumors into account when deforming atlases, for the purpose of healthy tissue segmentation [29], [30], [31]. However, this has not yet been specifically explored for OARs. An atlas-based method for segmenting OARs but not tumor has been evaluated on a limited dataset of tumor patients [32], with good performance on brainstem and eyes but lower performance on optic nerves and chiasm. A few discriminative methods specifically focused on brainstem segmentation have been evaluated on brains with tumors, showing good performance on a limited dataset, by using SVMs [10] and denoising autoencoders [33].

Segmentation of the optic system has generally received little attention because optic nerves and eyes are outside the brain, but a few methods have been developed for use in radiation therapy planning. Some focus solely on optic nerve segmentation by using e.g., SVMs [34] or multi-atlas approaches [35]. Geometric-driven methods have been successful in capturing the specific nature of the optic system (i.e., optic nerves have a tubular structure with a crossing at the chiasm) such as in [36], and also including eyes in [7], [37].

In this paper, we present a fully automated generative method for application in radiation therapy planning that is capable of simultaneously segmenting brain tumors and OARs. To the best of our knowledge, this is the first time a segmentation method has been presented that encompass both types of structures. The method builds on prior work from our group in whole-brain segmentation, which has been thoroughly validated on healthy brains in [38]. The method uses a mesh-based probabilistic atlas for detailed segmentation of healthy brain structures, including hippocampi and brainstem. This atlas is coupled with a Gaussian mixture model for modeling the distribution of image intensities together with a model for bias field artifacts. It was first presented in [39] and also used in [40], [41] for hippocampus and brainstem substructure segmentation. Because the model of brain anatomy is separated from the model of image intensity, the method can adapt to the intensity distribution of the images to be segmented. Thus, varying types of images can easily be included and no preprocessing is needed such as intensity normalization, bias field correction or skull-stripping.

To handle tumor segmentation in this paper, we extend this method by incorporating a model on tumor shape based on convolutional Restricted Boltzmann Machines (cRBMs). A cRBM is a *generative* neural network and thus, similar to CNNs, it can automatically learn features from training data. However, we learn features on the tumor segmentations from annotated training data, without using the corresponding *intensity* information. Early works of this model appeared in [42], [43]. In [42], we proved that cRBMs are suitable for brain tumor segmentation. As a proof-of-concept demonstration in [43], we demonstrated the feasibility of combining the whole-brain segmentation framework with the tumor shape model. Here, we extend the method to include the OARs optic nerves, optic chiasm and eyes; and update the intensity model to a more robust version. Furthermore, we add a detailed description of our method and validate it on more subjects from different imaging centers and with varying MR image contrasts.

II. Modeling framework

We aim to automatically partition a multi-contrast MR scan of a patient's head into K healthy tissue labels and two tumor-affected tissue labels. To reach this aim, we build a generative

model that describes the image formation process of an MR scan and then use this model to obtain a fully automated segmentation algorithm. To build the model, we combine a previously validated atlas-based approach for whole-brain segmentation [38] with a spatial prior on tumor shape.

We denote the MR scan $\mathbf{D} = (\mathbf{d}_i, ..., \mathbf{d}_I)$, where I is the number of voxels in the scan. The column vector $\mathbf{d}_i \in \mathbb{R}^N$ contains the log-transformed intensities at voxel i for N MR contrasts.

We call the two tumor-affected tissue labels core and edema. The core label corresponds to the MR-defined gross tumor volume (GTV) used in radiotherapy, while the edema label corresponds to the visible peritumoral edema surrounding the core. To capture these labels, we introduce two binary maps: $\mathbf{z} = (z_i, ..., z_I)^T$ and $\mathbf{y} = (y_i, ..., y_I)^T$. The $\mathbf{z}$ map indicates all tumor-affected voxels, and the $\mathbf{y}$ map indicates voxels inside the core. An edema voxel i will thus be indicated by $\{z_i = 1, y_i = 0\}$ and a core voxel i by $\{z_i = 1, y_i = 1\}$.

We represent the healthy segmentation with the vector $\mathbf{l} = (l_i, ..., l_I)^T$, where $l_i \in \{1, ..., K\}$. A voxel i with a label l_i will change its status to tumor when $z_i = 1$. As glioblastoma is an intracranial disease, we restrict this status change to labels located inside the brain.

The generative model consist of a prior and a likelihood. The prior is a probability distribution over the labels $p(\mathbf{l}, \mathbf{y}, \mathbf{z})$ combining two spatial distributions: one over the healthy tissue labels and one over the tumor tissue labels. The likelihood function $p(\mathbf{D}|\mathbf{l}, \mathbf{y}, \mathbf{z})$ connects the labels to the intensities in the MR scan. We will first describe the healthy tissue prior distribution in section II-A, then the tumor tissue prior distribution in section II-B and how the two are combined to form the full prior in section II-C. The likelihood is described in section II-D, followed by a description of the inference of the model in section II-E and implementation details in II-I.

A. Atlas-based prior on healthy tissue

To model the spatial configuration of healthy tissue, we use an atlas-based prior that was introduced in [39] and further validated in [38]. The model consist of a deformable mesh-based atlas. Each node in the tetrahedral mesh is associated with a probability vector containing the probabilities of healthy labels to occur at that node.

The probability of a label k to occur at voxel i is given by $\pi_i(k|\boldsymbol{\eta})$, where $\boldsymbol{\eta}$ contains the node positions of a deformed mesh. We obtain the voxel probability by interpolating from the nodes of the tetrahedron in which the voxel lie. Assuming that labels at different voxels are conditionally independent given the node positions, we have

$$p(\mathbf{l}|\boldsymbol{\eta}) = \prod_{i=1}^{I} \pi_i(l_i|\boldsymbol{\eta}).$$

Adding a topology-preserving deformation prior $p(\boldsymbol{\eta})$ (as in [39], [38]) the final healthy tissue prior is

$$p(\mathbf{l}) = \int_{\boldsymbol{\eta}} p(\mathbf{l}|\boldsymbol{\eta}) p(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta}. \quad (1)$$

The probability vectors and spatial configuration of the atlas mesh are automatically learned from manual segmentations obtained from training data, as described in section II-F.

B. Using convolutional RBMs to form a spatial prior on tumor tissue

We use restricted Boltzmann machines (RBMs) to design a spatial prior on the configuration of tumor tissue. An RBM is a Markov random field (MRF) that can model high-order interactions between visible units through local connections to hidden units [44]. We use the convolutional variant of the RBM (cRBM), where the weights of the connections are shared among all locations [45]. This allows us to infer over large images without a predefined size. We combine two separate cRBMs: one that models interactions between the binary units in the tumor-affected map $\mathbf{z}$, and one that models interactions between the binary units in the core map $\mathbf{y}$.

We start by describing the cRBM that models $\mathbf{z}$. A small example of the model can be seen in figure 1. This cRBM is defined by

$$p(\mathbf{z}) = \sum_{\mathbf{H}} p(\mathbf{z}, \mathbf{H})$$

with

$$p(\mathbf{z}, \mathbf{H}) \propto \mathrm{e}^{-E_z(\mathbf{z}, \mathbf{H})},$$

where $\mathbf{H}$ denotes the hidden units, which are binary, and $E_z(\mathbf{z}, \mathbf{H})$ is an energy term. Note that we present the model in 1D with the sole purpose of avoiding cluttered equations, but it directly generalizes to 3D images. The energy term is defined as

$$E_z(\mathbf{z}, \mathbf{H}) = -\sum_m \mathbf{h}_m \bullet (\mathbf{w}_m * \mathbf{z}) - \sum_k b_{zm} \sum_j h_{mj} - a_z \sum_i z_i,$$

where $\bullet$ denotes element-wise product followed by summation and $*$ denotes convolution. Each hidden group $\mathbf{h}_m \in \mathbf{H}$ is connected to the visible units in $\mathbf{z}$ with a convolutional filter $\mathbf{w}_m$. The filter models interactions between the hidden and visible units, effectively detecting specific features in $\mathbf{z}$. Furthermore, each hidden group has a bias b_{zm} and visible units have a bias a_z. These bias terms encourage units to be enabled or disabled.

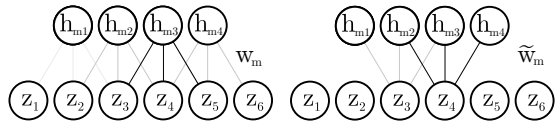


Fig. 1. A small 1D example of a cRBM. Visible units in $\mathbf{z}$ are connected to hidden units in a hidden group $\mathbf{h}_m$ through a convolutional filter $\mathbf{w}_m$ of size 3. The first illustration shows the model from the hidden layer's perspective. The second shows the model from the visible layer's perspective, where $\tilde{\mathbf{w}}_m$ is a mirror-reversed version of the filter. Note that boundary units in the visible layer are set to 0.

The computational appeal of this model is that no direct connections exist between two visible units or two hidden

units, so that the visible units are independent of each other given the state of the hidden ones, and vice versa:

$$p(\mathbf{z}|\mathbf{H}) = \prod_i p(z_i|\mathbf{H}) \quad \text{and} \quad p(\mathbf{H}|\mathbf{z}) = \prod_k \prod_j p(h_{mj}|\mathbf{z})$$

with

$$p(z_i = 1|\mathbf{H}) = \sigma\Big(\sum_m (\tilde{\mathbf{w}}_m * \mathbf{h}_m)_i + a_z\Big)$$

and

$$p(h_{mj} = 1|\mathbf{z}) = \sigma\Big((\mathbf{w}_m * \mathbf{z})_j + b_{zm}\Big),$$

where $\sigma(t) = 1/(1 + e^{-t})$ and $\tilde{\mathbf{w}}$ denotes a mirror-reversed version of the filter $\mathbf{w}$. This property enables efficient Markov chain Monte Carlo (MCMC) sampling of the model using a blocked Gibbs sampler.

For the tumor core map $\mathbf{y}$, we use a similar cRBM model:

$$p(\mathbf{y}) = \sum_{\mathbf{G}} p(\mathbf{y}, \mathbf{G}), \quad p(\mathbf{y}, \mathbf{G}) \propto \mathrm{e}^{-E_y(\mathbf{y},\mathbf{G})}$$

with hidden units $\mathbf{G}$ and an energy $E_y(\mathbf{y}, \mathbf{G})$ that depends on convolutional filters $\mathbf{u}_m$, a visible bias a_y and hidden biases b_{ym}.

As we detail in section II-H, the cRBM filter weights and biases can be learned automatically from a set of manually labeled tumor maps. '

C. A joined healthy and tumor tissue prior

To form the prior on all labels, we incorporate the two tumor tissue priors defined in section II-B into the healthy tissue prior in section II-A. As we are interested specifically in brain tumors, we assume that tumor tissue only will occur inside the brain. Furthermore, we restrict core tissue to only occur inside the tumor-affected region $\mathbf{z}$. Specifically, we define the following distribution

$$p(\mathbf{y}, \mathbf{z}, \mathbf{l}, \mathbf{G}, \mathbf{H}|\boldsymbol{\eta}) \propto \mathrm{e}^{-E(\mathbf{y},\mathbf{z},\mathbf{l},\mathbf{G},\mathbf{H}|\boldsymbol{\eta})}$$

with energy

$$\begin{aligned} E(\mathbf{y}, \mathbf{z}, \mathbf{l}, \mathbf{G}, \mathbf{H}|\boldsymbol{\eta}) &= E_z(\mathbf{z}, \mathbf{H}) + E_y(\mathbf{y}, \mathbf{G}) \\ &\quad - \sum_i \log \pi_i(l_i) + \sum_i f(y_i, z_i, l_i), \end{aligned}$$

where we have used a restricition function defined as

$$f(y, z, l) = \begin{cases} \infty & \text{if } z = 0 \text{ and } y = 1 \\ \infty & \text{if } z = 1 \text{ and } l \notin B \\ 0 & \text{otherwise,} \end{cases} \tag{2}$$

where B denotes the collection of healthy labels that are located inside the brain (see table I for a precise definition of healthy labels we categorize as being inside the brain). This function encodes that a core voxel can never appear outside the tumor-affected region $\mathbf{z}$, and that a tumor-affected voxel can never appear outside the brain.

The resulting joint prior on the segmentation labels $\mathbf{y}$, $\mathbf{z}$ and $\mathbf{l}$ is obtained as

$$p(\mathbf{y}, \mathbf{z}, \mathbf{l}) = \int_{\boldsymbol{\eta}} \sum_{\mathbf{G}} \sum_{\mathbf{H}} p(\mathbf{y}, \mathbf{z}, \mathbf{l}, \mathbf{G}, \mathbf{H}|\boldsymbol{\eta}) p(\boldsymbol{\eta}) \mathrm{d}\boldsymbol{\eta}.$$

Note that without the restriction function $f(y, z, l)$, this prior would simply devolve into $p(\mathbf{l}, \mathbf{z}, \mathbf{y}) = p(\mathbf{l})p(\mathbf{z})p(\mathbf{y})$.

D. Likelihood

The likelihood $p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y})$ links the labels in the model to MR intensities. As in [38], each label is associated with a Gaussian mixture model (GMMs) that models the intensity distribution of that label. MR scans are typically corrupted by a multiplicative low-frequency imaging artifact referred to as a bias field. By log-transforming the MR intensities, we can model the bias field as a linear combination of spatially smooth basis functions added to the log-transformed intensities.

Let x denote a label indicating which of $K + 2$ possible segmentation assignments are given to any given voxel: one of the K healthy tissues (when $z = 0$), edema ($z = 1$ and $y = 0$), or core ($y = 1$). (Note that the combination $(z = 0, y = 1)$ will never occur due to the restriction function of Eq. (2) in the prior.) We define a local Gaussian mixture model at voxel i as

$$p_i(\mathbf{d}_i|\boldsymbol{\theta}_x, \mathbf{C}) = \sum_{g=1}^{G_x} \gamma_{xg} \mathcal{N}(\mathbf{d}_i|\boldsymbol{\mu}_{xg} + \mathbf{C}\boldsymbol{\phi}_i, \boldsymbol{\Sigma}_{xg}),$$

where

$$\mathcal{N}(\mathbf{d}|\boldsymbol{\mu}, \boldsymbol{\Sigma}) = \frac{1}{\sqrt{(2\pi)^N |\boldsymbol{\Sigma}|}} e^{-\frac{1}{2}(\mathbf{d}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{d}-\boldsymbol{\mu})}$$

denotes a multivariate normal distribution with mean $\boldsymbol{\mu}$ and covariance $\boldsymbol{\Sigma}$; G_x is the number of Gaussian components connected to label x; and γ_{xg}, $\boldsymbol{\mu}_{xg}$ and $\boldsymbol{\Sigma}_{xg}$ are the weight, mean and covariance matrix of component g, jointly collected in the parameter set $\boldsymbol{\theta}_x = \{\gamma_{xg}, \boldsymbol{\mu}_{xg}, \boldsymbol{\Sigma}_{xg}, \forall g\}$. The weights satisfy the constraints $\gamma_{xg} \geq 0$ and $\sum_g^{G_x} \gamma_{xg} = 1$. Furthermore, the column vector $\boldsymbol{\phi}_i \in \mathbb{R}^P$ evaluates P bias field basis functions at voxel i and $\mathbf{C} = (\mathbf{c}_1, ..., \mathbf{c}_N)^T$ denotes the parameters of the bias field model, where $\mathbf{c}_n \in \mathbb{R}^P$ are the parameters of the bias field model for MR contrast n.

Letting $\boldsymbol{\theta} = \{\{\boldsymbol{\theta}_x, \forall x\}, \mathbf{C}\}$ collect all the bias field parameters and the mixture model parameters of all segmentation labels, we model the (log-transformed) MR intensities in all voxels to be conditionally independent given the segmentation labels:

$$p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta}) = \prod_i p_i(\mathbf{d}_i|l_i, z_i, y_i, \boldsymbol{\theta})$$

with

$$\begin{aligned} p_i(\mathbf{d}_i|l_i, z_i, y_i, \boldsymbol{\theta}) &= p_i(\mathbf{d}_i|\boldsymbol{\theta}_{l_i}, \mathbf{C})^{(1-z_i)} \\ &\quad p_i(\mathbf{d}_i|\boldsymbol{\theta}_e, \mathbf{C})^{z_i(1-y_i)} \\ &\quad p_i(\mathbf{d}_i|\boldsymbol{\theta}_c, \mathbf{C})^{y_i}, \end{aligned}$$

where e denotes the edema label, c denotes the core label and l_i is any of the healthy labels.

We use a restricted conjugate prior $p(\boldsymbol{\theta})$ on the parameters of the likelihood model:

$$p(\boldsymbol{\theta}) \propto \begin{cases} \prod_x \Big[\mathrm{Dir}(\boldsymbol{\gamma}_x|\alpha_0) \prod_{g=1}^{G_x} \mathrm{NIW}(\boldsymbol{\mu}_{xg}, \boldsymbol{\Sigma}_{xg}|\mathbf{m}_0, \kappa_0, \upsilon_0, \mathbf{S}_0)\Big] \\ \quad \text{if certain constraints on } \{\boldsymbol{\mu}_{xg}\} \text{ are satisfied} \\ 0 \quad \text{otherwise,} \end{cases} \tag{3}$$

where we have used a uniform prior on the bias field parameters $\mathbf{C}$ and conjugate priors on the mixture parameters, following the definitions in [46]: symmetric Dirichlet distributions $\mathrm{Dir}(\boldsymbol{\gamma}|\alpha_0)$ that discourage removal of components if $\alpha_0 > 0$, and normal-inverse-Wishart distributions $\mathrm{NIW}(\boldsymbol{\mu}, \boldsymbol{\Sigma}|\mathbf{m}_0, \kappa_0, \upsilon_0, \mathbf{S}_0)$ where the hyperparameters can be interpreted as prior mean $\mathbf{m}_0$ with strength κ_0 and prior scatter matrix $\mathbf{S}_0$ with strength υ_0. We add certain constraints on $\{\boldsymbol{\mu}_{xg}\}$ to encode prior knowledge about overall tumor appearance relative to normal brain tissue in typical MR sequences for brain tumor imaging. These constraints and the setting of the hyperparameters will be specified in section II-G.

The final likelihood function is given by

$$p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}) = \int_{\boldsymbol{\theta}} p(\mathbf{D}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \boldsymbol{\theta}) p(\boldsymbol{\theta}) \mathrm{d}\boldsymbol{\theta}.$$

E. Inference

To obtain a segmentation of an MR-scan using the described model we would like to evaluate the posterior $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D})$. However, exact inference of this segmentation posterior is intractable because it marginalizes over all of the uncertainty in the model parameters and the hidden units of the cRBM models. We therefore resort to Markov chain Monte Carlo (MCMC) techniques to sample from all unknown variables (except the atlas deformation parameters $\boldsymbol{\eta}$, as detailed below), followed by voxel-wise majority voting on the segmentation samples to obtain the final segmentation. Although it is possible to also sample from $\boldsymbol{\eta}$, as shown in [47], this is considerably more difficult to implement and was not attempted in this paper. Instead, we first estimate a reasonable estimate of good atlas deformation parameters $\hat{\boldsymbol{\eta}}$ using a simplified model, and subsequently estimate segmentations from the conditional posterior $p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \hat{\boldsymbol{\eta}})$ in which the uncertainty on deformations is removed, yielding the following two-step inference algorithm:

1) Step 1: For the purpose of estimating good atlas deformation parameters $\hat{\boldsymbol{\eta}}$, we use a simplified model in which the non-local dependencies between the voxels introduced by the cRMB shape models is removed. In particular, we set the filter weights $\mathbf{w}_m$ and $\mathbf{u}_m$ to zero values, effectively removing the hidden units from the model, and use log-odds visual bias values of $a_z = \log(w/(1-w))$ and $a_y = \log(v/(1-v))$, yielding an overall model in which a user-specified fraction w of healthy brain voxels is expected to be tumorous, and within these voxels a fraction v is expected to be tumor core. We used values of $w = 0.1$ and $v = 0.5$ throughout this paper. This reduces the model to the same form as in the one used in [38], and we can therefore use the same approach for optimization, i.e., by alternating between optimizing the atlas deformation parameters $\boldsymbol{\eta}$ with a conjugate gradient algorithm [48] and optimizing the likelihood parameters $\boldsymbol{\theta}$ with a generalized expectation-maximization (GEM) algorithm [49]. Further details on this algorithm are provided in appendix B.

After convergence, we record $\hat{\boldsymbol{\eta}}$ and compute the *maximum a posteriori* segmentation

$$\begin{aligned}\{\hat{\mathbf{l}}, \hat{\mathbf{z}}, \hat{\mathbf{y}}\} &= \arg\max_{\mathbf{l},\mathbf{z},\mathbf{y}} p(\mathbf{l}, \mathbf{y}, \mathbf{z}|\mathbf{D}, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}) \\ &= \arg\max_{\{l_i, z_i, y_i\}} \prod_i p(l_i, z_i, y_i|\mathbf{d}_i, \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}),\end{aligned}$$

which are used in the subsequent second step of inference.

2) Step 2: In the second inference step, we clamp the atlas deformation parameters to the values $\hat{\boldsymbol{\eta}}$ estimated during the first step, and generate samples $\{\mathbf{l}^{(s)}, \mathbf{z}^{(s)}, \mathbf{y}^{(s)}\}_{s=1}^{S}$ from $p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D}, \hat{\boldsymbol{\eta}})$ using a Markov chain Monte Carlo sampler initialized at $\{\hat{\mathbf{l}}, \hat{\mathbf{z}}, \hat{\mathbf{y}}\}$. The final segmentation is then obtained by performing voxel-wise majority voting across the collected samples: the i^{th} voxel is declared tumor core if $y_i = 1$ in half of the samples, else the voxel is declared edema if $z_i = 1$ in half of the samples, else the voxel is declared the most frequent healthy label in the samples.

In order to generate the required samples from $p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{D}, \hat{\boldsymbol{\eta}})$, we sample from $p(\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}, \boldsymbol{\theta}|\mathbf{D}, \hat{\boldsymbol{\eta}})$ using a partially collapsed blocked Gibbs sampler, and discard the samples of $\mathbf{G}, \mathbf{H}$ and $\boldsymbol{\theta}$. The sampler iteratively samples each set of variables from its conditional distribution given the other variables; with the exception of $\boldsymbol{\theta}$ this is straightforward to implement as each conditional distributions factorizes over its components:

$$p(\mathbf{H}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{G}, \boldsymbol{\theta}, \mathbf{D}, \hat{\boldsymbol{\eta}}) = \prod_m \prod_j p(h_{mj}|\mathbf{z}),$$

$$p(\mathbf{G}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \boldsymbol{\theta}, \mathbf{D}, \hat{\boldsymbol{\eta}}) = \prod_m \prod_j p(g_{mj}|\mathbf{y}),$$

and

$$p(\mathbf{l}, \mathbf{z}, \mathbf{y}|\mathbf{H}, \mathbf{G}, \boldsymbol{\theta}, \mathbf{D}, \hat{\boldsymbol{\eta}}) = \prod_i p(l_i, z_i, y_i|\mathbf{d}_i, \mathbf{H}, \mathbf{G}, \boldsymbol{\theta}, \hat{\boldsymbol{\eta}})$$

with

$$\begin{aligned}p(l_i, z_i, y_i|\mathbf{d}_i, \mathbf{H}, \mathbf{G}, \boldsymbol{\theta}, \hat{\boldsymbol{\eta}}) \propto\ & p(\mathbf{d}_i|l_i, z_i, y_i, \boldsymbol{\theta}) \pi_i(l_i) \\ & \exp\left[z_i \left(\sum_m (\tilde{\mathbf{w}}_m * \mathbf{h}_m)_i + a_z\right)\right] \\ & \exp\left[y_i \left(\sum_m (\tilde{\mathbf{u}}_m * \mathbf{g}_m)_i + a_y\right)\right] \\ & \exp\left[-f(l_i, z_i, y_i)\right].\end{aligned}$$

Sampling from the remaining conditional distribution

$$p(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{H}, \mathbf{G}, \mathbf{D}, \hat{\boldsymbol{\eta}}) = p(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$$

is more difficult due to interdependencies between the various components of $\boldsymbol{\theta}$, and is detailed in Appendix C.

F. Learning parameters for atlas-based prior on healthy tissue

To learn the parameters of the healthy tissue atlas described section II-A, we used brain segmentations from a dataset consisting of 39 subjects. This is the same training dataset used in [38] and the publicly available software package FreeSurfer [26]. The dataset include segmentations of 39 neuroanatomical structures within the brain, which were obtained using a validated semi-automated protocol developed at the Center for

Morphometric Anlysis (CMA), MGH, Boston [50], [51], [52]. The complete dataset consists of 28 healthy subjects and 11 subjects with varying severity of Alzheimer's disease, with an age range from under 30 years to over 60 years old [27]. The scans were acquired on a 1.5T Siemens Vision scanner with a voxel size of $1 \times 1 \times 1.5$ mm^3 We selected a representative subset of 10 subjects for the atlas building process. This number was chosen as the atlas building process is time consuming and Puonti et al. [38] have shown that adding more subjects does not substantially increase segmentation performance.

As we are specifically interested in structures applicable to radiation therapy, we merged some of the 39 segmented structures into larger labels before building the atlas. We keep the segmentations for the organs-at-risk *hippocampi* (HC), *brainstem* (BS) and *optic chiasm* (CH); and the background label. All other structures were merged into the following labels: cerebrospinal fluid (CSF), remaining white matter (WM) and remaining gray matter (GM). Two important organs-at-risk where not included in the original segmentations, as they are located outside of the brain – namely *optic nerves* (ON) and *eyes* (EB). We therefore performed additional manual delineations for these two extra structures. To provide some context around these structures, we also delineated the muscles and fat in the eye socket into two separate labels. Eyes were further separated into two labels: *eye fluid* describing the fluid and gel inside an eye and *eye tissue* describing the lens and the solid outer layer of an eye.

The parameters of the healthy tissue atlas were learned on these 10 updated segmentations, following the details in [39], [38]. When segmenting a subject, we added an additional label designed to capture healthy tissue that is not specified in the atlas, such as blood vessels. In each mesh node's probability vector, we added this label with a constant prior probability of 0.01 and re-normalized the probability vector to ensure that the values sum to one. Thus, we used $K = 13$ healthy labels in this paper, namely: hippocampi, brainstem, optic chiasm, optic nerves, eye fluid, eye tissue, eye socket fat, eye socket muscles, CSF, WM, GM, unspecified brain tissue and background.

G. Settings for likelihood

The settings in this section are mainly based on initial experiments on the training dataset of the brain tumor segmentation (BRATS) challenge that was held in conjunction with the BrainLes workshop at the 2015 MICCAI conference. It consists of 220 high-grade gliomas and 54 low-grade gliomas of varying types, with publicly available ground truth segmentations of tumor. 30 subjects were manually segmented (20 high-grade, 10 low-grade), while the rest had fused segmentations from highly ranked algorithms from previous challenges. The included MR contrasts are T2-weighted FLAIR (varying 2D acquisition), T2-weighted (2D acquisition), T1-weighted (2D acquisition) and T1-weighted with contrast enhancement (T1c, 3D-acquisition). The scans have been acquired at different centers, with varying signal strength and resolution.

The number of Gaussian components per label is shown in table I, together with the division between brain and non-brain structures. We group labels with similar intensity profiles into superstructures with shared Gaussian components. Note that optic nerves and chiasm are included into the same superstructures as surrounding tissue. This is because they are small and hence severely affected by partial volume effects. Some data we tested our method on are skull-stripped (see section III-A). For these images, non-brain structures are therefore missing. Hence, we include all non-brain labels into a non-brain superstructure when segmenting these subjects.

For tumor core, we use one Gaussian component in the first inference step and introduce two more Gaussian components in step 2. Tumor core can include regions with widely differing intensity distributions such as degraded blood-brain barrier, necrotic tissue and regions with surgically removed tumor. Some of these regions have an intensity distribution similar to edema and some a distribution similar to healthy tissues. They can be identified by their vicinity to each other. However, in the first inference step, the temporary tumor prior is spatially flat – which means that we rely on intensity alone in this step. The easiest region to recognize by intensity is the degraded blood-brain barrier. Therefore, we use just one Gaussian component for tumor core in step 1, specifically targeted at this region. In the second step of the algorithm, we then add two more Gaussian components to better represent other tumor core regions. These are initialized by randomly changing a fraction of edema voxels ($z_i = 1$ and $y_i = 0$) to core voxels ($z_i = 1$ and $y_i = 1$). The fraction is chosen so that the total fraction of core within the whole tumor equals the average fraction in the employed training data.

TABLE I
NUMBER OF GAUSSIAN COMPONENTS PER SUPERSTRUCTURE, FOR DATASETS WITH RAW MR-SCANS AND THE SKULL-STRIPPED DATASET

	Label	Raw MR-scan	Skull-stripped
non-brain	Background	3	2
	Eye socket fat	2	
	Eye socket muscle		
	Optic nerves (ON)	3	
	Eye tissue (part of EB)		
	Eye fluid (part of EB)		
brain	Chiasm	2	2
	Cerebrospinal fluid		
	Grey matter	1	
	Hippocampi (HC)		
	White matter	1	
	Brainstem (BS)		
	Unspecified brain tissue	1	
	Peritumoral edema	1	
	Tumor core (TC)	1 (step 1) / 3 (step 2)	

For all Gaussian components, we set the hyperparameters in $p(\boldsymbol{\theta})$, defined in equation 3, as

$$\mathbf{m}_0 = \frac{\sum_i \mathbf{d_i}}{I}, \quad \kappa_0 = 1, \quad \alpha_0 = 10^{-4} I, \quad \nu_0 = 10^{-3} I,$$

where I is the number of voxels. The prior scatter matrix is

set to

$$\mathbf{S}_0 = \frac{10^{-3}}{S^2}\mathrm{Dg}\Big(\sum_i (\mathbf{d}_i - \mathbf{m}_0)(\mathbf{d}_i - \mathbf{m}_0)^T\Big),$$

where S is the number of superstructures in the model and $\mathrm{Dg}(\square)$ denotes a diagonal matrix that only retains the diagonal elements of matrix $\square$. Because the unspecified brain tissue label should catch any unspecified brain tissue, we use a wider prior scatter matrix for this label, with S set to 1.

The linear constraints used in prior $p(\boldsymbol{\theta})$, defined in equation 3, can be seen in table II. Details on how these constraints are formulated in the model framework is given in appendix A. We set the constraints relating to tumor labels by building statistics of mean values from the training data. We estimated the mean values using our method, but with tumor labels fixed to the ground truth. The constraints relating to the unspecified brain tissue label were set as to ascertain that this label will not interfere with the tumor segmentation.

TABLE II
CONSTRAINTS ON MEAN VALUES OF GAUSSIAN COMPONENTS

Edema (TE)		
$\mu_{\text{FLAIR,TE}}$	$\geq$	$1.04\ \mu_{\text{FLAIR,WM}}$
$\mu_{\text{T2,TE}}$	$\geq$	$\mu_{\text{T2,WM}}$
Core, fist Gaussian component (denoted TC1)		
$\mu_{\text{T1c,TC1}}$	$\geq$	$1.01\ \mu_{\text{T1c,GM}}$
$\mu_{\text{T1c,TC1}}$	$\geq$	$1.01\ \mu_{\text{T1c,WM}}$
$\mu_{\text{FLAIR,TC1}}$	$\geq$	$\mu_{\text{FLAIR,WM}}$
$\mu_{\text{T2,TC1}}$	$\geq$	$\mu_{\text{T2,WM}}$
Unspecified brain tissue (US)		
$\mu_{\text{FLAIR,US}}$	$\leq$	$1.02\ \mu_{\text{FLAIR,WM}}$
$\mu_{\text{T1c,US}}$	$\leq$	$1.01\ \mu_{\text{T1c,GM}}$

H. Learning parameters of tumor shape prior

To learn suitable values for the filters and the biases of the cRBMs described in section II-B, we used the 30 manual tumor segmentations in the BRATS training dataset, defined in section II-G. As the number of subjects is small, we augmented the dataset by flipping the segmentations in 8 different directions, yielding a dataset of 240 tumor segmentations.

The values are learned through stochastic gradient ascent. To efficiently approximate the log-likelihood gradients, we use the contrastive divergence (CD) approximation with one block-Gibbs sampling step [44]. Additionally, we use the so-called enhanced gradient together with the CD approximation which have been shown to improve learning [53], [54].

We used the same settings for both the whole tumor cRBM for $\mathbf{z}$ and the tumore core cRBM for $\mathbf{y}$. The filter size and number of filters were set by initial experiments. Choosing a larger filter size will increase the number of parameters which may result in overfitting, while a smaller filter size might not capture long-range features. Therefrom, we found that by tying neighboring parameters in a filter we can reduce the number of parameters while still capturing long-range features. We tie filter parameters in $(2\times2\times2)$-blocks, effectively treating each block as one parameter. We use 40 filters of size $(14\times14\times14)$ (i.e. $7\times7\times7$ blocks) corresponding to 40 hidden groups. This configuration performed better than other combinations of 20, 30 and 40 filters of sizes between 10 and 18. Each cRBM was trained with 9600 gradient steps of size 0.1. A subset of 10 segmentations was used to compute the gradient at each step.

When segmenting a subject, we found it necessary to individualize the cRBM bias term a_z connected to $\mathbf{z}$ to better represent the tumor to be segmented. This is due to the large size variation of tumors. In step 2 of the inference, we therefore added $\log\left(\frac{p_{zs}(1-p_{zt})}{p_{zt}(1-p_{zs})}\right)$ to a_z. Here, p_{zs} denotes the fraction of tumor within the brain as segmented by step 1 and p_{zt} denotes the average tumor size in the data used to train the cRBM. We did the same for the bias term a_y connected to $\mathbf{y}$, matching it with the average fraction of core within whole tumor in the training data.

I. Implementation details for inference

The algorithm is initialized by affinely registering the atlas to the target image by using the method described in [55]. We then smooth the atlas map for the background class using a Gaussian kernel with a standard deviation of 4 mm, and subsequently exclude all voxels with a probability larger than 0.99 in this smoothed map.

The image intensities are log-transformed to accommodate the additive bias field model we use. To model the bias field, we use the lowest frequency components of the 3D discrete cosine transform (DCT) as basis functions. We set the number basis functions to 4 per dimension, which gives a total of $P = 4^3 = 64$.

The optimization in step 1 of the algorithm is performed on two resolution levels. For the first level, we smooth the atlas probabilities using a Gaussian kernel with a standard deviation of 2 mm to capture mesh deformations on a larger scale. In the second level, we refine the registration on a smaller scale without any smoothing. For this step, we use the same stopping criteria as in [38].

All Gaussian component parameters are initialized based on the atlas, except the mean values for tumor core and edema as these labels have spatially flat priors in the first step of the algorithm. We initialize these mean values in one image contrast based on the distance (measured in standard deviation) from the average data intensity in that contrast. Reasonable distance values can be found with knowledge of typical tumor intensities in an included image contrast or by estimates from subjects with already delineated edema and core. Based on initial experiments on the BRATS training dataset from section II-G, we set the distances for MR contrasts {FLAIR, T2, T1} to $\{1, 0.7, 0.2\}$ standard deviations for both tumor core and edema. For T1c we set the distance to 0.2 for edema and 1.5 for core (as we model specifically enhanced core in the first step of the algorithm). One of the datasets we test our method on includes the MR contrast T2-weighted DIR (see section III-A). For this contrast, we set the distance to the same as for FLAIR, without any initial experiments.

In step 2 of the algorithm, we generate 50 samples of $\mathbf{l}$, $\mathbf{z}$ and $\mathbf{y}$ after an initial burn-in period of 200 samples. The final

segmentation is then obtained by performing majority voting on these 50 samples.

The segmentation of one subject with our method took on average around two hours, with roughly equal time spent on the first and the second step of the algorithm. The algorithm was implemented in MATLAB 2015a, except for the atlas mesh deformation part which was implemented in C++, and performed on a Core i5-2400 3.1 GHz CPU with 8 GB of RAM.

III. Experiments

We will present segmentation results on three different datasets. In this section, we first describe these three datasets (section III-A), and then briefly describe the methods that we compare our method's performance against on tumor segmentation (section III-B). Finally, we describe the setup of the evaluation (section III-C).

A. Data

The first dataset consists of 42 glioblastoma patients that have undergone radiation therapy treatment at our institute (GTV size range: 6-130 cm^3). As part of their radiation therapy workup, these patients have been scanned with a CT scanner and a Siemens Magnetom Espree 1.5T MRI scanner. The dataset includes three MR-contrasts: T2-weighted FLAIR (transversal 2D-acquisition), T2-weighted (transversal 2D-acquisition) and T1-weighted with contrast enhancement (T1c, 3D-acquisition); with a voxel size of $1 \times 1 \times 3$ mm^3, $1 \times 1 \times 3$ mm^3 and $0.5 \times 0.5 \times 1$ mm^3 respectively. The CT scans have a voxel size of $0.5 \times 0.5 \times 3$ mm^3. As part of the treatment planning, the MR-defined GTV (corresponding to tumor core) and several organs-at-risk (including hippocampi, brainstem, eyes, optic nerves and chiasm) have been manually delineated on the transversal slices of the CT scan of each patient, with the MR-sequences registered and resampled to the CT scan. As the only pre-processing step, we co-register the MR scans and resample them to 1 mm isotropic resolution. We will refer to this dataset as the *clinical* dataset.

The second dataset is the test dataset of the brain tumor segmentation (BRATS) challenge that was held in conjunction with the BrainLes workshop at the 2015 MICCAI conference, where we participated with a previous version of our method [42]. The dataset is publicly available at the virtual skeleton online platform [56]. It consists of 53 patients from different imaging centers, with varying types of high- and low-grade gliomas. Some of the patients have had tumor resections. The included MR contrasts are T2-weighted FLAIR (varying 2D acquisition), T2-weighted (2D acquisition), T1-weighted (2D acquisition) and T1-weighted with contrast enhancement (T1c, 3D-acquisition). The scans have been acquired at different centers, with varying signal strength and resolution. All data were resampled to 1 mm isotropic resolution, aligned to the same anatomical template and skull-stripped by the challenge organizers. The dataset includes manual annotations of peritumoral edema and tumor core, which are not publicly available. Instead, the performance of a method can be evaluated by uploading segmentations to the online platform. We will refer to this dataset as the *BRATS test* dataset.

The third dataset consists of 7 patients with varying low- and high-grade gliomas. The patients have been scanned with a Siemens Trio 3T scanner at Institute of Neurology, UCL, London. No manual segmentation have been performed. The following MR-contrasts are included in the dataset, with 1 mm isotropic resolution: T2-weighted (3D acquisition), T2-weighted FLAIR (3D acquisition) and T2-weighted double inversion recovery (DIR, 3D-acquisition). In contrast to the other datasets, this dataset lacks T1-weighted contrasts and includes a new contrast: DIR. We will refer to this dataset as the *DIR* dataset.

B. Benchmark methods

We compare our method's performance in tumor segmentation to that of four methods that participated in the 2015 BRATS challenge. These four methods where all among the top-performing methods in the challenge.

1) Previous version of our method [42]: This method is a previous version of the method presented in this paper, focused on segmenting tumor only. The main difference is that the method uses an affinely registered atlas, instead of the mesh-based deformable atlas. It also differs in the initialization and prior distribution for the Gaussian parameters. The method segments a subject in around 30 minutes on a Nvidia TITAN black GPU.

2) GLISTRboost [25]: This semi-automated method is based on a modified version of the generative atlas-based method GLISTR [22], [23], which uses a tumor growth model to model the tumor. The method requires manual input of a seed-point for each tumor center and a radius of the extent of the tumor. To increase the segmentation performance, the method is extended with a discriminative post-processing step using a gradient boosting multi-label classification scheme followed by a patient-wise refinement step. The segmentation time was not reported, but GLISTR took around 85 minutes on a Intel Core i7 3.4 GHz CPU in [22].

3) Grade-specific CNNs [17]: This semi-automated method uses a discriminative 2D Convolutional Neural Network (CNN) approach. The method takes advantage of the fact that high- and low-grade tumors exhibit differences in intensity and spatial distribution. To do this, it uses two CNNs: one trained on high-grade tumors and one trained on low-grade tumors. The CNN to use for a specific subject is then chosen manually based on visual assessment, which is the only manual step in the method. The method is fast after the initial assessment, taking around 8 minutes to segment a tumor in a subject on a Nvidia 980 GPU.

4) Cascaded two-way CNN [18]: This fully automated method uses a discriminative 2D CNN approach as the previous method. The architecture has two pathways, where intensity features are automatically learned: one learning local details of tumor appearance and one for learning larger contexts. In the first part of the method, each voxel is predicted without taking local dependencies of labels into account. These voxel-wise predictions are then added as additional input to a

second CNN with the same architecture, forming a cascaded architecture. The method is fast, taking only 3 minutes to segment the tumor in a subject on a Nvidia TITAN black GPU.

C. Evaluation setup

We employ the two widely used metrics Dice score and Hausdorff distance to evaluate our method and compare its performance to that of other methods. For both metrics, we compare the automatic segmentation of a structure to the corresponding manual segmentation. Assuming that the manual segmentation captures the structure well, the obtained scores can be said to measure the method's performance.

The Dice score measures the overlap between two segmentations. Let $\mathbf{a}$ be a binary map of the automatically segmented structure and $\mathbf{m}$ be a binary map of the manually segmented structure, where 1 indicates presence of the structure. The Dice score is then computed as

$$\text{Dice}(\mathbf{a}, \mathbf{m}) = \frac{2\sum_i a_i m_i}{\sum_i a_i + \sum_i m_i},$$

where a score of zero means no overlap and a score of one means perfect overlap between the segmentations. The Dice score is more sensitive to segmentation differences in small structures than large structures, e.g., if a similar displacement is present in two structures the larger structure will normally still have a larger overlap.

In contrast, the Hausdorff distance evaluates the distance between the surfaces of two segmentations. Let A and M denote the sets of points on the two surfaces. We first find the shortest euclidean distance $d(a, m)$ for each point a in A to any point in set M and then find the longest of these distances. We do the same for set M and return the maximum distance given either from comparing A with M or M with A. This is the Hausdorff distance $d_{\text{H}}(A, M)$, defined by

$$d_{\text{H}}(A, M) = \max\{\max_{a \in A} \min_{m \in M} d(a, m), \max_{m \in M} \min_{a \in A} d(m, a)\}.$$

As the original Hausdorff metric can be substantially influenced by small outlying subregions, we instead use a more robust version of the metric. In this robust metric, the 95th percentile is returned instead of the maximum in the list of shortest distances.

For the clinical dataset, we will evaluate the segmentation of our method on tumor core (which corresponds to the manually segmented GTV) together with the following organs-at-risk: hippocampi (HC), brainstem (BS), chiasm (CH), optic nerves (ON) and eyes (EB). Before computing Dice score and Hausdorff distance, we transform and resample the automatic segmentations to the CT scan of each subject to be able to compare them to the manual segmentations. For the BRATS test dataset, we are especially interested in evaluating the segmentation of tumor core although we also evaluate the segmentation of the whole tumor region (i.e., both tumor core and edema). Note that enhanced core was also separately evaluated in the BRATS challenge, but it is not included in this paper as it is not directly relevant for radiation therapy. For this dataset, we will compare the scores obtained with our method to the scores from the benchmark methods presented in the previous section. As the DIR dataset contains no manual segmentations, we only do a visual inspection of the results – to indicate the ability of our method to adapt to varying situations.

IV. Results

In this section, we present the results on the three datasets described in the previous section. We first present the results for the BRATS test dataset where we compare to the benchmark methods (section IV-A), then we present the results for the clinical dataset (section IV-B), and finally we present the results for the DIR dataset (section IV-C).

A. BRATS test dataset

As the BRATS test dataset is skull-stripped, our method only segmented structures located within the brain (as specified in table I). Furthermore, we only evaluate this dataset for tumor segmentation, as manual segmentations for healthy structures are not available. Figure 2 shows box plots of the Dice scores and Hausdorff distances for this dataset. We show scores for our method (method 1) on tumor core and whole tumor together with the four benchmark methods that participated in the 2015 BRATS challenge. Note that the online platform does not report Hausdorff distance, while it was reported for the challenge. Thus, the plots of this metric only include the benchmark methods. The first benchmark method (method 2) is the previous version of our method that participated in the challenge. As expected, the average and median Dice scores show a similar performance for the method presented in this paper and the previous version. Comparing to the other benchmark methods, our method performs particularly well on tumor core and comparable to the other methods on whole tumor. The range of values are very large for all methods, illustrating the difficulty of this dataset. The dataset includes many subjects with large resections and a wide variety of tumors, e.g., low-grade tumors that has been shown to be difficult to segment in [8]. The Hausdorff distances for our previous version are somewhat worse than for at least two of the other methods, which could be explained by the post-processing used by the other methods.

Figure 3 shows slices of three representative segmentations with: FLAIR, T1c, T2 and T1, and segmentation by our method. Note that the manual segmentations compared against are not publicly available. We can see that the atlas deforms well to the subjects, and brainstem and hippocampi are well-captured. Furthermore, our method can segment brain tumors with large variations in size, location and appearance. Also note the low resolution and image quality in some of the images.

B. Clinical dataset

Figure 4 shows box plots of the Dice scores and Hausdorff distances for the clinical dataset, with the following structures: tumor core (TC), brainstem (BS), hippocampi (HC), eyes (EB), optic nerves (ON) and chiasm (CH). Note that this

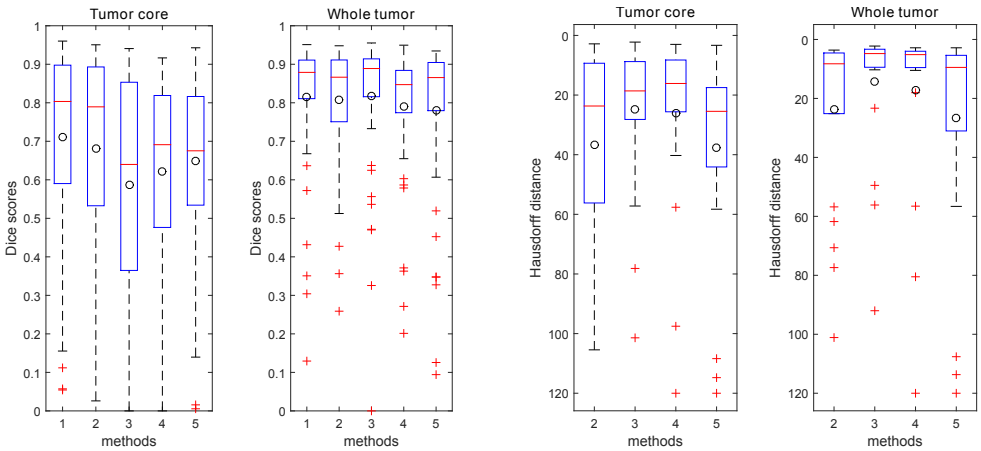


Fig. 2. Box plots of Dice scores and Haussdorf distances for tumor core and whole tumor (including edema and core) on the BRATS 2015 test dataset. Method 1 is the method presented in this paper, with scores as reported from the online platform (not including Hausdorff distance). The scores for the other methods are from the challenge. Method 2 is the previous version of our method used in the challenge [42], method 3 is the discriminative-generative method [25], method 4 is the grade-specific CNN method [17] and method 5 is the two-way CNN method [18]. On each box, the central line is the median, the circle is the mean and the edges of the box are the 25th and 75th percentiles. Outliers are shown with $+$.

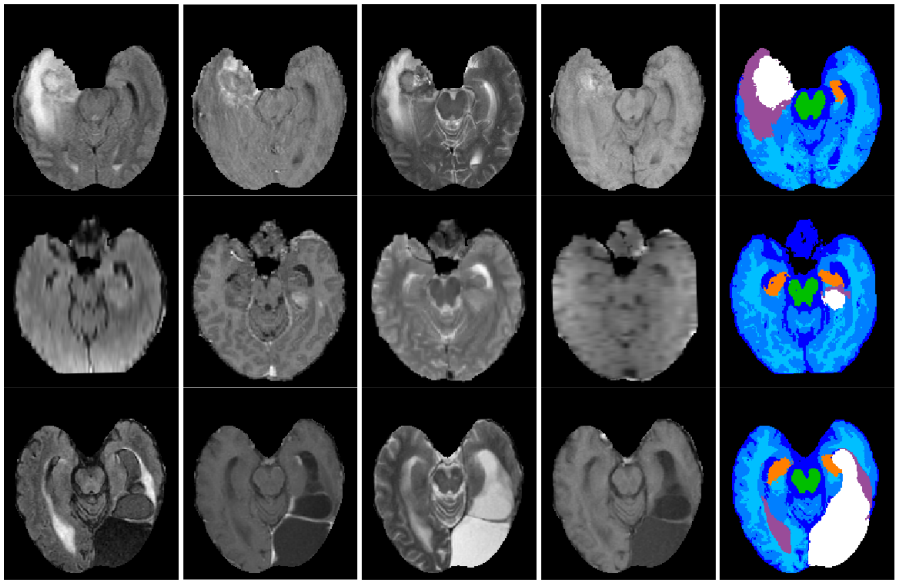


Fig. 3. Segmentations of 3 representative subjects of the BRATS test dataset. Slices of FLAIR, enhanced T1-weighted scan, T2-weighted scan and standard T1-weighted scan; and automatic segmentation. Label colors: white = tumor core, lilac = edema, green = brainstem, dark orange = hippocampi, shades of blue = other brain tissues. Note that the images are skull-stripped by the BRATS challenge organizers.

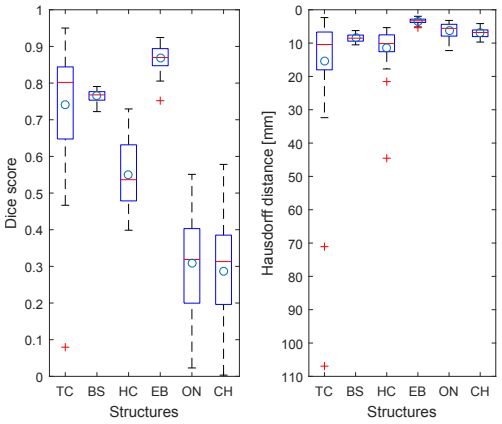


Fig. 4. Boxplots of Dice scores and Hausdorff distances for structures in clinical dataset. 42 subjects in total. On each box, the central line is the median, the circle is the mean and the edges of the box are the 25th and 75th percentiles. Outliers are shown with +.

dataset does not include the T1 image contrast, which is included in the BRATS test dataset, but the method readily adapts to this without the need for adjustment. Here, we have compared the automatic segmentations of our method to the manual segmentations that were used when planning the radiation therapy session of each subject. The median and average Dice scores for tumor core are consistent with the scores for the BRATS test dataset in figure 2. The range of Dice scores is smaller and only one outlier score is very low. The Hausdorff distances are also much lower than in figure 2, with only two outliers with high values. The smaller range is expected, as the dataset includes images with better quality and only includes glioblastomas – high-grade tumors with usually large contrast-enhanced tumor core regions. The Dice scores for eyes are high, except one outlier that was affected by a very thin outer eye wall, and the Hausdorff distances are low, indicating a good performance. The Dice scores for brainstem are not very high, but are consistent across the subjects as are the Hausdorff distances. Hippocampi, on the other hand, have a fairly large range and generally fairly low Dice scores. The Hausdorff distances are also fairly large, with two outliers. In the outliers, the method has segmented the hippocampus near to the tumor border while the manual segmentations lack the hippocampus in that brain hemisphere. Finally, the Dice scores for optic nerves and chiasm are generally low and with a large range. These structures are very small and thin, which significantly effect this metric. The Hausdorff distances for these structures are fairly low, which indicates that the manual and automatic segmentations are fairly close.

Figure 5 shows slices of seven representative segmentations with: T1c, FLAIR, T2, segmentation by the method and manual segmentation. From the images, we can see that the atlas deforms well to fit subjects with varying shapes. The method is capable of segmenting tumor cores with varying size, shape and intensity profile; although it underestimate the tumor size in some cases. Eyes, hippocampi and brainstem seem to be captured well. However, the method's hippocampus segmentation is larger than the manual segmentation. In the first case, neither the automatic or manual segmentation for optic nerves seems optimal, although the automatic segmentation seems better in this case. Furthermore, the MR slices show the ambiguity in the intensity profile for this structure, which is present in many subjects.

Figure 6 shows sagittal slices of three representative segmentations of hippocampi, together with surface plots of the segmentations. Note that the manual segmentations appear blocky because they have been delineated on the transversal slices of the CT scans. In all cases, the automatic segmentations are larger and capture the hippocampi better than the manual segmentations. As can be seen in the surface plots, the manual segmentations are not very consistent to each other. This explains to a large extent the fairly low and inconsistent Dice scores. However, our method slightly oversegments the hippocampi in some cases, as can be seen in the surface plot of the first case.

Figure 7 shows sagittal slices of three representative segmentations of brainstem, together with surface plots of the segmentations. The method captures the brainstem well, although it seems to slightly undersegment the structure. However, the method misses a large part of the midbrain that is present in the manual segmentations. This is because the protocols for brainstem segmentation are different for the manual segmentations and the healthy segmentations used to build the method's atlas. The protocol difference lowers the Dice scores, but keeps the scores consistent across subjects.

Figure 8 shows transversal slices of three representative segmentations of the optic system (including eyes, optic nerves and chiasm), together with surface plots of the segmentations. The method captures the eyes well, although in some cases the wall of the eye is slightly oversegmented. The method also has some difficulties when a subject has the eye lids open, as the solid wall between eye and air becomes very thin. The optic nerve (the thin nerve going from an eye in one end to the chiasm in the other end) is clearly a difficult structure to segment. Both the automatic segmentation and manual segmentation seems suboptimal. From the surface plots, it is also clear that although the two segmentations are fairly close and follow the same path, the overlap is often small. In most cases, the chiasm seems to be fairly well captured by the method, but as it is a small structure a small difference to the manual segmentation can have a large effect on the Dice scores. Furthermore, the segmentation protocol for the healthy subjects used to build the atlas seems to be slightly different than for the manual segmentations. The chiasm seem to extend further upwards in the brain in this protocol.

Figure 9 shows slices of two problematic tumor core segmentations, which can be seen as outliers in figure 4. In the first case, the method segments the tumor core well, but also captures an old infarct in another part of the brain. One other case had a similar problem, where the tumor core of interest is captured well but also a benign brain tumor in the cerebellum. The second case includes a very large resection at the border of the brain, which the method has difficulty to adapt to.

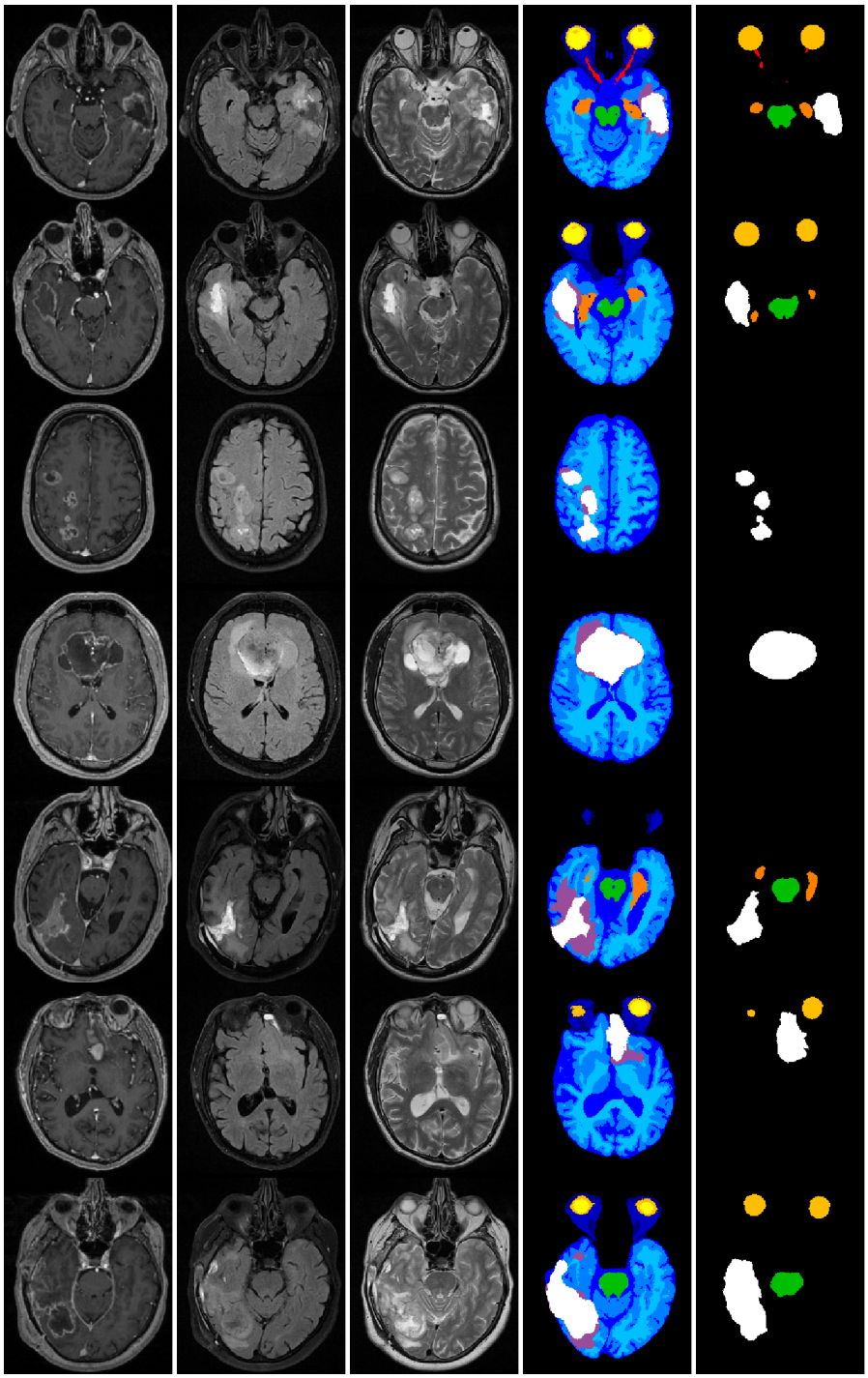


Fig. 5. Segmentations of 7 representative subjects of the clinical dataset. Slices of enhanced T1-weighted scan, FLAIR and T2-weighted scan, automatic segmentation and manual segmentation. Label colors: white = tumor core, lilac = edema, green = brainstem, dark orange = hippocampi, yellow/light orange = eyes, red = optic nerves, shades of blue = other brain tissues. For tumor core: Dice scores $= \{0.83, 0.91, 0.89, 0.92, 0.83, 0.63, 0.84\}$, Hausdorff distance $= \{11, 3, 3, 4, 6, 19, 7\}$.

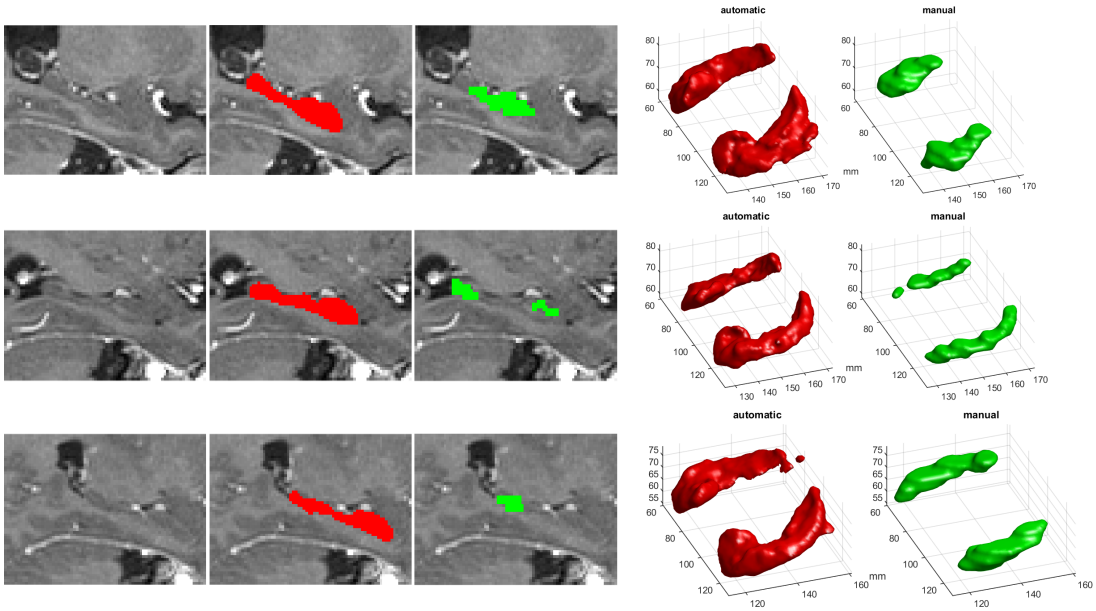


Fig. 6. Hippocampi on three representative subjects of the clinical dataset. Automatic segmentations in red and manual segmentations in green. Slice of segmentation overlayed on the T1-weighted scan and 3D surface plot of full structure. Dice score $= \{0.54, 0.45, 0.57\}$, Hausdorff distance $= \{12, 14, 11\}$.

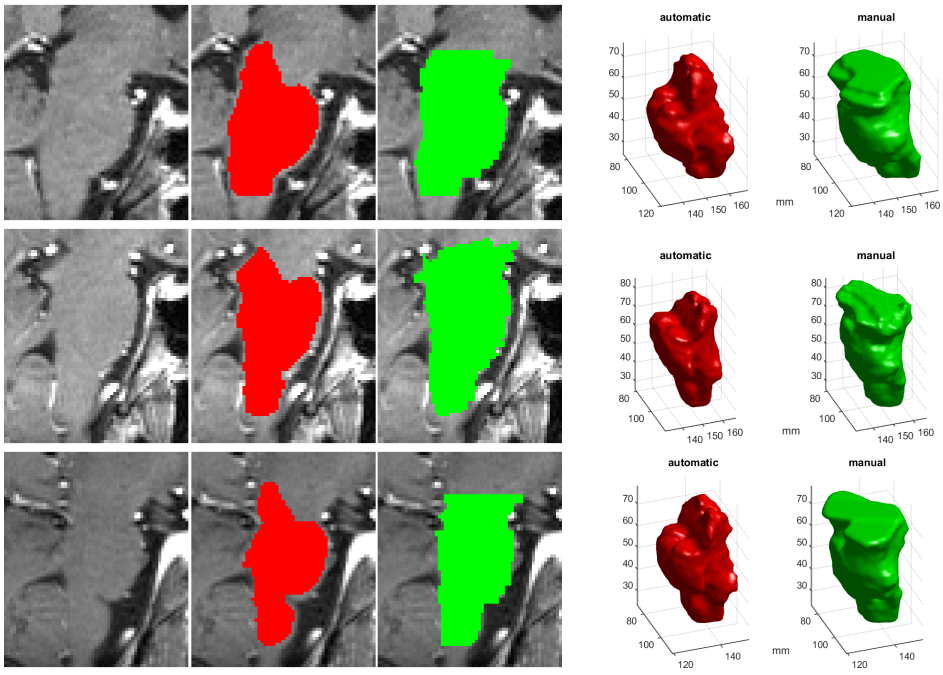


Fig. 7. Brainstem on three representative subjects of the clinical dataset. Automatic segmentations in red and manual segmentations in green. Slice of segmentation overlayed on the T1-weighted scan and 3D surface plot of full structure. Dice score $= \{0.77, 0.78, 0.78\}$, Hausdorff distance $= \{10, 10, 7\}$.

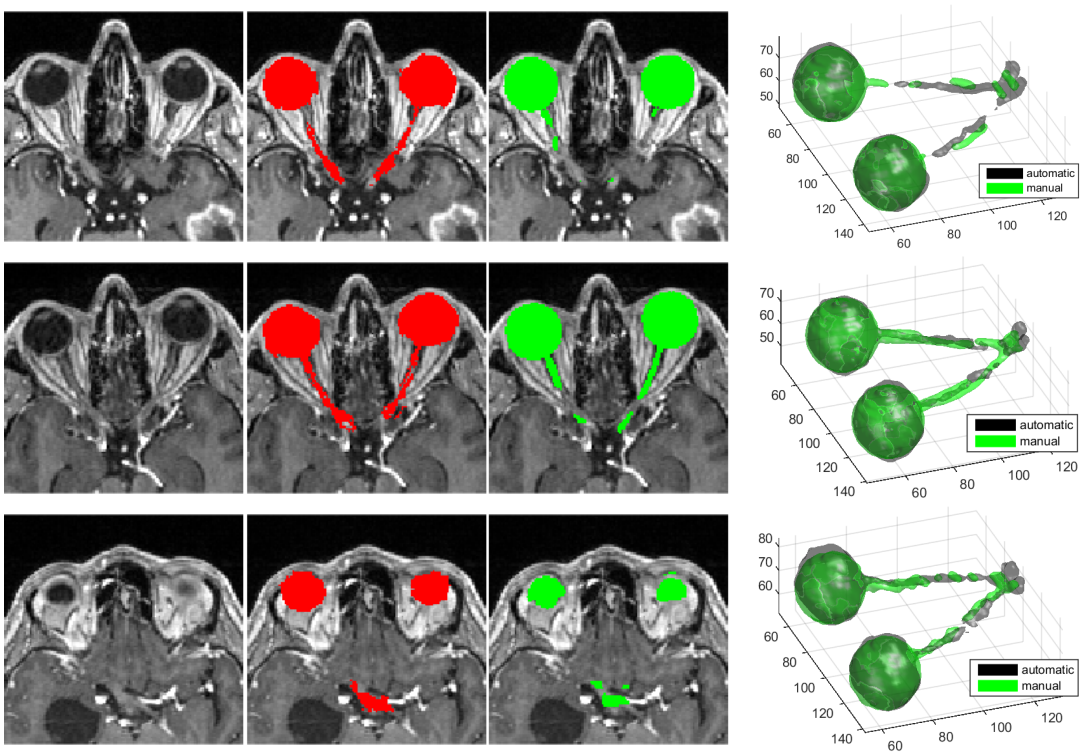


Fig. 8. Optic system on three representative subjects of the clinical dataset. Automatic segmentations in red and manual segmentations in green. Slice of segmentation overlayed on the T1-weighted scan and 3D surface plot of full structure. Scores for eyes, optic nerves and chiasm: Dice $= \{0.90, 0.91, 0.87\}$, $\{0.17, 0.41, 0.33\}$ and $\{0.32, 0.46, 0.38\}$; Hausdorff $= \{3, 4, 5\}$, $\{6, 3, 6\}$ and $\{6, 6, 5\}$.

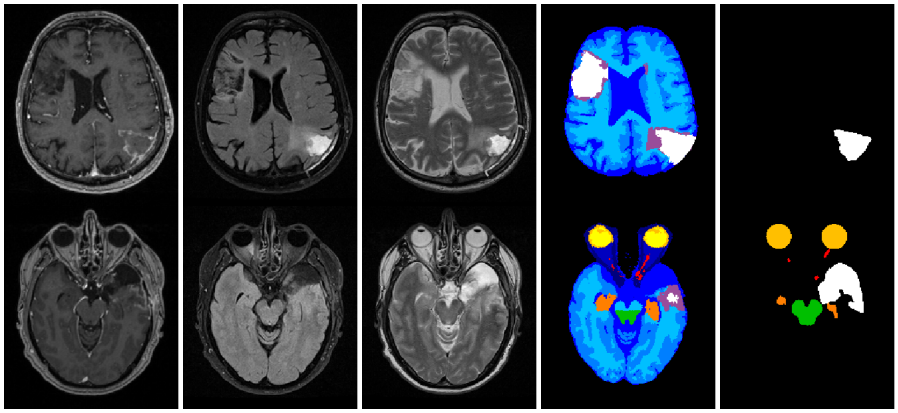


Fig. 9. Two problematic tumor core segmentations of the clinical dataset. Slices of enhanced T1-weighted scan, FLAIR, T2-weighted scan, automatic segmentation and manual segmentation. For tumor core: Dice scores $= \{0.47, 0.08\}$, Hausdorff distance $= \{107, 32\}$.

C. DIR dataset

Figure 10 shows slices of three representative segmentations with: DIR, FLAIR, T2 and segmentation by the method. The same structures were segmented as for the clinical dataset. As seen in the figure, our method can easily be made to segment MR-scans that lack a T1 contrast and include a DIR contrast. Visual inspection of all 7 segmentations revealed no significant deviations from other results presented in this section. However, a few observations can be made: the DIR images have a lower signal-to-noise ratio that some-what affects the segmentations, segmentation of high-grade gliomas typically benefit from a T1c contrast and the optic nerves seem to be better visualized on a T1-weighted contrast.

V. Discussion and conclusion

In this paper, we have presented a fully automated generative method for simultaneous segmentation of brain tumors and an extensive set of organs-at-risk (OARs) applicable to radiation therapy planning. The method combines a previously validated atlas-based model for detailed healthy tissue segmentation with a model for brain tumor segmentation based on convolutional restricted Boltzmann machines. The method separates the modeling of anatomy from the modeling of image intensities. Thus, the method is adaptable to varying image sequences and modalities. To learn the parameters of the anatomy model, just 10 training segmentations were used for the healthy anatomy and 30 training segmentations for the tumor anatomy (without using any *intensity* information from the training data). The small number of needed training subjects makes the method fairly easy to extend to handle new structures.

A few parameters relating to tumor tissue in the intensity model were set by using intensity information from training data. We found that this was necessary to guide the model to the correct intensities for tumor in the MR-contrasts FLAIR and T1c. Although this is a limitation, we can conclude from the validation results that the method is still adaptable to FLAIR and T1c images from varying imaging centers with varying sequence protocols. Furthermore, our experiments indicate that the method can adapt when new contrasts are included, such as double-inversion-recovery sequences, and when some are missing, such as T1 (cf. clinical dataset). To adapt the method to other image contrasts than those encountered in this paper should be easy, without the need for dedicated training data.

Our experiments show that the method's performance in segmenting tumors compares well to other state-of-the-art methods in brain tumor segmentation, while also being capable of segmenting the OARs hippocampi, brainstem, eyes, optic nerves and optic chiasm. We evaluated the OAR segmentations in 42 patients with manual segmentations used when planning a radiation therapy session. The evaluation showed a generally good performance in segmenting hippocampi (HC), brainstem (BS) and eyes (EB); but lower performance in segmenting the very small structures optic nerves (ON) and chiasm (CH). The performance can be compared to the inter-rater variability shown in [7], where 8 experts segmented OARs in 20 high-grade glioma patients. The inter-rater variability in [7] measured in average Dice scores was: BS: 0.83, EB: 0.84, ON: 0.50, CH: 0.39. This can be compared to the average Dice scores for our method for these structures: BS: 0.77, EB: 0.86, ON: 0.31, CH: 0.29. It is clear that the Dice scores for optic nerves and chiasm are low even for experts, although not as low as for our method.

In our experiments, the Dice score for brainstem was significantly affected by differing delineation protocols between the experts at the clinic and the segmentations used to train the atlas in our method. Differences in segmentation protocols are fairly common. As just 10 training subjects are used to build the atlas, it would be fairly feasible to update the segmentations of these subjects to adhere to a differing protocol. The manual segmentations at the clinic for hippocampi was found to be fairly variable in quality, which was also true for some optic nerve and chiasm segmentations. To use manual segmentations from radiation therapy planning as ground truth might not be optimal, as structures far away from a tumor might not be carefully delineated because they will not significantly affect the radiation therapy plan. Furthermore, because of the large inter-rater variability, a superior ground truth would be obtained by fusing segmentations from several experts.

It is clear that further research is needed to obtain optimal chiasm and optic nerve segmentations. At the moment, the atlas deformation process is the same for all structures. A way forward could be to incorporate geometrical information in the prior, e.g., about the tubular structure of the optic nerves which was successfully used by the discriminative approach in [36]. The tumor segmentation performance compares well to other brain tumor segmentation methods, but can still be improved. The tumor shape model is still fairly local, which can affect the segmentation when e.g., inner parts of the tumor have a similar appearance to healthy structures. A model with a deeper structure with more layers of hidden units, similar to the deep structure used in CNNs, could potentially improve the performance on a more global scale. Finally, a potential improvement could be achieved if the tumor shape model would inform the deformation of the atlas.

Appendix A
Constraints in prior

The linear constraints in the prior $p(\boldsymbol{\theta})$, in equation 3, is formulated in the form

$$\mathbf{A}\mathbf{v} \leq \mathbf{b}, \quad \text{where } \mathbf{v} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_S \end{pmatrix}, \tag{4}$$

Here, S denotes the total number of mean vectors in the model, i.e. for all labels. Let R denote the number of constraints. Matrix $\mathbf{A} = (\mathbf{a}_1, ..., \mathbf{a}_R)^T$, where column vector $\mathbf{a}_r \in \mathbb{R}^{(S \times N)}$ encodes one constraint. Column vector $\mathbf{b} \in \mathbb{R}^R$ contains zeros. The first constraint in table II will for example be encoded,

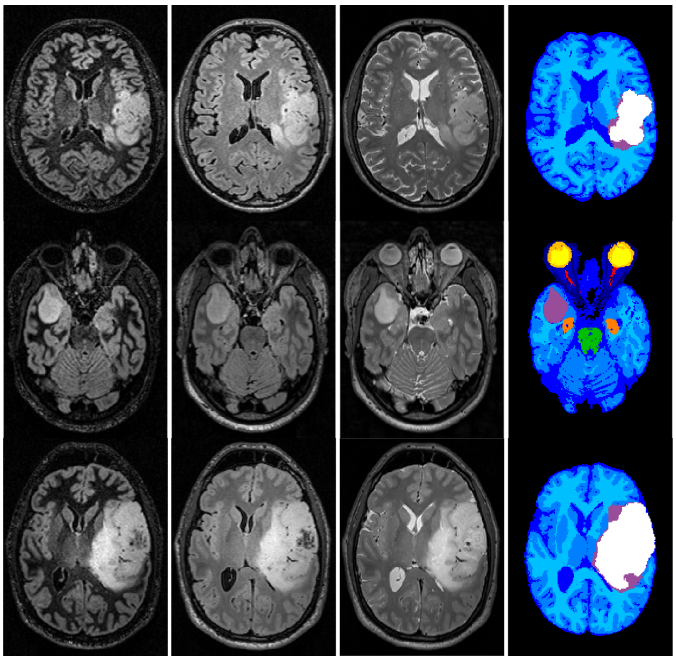


Fig. 10. Three representative segmentations from the DIR dataset. Slices of DIR, FLAIR, T2-weighted scan and automatic segmentation.

for $N = 3$ contrasts where the first contrast is FLAIR, as

$$\mathbf{a}_1^T = \begin{pmatrix} 1.04 & 0 & 0 & -1 & 0 & \cdots & 0 \end{pmatrix}, \quad \text{with } \mathbf{v} = \begin{pmatrix} \boldsymbol{\mu}_{\text{WM}} \\ \boldsymbol{\mu}_{\text{TE}} \\ \vdots \end{pmatrix}.$$

APPENDIX B
INFERENCE, STEP 1

Here we describe how we maximize $p(\boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{D})$ in the first step of the inference. Throughout this appendix, we use x to denote the $K + 2$ segmentation labels (K normal tissue classes, edema and tumor core) given by the combination $\{l, z, y\}$. Taking the logarithm of $p(\boldsymbol{\eta}, \boldsymbol{\theta}|\mathbf{D})$, we can rewrite this problem as a maximization of the following objective function:

$$\{\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\eta}}\} = \arg\max_{\{\boldsymbol{\theta}, \boldsymbol{\eta}\}} \Big[\sum_i \log \Big(\sum_{x=1}^{K+2} p_i(\mathbf{d}_i|x, \boldsymbol{\theta}) p_i(x|\boldsymbol{\eta}) \Big) + \log p(\boldsymbol{\eta}) + \log p(\boldsymbol{\theta}) \Big],$$

where $p(x|\boldsymbol{\eta})$ refers to the simplified model of the label prior.

We optimize the atlas deformation parameters $\boldsymbol{\eta}$ and the likelihood parameters $\boldsymbol{\theta}$ with a coordinate ascent algorithm as in [38], which alternates between optimizing each set of parameters. The atlas deformation parameters $\boldsymbol{\eta}$ are optimized with a standard conjugate gradient optimizer [48], while keeping $\boldsymbol{\theta}$ fixed. The likelihood parameters $\boldsymbol{\theta}$ are optimized with a generalized expectation-maximization (GEM) algorithm [49], while keeping $\boldsymbol{\eta}$ fixed. The GEM algorithm alternates between two steps: the E-step and the M-step. In the E-step, the following voxel-wise soft assignment is computed to each Gaussian component, based on the current parameter estimates:

$$q_i^{xg} = \frac{\gamma_{xg} \mathcal{N}(\mathbf{d}_i|\boldsymbol{\mu}_{xg} - \mathbf{C}\boldsymbol{\phi}_i, \boldsymbol{\Sigma}_{xg}) p_i(x|\boldsymbol{\eta})}{\sum_{x'=1}^{K+2} p_i(\mathbf{d}_i|x', \boldsymbol{\theta}) p_i(x'|\boldsymbol{\eta})}$$

In the M-step, the parameters are subsequently updated. In the following we have defined $N_{xg} = \sum_i q_i^{xg}$. The weights are updated according to

$$\gamma_{xg} \leftarrow \frac{N_{xg} + \alpha_0 - 1}{\sum_{g'=1}^{G_x} (N_{xg'} + \alpha_0 - 1)}.$$

For the mean vectors, we first define the updates that would be obtained if the constraints in prior $p(\boldsymbol{\theta})$ would be removed:

$$\bar{\boldsymbol{\mu}}_{xg} = \frac{\kappa_0 \mathbf{m}_0 + \sum_i q_i^{xg} (\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i)}{\kappa_0 + N_{xg}}.$$

We can then update the mean vectors according to

$$\mathbf{v} \leftarrow \arg\max_{\mathbf{v}} \Big[\frac{1}{2} \mathbf{v}^T \mathbf{Q} \mathbf{v} + \mathbf{f}^T \mathbf{v} \Big] \quad \text{s.t.} \quad \mathbf{A}\mathbf{v} \leq \mathbf{b}, \tag{5}$$

with

$$\mathbf{Q} = \begin{pmatrix} (N_1 + \kappa_0)\boldsymbol{\Sigma}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (N_2 + \kappa_0)\boldsymbol{\Sigma}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & (N_S + \kappa_0)\boldsymbol{\Sigma}_S^{-1} \end{pmatrix},$$

$$\mathbf{f} = -\mathbf{Q} \begin{pmatrix} \bar{\boldsymbol{\mu}}_1 \\ \vdots \\ \bar{\boldsymbol{\mu}}_S \end{pmatrix}, \quad \text{and} \quad \mathbf{v} = \begin{pmatrix} \boldsymbol{\mu}_1 \\ \vdots \\ \boldsymbol{\mu}_S \end{pmatrix},$$

where S denotes the total number of Gaussian components in the model, i.e., for all labels. Here, $\mathbf{A}$ and $\mathbf{b}$ relate to the constraints in the prior $p(\boldsymbol{\theta})$, and are defined in appendix A. Solving equation 5 is known as *quadratic programming*, for which an implementation is directly available in MATLAB. For a current estimate of the covariances $\{\boldsymbol{\Sigma}_s\}$, equation 5 is solved to obtain the updated means $\{\boldsymbol{\mu}_s\}$. With those updated mean values, the covariances are then updated according to

$$\boldsymbol{\Sigma}_{xg} \leftarrow \frac{\mathbf{S}_0 + \mathbf{S}_{xg} + \kappa_0 \mathbf{m}_0 \mathbf{m}_0^T - (\kappa_0 + N_{xg})\boldsymbol{\mu}_{xg}\boldsymbol{\mu}_{xg}^T}{\nu_0 + D + 2 + N_{xg}},$$
$$\text{where} \quad \mathbf{S}_{xg} = \sum_i q_i^{xg} (\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i)(\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i)^T.$$

The bias field parameters in $\mathbf{C}$ are updated according to

$$\begin{pmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_N \end{pmatrix} \leftarrow \begin{pmatrix} \boldsymbol{\Phi}^T\mathbf{S}_{1,1}\boldsymbol{\Phi} & \cdots & \boldsymbol{\Phi}^T\mathbf{S}_{1,N}\boldsymbol{\Phi} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}^T\mathbf{S}_{N,1}\boldsymbol{\Phi} & \cdots & \boldsymbol{\Phi}^T\mathbf{S}_{N,N}\boldsymbol{\Phi} \end{pmatrix}^{-1} \cdot \begin{pmatrix} \boldsymbol{\Phi}^T(\mathbf{S}_{1,1}\mathbf{r}_{1,1} + \cdots + \mathbf{S}_{1,N}\mathbf{r}_{1,N}) \\ \vdots \\ \boldsymbol{\Phi}^T(\mathbf{S}_{N,1}\mathbf{r}_1 + \cdots + \mathbf{S}_{N,N}\mathbf{r}_{N,N}) \end{pmatrix},$$

where

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_1^1 & \cdots & \phi_P^1 \\ \vdots & \ddots & \vdots \\ \phi_1^I & \cdots & \phi_P^I \end{pmatrix}, \quad \mathbf{S}_{m,n} = \text{diag}\,(s_i^{m,n})$$

$$\text{and} \quad \mathbf{r}_{m,n} = (r_1^{m,n}, ..., r_I^{m,n})^T, \quad \text{with}$$

$$s_i^{m,n} = \sum_x \sum_{g=1}^{G_x} s_{ixg}^{m,n}, \quad s_{ixg}^{m,n} = q_i^{xg} (\boldsymbol{\Sigma}_{xg}^{-1})_{m,n}$$

$$\text{and} \quad r_i^{m,n} = d_i^n - \frac{\sum_x \sum_{g=1}^{G_x} s_{ixg}^{m,n} (\boldsymbol{\mu}_{xg})_n}{\sum_x \sum_{g=1}^{G_x} s_{ixg}^{m,n}}.$$

It can be shown that the objective function is guaranteed to increase with respect to $\boldsymbol{\theta}$ with these update rules in each GEM iteration [49].

APPENDIX C
INFERENCE, STEP 2: SAMPLING

Here we describe how we sample from $p(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$. We first describe a sampler of a distribution $q(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$ that is identical to $p(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$, except that the constraints on the means in the prior $p(\boldsymbol{\theta})$, in equation 3, are removed. Using the notation $i \in x$ to denote which of the $K+2$ segmentation labels (one of the K normal tissue classes, edema, or core) the i^{th} voxel belongs to according to its combination $\{l_i, z_i, y_i\}$, this new distribution is given by

$$q(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) \propto \prod_x \left[\left(\prod_{i \in x} p_i(\mathbf{d}_i|\boldsymbol{\theta}_x, \mathbf{C}) \right) \text{Dir}(\boldsymbol{\gamma}_x|\alpha_0) \prod_{g=1}^{G_x} \text{NIW}(\boldsymbol{\mu}_{xg}, \boldsymbol{\Sigma}_{xg}|\mathbf{m}_0, \kappa_0, \upsilon_0, \mathbf{S}_0) \right].$$

Introducing the auxiliary variable $\mathbf{t} = (t_1, \ldots, t_I)^T$, with $t_i \in \{1, \ldots, G_x\}, i \in x$ indicating which individual Gaussian component in the relevant mixture model the i^{th} voxel is assocated with, this distribution is obtained as a marginal distribution

$$q(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) = \sum_{\mathbf{t}} q(\boldsymbol{\theta}, \mathbf{t}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$$

of

$$q(\boldsymbol{\theta}, \mathbf{t}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) \propto \prod_x \left[\left(\prod_{i \in x} \gamma_{xt_i} \mathcal{N}(\mathbf{d}_i|\boldsymbol{\mu}_{xt_i} + \mathbf{C}\boldsymbol{\phi}_i, \boldsymbol{\Sigma}_{xt_i}) \right) \text{Dir}(\boldsymbol{\gamma}_x|\alpha_0) \prod_{g=1}^{G_x} \text{NIW}(\boldsymbol{\mu}_{xg}, \boldsymbol{\Sigma}_{xg}|\mathbf{m}_0, \kappa_0, \upsilon_0, \mathbf{S}_0) \right].$$

Therefore, samples of $q(\boldsymbol{\theta}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$ can be obtained with a blocked Gibbs sampler of $q(\boldsymbol{\theta}, \mathbf{t}|\mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D})$ that cycles through the following conditional distributions, and discarding the samples of $\mathbf{t}$ (in the following, we have defined $N_{xg} = \sum_{i \in x} \mathbb{I}(t_i = g)$ and denote the inverse Wishart distribution as $\text{IW}(\cdot)$):

$$q(\mathbf{t}|\boldsymbol{\theta}, \mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) \propto \prod_x \prod_{i \in x} \gamma_{xt_i} \mathcal{N}(\mathbf{d}_i|\boldsymbol{\mu}_{xt_i} + \mathbf{C}\boldsymbol{\phi}_i, \boldsymbol{\Sigma}_{xt_i}),$$

$$q(\{\boldsymbol{\gamma}_x\}|\boldsymbol{\theta}_{\setminus\{\boldsymbol{\gamma}_x\}}, \mathbf{t}, \mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) = \prod_x \text{Dir}(\{\alpha_0 + N_{xg}\}_{g=1}^{G_x}),$$

$$q(\{\boldsymbol{\mu}_{xg}\}|\boldsymbol{\theta}_{\setminus\{\boldsymbol{\mu}_{xg}\}}, \mathbf{t}, \mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) = \prod_x \prod_g \mathcal{N}(\boldsymbol{\mu}_{xg}|\mathbf{m}_{xg}, \mathbf{V}_{xg})$$

$$\text{where} \quad \mathbf{V}_{xg} = \frac{1}{\kappa_0 + N_{xg}} \boldsymbol{\Sigma}_{xg}, \quad \mathbf{m}_{xg} = \frac{\kappa_0 \mathbf{m}_0 + N_{xg} \bar{\mathbf{d}}_{xg}}{\kappa_0 + N_{xg}}$$

$$\text{and} \quad \bar{\mathbf{d}}_{xg} = \frac{\sum_{i \in x} \mathbb{I}(t_i = g)(\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i)}{N_{xg}},$$

$$q(\{\boldsymbol{\Sigma}_{xg}\}|\boldsymbol{\theta}_{\setminus\{\boldsymbol{\Sigma}_{xg}\}}, \mathbf{t}, \mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) = \prod_x \prod_g \text{IW}(\boldsymbol{\Sigma}_{xg}|\mathbf{S}_{xg}, \upsilon_{xg})$$

where

$$\mathbf{S}_{xg} = \mathbf{S}_0 + \sum_{i \in x} \mathbb{I}(t_i = g)(\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i - \boldsymbol{\mu}_{xg})(\mathbf{d}_i - \mathbf{C}\boldsymbol{\phi}_i - \boldsymbol{\mu}_{xg})^T$$

$$\text{and} \quad \upsilon_{xg} = \upsilon_0 + N_{xg},$$

and finally

$$q(\mathbf{c}|\boldsymbol{\theta}_{\setminus\mathbf{c}}, \mathbf{t}, \mathbf{l}, \mathbf{z}, \mathbf{y}, \mathbf{D}) = \mathcal{N}(\boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c)$$

with

$$\mathbf{c} = \begin{pmatrix} \mathbf{c}_1 \\ \vdots \\ \mathbf{c}_N \end{pmatrix}, \quad \boldsymbol{\Sigma}_c = \begin{pmatrix} \boldsymbol{\Phi}^T\mathbf{S}_{1,1}\boldsymbol{\Phi} & \cdots & \boldsymbol{\Phi}^T\mathbf{S}_{1,N}\boldsymbol{\Phi} \\ \vdots & \ddots & \vdots \\ \boldsymbol{\Phi}^T\mathbf{S}_{N,1}\boldsymbol{\Phi} & \cdots & \boldsymbol{\Phi}^T\mathbf{S}_{N,N}\boldsymbol{\Phi} \end{pmatrix}^{-1}$$

$$\text{and} \quad \boldsymbol{\mu}_c = \boldsymbol{\Sigma}_c \begin{pmatrix} \boldsymbol{\Phi}^T(\mathbf{S}_{1,1}\mathbf{r}_1 + \cdots + \mathbf{S}_{1,N}\mathbf{r}_N) \\ \vdots \\ \boldsymbol{\Phi}^T(\mathbf{S}_{N,1}\mathbf{r}_1 + \cdots + \mathbf{S}_{N,N}\mathbf{r}_N) \end{pmatrix}.$$

Here,

$$\boldsymbol{\Phi} = \begin{pmatrix} \phi_1^1 & \cdots & \phi_P^1 \\ \vdots & \ddots & \vdots \\ \phi_1^I & \cdots & \phi_P^I \end{pmatrix}, \quad \mathbf{S}_{m,n} = \text{diag}\left(\left(\boldsymbol{\Sigma}_{xt_i}^{-1}\right)_{m,n} \right)$$

$$\text{and} \quad \mathbf{r}_n = (r_1^n, ..., r_I^n)^T, \quad \text{with} \quad r_i^n = d_i^n - (\boldsymbol{\mu}_{xt_i})_n.$$

Using this scheme to sample from the unconstrained distribution $q(\boldsymbol{\theta}|\mathbf{l},\mathbf{z},\mathbf{y},\mathbf{D})$, samples from the distribution of interest $p(\boldsymbol{\theta}|\mathbf{l},\mathbf{z},\mathbf{y},\mathbf{D})$ – which contains constraints on $\{\boldsymbol{\mu}_{xg}\}$ as specified in table II – can be obtained by performing rejection sampling in the sampling step of $q(\{\boldsymbol{\mu}_{xg}\}|\boldsymbol{\theta}_{\setminus\{\boldsymbol{\mu}_{xg}\}},\mathbf{t},\mathbf{l},\mathbf{z},\mathbf{y},\mathbf{D})$, i.e., by repeatedly generating unconstrained samples of $\{\boldsymbol{\mu}_{xg}\}$ until values are obtained for which the constraints on $\{\boldsymbol{\mu}_{xg}\}$ are fullfilled.

In our implementation, rather than repeating the Gibbs sampler steps described above until the Markov chain reaches equilibrium and an independent sample of $\boldsymbol{\theta}$ is obtained, we only make a single sweep before obtaining new samples of $\mathbf{G}$, $\mathbf{H}$, and $\{\mathbf{l},\mathbf{z},\mathbf{y}\}$ in the main loop described in section II-E, effectively implementing a so-called partially collapsed Gibbs sampler [57].

ACKNOWLEDGMENT

This research was supported by the NIH NCRR (P41-RR14075, 1S10RR023043), NIBIB (R01EB013565) and the Lundbeck foundation (R141-2013-13117).

REFERENCES

[1] M. Preusser, S. de Ribaupierre, A. Wöhrer, S. C. Erridge, M. Hegi, M. Weller, and R. Stupp, "Current concepts and management of glioblastoma," *Annals of neurology*, vol. 70, no. 1, pp. 9–21, 2011.

[2] O. L. Chinot, W. Wick, W. Mason, R. Henriksson, F. Saran, R. Nishikawa, A. F. Carpentier, K. Hoang-Xuan, P. Kavan, D. Cernea, *et al.*, "Bevacizumab plus radiotherapy–temozolomide for newly diagnosed glioblastoma," *New England Journal of Medicine*, vol. 370, no. 8, pp. 709–722, 2014.

[3] R. Shaffer, A. M. Nichol, E. Vollans, M. Fong, S. Nakano, V. Moiseenko, M. Schmuland, R. Ma, M. McKenzie, and K. Otto, "A comparison of volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for frontal and temporal high-grade gliomas," *International Journal of Radiation Oncology* Biology* Physics*, vol. 76, no. 4, pp. 1177–1184, 2010.

[4] P. Munck af Rosenschöld, S. Engelholm, L. Ohlhues, I. Law, I. Vogelius, and S. A. Engelholm, "Photon and proton therapy planning comparison for malignant glioma based on CT, FDG-PET, DTI-MRI and fiber tracking," *Acta Oncologica*, vol. 50, no. 6, pp. 777–783, 2011.

[5] P. Munck af Rosenschöld, J. Costa, S. A. Engelholm, M. J. Lundemann, I. Law, L. Ohlhues, and S. Engelholm, "Impact of [18F]-fluoro-ethyl-tyrosine PET imaging on target definition for radiation therapy of high-grade glioma," *Neuro-oncology*, vol. 17, no. 5, pp. 757–763, 2015.

[6] J. Dolz, L. Massoptier, and M. Vermandel, "Segmentation algorithms of subcortical brain structures on MRI for radiotherapy and radiosurgery: a survey," *IRBM*, vol. 36, no. 4, pp. 200–212, 2015.

[7] M. Deeley, A. Chen, R. Datteri, J. Noble, A. Cmelak, E. Donnelly, A. Malcolm, L. Moretti, J. Jaboin, K. Niermann, *et al.*, "Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study," *Physics in medicine and biology*, vol. 56, no. 14, p. 4557, 2011.

[8] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, R. Wiest, *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Transactions on Medical Imaging*, vol. 34, no. 10, pp. 1993–2024, 2015.

[9] N. Sauwen, M. Acou, S. Van Cauter, D. Sima, J. Veraart, F. Maes, U. Himmelreich, E. Achten, and S. Van Huffel, "Comparison of unsupervised classification methods for brain tumor segmentation using multi-parametric MRI," *NeuroImage: Clinical*, vol. 12, pp. 753–764, 2016.

[10] J. Dolz, A. Laprie, S. Ken, H.-A. Leroy, N. Reyns, L. Massoptier, and M. Vermandel, "Supervised machine learning-based classification scheme to segment the brainstem on MRI in multicenter brain tumor treatment context," *International journal of computer assisted radiology and surgery*, vol. 11, no. 1, pp. 43–51, 2016.

[11] N. Cordier, H. Delingette, and N. Ayache, "A patch-based approach for the segmentation of pathologies: Application to glioma labelling," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 1066–1076, 2016.

[12] N. J. Tustison, K. Shrinidhi, M. Wintermark, C. R. Durst, B. M. Kandel, J. C. Gee, M. C. Grossman, and B. B. Avants, "Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR," *Neuroinformatics*, vol. 13, no. 2, pp. 209–225, 2015.

[13] A. Islam, S. M. Reza, and K. M. Iftekharuddin, "Multifractal texture estimation for detection and segmentation of brain tumors," *IEEE transactions on biomedical engineering*, vol. 60, no. 11, pp. 3204–3215, 2013.

[14] D. Zikic, B. Glocker, E. Konukoglu, A. Criminisi, C. Demiralp, J. Shotton, O. Thomas, T. Das, R. Jena, and S. Price, "Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 369–376, Springer, 2012.

[15] O. Maier, M. Wilms, and H. Handels, "Image Features for Brain Lesion Segmentation Using Random Forests," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pp. 119–130, Springer, 2016.

[16] S. Bauer, L.-P. Nolte, and M. Reyes, "Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 354–361, Springer, 2011.

[17] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, vol. 9556, pp. 131–143, Lecture Notes in Computer Science, Springer, 2016.

[18] M. Havaei, F. Dutil, C. Pal, H. Larochelle, and P.-M. Jodoin, "A convolutional neural network approach to brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pp. 195–208, Springer, 2016.

[19] M. Lyksborg, O. Puonti, M. Agn, and R. Larsen, "An ensemble of 2D convolutional neural networks for tumor segmentation," in *Scandinavian Conference on Image Analysis*, pp. 201–211, Springer, 2015.

[20] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation," *arXiv preprint arXiv:1603.05959*, 2016.

[21] B. H. Menze, K. Van Leemput, D. Lashkari, M.-A. Weber, N. Ayache, and P. Golland, "A generative model for brain tumor segmentation in multi-modal images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 151–159, Springer, 2010.

[22] D. Kwon, R. T. Shinohara, H. Akbari, and C. Davatzikos, "Combining generative models for multifocal glioma segmentation and registration," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 763–770, Springer, 2014.

[23] A. Gooya, K. M. Pohl, M. Bilello, L. Cirillo, G. Biros, E. R. Melhem, and C. Davatzikos, "GLISTR: glioma image segmentation and registration," *IEEE transactions on medical imaging*, vol. 31, no. 10, pp. 1941–1954, 2012.

[24] B. H. Menze, K. Van Leemput, D. Lashkari, T. Riklin-Raviv, E. Geremia, E. Alberts, P. Gruber, S. Wegener, M.-A. Weber, G. Székely, *et al.*, "A Generative Probabilistic Model and Discriminative Extensions for Brain Lesion SegmentationWith Application to Tumor and Stroke," *IEEE transactions on medical imaging*, vol. 35, no. 4, pp. 933–946, 2016.

[25] S. Bakas, K. Zeng, A. Sotiras, S. Rathore, H. Akbari, B. Gaonkar, M. Rozycki, S. Pati, and C. Davatzikos, "GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pp. 144–155, Springer, 2016.

[26] B. Fischl, "FreeSurfer," *Neuroimage*, vol. 62, no. 2, pp. 774–781, 2012.

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

[28] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 3, pp. 611–623, 2013.

[29] S. Bauer, H. Lu, C. P. May, L.-P. Nolte, P. Büchler, and M. Reyes, "Integrated segmentation of brain tumor images for radiotherapy and neurosurgery," *International Journal of Imaging Systems and Technology*, vol. 23, no. 1, pp. 59–63, 2013.

[30] M. B. Cuadra, C. Pollo, A. Bardera, O. Cuisenaire, J.-G. Villemure, and J.-P. Thiran, "Atlas-based segmentation of pathological MR brain images using a model of lesion growth," *IEEE transactions on medical imaging*, vol. 23, no. 10, pp. 1301–1314, 2004.

[31] M. Conson, L. Cella, R. Pacelli, M. Comerci, R. Liuzzi, M. Salvatore, and M. Quarantelli, "Automated delineation of brain structures in patients undergoing radiotherapy for primary brain tumors: From atlas to dose–volume histograms," *Radiotherapy and Oncology*, vol. 112, no. 3, pp. 326–331, 2014.

[32] A. Isambert, F. Dhermain, F. Bidault, O. Commowick, P.-Y. Bondiau, G. Malandain, and D. Lefkopoulos, "Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context," *Radiotherapy and oncology*, vol. 87, no. 1, pp. 93–99, 2008.

[33] J. Dolz, N. Betrouni, M. Quidet, D. Kharroubi, H. A. Leroy, N. Reyns, L. Massoptier, and M. Vermandel, "Stacking denoising auto-encoders in a deep network to segment the brainstem on MRI in brain cancer patients: A clinical study," *Computerized Medical Imaging and Graphics*, vol. 52, pp. 8–18, 2016.

[34] J. Dolz, H.-A. Leroy, N. Reyns, L. Massoptier, and M. Vermandel, "A Fast and Fully Automated Approach to Segment Optic Nerves on MRI and its application to radiosurgery," in *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pp. 1102–1105, IEEE, 2015.

[35] S. Panda, A. J. Asman, M. P. DeLisi, L. A. Mawn, R. L. Galloway, and B. A. Landman, "Robust optic nerve segmentation on clinically acquired ct," in *SPIE Medical Imaging*, pp. 90341G–90341G, International Society for Optics and Photonics, 2014.

[36] J. H. Noble and B. M. Dawant, "An atlas-navigated optimal medial axis and deformable model algorithm (NOMAD) for the segmentation of the optic nerves and chiasm in MR and CT images," *Medical image analysis*, vol. 15, no. 6, pp. 877–884, 2011.

[37] G. Bekes, E. Mt, L. G. Nyl, A. Kuba, and M. Fidrich, "Geometrical model-based segmentation of the organs of sight on CT images," *Medical Physics*, vol. 35, no. 2, pp. 735–743, 2008.

[38] O. Puonti, J. E. Iglesias, and K. Van Leemput, "Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling," *NeuroImage*, vol. 143, pp. 235–249, 2016.

[39] K. Van Leemput, "Encoding probabilistic brain atlases using Bayesian inference," *IEEE Transactions on Medical Imaging*, vol. 28, no. 6, pp. 822–837, 2009.

[40] J. E. Iglesias, J. C. Augustinack, K. Nguyen, C. M. Player, A. Player, M. Wright, N. Roy, M. P. Frosch, A. C. McKee, L. L. Wald, *et al.*, "A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI," *Neuroimage*, vol. 115, pp. 117–137, 2015.

[41] J. E. Iglesias, K. Van Leemput, P. Bhatt, C. Casillas, S. Dutt, N. Schuff, D. Truran-Sacrey, A. Boxer, B. Fischl, A. D. N. Initiative, *et al.*, "Bayesian segmentation of brainstem structures in MRI," *NeuroImage*, vol. 113, pp. 184–195, 2015.

[42] M. Agn, O. Puonti, P. Munck af Rosenschöld, I. Law, and K. Van Leemput, "Brain Tumor Segmentation Using a Generative Model with an RBM Prior on Tumor Shape," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pp. 168–180, Springer, 2016.

[43] M. Agn, I. Law, P. M. af Rosenschöld, and K. Van Leemput, "A generative model for segmentation of tumor and organs-at-risk for radiation therapy planning of glioblastoma patients," in *SPIE Medical Imaging*, pp. 97841D–97841D, International Society for Optics and Photonics, 2016.

[44] A. Fischer and C. Igel, "Training restricted Boltzmann machines: An introduction," *Pattern Recognition*, vol. 47, no. 1, pp. 25–39, 2014.

[45] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng, "Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the ACM*, vol. 54, no. 10, pp. 95–103, 2011.

[46] K. P. Murphy, *Machine learning: a probabilistic perspective*. MIT press, 2012.

[47] J. E. Iglesias, E. Konukoglu, D. Zikic, B. Glocker, K. Van Leemput, and B. Fischl, "Is synthesizing MRI contrast useful for inter-modality analysis?," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 631–638, Springer, 2013.

[48] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain. Technical report," 1994.

[49] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38, 1977.

[50] V. S. Caviness, P. A. Filipek, and D. N. Kennedy, "Magnetic resonance technology in human brain science: blueprint for a program based upon morphometry," *Brain and Development*, vol. 11, no. 1, pp. 1–13, 1989.

[51] V. S. Caviness, J. Meyer, N. Makris, and D. N. Kennedy, "MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability," *Journal of Cognitive Neuroscience*, vol. 8, no. 6, pp. 566–587, 1996.

[52] D. N. Kennedy, P. A. Filipek, and V. S. Caviness, "Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging," *IEEE Transactions on Medical Imaging*, vol. 8, no. 1, pp. 1–7, 1989.

[53] K. Cho, T. Raiko, and A. Ilin, "Enhanced gradient for training restricted Boltzmann machines," *Neural computation*, vol. 25, no. 3, pp. 805–831, 2013.

[54] J. Melchior, A. Fischer, N. Wang, and L. Wiskott, "How to center binary restricted Boltzmann machines," *arXiv preprint arXiv:1311.1354*, 2013.

[55] E. DAgostino, F. Maes, D. Vandermeulen, and P. Suetens, "Non-rigid atlas-to-image registration by minimization of class-conditional image entropy," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 745–753, Springer, 2004.

[56] M. Kistler, S. Bonaretti, M. Pfahrer, R. Niklaus, and P. Büchler, "The virtual skeleton database: an open access repository for biomedical research and collaboration," *Journal of medical Internet research*, vol. 15, no. 11, 2013.

[57] D. A. Van Dyk and T. Park, "Partially collapsed Gibbs samplers: Theory and methods," *Journal of the American Statistical Association*, vol. 103, no. 482, pp. 790–796, 2008.

Chapter 10

Paper D

An Ensemble of 2D Convolutional Neural Networks for Tumor Segmentation

Mark Lyksborg( ), Oula Puonti, Mikael Agn, and Rasmus Larsen

Department of Applied Mathematics and Computer Science,
Technical University of Denmark, Kgs. Lyngby, Denmark
{mlyk,oupu,miag,rlar}@dtu.dk

Abstract. Accurate tumor segmentation plays an important role in radiosurgery planning and the assessment of radiotherapy treatment efficacy. In this paper we propose a method combining an ensemble of 2D convolutional neural networks for doing a volumetric segmentation of magnetic resonance images. The segmentation is done in three steps; first the full tumor region, is segmented from the background by a voxel-wise merging of the decisions of three networks learned from three orthogonal planes, next the segmentation is refined using a cellular automaton-based seed growing method known as growcut. Finally, within-tumor sub-regions are segmented using an additional ensemble of networks trained for the task. We demonstrate the method on the MICCAI Brain Tumor Segmentation Challenge dataset of 2014, and show improved segmentation accuracy compared to an axially trained 2D network and an ensemble segmentation without growcut. We further obtain competitive Dice scores compared with the most recent tumor segmentation challenge.

Keywords: Tumor segmentation · Convolutional neural network · Ensemble classification · Cellular automaton

1 Introduction

Segmentation of brain tumors plays a role in radiosurgery, radiotherapy planning, and for monitoring tumor growth. Segmentation is challenging since tumor location and appearance vary greatly between patients.

Many successful method for doing voxel-based segmentation are based on the random forest (RF) classification scheme which predicts segmentation labels from user engineered image features. Tustison et al. [15] proposed a two-stage RF approach, with features derived from a Gaussian mixture model followed by a Markov random field segmentation smoothing. The RF was also used by Reza et al. [12] who designed features using textons and multifractional Brownian motion. Menze et al. [10] proposed a generative probabilistic atlas-based model which adapts to the intensity distribution of different subjects and later combined it with the RF classifier [9]. An example of a successfull method that does not use a RF classifier is the patch-based approach [2]. Here voxels are

R.R. Paulsen and K.S. Pedersen (Eds.): SCIA 2015, LNCS 9127, pp. 201–211, 2015.
DOI: 10.1007/978-3-319-19665-7_17

segmented by comparing image patches to a dictionary consisting of training patches where the corresponding expert labels are used for segmentation.

In recent years and due to advancements in computational power, deep neural networks have been revived. In the most recent Brain Tumor Segmentation Challenge 2014 (BraTS2014), this was reflected by a number of contributions using deep neural networks. The work by Davy et al. [3] presented a 2D convolutional network trained from an axial perspective. Two others presented 3D networks [16], [18], and while their implementations differed, the results indicated a benefit of using 3D information. An important property of a network is that it learns image features relevant for the specific segmentation problem. This alleviate researchers from having to engineer such features.

We revisit the idea of Davy et al. [3] but instead of using one 2D network to do voxel-based segmentations, we learn an ensemble of networks, one for each of the axial, sagittal and coronal planes and fuse their segmentations into a more accurate 3D informed segmentation. Unlike previous works using convolutional networks we do not segment the tumor and its sub-regions using a single multi-label classifier. Instead, we split the problem into two sequential segmentation problems. The first segmentation separates tumor from healthy tissue and refine the segmentation using a growcut algorithm [17]. The second segmentation performs the within-tumor sub-region segmentation using the tumor mask of the first segmentation to select voxels of interest.

The method (Fig. 1) is demonstrated on the BraTS2014 dataset. We were able to achieve improved ground truth segmentation accuracy compared to a 2D axially trained network [3] and Dice scores [4] just below the top methods of the challenge leaderboard (https://www.virtualskeleton.ch/BRATS/Start2014).

2 Data

Two datasets were downloaded from the BraTS2014 website (November, 2014).

The first dataset (data1) consisted of 106 high grade glioma (HGG) and 25 low grade glioma (LGG) subjects (no longitudinal repetitions), all with ground truth segmentations of the tumors. It was randomly split into a training set of 76 HGG/15 LGG subjects, and the rest (30 HGG/10 LGG) were used as test data. For each subject, we used a set of multimodal magnetic resonance imaging (MRI) volumes, consisting of two T2-weighted images (Fluid-attenuated inversion recovery (FLAIR) and (T2)) and a T1-weighted image with gadolinium contrast (T1c). The MRIs were skull stripped, rigidly oriented according to MNI space and re-sliced to 1 mm^3 as described in [6]. The ground truth segmentation consisted of five labels (background=0, necrosis=1, edema=2, non-enhancing=3, enhancing=4).

The second dataset (data2) consisted of 187 multi-modal MRI volumes from 88 different subjects with 99 longitudinal repetitions. Since only the BraTS2014 challenge organizers know the ground truth segmentations, it allowed for a blinded segmentation evaluation via the challenge website.

3 Method

The proposed method, outlined in Fig. 1, consists of four steps. First, the MRI volumes are bias corrected for scanner field inhomogeneity and standardized to similar cross subject intensities. Second, an ensemble of convolutional networks segments the tumor from healthy tissue. The third step (growcut) post processes the segmentation to improve the segmentation. The fourth step does the within-tumor segmentation using an additional ensemble of networks. The four steps of the method are detailed successively in section 3.1-3.4.

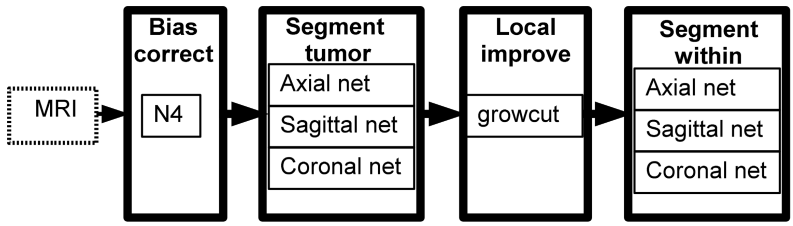


Fig. 1. Shows a schematic, outlining the pipeline of our method. The multi-modal MRI data is pushed through four successive stages of 1) bias correction, 2) whole tumor segmentation (tumor vs. none tumor), 3) localized post-processing of the segmentation and 4) a within-tumor segmentation stage.

3.1 Bias Correction and Standardization

MRI generally exhibits large intensity variations even within the same tissue type of a subject, largely due to field inhomogeneity of the scanner. To minimize this bias, the N4 method [14] was applied to each MRI.The N4 method works under the assumption that the bias field can be modeled by a smooth multiplicative model which is fitted iteratively to maximize the high frequency content of the MRI intensity distribution. To further standardize across different scanners, the maximum peak of each MRI intensity histogram was found, and the intensities scaled according to $I = I_c \cdot (I_b/I_p)$, where I_c is the N4 bias corrected image volume, I_p is the maximum peak intensity of I_c and I_b is a reference value which we fixed to $I_b = 200$. To achieve equal importance of the multi-modal MRI, their intensities were further standardised using a normal transformation applied to each of the different modalities.

3.2 Convolutional Network Ensemble: Whole Tumor

To segment tumor tissue, three convolutional neural networks were trained using a multi-modal image patch of dimension 46×46. Each 2D network learned to classify the same center voxel but viewed from an axial, sagittal and coronal perspective. Combining this ensemble of 2D networks enabled the segmentation method to become 3D aware.

The 2D networks are described by the architecture in Fig. 2. It shows a network consisting of 6 layers. Each perform an algebraic operation on the input data x and passes the result as input to the next layer. The process is repeated until reaching layer 6 which predicts the most probable classification label.

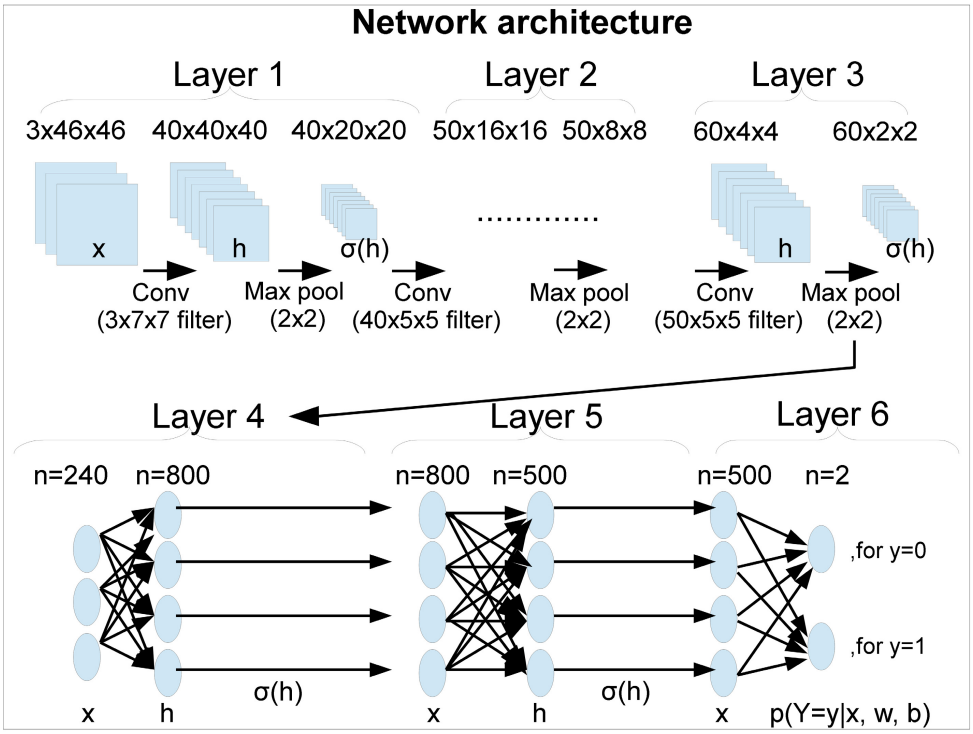


Fig. 2. Depicts a 2D deep neural network architecture consisting of six layers. The first three are convolutional layers, followed by two fully connected layers and a softmax layer where the arrows indicate the connections between layers. The squares illustrate the 2D nature of the input (x) and the intermediate representations (h) of the convolutional layers, where $x = [x_1...x_n]$ is a 3D matrix of n input patches and $h = [h_1...h_m]$, is the concatenation of m 2D filter response. The circles of the fully connected layers indicate its 1D nature with n being the number of neurons (=the circles), such that $x = [x_1...x_n]^T$ and $h = [h_1...h_n]^T$ are the 1D vector representations of the input and the neuronal activations.

Convolutional layers: The convolutional layers apply filtering and downsampling operations to image patches. The first layer uses a filter bank of size $40 \times 3 \times 7 \times 7$ which it applies to the $3 \times 46 \times 46$ image patch. This produces a feature map h of size $40 \times 40 \times 40$, where the first dimension indexes the feature maps, while the second and third dimensions indexes (row, column) coordinates. More specifically the j^{th} map is calculated by $h_j = b_j + \sum_{i=1}^{n}(w_{ij} * x_i)$, where i

indexes the input channel and a trainable filter w_{ij}, the $*$ operator denotes 2D convolution and $n = 3$ is the number of input channels. Subsequently a 2×2 max pooling strategy is used to downsample h to size $40 \times 20 \times 20$ and the rectified linear unit function, $\sigma(h) = max(0, h)$ is applied. The remaining convolutional layers (two and three) perform the same type of operations but using filter banks of size $50 \times 40 \times 5 \times 5$ and $60 \times 50 \times 5 \times 5$ for the respective layers. The application of these filters and downsampling steps result in a number of the intermediate feature maps with the dimensionalities listed in the top part of Fig. 2.

Fully connected layers: Layer 4, 5 and 6 are fully connected layers meaning each neuron is exposed to the full input x of the previoues layer. Each of the 800 neurons in layer 4, evaluates the product $h_j = w_j^T x + b_j$ and applies the non-linear activation function $\sigma(h_j)$. Thereby transforming the 240 dimensional vector x into an 800 dimensional vector $\sigma(h)$ which is passed to layer 5. Layer 5 works similar to layer 4, but now generating a 500 dimensional feature vector $\sigma(h)$ which is propagated to layer 6. Layer 6 evaluates the softmax function

$$p(Y = y|x, w, b) = \frac{e^{w_y x + b_y}}{\sum_j e^{w_j x + b_j}}, \tag{1}$$

generating posterior probabilities for a number of classification labels, $y = \{0, 1\}$. Here w_j refer to a vector of linear parameters for the j^{th} class, b_j is a bias weight and x is the 500 dimensional response vector from the previous layer.

Network Training Each of the 2D networks were trained by minimizing the following cost function

$$C(W, B) = \frac{1}{nd} \cdot \sum_{i=1}^{nd} -\ln(p(Y = y^i|x^i, W, B)) + \lambda \cdot \sum_{j=1}^{nw} W_j^2. \tag{2}$$

The first term of eq. (2) is the mean negative log-likelihood of the softmax probability and we have used capitalized (W, B) to indicate that it is a function of (w, b) parameters from different types of layers. Further, the training patches are denoted x^i, y^i, corresponding to the patch intensities and ground truth label of the i^{th} training example. The second term of eq. (2) is a regularization term that adds robustness to the optimization problem by limiting the solution space to models with smaller parameter weights. It does so by penalizing the 2-norm of the parameters and through experimentation we found $\lambda = 0.0001$ to be suitable.

The cost function was minimized using a stochastic gradient descent (SGD) which relied on the back propagation algorithm to estimate gradients. The SGD performed iterative updates based on gradients estimated from mini-batches with a batch size of 200 where an update occurred after each mini-batch. Each gradient update was further augmented by a moment based learning rule [13] which updated the parameters as a weighted combination of the current gradients and the gradients of previous iteration update. We used a momentum coefficient of 0.9. Layer 4 and 5 were trained using the dropout learning [5]

(dropout rate=0.5) which activates half the neurons for each training example. As a consequences the activations of these layers($\sigma(h)$) were divided by 2 when a network was applied to an unseen test image patch.

A GPU implementation for training the three 2D networks was achived using Theano [1].

Network Ensemble Merging Having learned the parameters of the three networks, their complementary decision information were merged. This was done using the posterior probablities of the last layer (layer 6). If the networks agreed on the same label we were highly confident in this classification and assigned the label of voxel x with probability $p(Y|x) = 1$. Otherwise a majority vote decided the class label and the probability was set to reflect this uncertainty by averaging the class probabilities of the three networks, $p(Y|x) = (1/3)\sum_{i=1}^{3} p_i(Y|x, w, b)$. The resulting label segmentations and their probabilities were then used as input for the growcut algorithm.

3.3 Cellular Automaton: Growcut

The growcut algorithm was initially proposed as a continuous state cellular automata method for automated segmentation based on user labeled seed voxels [17]. From these labels and a local intensity transition rule the algorithm decides whether voxels should be re-labelled.

We used the algorithmic formulation of [17] which we extended to 3D. The algorithm models each voxel as a cell with a state set $S(\Theta, l, C)$ consisting of a strength value $\Theta \in [0, 1]$, a label l and an intensity feature vector C. It is an iterative algorithm and for each iteration the strength and labels of the previous iteration remain fixed. During an iteration each image cell r is attacked by its neighboring cells $s \in N(r)$ where $N(r)$ denote the $3 \times 3 \times 3$ neighborhood of a volume and only if $g(C_r, C_s) \cdot \Theta_s > \Theta_r$, will Θ_r, and l_r be updated before the next iteration. The local transition rule is given by

$$g(C_1, C_2) = 1 - \frac{||C_1 - C_2||_2}{k} \tag{3}$$

Where we have normalized the intensities of C to be in the range [0, 1] such that for $k = \sqrt{3}$, the value of $g(C_1, C_2) \in [0, 1]$. Since $g(C_1, C_2)$ can never exceed 1, any cells with strength $\Theta = 1$ will remain constant throughout the algorithm.

To use the growcut on the ensemble segmentations, the feature vector C was set to the multi-modal MRI intensities and the values of l, Θ were initialized with the labels and probability maps of the convolutional network ensemble. This initialization served as a strong prior for growcut segmentation, assuming that the segmentation was already near optimal.

Once growcut converged to a stable segmentation (100 iterations), a heuristic rule was used to identify the tumor. It was based on a connected components analysis to remove any spatially coherent clusters of voxels which were less than 80% of the biggest cluster.

3.4 Convolutional Network Ensemble: Within-Tumor

This ensemble of convolutional networks was used to segment the within-tumor sub-regions. The architecture of each network is similar to the previously described, but considers a smaller image patch and has only two convolutional layers, two fully connected dropout layers and softmax probability layer. The input patch size is $3 \times 34 \times 34$ and the first convolutional layer uses a filter bank of size $50 \times 3 \times 7 \times 7$ while the second one uses a filter bank of size $60 \times 50 \times 5 \times 5$. The justification of choosing a smaller patch size is that the within-tumor segmentation uses information on a smaller scale compared to the whole tumor segmentation. As with the previously described networks, the fully connected layers use 800 and 500 neurons respectively while the softmax layer, predicts one of four possible classification labels. The SGD optimization was again used to train the networks but for these specific networks we used $\lambda = 0.00005$.

Network Ensemble Merging The voxel-based decisions of the ensemble of axial, sagittal and coronal networks were either set to the label they all agree on, or according to the most probable average probability of the softmax probability.

4 Results

4.1 Test and Phenotype Performance

Testing our method on the 40 left out subjects (data1), resulted in the segmentation performances of Table 1. This table shows ground truth scores for three methods; A 2D convolutional network applied to the axial plane similar to [3], a method using only the ensemble part of our method (ensem) and our full method which is ensem in combination with growcut (ensem+grow). The scores of the table are given for pathologically relevant tumor regions. These are the whole tumor (labels: necrosis, edema, non-enhancing, enhancing), the enhanced tumor region and the tumor core (labels: necrosis, non-enhancing, enhancing). We see that using an ensemble improved the segmentation relative to a 2D network and achieved further improvement by including growcut post-processing. As a visual comparison example, two tumor segmentations based on our method and their

Table 1. Average segmentation performance scores of three convolutional neural network methods evaluated on 40 subjects of data1. The scores (Dice, positive predictive and sensitivity) were calculated for the different tumor regions.

Method	Dice scores			Positive predictive			Sensitivity		
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.
axial	0.744	0.642	0.629	0.732	0.624	0.642	0.811	0.746	0.707
ensem	0.786	0.686	0.676	0.786	0.707	0.693	0.825	0.743	0.717
ensem+grow	0.810	0.697	0.681	0.833	0.718	0.701	0.825	0.750	0.720

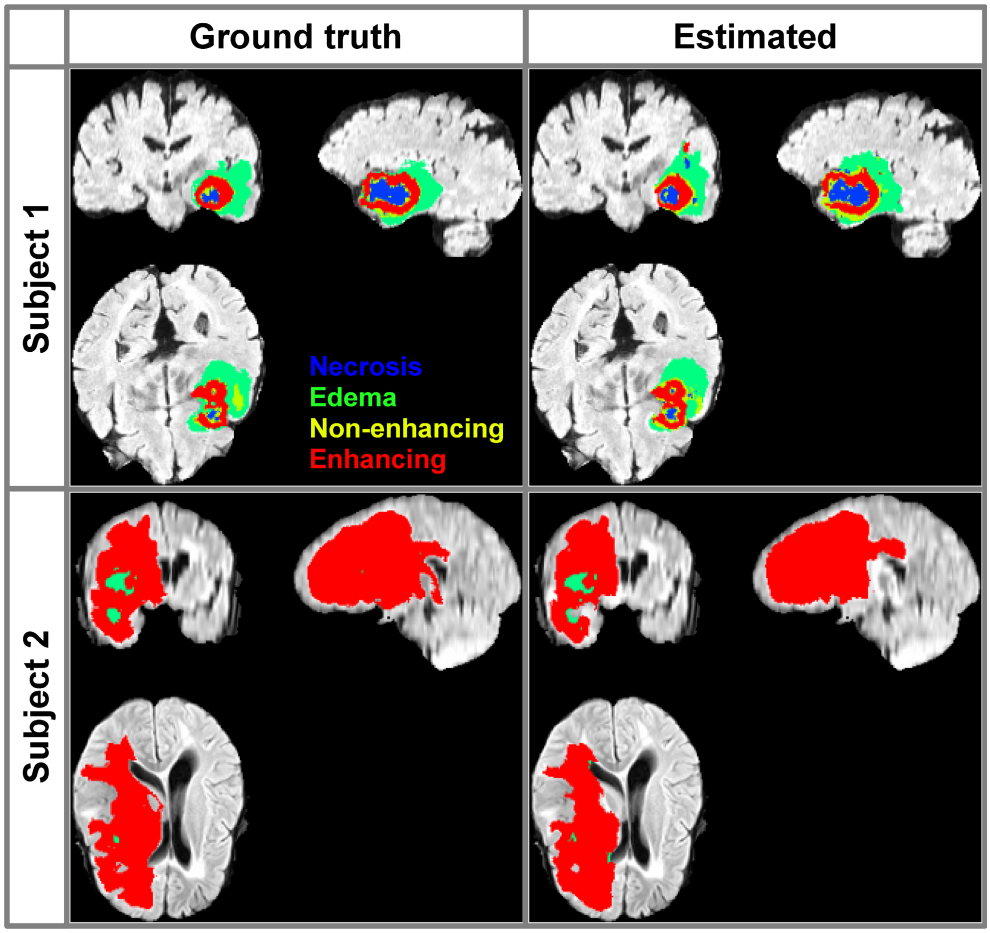


Fig. 3. This visual comparison shows both the proposed segmentation method and corresponding ground truth for two subjects. The Dice scores of subject 1 were 0.825 (whole), 0.795 (core) and 0.842 (enhanced) and for subject 2 they were, 0.892 (whole), 0.840 (core) and 0.854 (enhanced).

ground truth, are shown in Fig. 3. By dividing the test subjects based on tumor types (HGG/LGG), we evaluated their impact on method performance. This comparison (Fig. 4), reveals higher Dice scores with less variance for the HGGs, indicating a methodological bias towards the tumor type.

4.2 Blinded Challenge Performance

Testing our method on the blinded challenge dataset previously denoted data2 and performing an on-line evaluation of the segmentations, resulted in the average performance scores of Table 2. It lists the scores for the first time point of the 99 subjects (cross sectional) and the full challenge data (full data) where similar performances are achieved. It also includes the top 3 scores of the BraTS2014 challenge where our method is ranked amongst.

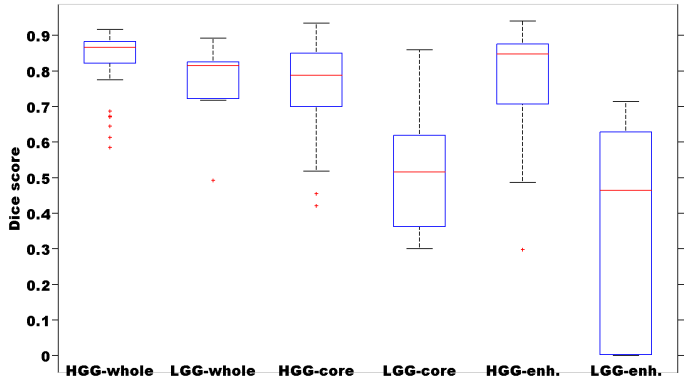


Fig. 4. Ground truth Dice scores performance for two different types of tumors (HGG and LGG). Red line indicate mean Dice score, blue boxes show the 25 and 75 percentiles of the scores while extreme observations are show with red dots.

Table 2. Shows the average segmentation performance scores of our method in grey (cross sectional and full data), for the BraTS2014 challenge data (data2). Also listed are the top three of the challenge (15/12-2014), ranked according to their whole tumor Dice scores. These are Urbag [16], Kleej [7], Dvorp [8].

Method	Dice scores			Positive predictive			Sensitivity		
	Whole	Core	Enh.	Whole	Core	Enh.	Whole	Core	Enh.
Cross sectional	0.801	0.637	0.586	0.803	0.682	0.554	0.857	0.715	0.745
Full data	0.799	0.631	0.625	0.783	0.629	0.580	0.861	0.736	0.776
Urbag	0.87	0.76	0.72	0.91	0.80	0.69	0.85	0.76	0.81
Kleej	0.87	0.76	0.73	0.90	0.73	0.66	0.85	0.83	0.87
Dvorp	0.60	0.30	0.29	0.86	0.58	0.56	0.53	0.27	0.28

5 Discussion

We have presented a method, combining an ensemble of 2D convolutional networks with the growcut method for making a 3D informed segmentation. It showed improved accuracy compared to a 2D network and an ensemble segmentation without growcut thereby validating the usefulness of the proposed method. The investigation of tumor type showed better performance for HGG, likely due to the imbalanced training data distribution (76 HGG/15 LGG). It could also indicate the presence of a measurable pathologic difference. If so, the training of a segmentation method for each type could lead to improved segmentations for both types. This would require knowing the tumor type in advance, information that was not readily available for the blinded challenge data. Our challenge results showed a nice performance although sub-par to the top two methods of the challenge but was superior to the remaining 11. It is noted that our methods performance is in the Dice score range that manual annotators can achieve according the results of [11]. They reported the Dice accuracy of

annotators to be in the range of (0.74-0.85). This is comparable to the proposed method. A simple strategy for improving our work would be to extend the ensemble to use 3D network (computationally costly) or to investigate the inclusion of networks trained from more than orthogonal planes. In addition, the usage of using longitudinal information could also play a role towards improving segmentations.

References

1. Bergstra, J., et al.: Theano: a CPU and GPU math expression compiler. In: Python for Scientific Computing Conference (SciPy) (2010)
2. Cordier, N., Menze, B., Delingette, H., Ayache, N.: Patch-based segmentation of brain tissues. In: MICCAI-BraTS (Challenge on Multimodal Brain Tumor Segmentation), pp. 6–17 (2013)
3. Davy, A., Havaei, M., Warde-Farley, D., Biard, A., Tran, L., Jodoin, P.M., Courville, A., Larochelle, H., Pal, C., Bengio, Y.: Brain tumor segmentation with deep neural networks. In: MICCAI-BraTS, pp. 1–5 (2014)
4. Dice, L.R.: Measures of the amount of ecologic association between species. Ecology (1945)
5. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors. CoRR (2012)
6. Jakab, A.: Segmenting brain tumors with the slicer 3d software. Tech. rep., University of Debrecen / ETH Zürich (2012)
7. Kleesiek, J., Biller, A., Urban, G., Kothe, U., Bendszus, M., Hamprecht, F.: Ilastik for multi-modal brain tumor segmentation. In: MICCAI-BraTS, pp. 12–17 (2014)
8. Kwon, D., Akbari, H., Da, X., Gaonkar, B., Davatzikos, C.: Multimodal brain tumor image segmentation using glistr. In: MICCAI-BraTS, pp. 18–19 (2014)
9. Menze, B., Geremia, E., Ayache, N., Szekely, G.: Segmenting glioma in multimodal images using a generative-discriminative model for brain lesion segmentation. In: MICCAI-BraTS, pp. 56–63 (2012)
10. Menze, B., Leemput, K.V., Lashkar, D., Weber, M., Ayache, N., Golland, P.: Segmenting glioma in multi-modal images using a generative model for brain lesion segmentation, pp. 49–55 (2012)
11. Menze, B.H., et al.: The multimodal brain tumor image segmentation benchmark (brats). IEEE Transactions on Medical Imaging (2014)
12. Reza, S., Iftekharuddin, K.: Improved brain tumor tissue segmentation using texture features. In: MICCAI-BraTS, pp. 27–30 (2014)
13. Sutskever, I., Martens, J., Dahl, G.E., Hinton, G.E.: On the importance of initialization and momentum in deep learning. In: 30th International Conference on Machine Learning (ICML 2013), vol. 28, pp. 1139–1147, May 2013
14. Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C.: N4ITK: Improved N3 Bias Correction. IEEE Trans. Med. Imaging **29**(6), 1310–1320 (2010)
15. Tustison, N., Wintermark, M., Durst, C., Avants, B.: Ants and arboles. In: MICCAI-BraTS, pp. 47–50 (2013)

16. Urban, G., Bendszus, M., Hamprecht, F.A., Kleesiek, J.: Multi-modal brain tumor segmentation using deep convolutional neural networks. In: MICCAI-BraTSs, pp. 31–35 (2014)
17. Vezhnevets, V., Konouchine, V.: GrowCut - interactive multi-label n-d image segmentation by cellular automata. In: Proceedings of Graphicon (2005)
18. Zikic, D., Ioannou, Y., Brown, M., Criminisi, A.: Segmentation of brain tumor tissues with convolutional neural networks. In: MICCAI-BraTS, pp. 36–39 (2014)

Bibliography

[Andreasen et al., 2015] Andreasen, D., Van Leemput, K., Hansen, R. H., Andersen, J. A., and Edmund, J. M. (2015). Patch-based generation of a pseudo CT from conventional MRI sequences for MRI-only radiotherapy of the brain. *Medical physics*, 42(4):1596–1605.

[Ashburner et al., 2000] Ashburner, J., Andersson, J. L., and Friston, K. J. (2000). Image registration using a symmetric prior—in three dimensions. *Human brain mapping*, 9(4):212–225.

[Ashburner and Friston, 1997] Ashburner, J. and Friston, K. (1997). Multimodal image coregistration and partitioning—a unified framework. *Neuroimage*, 6(3):209–217.

[Ashburner and Friston, 2005] Ashburner, J. and Friston, K. J. (2005). Unified segmentation. *Neuroimage*, 26(3):839–851.

[Asman and Landman, 2013] Asman, A. J. and Landman, B. A. (2013). Non-local statistical label fusion for multi-atlas segmentation. *Medical image analysis*, 17(2):194–208.

[Bakas et al., 2016] Bakas, S., Zeng, K., Sotiras, A., Rathore, S., Akbari, H., Gaonkar, B., Rozycki, M., Pati, S., and Davatzikos, C. (2016). GLISTRboost: Combining Multimodal MRI Segmentation, Registration, and Biophysical Tumor Growth Modeling with Gradient Boosting Machines for Glioma Segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pages 144–155. Springer.

[Bauer et al., 2013] Bauer, S., Lu, H., May, C. P., Nolte, L.-P., Büchler, P., and Reyes, M. (2013). Integrated segmentation of brain tumor images for radiotherapy and neurosurgery. *International Journal of Imaging Systems and Technology*, 23(1):59–63.

[Bauer et al., 2011] Bauer, S., Nolte, L.-P., and Reyes, M. (2011). Fully automatic segmentation of brain tumor images using support vector machine classification in combination with hierarchical conditional random field regularization. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 354–361. Springer.

[Bekes et al., 2008] Bekes, G., Máté, E., Nyúl, L. G., Kuba, A., and Fidrich, M. (2008). Geometrical model-based segmentation of the organs of sight on CT images. *Medical Physics*, 35(2):735–743.

[Bengio et al., 2013] Bengio, Y., Yao, L., Alain, G., and Vincent, P. (2013). Generalized denoising auto-encoders as generative models. In *Advances in Neural Information Processing Systems*, pages 899–907.

[Burnet et al., 2004] Burnet, N. G., Thomas, S. J., Burton, K. E., and Jefferies, S. J. (2004). Defining the tumour and target volumes for radiotherapy. *Cancer Imaging*, 4(2):153–161.

[Caviness et al., 1989] Caviness, V. S., Filipek, P. A., and Kennedy, D. N. (1989). Magnetic resonance technology in human brain science: blueprint for a program based upon morphometry. *Brain and Development*, 11(1):1–13.

[Caviness et al., 1996] Caviness, V. S., Meyer, J., Makris, N., and Kennedy, D. N. (1996). MRI-based topographic parcellation of human neocortex: an anatomically specified method with estimate of reliability. *Journal of Cognitive Neuroscience*, 8(6):566–587.

[Cedric and Tang, 2011] Cedric, X. Y. and Tang, G. (2011). Intensity-modulated arc therapy: principles, technologies and clinical implementation. *Physics in medicine and biology*, 56(5):R31.

[Chen and Dawant, 2015] Chen, A. and Dawant, B. (2015). A Multi-atlas Approach for the Automatic Segmentation of Multiple Structures in Head and Neck CT Images. *The MIDAS Journal - Head and Neck Auto Segmentation Challenge. Identifier: http://hdl.handle.net/10380/3540.*

[Chinot et al., 2014] Chinot, O. L., Wick, W., Mason, W., Henriksson, R., Saran, F., Nishikawa, R., Carpentier, A. F., Hoang-Xuan, K., Kavan, P., Cernea, D., et al. (2014). Bevacizumab plus radiotherapy–temozolomide for newly diagnosed glioblastoma. *New England Journal of Medicine*, 370(8):709–722.

[Cho et al., 2013] Cho, K., Raiko, T., and Ilin, A. (2013). Enhanced gradient for training restricted Boltzmann machines. *Neural computation*, 25(3):805–831.

[Conson et al., 2014] Conson, M., Cella, L., Pacelli, R., Comerci, M., Liuzzi, R., Salvatore, M., and Quarantelli, M. (2014). Automated delineation of brain structures in patients undergoing radiotherapy for primary brain tumors: From atlas to dose–volume histograms. *Radiotherapy and Oncology*, 112(3):326–331.

[Cordier et al., 2016] Cordier, N., Delingette, H., and Ayache, N. (2016). A patch-based approach for the segmentation of pathologies: Application to glioma labelling. *IEEE transactions on medical imaging*, 35(4):1066–1076.

[Coupé et al., 2011] Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., and Collins, D. L. (2011). Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage*, 54(2):940–954.

[Cuadra et al., 2004] Cuadra, M. B., Pollo, C., Bardera, A., Cuisenaire, O., Villemure, J.-G., and Thiran, J.-P. (2004). Atlas-based segmentation of pathological MR brain images using a model of lesion growth. *IEEE transactions on medical imaging*, 23(10):1301–1314.

[Deeley et al., 2011] Deeley, M., Chen, A., Datteri, R., Noble, J., Cmelak, A., Donnelly, E., Malcolm, A., Moretti, L., Jaboin, J., Niermann, K., et al. (2011). Comparison of manual and automatic segmentation methods for brain structures in the presence of space-occupying lesions: a multi-expert study. *Physics in medicine and biology*, 56(14):4557.

[Dempster et al., 1977] Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society. Series B (methodological)*, pages 1–38.

[Dolz et al., 2016a] Dolz, J., Betrouni, N., Quidet, M., Kharroubi, D., Leroy, H. A., Reyns, N., Massoptier, L., and Vermandel, M. (2016a). Stacking denoising auto-encoders in a deep network to segment the brainstem on MRI in brain cancer patients: A clinical study. *Computerized Medical Imaging and Graphics*, 52:8–18.

[Dolz et al., 2016b] Dolz, J., Laprie, A., Ken, S., Leroy, H.-A., Reyns, N., Massoptier, L., and Vermandel, M. (2016b). Supervised machine learning-based classification scheme to segment the brainstem on MRI in multicenter brain tumor treatment context. *International journal of computer assisted radiology and surgery*, 11(1):43–51.

[Dolz et al., 2015a] Dolz, J., Leroy, H.-A., Reyns, N., Massoptier, L., and Vermandel, M. (2015a). A Fast and Fully Automated Approach to Segment

Optic Nerves on MRI and its application to radiosurgery. In *2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI)*, pages 1102–1105. IEEE.

[Dolz et al., 2015b] Dolz, J., Massoptier, L., and Vermandel, M. (2015b). Segmentation algorithms of subcortical brain structures on MRI for radiotherapy and radiosurgery: a survey. *IRBM*, 36(4):200–212.

[Fennema-Notestine et al., 2006] Fennema-Notestine, C., Ozyurt, I. B., Clark, C. P., Morris, S., Bischoff-Grethe, A., Bondi, M. W., Jernigan, T. L., Fischl, B., Segonne, F., Shattuck, D. W., et al. (2006). Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location. *Human brain mapping*, 27(2):99–113.

[Fischer and Igel, 2014] Fischer, A. and Igel, C. (2014). Training restricted Boltzmann machines: An introduction. *Pattern Recognition*, 47(1):25–39.

[Fischl, 2012] Fischl, B. (2012). FreeSurfer. *Neuroimage*, 62(2):774–781.

[Fischl et al., 2002] Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., Van Der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., et al. (2002). Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron*, 33(3):341–355.

[Gooya et al., 2012] Gooya, A., Pohl, K. M., Bilello, M., Cirillo, L., Biros, G., Melhem, E. R., and Davatzikos, C. (2012). GLISTR: glioma image segmentation and registration. *IEEE transactions on medical imaging*, 31(10):1941–1954.

[Havaei et al., 2016] Havaei, M., Dutil, F., Pal, C., Larochelle, H., and Jodoin, P.-M. (2016). A convolutional neural network approach to brain tumor segmentation. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pages 195–208. Springer.

[Hinton, 2002] Hinton, G. E. (2002). Training products of experts by minimizing contrastive divergence. *Neural computation*, 14(8):1771–1800.

[Hinton, 2012] Hinton, G. E. (2012). A practical guide to training restricted boltzmann machines. In *Neural Networks: Tricks of the Trade*, pages 599–619. Springer.

[Ibragimov and Xing, 2016] Ibragimov, B. and Xing, L. (2016). Segmentation of organs-at-risks in head and neck CT images using convolutional neural networks. *Medical Physics*.

[Iglesias et al., 2015a] Iglesias, J. E., Augustinack, J. C., Nguyen, K., Player, C. M., Player, A., Wright, M., Roy, N., Frosch, M. P., McKee, A. C., Wald, L. L., et al. (2015a). A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: application to adaptive segmentation of in vivo MRI. *Neuroimage*, 115:117–137.

[Iglesias et al., 2013a] Iglesias, J. E., Konukoglu, E., Zikic, D., Glocker, B., Van Leemput, K., and Fischl, B. (2013a). Is synthesizing MRI contrast useful for inter-modality analysis? In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 631–638. Springer.

[Iglesias et al., 2015b] Iglesias, J. E., Sabuncu, M. R., Aganj, I., Bhatt, P., Casillas, C., Salat, D., Boxer, A., Fischl, B., and Van Leemput, K. (2015b). An algorithm for optimal fusion of atlases with different labeling protocols. *NeuroImage*, 106:451–463.

[Iglesias et al., 2013b] Iglesias, J. E., Sabuncu, M. R., Van Leemput, K., Initiative, A. D. N., et al. (2013b). Improved inference in Bayesian segmentation using Monte Carlo sampling: Application to hippocampal subfield volumetry. *Medical image analysis*, 17(7):766–778.

[Iglesias et al., 2015c] Iglesias, J. E., Van Leemput, K., Bhatt, P., Casillas, C., Dutt, S., Schuff, N., Truran-Sacrey, D., Boxer, A., Fischl, B., Initiative, A. D. N., et al. (2015c). Bayesian segmentation of brainstem structures in MRI. *NeuroImage*, 113:184–195.

[Isambert et al., 2008] Isambert, A., Dhermain, F., Bidault, F., Commowick, O., Bondiau, P.-Y., Malandain, G., and Lefkopoulos, D. (2008). Evaluation of an atlas-based automatic segmentation software for the delineation of brain organs at risk in a radiation therapy clinical context. *Radiotherapy and oncology*, 87(1):93–99.

[Islam et al., 2013] Islam, A., Reza, S. M., and Iftekharuddin, K. M. (2013). Multifractal texture estimation for detection and segmentation of brain tumors. *IEEE transactions on biomedical engineering*, 60(11):3204–3215.

[Kamnitsas et al., 2016] Kamnitsas, K., Ledig, C., Newcombe, V. F., Simpson, J. P., Kane, A. D., Menon, D. K., Rueckert, D., and Glocker, B. (2016). Efficient Multi-Scale 3D CNN with Fully Connected CRF for Accurate Brain Lesion Segmentation. *arXiv preprint arXiv:1603.05959*.

[Kennedy et al., 1989] Kennedy, D. N., Filipek, P. A., and Caviness, V. S. (1989). Anatomic segmentation and volumetric calculations in nuclear magnetic resonance imaging. *IEEE Transactions on Medical Imaging*, 8(1):1–7.

[Kingma and Welling, 2013] Kingma, D. P. and Welling, M. (2013). Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*.

[Kistler et al., 2013] Kistler, M., Bonaretti, S., Pfahrer, M., Niklaus, R., and Büchler, P. (2013). The virtual skeleton database: an open access repository for biomedical research and collaboration. *Journal of medical Internet research*, 15(11).

[Kwon et al., 2014] Kwon, D., Shinohara, R. T., Akbari, H., and Davatzikos, C. (2014). Combining generative models for multifocal glioma segmentation and registration. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 763–770. Springer.

[Lagendijk et al., 2016] Lagendijk, J., Vulpen, M., and Raaymakers, B. (2016). The development of the MRI linac system for online MRI-guided radiotherapy: a clinical update. *Journal of internal medicine*.

[Larsen et al., 2014] Larsen, C. T., Iglesias, J. E., and Van Leemput, K. (2014). N3 bias field correction explained as a bayesian modeling method. In *Bayesian and grAphical Models for Biomedical Imaging*, pages 1–12. Springer.

[Lee et al., 2011] Lee, H., Grosse, R., Ranganath, R., and Ng, A. Y. (2011). Unsupervised learning of hierarchical representations with convolutional deep belief networks. *Communications of the ACM*, 54(10):95–103.

[Lyksborg et al., 2015] Lyksborg, M., Puonti, O., Agn, M., and Larsen, R. (2015). An ensemble of 2D convolutional neural networks for tumor segmentation. In *Scandinavian Conference on Image Analysis*, pages 201–211. Springer.

[Maier et al., 2016] Maier, O., Wilms, M., and Handels, H. (2016). Image Features for Brain Lesion Segmentation Using Random Forests. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pages 119–130. Springer.

[Melchior et al., 2013] Melchior, J., Fischer, A., Wang, N., and Wiskott, L. (2013). How to center binary restricted Boltzmann machines. *arXiv preprint arXiv:1311.1354*.

[Menze et al., 2015] Menze, B. H., Jakab, A., Bauer, S., Kalpathy-Cramer, J., Farahani, K., Kirby, J., Burren, Y., Porz, N., Slotboom, J., Wiest, R., et al. (2015). The multimodal brain tumor image segmentation benchmark (BRATS). *IEEE Transactions on Medical Imaging*, 34(10):1993–2024.

[Menze et al., 2016] Menze, B. H., Van Leemput, K., Lashkari, D., Riklin-Raviv, T., Geremia, E., Alberts, E., Gruber, P., Wegener, S., Weber, M.-A., Székely, G., et al. (2016). A Generative Probabilistic Model and Discriminative Extensions for Brain Lesion Segmentation—With Application to Tumor and Stroke. *IEEE transactions on medical imaging*, 35(4):933–946.

[Menze et al., 2010] Menze, B. H., Van Leemput, K., Lashkari, D., Weber, M.-A., Ayache, N., and Golland, P. (2010). A generative model for brain tumor segmentation in multi-modal images. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 151–159. Springer.

[Munck af Rosenschöld et al., 2015] Munck af Rosenschöld, P., Costa, J., Engelholm, S. A., Lundemann, M. J., Law, I., Ohlhues, L., and Engelholm, S. (2015). Impact of [18F]-fluoro-ethyl-tyrosine PET imaging on target definition for radiation therapy of high-grade glioma. *Neuro-oncology*, 17(5):757–763.

[Munck af Rosenschöld et al., 2011] Munck af Rosenschöld, P., Engelholm, S., Ohlhues, L., Law, I., Vogelius, I., and Engelholm, S. A. (2011). Photon and proton therapy planning comparison for malignant glioma based on CT, FDG-PET, DTI-MRI and fiber tracking. *Acta Oncologica*, 50(6):777–783.

[Noble and Dawant, 2011] Noble, J. H. and Dawant, B. M. (2011). An atlas-navigated optimal medial axis and deformable model algorithm (NOMAD) for the segmentation of the optic nerves and chiasm in MR and CT images. *Medical image analysis*, 15(6):877–884.

[Nyúl et al., 2000] Nyúl, L. G., Udupa, J. K., and Zhang, X. (2000). New variants of a method of MRI scale standardization. *IEEE transactions on medical imaging*, 19(2):143–150.

[Orbes Arteaga et al., 2015] Orbes Arteaga, M., Cárdenas Peña, D., and Castellanos Dominguez, G. (2015). A Multi-atlas Approach for the Automatic Segmentation of Multiple Structures in Head and Neck CT Images. *The MIDAS Journal - Head and Neck Auto Segmentation Challenge. Identifier: http://hdl.handle.net/10380/3538*.

[Panda et al., 2014] Panda, S., Asman, A. J., DeLisi, M. P., Mawn, L. A., Galloway, R. L., and Landman, B. A. (2014). Robust optic nerve segmentation on clinically acquired ct. In *SPIE Medical Imaging*, pages 90341G–90341G. International Society for Optics and Photonics.

[Pereira et al., 2016] Pereira, S., Pinto, A., Alves, V., and Silva, C. A. (2016). Deep convolutional neural networks for the segmentation of gliomas in multi-sequence MRI. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, volume 9556, pages 131–143. Lecture Notes in Computer Science, Springer.

[Pohl et al., 2006] Pohl, K. M., Fisher, J., Grimson, W. E. L., Kikinis, R., and Wells, W. M. (2006). A Bayesian model for joint segmentation and registration. *NeuroImage*, 31(1):228–239.

[Poulsen et al., 2016] Poulsen, S. H., Urup, T., Grunnet, K., Christensen, I. J., Larsen, V. A., Jensen, M. L., Munck af Rosenschöld, P., Poulsen, H. S., and Law, I. (2016). The prognostic value of FET PET at radiotherapy planning in newly diagnosed glioblastoma. *European Journal of Nuclear Medicine and Molecular Imaging*, pages 1–9.

[Preusser et al., 2011] Preusser, M., de Ribaupierre, S., Wöhrer, A., Erridge, S. C., Hegi, M., Weller, M., and Stupp, R. (2011). Current concepts and management of glioblastoma. *Annals of neurology*, 70(1):9–21.

[Puonti et al., 2016] Puonti, O., Iglesias, J. E., and Van Leemput, K. (2016). Fast and sequence-adaptive whole-brain segmentation using parametric Bayesian modeling. *NeuroImage*, 143:235–249.

[Puonti and Van Leemput, 2016] Puonti, O. and Van Leemput, K. (2016). Simultaneous Whole-Brain Segmentation and White Matter Lesion Detection Using Contrast-Adaptive Probabilistic Models. In *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries: First International Workshop, Brainles 2015, Held in Conjunction with MICCAI 2015, Munich, Germany, October 5, 2015, Revised Selected Papers*, pages 9–20. Springer.

[Roy et al., 2013] Roy, S., Carass, A., and Prince, J. L. (2013). Magnetic resonance image example-based contrast synthesis. *IEEE transactions on medical imaging*, 32(12):2348–2363.

[Sabuncu et al., 2010] Sabuncu, M. R., Yeo, B. T., Van Leemput, K., Fischl, B., and Golland, P. (2010). A generative model for image segmentation based on label fusion. *IEEE transactions on medical imaging*, 29(10):1714–1729.

[Salakhutdinov et al., 2007] Salakhutdinov, R., Mnih, A., and Hinton, G. (2007). Restricted Boltzmann machines for collaborative filtering. In *Proceedings of the 24th international conference on Machine learning*, pages 791–798. ACM.

[Sanjuán et al., 2013] Sanjuán, A., Price, C. J., Mancini, L., Josse, G., Grogan, A., Yamamoto, A. K., Geva, S., Leff, A. P., Yousry, T. A., and Seghier, M. L. (2013). Automated identification of brain tumors from single MR images based on segmentation with refined patient-specific priors. *Frontiers in neuroscience*, 7.

[Sauwen et al., 2016] Sauwen, N., Acou, M., Van Cauter, S., Sima, D., Veraart, J., Maes, F., Himmelreich, U., Achten, E., and Van Huffel, S. (2016). Comparison of unsupervised classification methods for brain tumor segmentation using multi-parametric MRI. *NeuroImage: Clinical*, 12:753–764.

[Ségonne et al., 2004] Ségonne, F., Dale, A., Busa, E., Glessner, M., Salat, D., Hahn, H., and Fischl, B. (2004). A hybrid approach to the skull stripping problem in MRI. *Neuroimage*, 22(3):1060–1075.

[Shaffer et al., 2010] Shaffer, R., Nichol, A. M., Vollans, E., Fong, M., Nakano, S., Moiseenko, V., Schmuland, M., Ma, R., McKenzie, M., and Otto, K. (2010). A comparison of volumetric modulated arc therapy and conventional intensity-modulated radiotherapy for frontal and temporal high-grade gliomas. *International Journal of Radiation Oncology* Biology* Physics*, 76(4):1177–1184.

[Shewchuk, 1994] Shewchuk, J. R. (1994). An introduction to the conjugate gradient method without the agonizing pain. Technical report.

[Smolensky, 1986] Smolensky, P. (1986). Information processing in dynamical systems: Foundations of harmony theory. Technical report, DTIC Document.

[Tustison et al., 2010] Tustison, N. J., Avants, B. B., Cook, P. A., Zheng, Y., Egan, A., Yushkevich, P. A., and Gee, J. C. (2010). N4ITK: improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6):1310–1320.

[Tustison et al., 2015] Tustison, N. J., Shrinidhi, K., Wintermark, M., Durst, C. R., Kandel, B. M., Gee, J. C., Grossman, M. C., and Avants, B. B. (2015). Optimal symmetric multimodal templates and concatenated random forests for supervised brain tumor segmentation (simplified) with ANTsR. *Neuroinformatics*, 13(2):209–225.

[Van Leemput, 2009] Van Leemput, K. (2009). Encoding probabilistic brain atlases using Bayesian inference. *IEEE Transactions on Medical Imaging*, 28(6):822–837.

[Van Leemput et al., 2001] Van Leemput, K., Maes, F., Vandermeulen, D., Colchester, A., and Suetens, P. (2001). Automated segmentation of multiple sclerosis lesions by model outlier detection. *IEEE transactions on medical imaging*, 20(8):677–688.

[Van Leemput et al., 1999a] Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999a). Automated model-based bias field correction of MR images of the brain. *IEEE transactions on medical imaging*, 18(10):885–896.

[Van Leemput et al., 1999b] Van Leemput, K., Maes, F., Vandermeulen, D., and Suetens, P. (1999b). Automated model-based tissue classification of MR images of the brain. *IEEE transactions on medical imaging*, 18(10):897–908.

[Wang et al., 2013] Wang, H., Suh, J. W., Das, S. R., Pluta, J. B., Craige, C., and Yushkevich, P. A. (2013). Multi-atlas segmentation with joint label fusion. *IEEE transactions on pattern analysis and machine intelligence*, 35(3):611–623.

[Zikic et al., 2012] Zikic, D., Glocker, B., Konukoglu, E., Criminisi, A., Demiralp, C., Shotton, J., Thomas, O., Das, T., Jena, R., and Price, S. (2012). Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 369–376. Springer.