Multi-atlas patch-based segmentation and synthesis of brain tumor MR images

Nicolas Cordier

UNIVERSITÉ NICE SOPHIA ANTIPOLIS

ÉCOLE DOCTORALE STIC
SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

THÈSE

pour l'obtention du grade de

Docteur en Sciences

de l'Université Nice Sophia Antipolis
Domaine : Automatique, Traitement du Signal et des Images

présentée et soutenue par

Nicolas CORDIER

Approches Multi-Atlas fondées sur l'Appariement de Blocs de Voxels pour la Segmentation et la Synthèse d'Images par Résonance Magnétique de Tumeurs Cérébrales

Directeur : Hervé DELINGETTE
Co-Directeur : Nicholas AYACHE
préparée à Inria Sophia Antipolis, équipe ASCLEPIOS
soutenue le 2 décembre 2015

Jury :

Rapporteurs :	Isabelle BLOCH	-	Télécom ParisTech
	François ROUSSEAU	-	Télécom Bretagne
Directeur :	Hervé DELINGETTE	-	Inria Sophia Antipolis
Co-Directeur :	Nicholas AYACHE	-	Inria Sophia Antipolis
Examinateurs :	Christian BARILLOT	-	CNRS
	Bjoern MENZE	-	TU München

UNIVERSITY OF NICE - SOPHIA ANTIPOLIS

DOCTORAL SCHOOL STIC

SCIENCES ET TECHNOLOGIES DE L'INFORMATION
ET DE LA COMMUNICATION

P H D T H E S I S

to obtain the title of

PhD of Science

of the University of Nice - Sophia Antipolis

Specialty: Control, Signal and Image Processing

Defended by

Nicolas CORDIER

Multi-Atlas Patch-Based Segmentation and Synthesis of Brain Tumor MR Images

Thesis Advisor: Hervé DELINGETTE

Thesis Co-Advisor: Nicholas AYACHE

prepared at INRIA Sophia Antipolis, ASCLEPIOS Team

defended on December 2, 2015

Jury:

Reviewers:	Isabelle BLOCH	-	Télécom ParisTech
	François ROUSSEAU	-	Télécom Bretagne
Advisor:	Hervé DELINGETTE	-	Inria Sophia Antipolis
Co-Advisor:	Nicholas AYACHE	-	Inria Sophia Antipolis
Examinators:	Christian BARILLOT	-	CNRS
	Bjoern MENZE	-	TU München

Acknowledgments

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Je tiens tout d'abord à remercier mes deux directeurs de thèse pour la confiance qu'ils m'ont accordée. Merci à Hervé Delingette pour ses capacités d'écoute et de synthèse, qui ont permis de clarifier les développements de cette thèse. Je remercie aussi Hervé pour ses grandes qualités humaines, dont il a fait preuve à de nombreuses reprises au cours de ces dernières années. Merci aussi à Nicholas Ayache pour la direction ambitieuse qu'il a su insufflée dans cette thèse, pour sa patience lorsque les objectifs ne sont pas atteints dans les temps, et enfin pour sa capacité à déterminer ce qui est essentiel pour qu'un projet soit conduit à bon port.

Je tiens ensuite à remercier les rapporteurs du jury, Isabelle Bloch et François Rousseau, pour leur sympathie, et pour leur relecture attentive du manuscrit. Je remercie aussi les autres membre du jury. Merci à Christian Barillot de m'avoir accueilli quelques jours dans l'équipe VisAGES à Rennes en 2013. Cette rencontre a été déterminante, car elle m'a permis de mettre de côté un modèle de synthèse d'images incorrect. Merci aussi à Bjoern Menze pour l'organisation du challenge BraTS de segmentation de gliomes, et pour de longues discussions durant les premières années de thèse. C'est une de ces discussions qui m'a conduit à envisager sérieusement la piste des approches fondées sur l'appariement de blocs de voxels.

Merci aussi à tous les collègues qui ont partagé un bureau avec moi. Merci à Chloé et à Matthieu pour la bonne ambiance qui a régné dans ce bureau, et pour tous ces tableaux blancs qui ont été noircis d'équations, de graphes et dessins. Merci à Ezequiel et à Adityo qui m'ont accueilli dans leur bureau lorsque je commençais tout juste ma thèse. Ce sont eux qui m'ont appris à utiliser le centre de calcul, le début d'une longue histoire. Par ailleurs, c'est Adityo qui, le premier, m'a suggéré de m'intéresser aux méthodes d'appariement de blocs de voxels, mais j'étais trop concentré sur la méthode que je développais à l'époque pour me rendre compte qu'il avait mille fois raison. Merci à Thomas B. et Hugo pour leur énergie et leur enthousiasme, je pense que c'est contagieux, et c'est tant mieux ! Merci à Loïc C. et à Hakim : c'est au cours de mon passage dans leur bureau pendant quelques semaines que j'ai programmé l'algorithme de synthèse d'images, je les remercie donc à la fois pour les temps de rire et les temps de concentration studieuse.

Merci à Loïc L.F. pour les discussions que nous avons eues sur les modèles probabilistes, et qui ont influencé favorablement mon travail. Merci à Mehdi, qui a réussi à pointer du doigt les approximations successives qui étaient faites lors de mon pré-traitement des données : pour un algorithme de segmentation, ces étapes sont clés et doivent être effectuées de la façon la plus rigoureuse possible. Merci à Marco, Jan, Bishesh, Rocio pour leur curiosité scientifique et les échanges que nous avons eus. Merci à Vikash et Federico, avec qui j'ai effectué les trajets vers Auron, et avec qui j'ai bien ri. Je remercie ingénieurs, doctorants et stagiaires que je n'oublie pas

: Aurélie, Brina, Florent, Florian, Kristin, Aziz, Pauline, Anant, Arnaud, Christof, Erin, Jatin, Marc-Michel, Nina, Roch, Sophie, Thomas D., Héloïse, Raphaël, Sonia, etc.

Naturellement, je remercie les autre chercheurs de l'équipe (Xavier, Maxime, Benjamin, Alan, Hervé L., Nicolas D.) pour leurs conseils et remarques, et plus spécifiquement Xavier, grâce à qui j'ai eu l'envie de me présenter au Comité de Centre. Je remercie aussi notre assistante d'équipe, Isabelle S., pour l'organisation sans faille dont elle est la garante.

Merci à la LudAGOS : Marine, Ludovic, Laurent, Justine, Oleksandra, Sophie pour les parties. Merci aux membres Inria extérieurs à l'équipe : merci à Remy et à Karim pour les réunions des comités Inria ; et bien évidemment, merci à toute l'équipe très sympathique qui gère le centre de calcul, sans qui rien n'aurait été possible. J'ai une pensée toute particulière pour Francis, Marc, Nicolas N. et Rémi qui ont été disponibles à toute heure et tous les jours de l'année, weekend, jour férié, petites et grandes vacances comprises. Merci !

Enfin, je tiens à remercier chaleureusement ma famille et mes amis, qui ont dû supporter l'éloignement géographique l'espace d'une thèse.

Contents

Chapter 1

Introduction

Contents

This thesis introduces methods for the analysis and the synthesis of medical images of pathological cases. The focus is on Magnetic Resonance Images (MRI) acquired for tumor-bearing human brains, with a particular interest in the appearance of glioblastomas in MRI. In this thesis, the analysis of medical images refers to segmentation, i.e. the task of contouring and annotating regions based on the observation of medical image acquisitions. The synthesis of medical images refers to the numerical simulation of medical images, given a patient-specific virtual anatomy. A model which would unify analysis and synthesis is our objective, since such a model would allow a better understanding of the two most fundamental bricks in medical imaging, namely: what is it that we see on these given images of pathological cases? And what would an image acquisition of this given pathological case look like?

1.1 Context and motivation

1.1.1 Medical Image Analysis: from a clinical perspective

Gliomas are brain tumors which arise from glial cells. According to the World Health Organization, there are four grades of malignancy [Louis 2007], with glioblastomas being both the most common and the most aggressive kind of gliomas (grade IV). From a clinical perspective, the monitoring of the tumor growth is key for the management of the patient therapy [Mandonnet 2013], and for adequate decision-making in case of a transition from one grade to another. For this purpose, the most reliable monitoring technique consists in periodic acquisitions of multiple Magnetic Resonance (MR) channels (typically contrast-enhanced T_1-weighted, T_2-weighted and T_2-FLAIR MR images), and the subsequent segmentation of the tumor volumes to estimate the velocity of tumor diameter expansion [Mandonnet 2013].

Glioblastoma segmentation typically consists in the delineation of three mutually-inclusive tumor regions [Menze 2014]:

- the enhancing tumor core, or proliferative rim, is positively correlated with angiogenesis induced by high-grade gliomas. This region appears hyper-intense on contrast-enhanced T_1-weighted images due to blood-brain barrier leakage. Indeed, extremely proliferative tumor cells induce angiogenesis, the development of chaotic vessels starting from nearby blood vessels up to the tumor. The newly formed vessels are badly formed and the environment is acid. This results in a blood-brain barrier leakage, which appears hyper-intense on contrast-enhanced T_1-weighted images.

- the tumor core consists in tumor cells, either proliferative or necrotic. This region includes the proliferative rim and it is defined based on T_2-weighted and contrast-enhanced T_1-images. For most high-grade cases, the contrast-enhanced T_1-image helps to detect the tumor core. For low-grade cases and for some high-grade cases, T_2-weighted images are required to determine the fuzzier borders of the tumor core.

- the complete tumor consists of the tumor core and the surrounding edema. This region usually appears hyper-intense in T_2-weighted and T_2-FLAIR MR images.

As could be expected, the manual segmentation of these three tumor regions is tedious and time-consuming, and therefore is the ideal candidate for the development of automatic segmentation algorithms.

1.1.2 Medical Image Synthesis: the importance of annotated data

Many automatic approaches for medical image analysis rely on the learning of model parameters directly from the data, e.g. with statistical learning or generative probabilistic models. Intuitively, these approaches should benefit from training on more data, and be able to leverage large databases of annotated pathological cases to learn models which improve on the state-of-the-art for automatic segmentation. However, the catch is that the learning is done in a supervised way, which means databases should be large and annotated. Such databases are expensive to build and there are only a few attempts at sharing them, for instance the MICCAI Multimodal Brain Tumor Image Segmentation benchmark challenge [Menze 2014].

An original approach consists in building these annotated databases of images from the ground with virtual patients, i.e. atlases completely generated by a numerical biophysical model. This approach was successfully applied in the past: [Geremia 2013] trained random forest on a large dataset consisting of several hundreds of synthetic cases [Prastawa 2009] and showed good automatic segmentation results on a few real cases. [van Tulder 2015] reported two explanations to the benefits of having synthetic medical images for segmentation: *i*) a simple segmentation model builds on the synthesis model, and therefore becomes more complex, and *ii*), a complex segmentation model benefits from having more training data. This kind of approach has lead to an on-going interest in the medical imaging field for

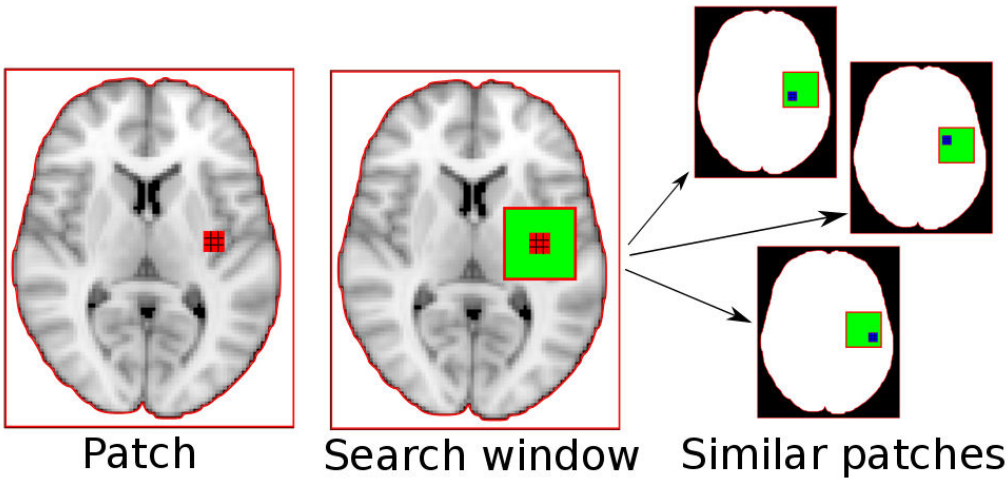

Figure 1.1: Concept of local search window. Left: T_1-weighted image in the MNI space [Mazziotta 2001], with a patch in red. Middle: local search window in green around the patch location. Right: Similar patches, in blue, found in other cases via a search constrained to the local search window location.

more realistic image synthesis algorithms [Ye 2013, Iglesias 2013, Cardoso 2015, van Tulder 2015].

1.1.3 A patch at the intersection of analysis and synthesis

Multi-atlas methods aim at transferring knowledge from multiple annotated templates, known as atlases, to a target case. These approaches can build upon small databases of annotated cases to segment healthy brain structures [Heckemann 2006, Aljabar 2009, Iglesias 2015]. Regarding an application to the analysis of MRI of tumor-bearing brains, [Asman 2012] use a multi-atlas approach to detect outliers on MRI: the target case presents a brain tumor, atlases are images of healthy brains, so the tumor should be identified as an outlier region. One limitation of the original multi-atlas approaches is the expensive computational cost of warping the atlases to the target with deformable registration.

Multi-atlas patch-based methods make use of local intensity patterns, represented by blocks of voxels known as patches. Recently, multi-atlas patch-based methods have been successfully applied for the segmentation of healthy brain structures [Rousseau 2011, Coupé 2011]. The core assumption is that the central voxels of similar patches should have a similar label. To segment healthy structures of the brain, the concept of local search window is introduced: it is sufficient to compare a test patch to annotated patches positioned in a neighbourhood of the position of the test patch, as shown in Figure 1.1. This allows multi-atlas patch-based methods to be computationally efficient.

Even more recently, multi-atlas patch-based methods have been adapted for the

synthesis of an image modality based on another one, known as Modality Propagation [Ye 2013]. Therefore, the patch lies at the intersection of analysis and synthesis, which is of particular interest for this thesis.

However, the application of patch-based methods to pathological cases requires avoiding to rely on local search windows. Prior works [Wang 2013b, Wang 2014, Wang 2013a] expressed a similar interest and suggested different methods.

[Wang 2013b] perform patch-based multi-label segmentation without registration for the segmentation of bones and cartilage on knee MRI. Bones appear hypointense on knee MRI, while cartilage appears hyperintense and is located at the extremities of the bones. To compensate for the absence of registration, patch features are augmented with spatial context, the distance to the center of the image. Similarly, in [Wang 2014], features are augmented with spatial context using geodesic distances.

[Wang 2013a] perform label transfer without registration for brain tumor segmentation. The approach is a hierarchical patch-based technique relying on an initial over-segmentation into small homogeneous regions, known as super-voxels. The coarsest scale is described with super-voxels, while the finest scale is described with patches. Given a test super-voxel, similar super-voxels are found in the atlases. Then, a patch-based technique restricted to the matched super-voxels is performed. Super-voxels are surrogate for non-local search windows to constrain the patch match and allow short computation time.

1.2 Thesis overview

With this thesis, our motivation is to apply multi-atlas patch-based methods to the analysis and simulation of pathological cases, therefore without relying on the concept of local search window. Several questions are considered:

- Can multi-atlas patch-based methods be adapted to the segmentation of pathological cases? Which concept should help replace the assumption of local search? How does the proposed method fit in the probabilistic framework for multi-atlas segmentation? Can state-of-the-art results be achieved for glioma segmentation? Can the computational cost and running time be kept reasonable?

- Is the proposed patch-based segmentation method subject to overfit? How critical are down-sampling, pre-processing and post-processing? What is the effect of a change of definition for patch similarity, or a change of definition for the decay parameter in the patch-based vote? Are there subtle effects due to feature augmentation?

- If the generative probabilistic model for segmentation achieves state-of-the-art results, can the mirrored generative model be used for image synthesis of pathological cases? Can realistic images be synthesized if a label map is the only input? If so, what is the optimal encoding of input label maps to

perform patch matching? Is spatial consistency of synthetic images a given, or should an iterative strategy be introduced? Can uncertainty of the image synthesis model be estimated? How realistic are synthetic images? How does the proposed method compare to state-of-the-art in terms of image synthesis and modularity?

Chapter 2 introduces a patch-based approach for the segmentation of pathologies on multi-channel MRI. A generative probabilistic model is detailed, and the approach consists in a smart approximation of the cost function so that the computations become tractable. Several improvements are proposed to obtain segmentation results in competitive time, including atlas and patch pre-selection criteria. The method is validated on a publicly available benchmark dataset and achieves state-of-the-art results. This chapter is based on [Cordier 2013, Cordier 201X].

Chapter 3 questions the validation framework, generalization, and implicit choices regarding the method described in Chapter 2. Several meta-experiments were conducted. Results and conclusions are discussed.

Chapter 4 introduces a patch-based approach for the synthesis of multi-channel MRI exhibiting a pathology. A novel generative probabilistic model, which mirrors the model presented in Chapter 2, is detailed, and a heuristic optimization procedure is proposed to estimate the maximum a posteriori. Experiments are conducted to assess the quality of the synthetic images with respect to real MRI, to compare to the state-of-the-art method, and to illustrate the modularity of our approach. This Chapter is based on [Cordier 201Y].

Chapter 5 concludes the thesis with a discussion of the main contributions, and finally suggests future research directions.

Appendix A showcases segmentation results obtained with the method described in Chapter 2.

Appendix B showcases synthetic MRI obtained with the patch-based method described in Chapter 4.

Appendix C presents Student distributions of multi-modal image intensity fitted in Chapter 2 and mentioned in Chapter 4.

Chapter 2

A Patch-Based Approach for the Segmentation of Pathologies

Contents

Abstract

In this chapter, we describe a novel and generic approach to address fully-automatic segmentation of brain tumors by using multi-atlas patch-based voting techniques. In addition to avoiding the local search window assumption, the conventional patch-based framework is enhanced through several simple procedures: an improvement of the training dataset in terms of both label purity and intensity statistics, augmented features to implicitly guide the nearest-neighbor-search, multi-scale patches, invariance to cube isometries, stratification of the votes with respect to cases and labels. A probabilistic model automatically delineates regions of interest enclosing high-probability tumor volumes, which allows the algorithm to achieve highly competitive running time despite minimal processing power and resources. This method was evaluated on Multimodal Brain Tumor Image Segmentation challenge datasets. State-of-the-art results are achieved, with a limited learning stage thus restricting the risk of overfit. Moreover, segmentation smoothness does not involve any post-processing.

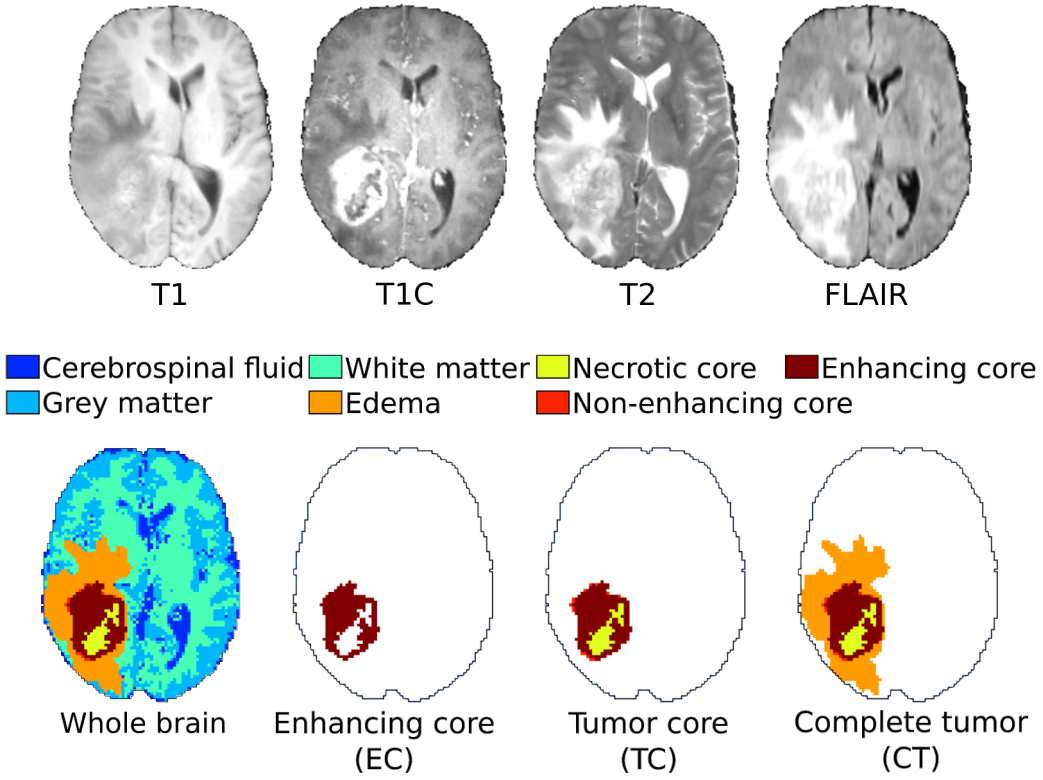

Figure 2.1: MR channels (top row) ; whole brain segmentation and mutually-inclusive pathological regions (bottom row).

2.1 Introduction

2.1.1 Motivation

Glioblastoma is the most severe case of brain tumors. Clinical guidelines such as RECIST [Eisenhauer 2009] or RANO [Wen 2010] are limited to 1D or 2D analysis (maximal diameter and possibly second diameter) of the lesions. However, from tumor growth monitoring to radiotherapy planning, 3D analysis is crucial in the clinical pipeline [Angelini 2007, Mandonnet 2013]. Glioblastoma segmentation consists in a 3D delineation of the pathological compartments [Menze 2014] shown in Figure 2.1. Manual segmentation is usually complex, subjective and time-consuming. First, glioblastoma exhibit high tumor shape variability. Second, the border between compartments can appear fuzzy, which can lead to a debatable segmentation: inter-rater variability of manual segmentations is in the range 74-85% (Dice overlap) [Menze 2014]. Third, the segmentation task requires the simultaneous screening of 3D images acquired with multiple Magnetic Resonance (MR) sequences (Figure 2.1). This explains the ongoing interest for automatic segmentation algorithms, notably within the Multimodal Brain Tumor Image Segmentation (BraTS) benchmark challenge [Menze 2014].

2.1.2 Related work

2.1.2.1 Glioma segmentation

Most of the automatic glioma segmentation approaches learn offline a discriminative model [Menze 2014, Zikic 2012, Tustison 2014]: image intensity features are computed, then a machine learning algorithm is trained offline. Most computation time is spent during the learning stage, which should be run again if newly acquired data is annotated. Moreover, results are highly dependent on the choice of features [Tustison 2014], and feature extraction has to be performed at test time. For instance, [Tustison 2014] introduce a cascade of random forest classifiers based on a set of intensity, geometry and asymmetry features. The segmentation is refined using Markov random field regularization. Feature extraction takes about 90 minutes (single-thread implementation) [Menze 2014], mostly due to the computation of the most relevant features (asymmetry) [Tustison 2014].

The generative approach builds a probabilistic model of observed image intensity given the tissue type. The latent variable is the spatial distribution of healthy tissues and tumor compartments. Prior knowledge includes the location and spatial extent of healthy tissues in an atlas. [Menze 2010] introduce a generative model for channel-specific pathology segmentation: the appearance of healthy tissues is modelled as a multivariate distribution for all channels simultaneously, while tumor appearance is channel-specific. Tumor location is a latent variable shared between MR channels, which results in tumor segmentations consistent across channels. Since tumor compartments are not explicitly modelled in [Menze 2010], a discriminative classifier is learnt to transform channel-specific abnormality probability maps into tumor compartment delineations [Menze 2014]. To reach state-of-the-art results, a strong post-processing is required to deal with the high number of false positives. Running time is 20 minutes [Menze 2014].

Recently, Kwon [Kwon 2014] achieved top rank among the BraTS benchmark competitors with a semi-automatic framework. First, an expert inputs multiple seed points and radii, and samples one point from each tissue class. A tumor shape prior is generated from the seed points via random walk. A pathological atlas is synthesized by growing tumors from the seed points and combining the result with a normal atlas. Average running time is less than 10 minutes for the user inputs, plus 85 minutes for the remaining automatic pipeline on a machine with 8 cores.

2.1.2.2 Multi-atlas segmentation

When applied to glioma segmentation, machine learning techniques are confronted with two major problems. First, the amount of training data is usually small: for instance, there are only 20 high-grade training cases for the 2013 BraTS benchmark [Menze 2014]. Second, most algorithms require a computationally intensive offline learning stage, which can be subject to overfit. Multi-atlas segmentation methods are appealing as they can cope with a small training dataset, and are performed online, which allows a seamless integration of new cases into the training

dataset.

Atlas-guided segmentation consists in encoding the relationship between image intensities and labels through the use of an atlas, an expert-annotated image [Iglesias 2015]. The segmentation of a test case relies on label propagation, the deformable registration of the atlas to the test case. Recently, multi-atlas segmentation [Rohlfing 2003], which uses several atlases to best capture anatomical variability, has proven successful for the segmentation of healthy brain structures [Heckemann 2006, Aljabar 2009, Cardoso 2013, Iglesias 2015]. However, limitations include *i)* the high computational cost of non-linear registration, and *ii)* the assumption that a one-to-one mapping exists between atlas and patient.

The development of patch-based segmentation alleviates these two limitations by performing only affine registration and by introducing the concept of local search window to take into account registration errors: since brain structures should be located around the same position after registration, it is sufficient to compare a test patch to annotated patches positioned in a spatial neighbourhood. Patch-based methods have achieved state-of-the-art results for the segmentation of brain structures [Rousseau 2011, Coupé 2011, Romero 2015]. However, local search cannot be used for the segmentation of pathologies: tumor patches may not be located in the same region of the human body.

In this chapter, we address the automatic segmentation of brain tumors by using multi-atlas patch-based techniques, without any assumption of local search, so that *i)* minimal learning is required, which decreases the risk of overfit, especially on a small training dataset, *ii)* minimal post-processing is required, since segmentation is robust, and spatially consistent labelling is guaranteed by patch overlap, *iii)* competitive running times are achieved using a single core, and the algorithm could benefit from a computer grid as it is embarrassingly parallel. To obtain a reliable patch-based method which does not rely on the concept of local search window, we combine several methodological advances: feature augmentation based on spatial prior and robust intensity likelihood, selection of training dataset based on label purity and intensity statistics, and a stratification of the votes with respect to cases and labels. In the following, an enhanced patch-based segmentation framework is introduced to accommodate with the presence of a pathology (Section 2.2). Then the algorithm is evaluated on publicly available MR images, and its performance is benchmarked against state-of-the-art methods (Section 2.3). Finally, results and perspectives are discussed (Section 2.4).

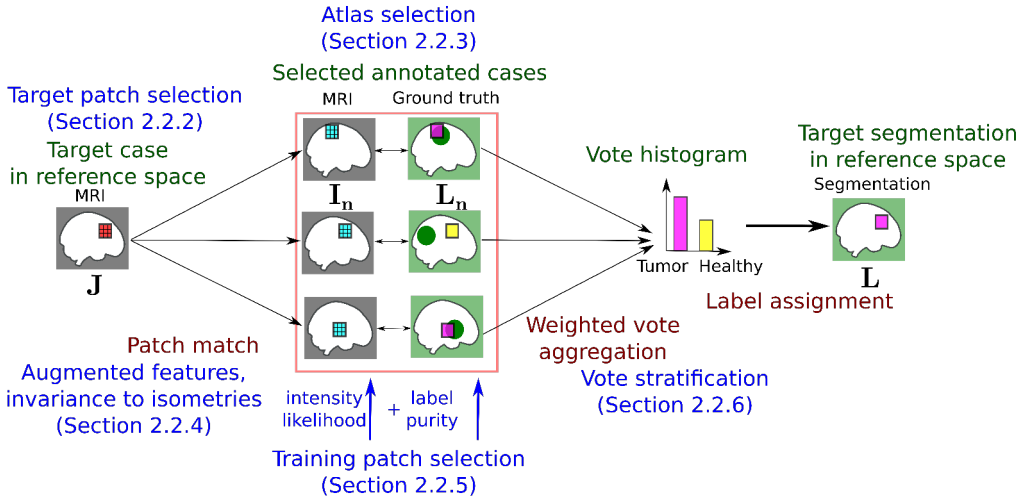

Figure 2.2: Patch-based glioma segmentation. Our contributions are mentioned in blue. Green disks mark known tumor locations.

2.2 Patch-based glioma segmentation

Conventional patch-based segmentation without any search window constraint is illustrated in Figure 2.2, along with our contributions. The fundamental assumption is that central voxels of similar patches should have similar labels. Consequently, the label of a test patch could be inferred by finding similar patches in a subset of the training cases and performing a weighted vote: each patch votes for its label with a weight depending on its similarity to the target patch.

A probabilistic model is first described to automatically define a region-of-interest (ROI) within the target image. Then three improvements to the conventional framework are proposed: feature augmentation and invariance, training patch selection, and stratification of votes.

2.2.1 Notations

Patch-based segmentation relies on a set of training cases $\{I_n, L_n\}_n$ where I_n denotes multi-channel MRI and L_n is a label map. A label l is assigned to a test patch $S(J, x)$ based on comparisons with a database of annotated patches $\{S(I_n, y)\}_{(n,\, y)}$, where y indexes every spatial position in the reference space Ω. Conventionally, a mono-channel patch $S(I, x)$ is a cube of edge length $2r + 1$, centered at spatial position $x \in \mathbb{R}^3$, consisting of image intensity values taken from a mono-channel MR image I. The distance $d(S(I, x),\ S(J, y))$ between 3D mono-channel patches of identical edge length is the canonical L_2-norm on $\mathbb{R}^{(2r+1)^3}$. A multi-channel patch is the concatenation of mono-channel patches. The squared distance between multi-channel patches is the sum of the squared distances between mono-channel patches.

2.2.2 Patch selection: a probabilistic model for glioma detection

2.2.2.1 Motivation: segmentation speed-up

The automatic definition of ROI, enclosing high-probability tumor regions, allows patch-based segmentation to achieve competitive running time, thanks to patch selection ahead of time. Indeed, patches outside of the target ROI in the target case are directly discarded, which cuts down on time for both patch extraction and patch match. Moreover, patches outside of the training ROI in the training case are similarly discarded since they are unlikely to be relevant matches. In practice, target patch selection has close to no influence on segmentation results: discarded target patches are unlikely to trigger any alarm during segmentation, since the features used for detection (average intensity over a patch) are less specific than the patches used for segmentation.

2.2.2.2 Definition of a probabilistic model

The idea is to perform a robust clustering of patches into 7 clusters based on average patch intensity. Clusters include cerebrospinal fluid (CSF), grey matter (GM), white matter (WM), necrotic core (Nec.), edema, non-enhancing tumor core (NETC), and enhancing core (EC). Robustness to outliers arising from imaging artefacts, bias field or ground truth mislabelling, is achieved through the use of multivariate Student distributions to model the heavy-tailed distribution of multi-channel intensity[1]. The proposed model is a mixture of Student distributions [Svensén 2005], with fixed and spatially-varying mixing coefficients, similar to [Shiee 2011]. Each Student distribution is parametrized by $\theta_l = (\mu_l, \Sigma_l, \nu_l)$, where μ_l is the mean, Σ_l the covariance matrix, and ν_l the number of degrees of freedom. The graphical model for glioma detection consists of a directed graph with 3 nodes: $\mathbf{X} \rightarrow \mathbf{L} \rightarrow \mathbf{J}$, as presented in Figure 2.3. The joint probability is:

$$p(\mathbf{J}, \mathbf{L}, \mathbf{X}) = p(\mathbf{J} \mid \mathbf{L}) p(\mathbf{L} \mid \mathbf{X}) p(\mathbf{X})$$

where $\mathbf{X}$ indexes the coordinates of the brain in the MNI space Ω. The label map $\mathbf{L}$ is assumed to be sampled independently for each position x from a generalized Bernoulli distribution with parameters $\{\pi_x(l)\}_l$, our prior regarding the spatial extent of tissue classes. Given a class $\mathbf{L}(x) = l$, the multimodal image intensity $\mathbf{J}(x) \in \mathbb{R}^4$ is sampled independently from a Student distribution with parameters θ_l.

[1] A Student distribution is chosen as it generalizes Gaussian distribution with the ability of cope with heavy tail distributions. This translates into a more robust estimation in the presence of outliers. We have tried to fit Gaussian distributions first, but distribution statistics (mean, covariance) would be very poorly captured. Overall, Student distributions have this extra-flexibility to fit our data.

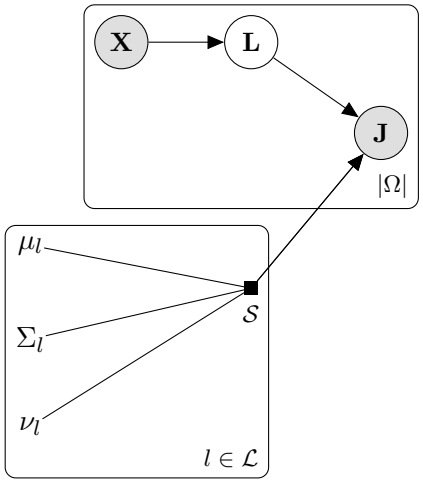

Figure 2.3: Graphical model for glioma detection: mixture of Student distributions with fixed and spatially varying mixing coefficients. The label $\mathbf{L}(x)$ is sampled from a generalized Bernoulli distribution at position x in the MNI space Ω indexed by $\mathbf{X}$. The observed intensity $\mathbf{J}(x)$ is sampled from a multivariate Student distribution $\mathcal{S}$ parametrized by $\theta_l = (\mu_l, \Sigma_l, \nu_l)$.

2.2.2.3 Estimation of the spatial probabilistic prior

To build atlases of label spatial distribution, every training case is affinely registered to a reference space [Mazziotta 2001], then, for each label, an atlas of spatial distribution is defined as the voxel-wise average of warped label maps. Atlases are finally symmetrized with respect to the mid-sagittal plane and smoothed by a Gaussian convolution (see examples in Figure 2.4). For this task, 314 ground truth segmentations from BraTS are used, and the standard deviation of the smoothing Gaussian kernel is 10 mm.

2.2.2.4 Estimation of the parameters of Student distributions

Multi-channel patch information is summarized by the average intensity of mono-channel patches. Multivariate Student distributions are fitted using Expectation-Maximization [Liu 1995] on $6 \times 6 \times 6$ mm^3 pure patches, i.e. for which all voxels belong to the same class. Pure patches are only considered because depending on the quality of the training segmentation, pre-processing (registration, interpolation), patches at the border between labels could be mislabelled. By restricting the number of patches, we make this estimation more tractable in terms of memory size. Given the optimal parameters θ_l^*, class conditional Student intensity likelihood can be computed in closed form for any patch (see Figures 2.4 and 2.5).

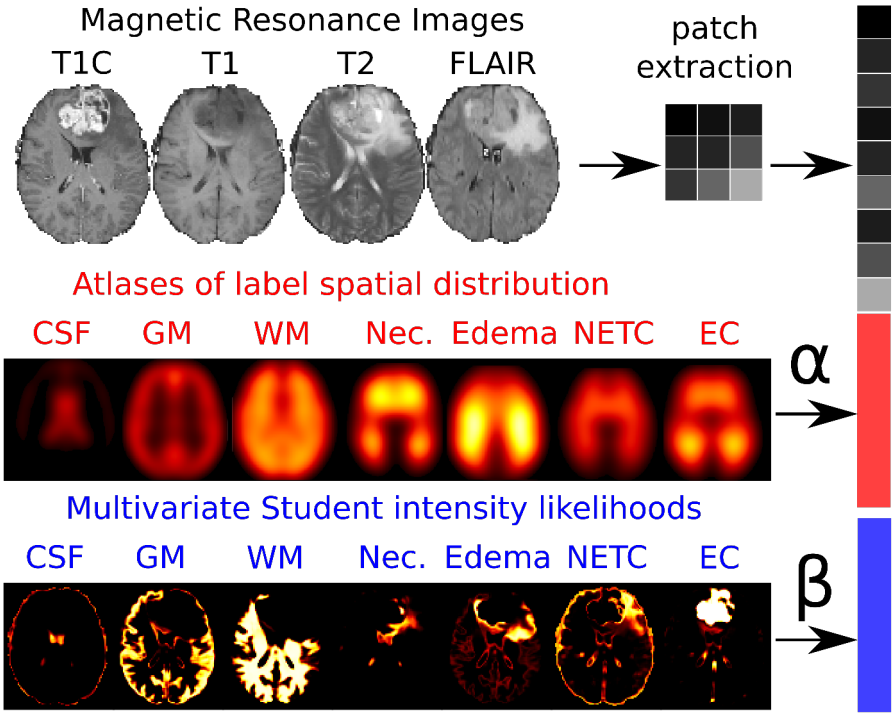

Figure 2.4: Feature augmentation. A feature vector is composed of *i)* intensity patches extracted from 4 MR channels, *ii)* scalar values from 7 atlases of label spatial distribution (weight α), and *iii)* scalar values from 7 class conditional Student intensity likelihoods (weight β).

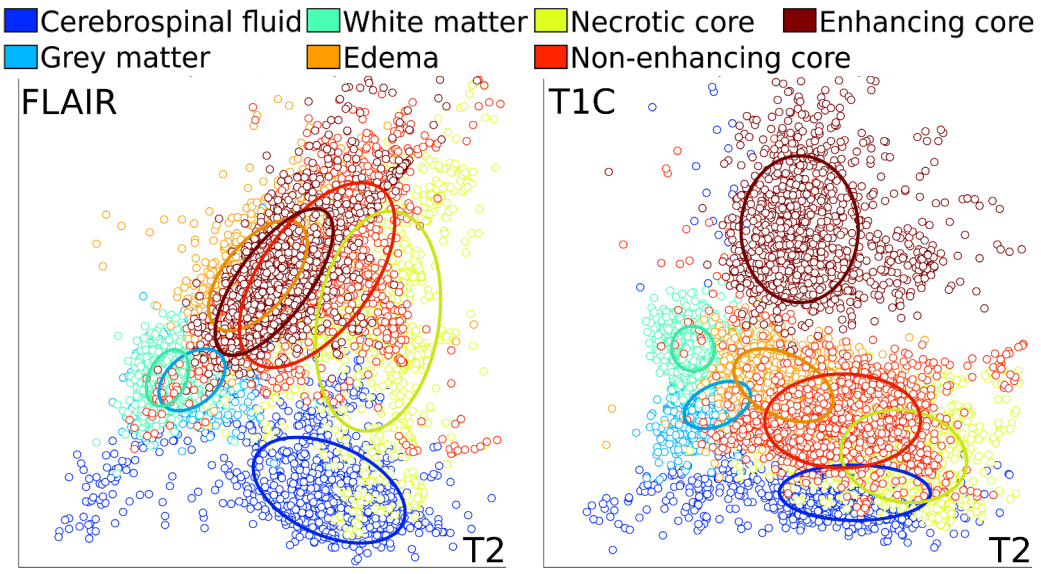

Figure 2.5: 2D projections of Student distributions fitted to average intensity of pure patches, with 60% quantiles overlayed. Appendix C presents other projections.

2.2.2.5 Estimation of the posterior probabilities

The posterior probability of observing label $\mathbf{L}$ is given by Bayes' rule and the conditional dependence given by the factorization from the graphical model:

$$p(\mathbf{L} \mid \mathbf{J}, \mathbf{X}) \propto p(\mathbf{J} \mid \mathbf{L}) p(\mathbf{L} \mid \mathbf{X})$$

Indeed, Bayes' rule gives:

$$p(\mathbf{L}|\mathbf{J}, \mathbf{X}) p(\mathbf{J}|\mathbf{X}) = p(\mathbf{J}|\mathbf{L}, \mathbf{X}) p(\mathbf{L}|\mathbf{X})$$

And the conditional dependence given by the factorization from the graphical model gives:

$$p(\mathbf{J}|\mathbf{L}, \mathbf{X}) = p(\mathbf{J}|\mathbf{L})$$

which results in:

$$\begin{aligned} \arg\max_{\mathbf{L}} p(\mathbf{L}|\mathbf{J}, \mathbf{X}) &= \arg\max_{\mathbf{L}} p(\mathbf{L}|\mathbf{J}, \mathbf{X}) p(\mathbf{J}|\mathbf{X}) \\ &= \arg\max_{\mathbf{L}} p(\mathbf{J}|\mathbf{L}, \mathbf{X}) p(\mathbf{L}|\mathbf{X}) \\ &= \arg\max_{\mathbf{L}} p(\mathbf{J}|\mathbf{L}) p(\mathbf{L}|\mathbf{X}) \end{aligned}$$

2.2.2.6 Automatic glioma detection

To cope with the lack of specificity of detection features, the maximal posterior map is post-processed by thresholding signed Euclidean distance maps, as shown in Figure 2.6. The goal is to discard thin layers of tumor-detected voxels which are adjacent to healthy-detected tissues, e.g. false positives at the outline of ventricles. Three distance thresholds are used: *i)* 3 mm from each healthy tissue (CSF, GM, WM), *ii)* 6 mm from the intersection of the binary masks obtained in step *i*, and *iii)* 6 mm from the detected tumor binary mask obtained in step *ii*.

2.2.3 Atlas selection: segmentation scaling to large data sets

Atlas selection, i.e. the pre-selection of a fixed number of training cases similar to the target case, is one solution to scale to large datasets. Higher priority is given to training cases with the same grade as the target. To sort training cases of identical grade, the distance between target case $\mathbf{J}$ and training case $\mathbf{I_n}$ is the arithmetic mean, for every label l, of the class-specific Hellinger distances between histograms $\left(p\left(\mathbf{L} = l \mid \mathbf{J}, \mathbf{X} = x\right)\right)_{x \in \Omega}$ and $\left(p\left(\mathbf{L} = l \mid \mathbf{I_n}, \mathbf{X} = x\right)\right)_{x \in \Omega}$. For this purpose, the posterior probabilities detailed in Section 2.2.2.5 are used. This is a very simple and efficient procedure. The intuition is that similar cases should have equivalent volumes for each label, no matter the spatial configuration.

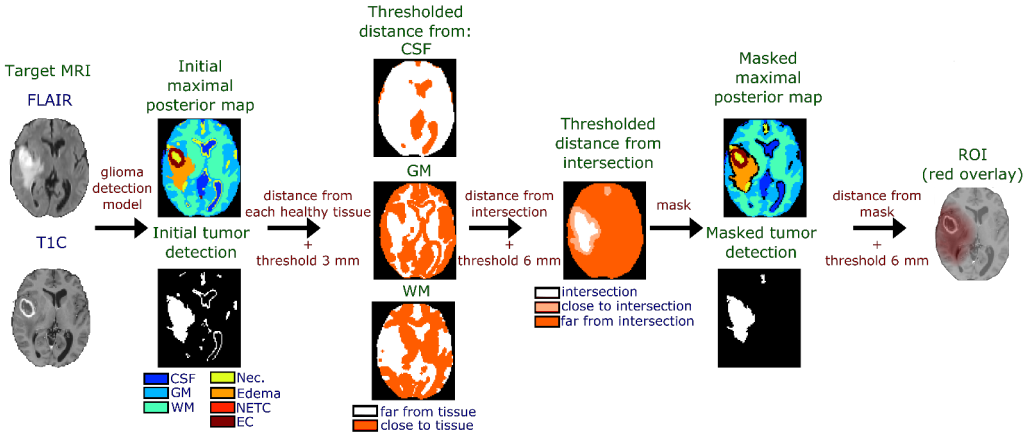

Figure 2.6: Glioma detection post-processing. Euclidean distances from each healthy tissue are thresholded and intersected to determine the tumor bulk. The ROI is the result of a dilatation of the tumor mask using two additional distance maps.

2.2.4 Feature augmentation for a guided patch match

Patch match consists in finding the most similar patches to a target patch. An augmentation of patch features with additional contextual features can implicitly guide patch match. Three feature augmentations and one feature invariance are proposed.

2.2.4.1 Multi-scale patches

Conventional patches constrain our vision of structures to the scale of the patch. However, gliomas exhibit structures at scales varying from case to case: *i)* the edema often has the largest extent for high-grade cases, but can be very small compared to the tumor core for low-grade cases ; *ii)* the distinction between necrotic core and enhancing core may be challenging due to complex borders.

As explained in Figure 2.7, we have adopted inherently multi-scale patches [Wachinger 2014a] with two scales: 6 mm (3x3x3 patch) and 18 mm (9x9x9 patch). The central part of the patch is described precisely, while the peripheral part is described by average intensity values over smaller patches, by analogy with the foveal vision. This idea allows to capture longer-range image information since it is computationally efficient: conventional 3x3x3 and multi-scale 9x9x9 patches have a similar memory footprint.

More precisely, conventional mono-channel patches are represented as vectors of the voxel-wise intensity values, e.g. L^n scalar values for an n-D patch of length L. However, for 3D multi-scale mono-channel patches: i) from a given voxel, we consider its 26 neighbouring voxels making a 3x3x3 image patch (this represents 27 scalar values) ; ii) we then consider all the 26 neighbouring conventional 3x3x3

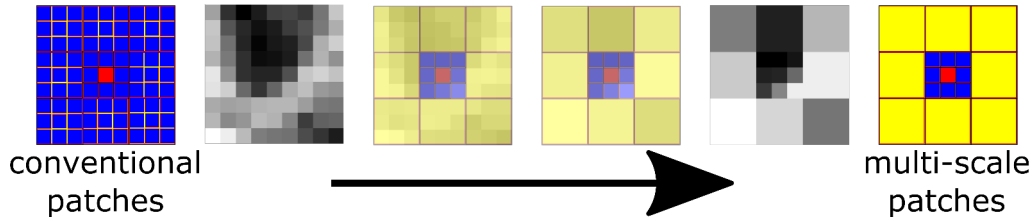

Figure 2.7: Multi-scale patch as described in [Wachinger 2014a]. In red: central voxel. In blue: intensity described at the same scale as the central voxel. In yellow: intensity described at a coarser scale, typically average intensity over the coarser area.

patches and compute the average intensity for all these neighbouring patches, thus providing 26 additional scalar values. In the end, the feature vector corresponding to a 3D multi-scale and mono-channel patch is composed of 53 scalar intensity values. We illustrate the concept of 3D multi-scale mono-channel patches in Figure 2.8. This process is repeated for each MR channel to process 3D multi-scale and multi-channel patches.

2.2.4.2 Atlases of label spatial distribution

One aspect to study glioma pathogenesis is the analysis of the distribution of gliomas by anatomic locations [Larjavaara 2007, Duffau 2004, Parisot 2011]. Based on 331 cases among which 47% were glioblastomas, [Larjavaara 2007] report the majority of gliomas in the frontal and temporal lobes, with 29% and 14% occurrences respectively. Similarly, based on 314 cases among which 82% are glioblastomas, most gliomas in the BraTS benchmark dataset are found in the temporal lobe, as illustrated with the atlas in Figure 2.9. We augment patch features with the atlases of spatial distribution for each of the 7 labels, in order to implicitly encode patch position in the reference space. The additional feature is a vector in $\mathbb{R}^7$, which is first normalized so that its sum is equal to the average intensity value reported in Section 2.3.2, then weighted by a coefficient $\alpha \geq 0$. The distance between augmented features is the L_2-norm on concatenation of patch values and weighted additional features.

2.2.4.3 Multivariate Student intensity likelihoods

Our suggestion is to augment patch features with robust Student intensity likelihoods, which encode the relative global intensity distance of a patch, with respect to each class taken as a whole. The fit of Student distributions is illustrated in Figure 2.5. Exactly as in Section 2.2.4.2, the additional feature is a vector in $\mathbb{R}^7$, which is first normalized (unit normalization, followed by the multiplication by a scalar value so that the L_1 norm is the average intensity value reported in Section 2.3.2) and then weighted by a coefficient $\beta \geq 0$.

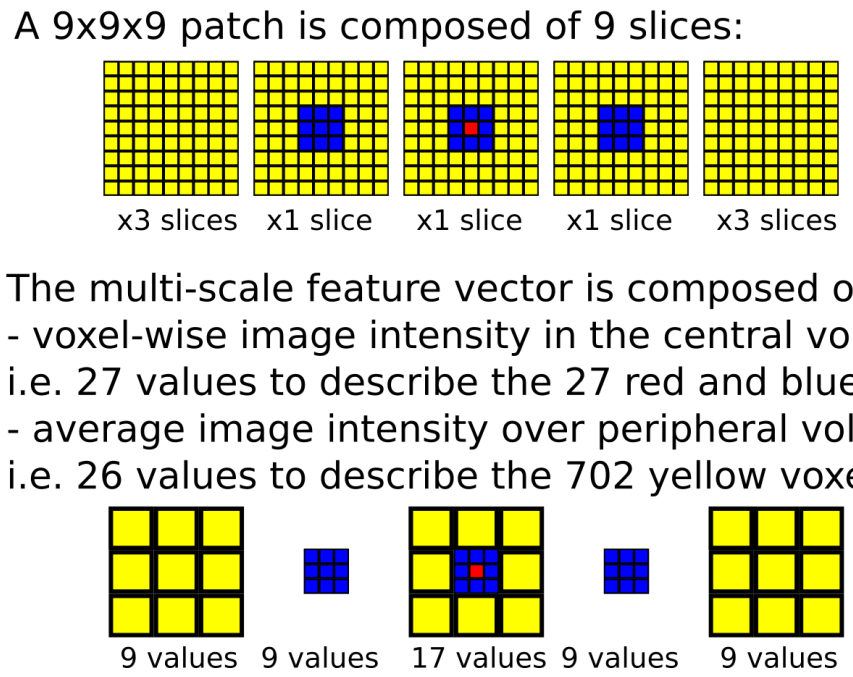

Figure 2.8: Definition of a 3D multi-scale patch. In red: central voxel. In blue: intensity described at the same scale as the central voxel. In yellow: intensity described at a coarser scale, by taking the average intensity over the coarser area. Each square in the bottom row corresponds to one of the 53 scalar values found in the patch feature vector.

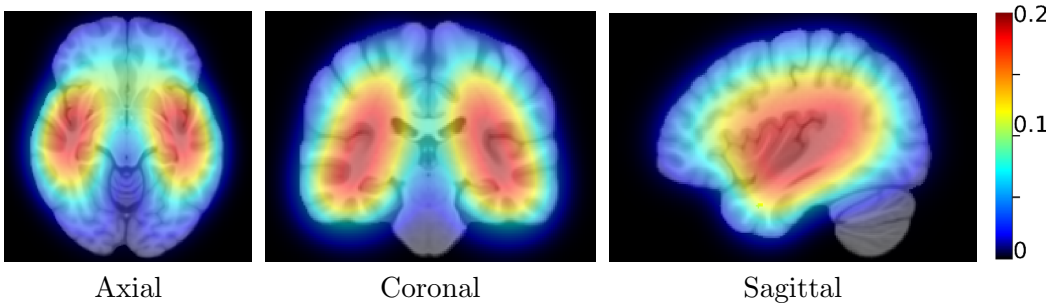

Figure 2.9: Atlas of spatial distribution of complete tumor superimposed on a template. Colormap: percentage of occurrences.

2.2.4.4 Invariance to cube isometries

The canonical L_2 distance is sensitive to rotation or symmetry of the patches. This sensitivity is justified for healthy brain structures since rotated patches can be unrealistic. However, since the brain is roughly symmetric with respect to the mid-saggital plane, the application of sagittal plane symmetry to healthy training patches results in plausible healthy patches. For pathological training patches, it is desirable that the similarity measure is insensitive to rotation and symmetry: gliomas do not seem to exhibit any general trend in terms of texture anisotropy. Moreover, the training dataset is small, so plausible configuration of pathological patches are missing. Cube isometries consist of certain rotations under which the cube is invariant, plus their composition with central symmetry. The application of the 48 cube isometries to pathological training patches allows to generate additional plausible configurations, which leads to an invariant patch distance.

2.2.5 Training patch selection for a robust patch match

Manual segmentations are prone to errors due to tumor complex appearance and shape. Training patch selection consists in trimming the training data, so as to increase its robustness to mislabelled patches.

2.2.5.1 Data pruning based on class conditional Student intensity likelihood

The most representative patches lie near their class centroid, so patches whose Student likelihood lie outside of a 60%-quantile could be discarded as dubiously labelled or least representative (Figure 2.5).

2.2.5.2 Data pruning based on patch label purity

Patch label purity is defined as the percentage of voxels sharing the same label as the central voxel in a patch. The fundamental assumption of patch-based methods may not hold for patches of low purity. Different purity thresholds are tested in Section 2.3.6.4.

2.2.6 Vote stratification for a robust vote aggregation

2.2.6.1 Conventional model for patch-based segmentation

Conventional patch-based segmentation is a variation of local weighted voting [Bai 2013, Iglesias 2015] within Bayesian modeling of multi-atlas segmentation [Sabuncu 2010], as recalled in Figure 2.10. The target image is the result of sampling patches from atlases $(\mathbf{I_n},\ \mathbf{L_n})$ at different positions y in the reference space Ω. The membership index $\mathbf{M}$ encodes, for every position $x \in \Omega$, both the atlas index n and the position y to sample from: $\mathbf{M}(x) = (n,\ y)$. If $\mathbf{M}$ were known,

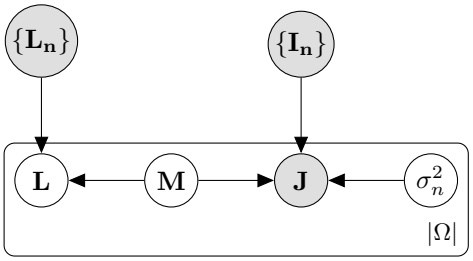

Figure 2.10: Graphical model for patch-based segmentation. A membership variable $\mathbf{M} : \Omega \rightarrow \{1, \ \ldots, \ N\} \times \Omega$ is sampled at every position x in the MNI space Ω to encode the training case $\mathbf{n}$ to sample from, and the spatial offset y. The observed intensity $\mathbf{J}(x)$ results from the sampling of a patch $S(\mathbf{J}, x)$ from a Gaussian distribution with $S(\mathbf{I_n}, y)$ for mean and isotropic scaling $\sigma_n^2(x)$ for variance. The label $\mathbf{L}$ at position x is the central label $\mathbf{L_n}(y)$.

the target segmentation $\mathbf{L}$ would be given by labels $\mathbf{L_n}(y)$ at central voxels of sampled patches. However, since $\mathbf{M}$ is not observed, the inference has to be performed by marginalizing over $\mathbf{M}$ [Sabuncu 2010]. The exact marginalization consists in a weighted vote involving patches at every position y in every atlas n, with weights proportional to both the probability that the training patch belongs to class $\mathbf{L_n}(y)$, and the intensity likelihood of the training patch.

Assuming the membership index $\mathbf{M}(x)$ is independent and identically uniformly distributed $p(\mathbf{M}) = (N \times |\Omega|)^{-|\Omega|}$, the marginalization over $\mathbf{M}$ [Sabuncu 2010] is given voxel-wise by:

$$\widehat{\mathbf{L}}(x) = \underset{l \in \{1, \ ..., \ \mathcal{L}\}}{\operatorname{argmax}} \sum_{n=1}^{N} \sum_{y \in \Omega} p_n \left(\mathbf{L}(x) = l \mid \mathbf{L_n}, \ y\right) \times p_n \left(S(\mathbf{J}, x) \mid \mathbf{I_n}, \ y, \ \sigma_n^2(x)\right)$$

where: *i)* since label maps of training cases are warped from patient space into a reference space, the ground truth can be represented as probabilistic label map $\widetilde{\mathbf{L_n}}(y)$, or as a binary label map $\mathbf{L_n}(y)$ obtained by binarizing $\widetilde{\mathbf{L_n}}(y)$. The probability of belonging to class $\mathbf{L_n}(y)$ is $p_n\left(\mathbf{L}(x) = l \mid \mathbf{L_n}, \ y\right)$, which is either the l-th component of $\widetilde{\mathbf{L_n}}(y)$ (probabilistic label), or the indicator function $\mathbb{1}_{\mathbf{L_n}(y)=l}$ (binarized label),

and *ii)* a multivariate Gaussian probability density function $p_n\left(S(\mathbf{J}, x) \mid \mathbf{I_n}, \ y, \ \sigma_n^2(x)\right)$, with $S(\mathbf{I_n}, y)$ for mean and isotropic scaling $\sigma_n^2(x)$ for variance, comprises the similarity between intensity patches:

$$p_n \left(S(\mathbf{J}, x) \mid \mathbf{I_n}, \ y, \ \sigma_n^2(x)\right) \propto \exp\left(-\frac{1}{2\sigma_n^2(x)} d^2\left(S(\mathbf{J}, x), \ S(\mathbf{I_n}, y)\right)\right)$$

with $\sqrt{2}\sigma_n(x) = \min_y d\left(S(\mathbf{J}, x), \ S(\mathbf{I_n}, y)\right)$

2.2.6.2 Vote stratification for an approximate marginalization

The exact marginalization over $\mathbf{M}$ would require a high number of comparisons between image patches, which would result in prohibitive computation time. Conventional patch-based methods approximate the marginalization step with K-nearest-neighbor patch match or local search window.

For pathology segmentation, we propose to use a stratified K-nearest neighbor approach in order to improve the robustness of the approximate marginalization. More precisely, the set of all patches from all training patients may be partitioned into Q sets of patches which match a number of criteria such as the patient Id, and label type. Given a test patch, the final vote for its label combines the contributions of the K closest patches within each of the Q sets leading to $K \times Q$ votes. Vote stratification can be seen as a way to normalize the vote to limit bias in the training dataset[2]. For this application, we chose to stratify the sets according to patient Id and label type such that $Q = N \times \mathcal{L}$ and $K = 1$. Indeed, this allows to cope with the variability in tumor size between patients, and the unbalanced number of patches between labels. Moreover, this approximation sounds reasonable since, in practice, we verified that taking into account strictly more than $K = 1$ neighbor does not affect segmentation results. In the case of binary label maps, the proposed approximate marginalization consists in:

$$\widehat{\mathbf{L}}(x) = \underset{l \in \{1,\ ...,\ \mathcal{L}\}}{\operatorname{argmax}} \sum_{n=1}^{N} p_n \left(S(\mathbf{J}, x) \mid \mathbf{I_n},\ y_{n,1}(l),\ \sigma_n^2(x) \right)$$

where $y_{n,1}(c)$ is the position in reference space of the closest patch of label c found in the training case $(\mathbf{I_n},\ \mathbf{L_n})$. An implementation of this stratification scheme is given in Algorithm 1.

In the case of probabilistic label maps, the equation for an approximate marginalization differs slightly due to the fact that $p_n\left(\mathbf{L}\left(x\right) = l \mid \mathbf{L_n},\ y\right)$ is no longer an indicator function:

$$\widehat{\mathbf{L}}(x) = \underset{l \in \{1,\ ...,\ \mathcal{L}\}}{\operatorname{argmax}} \sum_{n=1}^{N} \sum_{c=1}^{\mathcal{L}} p_n\left(\mathbf{L}\left(x\right) = l \mid \mathbf{L_n},\ y_{n,1}(c)\right) \times p_n \left(S(\mathbf{J}, x) \mid \mathbf{I_n},\ y_{n,1}(c),\ \sigma_n^2(x) \right)$$

In practice, the only impact of considering probabilistic label maps $\widetilde{\mathbf{L_n}}$ lies in the vote accumulation step in Algorithm 1. The update becomes a for-loop over classes $c \in \{1,\ \ldots,\ \mathcal{L}\}$:

$$v_{x,j}(c) \leftarrow v_{x,j}(c) + p_j\left(\mathbf{L}\left(x\right) = c \mid \mathbf{L_j},\ y_{j,k}\right) \times \exp(-\frac{1}{2\sigma_j^2(x)} ||S(\mathbf{J}, x) - S(\mathbf{I_j}, y_{j,k})||_2^2)$$

[2]Indeed, the summation over all the training cases ensures that the vote takes into account patches from every training case. Otherwise, a case with a large tumor could provide more patch matches than cases with smaller tumors, which would bias the vote towards the training cases with larger tumors.

Algorithm 1: Stratified patch-based segmentation. The blue dashed line marks an embarassingly parallel for-loop.

pre-select N training cases similar to $\mathbf{J}$
foreach voxel $x \in \Omega$ **do**
  evaluate the features $S(\mathbf{J}, x)$
  foreach training case j **do**
    initialize the vote vector $v_{x,j} = 0 \in \mathbb{R}^{\mathcal{L}}$
    foreach class $l \in \{1, \ldots, \mathcal{L}\}$ **do**
      pre-select K closest patches of label l in $\mathbf{I_j}$
      foreach annotated patch $k \in \{1, \ldots, K\}$ **do**
        evaluate the features $S(\mathbf{I_j}, y_{j,k})$
        accumulate the vote:
        $v_{x,j}(l) \leftarrow v_{x,j}(l) + \exp(\frac{-||S(\mathbf{J},x) - S(\mathbf{I_j}, y_{j,k})||_2^2}{2\sigma_j^2(x)})$
    foreach class $l \in \{1, \ldots, \mathcal{L}\}$ **do**
      evaluate the probability $p_{x,j}(l) = \frac{v_{x,j}(l)}{\Sigma_l v_{x,j}(l)}$
  foreach class $l \in \{1, \ldots, \mathcal{L}\}$ **do**
    evaluate the probability $p_x(l) = \frac{1}{N} \Sigma_j p_{x,j}(l)$
  set the label: $\widehat{\mathbf{L}}(x) = \underset{l \in \{1, \ldots, \mathcal{L}\}}{\operatorname{argmax}} \; p_x(l)$

2.3 Results

2.3.1 Dataset and Evaluation

Four MR channels commonly acquired for glioma assessment are available in 1 mm isotropic resolution for each case: pre-contrast T_1-weighted image (T1), contrast-enhanced T_1-weighted (T1C), T_2-weighted (T2), and T_2-FLAIR MR images. Details about MRI acquisition and processing, manual delineation protocol, and expert consensus can be found in [Menze 2014].

The proposed glioma segmentation algorithm is evaluated on publicly available MR images of high-grade (HG) and low-grade (LG) gliomas, made available for training and testing in the 2013 and 2014 editions of the MICCAI BraTS benchmark [Menze 2014]. Ground truth is only publicly available for 2013 and 2014 Evaluation datasets. The 2014 Evaluation dataset is used for the construction of atlases of label spatial extent and to fit Student distributions. Unless explicitly mentioned otherwise, the training dataset used for patch match in this chapter is the 2013 Evaluation dataset for a fair comparison to other competitors. For validation on training cases, a comprehensive leave-one-out procedure is applied to exclude the target image from the training dataset at every stage of the method, namely *i*) for atlas of label spatial extent construction, *ii*) to fit Student distributions, *iii*) during patch-match. This way, the target case cannot result in any overtraining that would be due to a favourably-biased glioma detection, atlas selection, feature augmentation, or patch match.

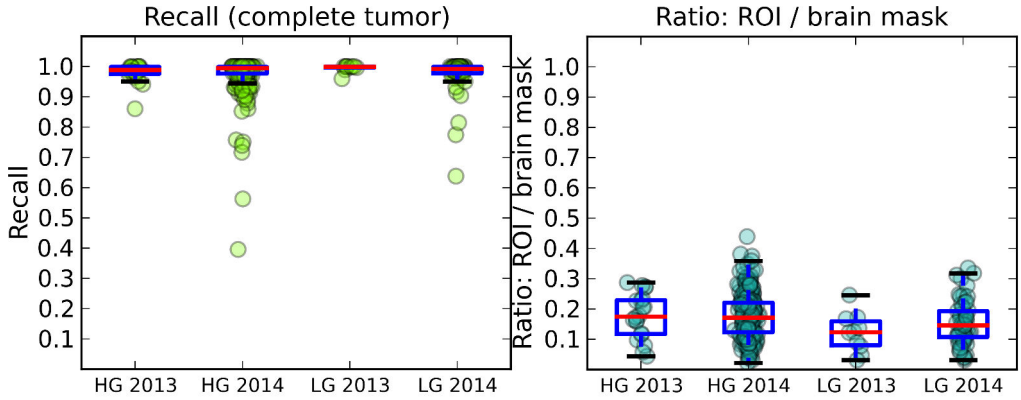

Figure 2.11: Left: recall of glioma detection for the complete tumor, reported for cases grouped by year and grade. Right: volume ratio of detected ROI over the brain mask.

2.3.2 Pre-processing pipeline

Every image is affinely registered to an MNI atlas and warped to the same reference space. A resampling to 2 mm isotropic resolution is performed to decrease computation time at little accuracy cost: most MR channels were not acquired in 1 mm isotropic resolution, and interpolation artefacts are already visible on few MRI. A robust image normalization ensures that inter-patient patch distance is meaningful: non-zero intensity are clipped below 1% and above 99% quantiles, then mean and standard deviation are set to the values reached by averaging over training data, respectively 360 and 120. Image normalization is performed separately for each MR channel, to allow an equal contribution of each channel in the patch distance. Finally, for the training cases, only the segmentation of the tumor compartments is originally available, so healthy tissues are automatically segmented using FSL FAST [Zhang 2001] to avoid under-segmentation by partitioning the background [Wachinger 2014b].

2.3.3 Glioma detection benchmark

ROI are automatically defined by the glioma detection algorithm described in Section 2.2.2. Recall, i.e. the proportion of tumor voxels which are successfully enclosed in the ROI, is evaluated on the 2013 and 2014 training datasets. Considering all cases, *i)* regions-of-interest occupy less than 17% of the brain masks, which effectively results in lower computational burden, *ii)* recall is higher than 97% for every tumor region. Individual results are summarized in Figure 2.11.

2.3.4 Post-processing pipeline for glioma segmentation

Segmentation results are smooth by design due to patch overlap, so post-processing is limited. The pipeline consists in warping back the probabilistic label maps to the 1 mm isotropic patient space, and then keeping at most the two biggest connected

components of the segmentation: if the volume of the second component is less than 3 cm^3, only the first component is kept.

2.3.5 Segmentation benchmark

The proposed approach is benchmarked on the 2013 Challenge dataset against the top-performing documented methods, according to the ranking found on the MICCAI 2013 BraTS benchmark website [Menze 2014] on April 28, 2015: [Kwon 2014] is the highest ranked semi-automatic method, with Dice overlap of 88%/83%/72% for CT/TC/EC respectively ; [Tustison 2014] is the highest ranked automatic method, with Dice overlap of 87%/78%/74%. Our automatic method achieves state-of-the-art performance, with Dice overlap of 87%/77%/73%, in a shorter running time. Fast approximate nearest-neighbour search, especially multiple randomized k-d trees for high dimensional data [Muja 2014], allows a single-thread implementation of the proposed patch-based segmentation to achieve competitive running times: *i)* 16 minutes on average for HG, and *ii)* about 1 minute for LG due to fewer selected atlases, using a single core (2.66 GHz) and less than 6 GB of RAM. Moreover, given access to a computer grid with at least as many cores as there are selected atlases, the proposed stratification scheme allows to achieve running times close to one minute. Segmentation results[3] are shown in Figure 2.12, and box plots[4] are displayed for HG in Figure 2.13 and Figure 2.14.

2.3.6 Parameter setting

The optimal set of parameters is determined, using leave-one-out on the 2013 training dataset of HG cases, by changing one parameter at a time and defining its optimal value based on Dice overlap. No post-processing is applied. Moreover, binary label maps are used, which results in lower Dice overlap than with probabilistic label maps.

2.3.6.1 Atlas selection: number of selected training cases

Given a target case, a number of similar training cases are selected to perform patch-based segmentation. The selection first considers training cases of the same grade as the target, and once the training dataset is exhausted, training cases of a different grade start to be selected. The influence of atlas selection count, i.e. the number of selected training cases, is studied in Figure 2.15. For HG target, an optimum of Dice overlap is reached for 28 training cases if we only consider the complete tumor, and for 24 training cases (20 HG, 4 LG) if we simultaneously consider all the tumor regions. For LG target, Dice overlap is optimal once 3 LG training cases are selected, and then tends to decrease, especially as HG cases are selected: it is mostly detrimental to segment LG tumor core based on HG training cases.

[3] All the segmentation results are shown in Appendix A for the 2013 validation dataset.
[4] A table with average assessment measures is given in Appendix A for all the datasets.

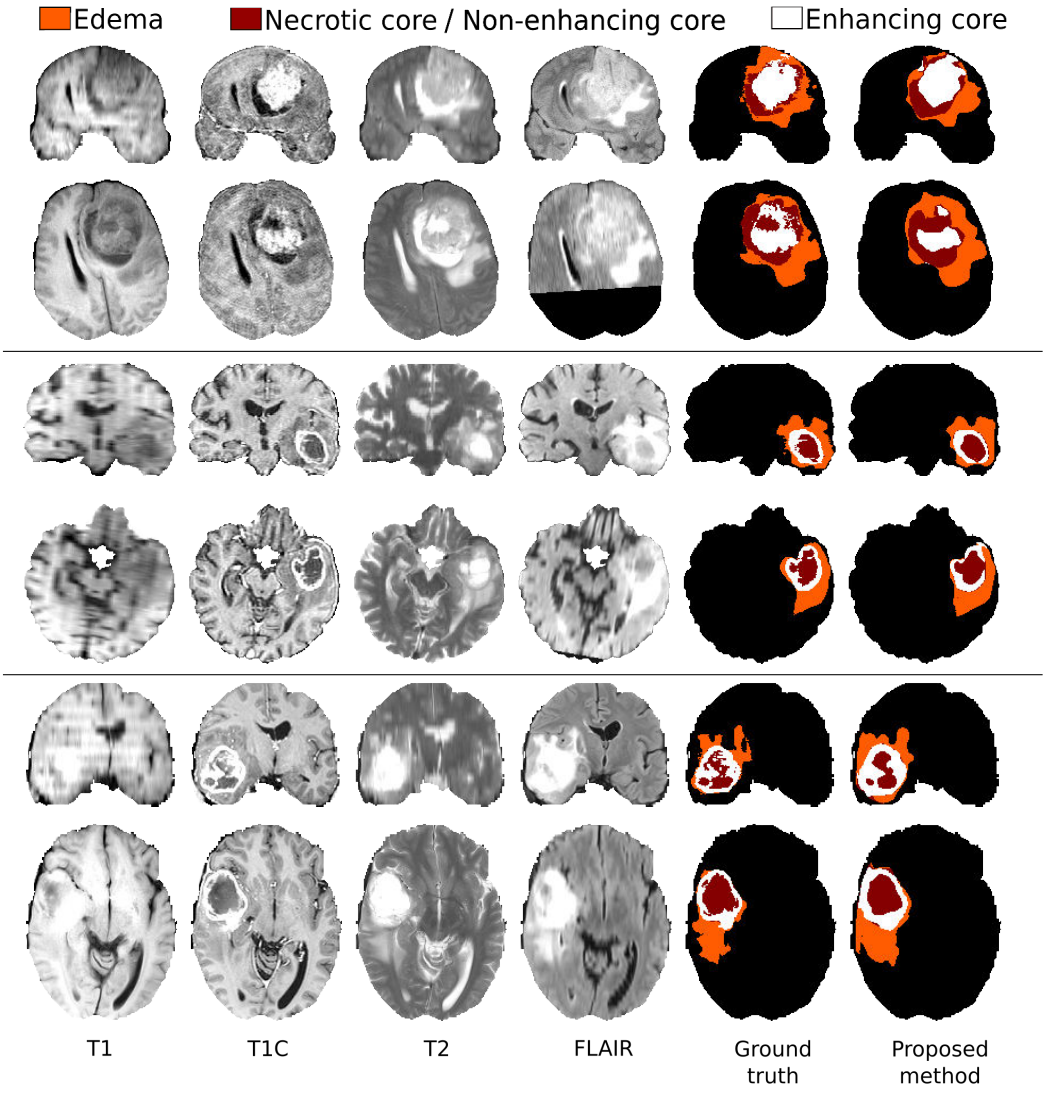

Figure 2.12: Segmentation of 3 HG cases (coronal and axial views).

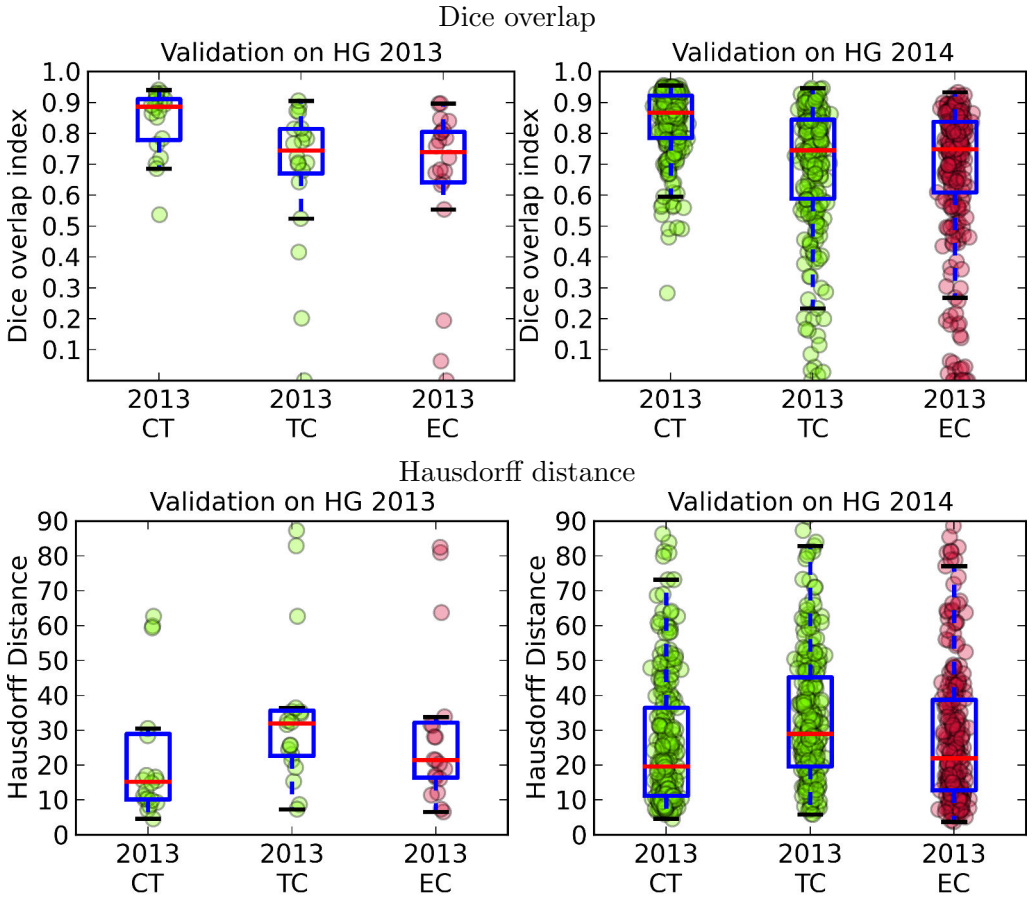

Figure 2.13: Box plots of Dice overlap and Hausdorff distance for HG cases. The test dataset consists of the 2013 (left) and 2014 (right) MICCAI BraTS benchmark Evaluation datasets. In abscissa of each plot, 2013 indicates the year used for the training atlases, here the 2013 Evaluation dataset ; CT/TC/EC denotes respectively complete tumor, tumor core, and enhancing core.

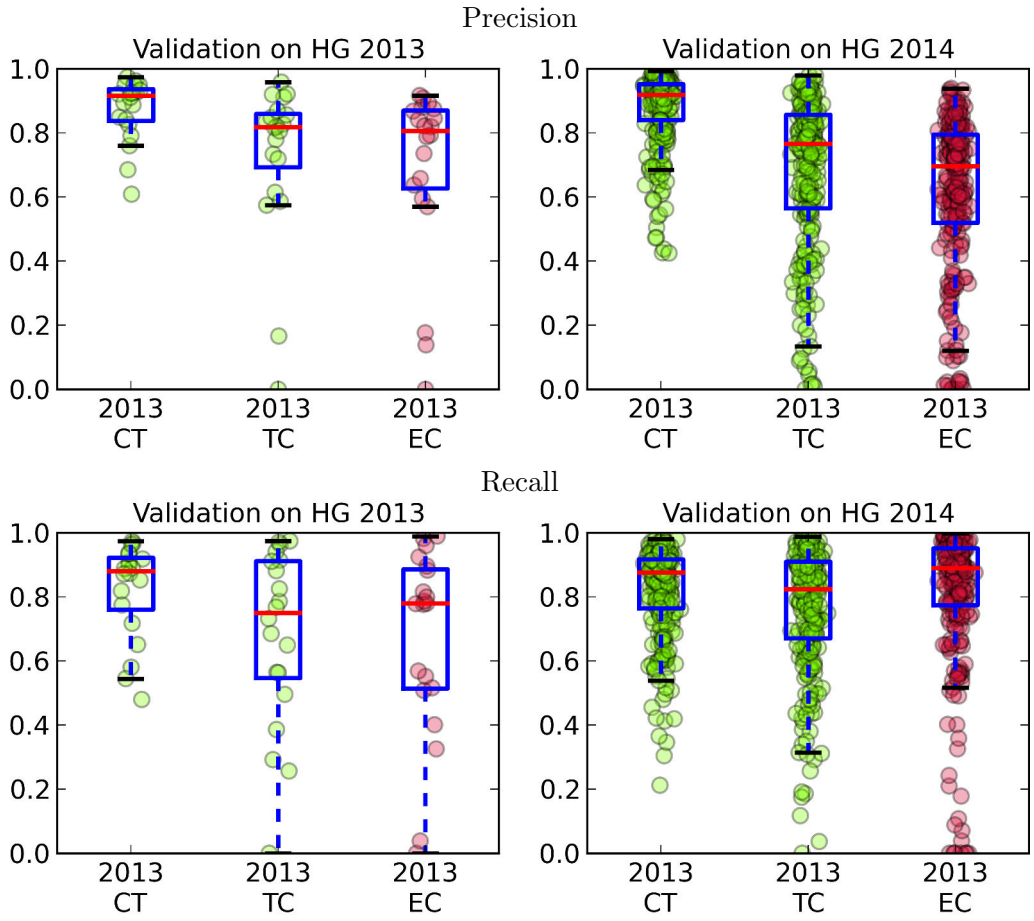

Figure 2.14: Box plots of Precision and Recall for HG cases. The test dataset consists of the 2013 (left) and 2014 (right) MICCAI BraTS benchmark Evaluation datasets. In abscissa of each plot, 2013 indicates the year used for the training atlases, here the 2013 Evaluation dataset ; CT/TC/EC denotes respectively complete tumor, tumor core, and enhancing core.

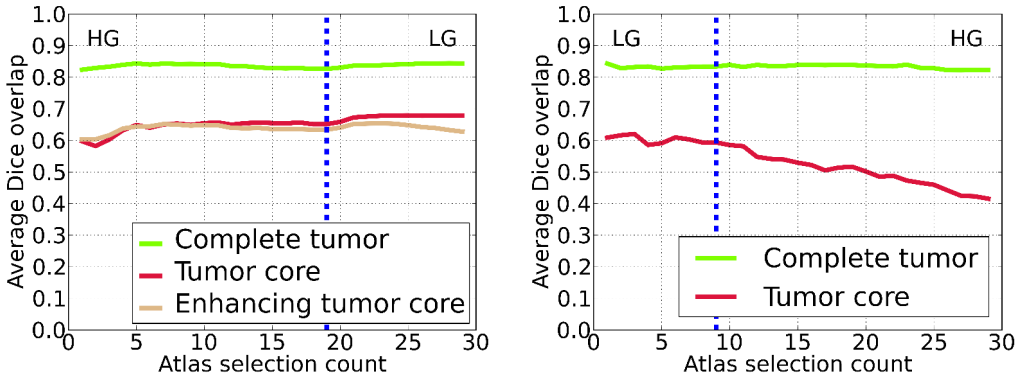

Figure 2.15: Influence of atlas selection count, in leave-one-out. Vertical lines mark transitions between tumor grades of selected cases.

2.3.6.2 Feature: patch width, multi-scale patches, and invariance to cube isometries or sagittal plane symmetry

Different patch widths were tested, as shown in Figure 2.16. In the left pane for which features are not invariant to cube isometries or sagittal plane symmetry, the optimal patch width is 3, which is the smallest tested width. The use of multi-scale patches, including a 3-voxel-wide central patch, results in Dice overlaps similar to these obtained with conventional 5-voxel-wide patches. In the right pane for which features are invariant, the optimal patch width for conventional patches is 5. Comparable yet slightly better results are obtained with multi-scale patches. Invariance decreases Dice overlap for the smallest conventional patches, but increases Dice overlap for larger conventional patches and for multi-scale patches. Intuitively, matching larger patches is sufficiently constrained so that patch match actually benefits from an augmentation of the training dataset via cube isometries or sagittal plane symmetry. In the end, 9-voxel-wide multi-scale patches, with a 3-voxel-wide central patch, are used along invariance to cube isometries for pathological training patches, and invariance to sagittal plane symmetry for healthy training patches.

2.3.6.3 Feature augmentation weights α and β

With feature augmentation, a weighted L_2-norm replaces patch distance, with weights α and β respectively for atlases of label spatial distribution and for class conditional Student intensity likelihoods. The influence of both weights is studied in Figure 2.17. Increasing α leads to higher Dice overlap, with visually similar segmentation results in the range $[8,\ 14]$. When α tends to infinity, the segmentation is only driven by label spatial priors, and Dice overlap measures tend to 0. Decreasing β leads to higher Dice overlap. Indeed, class conditional Student intensity likelihoods do not allow to discriminate between classes in areas where clusters overlap (Figure 2.5). When β tends to infinity, Dice overlap measures fall off to

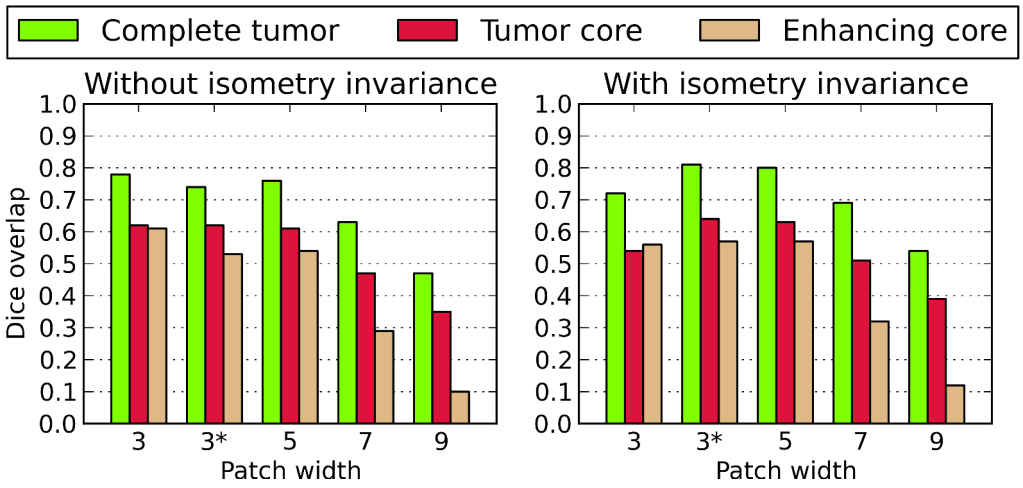

Figure 2.16: Influence of patch width, multi-scale enhancement, and invariance to cube isometries on the average Dice overlap for HG target cases. The "3*" index corresponds to multi-scale patches with a central patch of indicated width. The right part uses invariance to cube isometries (for pathological patches) and sagittal plane symmetry (for healthy patches).

42%, 26% and 32% for complete tumor (CT), tumor core (TC), and enhancing core (EC) respectively. In the end, $\alpha = 12$ and $\beta = 2$ are chosen, since a low non-zero value for β seems to result in slightly smoother segmentations.

2.3.6.4 Training patch selection: quantile of Student distributions and minimal patch label purity

The effect of training patch selection is mostly visible if the pre-processing pipeline accumulates approximations (interpolation of images, and binarization of warped label maps). In our case, its influence on Dice overlap is marginal for conventional or multi-scale patches. Dice overlap for enhancing core could benefit from a higher

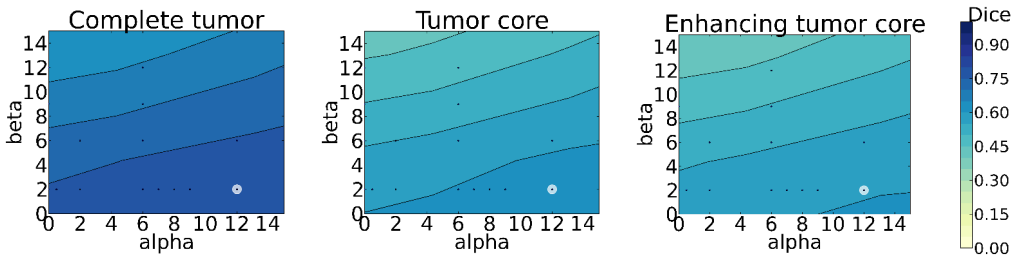

Figure 2.17: Influence of feature augmentation weights (α, β) on the average Dice overlap for HG target cases. Chosen parameter values (α, β) are circled in white.

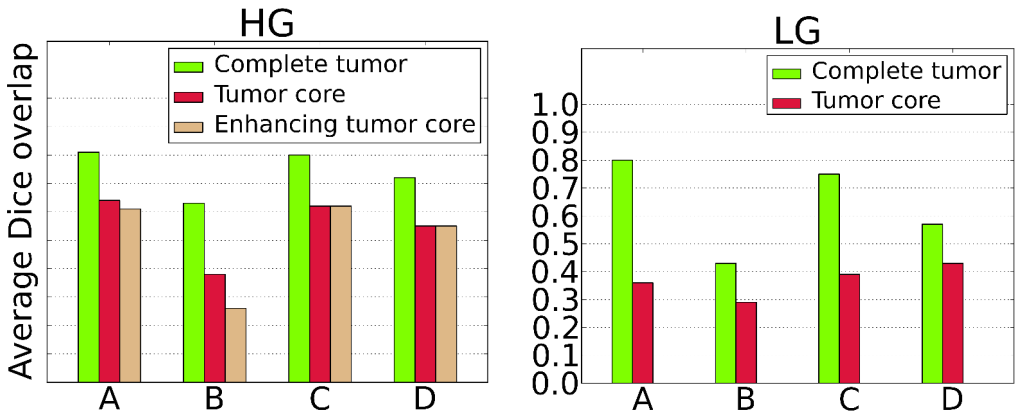

Figure 2.18: Comparison of stratification schemes A to D.

minimal patch label purity, but this would decrease Dice overlap for complete tumor and tumor core. In the end, the quantile of Student distributions is equal to 60%, and the minimal patch label purity is equal to 2/3, to ensure shorter running time at the cost of a marginal Dice overlap decrease.

2.3.6.5 Vote aggregation: stratification scheme

The proposed stratification scheme is compared to simpler schemes, with the same number of patch matches for a fair comparison. No atlas selection is performed: the number of training cases is 29 due to leave-one-out. The four schemes are: *A)* stratification with respect to training cases (29) and labels (7): 1 patch match, *B)* stratification only with respect to labels (7): 29 patch matches, *C)* stratification only with respect to training cases (29): 7 patch matches, *D)* no stratification: 203 patch matches.

Scheme C is comparable to our naive patch-based method [Cordier 2013], which ranked 5th in 2013, and would rank 18th as of April 2015. The definition of σ_n^2 was different, votes were heuristically scaled, and results required stronger post-processing. On the 2013 training data in leave-one-out, Dice overlap measures were similar to those reported for scheme C: 79%, 60% and 59% for CT, TC and EC respectively.

As shown in Figure 2.18, a stratification with respect to labels (scheme B) decreases Dice overlap as compared to no stratification (scheme D). On the other hand, a stratification with respect to training cases (scheme C) consistently leads to higher Dice overlap. The situation regarding the tumor core on LG cases is particular since the distinction between edema and tumor core for this grade is especially debatable. Finally, for HG cases, a stratification with respect to both labels and training cases (scheme A) leads to 1% and 2% higher Dice overlap for CT and TC respectively, and 1% lower Dice overlap for EC. In the end, the chosen scheme is A.

2.4 Discussion

Overall running time is highly competitive, mostly due to resampling to 2 mm isotropic resolution, target and training patch selection, and atlas selection. However, these methods show some limitations. First, image resampling leads to a loss of the finer structures, which can be detrimental to the segmentation of the enhancing core. For instance, in the second row in Figure 2.12, the enhancing core signal is subtle in T1C and only partially retrieved by our method. Second, target patch selection depends on a few automatic post-processing steps which could fail and hinder the segmentation process. Third, training patch selection leads to a decrease of Dice overlap, although very marginal in our experiments. Fourth, atlas selection could not be properly studied since the 2013 training dataset is small: the optimal atlas selection count would differ with a larger training dataset.

Patch distance is ultimately the canonical L_2-norm. Based on the mean and covariance of the Student distributions, Mahalanobis patch distances have been tested but resulted in a radical drop of performance. This might be due to an inconsistent definition of the minimal patch distance $\sigma_n^2(x)$ in the case of class-specific patch distances.

Patch matching heavily relies on fast approximate nearest neighbor search for high-dimensional data. In our experiments, approximate search resulted in segmentation performance indistinguishable from exact nearest neighbor search. However, this might be due to the fact that our approach only considers distances between patches to perform the segmentation. This could lead to drastically different results for any application which would make use of the nearest-neighbor per-se.

Finally, the proposed approach has lead to state-of-the-art brain tumor segmentation results on the BraTS benchmark, and appears to work fine on the clinical data of our collaborators. However, a larger test dataset, with a reliable and hidden expert ground truth, would be necessary to be able to distinguish between the performances of the top-ranked algorithms of the BraTS benchmark. More importantly, given the amount of machine learning segmentation methods, it is essential that test datasets are made available through third-parties as is the case with the BraTS benchmark: it is not uncommon that learning-based procedures overfit, which makes irrelevant any comparison of segmentation results on the training datasets.

2.5 Summary and Conclusions

We presented a generic approach to adapt patch-based techniques to the segmentation of pathological cases. This approach has lead to state-of-the-art brain tumor segmentation results in a fully automatic setting. The procedure requires very limited to no prior learning, which limits the risk of overfit and can take advantage of a constantly growing database of annotated cases. Due to patch overlap, segmentation results only require minimal post-processing, which confirms the robustness of the proposed segmentation approach. Patch extraction and matching benefits from automatic patch selection ahead of time, and fast approximate nearest-neighbor search, which results in a highly competitive overall running time. With stratification, patch match is effectively embarrassingly parallel, which results in even shorter running times (close to one minute) using a computer grid.

The possibility to automatically define precise regions of interest in a matter of seconds, using a probabilistic model similar to the one proposed for glioma detection, with a spatially-varying prior learnt on the training data, could be of interest to other medical imaging algorithms. The proposed approach could also be incorporated within the Modality Propagation framework [Ye 2013], therefore opening new perspectives for the application of patch-based methods to the simulation of medical images including brain pathologies.

Chapter 3

Further Analysis regarding our Glioma Segmentation Approach

Contents

Abstract

In this chapter, we propose to question and analyze further a few choices and parameter settings regarding the glioma segmentation approach described in Chapter 2.

Some of these interrogations are tied to the validation framework and concern the generalization of the method: how important is it to perform leave-one-out validation at every possible stage of the method? Would we be able to retrieve the optimal parameters and a good approximation of our Dice score, if we were to perform leave-one-out in a less comprehensive manner[1]? How reliant is the method with respect to pre-processing and post-processing of images? Images were downsampled to 2 mm resolution, how good are segmentation results with the original 1 mm resolution? How are the results affected by a change of the training dataset?

Other interrogations deal with implicit choices regarding the method in itself: what happens if the similarity measure between patches is not the L_2 norm? What happens if the decay parameter $\sigma_n^2(x)$ is defined differently, for instance as $\sigma^2(x)$? What if the approximated marginalization were to take into account more patch matches? How good is the glioma detection framework for segmentation, wouldn't it be enough to achieve state-of-the-art results? Finally, is the weight β which is used for feature augmentation really necessary? We detail a subtle and somewhat striking impact of feature augmentation on the optimal number of atlases.

[1] For instance, would it be sufficient to leave the target case out only during patch match?

3.1 Questioning the validation framework and generalization of the proposed method

3.1.1 Importance of leave-one-out for parameter setting

In Chapter 2, parameters such as the optimal number of atlases are set through a comprehensive leave-one-out validation framework. Indeed, the current target case, which is to be segmented, is excluded from the set of training cases *i*) for the construction of atlases of label spatial extent, *ii*) to fit Student distributions (for glioma detection, atlas selection, and feature augmentation), *iii*) and during patch-match. This can be a tedious procedure when evaluating hundreds of cases, as with the 2014 Evaluation dataset, especially due to steps *i* and *ii*.

However, our intuition is that leave-one-out should mostly matter during patch match. Indeed, the Student distributions and atlases of label distributions are built based on the processing of 314 cases (and 313 with the leave-one-out), so they should not be significantly influenced by the removal of a few images. To confirm this intuition, we have conducted an experiment with comprehensive leave-one-out and with leave-one-out only during patch match. As expected, this has led to equivalent quantitative results. Moreover, the optimal number of atlases for a high-grade target case is the same (24 atlases are selected), as shown in Figure 3.1, and for a low-grade target case almost the same (3 atlases with comprehensive leave-one-out, 2 atlases otherwise). When leave-one-out is performed comprehensively, average Dice score is 1% lower for high-grade target cases and 3% lower for low-grade target cases. In the end, we would retrieve the optimal parameters and a good approximation of our Dice score, if we were to perform leave-one-out in a less comprehensive manner. This experiment confirms that a simpler validation framework, in which leave-one-out is only performed during patch match, could be used to look for optimal parameters.

3.1.2 Pre-processing: image standardization

We do not provide any quantitative result regarding pre-processing. However, we would like to stress out how important pre-processing and image standardization are. As shown in Figure 3.2, image intensity can vary tremendously from one image to another, even among cases from the same database. The combination of a translation and a global scaling of image intensity is sufficient to make image intensity distributions look more similar. More complex procedures can be detrimental, since image standardization remains a structural problem for any segmentation algorithm working with pathological cases imaged with Magnetic Resonance Images. For instance, our pre-processing pipeline, recalled in Figure 3.3, does not attempt to correct the bias field since it usually tends to smooth out the tumor signal, especially on T_2-weighted and T_2-FLAIR MR images. With a different dataset, bias field correction of at least the T_1-weighted images might me mandatory to achieve acceptable results.

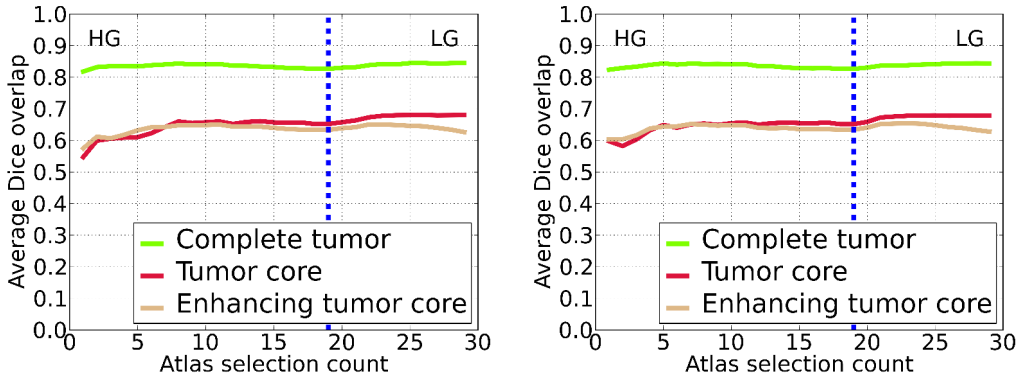

Figure 3.1: Influence of a comprehensive leave-one-out on atlas selection count. Leave-one-out is performed either only during patch match (left) or at every possible stage of the method (right) Vertical lines mark transitions between tumor grades of selected cases.

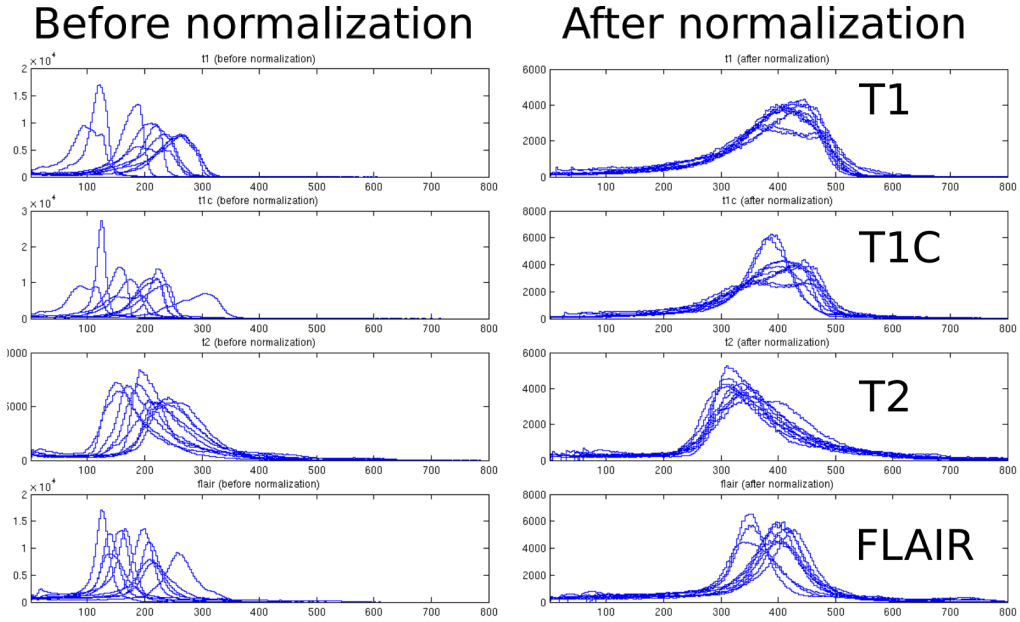

Figure 3.2: Effect of intensity normalization on histograms of image intensity for 10 cases of the MICCAI 2013 BRATS benchmark (Challenge dataset). Before (left) and after (right) intensity normalization. From top to bottom, MR sequences: T_1, T_1 with Gadolinium, T_2, T_2-FLAIR. After normalization, histograms have the same robust mean and robust variance. The skewness is not corrected and mostly differs from one MR sequence to another.

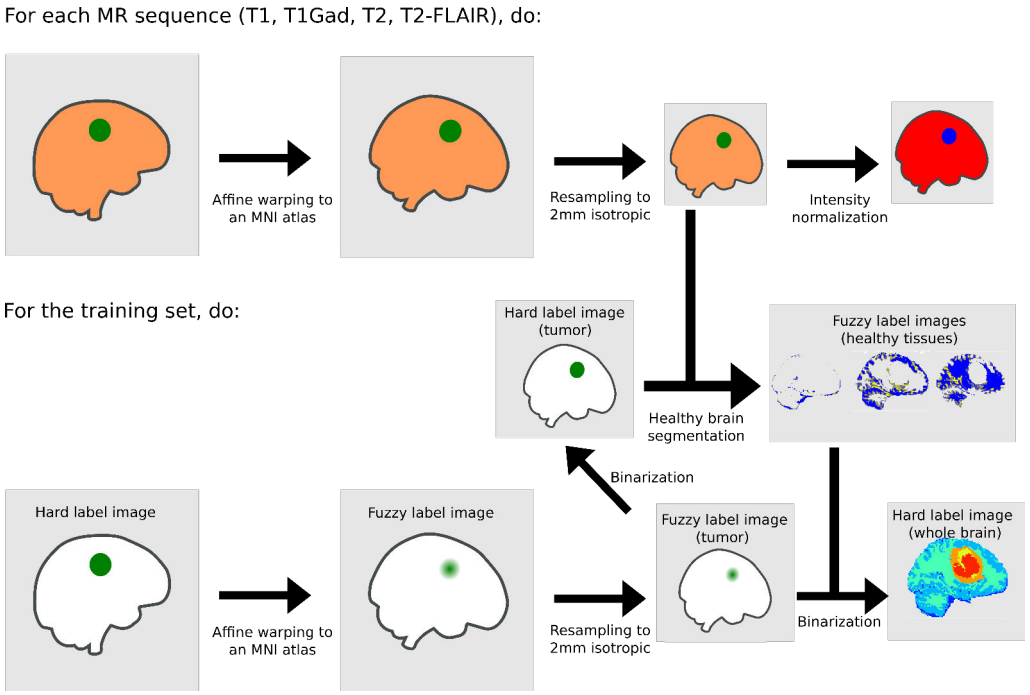

Figure 3.3: Pre-processing pipeline. Images are warped to the MNI space, and subsampled to 2 mm isotropic resolution. MR images are normalized so that robust mean and robust variance are consistent inter-modality and inter-patient. The ground truth consists of a hard label image of the tumor compartments in the patient space. A fuzzy and a hard segmentation of the whole brain, including healthy tissues and tumor compartments, are computed in the MNI space.

3.1.3 Importance of post-processing

For post-processing, only the two biggest connected components are kept as a brain tumor segmentation. Indeed, most tumors present in the BraTS datasets are mono-focal, some are bi-focal, none present more than two focii. However, on a different dataset, post-processing might be detrimental, so it is legitimate to question the post-processing. We have evaluated our segmentation method with and without post-processing. In the end, Dice scores would be about 0.5% higher (training data in leave-one-out, but also test data) with post-processing. So the post-processing is situational: it could be slightly beneficial for a given dataset, but never mandatory to achieve good results with the proposed method.

A few other post-processing techniques were tried but were not kept since they did not improve much the results. For instance, depending on the clinician's expectation, the segmentation of the enhancing rim can be visually refined with a heuristic procedure. Indeed, the proposed patch-based approach analyses 2 mm images, thus finer regions of the enhancing rim, more clearly delineated on 1 mm images, may appear smooth and coarsely segmented. To alleviate this problem, a robust image intensity threshold can be defined based on the segmentation of the tumor core compartments and the 1-mm isotropic contrast-enhanced T_1-weighted MR image: among the voxels assigned to the tumor core, any voxel with an intensity above the threshold is specifically assigned to the enhancing tumor core, while any voxel with an intensity below the threshold is assigned to the non-enhancing tumor core. On average, Dice score for the enhancing rim is 1% higher, with a debatably more appealing segmentation of the enhancing rim, as illustrated in Figure 3.4.

3.1.4 Down-sampling

The impact of down-sampling to 2 mm was quantified in recent experiments. Down-sampling results in a significant decrease of the average Dice score for complete tumor (CT) and enhancing core (EC), and almost an identical average Dice score for tumor core (TC). More precisely, with 1 mm resolution, the optimal number of atlases is lower for high-grade target cases (14 atlases are selected, instead of 24 atlases for 2 mm resolution) and identical for low-grade target cases (3 atlases are selected), as shown in Figure 3.5. The average Dice score obtained with leave-one-out on the training dataset is:

- for HG target cases: 1.4% higher for CT, 0.1% higher for TC, and 4.4% higher for EC,
- for LG target cases: 0.6% lower for CT, 2% lower for TC.

In the end, the impact of down-sampling is important for the segmentation of the enhancing tumor core for high-grade target cases. Qualitatively, segmentations look similar, with slightly more precise borders for the enhancing tumor core, as illustrated with the exampled in Figure 3.6. However, the running time is longer, which defeats one of the purposes of our patch-based method.

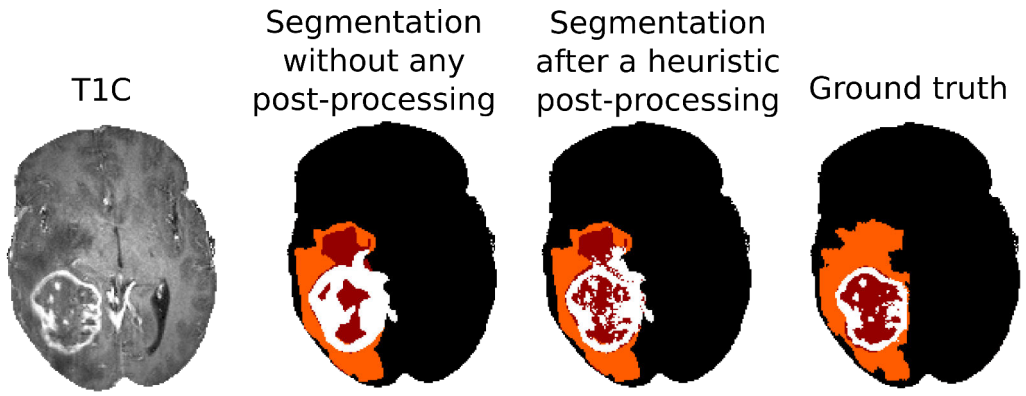

Figure 3.4: Effect of a heuristic post-processing procedure (automatic thresholding). From left to right: *i*) 1-mm isotropic T_1-weighted MRI with Gadolinium, *ii*) our segmentation without any post-processing, *iii*) our segmentation after a heuristic post-processing procedure, *iv*) ground truth, which is a consensus of four expert manual segmentations. The original segmentation is smooth (*ii*). After the heuristic automatic post-processing, the segmentation of the enhancing rim appears, in this specific case, more precise. However, a lack of smoothness can make the segmentation less appealing. Moreover, on average, assessment measures such as Dice scores are almost not effected by this procedure.

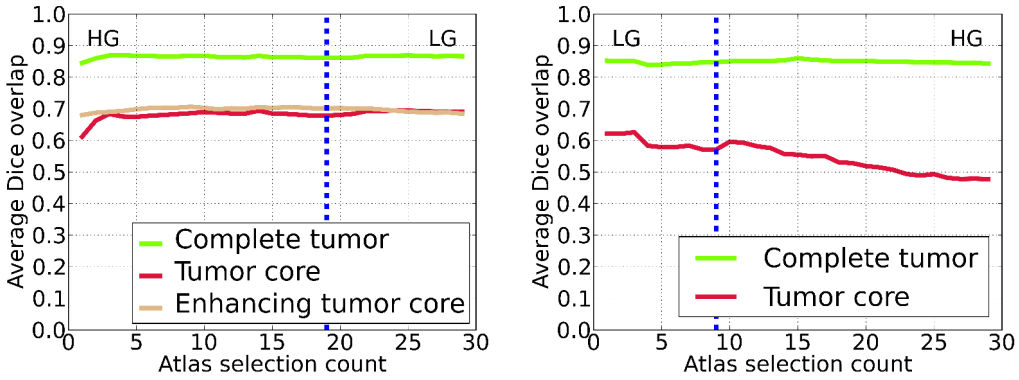

Figure 3.5: Influence of atlas selection count, in leave-one-out. Dataset is processed at a 1 mm isotropic resolution. Vertical lines mark transitions between tumor grades of selected cases.

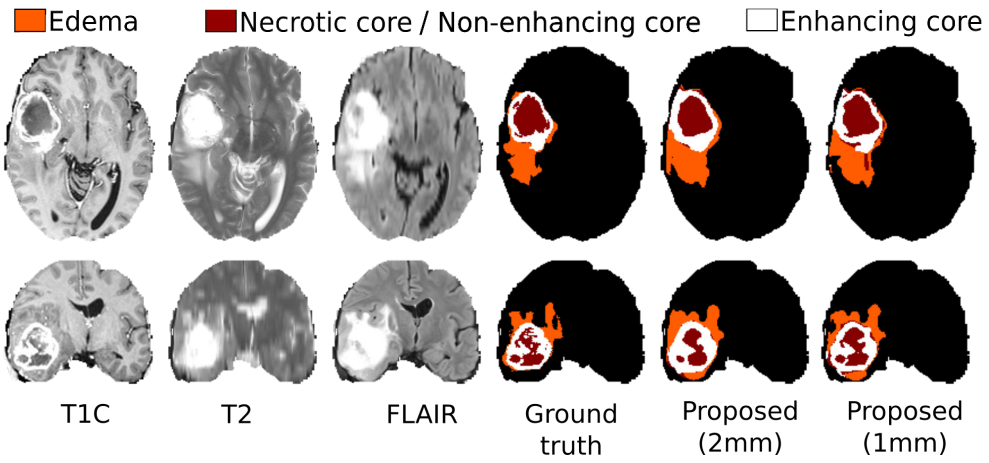

Figure 3.6: Influence of the resolution at which the segmentation is performed, observed on a high-grade case (HG-0001).

3.1.5 Changing the training dataset

Our approach is based on patches extracted from the 2013 training dataset, for which the ground truth is obtained as a consensus of four manual expert segmentations. The 2014 training dataset is only used for validation purpose, and for a few learning stages: the creation of atlases of label spatial distribution, and the fit of Student distributions to average patch intensity. As shown in Figure 3.7, we tried to base our approach on patches extracted from the 2014 training dataset, but segmentation results were less satisfactory: average Dice overlap was slightly lower for 2013 target cases, and slightly higher for 2014, however Dice overlap variance was higher for both years. One possible explanation is that the ground truth of the 2014 training dataset, which was obtained as a consensus of randomly-selected top-ranking algorithms, could be less reliable. Another explanation is that the two types of delineations (BraTS 2013 and 2014) are based on slightly different assumptions. Nevertheless, segmentation results obtained on the 2 Challenge datasets are consistent, and assessment measures are similar, irrespective of the choice of the training dataset (Figure 3.7), which seems to indicate that the proposed method does not suffer from overfitting.

3.2 Questioning implicit choices regarding the method in itself

3.2.1 Mahalanobis distances for patch similarity

For feature augmentation, weights α and β were introduced. Similarly, for patch features, each MR channel could be weighted differently, when a feature vector is built from a multi-channel patch. Ideally, the weight assigned to each modality would

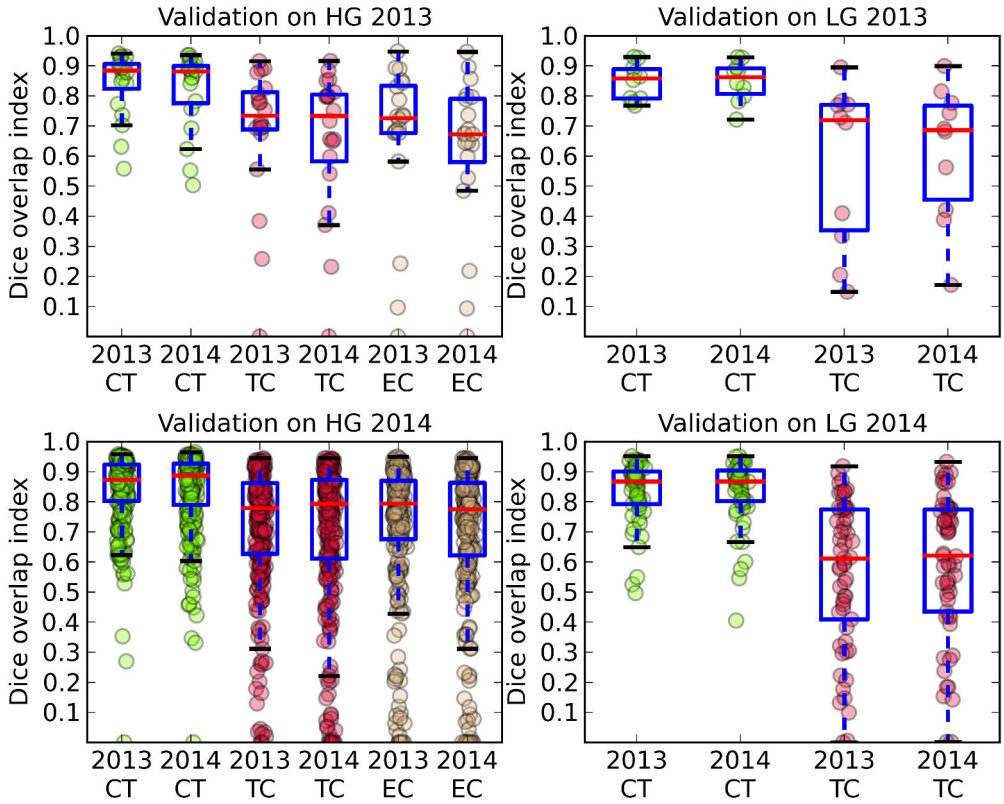

Figure 3.7: Box plots of Dice overlap. In abscissa of each plot, 2013/2014 indicates the year used for the training atlases, CT/Core/ET denotes respectively complete tumor, tumor core, and enhancing tumor core. The test dataset consists of the 2013 (top row) and 2014 (bottom row) MICCAI BRATS benchmark Evaluation datasets. Left column: high-grade cases ; right column: low-grade cases.

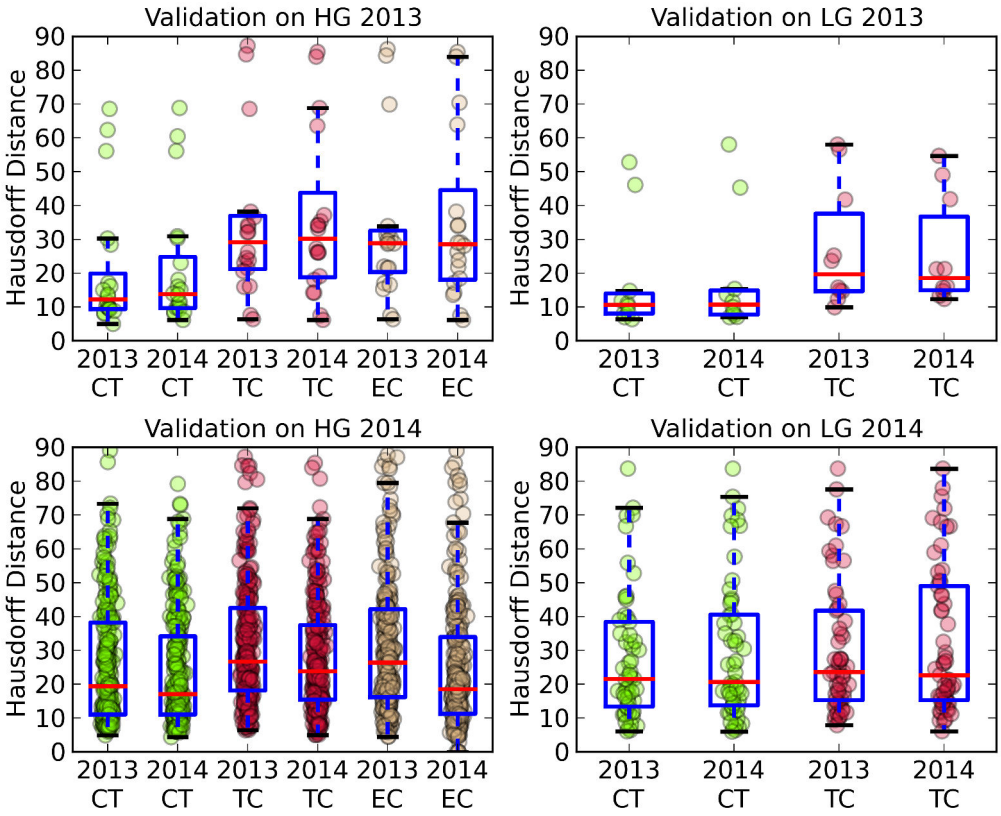

Figure 3.8: Box plots of Hausdorff Distance. In abscissa of each plot, 2013/2014 indicates the year used for the training atlases, CT/Core/ET denotes respectively complete tumor, tumor core, and enhancing tumor core. The test dataset consists of the 2013 (top row) and 2014 (bottom row) MICCAI BRATS benchmark Evaluation datasets. Left column: high-grade cases ; right column: low-grade cases.

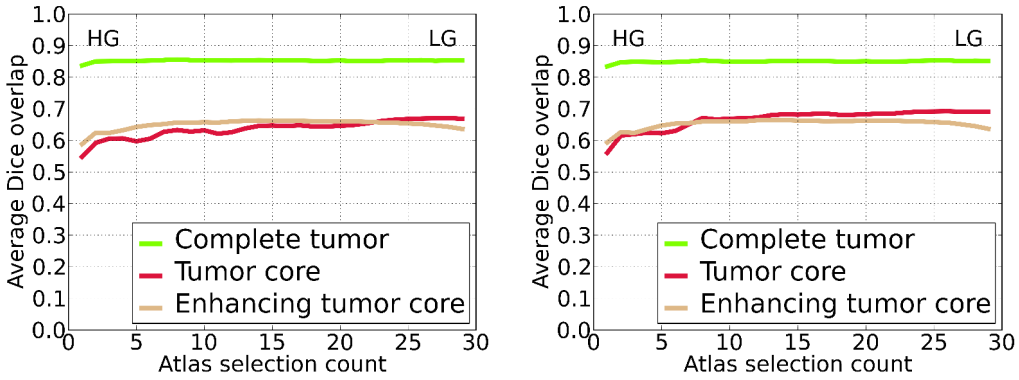

Figure 3.9: Influence of the definition of σ as $\sigma(x)$ (left) or $\sigma_n(x)$ (right) on the Dice overlap. Results are presented for different atlas selection counts. Leave-one-out is performed during patch match. Vertical lines mark transitions between tumor grades of selected cases.

be learnt by using training images. However, the problem of optimizing weights for each modality is computationally expensive. We tried a simpler approach instead: we replaced the L_2 norm used for patch distance with Mahalanobis norm taking into account the relative variability of each modality. This can be achieved with the means and covariance matrices of the fitted Student distributions. However, this approach was not successful leading to numerous false positives and worse distinction between tumor compartments. Our belief is that the procedure to define the decay $\sigma(x)$ used for patch-based voting becomes more complex than expected when patch similarity depends on the label assigned to the current training patch.

3.2.2 A different definition for sigma

σ is estimated on the fly for each target case as $\sigma_n(x)$ such that:

$$\sqrt{2}\sigma_n(x) = \min_y d\left(S(\mathbf{J}, x),\ S(\mathbf{I_n}, y)\right)$$

where n indexes atlases and y spatial position in the MNI space. However, another natural definition for σ could be $\sigma(x)$ such that:

$$\sqrt{2}\sigma(x) = \min_{n,y} d\left(S(\mathbf{J}, x),\ S(\mathbf{I_n}, y)\right)$$

We tested both $\sigma(x)$ and $\sigma_n(x)$. The evaluation of the optimal number of atlases is reported in Figure 3.9. The average Dice scores are identical except for the tumor core for high-grade target cases: using $\sigma_n(x)$ leads to a 3% increase of Dice score.

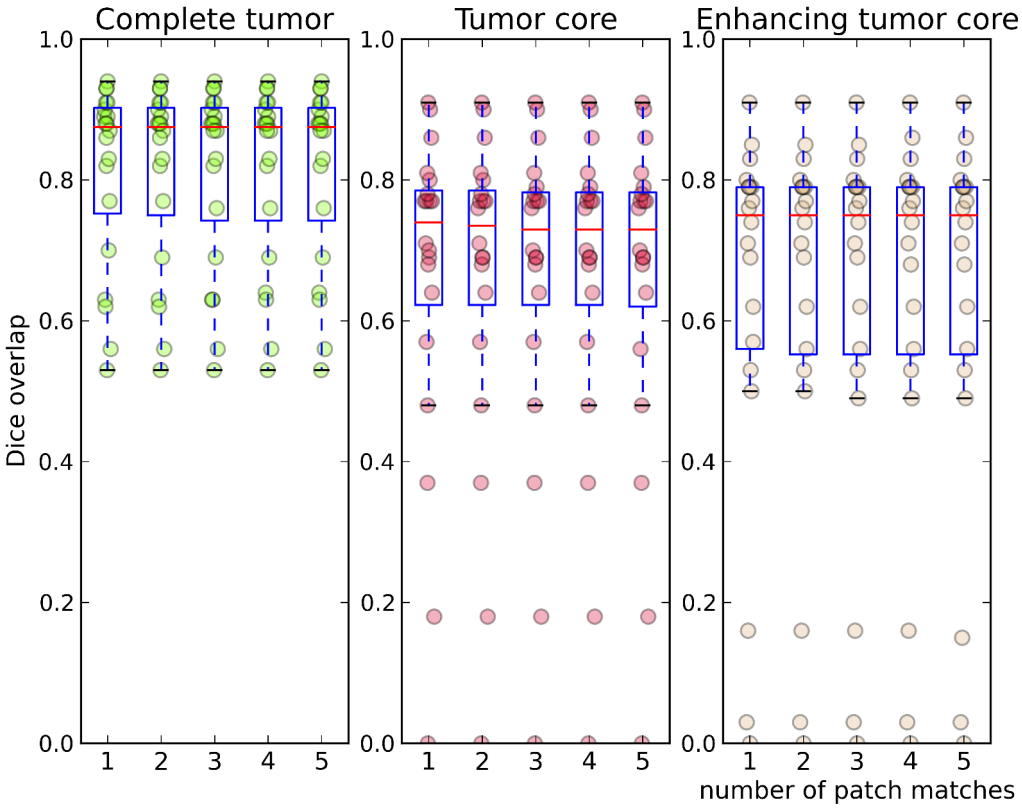

Figure 3.10: Influence of the number K of matched patches on the average Dice overlap for high-grade target cases. From left to right: complete tumor, tumor core, enhancing tumor core. From left to right inside each plot: increasing number of matched patches, from $K = 1$ to $K = 5$ closest patches per pair of training cases and labels. Superimposed as disks are individual results obtained for every case of the training dataset. The Dice overlap is stable.

3.2.3 Approximate marginalization: number K of closest patches

With vote stratification, each pair of training case and label results in $K = 1$ matched patch, and therefore one value in the approximate marginalization. To study the sensitivity of the method to the number of matched patches, we tested every value between $K = 1$ and $K = 30$ closest patches, for a number of training cases equal to 10, 24 and 29: the average Dice score for the high-grade target cases is the same for the complete tumor, and decreases by at most 1% for the other labels. Figure 3.10 shows the results obtained for $K = 1$ to $K = 5$ matched patches. In the end, it is sufficient to consider $K = 1$ matched patch per pair of training case and label, as it is the case for the method described in Chapter 2.

3.2.4 Using the glioma detection graphical model for segmentation

With the glioma detection graphical model, posterior probability estimates are obtained for each label. A natural question would be: what is the average Dice score when directly using these posterior probability estimates for segmentation? Actually, in this case, the Dice score is the limit when β tends to $+\infty$, which is low (cf. Section 2.3.6.3). This is due to the large amount of false positives. Moreover, the false positive removal which is applied at the end of glioma detection is only relevant for the definition of a ROI: if applied for segmentation purposes, the number of false negatives would obviously increase by a large margin. This would not allow to use the posterior probabilities obtained for glioma detection to perform state-of-the-art glioma segmentation.

3.2.5 Influence of feature augmentation on atlas selection

When adding new atlases, average Dice score remarkably decreases for tumor core of low-grade target cases, as recalled in Figure 3.11 (already shown in Chapter 2). This is most likely due to the difficulty to distinguish between edema and tumor core for low-grade cases, which do not show any enhancement on the enhanced T_1-weighted MRI, and could lead to an apparent inconsistency of the ground truth segmentation in the training dataset.

Moreover, the decrease is stronger after we have used 9 atlases. This threshold of 9 atlases corresponds to using all low-grade atlases, so any number of atlases lower than 9 means that only low-grade atlases are considered, whereas when more than 9 atlases are used, high-grade atlases are also included. Basically, the explanation is that it is detrimental to try to segment the tumor core of low-grade target cases with the knowledge given by the tumor core segmentations of high-grade atlases. This is in agreement with the fact that the tumor core is not easy to define for low-grade cases: it amounts to differentiating tumor core from edema in absence of any visual clue of enhancing or necrotic tumor cores.

We do observe a very subtle yet similar phenomenon for the enhancing tumor core (EC) of high-grade target cases. In Figure 3.12, average Dice score for EC has a negative slope for a number of atlases between 25 and 29. If the weight β is increased, and therefore patch features are more strongly augmented, this phenomenon is less marked. In the end, the augmentation of features can lead to subtle impacts such as allowing us to select slightly more atlases for the segmentation, which ultimately can lead to better segmentation results.

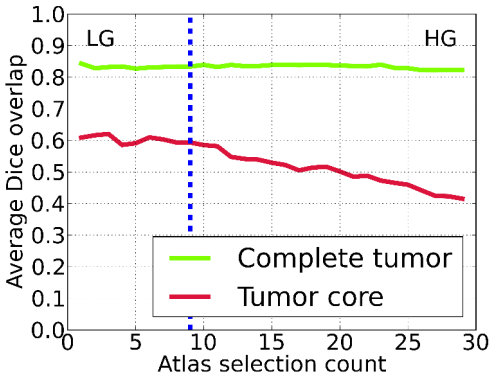

Figure 3.11: Influence of atlas selection count for the segmentation of low-grade gliomas, in leave-one-out. Vertical lines mark transitions between tumor grades of selected cases.

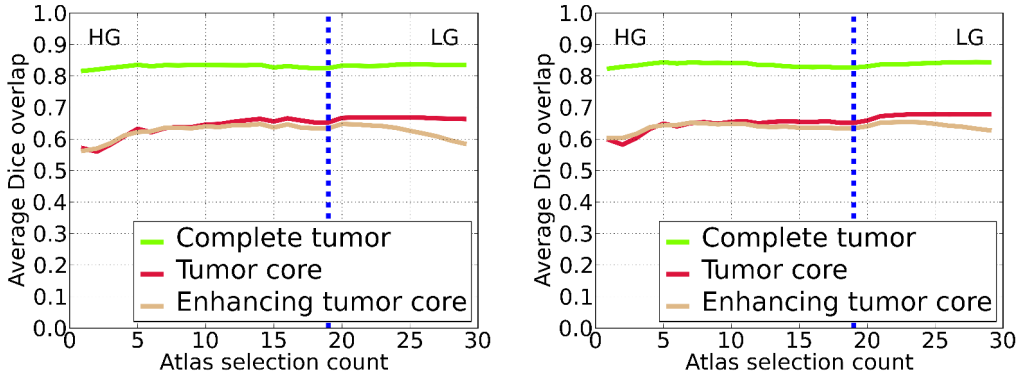

Figure 3.12: Influence of feature augmentation on atlas selection count. The weight for class conditional Student intensity likelihoods is either $\beta = 1$ (left) or $\beta = 2$ (right). Leave-one-out is performed comprehensively. Vertical lines mark transitions between tumor grades of selected cases.

Chapter 4

Magnetic Resonance Image Synthesis of Pathological Cases

Contents

Abstract

This chapter describes a novel generative model for the synthesis, based on a single label map, of multi-modal medical images of pathological cases. Our model builds upon *i)* a generative model commonly used for label fusion and multi-atlas patch-based segmentation of healthy anatomical structures, *ii)* the Modality Propagation iterative strategy used for a spatially-coherent synthesis of subject-specific scans of desired image modalities. The expression Extended Modality Propagation is coined to refer to the extension of Modality Propagation to the synthesis of images of pathological cases. Moreover, image synthesis uncertainty is estimated. An application to Magnetic Resonance Imaging synthesis of glioma-bearing brains is *i)* validated on the training dataset of a Multimodal Brain Tumor Image Segmentation challenge, *ii)* compared to the state-of-the-art method in glioma image synthesis, and *iii)* illustrated using the output of two different tumor growth models. Such a generative model allows the generation of a large dataset of synthetic cases, which could prove useful for the training, validation, or benchmarking of image processing algorithms.

4.1 Introduction

4.1.1 Motivation

The availability of public datasets [Jack 2008, Regan 2011, Menze 2014] of annotated medical images is a determinant factor in the development of medical image computing. For instance, the organization of the Multimodal Brain Tumor Image Segmentation (BraTS) benchmark challenge [Menze 2014] has lead to an important and fruitful research activity in glioma segmentation. However, the creation of a benchmark dataset is costly for obvious reasons: *i)* a large number of cases is required in the training and testing datasets to capture the variability of structures or pathologies, *ii)* images should be annotated by experts, which requires a complex and time-consuming manual work, and can still lead to debatable results (e.g. inter-rater variability in the range 74-85% for glioblastoma [Menze 2014], datasets can contain incorrect segmentations [Wachinger 2015]), *iii)* the distribution of medical images leads to ethical concerns, since the identification of patients may be possible despite anonymization steps such as skull-stripping [Wachinger 2015], *iv)* and finally, ensuring the continued quality of a very large dataset of anonymized images is complex, due to the presence of longitudinal data and duplicates [Toews 2015]. The development of image synthesis could allow *i)* the augmentation of a dataset by including new realistic synthetic cases, and *ii)* the creation of large, annotated, unbiased, anonymized, and easy-to-maintain datasets of synthetic medical images of virtual patients (namely atlases completely generated by a numerical biophysical model). The availability of a large dataset is especially important for pathologies such as glioblastoma which exhibit a high variability of shape and appearance. For reference, the 2013 BraTS challenge only consisted of 20 real high-grade cases for training, and 10 real high-grade cases for benchmarking [Menze 2014]). The 2014 BraTS challenge included about 250 additional cases obtained from The Cancer Imaging Archive [Clark 2013], however other problems arose: the ground truth was unavailable and had to be obtained through a consensus of automatic glioma segmentation algorithms, which resulted in a ground truth of lower reliability [Havaei 2015].

4.1.2 Related work

The problem which is tackled in this chapter is the synthesis of multi-modal medical images of pathological cases, based on a single label map, as illustrated in Figure 4.1 for brain tumors imaged with Magnetic Resonance Imaging (MRI). In the following, the focus is on related work regarding MRI synthesis of healthy and tumor-bearing brains.

4.1.2.1 MRI synthesis of healthy brains

Original attempts at generating synthetic MRI of healthy brains relied on a numerical simulation of MR acquisition physics. Given MR scanner parameters such

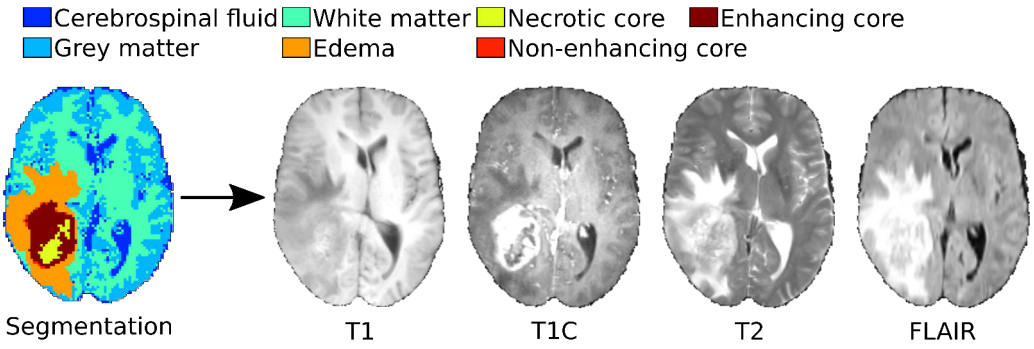

Figure 4.1: The generative model aims at synthesizing subject-specific MRI, provided a label map of a pathological brain.

as echo time and relaxation time, a discrete-event simulation model [Kwan 1999, Benoit-Cattin 2005, Glatard 2013] describes the dynamics of magnetization vectors, at each spatial position, according to Bloch equations [Bloch 1946]. Such numerical simulations are computationally expensive, except for specific cases for which there exists a closed-form solution to Bloch equations, such as for spin-echo or gradient-echo MR sequences [Kwan 1999, Maitra 2010, Iglesias 2010].

To describe the imaged object, two strategies are possible. The first strategy consists in describing the geometry of the imaged tissues with tissue-specific templates, and providing the tissue-specific biophysical properties (spin density and relaxation times). This allows the synthesis of MRI acquired on virtual geometrical templates. However, it requires that tissue-specific biophysical properties are precisely referenced in the literature [Kwan 1999], which is not the case for glioblastoma compartments. The second strategy does not rely on the definition of tissues: biophysical properties are specified voxel-wise, after an estimation from several MR scans obtained in a short time-frame with a very strict acquisition protocol (quantitative MRI or relaxometry), or by optimization methods [Maitra 2010, Iglesias 2010]. This strategy does not allow to generate synthetic images on virtual patients. Moreover, it cannot be applied to tumor scans when a proton density map is not available.

Recently, an iterative patch-based heuristic [Ye 2013] was used for Modality Propagation, i.e. the synthesis of a realistic subject-specific scan of one modality given a scan acquired with another modality. The core of the algorithm consists in modelling the covariation of local intensity patterns across modalities, in a fashion inspired by multi-atlas patch-based segmentation of healthy brain structures [Rousseau 2011, Coupé 2011]. Successful applications of Modality Propagation or related methods include synthesis of a patient-specific attenuation map for hybrid MR-PET [Burgos 2014], more accurate registration [Iglesias 2013], super-resolution [Roy 2013], and outlier detection [Ye 2013, Roy 2013, Cardoso 2015]. However, Modality Propagation shows two limitations for the synthesis of MRI for

virtual pathological cases: *i)* the local search window framework is adopted, which restricts the method to the image synthesis of healthy cases, *ii)* the input is a real image, which is assumed not to be available for virtual cases.

4.1.2.2 MRI synthesis of tumor-bearing brains

Previous works regarding MRI synthesis of tumor-bearing brains [Rexilius 2004, Prastawa 2005, Prastawa 2009] build upon tissue-specific templates, as in the framework introduced by Brainweb [Kwan 1999]. Typically, a tumor seed is artificially positioned in a healthy brain atlas, and then a tumor growth model simulates the tumor extension over time and its mass effect, i.e. the displacement of neighboring healthy tissues, which results in templates of healthy tissues and tumor compartments. However, in contrast with Brainweb, tissue-specific biophysical properties are not considered in favor of direct specifications of tissue-specific average MR intensities [Rexilius 2004, Prastawa 2005]. In [Prastawa 2009], textures are learnt for each healthy tissue and tumor compartment, which results in a set of 3D texture images. The only difference between different synthetic MRI lies in the tissue-specific templates: tissue-specific average MR intensities and 3D texture images are therefore always identical.

This approach is applicable to a wide variety of cases, but also bears some limitations: *i)* the complex distribution of image intensities for tumor compartments is summarized by its expectation, which is oblivious of multi-modal intensity distributions, *ii)* the inter-patient MR normalization procedure is not specified, which makes it difficult to standardize real MRI so that they look like synthetic MRI, typically for the training of machine learning algorithms [Geremia 2013], *iii)* simulated images do not show the variability of intensity of realistic MR scans, and the addition of a very high Gaussian noise only limits this effect.

The simulator of synthetic pathological MRI[1] described in [Prastawa 2009] has been used in a number of research articles mostly for prototyping and validation, in the context of glioma segmentation [Hamamci 2012, Geremia 2013, Menze 2014], outlier detection algorithm [Galimzianova 2015], registration of a healthy brain atlas to a tumor-bearing patient image [Bauer 2012], and construction of a brain atlas [Liu 2015]. Other applications include the training of machine-learning algorithms for glioma segmentation. [Geremia 2013] trained random forest on a large dataset consisting of 740 synthetic cases and showed good segmentation results on a few real cases for testing. [Galimzianova 2015] performed a thorough validation of an outlier detection algorithm, based on 100 MR scans synthesized with different tumor volumes to test the robustness of the algorithm to the amount of outliers. Image synthesis could also be useful in the context of tumor growth modeling: in [Gu 2012], PET images are synthesized for tumor-bearing brains, so as to allow clinically relevant interpretations of tumor growth model outputs. Although the applications of MRI synthesis of pathological cases are numerous, the competitors

[1] Freely available online at http://www.nitrc.org/projects/tumorsim

in the BraTS benchmark challenge tend to completely ignore the provided synthetic cases [Menze 2014], and we believe this is due to their lack of realism.

In this chapter, we address the image synthesis of pathological cases by using an iterative multi-atlas patch-based algorithm, inspired by recent successful algorithms in MRI synthesis of healthy brains. The expression Extended Modality Propagation (EMP) is coined for two reasons: *i*) to refer to the extension of Modality Propagation for the synthesis of images of pathological cases, *ii*) in contrast to Modality Propagation, EMP can deal with label maps in addition to image intensity. The synthesis process only requires a single label map, or the output of any tumor growth model in terms of the usual tumor compartments. Realistic synthetic images are obtained in the specific application of our algorithm to MRI synthesis of tumor-bearing brains. Our contributions include a novel generative model, an heuristic iterative algorithm to solve for the posterior distribution of multi-channel MR intensities, the estimation of image synthesis uncertainty, an analysis of Modality Propagation iterative feature augmentation, and the public availability of a large dataset of annotated and realistic MRI exhibiting gliomas. In the following, the generative model is described and solved for the maximum a posteriori and for uncertainty estimation (Section 4.2). Then synthesized MRI are validated with real MRI from the training dataset of BraTS benchmark challenge, compared to the state-of-the-art method in glioma image synthesis, and illustrated using the output of two different tumor growth models (Section 4.3). Finally, results and perspectives are discussed (Sections 4.4 and 4.5).

4.2 Extended Modality Propagation

In the following, a label map describing the anatomy of a tumor-bearing brain is assumed to be known. The goal of Extended Modality Propagation (EMP) is to synthesize medical images corresponding to the same anatomy. To achieve this goal, EMP relies on a set of training cases $\{L_n, I_n\}_n$ where L_n is a label map and I_n denotes multi-channel MRI. In the case of label maps, label values could be represented either as log-Odds [Sabuncu 2010], or as a vector of probabilistic labels of length $\mathcal{L} = 7$ (cerebrospinal fluid, grey matter, white matter, necrotic core, edema, non-enhancing tumor core, and enhancing core). Conventionally, a patch $S(I, x)$ is a cube of edge length $2r+1$, centered at spatial position $x \in \mathbb{R}^3$, consisting of image intensity values taken from an image I. The distance $d(S(I, x),\ S(J, y))$ between 3D patches of identical edge length is the canonical L_2-norm on $\mathbb{R}^{(2r+1)^3}$.

The EMP probabilistic generative model is first introduced to describe the synthesis of images. Second, an approximation of the marginalization step is introduced to estimate the maximum a posteriori and the image synthesis model uncertainty. Finally, more consistent patch matching is obtained by enhancing the patch features: *i*) Log-Odds are used to represent label maps, *ii*) multi-scale patches allow to consider larger patches in a computationally efficient manner, and *iii*) a patch similarity invariant to cube isometries allows to consider larger patches without

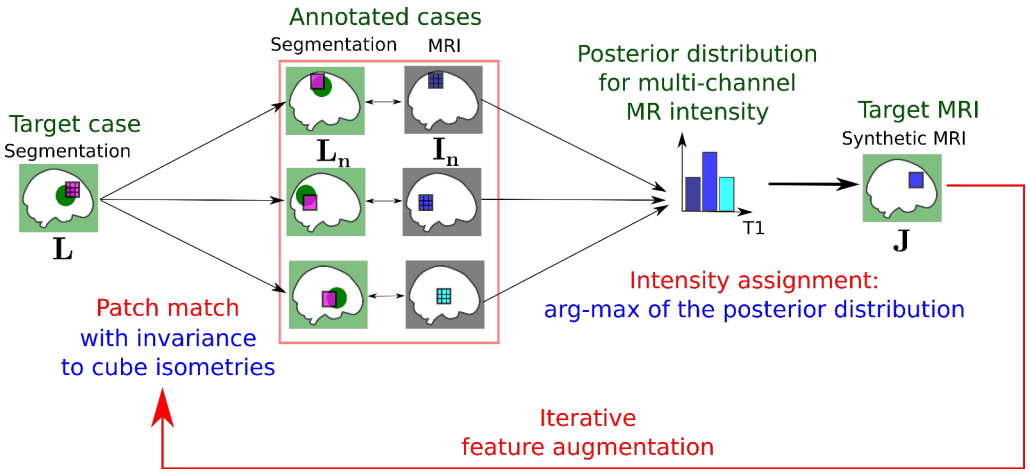

Figure 4.2: Pipeline for Extended Modality Propagation (EMP). Green disks mark known tumor locations. The fundamental assumption of EMP is that central voxels of similar label patches should have similar multi-channel image intensity. During the first iteration, multi-channel intensity j is assigned to a test patch $S(L, x)$ based on comparisons with a database of label patches $\{S(L_n, y)\}_{(n,\ y)}$, where y indexes every spatial position in the reference space Ω. During subsequent iterations, multi-channel intensity is updated based on comparisons between augmented features, which are the concatenation of label patches and multi-channel image patches synthesized during the previous iteration (feedback loop).

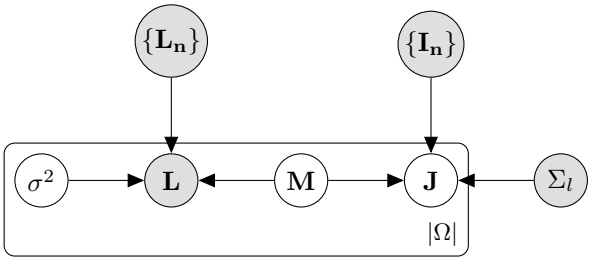

Figure 4.3: Graphical model for Extended Modality Propagation. A membership variable $\mathbf{M} : \Omega \rightarrow \{1, \ \ldots, \ N\} \times \Omega$ is sampled at every position x in the MNI space Ω to encode the training case $\mathbf{n}$ to sample from, and the spatial offset y. The observed probabilistic label $\mathbf{L}(x)$ results from the sampling of a patch $S(\mathbf{L}, x)$ from a Gaussian distribution with $S(\mathbf{L_n}, y)$ for mean and isotropic scaling $\sigma^2(x)$ for variance. The intensity $\mathbf{J}$ at position x is sampled from a Gaussian distribution with the central value $\mathbf{I_n}(y)$ for mean and covariance matrix $\Sigma_{\mathbf{L}}(x)$.

being impacted by a smaller sample size, and *iv*) as for Modality Propagation, an iterative patch match procedure results in smoother and more consistent synthetic images. The complete EMP pipeline is detailed in Figure 4.2.

4.2.1 Definition of a probabilistic generative model

The proposed generative model for Extended Modality Propagation builds upon the Bayesian modeling of local weighted voting for multi-atlas segmentation [Sabuncu 2010], which was then adapted for patch-based segmentation [Bai 2013, Iglesias 2015]. The proposed graphical model appears in Figure 4.3. This model is a mirrored version of the graphical model described in Chapter 2 for multi-atlas patch-based segmentation of pathological cases.

The target segmentation L is the result of sampling patches from atlases $(\mathbf{L_n}, \ \mathbf{I_n})$ at different positions y in the reference space Ω. The membership index $\mathbf{M}$ encodes, for every position $x \in \Omega$, both the atlas index n and the position y to sample from: $\mathbf{M}(x) = (n, \ y)$. If $\mathbf{M}$ were known, the target image $\mathbf{J}$ would be given by labels $\mathbf{I_n}(y)$ at central voxels of sampled patches. However, since $\mathbf{M}$ is not observed, the inference has to be performed by marginalizing over $\mathbf{M}$ [Sabuncu 2010]. The exact marginalization consists in a weighted vote involving patches at every position y in every atlas n, with weights proportional to both the probability that the target label patch $S(\mathbf{L}, x)$ is sampled from $S(\mathbf{L_n}, y)$, controlled by $\sigma^2(x)$, and the intensity likelihood of the target image patch, controlled by $\Sigma_{\mathbf{L}(x)}$.

4.2.2 Estimation of the posterior distribution

Variance parameters $\{\Sigma_l\}_{l \in \{1, \ ..., \ \mathcal{L}\}}$ and $\sigma^2(x)$ are first estimated ; then the problem of estimating $\mathbf{J}$ is stated by marginalizing over $\mathbf{M}$; finally a strategy to determine the arg-max of the posterior distribution is proposed.

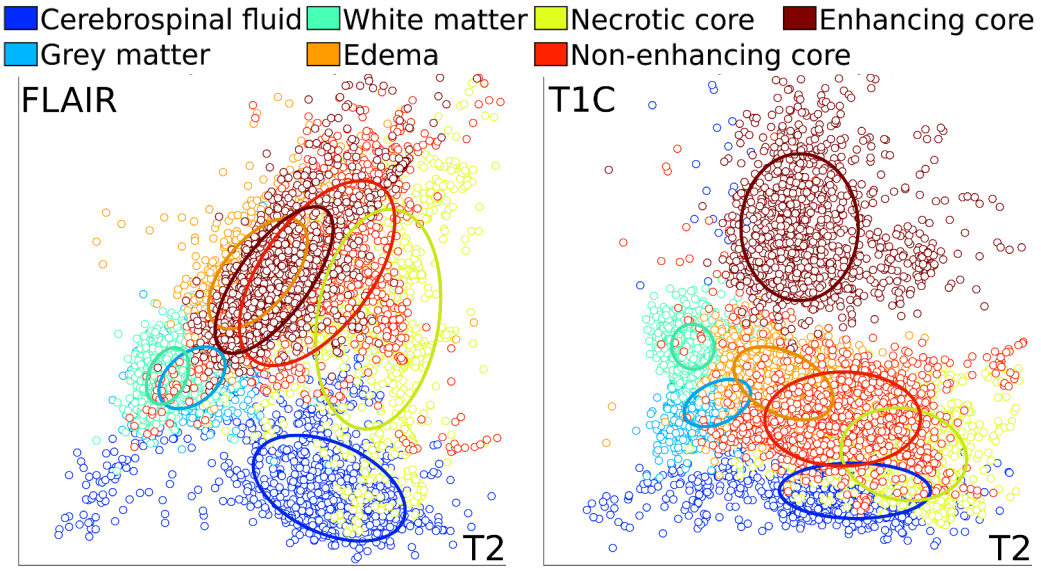

Figure 4.4: 2D projections of Student distributions fitted to average intensity of pure patches, with 60% quantiles overlayed. Appendix C presents other projections.

4.2.2.1 Estimation of the variance parameters

For each label $l \in \{1, \ldots, \mathcal{L}\}$, a multivariate Student distribution is fitted using Expectation-Maximization [Liu 1995] to average intensity of pure patches, i.e. for which all voxels belong to the same class, as described in Chapter 2. Robustness to outliers is achieved with Student distributions to model the heavy-tailed distributions of multi-channel image intensity. Figure 4.4 and Appendix C present 2D projections of obtained Student distributions. Appendix C also compares Student distributions obtained when fitting average intensity of pure patches, and when directly fitting voxel-wise image intensity $\{\mathbf{I_n}(y)\}_{(n,y)}$ labelled as l.

As with patch-based segmentation, the variance $\sigma^2(x)$ for label patch is spatially varying. In contrast to Chapter 2, $\sigma^2(x)$ is defined such that:

$$\sqrt{2}\sigma\,(x) = \min_{n,y} d\,(S(\mathbf{L}, x),\ S(\mathbf{L_n}, y))$$

4.2.2.2 Exact and approximate marginalizations

Assuming the membership index $\mathbf{M}(x)$ is independent and identically uniformly distributed $p\,(\mathbf{M}) = (N \times |\Omega|)^{-|\Omega|}$, the marginalization over $\mathbf{M}$ [Sabuncu 2010] is given voxel-wise by:

$$\widehat{\mathbf{J}}(x) = \underset{j}{\operatorname{argmax}} \sum_{n=1}^{N} \sum_{y \in \Omega} p_n\left(S(\mathbf{L}, x) \mid \mathbf{L_n},\ y,\ \sigma^2(x)\right) \times p_n\left(j \mid \mathbf{I_n},\ y,\ \Sigma_{\mathbf{L_n}(y)}\right)$$

where: *i)* $p_n\left(S(\mathbf{L},x) \mid \mathbf{L_n},\ y,\ \sigma^2(x)\right)$, which comprises the similarity between label patches, is the probability that the target label patch $S(\mathbf{L},x)$ is sampled from a multivariate Gaussian probability density function, with $S(\mathbf{L_n},y)$ for mean and isotropic scaling $\sigma^2(x)$ for variance:

$$p_n\left(S(\mathbf{L},x) \mid \mathbf{L_n},\ y,\ \sigma^2(x)\right) \propto \exp\left(-\frac{1}{2\sigma^2\left(x\right)} d^2\left(S(\mathbf{L},x),\ S(\mathbf{L_n},y)\right)\right)$$

and *ii)* $p_n\left(j \mid \mathbf{I_n},\ y,\ \Sigma_l\right)$ is the probability that the target intensity j is sampled from a multivariate Gaussian probability density function, with $\mathbf{I_n}(y)$ for mean and $\Sigma_{\mathbf{L_n}(y)}$ for covariance matrix:

$$p_n\left(j \mid \mathbf{I_n},\ y,\ \Sigma_{\mathbf{L_n}(y)}\right) \propto \exp\left(-\left(j-\mathbf{I_n}(y)\right)^T \Sigma^{-1}_{\mathbf{L_n}(y)}\left(j-\mathbf{I_n}(y)\right)\right)$$

or in the case of a multivariate Student probability density function with $\nu = \nu_{\mathbf{L_n}(y)}$ degrees of freedom:

$$p_n\left(j \mid \mathbf{I_n},\ y,\ \Sigma_{\mathbf{L_n}(y)}\right) \propto \left(1+\frac{1}{\nu}\left(j-\mathbf{I_n}(y)\right)^T \Sigma^{-1}_{\mathbf{L_n}(y)}\left(j-\mathbf{I_n}(y)\right)\right)^{-\frac{\nu+4}{2}}$$

The exact marginalization over $\mathbf{M}$ would require a high number of comparisons between patches, which is infeasible in practice. However, the marginalization step could be approximated, in a similar fashion as in Chapter 2, or as conventional patch-based methods do with K-nearest-neighbor patch match or local search window. Let $y_n(l)$ be the position in reference space of the closest patch of central label l found in the atlas $(\mathbf{L_n},\ \mathbf{I_n})$. For image synthesis of pathological cases, we propose to marginalize over only N patches, where N is the number of atlases: given a target label patch of central label $\mathbf{L}(x)$, the approximate marginalization only includes patches found at the spatial position $y_n(\mathbf{L}(x))$ in atlas n, which results in:

$$\widehat{\mathbf{J}}(x) = \underset{j}{\operatorname{argmax}} \sum_{n=1}^{N} p_n\left(S(\mathbf{L},x) \mid \mathbf{L_n},\ y_n(\mathbf{L}(x)),\ \sigma^2(x)\right) \times p_n\left(j \mid \mathbf{I_n},\ y_n(\mathbf{L}(x)),\ \Sigma_{\mathbf{L}(x)}\right)$$

4.2.2.3 Arg-max of the posterior distribution

With the approximate marginalization, our original belief was that $\widehat{\mathbf{J}}(x)$ could be efficiently computed thanks to Banach Fixed Point Theorem. Indeed, a necessary condition for j^* to be an optimal solution is given by gradient cancelling. Let $F:\mathbb{R}^4 \mapsto \mathbb{R}$ given by:

$$F(j) = \sum_{n=1}^{N} w_n \times p_n\left(j \mid \mathbf{I_n},\ y_n(\mathbf{L}(x)),\ \Sigma_{\mathbf{L}(x)}\right)$$

where:

$$w_n = p_n\left(S(\mathbf{L}, x) \mid \mathbf{L_n},\ y_n(\mathbf{L}(x)),\ \sigma^2(x)\right) > 0$$

Then the gradient of f with respect to $j \in \mathbb{R}^4$ is given by:

$$\vec{\nabla} F(j) = \sum_{n=1}^{N} w_n \ \times p_n\left(j \mid \mathbf{I_n},\ y_n(\mathbf{L}(x)),\ \Sigma_{\mathbf{L}(x)}\right) \times \Sigma^{-1}_{\mathbf{L}(x)}\left(\mathbf{I_n}\left(y_n(\mathbf{L}(x))\right) - j\right)$$

Gradient cancelling gives the necessary condition that j^* is a fixed point of a function T:

$$j^* = T(j^*)$$

where the function $T : \mathbb{R}^4 \mapsto \mathbb{R}^4$ is given by:

$$T(j) = \sum_{n=1}^{N} \lambda_n(j) \times \mu_n$$

where:

$$\mu_n = \mathbf{I_n}\left(y_n(\mathbf{L}(x))\right)$$

$$\lambda_n(j) = \frac{w_n \times p_n\left(j \mid \mathbf{I_n},\ y_n(\mathbf{L}(x)),\ \Sigma_{\mathbf{L}(x)}\right)}{\sum_{n=1}^{N} w_n \times p_n\left(j \mid \mathbf{I_n},\ y_n(\mathbf{L}(x)),\ \Sigma_{\mathbf{L}(x)}\right)}$$

$$\forall n,\ \lambda_n(j) \geq 0$$

$$\sum_{n=1}^{N} \lambda_n(j) = 1$$

With Gaussian density functions for $p_n\left(j \mid \mathbf{I_n},\ y,\ \Sigma_{\mathbf{L}(x)}\right)$, it can be shown that the Lipschitz constant of T is less than or equal to:

$$\max_{n=1,\ldots,N} ||\mu_n|| \max_{k=1,\ldots,N} ||\Sigma^{-1}_{\mathbf{L}(x)}\left(\mu_n - \mu_k\right)||$$

By applying Banach Fixed Point Theorem, we are certain that, for voxels x for which T is a contraction mapping, j^* is unique, and that for any initialization $j_0 \in \mathbb{R}^4$, any sequence $(j_k)_{k\in\mathbb{N}}$ such that $j_{k+1} = T\left(j_k\right)$ converges to the fixed point j^* with a geometric rate. An implementation of the EMP algorithm is given in

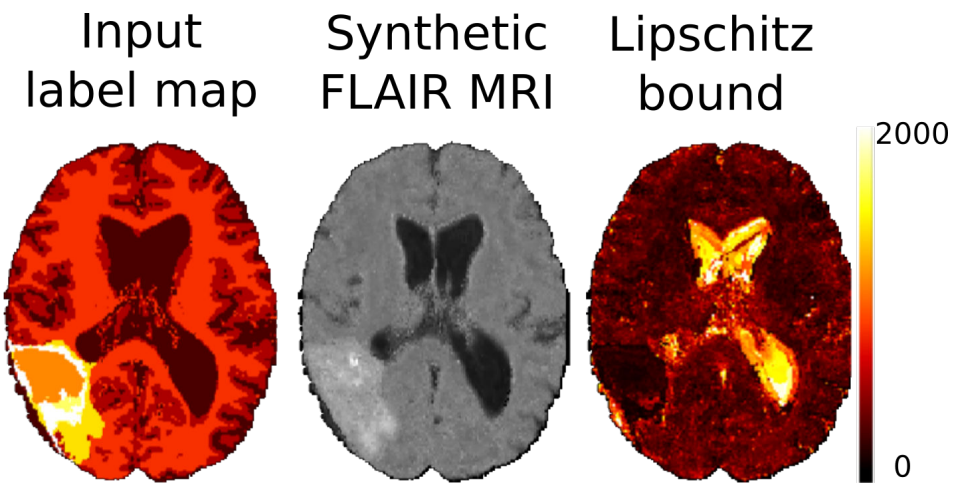

Figure 4.5: Visual representation of the bound which we reported for the Lipschitz constant of T. For most voxels, we could not prove that T is a contraction mapping. The synthetic FLAIR image (middle) appears noisy because it was obtained after a single iteration (no feedback loop).

Algorithm 2. Unfortunately, in our experiments, for most voxels x, the bound of the Lipschitz constant of T, which is reported above, is not strictly less than 1, as illustrated in Figure 4.5. Therefore, we could not prove that T is a contraction mapping. However, for all our test cases, the algorithm still does converge in roughly 40 iterations for 99.9% of the voxels of the brain mask.

4.2.2.4 Estimation of the image synthesis uncertainty

With the Fixed Point procedure, the estimation of the arg-max of the posterior image intensity distribution involves the iterative estimation of j^* as a weighted average of intensities μ_n found in the atlases. Therefore, the uncertainty of the image synthesis process can be similarly estimated as the weighted sample variance $\widehat{\sigma}^2$, thanks to the optimal weights λ_n^*. Uncertainty estimation allows to pin-point regions with unlikely image intensity based on the proposed generative model. This is illustrated in Figure 4.6. Uncertainty estimation allows to detect outliers in the images without having to identify outliers with the pathology as is the case in [Ye 2013, Cardoso 2015].

4.2.3 Consistent patch match

Patch match can be more consistent by relying on enhanced features and by adopting an iterative feature augmentation strategy.

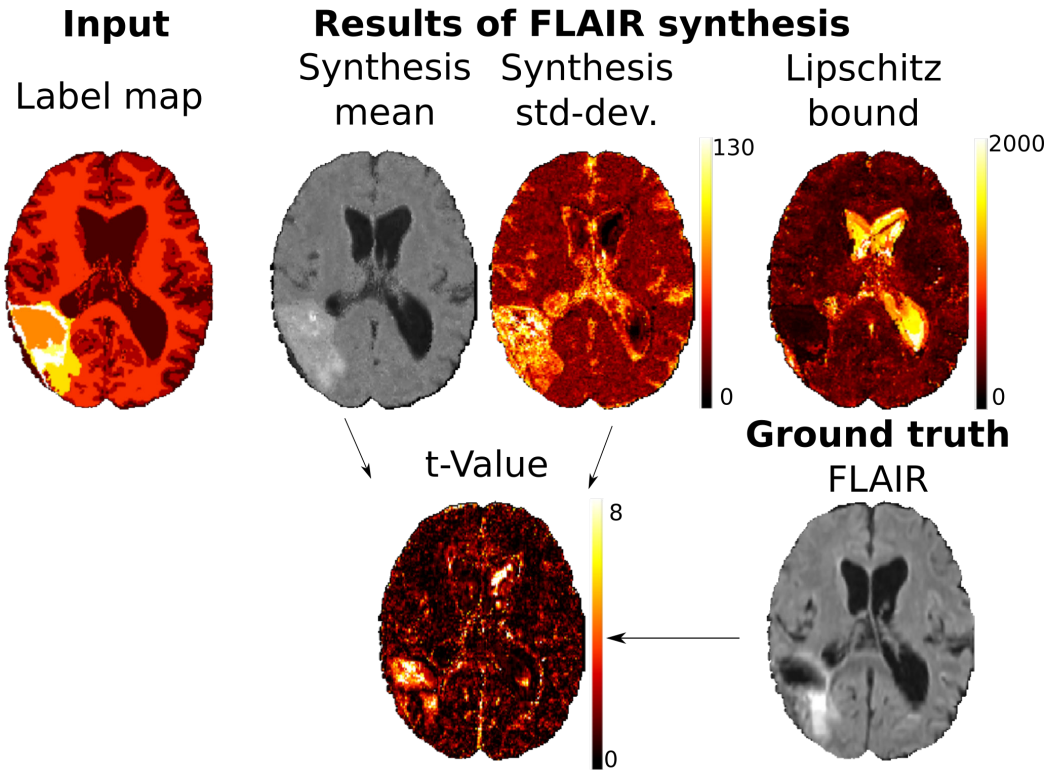

Figure 4.6: Illustration of image synthesis uncertainty. Along with the model prediction $\widehat{\mathbf{J}}(x)$ (synthesis mean), the knowledge of the model uncertainty $\widehat{\sigma}^2$ (synthesis variance) allows to compute the t-value $\sqrt{\widehat{n}}\left(\mathbf{J}(x) - \widehat{\mathbf{J}}(x)\right) \;/\; \widehat{\sigma}$, with the sample size $\widehat{n}$ being the number of atlases.

Algorithm 2: Extended Modality Propagation.

pre-select N training cases similar to $\mathbf{L}$
foreach voxel $x \in \Omega$ **do**
  evaluate the features $S(\mathbf{L}, x)$
  let $l = \mathbf{L}(x)$ be the central label in $S(\mathbf{L}, x)$
  foreach training case n **do**
    find the closest patch of central label l in $\mathbf{L_n}$
    let $y_n(l)$ be the spatial position of this patch
    evaluate the features $S(\mathbf{L_n}, y_n(l))$
    evaluate the intensity $\mathbf{I_n}(y_n(l))$
  set the variance $\sigma^2(x)$ based on the features
  set the intensity $\widehat{\mathbf{J}}(x)$ by applying a Fixed Point procedure as if T were a contraction mapping:

$$j_0 = \frac{\sum_{n=1}^{N} w_n \times \mu_n}{\sum_{n=1}^{N} w_n} \in \mathbb{R}^4$$
$$\forall k \in \mathbb{N},\ j_{k+1} = T(j_k)$$
$$\widehat{\mathbf{J}}(x) = \lim_{k \to +\infty} j_k$$

4.2.3.1 Log-Odds

During the fist iteration of image synthesis, patch match only relies on a label map as input. It is crucial to choose the best representation for this single input. A label map could be described as a set of probabilistic tissue templates: for each voxel, a vector contains the probability of belonging to each tissue. However, this representation is identical for patches completely enclosed in a single tissue. A better strategy, as recommended in [Sabuncu 2010], consists in replacing binary labels with Log-Odds, so as to take into account the distance to borders between tissues and to be able to distinguish different patches among patches completely enclosed in a tissue. Log-Odds distances are defined as proportional to $\exp\left(\rho \widetilde{D}^l(L, x)\right)$, where $\rho > 0$ is the slope constant (in our experiments, $\rho = 1$ as in [Sabuncu 2010]), and $\widetilde{D}^l(L, x)$ is the signed distance transform of label l for label map L at spatial position x, which is positive inside structures labelled as l.

4.2.3.2 Multi-scale patches

As shown in Chapter 2, multi-scale patches are computationally efficient representations for large patches, which allows more effective patch match. The central part of the patch is described voxel-wise, while the peripheral part is described by average intensity values over neighboring patches, by analogy with the foveal vision. Similar ideas regarding foveation and non-uniform sampling are detailed in [Ciresan 2012]. In the following, we have chosen two scales: 6 mm (3x3x3 patch at the center) and 18 mm (the volume described by the multi-scale patch corresponds to a 9x9x9 conventional patch), with only twice the memory footprint of conventional

3x3x3 patches.

4.2.3.3 Patch match with invariance properties

Every atlas is affinely registered the MNI space [Mazziotta 2001] so that patch extraction is invariant to the pose of the subject. Moreover, the canonical L_2 distance used for patch matching is sensitive to rotation or symmetry of the patches. As shown in Chapter 2, the combination of multi-scale patches and invariance to cube isometries yields better segmentation results, and thus could be beneficial for patch-based image synthesis as well. The 48 cube isometries consist of certain rotations under which the cube is invariant, plus their composition with central symmetry. For healthy brain structures, rotated patches can be unrealistic, only sagittal plane symmetry is considered. For pathological patches, the similarity measure should be insensitive to rotation and symmetry: the number of atlases is small, and using multi-scale patches further decreases the sample size. Therefore, the addition of plausible configurations of pathological patches generated by applying all the cube isometries to observed pathological patches is relevant, and ultimately leads to an invariant patch distance.

4.2.3.4 Iterative feature augmentation

As with Modality Propagation [Ye 2013], an iterative procedure guarantees a more coherent patch match, which results in an overall more coherent image synthesis. As shown in 4.2, during the first iteration $t = 1$, multi-channel intensity is assigned to a test patch based on comparisons with a database of label patches. During subsequent iterations $t > 1$, multi-channel intensity is updated based on comparisons between augmented features, which are the concatenation of *i)* label patches, weighted with $1 - \alpha_t$, and *ii)* multi-channel image patches synthesized during the previous iteration $t - 1$, weighted with α_t. As in [Ye 2013], the feedback weight α_t increases with the number of iterations, starting with $\alpha_1 = 0$ for iteration $t = 1$. There are different possible definitions for weight α_t. Unless mentioned, we choose a maximal number of iterations $t_{\max} = 3$, and let $\alpha_t = (t-1)/t_{\max}$, so that the weight of label patches is never zero and decreases linearly with the number of iterations. The effect of iterations is illustrated in Figure 4.7.

4.3 Results

Our goal is to synthesize four MR channels commonly acquired for glioma assessment: pre-contrast T_1-weighted image (T1), contrast-enhanced T_1-weighted (T1C), T_2-weighted (T2), and T_2-FLAIR MR images. Fast approximate nearest-neighbour search are used for patch match, more precisely multiple randomized k-d trees for high dimensional data [Muja 2014].

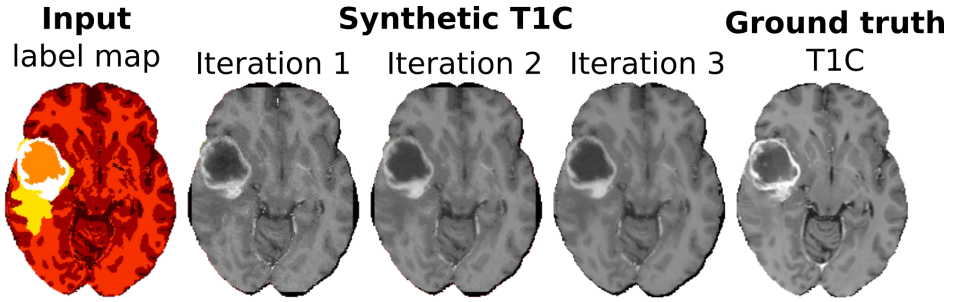

Figure 4.7: From left to right: input label map, synthetic contrast-enhanced T_1-weighted MRI after iterations $t = 1,\ 2,\ 3$, and ground truth (real MRI). The image parts at the border of the brain are better synthesize due to the iterative feedback of previously-synthesized images.

4.3.1 Pre-processing pipeline

Every image is affinely registered to an MNI atlas and warped to the same reference space. For experiments which require strictly more than 3 iterations, a resampling to 2 mm isotropic resolution is performed to decrease computation time, otherwise we keep the original 1 mm isotropic resolution. A resampling to 2 mm resolution was used by the top-performing glioma segmentation method at the 2012 BraTS challenge, which indicates that the generation of a synthetic dataset of 2 mm resolution images could already be of interest for applications other than medical image synthesis. Atlas images I_n are standardized with a global affine intensity transform, applied to each modality independently: non-zero intensity are clipped below 1% and above 99% quantiles, then mean and standard deviation are set to the values reached by averaging over training data, respectively 360 and 120. Except for the clipping, all the proposed pre-processing operations can be reverted, which eases the adaptation of such a synthetic dataset for the analysis of another real dataset.

4.3.2 Validation: an experiment using real cases

4.3.2.1 Dataset and Evaluation

The dataset consists of MRI of 20 high-grade (HG) and 10 low-grade (LG) cases, made publicly available as training dataset in the 2013 edition of the MICCAI BraTS benchmark [Menze 2014]. A leave-one-out procedure is always applied to exclude the target image from the training dataset. There is no atlas pre-selection and all the training cases, remaining after leave-one-out, are used as atlases for the synthesis of MRI of cases of the same grade. Only the segmentation of the tumor compartments is originally available [Menze 2014], so healthy tissues are automatically segmented by applying FSL FAST [Zhang 2001] to the T1C image, which consistently has the highest resolution among the different MR channels.

The validation of the experiment consists in comparing real MRI with synthetic MRI generated solely based on a segmentation of the brain and the tumor. Quan-

titative assessment include Peak Signal to Noise Ratio (PNSR) to quantify the quality of the synthesis, Mean-Absolute-Error (MAE) to quantify bias, Hellinger distances between tissue-specific intensity distributions to quantify how well the model captures the variability of real image intensity, mean Structural Similarity (SSIM) [Wang 2004] (with default values) as an additional image reconstruction criterion, and finally bias with respect to the mean image intensity, i.e. difference of mean image intensity for the real and the synthetic image, and the bias with respect to the variance of image intensity. All these measures are classically used for assessment of image reconstruction or synthesis [Manjón 2010, Konukoglu 2013, Cardoso 2015, Gómez 2015]. Whenever an average is computed, only the voxels which belong to a region-of-interest (ROI) are considered: the ROI is the whole brain mask as defined by the input label map (segmentation of healthy and pathological tissues).

4.3.2.2 Results of the first experiment

A visual inspection of image synthesis results can be performed in axial views for a high-grade glioma in Figure 4.8 and for a low-grade glioma in Figure 4.9. Axial, coronal and sagittal views are also presented in smaller displays in Figure 4.10 and in Appendix B. The model does not replicate artifacts or croppings present on some real cases, as shown in Figure 4.11. This is expected, since the image synthesis model is based on a label map, which was created from a high-resolution T1C image of a whole brain.

The influence of the feedback weight α_t is studied in Figure 4.12. The protocol is the following: different values are chosen for feedback weights, then Extended Modality Propagation is run for 10 iterations with $\alpha_1 = 0$ and subsequent α_t fixed at the chosen value. The highest PSNR (18.6) and highest mean-SSIM (0.63) are obtained for the smallest non-zero feedback weight ($\frac{1}{9}$). The variance of the PSNR is lower when a non-zero feedback weight is chosen, however the median PNSR is nearly the same for all feedback weight values. MAE increases when the weight increases. Other measures do not seem to show any remarkable phenomenon. In the end, either the range of chosen weight values is too narrow and biases our analysis, or any small non-zero feedback weight is reasonable for image synthesis.

The influence of the number of iterations of EMP is studied in Figure 4.13. Here again, PSNR variance decreases as soon as a non-zero feedback weight is introduced (second iteration). The highest PSNR (18.7) and highest mean SSIM (0.64) are obtained at the end of the second iteration. MAE increases notably with the number of iterations, although the bias with respect to mean image intensity and variance is relatively constant. In the end, a small number of iterations should be preferred, typically 2 or 3 iterations. The effect of the number of iterations is more marked here due to a high constant feedback weight fixed at $\frac{2}{3}$. It is reasonable to perform 2 to 3 iterations with a linearly increasing feedback weight α_t.

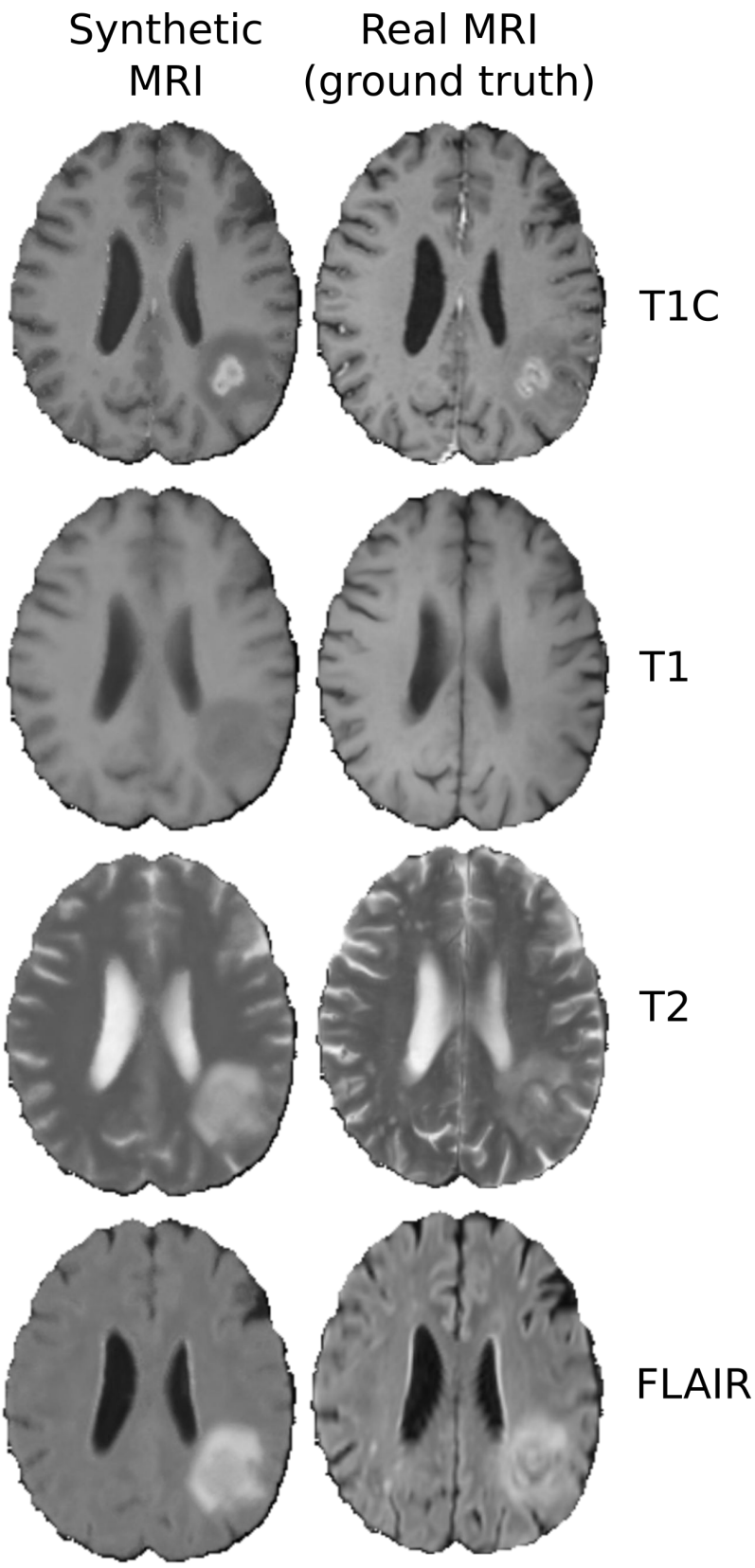

Figure 4.8: Qualitative evaluation of our synthesis method after 3 iterations on a high-grade case (HG-0010). Axial view.

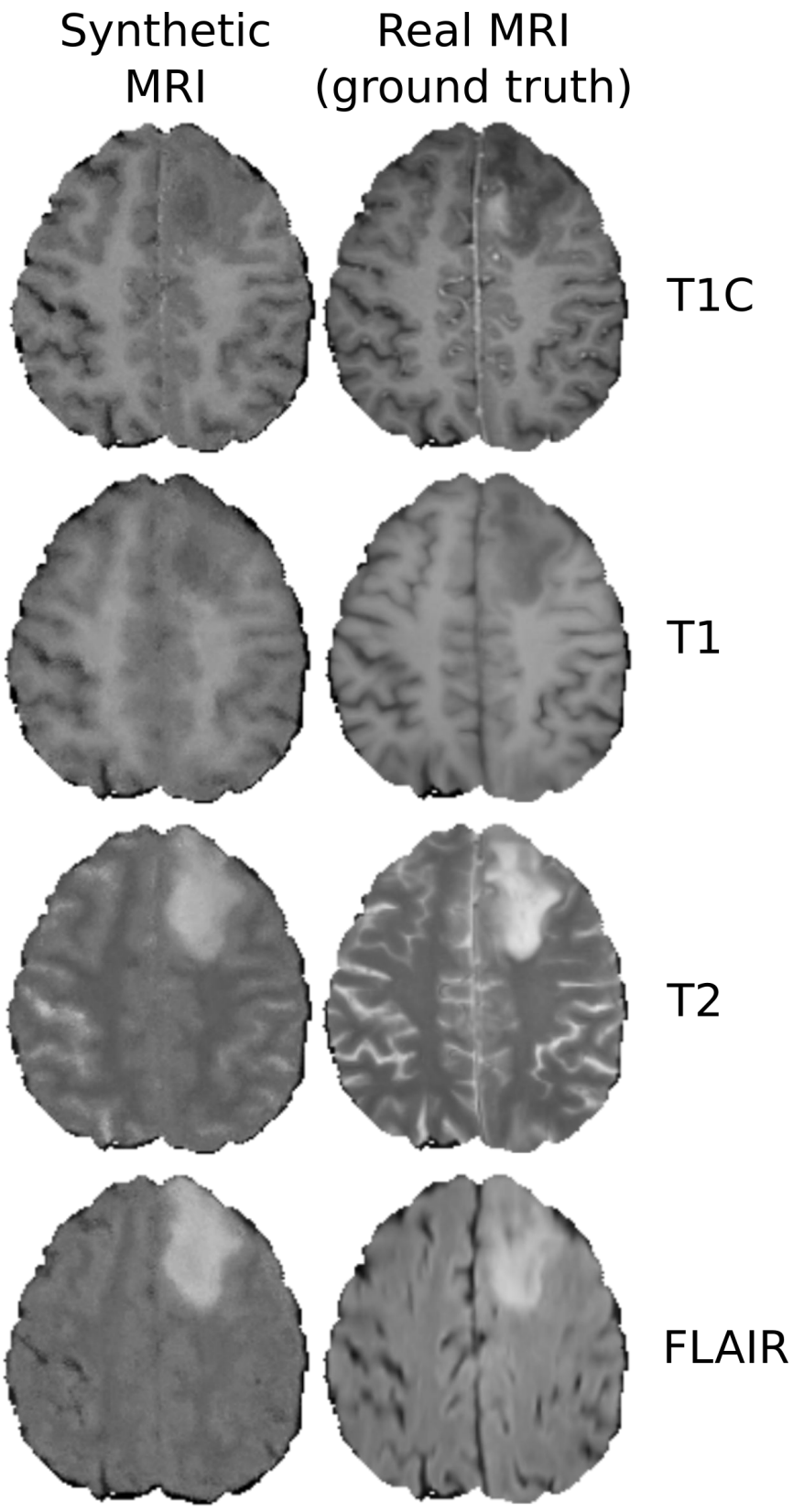

Figure 4.9: Qualitative evaluation of our synthesis method after 3 iterations on a low-grade case (LG-0006). Axial view.

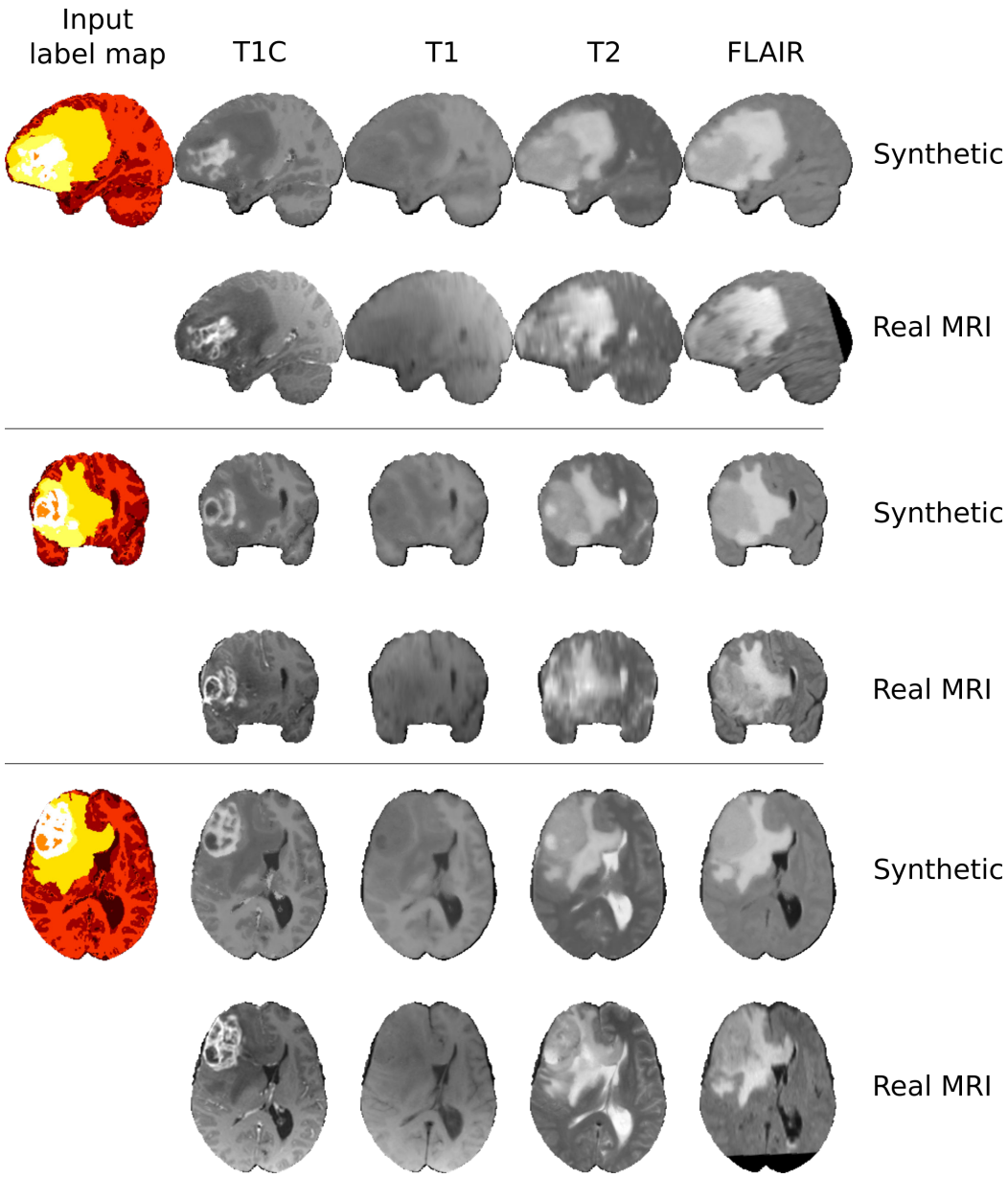

Figure 4.10: Qualitative evaluation of our synthesis method after 3 iterations on a high-grade case (HG-0009). Top to bottom: sagittal, coronal and axial views, for synthetic (ours) and real MRI on even and odd rows respectively.

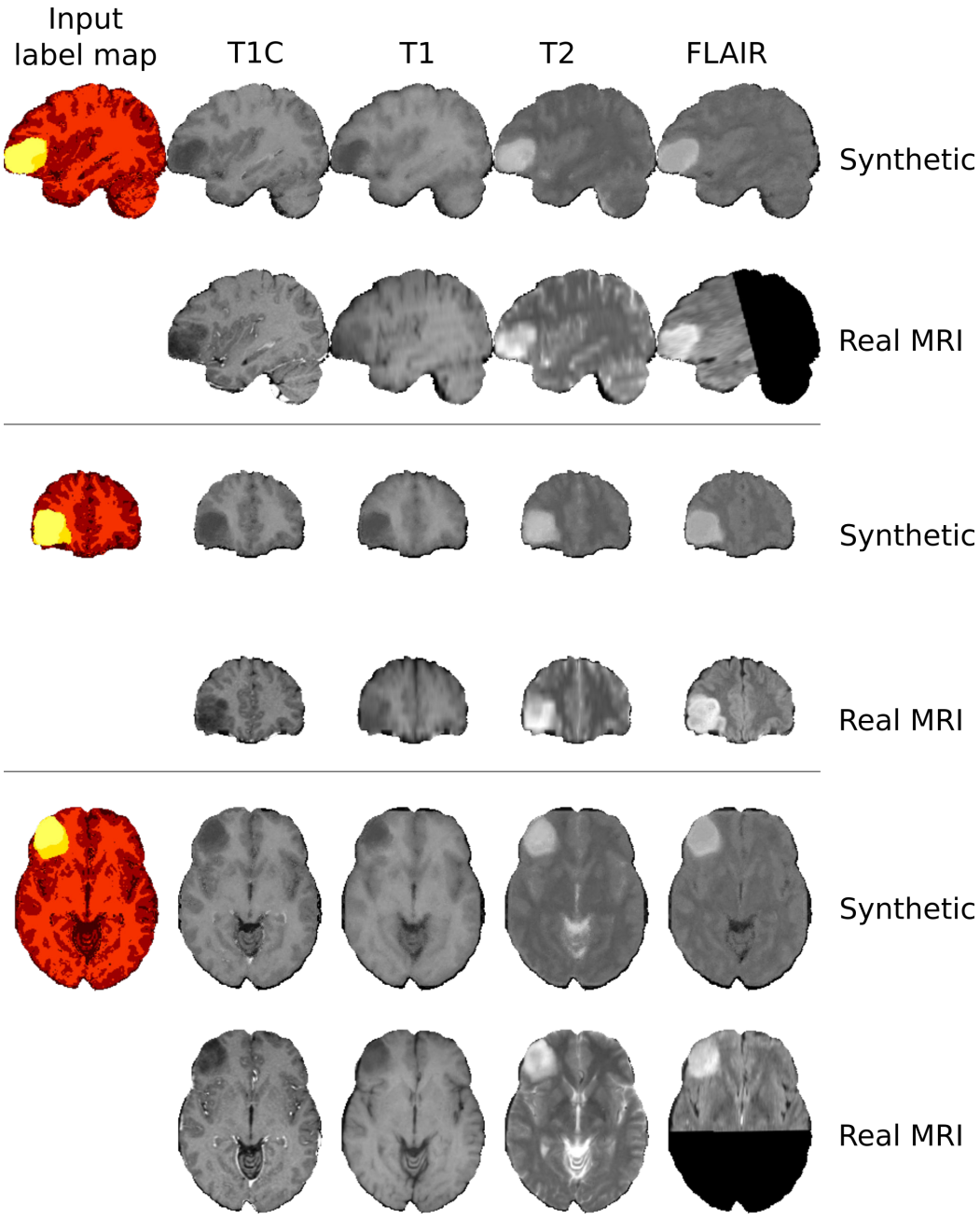

Figure 4.11: Focus on a low-grade case (LG-0008) with an artifact (sagittal view in T1, best viewed by zooming on a computer screen) and cropping (axial and sagittal views in FLAIR). Validation of our method after 3 iterations. Top to bottom: sagittal, coronal and axial views, for synthetic (ours) and real MRI on even and odd rows respectively.

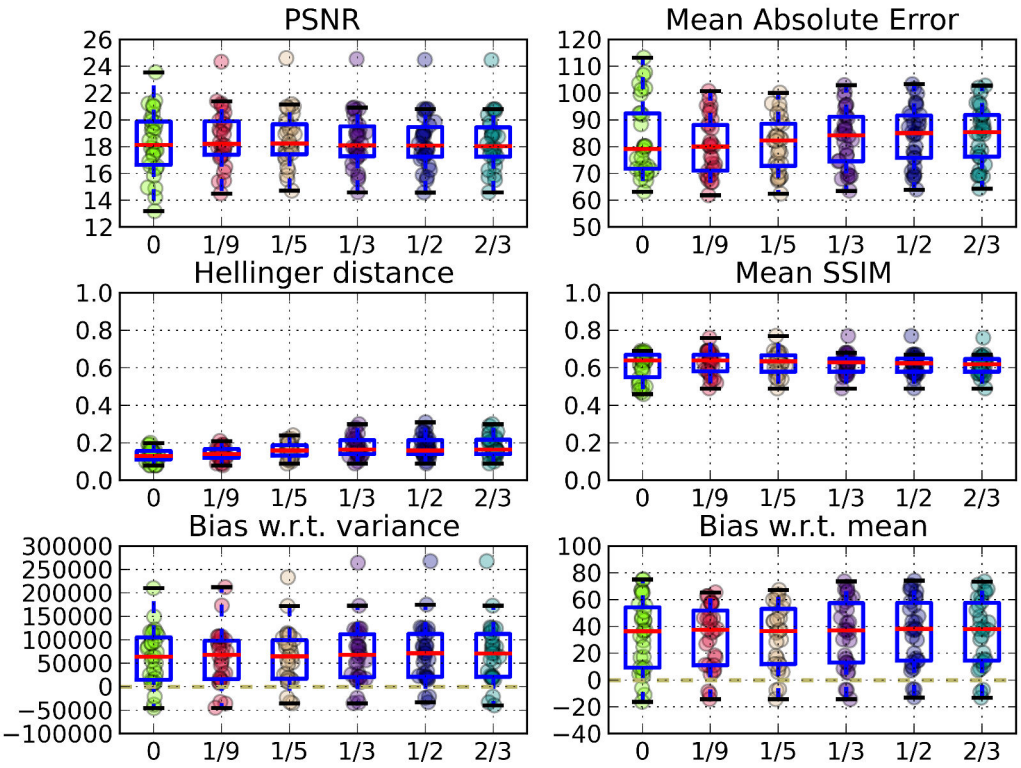

Figure 4.12: Influence of the feedback weight α_t on assessment measures. Measures are evaluated in leave-one-out on FLAIR images present in the 2013 MICCAI BraTS benchmark Evaluation dataset. The weight is fixed and appears in abscissa of the graphs. The number of iterations is 10 so that the influence of the weight is emphasized. PSNR variance decreases with stronger feedback.

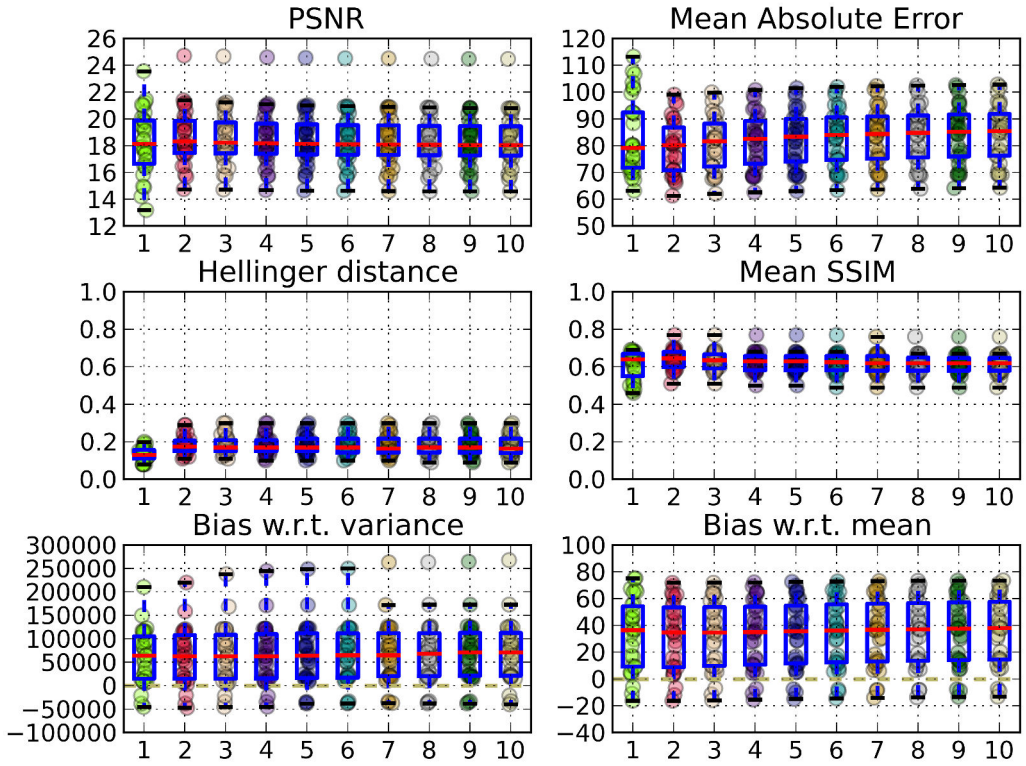

Figure 4.13: Influence of the number of iterations on assessment measures. Measures are evaluated in leave-one-out on FLAIR images present in the 2013 MICCAI BraTS benchmark Evaluation dataset. The number of iterations appears in abscissa of the graphs. The feedback weight is fixed at $\frac{2}{3}$. An optimum is quickly reached and then it could be detrimental to iterate further.

4.3.3 Comparison: an experiment using synthetic cases

4.3.3.1 Dataset and Evaluation

The dataset consists of the segmentations used in the previous experiment on real cases (Section 4.3.2.2). Our image synthesis model is compared to the image synthesis procedure given in [Prastawa 2009]. In contrast to the previous experiment, the feedback weight increases linearly with the number of iterations.

For a fair comparison, synthetic MRI generated with [Prastawa 2009] are standardized as described in Section 4.3.1, so that the average image intensity and the variance corresponds to the average and variance of the real image. We do not perform this procedure on images which are synthesized with our method, since the range of intensity of synthetic images is already comparable to the range of intensity of the standardized real image. If we were to standardize images synthesized with our method, the conclusion of the experiments would not change.

4.3.3.2 Results of the second experiment

For the sake of comparison, synthesized images based on [Prastawa 2009] are shown in Figure 4.14 and Figure 4.15, using the same input label maps as for our method in Figure 4.10 and Figure 4.11. Quantitative results are reported for each MR channel in Figure 4.16 (FLAIR), Figure 4.17 (T1C), Figure 4.18 (T1), and Figure 4.19 (T2). A summary of assessment measures is presented in Table 4.1. Since images synthesized with [Prastawa 2009] were standardized, the bias with respect to mean intensity and variance is always better for [Prastawa 2009]. However, PSNR, MAE, mean SSIM and Hellinger distances are always higher for the proposed image synthesis method. Moreover, with the linearly increasing feedback weight, most results are improved with additional iterations.

4.3.4 Illustration: synthetic MRI using 2 tumor growth models

In this section, we use two different tumor growth models to generate probabilistic label maps, based on which we create a hard label map. The proposed image synthesis model is then used to generate synthetic MRI of these virtual patients. In Figure 4.20, the first tumor growth model is a reaction-diffusion model with mass-effect, as proposed in [Prastawa 2009]. The second tumor growth model is a multi-population cell model driven by angiogenesis and vascularization, called the Proliferation Invasion Hypoxia Necrosis Angiogenesis (PIHNA) model. The PIHNA model is described in 1D in [Swanson 2011], and was implemented in 3D by a colleague of mine to simulate tumor growth in an atlas of a healthy brain. Synthetic MRI based on the PIHNA model are shown in Figure 4.21 and Figure 4.22, respectively after one and two iterations of the proposed image synthesis algorithm.

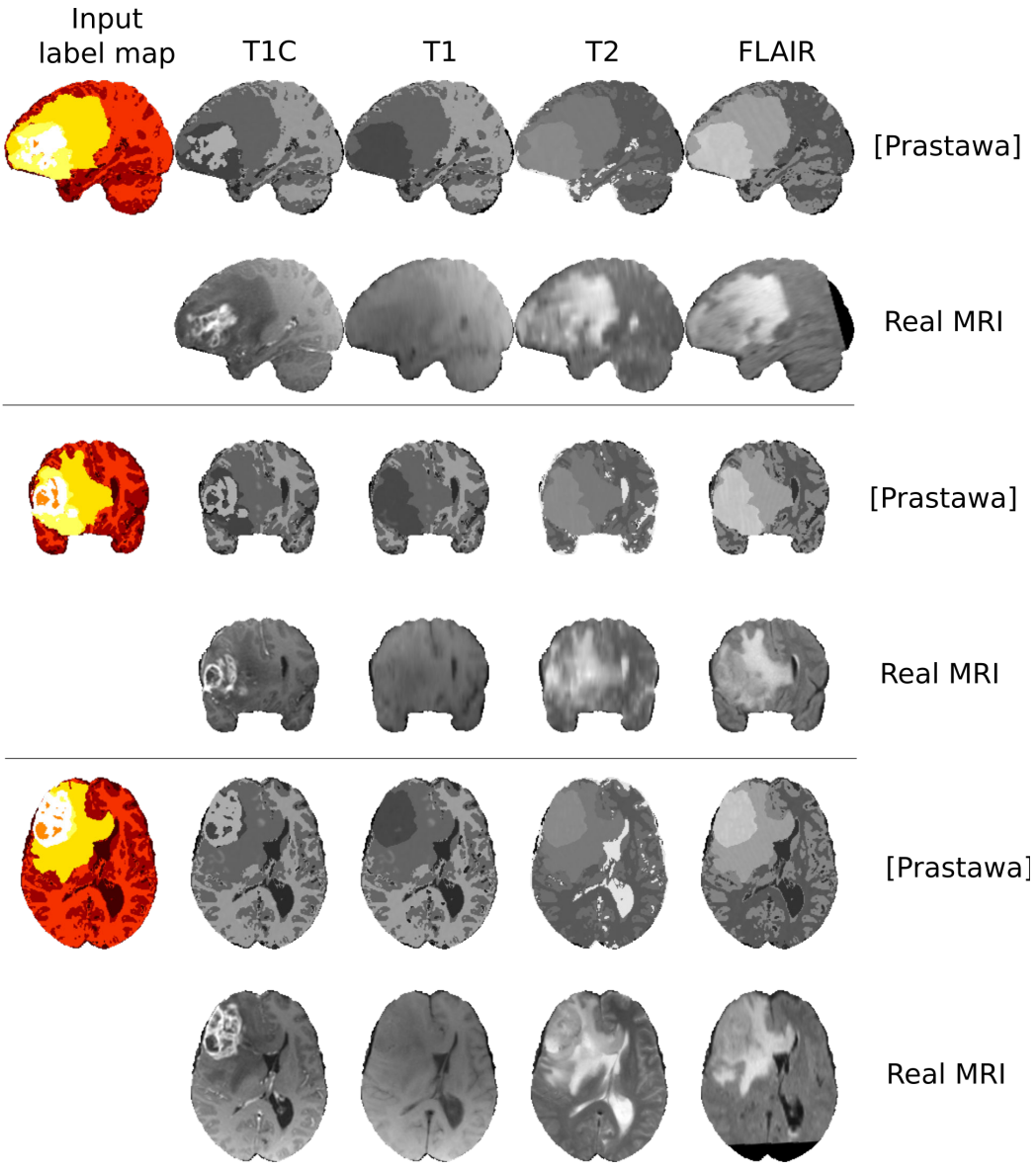

Figure 4.14: High-grade HG-0009. Top to bottom: sagittal, coronal and axial views, for Prastawa's synthetic and real MRI on even and odd rows respectively.

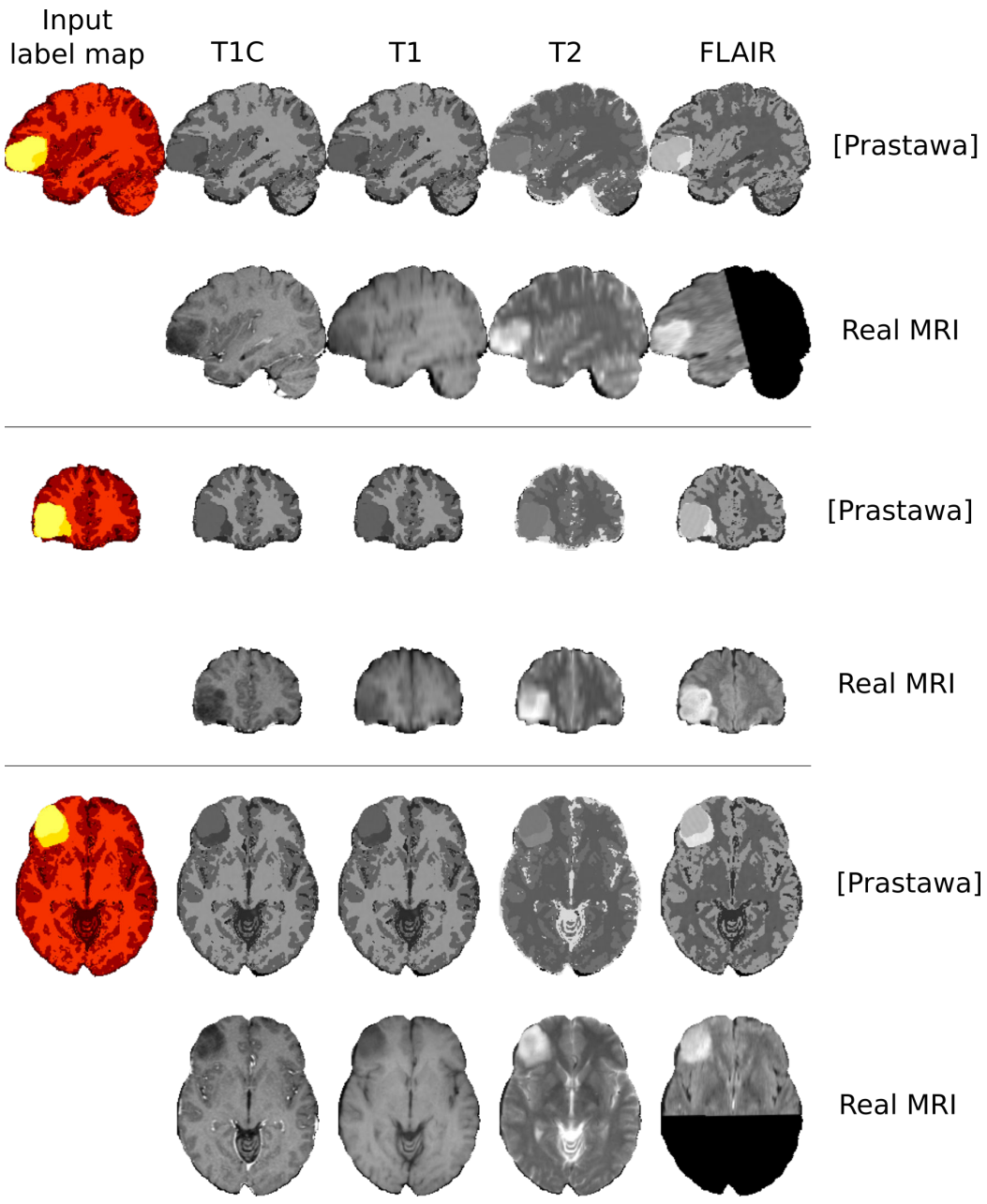

Figure 4.15: Low-grade LG-0008. Top to bottom: sagittal, coronal and axial views, for Prastawa's synthetic and real MRI on even and odd rows respectively.

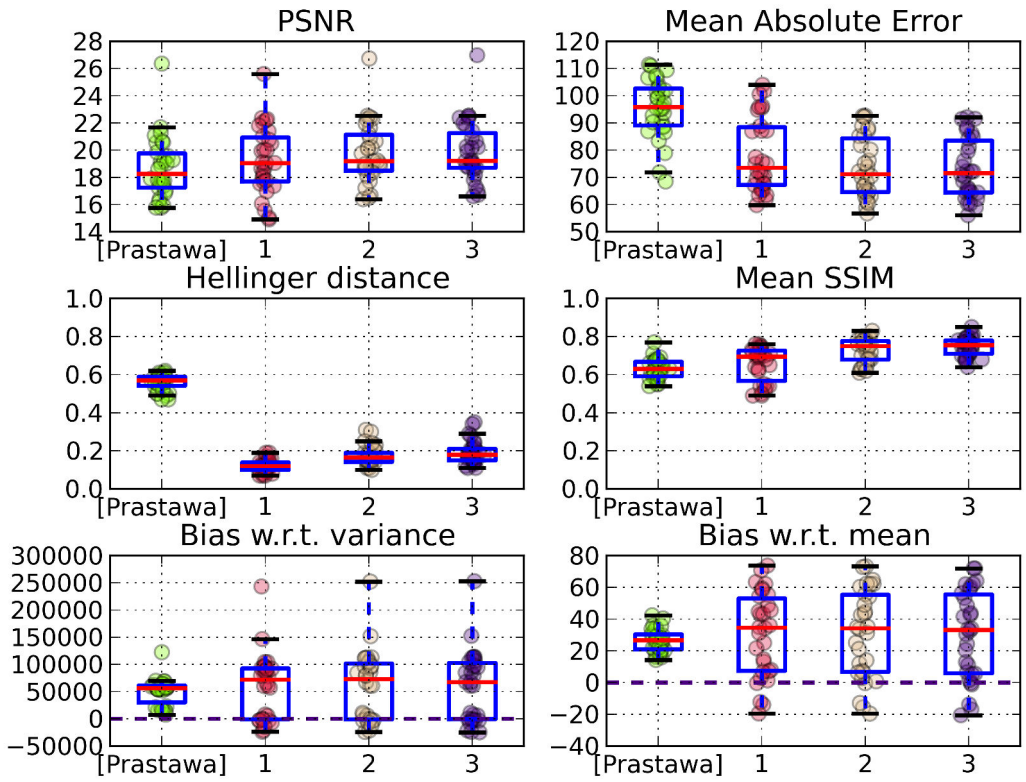

Figure 4.16: Comparison of synthetic FLAIR MRI to ground truth MRI: [Prastawa 2009] (left) and the three iterations of our method.

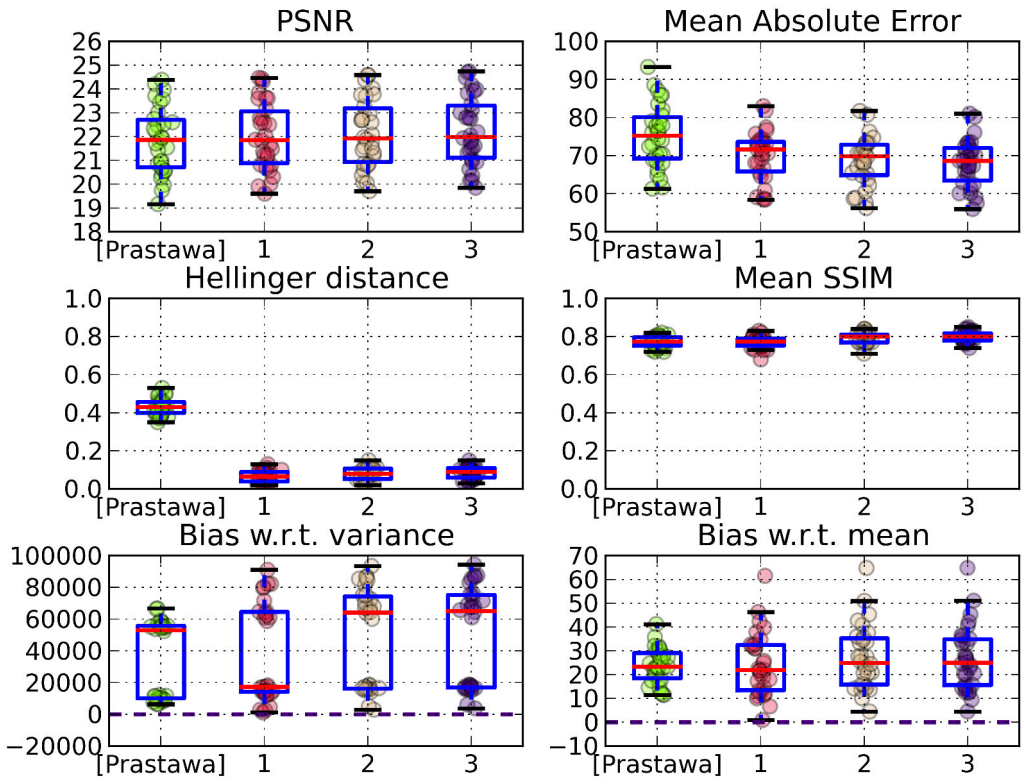

Figure 4.17: Comparison of synthetic T1C MRI to ground truth MRI: [Prastawa 2009] (left) and the three iterations of our method.

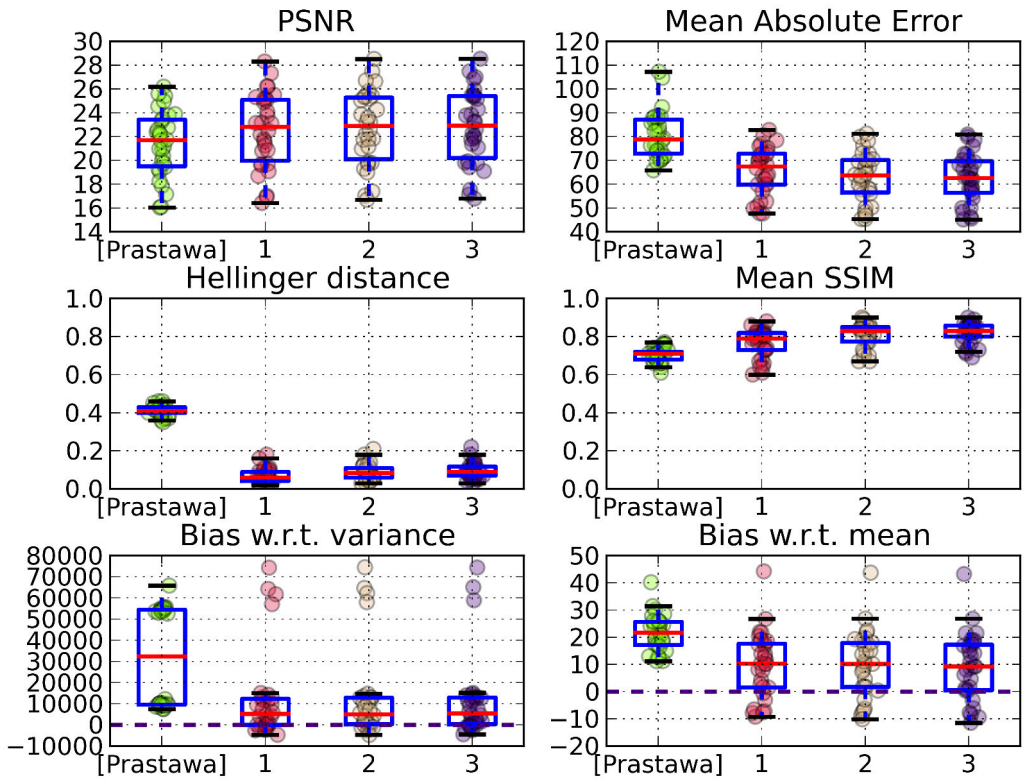

Figure 4.18: Comparison of synthetic T1 MRI to ground truth MRI: [Prastawa 2009] (left) and the three iterations of our method.

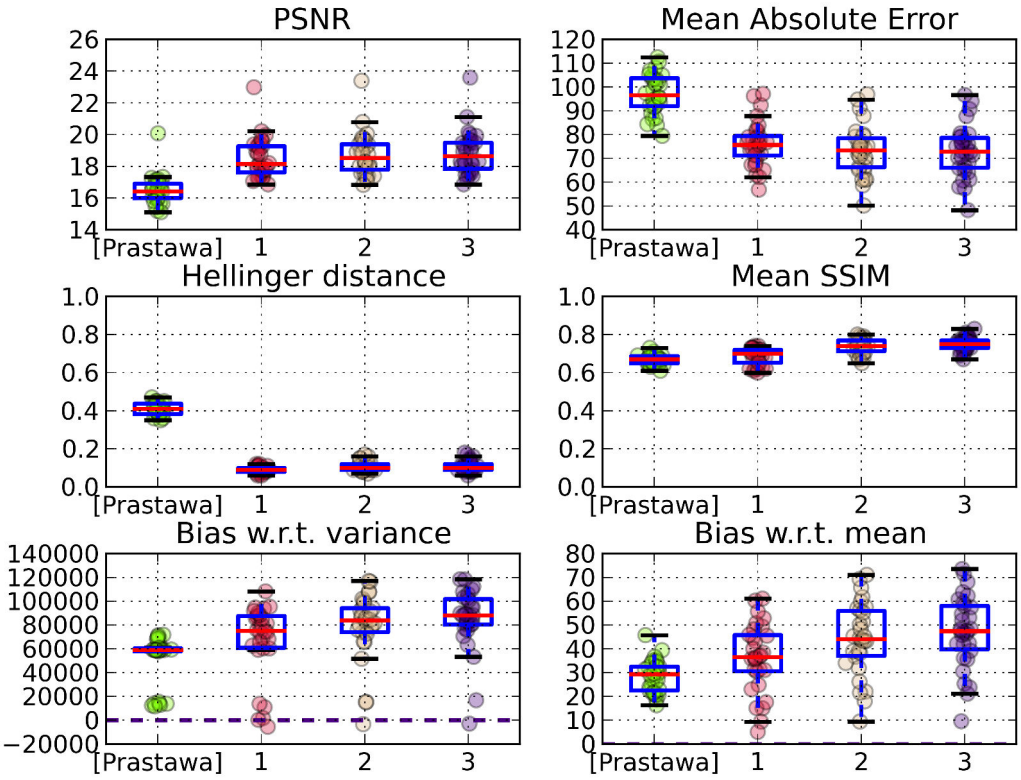

Figure 4.19: Comparison of synthetic T2 MRI to ground truth MRI: [Prastawa 2009] (left) and the three iterations of our method.

Table 4.1: Average assessment measures for image synthesis, based on the 2013 BRATS Evaluation dataset.

FLAIR	**PSNR**	**mean SSIM**	**MAE**	**Hellinger distance**
[Prastawa 2009]	18.7	0.63	95.4	0.56
Ours (iteration 1)	19.3	0.66	78.2	**0.12**
Ours (iteration 2)	19.8	0.73	73.6	0.17
Ours (iteration 3)	**19.9**	**0.75**	**73.1**	0.19
T1C	**PSNR**	**mean SSIM**	**MAE**	**Hellinger distance**
[Prastawa 2009]	21.9	0.77	75.2	0.43
Ours (iteration 1)	22.0	0.76	70.1	**0.07**
Ours (iteration 2)	22.1	0.81	68.6	0.08
Ours (iteration 3)	**22.2**	**0.82**	**68.0**	0.08
T1	**PSNR**	**mean SSIM**	**MAE**	**Hellinger distance**
[Prastawa 2009]	21.4	0.70	80.7	0.41
Ours (iteration 1)	22.5	0.76	65.6	**0.07**
Ours (iteration 2)	22.7	0.81	63.0	0.09
Ours (iteration 3)	**22.8**	**0.82**	**62.6**	0.10
T2	**PSNR**	**mean SSIM**	**MAE**	**Hellinger distance**
[Prastawa 2009]	16.5	0.67	97.2	0.41
Ours (iteration 1)	18.5	0.69	76.0	**0.09**
Ours (iteration 2)	18.7	0.74	73.3	0.11
Ours (iteration 3)	**18.8**	**0.75**	**73.0**	0.11

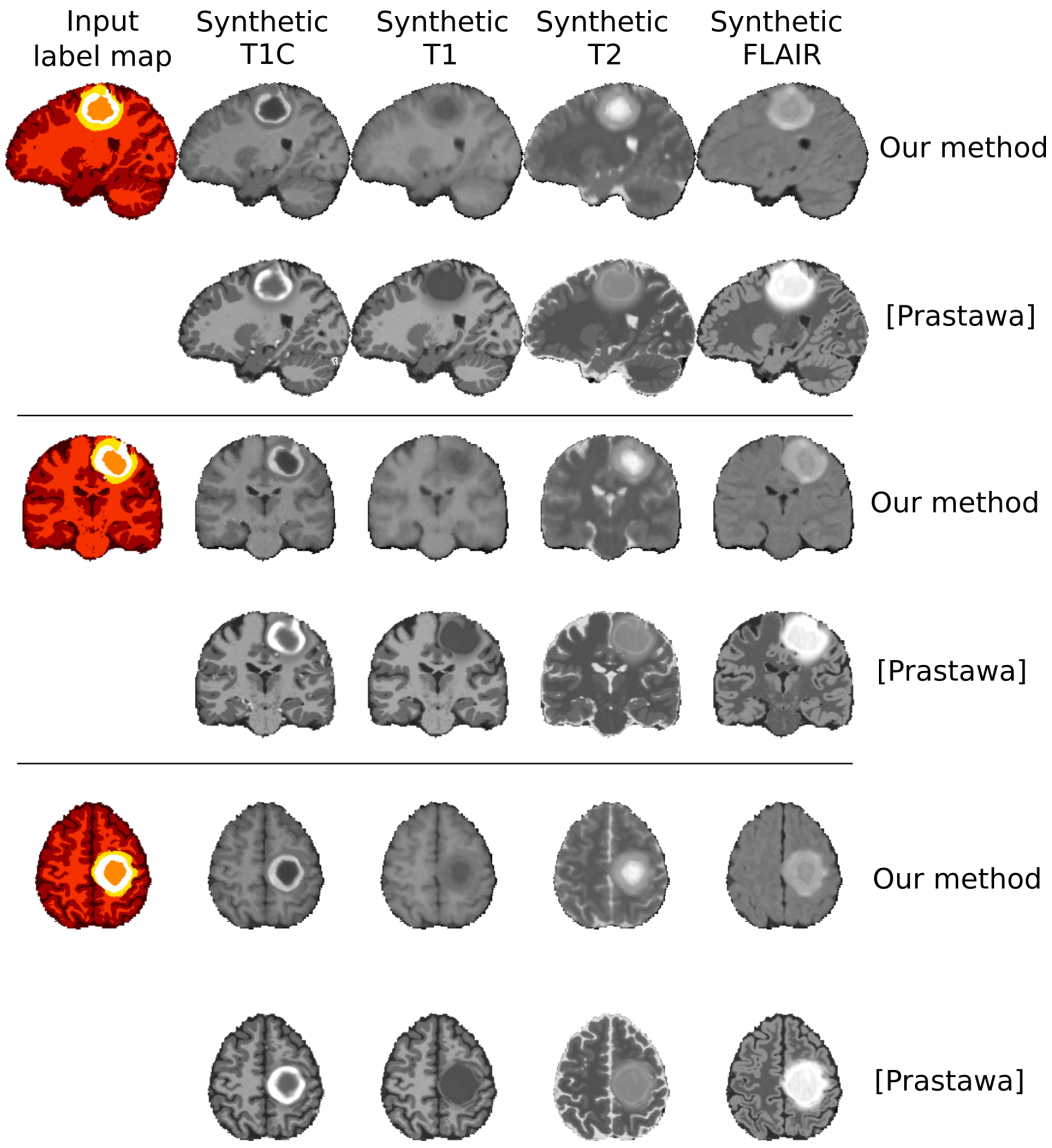

Figure 4.20: Illustration of the modularity of our image synthesis model. Synthetic 1 mm isotropic MRI based on a label map generated by a reaction-diffusion model with mass effect [Prastawa 2009]. Top to bottom: sagittal, coronal and axial views, for synthetic MRI, respectively our method on even rows and [Prastawa 2009] on odd rows. The results of our method were obtained after two iterations (one feedback loop). Please note that in this specific example, probabilistic label maps were used as input for [Prastawa 2009], but not for our method.

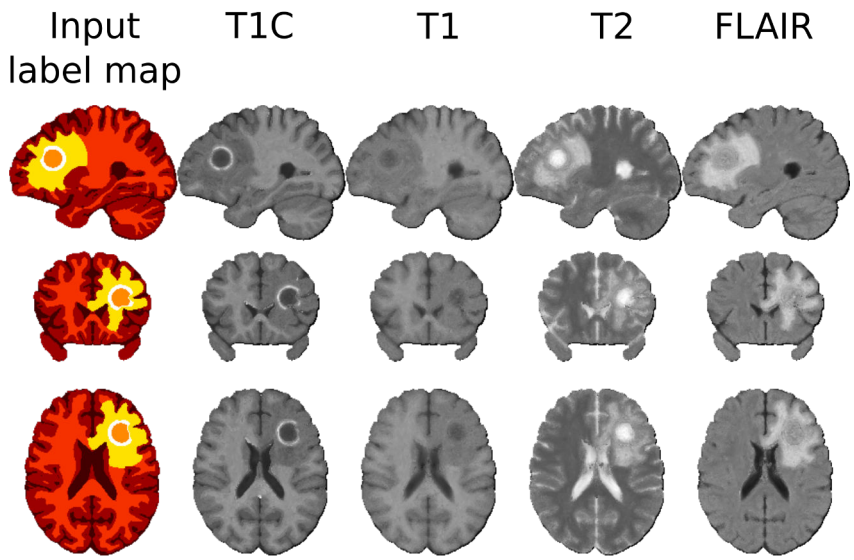

Figure 4.21: Illustration of the modularity of our image synthesis model. Synthetic 1 mm isotropic MRI based on a label map generated by the PIHNA tumor growth 3D model. Obtained after one iteration (no feedback).

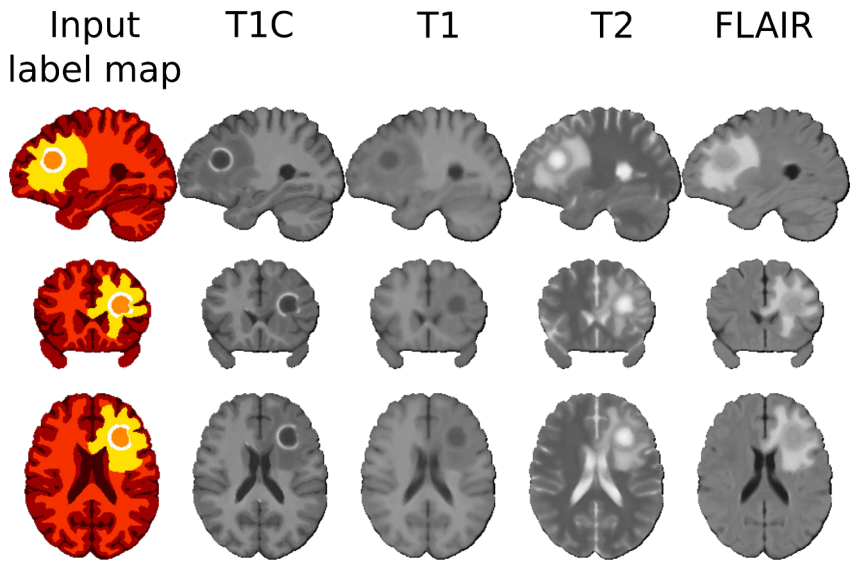

Figure 4.22: Illustration of the modularity of our image synthesis model. Synthetic 1 mm isotropic MRI based on a label map generated by the PIHNA tumor growth 3D model. Obtained after three iterations. Results look more consistent due to the feedback loop.

4.4 Discussion

In the experiments presented in this chapter, the iterative nature of the proposed image synthesis method allows to reach higher PSNR and synthesize qualitatively more realistic images. The optimal number of iterations seems to be about 2 or 3. The conclusion regarding feedback weights is not straightforward: it seems a small non-zero weight is to be preferred, and to this regard, the strategy, hinted in [Ye 2013], which consists in progressively increasing the feedback weight with the number of iterations seems relevant.

Several MR channels are simultaneously synthesized, which leads to a more constrained patch match (for iterations $t > 1$), and ultimately to more consistent synthetic MRI. Less consistent results could arise when trying to synthesize a single image modality.

One limitation of the proposed method is the running time: with 20 atlases, one iteration to simultaneously synthesize the 4 MR channels (for the whole brain) requires about 2 days of computation (16 cores, 100 GB of RAM) for 1 mm isotropic MRI, and about 3 hours of computation (3 cores, 20 GB of RAM) for 2 mm isotropic MRI. A visual comparison of the synthesis of 1 mm and 2 mm MRI is presented in Figure 4.23. Typically, sulci are less visible on 2 mm synthetic MRI. We have not tried to decrease the number of atlases since this would require a strategy to pre-select relevant atlases based on some similarity criterion between label maps. We have not tried to specifically optimize the implementation of the algorithm, mostly due to the prohibitive running time which renders any attempt at optimizing the implementation tedious and susceptible of introducing bugs in the program.

Regarding running time in the perspective of generating a large database of synthetic cases, a region-of-interest enclosing the tumor could be defined such that the rest of the synthetic image is supplied by a template. Indeed, tumor growth models are based on a healthy atlas in which a tumor seed is placed, and a T1 template is available for the healthy atlas [Mazziotta 2001]. In this case, image intensity does not have to be optimized for voxels outside the ROI, since these voxels are assumed not to present intensity abnormalities.

All the experiments were conducted with a label map as single input. Due to the iterative process described in this chapter, it would be straightforward to provide as input an MR channel on top of the label map. Typically, if a contrast-enhanced T1 MR image is available on top of the label map, we could synthesize MR channels such as FLAIR or T2. Additional inputs are merely additional constraints for the patch match.

4.5 Summary and Conclusions

We presented a generative model for the synthesis of multi-modal medical images of pathological cases by using an iterative multi-atlas patch-based algorithm. The method was illustrated on MRI synthesis of glioma-bearing brains. A single label maps allows to synthesize realistic images, with clear improvements with regard to

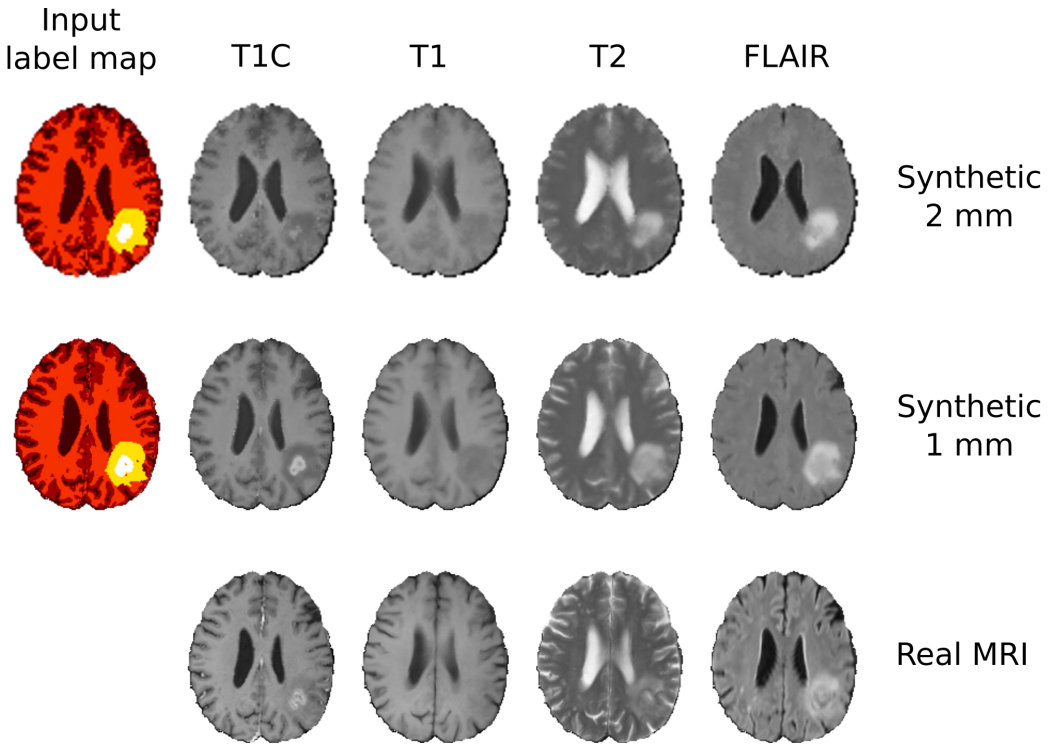

Figure 4.23: Influence of the resolution at which the image synthesis is performed, observed on a high-grade case (HG-0010). Top to bottom: axial views for synthetic 2 mm, synthetic 1 mm and real MRI. Sulci are less visible on MRI synthesized at 2 mm resolution.

the state-of-the-art method. The approach is flexible and can be applied *i)* either to label maps of real cases to synthesize artifact-free MRI, *ii)* or to label maps simulated by any tumor growth model. The standardization of images is made transparent and reversible, so that synthetic data can be used for any application with any other dataset, including intensity inhomogeneity or different levels of Rician noise. Moreover, due to the strategy used to solve for the maximum a posteriori, estimation of image synthesis uncertainty is straightforward.

Such a generative model would allow the availability of large, public and annotated datasets of synthetic medical images obtained with virtual patients, which could prove useful for the validation or benchmarking of image processing algorithms. Promising results have been obtained on real data after training machine learning algorithms on synthetic data [Geremia 2013], which could be further improved by using a more realistic synthetic dataset, generated with the proposed image synthesis model. Finally, image synthesis of pathological cases bridges the gap between outputs of tumor growth model and medical images, which are the most clinically relevant observations. The development of image synthesis could benefit to interpretability and personalization of tumor growth models.

Chapter 5

Conclusion

Contents

The main objective of this thesis was the development of unifying patch-based methods for the analysis and the synthesis of medical images of pathological cases. Two different generative probabilistic models have been proposed: *i*), a patch-based model for the automatic segmentation of brain tumor regions from MRI, and *ii*) a mirrored model for the synthesis of realistic MRI of pathological brains. For this purpose, the proposed methods apply patch-based techniques without relying on the concept of local search window.

5.1 Contributions

The first contribution of this thesis is a generic approach to perform fully-automatic segmentation of pathologies on brain medical images by using multi-atlas patch-based voting techniques, without relying on a local search window. Segmentation results are accurate thanks to the combination of an efficient representation of patches at multiple scales, the augmentation of the patch dataset by applying cube isometries to the observed patches, and finally an augmentation of patch features with intensity and location statistics to implicitly guide the patch match. The algorithm is computationally efficient due several enhancements: the automatic detection of a small region-of-interest enclosing high-probability tumor volumes based on a spatially-varying prior, the pre-selection of patches based on label purity and intensity statistics, and vote stratification with respect to cases and labels. Moreover, thanks to stratification, patch match is effectively embarrassingly parallel, which results in even shorter running times using a computer grid.

This method has been applied for three consecutive years to the publicly available datasets of the MICCAI Multimodal Brain Tumor Image Segmentation benchmark challenge. State-of-the-art results are achieved in a fully automatic setting, with very limited to no prior learning, which limits the risk of overfit. Moreover, due to patch overlap, segmentation results only require minimal post-processing, which confirms the robustness of the proposed segmentation approach. Numerous experiments were conducted to help better understand the model and the impact of method parameters.

The second contribution of this thesis is a generative model for the synthesis, based on a single label map, of realistic multi-modal medical images of pathological cases. The model builds upon *i)* the generative model proposed for multi-atlas patch-based segmentation, *ii)* the Modality Propagation iterative strategy for a spatially-coherent synthesis. The approach is flexible and can be applied *i)* either to label maps of real cases to synthesize artifact-free MRI, *ii)* or to label maps simulated by any tumor growth model. Due to the strategy used to solve for the maximum a posteriori, estimation of image synthesis uncertainty is straightforward. Moreover, the standardization of images is made transparent and reversible, so that synthetic data can be used for any application with any other dataset.

An application to Magnetic Resonance Imaging synthesis of glioma-bearing brains was *i)* validated on the training dataset of the MICCAI Multimodal Brain Tumor Image Segmentation challenge, *ii)* compared to the state-of-the-art method in glioma image synthesis, and *iii)* illustrated using the output of two different tumor growth models. Such a generative model would allow the generation of a large, public and annotated dataset of synthetic medical images, which could prove useful for the training, validation, or benchmarking of image processing algorithms.

A strength of the proposed models is their consistency: from a perspective, the two generative models mirror one another, so the model for synthesis can benefit from the experience that was gained when studying the model for segmentation.

5.2 Future work

In this thesis, we considered 4D signals, composed of 3D patches extracted from multiple MR channels (T1, T1C, T2, FLAIR). The addition of a 5th dimension, namely the temporal dimension, would allow to perform a consistent segmentation of longitudinal data. Indeed, image datasets of tumor-bearing human brains include time series of MRI acquisitions for patients who are periodically scanned for tumor growth monitoring. It would also be of interest to include additional information coming from MR spectroscopy on top of the four MR channels which we considered.

Glioblastoma exhibit a phenomenon known as mass effect: healthy brain tissues are displaced due to the tumor growth. It would be relevant to take into account the deformation of image patches due to mass effect. There are several directions to do so: *i*) add cases with an important mass effect to the training dataset, *ii*) consider a bio-mechanical model to cancel the mass effect, as pre-processing step, and *iii*) augment patch features with features which reflect the mass effect, such as symmetry features, e.g. symmetry with respect to the mid-sagittal plane.

The iterative strategy proposed for image synthesis resulted in quantitatively more realistic synthetic images. Since the models for segmentation and synthesis are mirrored, it could be of interest to perform a similar iterative strategy for segmentation: features would be augmented with the probabilistic segmentation obtained at the previous iteration. This is a process called auto-context [Tu 2010], which could lead to better segmentation results.

The proposed segmentation method could be applied to other pathologies, such as pulmonary nodules for the lungs. We could build atlases of label spatial extent for the lungs, as it was done in this thesis for the brain. We could also use augmented features, e.g. an augmentation with the distance to the bronchial tree. Other applications also include other brain pathologies, e.g. meningioma.

A method to estimate the uncertainty within the patch-based image synthesis framework was proposed in this thesis. An interesting development would be the application of uncertainty estimates to image intensity outlier detection. Indeed, in contrast to [Cardoso 2015], our approach does not assume that the pathology corresponds to outliers, so outlier detection based on our approach would identify regions of the image, irrespectively of their healthy or pathological state, which are not well explained by the annotated image dataset. A straightforward implication is that we could expand a training dataset in a wise manner, by focusing on previously unseen regions of interest, and by avoiding the addition of redundant image information to the training dataset.

There are different methods to randomly sample synthetic images based on the proposed synthesis framework. First, in this thesis, we always considered the whole set of atlases when performing image synthesis. We could obtain different synthetic images, based on the same input label map, by bootstrapping, i.e. randomly sampling (with replacement) a subset of atlases and then performing image synthesis. Second, with our synthesis model, we obtain a synthetic image as well as an estimation of synthesis uncertainty. We could add a spatial constraint with a Markov Random Field, and then sample spatially-consistent synthetic images based on both intensity (given by our model) and regional constraints (given by the Markov property).

The computational cost of the proposed iterative synthesis method is really high. A future direction to solve this issue would be a coarse-to-fine approach, with typically the scale getting finer with the number of iterations. This would allow to obtain a high-resolution synthetic image with only the last iteration performed at the finest scale.

Synthetic images tend to miss fine details such as some of the brain sulci. This problem could be solved by considering fewer patch matches for the approximate marginalization. This work direction is akin to selecting either globally or locally the best atlases for the synthesis. This would result in sharper synthetic images, while decreasing the computational cost of synthesis.

With patch-based methods, we assume that the function that we want to predict (probabilistic labels for segmentation, multi-channel image intensity for synthesis) could be inferred at the center of the target patch based on a measure of patch similarity, therefore we ultimately perform a point-wise estimation, i.e. we only estimate the value at the center of the target patch. Another estimation method would consist in estimating values for the whole target patch, this is known as multi-point estimation. According to [Rousseau 2011], multi-point estimators tend to lead to better segmentation results. It would be of interest to compare point-wise and multi-point estimation for image synthesis. Indeed, synthetic images could

present more texture and finer details when using multi-point estimates. Moreover, an intermediate estimate between point-wise and multi-point estimators could be introduced, thanks to the multi-scale patches presented in this thesis: an estimation could be performed for the whole conventional patch which lies at the center of the multi-scale patch. This could lead to finer synthetic images, as the intuition with patch-based methods is that the estimation is more accurate near the center of the target patch.

Finally, in [Geremia 2013], promising results were obtained on real data after training machine learning algorithms on synthetic data. The benefit for segmentation of the image synthesis of missing modalities is analyzed in [van Tulder 2015]. A straightforward future development of the work presented in this thesis would be *i*) the generation of a synthetic dataset of realistic annotated cases, *ii*) an in-depth analysis of the impact of using this realistic synthetic dataset for segmentation. This would allow to understand better the intricacies between the proposed models for segmentation and synthesis, and pave the way for an even more unified methodology.

Appendix A

Appendix: segmentation of 2 mm isotropic MR Images

In this appendix, we present segmentation results obtained with the multi-atlas patch-based approach presented in Chapter 2. The input consists of multi-modal MR images (T1C, T1, T2, FLAIR), with a 2 mm isotropic resolution, provided for the MICCAI 2013 Multimodal Brain Tumor Image Segmentation benchmark challenge (BraTS).

Position	User	Dice			Positive Predictive Value			Sensitivity		
		complete	core	enhancing	complete	core	enhancing	complete	core	enhancing
1	kwond1	0.88 (1)	0.83 (1)	0.72 (8)	0.92 (4)	0.90 (2)	0.74 (9)	0.84 (14)	0.78 (8)	0.72 (11)
2	tustn1	0.87 (4)	0.78 (5)	0.74 (1)	0.85 (13)	0.74 (21)	0.69 (18)	0.89 (5)	0.88 (1)	0.83 (2)
3	al-ss1	0.87 (2)	0.78 (4)	0.70 (11)	0.89 (5)	0.83 (6)	0.75 (6)	0.86 (13)	0.78 (7)	0.70 (16)
4	anon2	0.87 (3)	0.77 (7)	0.73 (6)	0.84 (14)	0.79 (15)	0.73 (10)	0.90 (4)	0.78 (9)	0.75 (7)
5	pinta1	0.83 (17)	0.78 (3)	0.73 (3)	0.86 (10)	0.86 (5)	0.71 (13)	0.82 (16)	0.75 (14)	0.79 (3)
6	kleej1	0.86 (6)	0.73 (15)	0.74 (2)	0.83 (18)	0.70 (23)	0.76 (5)	0.91 (3)	0.80 (5)	0.75 (8)
7	davya1	0.85 (8)	0.78 (6)	0.73 (5)	0.95 (1)	0.81 (8)	0.70 (17)	0.78 (22)	0.76 (12)	0.77 (6)

Figure A.1: Snapshot as of April 28, 2015 of the top-ranking methods on the MICCAI 2013 BRATS benchmark website. Kwon's submission (kwond1) is semi-automatic, Tustison's submission (tustn1) is automatic, other methods are unknown. Our submission (anon2) is automatic and highlighted in blue. Specific rankings appear in brackets for each measure.

Table A.1: Average assessment measures for BraTS 2013, 2014 & 2015 benchmarks. The training dataset is always the 2013 Evaluation dataset. Tumor regions are: complete tumor (CT), tumor core (TC), and enhancing core (EC). Average symmetric surface distance (ASSD) and Hausdorff distance (HD) are expressed in millimeters. Grade counts are indicated in the first column as (HG / LG). The 2014 Challenge dataset consists of 97 cases, presumably all HG. For other datasets, tumor grades are known, as given by the benchmark organizers. The 2013 Leaderboard dataset is a compromised test dataset, since part of the ground truth was mistakenly available online for the first few hours, which flaws comparisons with other competitors on this dataset.

2013	**Region**	**Grade**	**Dice**	**Precision**	**Recall**	**ASSD**	**HD**
Challenge	CT	HG	0.87	0.85	0.89		
(10 / 0)	TC	HG	0.77	0.80	0.76		
	EC	HG	0.72	0.71	0.77		
Leaderboard	CT	HG / LG	0.79 / 0.43	0.76 / 0.35	0.84 / 0.74		
(21 / 4)	TC	HG / LG	0.59 / 0.31	0.60 / 0.25	0.64 / 0.73		
	EC	HG	0.55	0.48	0.69		
Evaluation	CT	HG / LG	0.84 / 0.83	0.87 / 0.82	0.82 / 0.87	2.9 / 4.5	25.8 / 27.9
(20 / 10)	TC	HG / LG	0.68 / 0.62	0.73 / 0.62	0.68 / 0.72	9.8 / 5.1	38.3 / 36.1
	EC	HG	0.65	0.69	0.66	9.7	33.8
2014	**Region**	**Grade**	**Dice**	**Precision**	**Recall**	**ASSD**	**HD**
Challenge	CT	HG	0.84	0.89	0.82		
(97 / 0)	TC	HG	0.68	0.71	0.71		
	EC	HG	0.67	0.62	0.81		
Evaluation	CT	HG / LG	0.84 / 0.80	0.88 / 0.86	0.82 / 0.78	2.2 / 3.5	26.4 / 31.7
(257 / 57)	TC	HG / LG	0.70 / 0.53	0.68 / 0.53	0.76 / 0.66	3.9 / 4.4	33.9 / 35.9
	EC	HG	0.68	0.62	0.81	2.9	27.5
2015	**Region**	**Grade**	**Dice**	**Precision**	**Recall**	**ASSD**	**HD**
Evaluation	CT	HG / LG	0.74 / 0.63	0.87 / 0.80	0.67 / 0.54	5.2 / 9.3	35.3 / 55.0
(220 / 54)	TC	HG / LG	0.58 / 0.21	0.52 / 0.17	0.73 / 0.48	6.4 / 11.6	37.9 / 52.0
	EC	HG	0.62	0.59	0.71	5.7	32.0

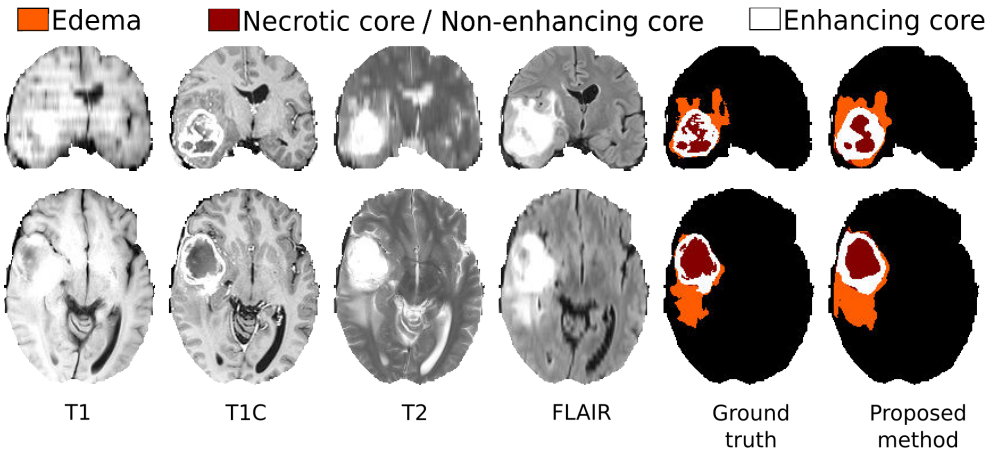

Figure A.2: Segmentation of HG-0001 (coronal and axial views) using 2 mm data.

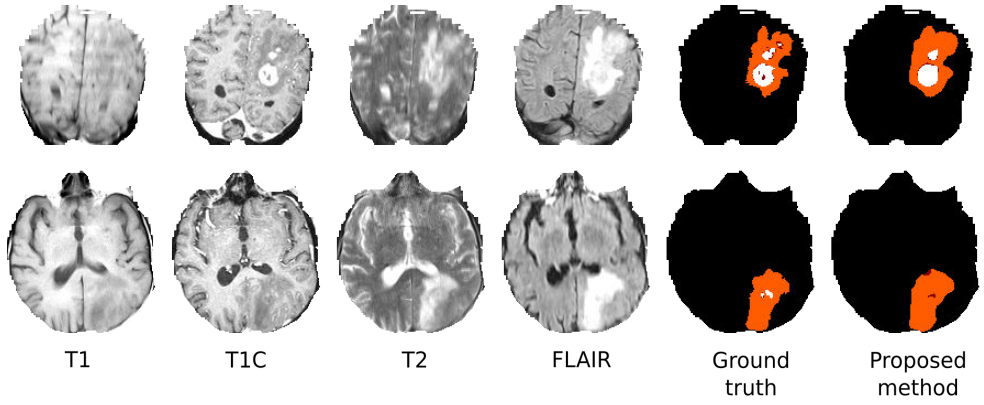

Figure A.3: Segmentation of HG-0002 (coronal and axial views) using 2 mm data.

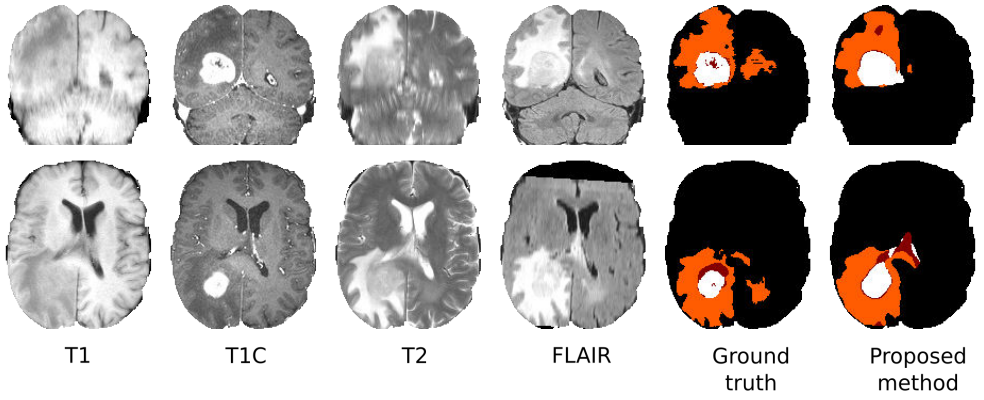

Figure A.4: Segmentation of HG-0003 (coronal and axial views) using 2 mm data.

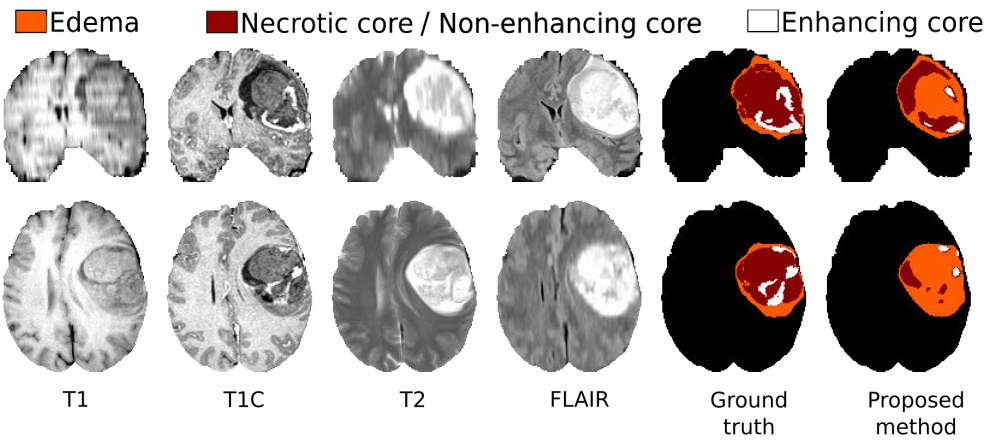

Figure A.5: Segmentation of HG-0004 (coronal and axial views) using 2 mm data.

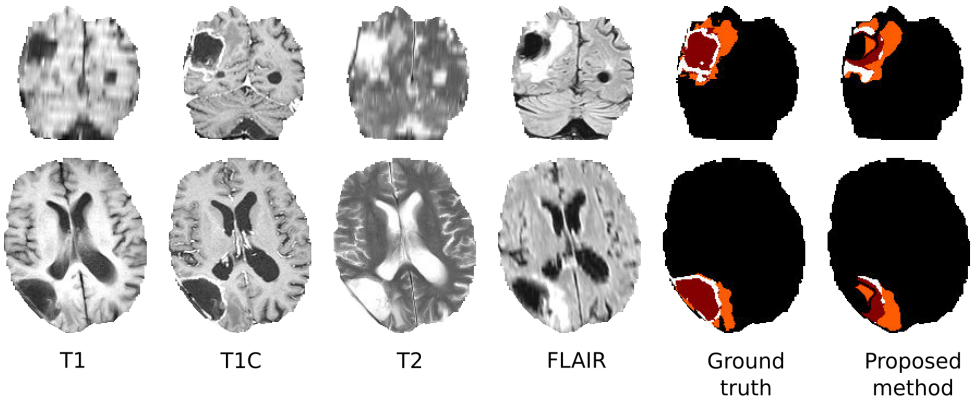

Figure A.6: Segmentation of HG-0005 (coronal and axial views) using 2 mm data.

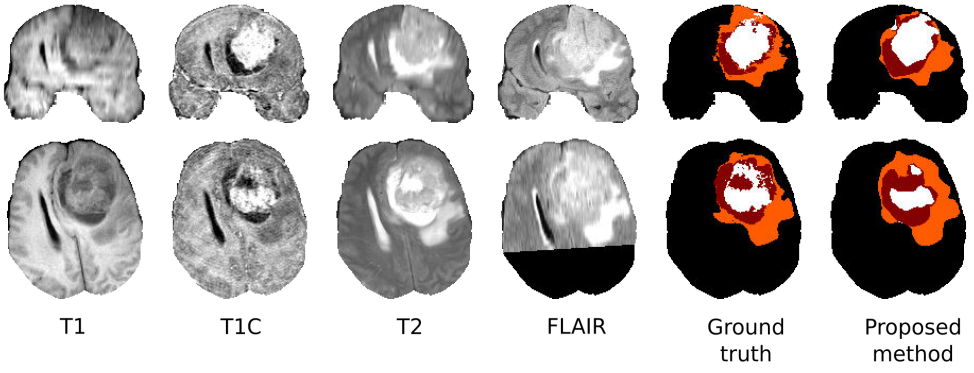

Figure A.7: Segmentation of HG-0006 (coronal and axial views) using 2 mm data.

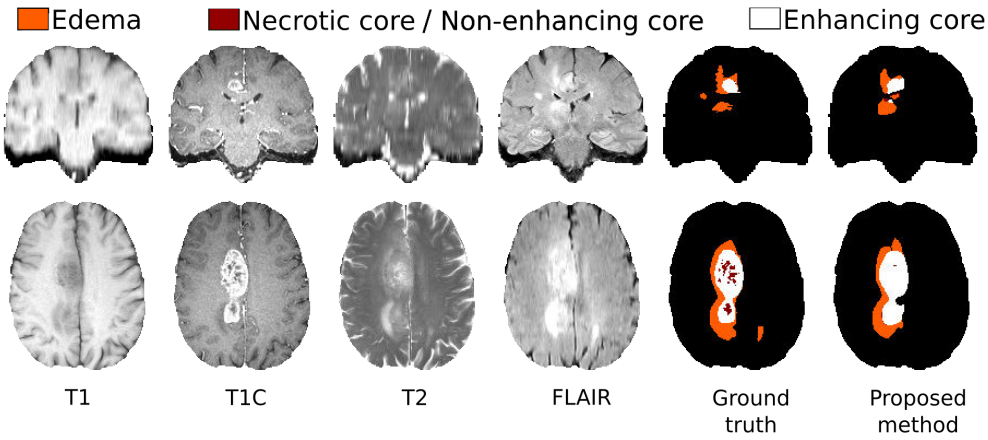

Figure A.8: Segmentation of HG-0007 (coronal and axial views) using 2 mm data.

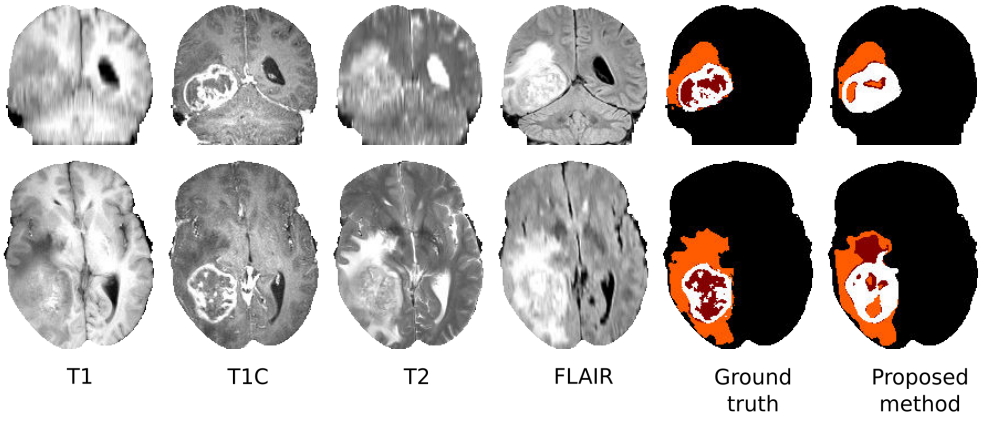

Figure A.9: Segmentation of HG-0008 (coronal and axial views) using 2 mm data.

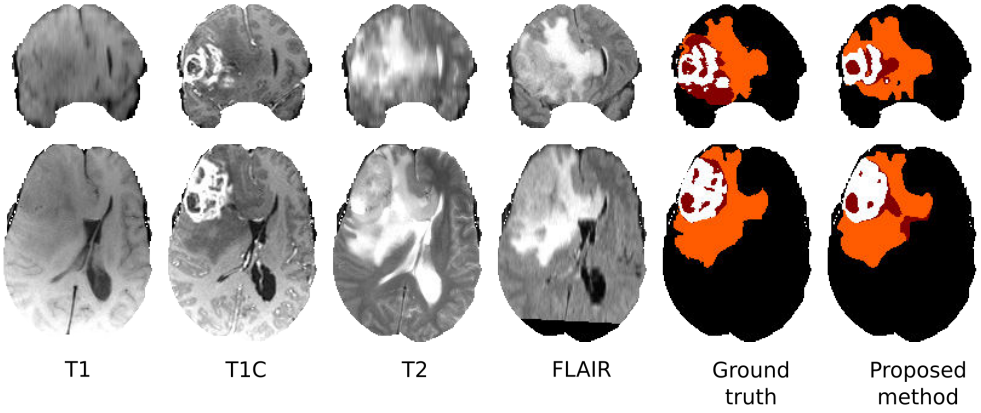

Figure A.10: Segmentation of HG-0009 (coronal and axial views) using 2 mm data.

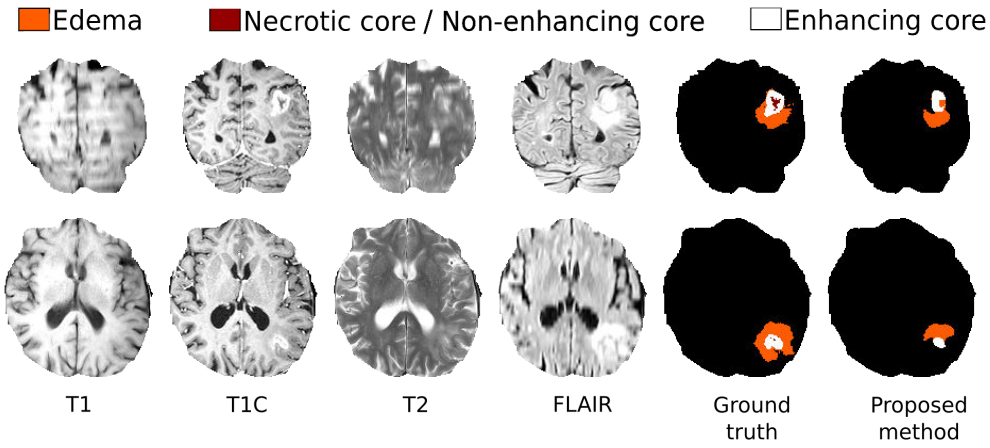

Figure A.11: Segmentation of HG-0010 (coronal and axial views) using 2 mm data.

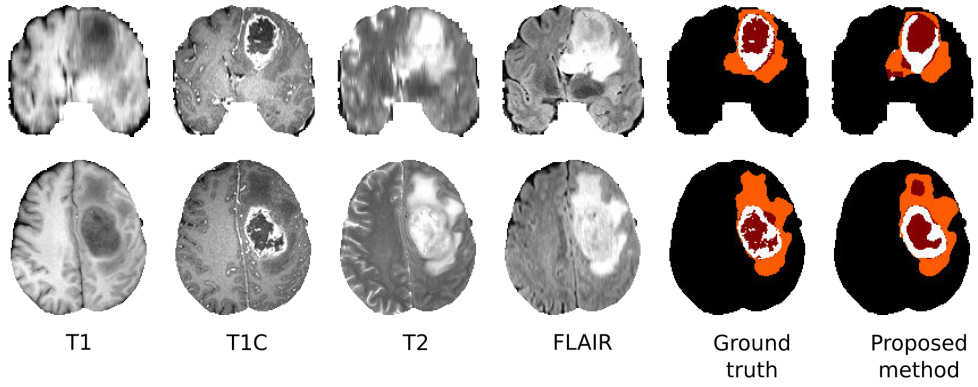

Figure A.12: Segmentation of HG-0011 (coronal and axial views) using 2 mm data.

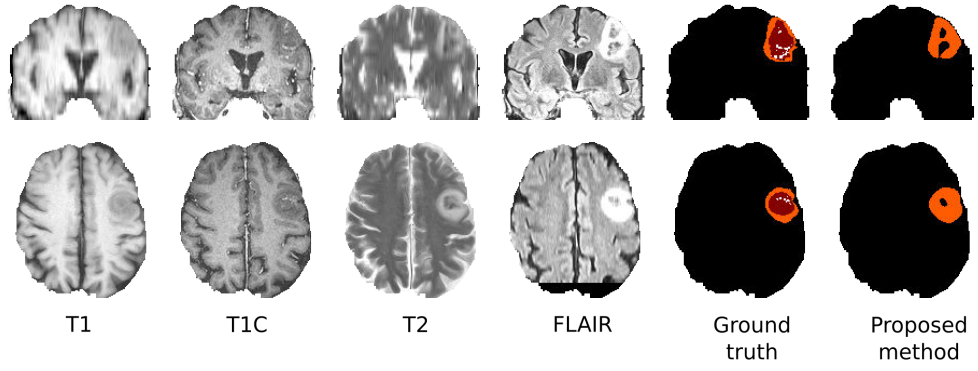

Figure A.13: Segmentation of HG-0012 (coronal and axial views) using 2 mm data.

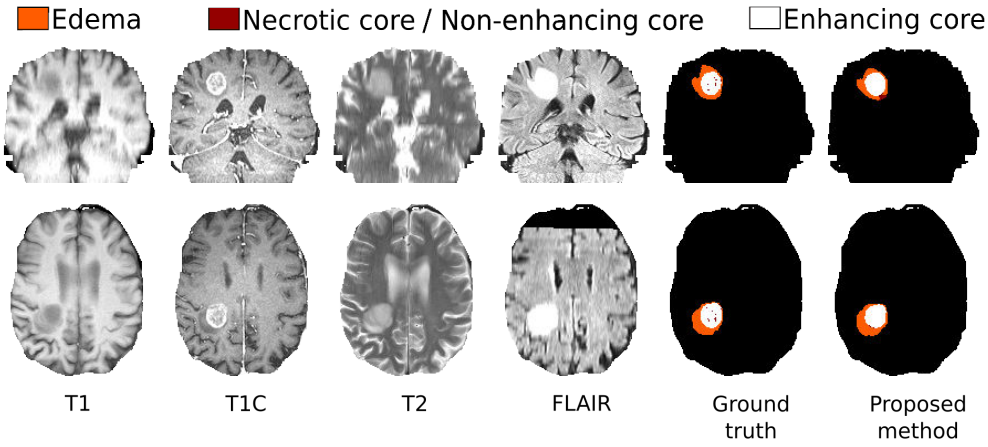

Figure A.14: Segmentation of HG-0013 (coronal and axial views) using 2 mm data.

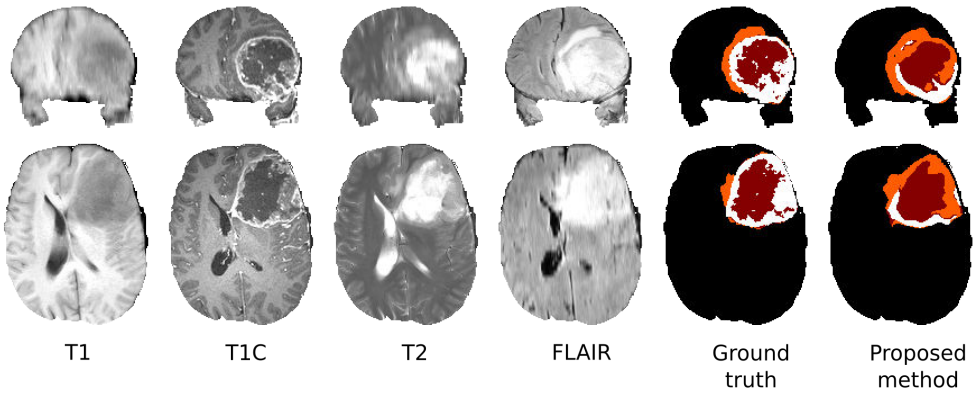

Figure A.15: Segmentation of HG-0014 (coronal and axial views) using 2 mm data.

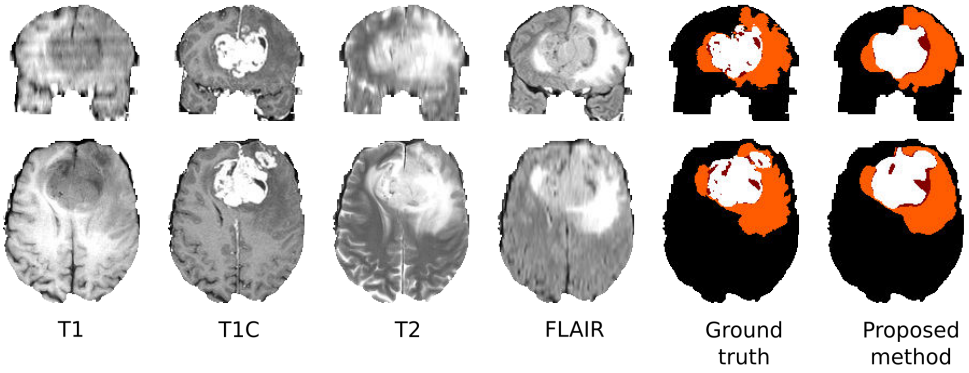

Figure A.16: Segmentation of HG-0015 (coronal and axial views) using 2 mm data.

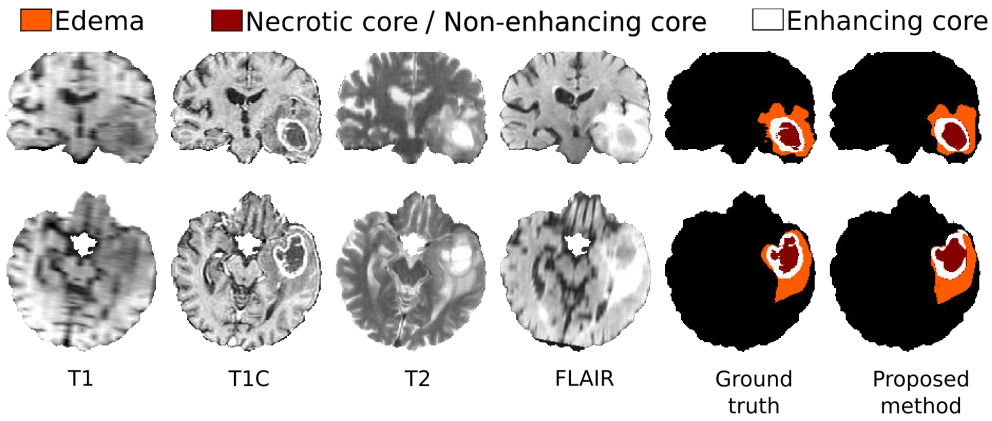

Figure A.17: Segmentation of HG-0022 (coronal and axial views) using 2 mm data.

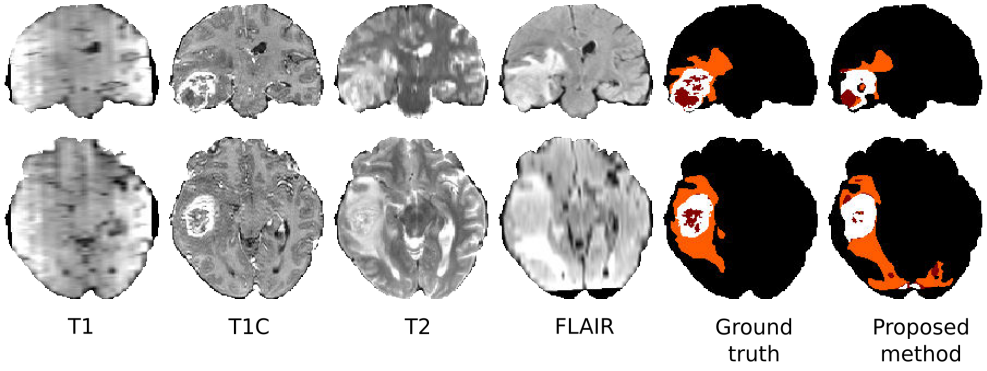

Figure A.18: Segmentation of HG-0024 (coronal and axial views) using 2 mm data.

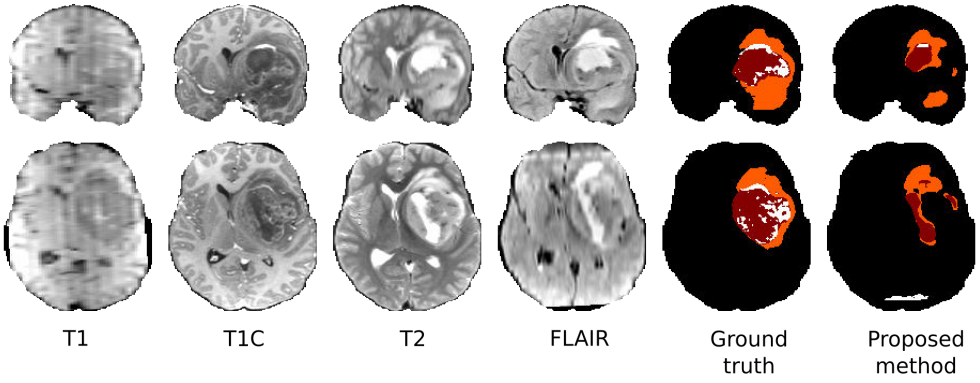

Figure A.19: Segmentation of HG-0025 (coronal and axial views) using 2 mm data.

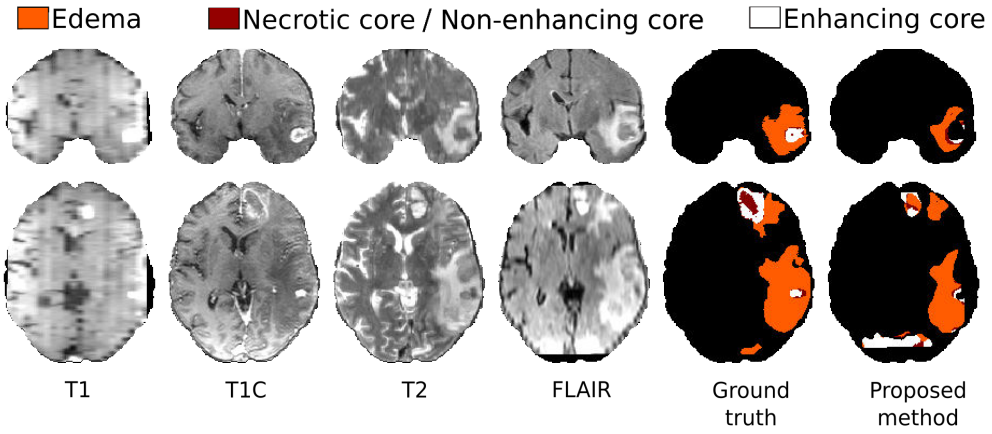

Figure A.20: Segmentation of HG-0026 (coronal and axial views) using 2 mm data.

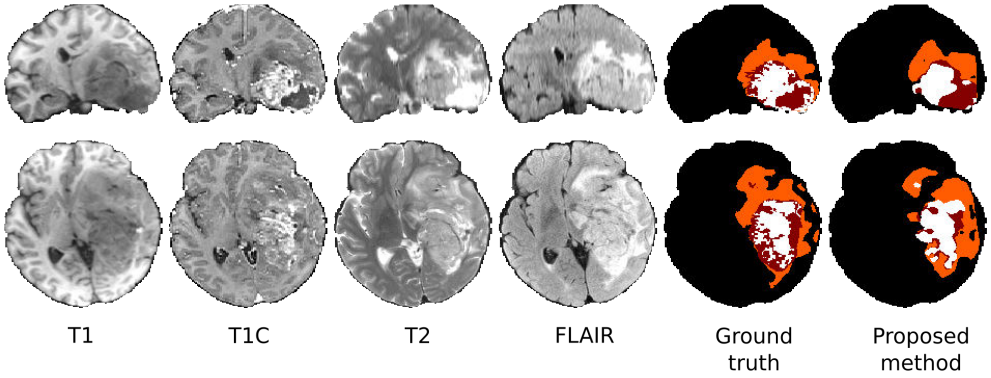

Figure A.21: Segmentation of HG-0027 (coronal and axial views) using 2 mm data.

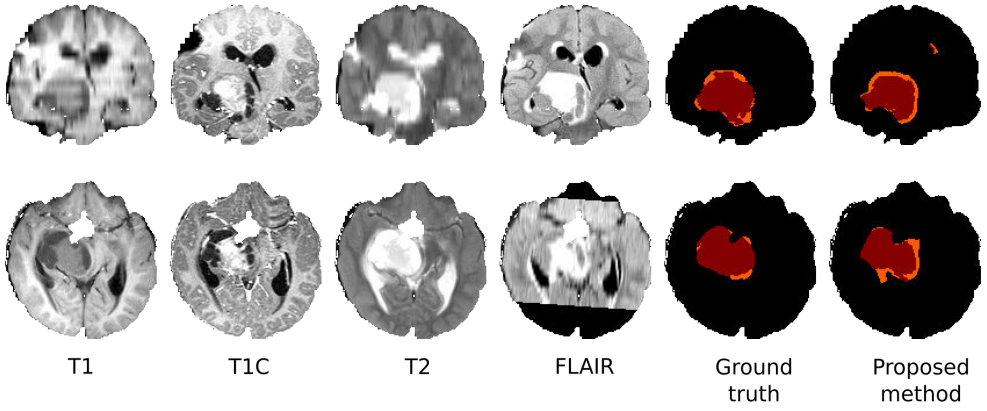

Figure A.22: Segmentation of LG-0001 (coronal and axial views) using 2 mm data.

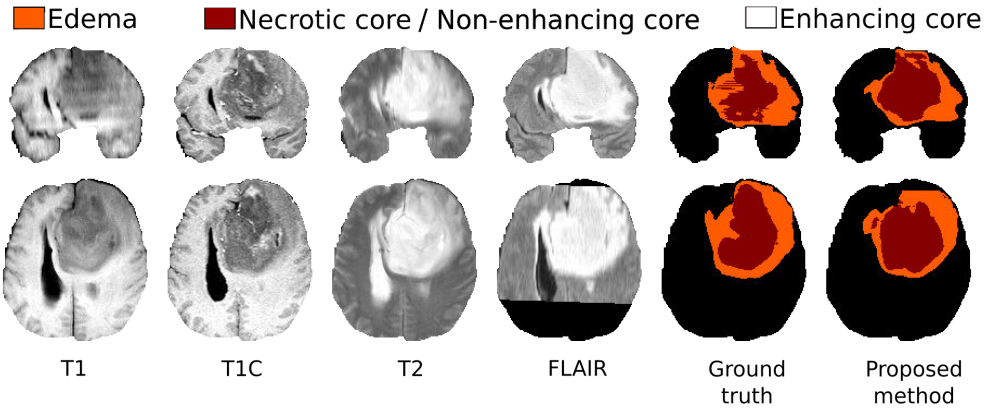

Figure A.23: Segmentation of LG-0002 (coronal and axial views) using 2 mm data.

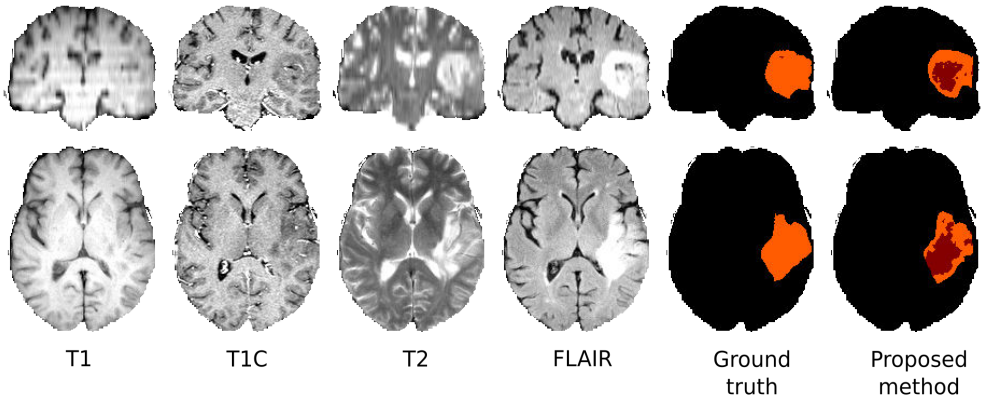

Figure A.24: Segmentation of LG-0004 (coronal and axial views) using 2 mm data.

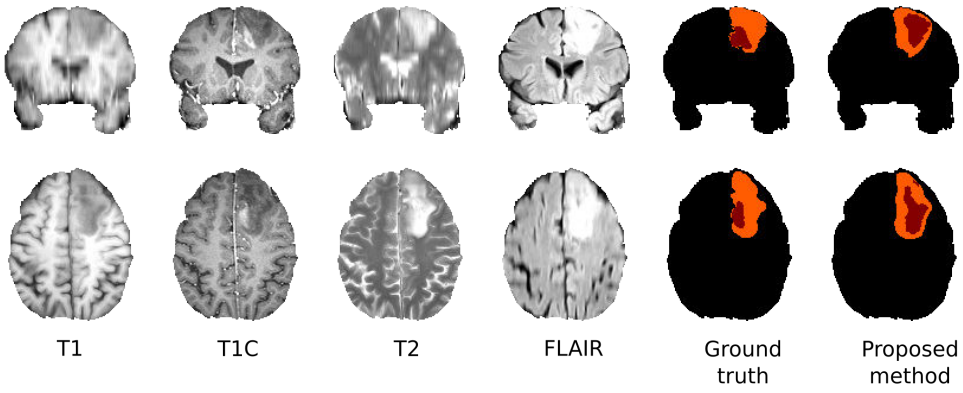

Figure A.25: Segmentation of LG-0006 (coronal and axial views) using 2 mm data.

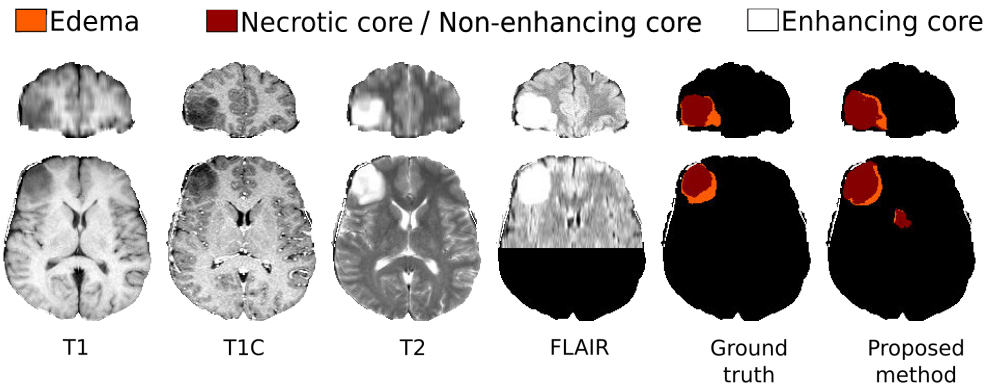

Figure A.26: Segmentation of LG-0008 (coronal and axial views) using 2 mm data.

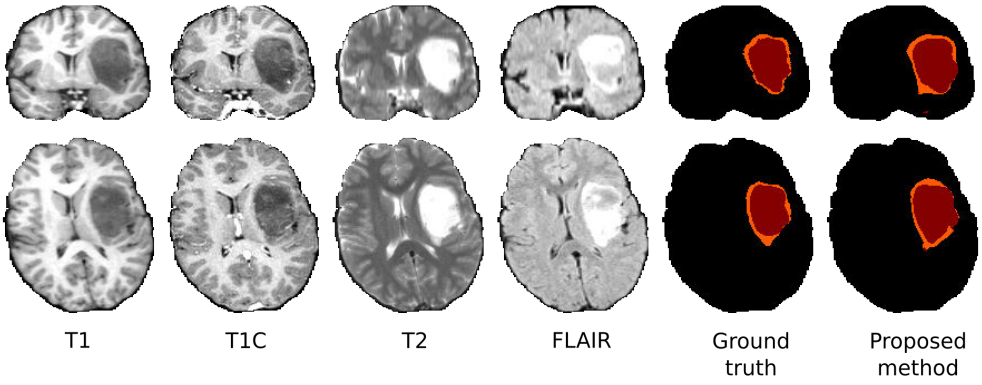

Figure A.27: Segmentation of LG-0011 (coronal and axial views) using 2 mm data.

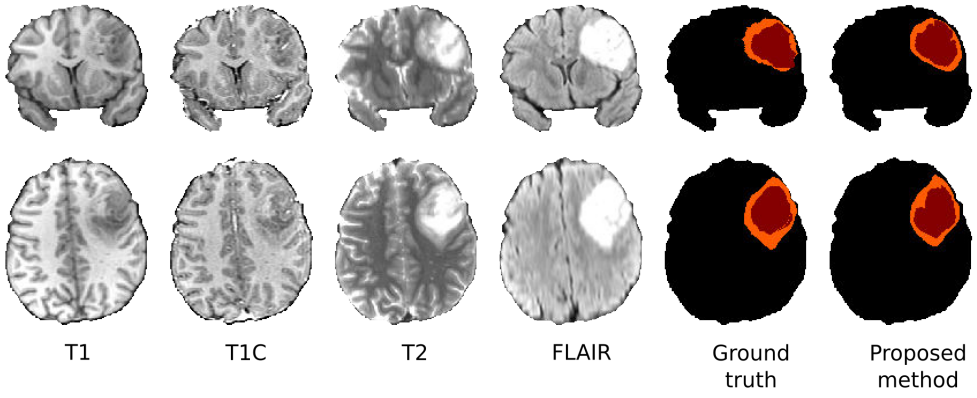

Figure A.28: Segmentation of LG-0012 (coronal and axial views) using 2 mm data.

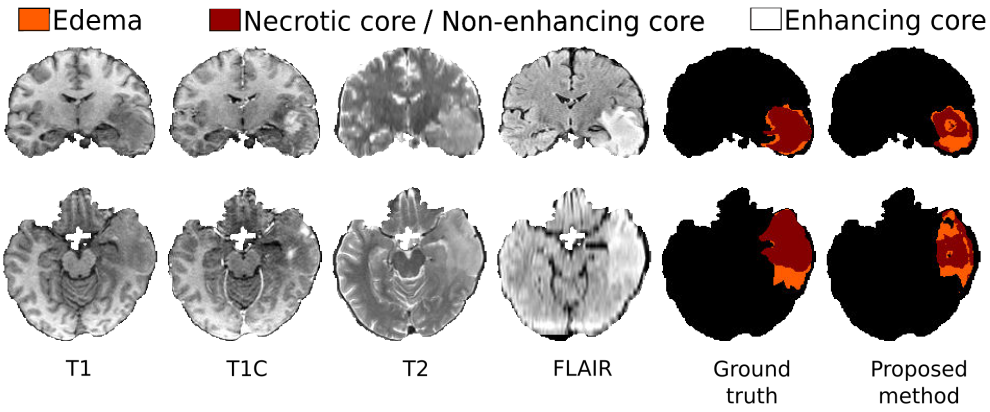

Figure A.29: Segmentation of LG-0013 (coronal and axial views) using 2 mm data.

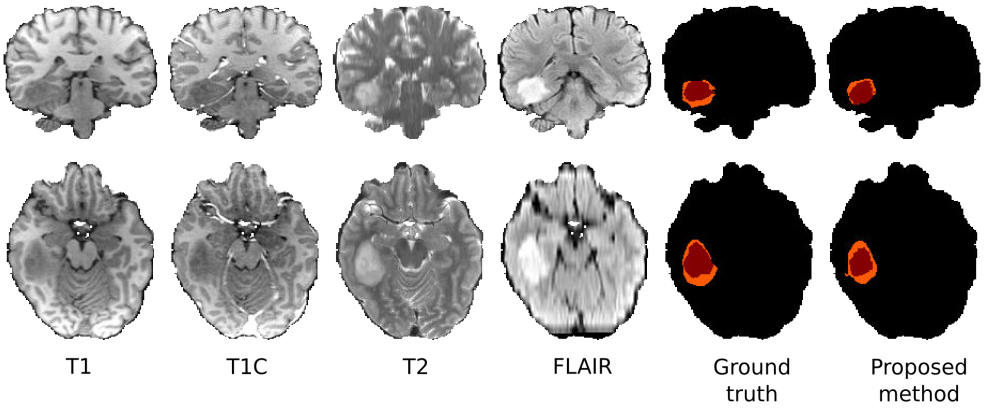

Figure A.30: Segmentation of LG-0014 (coronal and axial views) using 2 mm data.

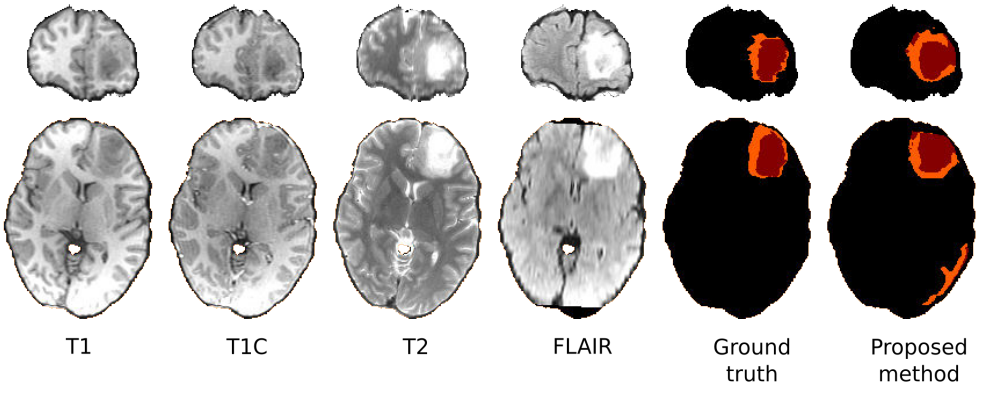

Figure A.31: Segmentation of LG-0015 (coronal and axial views) using 2 mm data.

Appendix B

Appendix: synthesis of 1 mm isotropic MR Images

In this appendix, we present the 1 mm isotropic synthetic MRI obtained after 3 iterations of the image synthesis process described in Chapter 4. The feedback weight α_t increases linearly with the number of iterations: α_t is equal to 0, $\frac{1}{3}$, $\frac{2}{3}$ respectively for the iterations $t = 1$, 2 and 3. The input of the program is a label map obtained from the MICCAI 2013 Multimodal Brain Tumor Image Segmentation benchmark challenge (BraTS). The output of the program is composed of four synthetic MR sequences (T1C, T1, T2, FLAIR). For each of the cases on which we test the algorithm, we provide the ground truth, i.e. MRI provided by the BraTS Challenge organizers, for visual comparison.

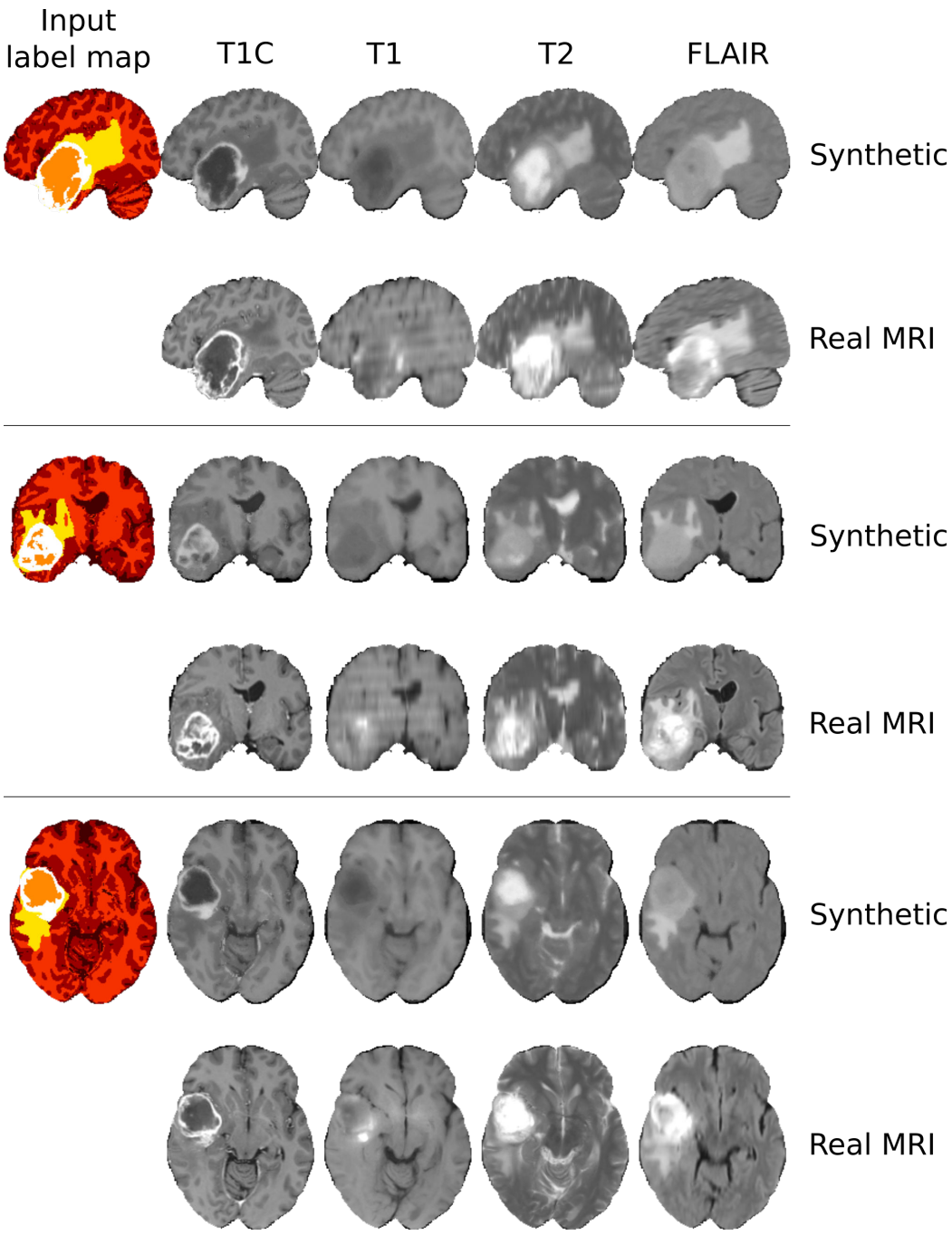

Figure B.1: High-grade HG-0001. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

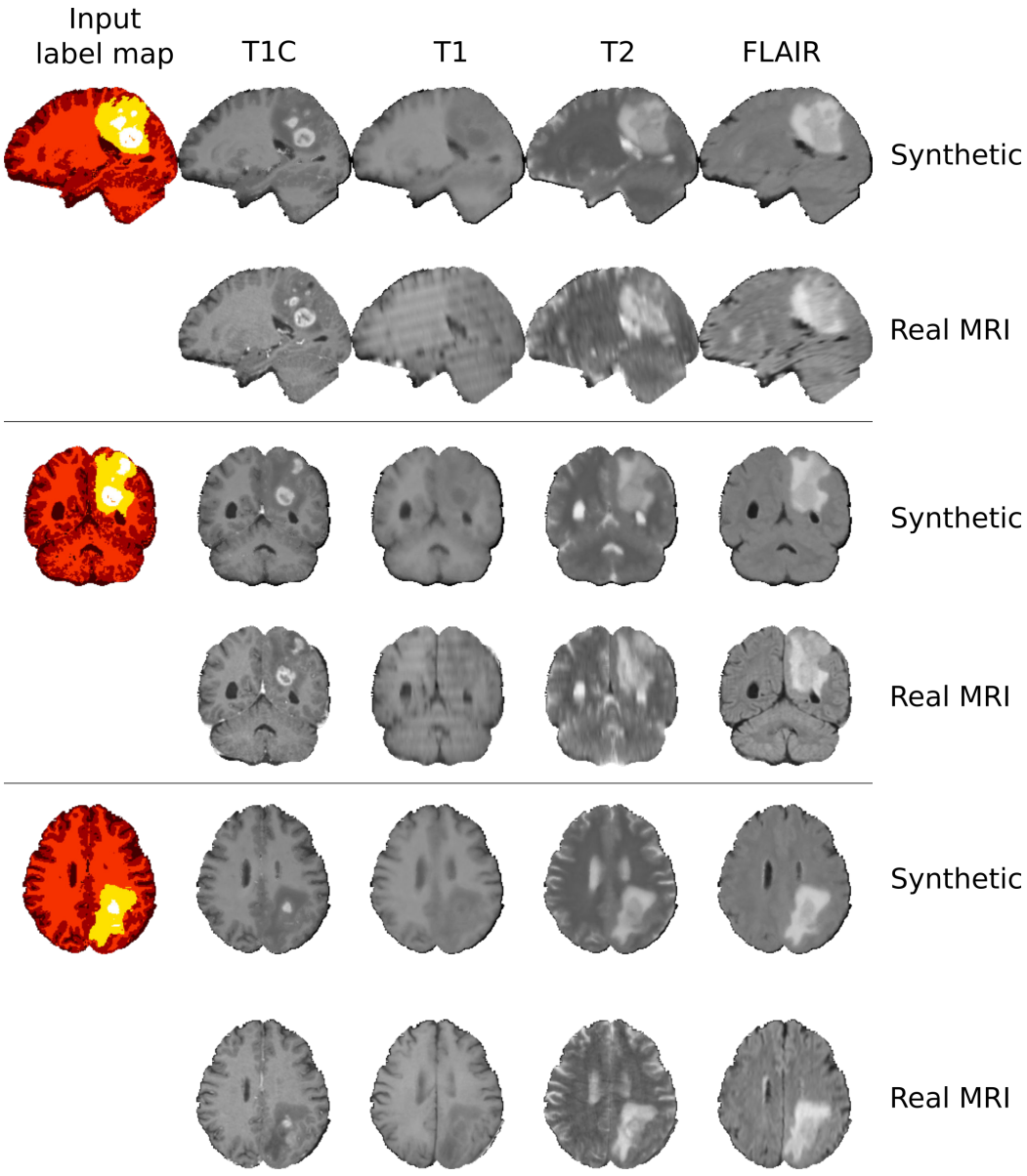

Figure B.2: High-grade HG-0002. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

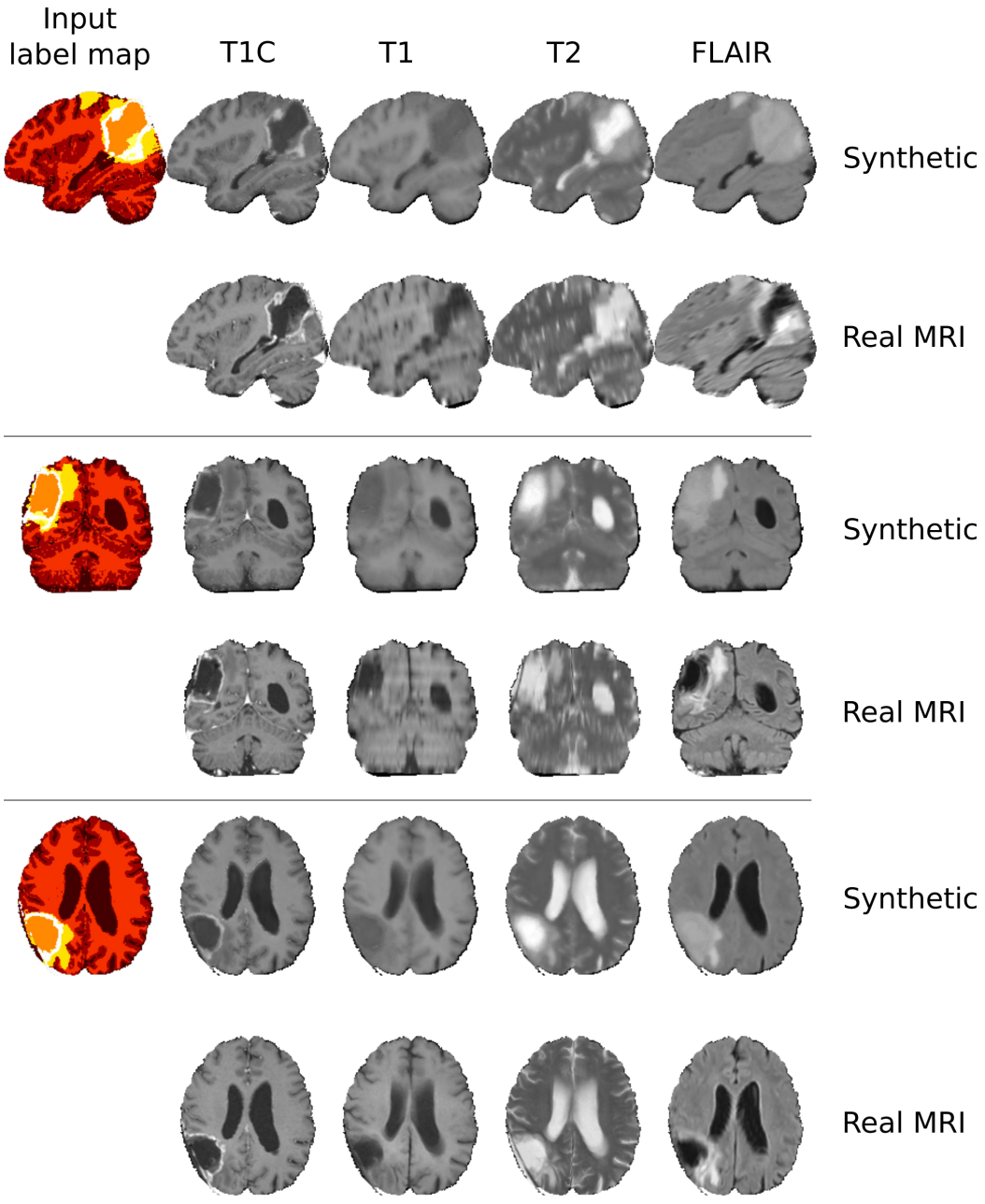

Figure B.3: High-grade HG-0005. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

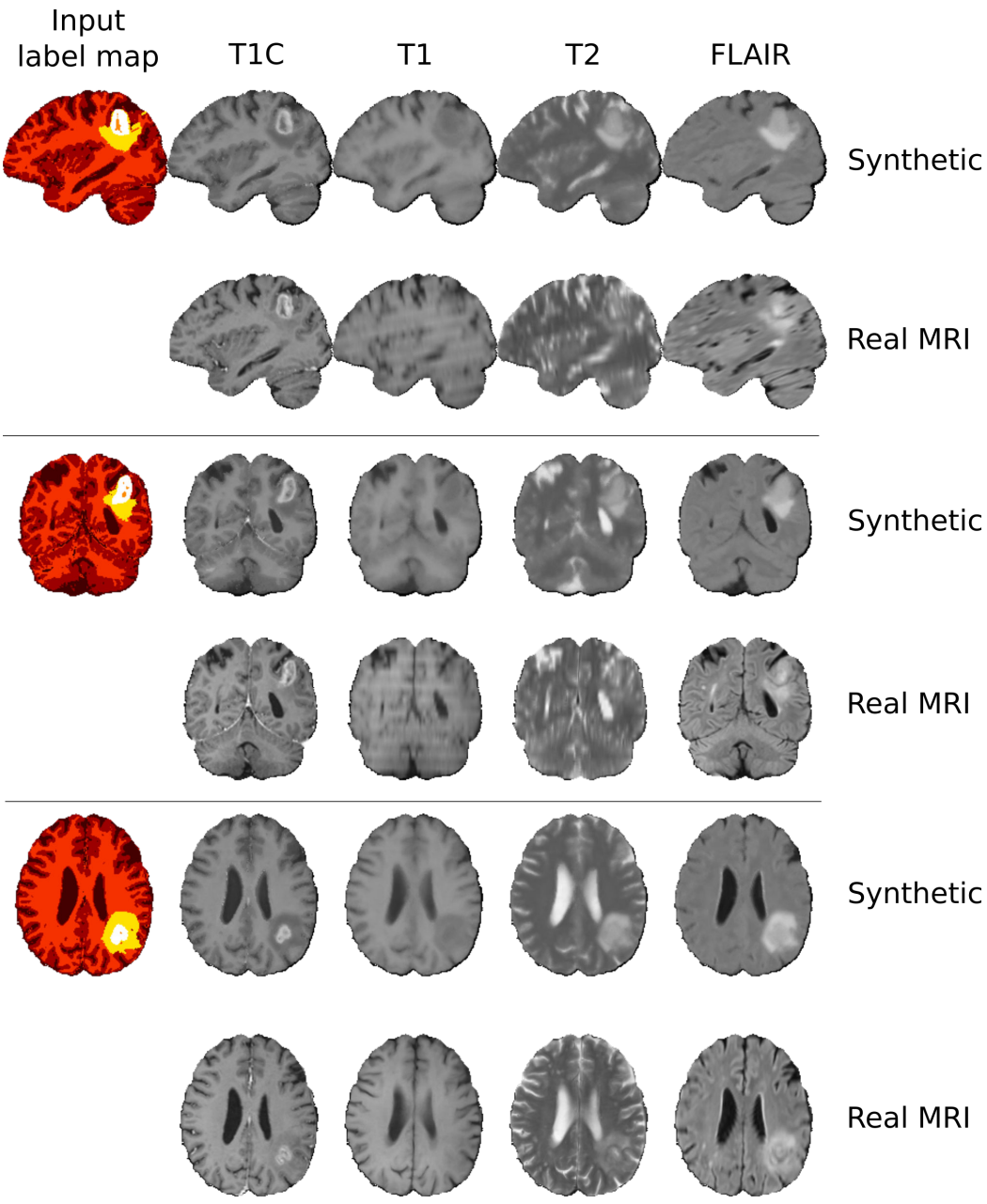

Figure B.4: High-grade HG-0010. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

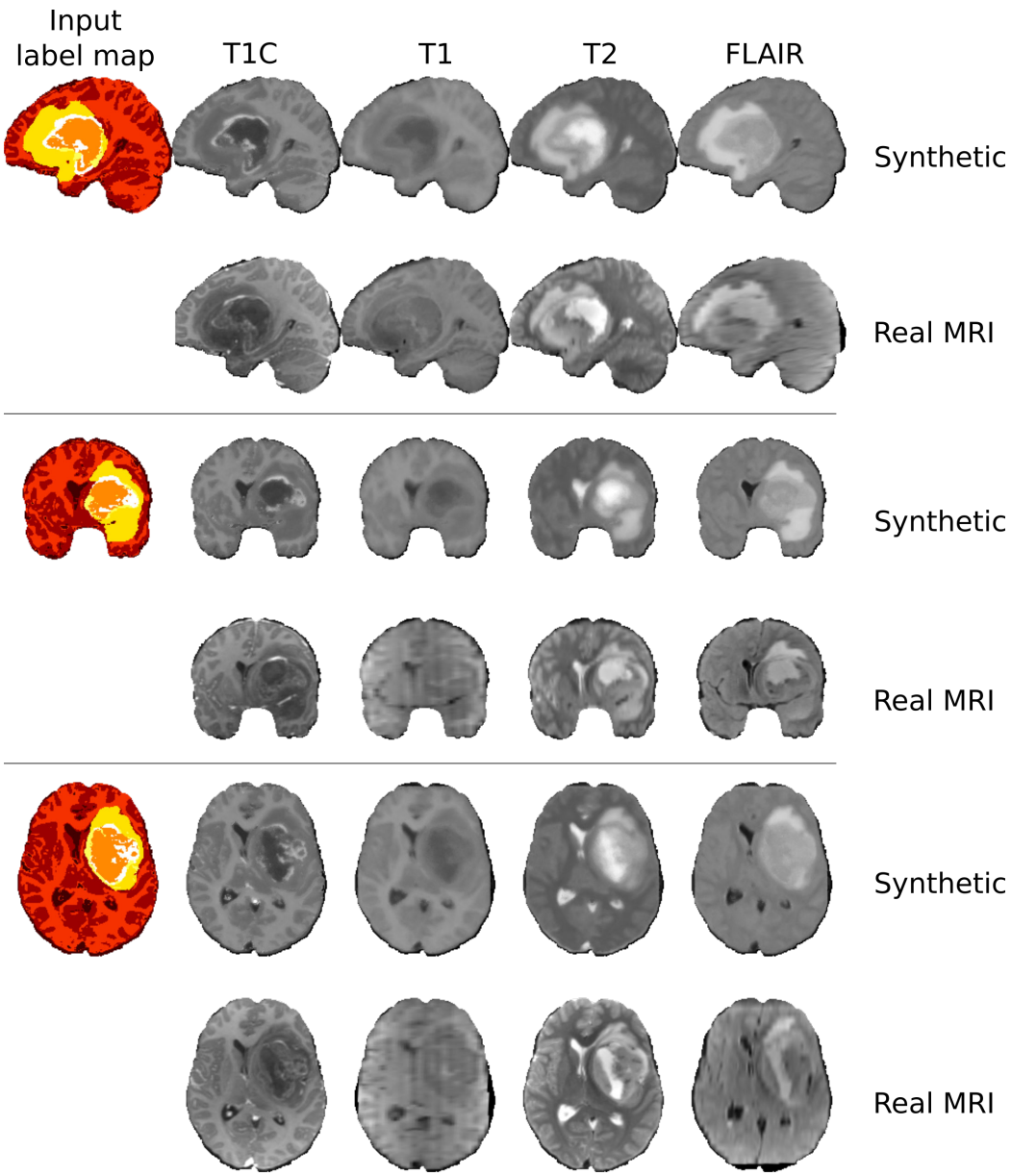

Figure B.5: High-grade HG-0025. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

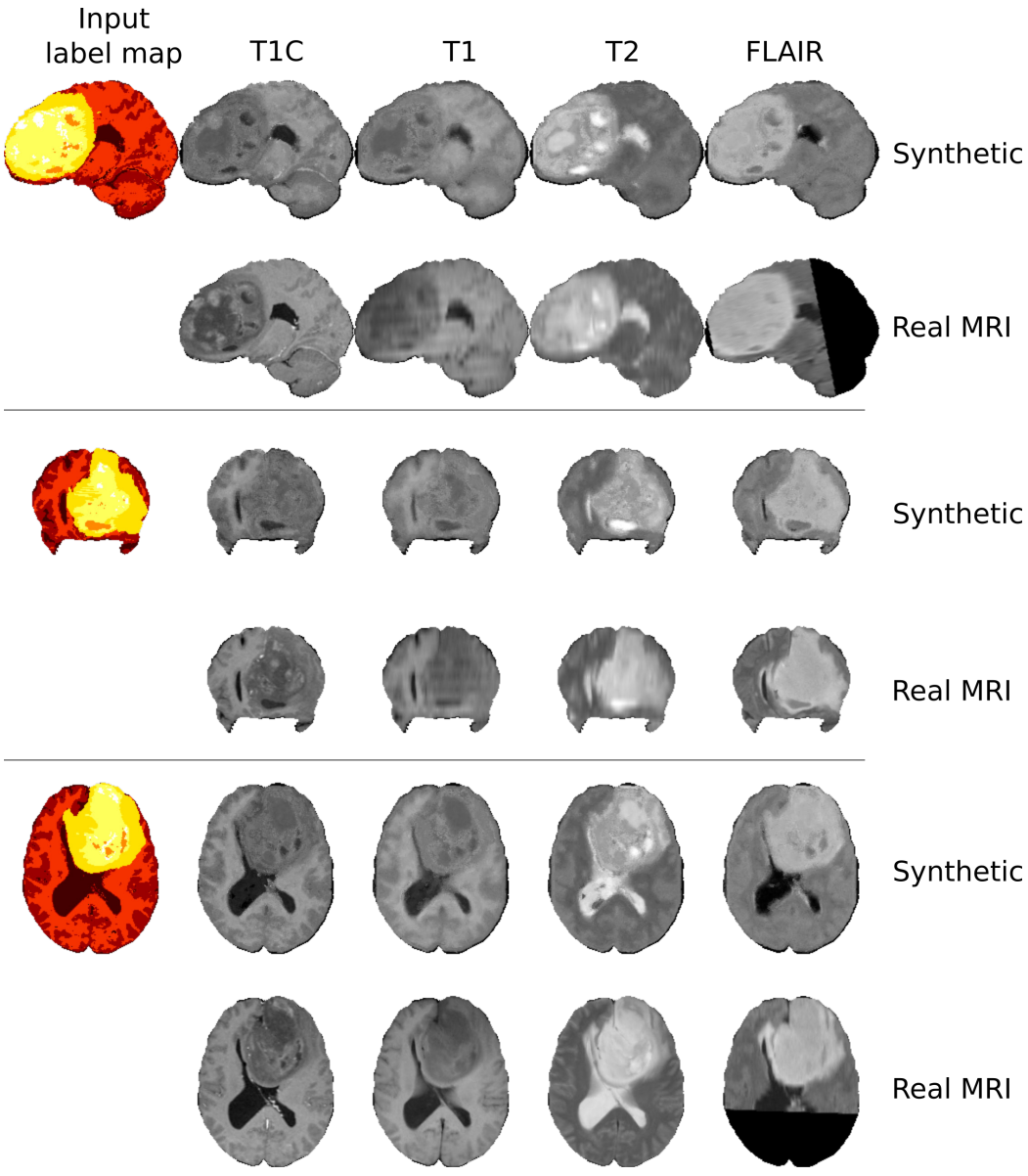

Figure B.6: Low-grade LG-0002. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

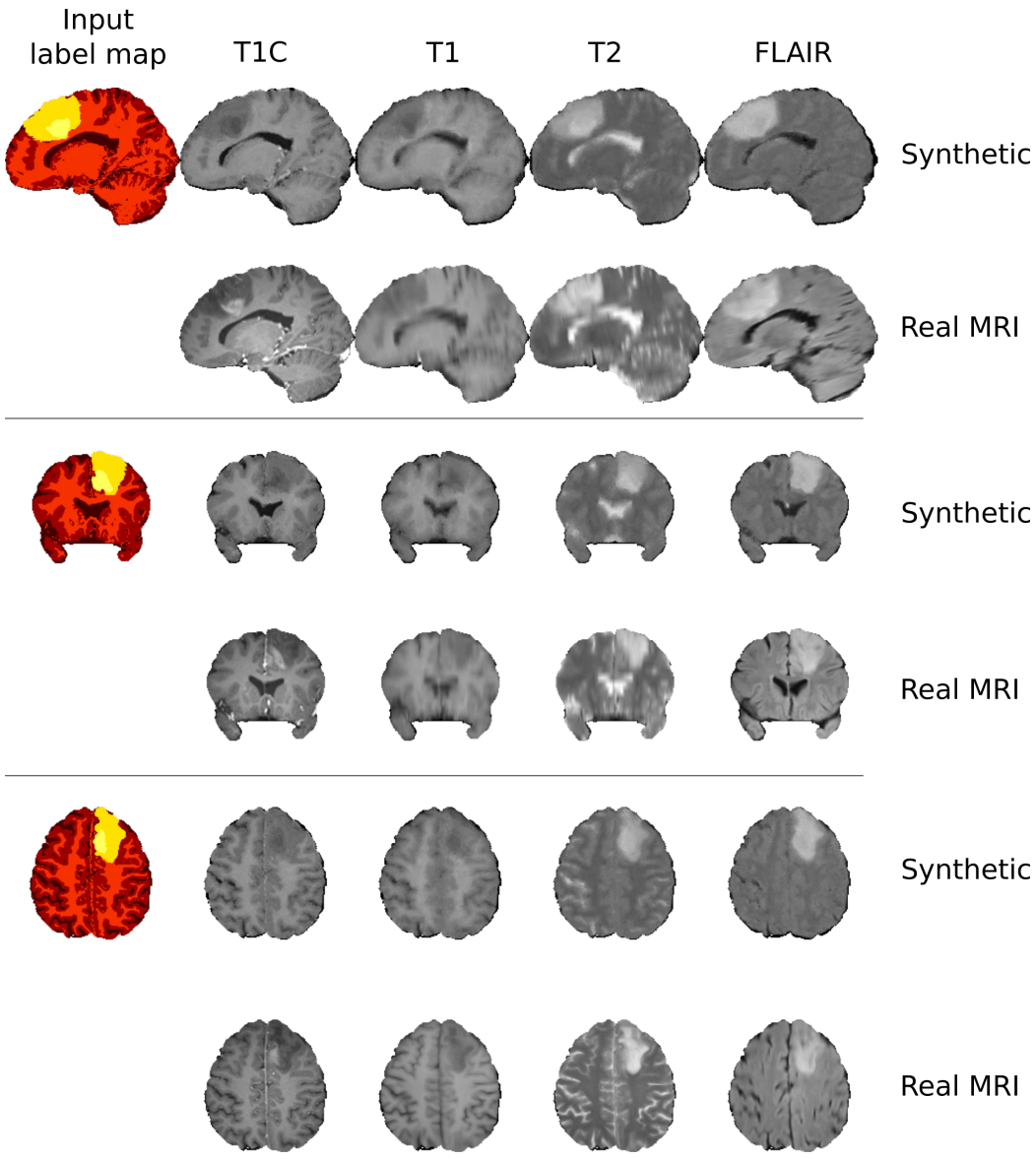

Figure B.7: Low-grade LG-0006. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

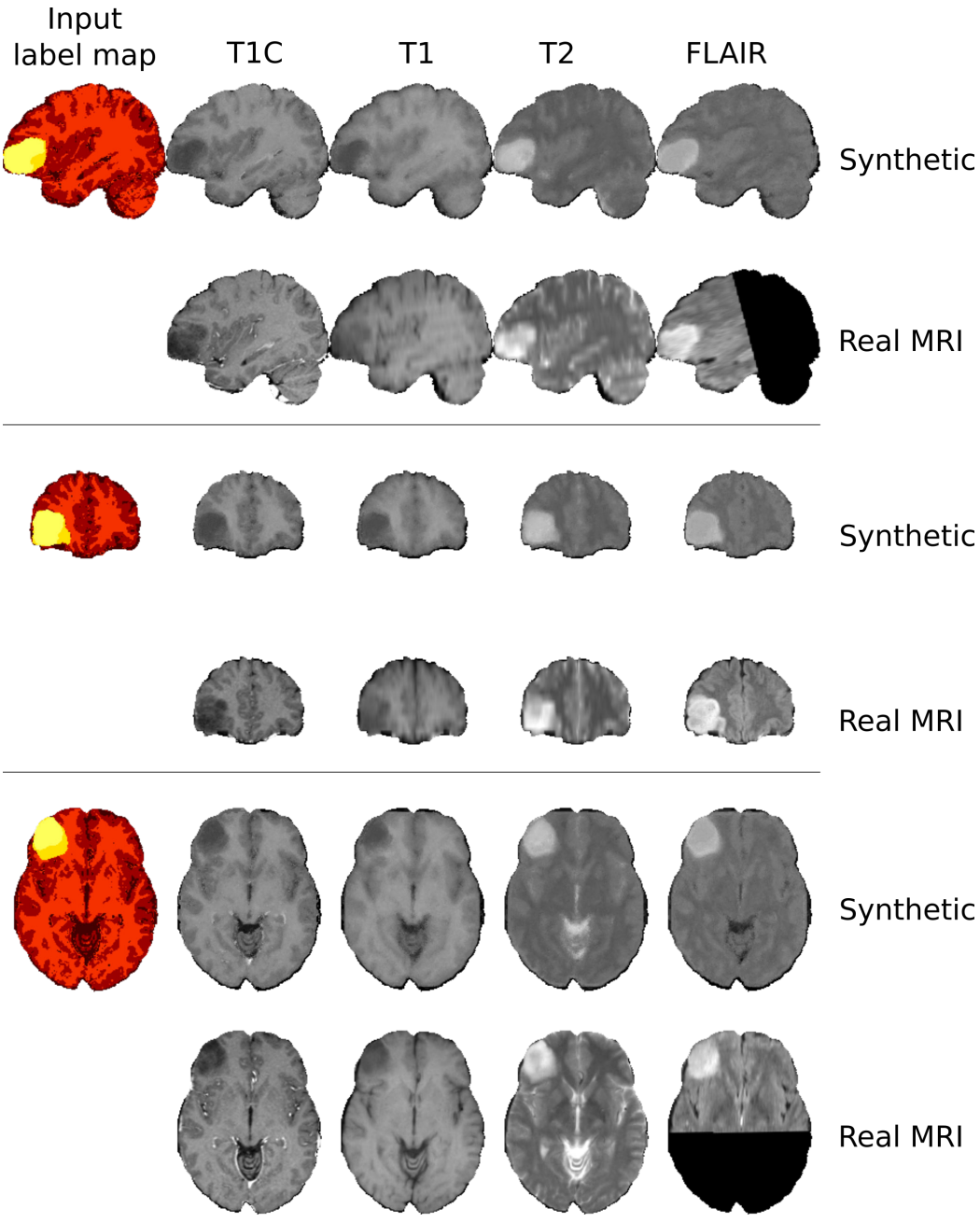

Figure B.8: Low-grade LG-0008. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

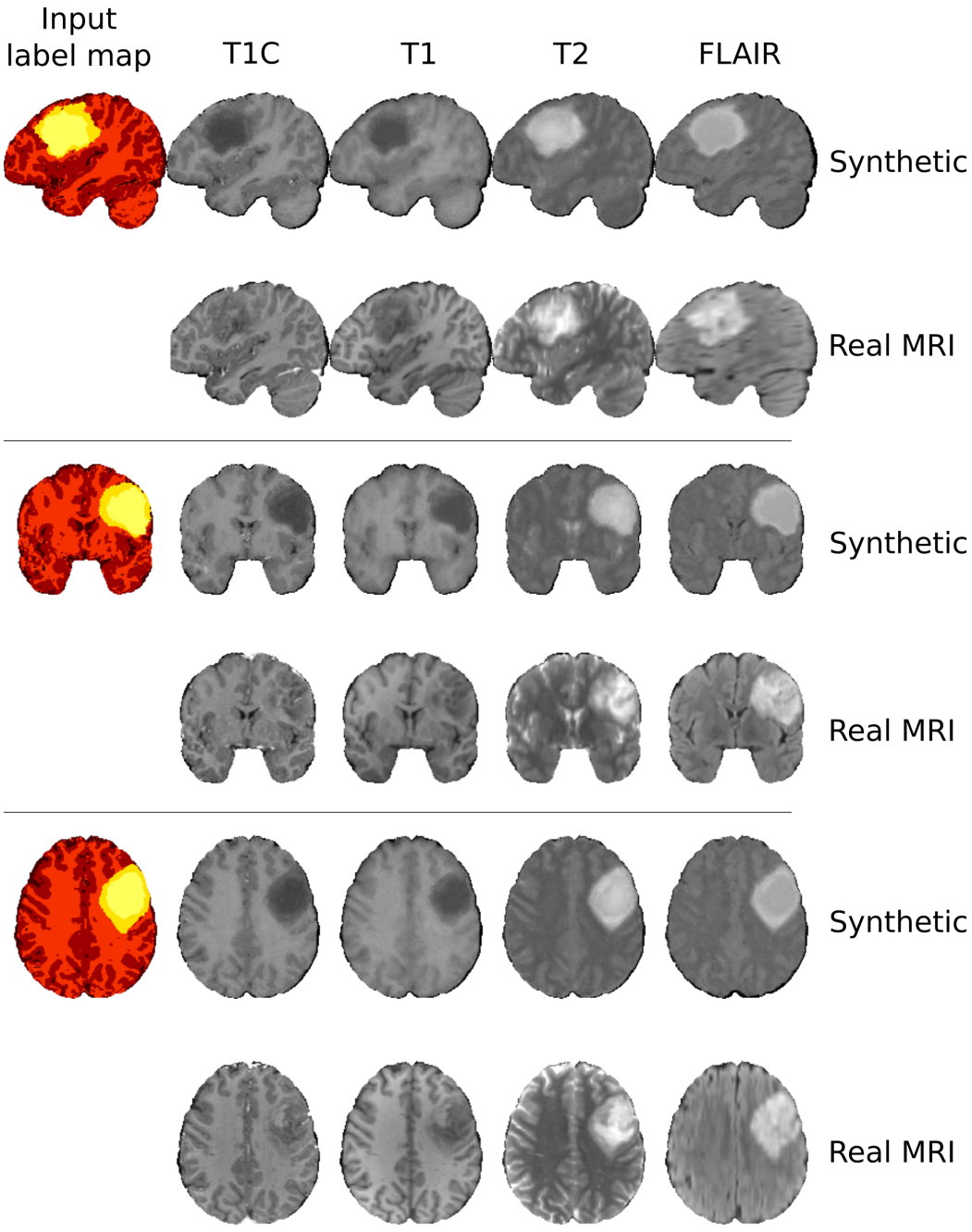

Figure B.9: Low-grade LG-0012. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

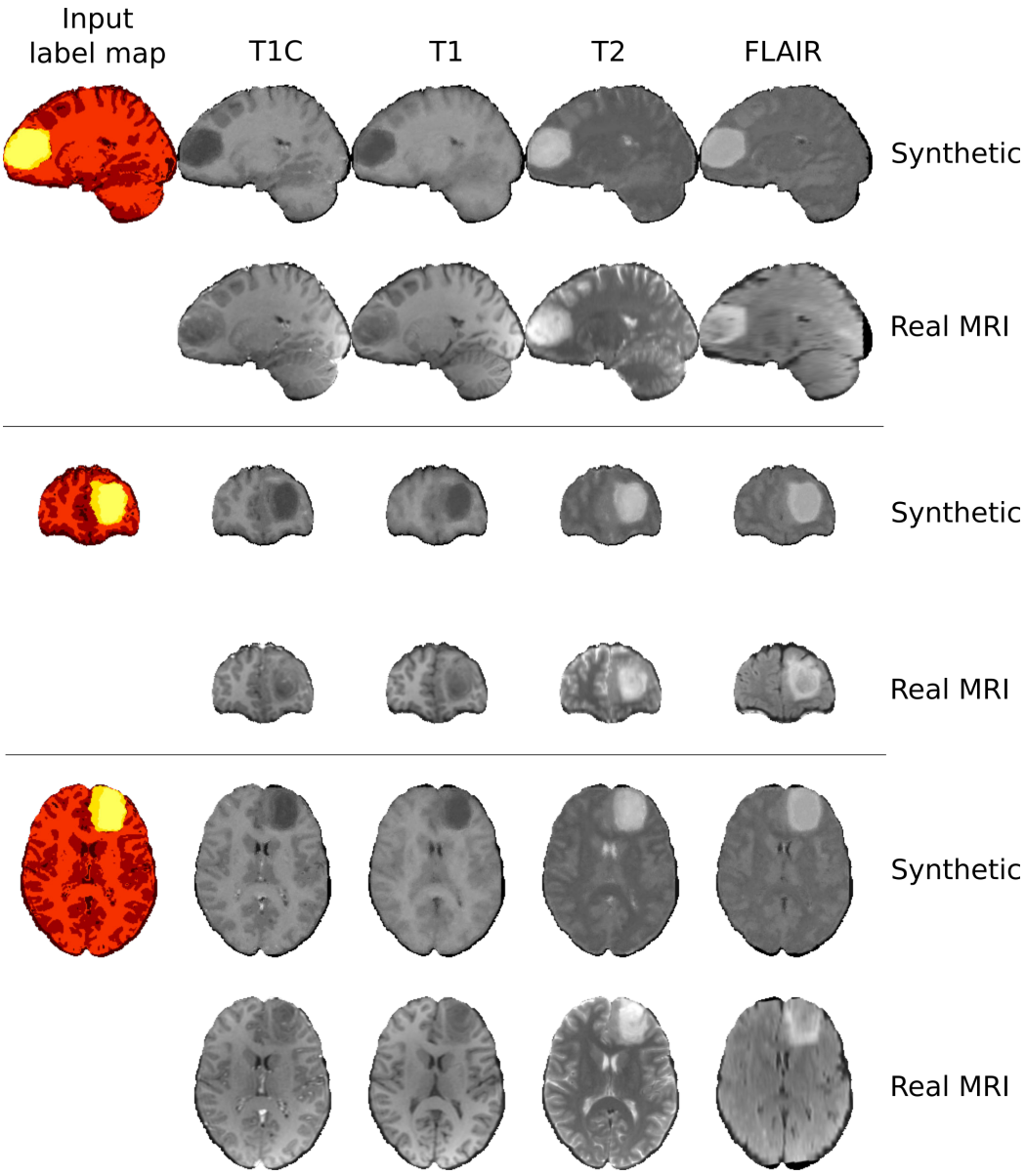

Figure B.10: Low-grade LG-0015. Top to bottom: sagittal, coronal and axial views, for synthetic and real MRI on even and odd rows respectively.

Appendix C

Appendix: distributions of multi-modal image intensity

In this appendix, we show 2D projections of Student distributions fitted to average intensity of pure patches, as described in Chapter 2 and mentioned in Chapter 4. For comparison, we also show 2D projections of Student distributions fitted to voxel-wise image intensity as mentioned in Chapter 4. The use of average intensity of pure patches seems to yield more discriminative Student fits than the use of voxel-wise image intensity.

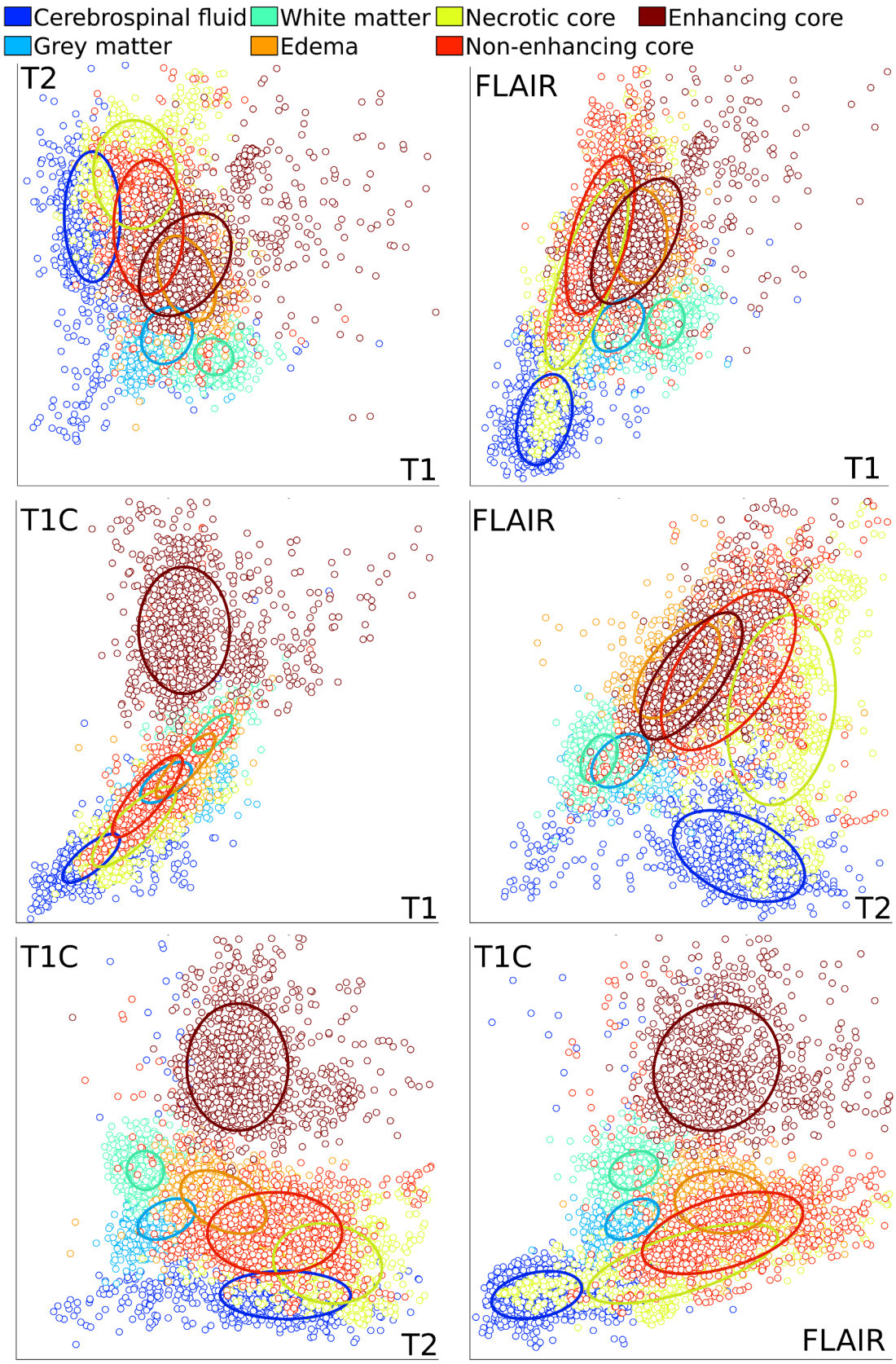

Figure C.1: 2D projections of Student distributions fitted to average intensity of pure patches, with 60% centiles overlayed.

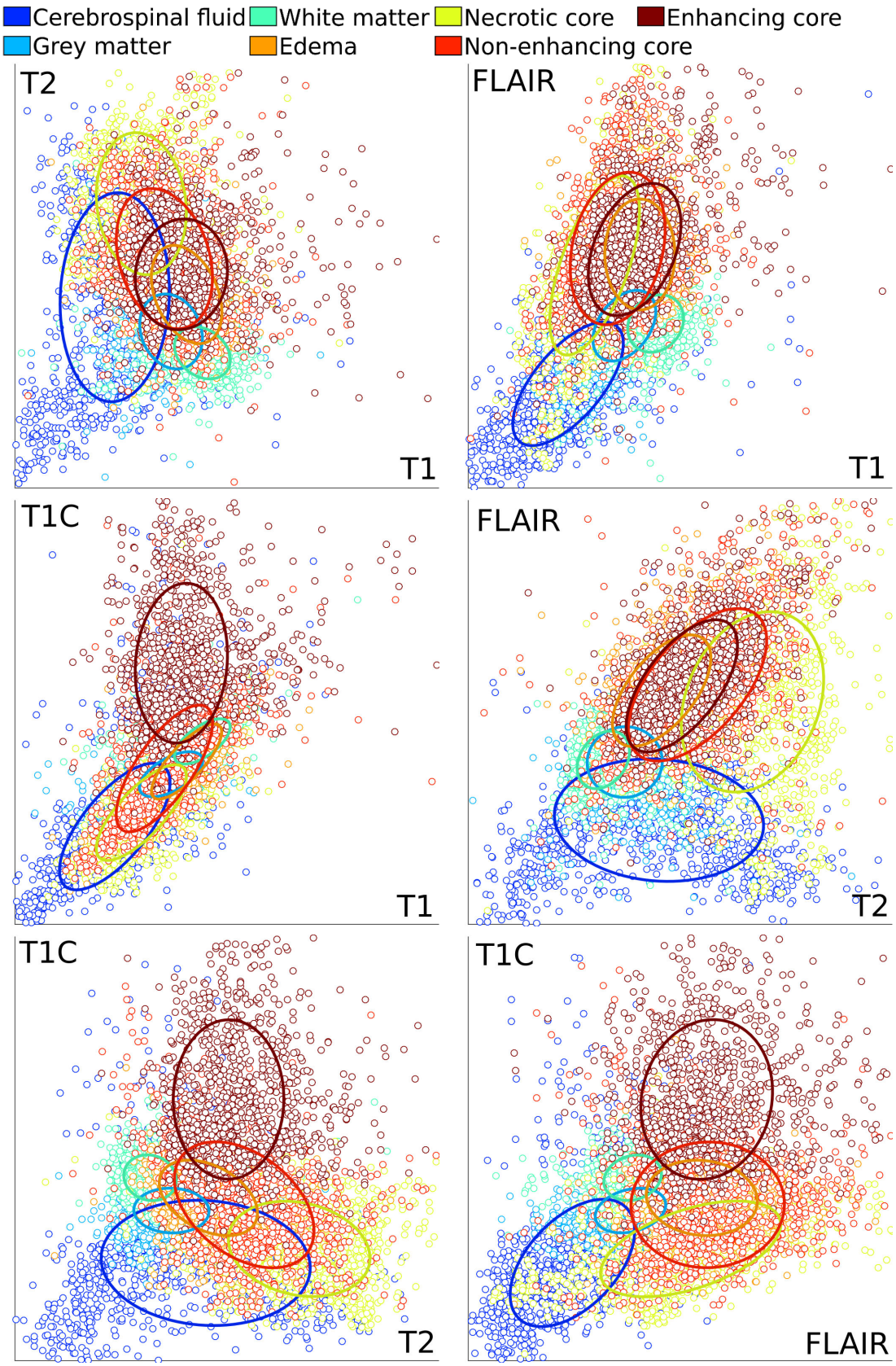

Figure C.2: 2D projections of Student distributions fitted to voxel-wise image intensity, with 60% centiles overlayed.

Bibliography

[Aljabar 2009] Paul Aljabar, Rolf A Heckemann, Alexander Hammers, Joseph V Hajnal and Daniel Rueckert. *Multi-atlas based segmentation of brain images: atlas selection and its effect on accuracy.* Neuroimage, vol. 46, no. 3, pages 726–738, 2009. (Cited on pages 3 and 13.)

[Angelini 2007] Elsa D Angelini, Olivier Clatz, Emmanuel Mandonnet, Ender Konukoglu, Laurent Capelle and Hugues Duffau. *Glioma dynamics and computational models: a review of segmentation, registration, and in silico growth algorithms and their clinical applications.* Current Medical Imaging Reviews, vol. 3, no. 4, pages 262–276, 2007. (Cited on page 11.)

[Asman 2012] Andrew J Asman, Bennett Landman*et al.* *Out-of-atlas labeling: A multi-atlas approach to cancer segmentation.* In Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on, pages 1236–1239. IEEE, 2012. (Cited on page 3.)

[Bai 2013] Wenjia Bai, Wenzhe Shi, Declan P O'Regan, Tong Tong, Haiyan Wang, Shahnaz Jamil-Copley, Nicholas S Peters and Daniel Rueckert. *A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: application to cardiac MR images.* Medical Imaging, IEEE Transactions on, vol. 32, no. 7, pages 1302–1315, 2013. (Cited on pages 22 and 61.)

[Bauer 2012] Stefan Bauer, Christian May, Dimitra Dionysiou, Georgios Stamatakos, Philippe Büchler and Mauricio Reyes. *Multiscale modeling for image analysis of brain tumor studies.* Biomedical Engineering, IEEE Transactions on, vol. 59, no. 1, pages 25–29, 2012. (Cited on page 59.)

[Benoit-Cattin 2005] H Benoit-Cattin, G Collewet, B Belaroussi, H Saint-Jalmes and C Odet. *The SIMRI project: a versatile and interactive MRI simulator.* Journal of Magnetic Resonance, vol. 173, no. 1, pages 97–115, 2005. (Cited on page 57.)

[Bloch 1946] Felix Bloch, WW Hansen and Martin Packard. *Nuclear induction.* Physical review, vol. 70, no. 7-8, pages 460–474, 1946. (Cited on page 58.)

[Burgos 2014] Ninon Burgos, M Jorge Cardoso, Kris Thielemans, Marc Modat, Stefano Pedemonte, John Dickson, Anna Barnes, Rizwan Ahmed, Colin J Mahoney, Jonathan M Schott*et al.* *Attenuation correction synthesis for hybrid PET-MR scanners: application to brain studies.* Medical Imaging, IEEE Transactions on, vol. 33, no. 12, pages 2332–2341, 2014. (Cited on page 58.)

[Cardoso 2013] M Jorge Cardoso, Kelvin Leung, Marc Modat, Shiva Keihaninejad, David Cash, Josephine Barnes, Nick C Fox, Sebastien Ourselin, Alzheimer's Disease Neuroimaging Initiative*et al. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcelation.* Medical image analysis, vol. 17, no. 6, pages 671–684, 2013. (Cited on page 13.)

[Cardoso 2015] M Jorge Cardoso, Carole H Sudre, Marc Modat and Sebastien Ourselin. *Template-Based Multimodal Joint Generative Model of Brain Data.* In Information Processing in Medical Imaging, pages 17–29. Springer International Publishing, 2015. (Cited on pages 2, 58, 66, 70 and 93.)

[Ciresan 2012] Dan Ciresan, Alessandro Giusti, Luca M Gambardella and Jürgen Schmidhuber. *Deep neural networks segment neuronal membranes in electron microscopy images.* In Advances in neural information processing systems, pages 2843–2851, 2012. (Cited on page 68.)

[Clark 2013] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle*et al. The Cancer Imaging Archive (TCIA): maintaining and operating a public information repository.* Journal of digital imaging, vol. 26, no. 6, pages 1045–1057, 2013. (Cited on page 57.)

[Cordier 2013] Nicolas Cordier, Bjoern Menze, Hervé Delingette and Nicholas Ayache. *Patch-based Segmentation of Brain Tissues.* In MICCAI Challenge on Multimodal Brain Tumor Segmentation, pages 6–17, 2013. (Cited on pages 5 and 33.)

[Cordier 201X] Nicolas Cordier, Hervé Delingette and Nicholas Ayache. *A patch-based approach for the segmentation of pathologies: Application to glioma labelling.* Medical Imaging, IEEE Transactions on, 201X. (Cited on page 5.)

[Cordier 201Y] Nicolas Cordier, Hervé Delingette and Nicholas Ayache. *Extended Modality Propagation: Image Synthesis of Pathological Cases.* Medical Image Analysis, 201Y. (Cited on page 5.)

[Coupé 2011] Pierrick Coupé, José V Manjón, Vladimir Fonov, Jens Pruessner, Montserrat Robles and D Louis Collins. *Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation.* NeuroImage, vol. 54, no. 2, pages 940–954, 2011. (Cited on pages 3, 13 and 58.)

[Duffau 2004] Hugues Duffau and Laurent Capelle. *Preferential brain locations of low-grade gliomas.* Cancer, vol. 100, no. 12, pages 2622–2626, 2004. (Cited on page 20.)

[Eisenhauer 2009] EA1 Eisenhauer, Patrick Therasse, Jan Bogaerts, LH Schwartz, D Sargent, Robert Ford, J Dancey, S Arbuck, S Gwyther, M Mooney*et al.*

New response evaluation criteria in solid tumours: revised RECIST guideline (version 1.1). European Journal of Cancer, vol. 45, no. 2, pages 228–247, 2009. (Cited on page 11.)

[Galimzianova 2015] Alfiia Galimzianova, Franjo Pernus, Bostjan Likar and Ziga Spiclin. *Robust estimation of unbalanced mixture models on samples with outliers*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. PP, no. 99, 2015. (Cited on page 59.)

[Geremia 2013] Ezequiel Geremia, Bjoern H Menze, Marcel Prastawa, M-A Weber, Antonio Criminisi and Nicholas Ayache. *Brain tumor cell density estimation from multi-modal MR images based on a synthetic tumor growth model*. In Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging, pages 273–282. Springer, 2013. (Cited on pages 2, 59, 89 and 94.)

[Glatard 2013] Tristan Glatard, Carole Lartizien, Bernard Gibaud, Rafael Ferreira da Silva, Germain Forestier, Frédéric Cervenansky, Martino Alessandrini, Hugues Benoit-Cattin, Olivier Bernard, Sorina Camarasu-Pop*et al. A Virtual Imaging Platform for multi-modality medical image simulation*. Medical Imaging, IEEE Transactions on, vol. 32, no. 1, pages 110–118, 2013. (Cited on page 57.)

[Gómez 2015] Pedro A Gómez, Jonathan I Sperl, Tim Sprenger, Claudia Metzler-Baddeley, Derek K Jones and Philipp Saemann. *Joint Reconstruction of Multi-Contrast MRI for Multiple Sclerosis Lesion Segmentation*. In Bildverarbeitung für die Medizin 2015, pages 155–160. Springer, 2015. (Cited on page 70.)

[Gu 2012] Stanley Gu, Gargi Chakraborty, Kyle Champley, Adam M Alessio, Jonathan Claridge, Russell Rockne, Mark Muzi, Kenneth A Krohn, Alexander M Spence, Ellsworth C Alvord*et al. Applying a patient-specific bio-mathematical model of glioma growth to develop virtual [18F]-FMISO-PET images*. Mathematical Medicine and Biology, vol. 29, no. 1, pages 31–48, 2012. (Cited on page 59.)

[Hamamci 2012] Andac Hamamci, Nadir Kucuk, Kutlay Karaman, Kayihan Engin and Gozde Unal. *Tumor-cut: segmentation of brain tumors on contrast enhanced MR images for radiosurgery applications*. Medical Imaging, IEEE Transactions on, vol. 31, no. 3, pages 790–804, 2012. (Cited on page 59.)

[Havaei 2015] Mohammad Havaei, Axel Davy, David Warde-Farley, Antoine Biard, Aaron Courville, Yoshua Bengio, Chris Pal, Pierre-Marc Jodoin and Hugo Larochelle. *Brain Tumor Segmentation with Deep Neural Networks*. arXiv preprint arXiv:1505.03540, 2015. (Cited on page 57.)

[Heckemann 2006] Rolf A Heckemann, Joseph V Hajnal, Paul Aljabar, Daniel Rueckert and Alexander Hammers. *Automatic anatomical brain MRI segmentation combining label propagation and decision fusion.* NeuroImage, vol. 33, no. 1, pages 115–126, 2006. (Cited on pages 3 and 13.)

[Iglesias 2010] Juan Iglesias, Ivo Dinov, Jaskaran Singh, Gregory Tong and Zhuowen Tu. *Synthetic MRI signal standardization: application to multi-atlas analysis.* MICCAI 2010, pages 81–88, 2010. (Cited on page 58.)

[Iglesias 2013] Juan Eugenio Iglesias, Ender Konukoglu, Darko Zikic, Ben Glocker, Koen Van Leemput and Bruce Fischl. *Is synthesizing MRI contrast useful for inter-modality analysis?* In MICCAI 2013, pages 631–638. Springer, 2013. (Cited on pages 2 and 58.)

[Iglesias 2015] Juan Eugenio Iglesias and Mert Rory Sabuncu. *Multi-Atlas Segmentation of Biomedical Images: A Survey.* Medical Image Analysis, 2015. (Cited on pages 3, 13, 22 and 61.)

[Jack 2008] Clifford R Jack, Matt A Bernstein, Nick C Fox, Paul Thompson, Gene Alexander, Danielle Harvey, Bret Borowski, Paula J Britson, Jennifer L Whitwell, Chadwick Ward*et al.* *The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods.* Journal of Magnetic Resonance Imaging, vol. 27, no. 4, pages 685–691, 2008. (Cited on page 57.)

[Konukoglu 2013] Ender Konukoglu, Andre van der Kouwe, Mert Rory Sabuncu and Bruce Fischl. *Example-based restoration of high-resolution magnetic resonance image acquisitions.* In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, pages 131–138. Springer, 2013. (Cited on page 70.)

[Kwan 1999] Remi KS Kwan, Alan C Evans and G Bruce Pike. *MRI simulation-based evaluation of image-processing and classification methods.* Medical Imaging, IEEE Transactions on, vol. 18, no. 11, pages 1085–1097, 1999. (Cited on pages 57, 58 and 59.)

[Kwon 2014] Dongjin Kwon, Russell T Shinohara, Hamed Akbari and Christos Davatzikos. *Combining generative models for multifocal glioma segmentation and registration.* In MICCAI 2014, pages 763–770. Springer, 2014. (Cited on pages 12 and 27.)

[Larjavaara 2007] Suvi Larjavaara, Riitta Mäntylä, Tiina Salminen, Hannu Haapasalo, Jani Raitanen, Juha Jääskeläinen and Anssi Auvinen. *Incidence of gliomas by anatomic location.* Neuro-oncology, vol. 9, no. 3, pages 319–325, 2007. (Cited on page 20.)

[Liu 1995] Chuanhai Liu and Donald B Rubin. *ML estimation of the t distribution using EM and its extensions, ECM and ECME.* Statistica Sinica, vol. 5, no. 1, pages 19–39, 1995. (Cited on pages 16 and 62.)

[Liu 2015] Xiaoxiao Liu, Marc Niethammer, Roland Kwitt, Nikhil Singh, Matt McCormick and Stephen Aylward. *Low-rank Atlas Image Analyses in the Presence of Pathologies*. Medical Imaging, IEEE Transactions on, vol. PP, no. 99, 2015. (Cited on page 59.)

[Louis 2007] David N Louis, Hiroko Ohgaki, Otmar D Wiestler, Webster K Cavenee, Peter C Burger, Anne Jouvet, Bernd W Scheithauer and Paul Kleihues. *The 2007 WHO classification of tumours of the central nervous system*. Acta neuropathologica, vol. 114, no. 2, pages 97–109, 2007. (Cited on page 1.)

[Maitra 2010] Ranjan Maitra and John J Riddles. *Synthetic magnetic resonance imaging revisited*. Medical Imaging, IEEE Transactions on, vol. 29, no. 3, pages 895–902, 2010. (Cited on page 58.)

[Mandonnet 2013] E Mandonnet, S Wait, L Choi and C Teo. *The importance of measuring the velocity of diameter expansion on MRI in upfront management of suspected WHO grade II glioma–Case report*. Neurochirurgie, vol. 59, no. 2, pages 89–92, 2013. (Cited on pages 1 and 11.)

[Manjón 2010] José V Manjón, Pierrick Coupé, Antonio Buades, Vladimir Fonov, D Louis Collins and Montserrat Robles. *Non-local MRI upsampling*. Medical image analysis, vol. 14, no. 6, pages 784–792, 2010. (Cited on page 70.)

[Mazziotta 2001] John Mazziotta, Arthur Toga, Alan Evans, Peter Fox, Jack Lancaster, Karl Zilles, Roger Woods, Tomas Paus, Gregory Simpson, Bruce Pike*et al.* *A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)*. Philosophical Transactions of the Royal Society B: Biological Sciences, vol. 356, no. 1412, pages 1293–1322, 2001. (Cited on pages 3, 16, 68 and 87.)

[Menze 2010] Bjoern H Menze, Koen Van Leemput, Danial Lashkari, Marc-André Weber, Nicholas Ayache and Polina Golland. *A generative model for brain tumor segmentation in multi-modal images*. In MICCAI 2010, pages 151–159. Springer, 2010. (Cited on page 12.)

[Menze 2014] B. Menze, M. Reyes, K. Van Leemput*et al.* *The Multimodal Brain Tumor Image Segmentation Benchmark (BraTS)*. Medical Imaging, IEEE Transactions on, vol. PP, no. 99, pages 1–33, December 2014. (Cited on pages 1, 2, 11, 12, 25, 27, 57, 59 and 70.)

[Muja 2014] Marius Muja and David G. Lowe. *Scalable Nearest Neighbor Algorithms for High Dimensional Data*. Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 36, 2014. (Cited on pages 27 and 69.)

[Parisot 2011] Sarah Parisot, Hugues Duffau, Stéphane Chemouny and Nikos Paragios. *Graph based spatial position mapping of low-grade gliomas*. In MICCAI 2011, pages 508–515. Springer, 2011. (Cited on page 20.)

[Prastawa 2005] Marcel Prastawa, Elizabeth Bullitt and Guido Gerig. *Synthetic ground truth for validation of brain tumor MRI segmentation.* In MICCAI 2005, pages 26–33. Springer, 2005. (Cited on page 59.)

[Prastawa 2009] Marcel Prastawa, Elizabeth Bullitt and Guido Gerig. *Simulation of brain tumors in MR images for evaluation of segmentation efficacy.* Medical Image Analysis, vol. 13, no. 2, page 297, 2009. (Cited on pages 2, 59, 77, 80, 81, 82, 83, 84 and 85.)

[Regan 2011] Elizabeth A Regan, John E Hokanson, James R Murphy, Barry Make, David A Lynch, Terri H Beaty, Douglas Curran-Everett, Edwin K Silverman and James D Crapo. *Genetic epidemiology of COPD (COPDGene) study design.* COPD: Journal of Chronic Obstructive Pulmonary Disease, vol. 7, no. 1, pages 32–43, 2011. (Cited on page 57.)

[Rexilius 2004] Jan Rexilius, Horst K Hahn, Mathias Schlüter, Sven Kohle, Holger Bourquain, Joachim Böttcher and Heinz-Otto Peitgen. *A framework for the generation of realistic brain tumor phantoms and applications.* In MICCAI 2004, pages 243–250. Springer, 2004. (Cited on page 59.)

[Rohlfing 2003] Torsten Rohlfing, Daniel B Russakoff and Calvin R Maurer Jr. *Extraction and application of expert priors to combine multiple segmentations of human brain tissue.* In MICCAI 2003, pages 578–585. Springer, 2003. (Cited on page 13.)

[Romero 2015] José E Romero, José V Manjón, Jussi Tohka, Pierrick Coupé and Montserrat Robles. *NABS: Non-local Automatic Brain Hemisphere Segmentation.* Magnetic resonance imaging, 2015. (Cited on page 13.)

[Rousseau 2011] François Rousseau, Piotr A Habas and Colin Studholme. *A supervised patch-based approach for human brain labeling.* Medical Imaging, IEEE Transactions on, vol. 30, no. 10, pages 1852–1862, 2011. (Cited on pages 3, 13, 58 and 93.)

[Roy 2013] Sandip Roy, Aaron Carass and Jerry L Prince. *Magnetic Resonance Image Example-Based Contrast Synthesis.* Medical Imaging, IEEE Transactions on, vol. 32, no. 12, pages 2348–2363, 2013. (Cited on page 58.)

[Sabuncu 2010] Mert R Sabuncu, BT Thomas Yeo, Koen Van Leemput, Bruce Fischl and Polina Golland. *A generative model for image segmentation based on label fusion.* Medical Imaging, IEEE Transactions on, vol. 29, no. 10, pages 1714–1729, 2010. (Cited on pages 22, 23, 61, 62, 66 and 68.)

[Shiee 2011] Navid Shiee, Pierre-Louis Bazin, Jennifer L Cuzzocreo, Ari Blitz and Dzung L Pham. *Segmentation of brain images using adaptive atlases with application to ventriculomegaly.* In Information Processing in Medical Imaging, pages 1–12. Springer, 2011. (Cited on page 15.)

[Svensén 2005] Markus Svensén and Christopher M Bishop. *Robust Bayesian mixture modelling.* Neurocomputing, vol. 64, pages 235–252, 2005. (Cited on page 15.)

[Swanson 2011] Kristin R Swanson, Russell C Rockne, Jonathan Claridge, Mark A Chaplain, Ellsworth C Alvord and Alexander RA Anderson. *Quantifying the role of angiogenesis in malignant progression of gliomas: in silico modeling integrates imaging and histology.* Cancer Research, vol. 71, no. 24, pages 7366–7375, 2011. (Cited on page 77.)

[Toews 2015] Matthew Toews, Christian Wachinger, Raul San Jose Estepar and William M Wells III. *A Feature-based Approach to Big Data Analysis of Medical Images.* In Information Processing in Medical Imaging. Springer Berlin Heidelberg, 2015. (Cited on page 57.)

[Tu 2010] Zhuowen Tu and Xiang Bai. *Auto-context and its application to high-level vision tasks and 3d brain image segmentation.* Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 32, no. 10, pages 1744–1757, 2010. (Cited on page 92.)

[Tustison 2014] Nicholas J Tustison, KL Shrinidhi, Max Wintermark, Christopher R Durst, Benjamin M Kandel, James C Gee, Murray C Grossman and Brian B Avants. *Optimal Symmetric Multimodal Templates and Concatenated Random Forests for Supervised Brain Tumor Segmentation (Simplified) with ANTsR.* Neuroinformatics, pages 1–17, 2014. (Cited on pages 12 and 27.)

[van Tulder 2015] Gijs van Tulder and Marleen de Bruijne. *Why Does Synthesized Data Improve Multi-sequence Classification?* In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015. Springer, 2015. (Cited on pages 2 and 94.)

[Wachinger 2014a] Christian Wachinger, Matthew Brennan, Greg C Sharp and Polina Golland. *On the Importance of Location and Features for the Patch-Based Segmentation of Parotid Glands.* In MICCAI Workshop on Image-Guided Adaptive Radiation Therapy. Springer, 2014. (Cited on pages 19 and 20.)

[Wachinger 2014b] Christian Wachinger and Polina Golland. *Atlas-Based Under-Segmentation.* In MICCAI 2014, pages 315–322. Springer, 2014. (Cited on page 26.)

[Wachinger 2015] Christian Wachinger, Polina Golland, William Kremen, Bruce Fischl, Martin Reuter, Alzheimer's Disease Neuroimaging Initiative*et al. BrainPrint: A discriminative characterization of brain morphology.* NeuroImage, vol. 109, pages 232–248, 2015. (Cited on page 57.)

[Wang 2004] Zhou Wang, Alan C Bovik, Hamid R Sheikh and Eero P Simoncelli. *Image quality assessment: from error visibility to structural similarity.* Image Processing, IEEE Transactions on, vol. 13, no. 4, pages 600–612, 2004. (Cited on page 70.)

[Wang 2013a] Hongzhi Wang and Paul A Yushkevich. *Multi-atlas Segmentation without Registration: A Supervoxel-Based Approach.* In MICCAI 2013, pages 535–542. Springer, 2013. (Cited on page 4.)

[Wang 2013b] Zehan Wang, Claire Donoghue and Daniel Rueckert. *Patch-based segmentation without registration: application to knee MRI.* In Machine Learning in Medical Imaging, pages 98–105. Springer, 2013. (Cited on page 4.)

[Wang 2014] Zehan Wang, Kanwal K Bhatia, Ben Glocker, Antonio Marvao, Tim Dawes, Kazunari Misawa, Kensaku Mori and Daniel Rueckert. *Geodesic patch-based segmentation.* In MICCAI 2014, pages 666–673. Springer International Publishing, 2014. (Cited on page 4.)

[Wen 2010] Patrick Y Wen, David R Macdonald, David A Reardon, Timothy F Cloughesy, A Gregory Sorensen, Evanthia Galanis, John DeGroot, Wolfgang Wick, Mark R Gilbert, Andrew B Lassman*et al. Updated response assessment criteria for high-grade gliomas: response assessment in neuro-oncology working group.* Journal of Clinical Oncology, vol. 28, no. 11, pages 1963–1972, 2010. (Cited on page 11.)

[Ye 2013] Dong Hye Ye, Darko Zikic, Ben Glocker, Antonio Criminisi and Ender Konukoglu. *Modality Propagation: coherent synthesis of subject-specific scans with data-driven regularization.* In MICCAI 2013, pages 606–613. Springer, 2013. (Cited on pages 2, 3, 35, 58, 66, 68 and 87.)

[Zhang 2001] Yongyue Zhang, Michael Brady and Stephen Smith. *Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm.* Medical Imaging, IEEE Transactions on, vol. 20, no. 1, pages 45–57, 2001. (Cited on pages 26 and 70.)

[Zikic 2012] Darko Zikic, Ben Glocker, Ender Konukoglu, Antonio Criminisi, C Demiralp, Jamie Shotton, OM Thomas, T Das, R Jena and SJ Price. *Decision forests for tissue-specific segmentation of high-grade gliomas in multi-channel MR.* In MICCAI 2012, pages 369–376. Springer, 2012. (Cited on page 12.)

Multi-Atlas Patch-Based Segmentation and Synthesis of Brain Tumor MR Images

Abstract:

This thesis focuses on the development of automatic methods for the segmentation and synthesis of brain tumor Magnetic Resonance images. The main clinical perspective of glioma segmentation is growth velocity monitoring for patient therapy management.

To this end, the thesis builds on the formalization of multi-atlas patch-based segmentation with probabilistic graphical models.

A probabilistic model first extends classical multi-atlas approaches used for the segmentation of healthy brains structures to the automatic segmentation of pathological cerebral regions. An approximation of the marginalization step replaces the concept of local search windows with a stratification with respect to both atlases and labels. A glioma detection model based on a spatially-varying prior and patch pre-selection criteria are introduced to obtain competitive running times despite patch matching being non local. This work is validated and compared to state-of-the-art algorithms on publicly available datasets.

A second probabilistic model mirrors the segmentation model in order to synthesize realistic MRI of pathological cases, based on a single label map. A heuristic method allows to solve for the maximum a posteriori and to estimate uncertainty of the image synthesis model. Iterating patch matching reinforces the spatial coherence of synthetic images. The realism of our synthetic images is assessed against real MRI, and against outputs of the state-of-the-art method. The junction of a tumor growth model to the proposed synthesis approach allows to generate databases of annotated synthetic cases.

Keywords: patch-based, multi-atlas, glioma, segmentation, probabilistic generative model, medical image simulation, modality synthesis.

Approches Multi-Atlas fondées sur l'Appariement de Blocs de Voxels pour la Segmentation et la Synthèse d'Images par Résonance Magnétique de Tumeurs Cérébrales

Résumé :

Cette thèse s'intéresse au développement de méthodes automatiques pour la segmentation et la synthèse d'images par résonance magnétique de tumeurs cérébrales. La principale perspective clinique de la segmentation des gliomes est le suivi de la vitesse d'expansion diamétrique dans le but d'adapter les solutions thérapeutiques. A cette fin, la thèse formalise au moyen de modèles graphiques probabilistes des approches de segmentation multi-atlas fondées sur l'appariement de blocs de voxels. Un premier modèle probabiliste prolonge à la segmentation automatique de régions cérébrales pathologiques les approches multi-atlas classiques de segmentation de structures anatomiques. Une approximation de l'étape de marginalisation remplace la notion de fenêtre de recherche locale par un tamisage par atlas et par étiquette. Un modèle de détection de gliomes fondé sur un a priori spatial et des critères de pré-sélection de blocs de voxels permettent d'obtenir des temps de calcul compétitifs malgré un appariement non local. Ce travail est validé et comparé à l'état de l'art sur des bases de données publiques.

Un second modèle probabiliste, symétrique au modèle de segmentation, simule des images par résonance magnétique de cas pathologiques, à partir d'une unique segmentation. Une heuristique permet d'estimer le maximum a posteriori et l'incertitude du modèle de synthèse d'image. Un appariement itératif des blocs de voxels renforce la cohérence spatiale des images simulées. Le réalisme des images simulées est évalué avec de vraies IRM et des simulations de l'état de l'art. Le raccordement d'une modèle de croissance de tumeur permet de créer des bases d'images annotées synthétiques.

Mots-clés : appariement de blocs de voxels, multi-atlas, gliome, segmentation, modèle génératif probabiliste, simulation d'image médicale, synthèse de modalité.