

Semantic parsing of brain MR images

Christian Ledig, Daniel Rueckert

*Biomedical Image Analysis Group, Department of Computing,
Imperial College London, 180 Queen's Gate, London, UK*

Abstract

The semantic parsing of brain MR images is an important step in many applications that require the segmentation of medical images. In this chapter we describe the most commonly used approach for image segmentation, namely atlas-based segmentation. We discuss the main components of atlas-based segmentation, including atlas-to-image registration, label fusion and atlas selection. In addition, we also review different brain atlases that are frequently used as anatomical priors in image segmentation of neuroanatomical structures.

Keywords: registration, segmentation, MR imaging, machine learning, computational atlases

Please cite as:

C. Ledig and D. Rueckert, "Semantic Parsing of Brain MR Images", In: Zhou, K. ed. Medical Image Recognition, Segmentation and Parsing: Machine Learning and Multiple Object Approaches (1st Edition), Academic Press, pp. 307-336, 2015.

1. Introduction

The extraction of a single or multiple anatomical meaningful regions of interest (ROIs) from a subject’s brain magnetic resonance (MR) image is essential for computer-aided diagnosis and therapy planning as well as for the extraction of biomarkers which are clinically useful (e.g. in clinical trials). The process of subdividing an image into distinct regions is referred to as *segmentation*. Once a brain MR image is segmented into its individual anatomical components, volumes or shape related measures can be readily quantified. In addition, the microstructure tissue properties of anatomical regions can be assessed via multi-modal MR imaging, *e.g.* diffusion weighted MR imaging.

The process of semantic segmentation refers to the labelling of pixels or voxels into anatomical meaningful regions. In the context of the segmentation of brain MR images this often includes tissue segmentation as well as the parcellation of tissues into distinct regions. For example, grey matter (GM) maybe subdivided into cortical GM and sub-cortical GM. The cortical GM maybe further subdivided into the temporal lobe, occipital lobe, parietal lobe and frontal lobe. Each lobe can be further subdivided in line with knowledge from anatomical or cytoarchitectonic brain atlases such as the Brodmann atlas (Brodmann, 1909). Similarly, sub-cortical GM can be further differentiated into structures such as thalamus, putamen and basal ganglia. The semantic segmentation is heavily dependent on a-priori knowledge about the location and relationship of different anatomical structures in the brain. In most cases this knowledge is represented in form of atlases of the human brain.

In this chapter, we provide an overview of existing atlas-based segmentation methods (Section 2). Even though most techniques are applicable to the segmentation of other organs or structures as well as other imaging modalities, particular focus is placed on methods for the segmentation of T1-weighted (T1w) MR brain images into anatomical structures. In Section 3 we also describe several publicly available brain atlases that are essential for the incorporation of semantic knowledge into the segmentation process.

2. Atlas-based segmentation methods

In atlas-based methods the segmentation of a brain image is inferred by aligning M brain atlases denoted by \mathbf{A}_m with $m = 1, \dots, M$. A brain atlas is

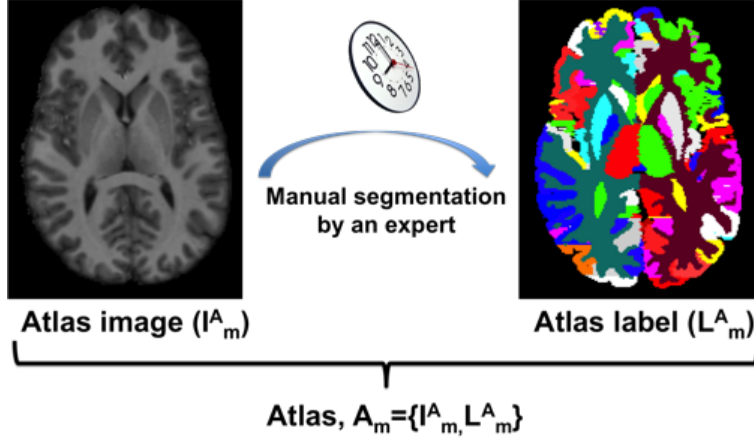


Figure 1: A brain atlas \mathbf{A}_m consists of an intensity image \mathbf{I}_m^A with corresponding segmentation \mathbf{L}_m^A . The reference segmentation is ideally created manually by an expert who follows a detailed segmentation protocol. This process is very time consuming and not scalable as it can only be performed on a small number of images.

usually created by an expert by manually or semi-automatically annotating a given volumetric brain image. The term atlas comprises the annotated atlas intensity image, \mathbf{I}_m^A , and the corresponding reference segmentation, \mathbf{L}_m^A , thus $\mathbf{A}_m = \{\mathbf{I}_m^A, \mathbf{L}_m^A\}$. A single brain atlas is shown in Figure 1. A detailed overview of commonly used brain atlases is provided in Section 3.

The aim of whole-brain segmentation is to segment a target image \mathbf{I}_{tgt} into K distinct structural ROIs. The unsegmented image $\mathbf{I}_{\text{tgt}} \in \Omega$ is indexed as $\mathbf{I}_{\text{tgt}} = \{y_1, y_2, \dots, y_N\}$ where $y_i \in \mathbb{R}^+$. Here y_i ($i = 1, \dots, N$) denotes the intensity value of the i -th voxel. The corresponding, inferred label map \mathbf{L}_{tgt} is indexed accordingly as $\mathbf{L}_{\text{tgt}} = \{l_1, l_2, \dots, l_N\}$. The probabilistic (or soft) segmentation is denoted by $\mathbf{P}_{\text{tgt}} = \{\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_N\}$, where \mathbf{p}_i is a k -valued vector with $\|\mathbf{p}_i\|_1 = 1$ and $\mathbf{p}_i(k) \geq 0$ describes the probability of voxel i belonging to structure $k \in \{1, \dots, K\}$. It can be observed that the label map \mathbf{L}_{tgt} is defined as:

$$l_i = \arg \max_k \mathbf{p}_i(k) \quad (1)$$

If \mathbf{p}_i is multi-modal l_i can be chosen as any of the modes, for example at random (Heckemann et al., 2006).

In the following an overview over widely used atlas-based segmentation approaches and their essential building blocks, *brain extraction*, *atlas align-*

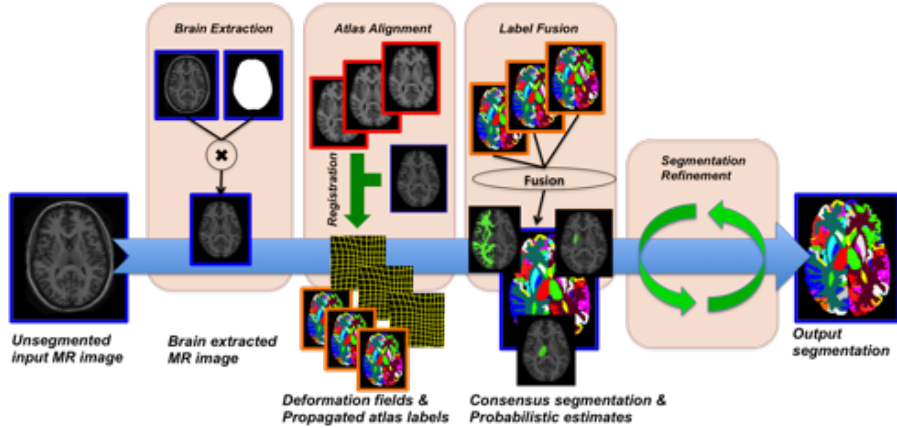


Figure 2: Schematic illustration of a typical segmentation pipeline with the common building blocks: brain extraction, atlas alignment, label fusion and segmentation refinement. The unsegmented target image and the estimated segmentation are outlined in blue, the atlas images are outlined in red and the propagated atlas labels in orange respectively.

ment, *label fusion* and *segmentation refinement* is provided. A schematic illustration of this procedure is provided in Figure 2.

2.1. Brain extraction

A prerequisite of most anatomical segmentation approaches is the availability of a binary brain mask discriminating the actual brain from non-brain tissue such as the skull, the neck or the eyes. The calculation of this mask is, however, not trivial. Many brain extraction methods tend to produce either too restrictive or too generous masks and there is no consensus of what constitutes an ideal brain mask (Eskildsen et al., 2012).

A rough categorisation can be made by distinguishing manual approaches (Eritaia et al., 2000), semi-automatic approaches (Freeborough et al., 1997) and automatic approaches (Sandor and Leahy, 1997; Smith, 2002; Ségonne et al., 2004; Leung et al., 2011; Eskildsen et al., 2012; Manjón et al., 2014). Approaches that require user interaction usually require substantial expertise, have poor inter- and intra-rater reliability and are very time consuming, which can be prohibitive when large databases are to be analysed (Freeborough et al., 1997; Warfield et al., 2004; Eskildsen et al., 2012). Skull-stripping is not the focus of this work, but a brief overview over existing approaches is provided in the following.

2.1.1. Deformable models

[Sandor and Leahy (1997)] (Brain Surface Extractor (BSE)) employ an edge-detector and morphological operations to calculate a brain mask. In [Smith (2002); Ségonne et al. (2004)] a deformable model is automatically fit to the brain surface. [Smith (2002)] (Brain Extraction Tool (BET)) calculate the center of gravity and an approximate radius of the head by thresholding the image using robust minimum/maximum estimates of the image intensities. Based on this information a brain surface tessellation is initialised that iteratively evolves. The process is constraint by both intensity and smoothness terms. [Ségonne et al. (2004)] obtain an initial estimate of the brain mask using a watershed algorithm. The surface of this approximate but robust estimate is then employed as initialisation of an active contour model [Kass et al. (1988)] that integrates both geometric and atlas-based information [Ségonne et al. (2004)]. A quantitative comparison of these methods [Sandor and Leahy, 1997; Smith, 2002; Ségonne et al. (2004)] can be found in [Fennema-Notestine et al. (2006)].

2.1.2. Database of extracted reference images

More recent and potentially more accurate methods for brain extraction rely on a database of brain extracted reference magnetic resonance images (MRIs) [Leung et al. (2011); Eskildsen et al. (2012); Manjón et al. (2014)]. In [Leung et al. (2011)] the authors nonrigidly align the reference images to the subject that is to be extracted (*cf.* Section 2.2) and perform label fusion (*cf.* Section 2.3.1). Other approaches such as [Eskildsen et al. (2012); Manjón et al. (2014)] follow a patch-based label fusion approach using linear registration only (*cf.* Section 2.3.4). Approaches for brain extraction tend to employ similar methodology to the techniques for the segmentation of anatomical structures. A comprehensive summary of relevant literature is provided in [Leung et al. (2011)] or [Eskildsen et al. (2012)].

In the remainder of this review it is assumed that \mathbf{I}_{tgt} and the atlas images $\mathbf{I}_{\text{m}}^{\text{A}}$ are brain extracted. Skull-stripping the atlas images is often trivial as the corresponding expert label maps $\mathbf{L}_{\text{m}}^{\text{A}}$ can be binarised and employed as mask. The calculation of accurate brain masks for images which contain pathologies is, however, more complicated.

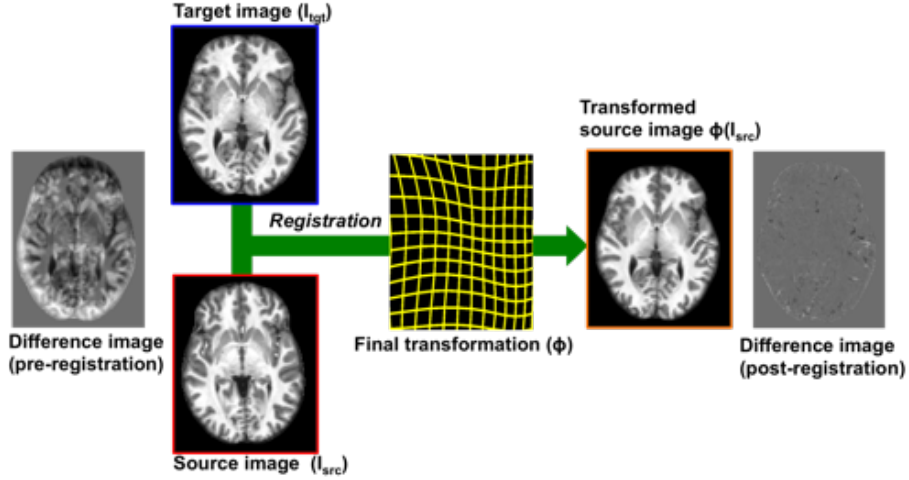


Figure 3: Schematic process of registering a source image \mathbf{I}_{src} to a target image \mathbf{I}_{tgt} with calculated transformation ϕ . The difference image of \mathbf{I}_{src} , $\phi(\mathbf{I}_{\text{src}})$, and \mathbf{I}_{tgt} is shown before and after registration respectively. The target image is outlined in blue, the source image is outlined in red (before) and orange (after registration).

2.2. Atlas-to-image registration

To exploit the labelling information encoded in the atlases most segmentation methods rely on aligning each individual atlas $\mathbf{A}_{\mathbf{m}}$ with \mathbf{I}_{tgt} . This process is commonly referred to as *image registration* and usually driven by the image intensities of the corresponding images. Image registration is highly complex and an active area of research that received continuous attention over the last decades. Therefore, it is only briefly discussed here. A comprehensive overview and evaluation of established registration approaches can be found in [Zitová and Flusser \(2003\)](#); [Sotiras et al. \(2013\)](#) and [Klein et al. \(2009\)](#).

2.2.1. Formulation of the registration problem

In general, image registration seeks to calculate a transformation $\phi : \Omega \rightarrow \Omega$ that transforms a source (moving) image \mathbf{I}_{src} such that its similarity with a target (fixed) image \mathbf{I}_{tgt} is maximised. The process of image registration is illustrated in Figure [3](#). For a given similarity measure $\mathcal{S}(\cdot, \cdot) : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$

the optimisation problem can be formulated as:

$$\hat{\phi} = \arg \min_{\phi} [-\mathcal{S}(\mathbf{I}_{\text{tgt}}, \phi(\mathbf{I}_{\text{src}})) + \mathcal{R}(\phi)] \quad (2)$$

Here \mathcal{R} is a regularisation term. In nonrigid registration, regularisation of the usually ill-posed optimisation problem is an important factor to encourage smooth and ideally even diffeomorphic transformations (Rueckert et al., 1999; Vercauteren et al., 2009).

2.2.2. Similarity measures

To quantify image similarity several measures have been proposed. Common choices for \mathcal{S} include sum of squared differences (SSD), mutual information (Collignon et al., 1995; Viola and Wells III, 1997) (MI), normalised mutual information (Studholme et al., 1999) (NMI) or cross-correlation (CC). A definition of these similarity measures is provided in Table 2.2.2. The basic assumption of SSD is that intensities of perfectly matched images are identical (Roche et al., 1999). This assumption can be relaxed when using CC (affine relationship) or MI/NMI (statistical relationship) (Roche et al., 1999; Artaechevarria et al., 2009). SSD and CC are often used to register images from the same modality/sequence, while the statistical measures have been successfully applied for multi-modal registration.

Measure	Definition	Value range	Similarity/ Dissimilarity
SSD(\mathbf{I}, \mathbf{J})	$\sum_{i=1}^N (I(i) - J(i))^2$	$\text{SSD} \geq 0$	Dissimilarity
MI(\mathbf{I}, \mathbf{J})	$H(\mathbf{I}) + H(\mathbf{J}) - H(\mathbf{I}, \mathbf{J})$	$\text{MI} \geq 0$	Similarity
NMI(\mathbf{I}, \mathbf{J})	$\frac{H(\mathbf{I}) + H(\mathbf{J})}{H(\mathbf{I}, \mathbf{J})}$	$1 \leq \text{NMI} \leq 2$	Similarity
CC(\mathbf{I}, \mathbf{J})	$\frac{\text{Cov}(\mathbf{I}, \mathbf{J})}{\sqrt{\text{Var}(\mathbf{I})} \sqrt{\text{Var}(\mathbf{J})}}$	$-1 \leq \text{CC} \leq 1$	Similarity

Table 1: Overview over possible measures to quantify similarity between two images \mathbf{I} and \mathbf{J} : SSD, MI, NMI and CC. Note that SSD is a measure of dissimilarity. MI and NMI require the calculation of individual histograms (h_I, h_J) and the joint histogram ($h_{I,J}$) of the intensity images. The measures can then be calculated based on the joint, $H(\mathbf{I}, \mathbf{J})$, and the marginal entropies, $H(\mathbf{I})$ and $H(\mathbf{J})$.

2.2.3. Transformation models

In image registration the degrees of freedom of ϕ are essential as they determine the level of detail of the computed transformations. Generally, it can be distinguished between *rigid*, *affine* and *non-rigid* image alignment.

Rigid registration establishes the optimal alignment of two images by allowing translations and rotations only. Affine registration additionally compensates for transformations that include shear and scale. The recovery of both rigid and affine transformations requires the solution of a low-dimensional optimisation problem that can be solved efficiently. An accurate affine transformation is usually a crucial prerequisite for a subsequent non-rigid registration (Rueckert et al., 1999).

The aim of non-rigid registration is the recovery of a deformation field that brings both images into optimal alignment. Here, the deformation field can be described by a transformation model which is often built on cosine or B-spline basis functions (Ashburner and Friston, 1999; Rueckert et al., 1999; Andersson et al., 2007; Klein et al., 2009; Modat et al., 2010). Next to these approaches, non-parametric methods that do not model the transformation explicitly have been successfully applied. Popular examples include the so-called Demons algorithm (Thirion, 1998; Vercauteren et al., 2009) or the symmetric registration method proposed in (Avants et al., 2008). The parameter space of the resulting optimisation problem is large and can exceed millions of parameters, especially if no transformation model is used (Klein et al., 2009; Vercauteren et al., 2009).

In the context of atlas alignment for image segmentation it should be noted that non-rigid transformations can be calculated at different levels of detail. For example, when an explicit transformation model based on B-spline basis functions is used, the control point spacing and thus the number of parameters can be varied. Usually a finer control point spacing yields a more accurate registration result, however, it comes at a substantially increased computational cost.

2.2.4. Large deformation registration problem

The accurate registration of an atlas and unsegmented MR image can be difficult if the target image differs from the available atlases due to general anatomical variability or pathological changes. Also intensity inhomogeneities or noise can facilitate an accurate registration difficult (Vovk et al., 2007).

There have been several methods proposed to address the large deformation registration problem in adult brains in the context of Alzheimer’s disease (AD). In (Heckemann et al., 2010) the authors proposed Multi-Atlas Propagation with Enhanced Registration (Heckemann et al., 2010, 2011) (MAPER). MAPER employs automatically calculated brain tissue segmentations to guide the registration process. This allows a robust image alignment, even if the target image shows severe brain atrophy. In another approach, (Wolz et al., 2010a) described an iterative approach to improve segmentation accuracy by propagating atlas labels over a learned manifold while refining intermediate segmentations based on image intensities using graph-cuts (GC) (*cf.* Section 2.4). It was shown by (Gerber et al., 2010) that coordinates within a low-dimensional space, a nonlinear manifold, allow for a meaningful image comparison and statistical tests.

A manifold of the anatomical variation in a given data set is also learned in (Hamm et al., 2010). In this work it was shown that the problem of recovering a large deformation between two images can be simplified by solving a series of small deformation registration problems (Hamm et al., 2010). The small deformations are calculated along the shortest path between the images on the learned manifold (Hamm et al., 2010).

In the following, ϕ_m denotes the calculated transformation from the atlas space of \mathbf{A}_m to the coordinate system of \mathbf{I}_{tgt} and \mathbf{A}_m^ϕ denotes the propagated atlas respectively. Once the atlas label maps, \mathbf{L}_m^A , reside in the same coordinate system as the unsegmented target image, \mathbf{I}_{tgt} , a consensus segmentation can be inferred using a variety of label fusion techniques (*cf.* Section 2.3).

2.3. Label fusion

2.3.1. Majority vote fusion

In majority vote fusion (MVF), also called “vote-rule” or “decision” fusion, each of the M propagated atlas labels \mathbf{L}_m^A contributes equally to the final segmentation. A certain voxel is thus labelled according to the opinion on which the majority of the propagated atlas segmentations agree. Formally the probability of voxel i being labelled as structure k can be calculated as:

$$\mathbf{p}_i(k) = \frac{1}{M} \sum_{m=1}^M \delta(L_m^A(i), k) \quad (3)$$

Here, $\delta(\cdot, \cdot)$ is the Kronecker delta defined as:

$$\delta(v, w) = \begin{cases} 1 & \text{if } v = w \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The actual segmentation estimate, \mathbf{L}_{tgt} , is then readily inferred through Equation 1.

MVF was first described by Rohlfing et al. (2004) where it was applied to the segmentation of bee brains. Heckemann et al. (2006) successfully employed MVF to segment MR brain images and presented a model that describes the basic assumption of label fusion. In this model it is assumed that there are two major sources of segmentation errors: 1) *Systematic errors* (ϵ_{sys}) due to deviations of the (manual) reference segmentation from the true segmentation or due to consistent registration bias. 2) *Random errors* (ϵ_{rand}) introduced due to inaccuracies in individual reference labels or image registrations (Heckemann et al., 2006). With this assumption random errors can be corrected by fusing several label estimates. However, the systematic errors asymptotically limit the segmentation accuracy as the number of fused atlases increases (Heckemann et al., 2006). In Heckemann et al. (2006) the following relation is suggested to model the segmentation accuracy SI_{model} , measured as Dice coefficient (Dice, 1945) or so-called similarity index (SI), dependent on the number of atlases K :

$$SI_{\text{model}} = 1 - \epsilon_{\text{sys}} - \frac{\epsilon_{\text{rand}}}{\sqrt{K}} \quad (5)$$

Figure 4 illustrates the schematic process and the qualitative behaviour of label fusion approaches using the model from Equation 5. According to Artachevarria et al. (2009) MVF is the most simple and most popular label fusion strategy, and does not require any a priori knowledge except for the actual segmentations that are to be fused. MVF has its origins in the more general combination of classifiers as for example Xu et al. (1992), Kittler et al. (1998) or Kuncheva (2004).

2.3.2. Atlas selection

If only a single atlas from a given atlas database is to be used, a carefully selected atlas that is similar to \mathbf{L}_{tgt} allows a higher segmentation accuracy than choosing an atlas at random. This was confirmed in a variety of applications, such as the segmentation of computed tomography (CT) images

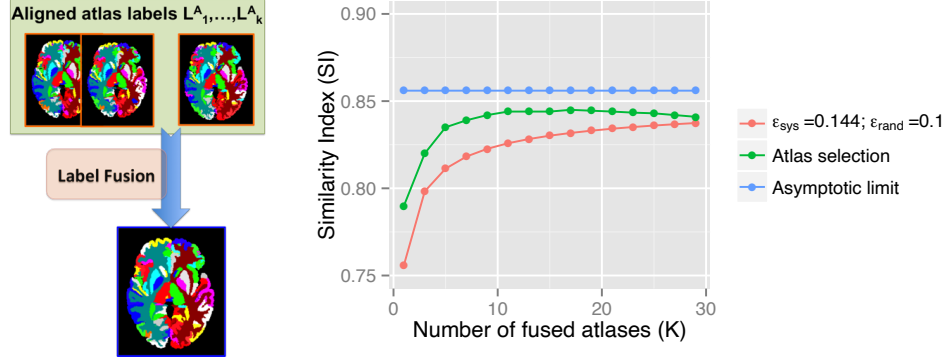


Figure 4: Left: Schematic process of fusing K atlas labels that have been aligned to the target image into a consensus segmentation. Right: Illustrations of the increasing segmentation accuracy with increasing number of fused atlas labels. The model described in Equation 5 is shown for $\epsilon_{\text{sys}} = 0.144, \epsilon_{\text{rand}} = 0.1$ (red) with corresponding asymptotic limit (blue). A qualitative plot indicating the increased accuracy when atlas selection is used is also shown (green). The diagram is based on and adapted from Heckemann et al. (2006); Aljabar et al. (2009)

of the head (Han et al., 2008) or the segmentation of bee brains (Rohlfing et al., 2004) and human brains in MRI (Wu et al., 2007).

However, the fusion of multiple atlases (*e.g.* using MVF) substantially outperforms the segmentation based on a single atlas only (Rohlfing et al., 2004; Heckemann et al., 2006; Aljabar et al., 2009). Klein et al. (2007) and Aljabar et al. (2009) thus suggested to combine these approaches and perform label fusion on a *selection* of atlases from a larger database.

As selection criterion, the image similarity between the aligned atlas images and \mathbf{I}_{tgt} can be calculated using a variety of similarity measures such as SSD, MI (Collignon et al., 1995; Viola and Wells III, 1997; Klein et al., 2007), NMI (Studholme et al., 1999; Wu et al., 2007), or CC (Aljabar et al., 2009) (*cf.* Table 2.2.2). Further, image similarity can be assessed on different alignment levels, *e.g.* after affine or nonlinear registration (Rohlfing et al., 2004; Klein et al., 2007; Wu et al., 2007), but also in a template space (Aljabar et al., 2009). When the selection is performed in a template space all atlas images can be pre-registered to this template. This reduces the computational burden dramatically as during atlas selection a single registration, aligning \mathbf{I}_{tgt} to the template, suffices. Only the selected atlases are then aligned with \mathbf{I}_{tgt} in a non-rigid fashion (Aljabar et al., 2009). In addition to image

similarity, other criteria such as characteristics of the required deformation to align the images (Rohlfing et al., 2004) or meta-information (age, sex, etc.) (Aljabar et al., 2009) can be used.

In summary, atlas selection addresses two limitations of standard multi-atlas label fusion: First, the accurate nonlinear alignment (*cf.* Section: 2.2) of numerous atlases to the unsegmented image, \mathbf{I}_{tgt} , is computationally intensive. Second, it was shown that there is an asymptotic limit of the segmentation accuracy caused by systematic and anatomical variations (Aljabar et al., 2009; Heckemann et al., 2006). Segmentation results can substantially improve up to around 20 selected atlases (Heckemann et al., 2006; Aljabar et al., 2009) (*cf.* Figure 4). However, using more atlases with an anatomy that increasingly differs from \mathbf{I}_{tgt} might even deteriorate segmentation accuracy (Aljabar et al., 2009). Nevertheless it should be noted that the ideal number of selected atlases might be strongly dependent on the quality of the atlases, the anatomical variability of the subjects and the ROI that is to be segmented.

2.3.3. Weighted vote fusion

Instead of selecting atlases, and thus potentially neglecting relevant information, another approach is to weigh the contribution of each individual atlas. A detailed description and comparison of this category of label fusion can be found in (Artaechevarria et al., 2009).

In globally weighted fusion (GWF) the probabilistic label estimates are calculated based on *global*, atlas-dependent voting weights w_m^g as:

$$\mathbf{p}_i(k) = \frac{\sum_{m=1}^M w_m^g \delta(L_m^A(i), k)}{\sum_{k'=1}^K \sum_{m=1}^M w_m^g \delta(L_m^A(i), k')} \quad (6)$$

Here, the voting weight, w_m^g , of each atlas is generally determined by the image similarity of the transformed atlas MRIs, \mathbf{I}_m^A , and the subject image \mathbf{I}_{tgt} . Potential measures to quantify this similarity of intensity images include SSD, MI, NMI or CC (Artaechevarria et al., 2009). The calculation of global voting weights based on these measures, w_m^{gSSD} , w_m^{gCC} , w_m^{gMI} , w_m^{gNMI} , is described in Table 2.2.2. As the SSD increases with increasing image *dissimilarity* the voting weight is defined as the inverse of the actual SSD, $w_m^{\text{gSSD}} = \text{SSD}(\mathbf{I}_m^A, \mathbf{I}_{\text{tgt}})^{-1}$. (Artaechevarria et al., 2009) further introduced a gain parameter, p , that potentiates the respective weights, *e.g.* $(w_m^{\text{gNMI}})^p$. However it was found that the segmentation result is usually not very sensi-

tive towards this parameter and the optimal choice depends on the dataset (Artaechevarria et al., 2009).

Segmentation accuracy can be further increased by calculating *local* voting weights, $w_m^l(i)$, within a location-specific region (Artaechevarria et al., 2009). These regions can be for example a spherical or cubical neighbourhood of the voxel under consideration. The size of the neighbourhood is a tuneable parameter, however, a neighbourhood radius of $r = 5$ was shown to yield good results (Artaechevarria et al., 2009). In locally weighted fusion (LWF) the probabilistic segmentation estimates are calculated as:

$$p_i(k) = \frac{\sum_{m=1}^M w_m^l(i) \delta(L_m^A(i), k)}{\sum_{k'=1}^K \sum_{m=1}^M w_m^l(i) \delta(L_m^A(i), k')} \quad (7)$$

It was shown that weighted fusion overall outperforms MVF (Artaechevarria et al., 2009; Sabuncu et al., 2010). LWF is usually superior to GWF and particularly useful in image regions with high contrast (Artaechevarria et al., 2009; Sabuncu et al., 2010). The ideal fusion strategy is thus dependent on the brain anatomy that is to be segmented. LWF with SSD often performs well (Artaechevarria et al., 2009; Wang et al., 2013). However, when compared to NMI, SSD is more sensitive to noise in low-contrast regions and relies on the assumption of similar intensity profiles of the images (Roche et al., 1999; Artaechevarria et al., 2009).

2.3.4. Patch-based label fusion

Inspired by the work on non-local means filtering for image denoising (Buades et al., 2005; Coupé et al., 2008), Coupé et al. (2010) proposed a *patch-based* approach to address the problem of label fusion. Patch-based label fusion was subsequently described thoroughly in Rousseau et al. (2011) and Coupé et al. (2011). In patch-based label fusion techniques image patches of the unsegmented image are locally compared to image patches in the atlas images. Usually patches are defined as three-dimensional (3D) volumes of a given diameter. For a fixed patch in \mathbf{I}_{tgt} with centre voxel i , $P(\mathbf{I}_{\text{tgt}}, i)$, numerous patches in $\mathbf{I}_{\text{m}}^{\text{A}}$ are taken into account. Specifically, all patches with centre j , $P(\mathbf{I}_{\text{m}}^{\text{A}}, j)$, with the spatial constraint that j lies within a defined neighbourhood, \mathcal{N}_i , of i are considered. Both the size of the 3D patches (patch size) and the size of the neighbourhood (search window size) are parameters that need to be chosen. An illustration of this process is provided in Figure 5. The exact label fusion procedure can be formalised as:

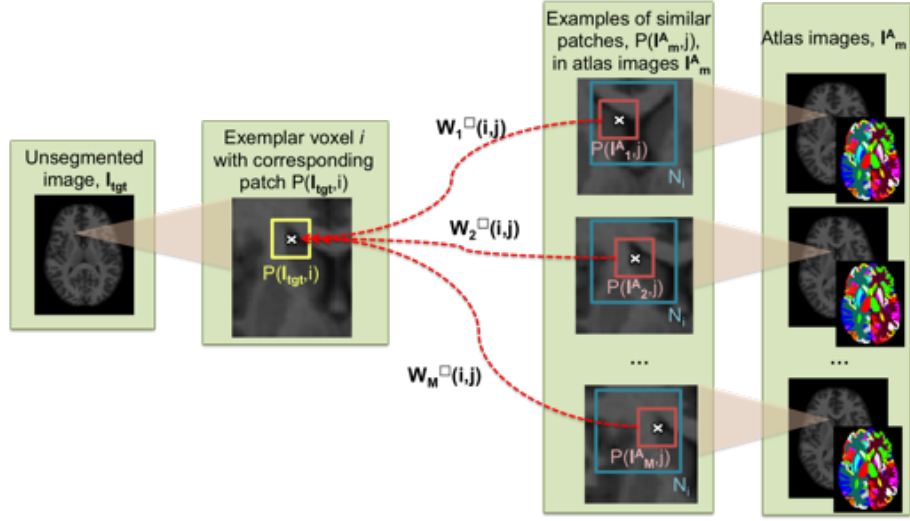


Figure 5: Illustration of the weight calculation when segmenting one example voxel i . The weights, $w_m^{\square}(i, j)$, are determined between the fixed reference patch, $P(I_{tgt}, i)$ (yellow), and atlas patches, $P(I_m^A, j)$ (red). For each atlas image several atlas patches that are shifted within a local search neighbourhood (cyan) are considered. With a weighted voting approach a probabilistic segmentation (cf. Eq. 8) is obtained. Illustration adapted and modified from [Coupé et al. \(2011\)](#); [Rousseau et al. \(2011\)](#).

$$\mathbf{p}_i(k) = \frac{\sum_{m=1}^M \sum_{j \in \mathcal{N}_i} w_m^\square(i, j) \delta(L_m^A(j), k)}{\sum_{k'=1}^K \sum_{m=1}^M \sum_{j \in \mathcal{N}_i} w_m^\square(i, j) \delta(L_m^A(j), k')} \quad (8)$$

Here $w_m^\square(i, j)$ is a weight that is calculated based on the similarity of the patch $P(\mathbf{I}_{\text{tgt}}, i)$, centred at voxel i in \mathbf{I}_{tgt} , and the patch $P(\mathbf{I}_{\mathbf{m}}^A, j)$, centred at voxel j in atlas $\mathbf{I}_{\mathbf{m}}^A$. A common definition (Buades et al., 2005; Coupé et al., 2011; Rousseau et al., 2011) of this similarity measure is:

$$w_m^\square(i, j) = \exp \frac{-\|P(\mathbf{I}_{\text{tgt}}, i) - P(\mathbf{I}_{\mathbf{m}}^A, j)\|^2}{Nh} \quad (9)$$

Here $\|\cdot\|$ is the L_2 -norm and N is the number of voxels in the patches. Further, h is a decay parameter that requires tuning (Coupé et al., 2011). If $h \rightarrow \infty$ all patches are weighted equally, if $h \rightarrow 0$ only the best matching patch is selected. Potentially, $w_m^\square(i, j)$ might be set to zero below a given threshold (Coupé et al., 2011).

The main motivation for patch-based label fusion is to remove the computationally expensive requirement of nonrigidly aligning the reference atlases (*cf.* Sec. 2.2). The patch-based formulation further relaxes the assumption of an one-to-one mapping between the atlas images, $\mathbf{I}_{\mathbf{m}}^A$, and the unsegmented image, \mathbf{I}_{tgt} (Rousseau et al., 2011). An approximate alignment, using for example affine registration (Coupé et al., 2011), is sufficient to establish reasonable patch correspondences. Nevertheless, it must be noted that nonrigid alignment can further improve patch-based segmentation as it allows the more rigorous incorporation of topological constraints (Coupé et al., 2011; Rousseau et al., 2011).

In patch-based label fusion one can discriminate *point-wise* and *multipoint* estimators (Rousseau et al., 2011). In Equation 8 a point-wise estimator is described, in which the label probability of each individual voxel (or point) i is estimated individually. In contrast to this, multipoint estimators calculate at each individual voxel a label patch estimate (Rousseau et al., 2011). Specifically, instead of considering only the centre label, $L_m^A(j)$ (*cf.* Equation 8), label information from the whole patch centred at j is fused into a patch estimate. While point-wise estimators are commonly used and computationally more efficient, it was shown that multipoint estimators can potentially increase segmentation accuracy (Rousseau et al., 2011). More details on multipoint estimators can be found in Katkovnik et al. (2010; Rousseau et al., 2011).

Recent advances in patch-based segmentation include the application of

sparsity techniques (Tong et al., 2013) or the efficient search of patch correspondences using the PatchMatch algorithm (Barnes et al., 2009; Shi et al., 2014; Ta et al., 2014). Next to image segmentation, it was proposed to employ patch-based methods to grade subjects according to clinical variables (Coupé et al., 2012; Coupé et al., 2012). In this patch-based grading approach a meaningful clinical label, *e.g.* AD or healthy control (HC), is associated with each patch of a reference database. Instead of fusing anatomical labels, $L_m^A(j)$, these clinical labels are fused in Equation 8 within a region. It was shown that the resulting grading value allows the accurate discrimination of different AD disease states (Coupé et al., 2012; Coupé et al., 2012).

2.3.5. Joint label fusion

The label fusion techniques that were presented so far treat each atlas independent from the others for the calculation of voting weights. Thus, none of these approaches takes into account that similar or, in general, correlated atlases might produce similar segmentation errors (Wang et al., 2013). To address this limitation Wang et al. (2013) introduced an approach called joint label fusion (JLF) in which the joint probability distribution of two atlases producing the same labelling error is explicitly modelled.

In particular the authors aim to minimise the expected error between the true, yet unknown segmentation of \mathbf{I}_{tgt} and the consensus segmentation estimate \mathbf{L}_{tgt} . This optimisation problem can be solved once dependency matrices \mathbf{M}_i of size $M \times M$ are established that model the likelihood that a pair of the M available atlases make an identical error at voxel i . In Wang et al. (2013), the authors calculate the correlation of the intensity similarity of two atlases with respect to \mathbf{I}_{tgt} . A local estimation of \mathbf{M}_i is obtained by performing the similarity computation in a local neighbourhood of the corresponding voxel i . More details on JLF can be found in Wang et al. (2013).

Many approaches ranking among the top performing methods are based on JLF strategies that incorporate the non-local patch-based concept. This was confirmed for the segmentation of various anatomies in recent Segmentation Challenges (Landman and Warfield, 2012; Asman et al., 2013). For example, JLF performs very well for the segmentation of the brain into distinct ROIs (Landman and Warfield, 2012; Asman et al., 2013), but also for the segmentation of the myocardium (Asman et al., 2013; Bai et al., 2015).

2.3.6. Statistical label fusion

With origins in the popular Simultaneous Truth And Performance Level Evaluation (Warfield et al., 2004) (STAPLE) algorithm there is another category of statistical label fusion approaches. Given several manual label sets, STAPLE was proposed to simultaneously estimate the performance level parameters of each expert rater and the most probable ground truth. The STAPLE algorithm computes a probabilistic estimate of the true segmentation (Warfield et al., 2004). For the sake of consistency with existing literature the commonly used notation is employed to outline the STAPLE algorithm. This notation is slightly different in comparison to the rest of this section. This also accommodates the fact that STAPLE was originally described for the estimation of a ground truth from several expert reference segmentations rather than for the fusion of automatic label estimates. The following description is adapted from Warfield et al. (2004), where more details about the exact implementation of STAPLE can be found.

Let \mathbf{D} be a $N \times M$ matrix that describes the decisions of M raters (or atlases) at N voxels. Let \mathbf{T} be a vector of size N indicating the true but unknown segmentation. For K possible labels \mathbf{D} and \mathbf{T} are indexed as $D_{im} \in \{1, \dots, K\}$ and $T_i \in \{1, \dots, K\}$. The goal of the STAPLE algorithm is to estimate a performance tensor $\hat{\theta}$ of size $M \times K \times K$ that maximises the log-likelihood of observing the complete data (\mathbf{D}, \mathbf{T}) . Here, element $\theta_{ms's}$ quantifies the probability that rater m decides on label s' given that the true label is s . For a probability mass function $f(\mathbf{D}, \mathbf{T}|\theta)$ that describes the probability of observing the complete data, $\hat{\theta}$ can be calculated by solving:

$$\hat{\theta} = \arg \max_{\theta} \ln f(\mathbf{D}, \mathbf{T}|\theta) \quad (10)$$

The solution of this problem would be trivial if the true segmentation \mathbf{T} was known. As \mathbf{T} is however unknown Warfield et al. (2004) suggested to employ the expectation-maximisation (EM) algorithm to solve Equation 10. In this approach, a ground truth \mathbf{T} is estimated in the Expectation-step and the ideal performance parameters $\hat{\theta}$ found in the Maximisation-step. This is repeated iteratively until the model converges. Convergence to a local maximum of Equation 10 is guaranteed (Warfield et al., 2004). There are different strategies to initialise θ in the first iteration (Warfield et al., 2004). Furthermore, spatial prior information can be incorporated through probabilistic priors or smoothness constraints based on Markov random fields(MRFs) (Warfield et al., 2004).

Next to the basic STAPLE algorithm outlined above, there have been several

successful attempts to incorporate both image intensity information (Cardoso et al., 2013; Asman and Landman, 2013) and the concept of non-local patch-based approaches into the STAPLE framework (Asman and Landman, 2013). Many other extensions include formulations that allow the application to data with missing labels (Landman et al., 2012), local estimation of rater performances (Commowick et al., 2012; Asman and Landman, 2012) or the fusion of probabilistic decisions (Akhondi-Asl and Warfield, 2013).

2.4. Refinement of image segmentations

The segmentation estimates obtained by multi-atlas label fusion, as described in the previous Section 2.3, can be further improved. In the following two successful strategies to the refinement of image segmentations are presented: the refinement of segmentations based on image intensities and the refinement based on classifiers that were trained to correct systematic bias of a segmentation method.

2.4.1. Intensity-based refinement

The probabilistic segmentation estimates, \mathbf{P}_{tgt} , obtained by automatic label fusion techniques (cf. Section 2.3) can be refined based on the image intensities of \mathbf{I}_{tgt} . It was shown that modifying image segmentations based on the actual image intensities can substantially improve segmentation results (van der Lijn et al., 2008; Wolz et al., 2009; Lötjönen et al., 2010; Ledig et al., 2015).

Using Bayes' theorem the refinement task is often modelled as calculating the maximum a posteriori (MAP) estimate as:

$$\mathbf{L}_{\text{tgt}} = \arg \max_{\mathbf{L}} p(\mathbf{L} | \mathbf{I}_{\text{tgt}}) = \arg \max_{\mathbf{L}} \frac{p(\mathbf{I}_{\text{tgt}} | \mathbf{L}) p(\mathbf{L})}{p(\mathbf{I}_{\text{tgt}})} = \arg \max_{\mathbf{L}} p(\mathbf{I}_{\text{tgt}} | \mathbf{L}) p(\mathbf{L}) \quad (11)$$

Assuming voxel-wise independence the image likelihood, $p(\mathbf{I}_{\text{tgt}} | \mathbf{L})$ can be calculated as $\prod_{i=1}^N p(y_i | l_i)$ where $p(y_i | l_i)$ is given by a predefined intensity model. The probability of a segmentation, $p(\mathbf{L})$, is often modelled based on spatial prior knowledge and smoothness constraints between adjacent labels. Since the target image is fixed, the optimisation problem is independent of $p(\mathbf{I}_{\text{tgt}})$.

Two popular approaches to solve the optimisation problem stated in Equation 11 are based on graph-cuts (GC) (Greig et al., 1989; Boykov et al., 2001;

van der Lijn et al., 2008; Wolz et al., 2009) or the expectation-maximisation (EM) algorithm (Van Leemput et al., 1999; Lötjönen et al., 2010; Ledig et al., 2015).

Expectation-maximisation optimisation

The widely used EM-optimisation was presented for image segmentation by Van Leemput et al. (1999). For the sake of consistency with existing literature, in the following paragraph the notation is employed that was also used in Van Leemput et al. (1999); Cardoso et al. (2011); Ledig et al. (2015).

Based on a Gaussian mixture model (GMM) it is assumed that given the intensity characteristics $\Phi = \{(\mu_1, \sigma_1), (\mu_2, \sigma_2), \dots, (\mu_K, \sigma_K)\}$ of K structural classes the likelihood of observing intensity y_i at voxel i is given as:

$$f(y_i|\Phi) = \sum_k f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi) f(\mathbf{z}_i = \mathbf{e}_k) \quad (12)$$

Here it is assumed that the probability of a voxel i to have intensity y_i , $f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi)$, given that it belongs to class k , ($\mathbf{z}_i = \mathbf{e}_k$), is described by a normal distribution (Wells III et al., 1996; Van Leemput et al., 1999; Zhang et al., 2001; Cardoso et al., 2011; Ledig et al., 2015). Thus $f(y_i|\mathbf{z}_i)$ is modelled as $f(y_i|\mathbf{z}_i = \mathbf{e}_k, \Phi) = \mathcal{G}_k(y_i)$ where \mathcal{G}_k denotes the Gaussian distribution with corresponding parameters (μ_k, σ_k) . The prior probability $f(\mathbf{z}_i = \mathbf{e}_k)$ that a voxel i belongs to structure k is given by the probabilistic label estimates after multi-atlas label propagation, $\mathbf{P}_{\text{tgt}}^{\text{prior}}$ (cf. Section 2.3). By assuming that voxels are statistically independent, the probability of observing the complete image \mathbf{I}_{tgt} , given that the model parameters Φ are known in iteration m , is given by $f(\mathbf{I}_{\text{tgt}}|\Phi^{(m)}) = \prod_i f(y_i|\Phi^{(m)})$. In the EM approach this model is solved by interleaving the expectation of the probabilities of each voxel i to belong to structure k , $p_{ik}^{(m)}$, and the maximisation of the model by updating the model parameters $\Phi^{(m)}$. It is assumed that the probabilities, $p_{ik}^{(m+1)}$, are known in iteration $(m+1)$, so that the model parameters, Φ , can be updated as:

$$\mu_k^{(m+1)} = \frac{\sum_{i=1}^N p_{ik}^{(m+1)} y_i}{\sum_{i=1}^N p_{ik}^{(m+1)}} \quad , \quad \sigma_k^{(m+1)} = \sqrt{\frac{\sum_{i=1}^N p_{ik}^{(m+1)} (y_i - \mu_k^{(m+1)})^2}{\sum_{i=1}^N p_{ik}^{(m+1)}}} \quad (13)$$

Given the updated model parameters the estimate of the class probabilities

in the next iteration is given as:

$$p_{ik}^{(m+1)} = \frac{f(y_i | \mathbf{z}_i = \mathbf{e}_k, \Phi^{(m)}) f(\mathbf{z}_i = \mathbf{e}_k)}{\sum_{k'=1}^K f(y_i | \mathbf{z}_i = \mathbf{e}_{k'}, \Phi^{(m)}) f(\mathbf{z}_i = \mathbf{e}_{k'})} \quad (14)$$

Usually the model converges after a few iterations.

Smoothness of the final segmentation can be enforced with a global and stationary MRF, which can be integrated using the mean field approximation (Zhang, 1992), following the example of Van Leemput et al. (1999), Cardoso et al. (2011) or Ledig et al. (2015). This also allows the incorporation of topological knowledge next to the spatial information provided by the prior estimates. The MRF energy function is usually calculated based on the probabilistic label estimates in iteration m , in the first-order neighbourhood of each image voxel. A connectivity matrix, G of size $K \times K$, can be defined that describes the connectivity between class k and j . Usually G is defined as:

$$G(k, j) = \begin{cases} 0, & \text{if } k = j \\ \beta, & \text{if structures } k \text{ and } j \text{ share a boundary} \\ \gamma, & \text{if structures } k \text{ and } j \text{ are distant} \end{cases} \quad (15)$$

Here β and γ , with $0 \leq \beta \leq \gamma$, are parameters describing the penalty for certain neighbourhood configurations.

Images of substantially deformed or abnormal brains pose great challenges to the label fusion techniques described in Section 2.3 as the atlas images are difficult to align due to *e.g.* existing pathology in the target. This results in inaccurate spatial a-priori information limiting the potential of refinement techniques based on intensities. It was shown that explicitly relaxing the spatial priors $\mathbf{P}_{\text{tgt}}^{\text{prior}}$ based on image intensities can substantially improve segmentation results (Cardoso et al., 2011; Ledig et al., 2015). In Ledig et al. (2015), the authors proposed a method called Multi-Atlas Label Propagation with Expectation-Maximisation based refinement (MALP-EM) with a prior relaxation technique to successfully segment abnormal MR images.

Graph-cut optimisation

The optimisation problem in Equation 11 can be rewritten (Greig et al., 1989; van der Lijn et al., 2008) as a minimisation problem by taking the negative logarithm as:

$$\mathbf{L}_{\text{tgt}} = \arg \min_{\mathbf{L}} (-\ln p(\mathbf{I}_{\text{tgt}} | \mathbf{L}) - \ln p(\mathbf{L})) = \arg \min_{\mathbf{L}} (E_{\text{intensity}}(\mathbf{L}) + E_{\text{prior}}(\mathbf{L})) \quad (16)$$

This amended formulation allows the definition of a network graph for which a minimum cut can be calculated based on the Ford-Fulkerson max-flow/min-cut algorithm (Greig et al., 1989). In Equation 16 the intensity (or data) term, $E_{\text{intensity}}(\mathbf{L}) = -\sum_{i=1}^N \ln p(y_i|l_i)$, quantifies the agreement of the image data, \mathbf{I}_{tgt} , with the intensity model. A common choice for the intensity model of $p(y_i|l_i)$ is a Gaussian probability distribution with label-specific parameters $(\mu_{l_i}, \sigma_{l_i})$ (van der Lijn et al., 2008; Wolz et al., 2010b). The second term, $E_{\text{prior}}(\mathbf{L})$, incorporates both spatial prior information provided through the probabilistic segmentation estimates (*e.g.* obtained through label fusion) and smoothness constraints which are often modelled through MRFs (van der Lijn et al., 2008; Wolz et al., 2010b). The cost function that is minimised by GC can then be summarised as:

$$E(\mathbf{L}) = \underbrace{-\sum_{i=1}^N \ln p(y_i|l_i)}_{E_{\text{intensity}}} - \underbrace{\sum_{i=1}^N \ln p(l_i)}_{E_{\text{spatial prior}}} + \underbrace{\sum_{i=1}^N \sum_{j \in \mathcal{N}_i} G(l_i, l_j)}_{E_{\text{smoothness prior}}} \quad (17)$$

Here $p(l_i)$ is the prior probability that voxel i has label l_i . It can be calculated using label fusion (*cf.* Section 2.3) so that $p(l_i) = \mathbf{p}_i(l_i)$. Furthermore \mathcal{N}_i is the set of voxels neighbouring voxel i and as in Equation 15 $G(l_i, l_j)$ penalises non-smooth label configurations ($G(l_i, l_j) > 0$ for $l_i \neq l_j$). The individual energy terms can be multiplied with weighting factors to control their individual contribution (Song et al., 2006; van der Lijn et al., 2008; Wolz et al., 2010b).

GC optimisation was first introduced for binary segmentation problems (Greig et al., 1989; Boykov et al., 2001; Boykov and Kolmogorov, 2004) for which a global optimum can be found. However, the optimisation problem can also be formulated as a multiway cut problem and employed for the segmentation of multiple labels (Boykov et al., 2001; Boykov and Kolmogorov, 2004; Song et al., 2006). For multiway cut problems, Boykov et al. (2001) presented an algorithm that efficiently calculates approximate solutions of the global minima with optimality bounds. GC optimisation typically assigns one strict label to each voxel.

Based on GC, Wolz et al. (2010b) proposed a framework to measure longitudinal changes by building a four-dimensional (4D) graph based on edges between both spatially and temporally neighbouring voxels.

2.4.2. Learning-based refinement

Learning-based refinement is based on the assumption that a substantial fraction of falsely labelled image voxels is due to systematic bias of the employed segmentation method (Wang et al., 2011). Systematic bias often originates in different definitions of manual segmentation protocols or inaccurate translation of the manual protocol to the automatic method (Wang et al., 2011). The assumption made by Wang et al. (2011) is similar to the one made by Heckemann et al. (2006): In the work by Heckemann et al. (2006) the authors described segmentation inaccuracies in the context of multi-atlas label propagation as a combination of random variability of propagated labels and systematic errors.

The main contribution of Wang et al. (2011) is the proposal to explicitly learn systematic segmentation errors with respect to reference segmentations using machine learning. Specifically, features based on image intensities and label context are extracted in a local neighbourhood of each voxel and combined with spatial information. These features are subsequently employed to train error detection and error correction classifiers using AdaBoost (Freund and Schapire, 1995) with respect to reference segmentations (Wang et al., 2011). These classifiers have been shown to allow a significant reduction of systematic bias in the context of many applications. In the original work, Wang et al. (2011) presented substantial improvements when the method is employed to correct automatically calculated hippocampus segmentations, brain masks and brain tissue segmentations.

A schematic illustration of the method is provided in Figure 6. More details on learning-based refinement can be found in Wang et al. (2011).

2.5. Remarks

Recently several segmentation challenges were held in conjunction with international conferences. The methods with particularly good performance for segmenting brain data were often based on JLF (*cf.* Section 2.3.5) and corrected for systematic segmentation bias (*cf.* Section 2.4.2) (Landman and Warfield, 2012; Asman et al., 2013).

The careful design and implementation of each individually outlined building block, *brain extraction*, *atlas alignment*, *label fusion*, *label refinement*, is essential and critical to realise a segmentation approach that is both accurate and robust. Failure or inaccuracies within a single of these 'modules' is likely

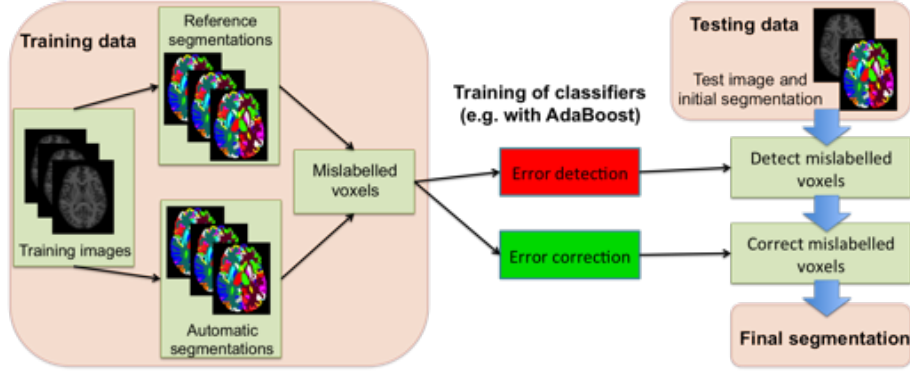


Figure 6: Schematic process of training classifiers for the detection and correction of systematic segmentation bias. This illustration corresponds to the variant named *explicit bias correction* in Wang et al. (2011). Note that for binary segmentation tasks the error detection and error correction classifiers are identical. Illustration adapted and modified from Wang et al. (2011).

to negatively impact the final segmentation result. It is tempting to consider each of these blocks separately to reduce the complexity of the task. This further allows the independent implementation, tuning and validation of each module. It should be mentioned, however, that there is some evidence that more complex approaches that model and solve several building blocks jointly (*e.g.* registration and segmentation) can increase accuracy (Ashburner and Friston, 2005). Nevertheless, in practice the theoretical advantages of these joint models might be outweighed by the above-mentioned merits of a modular approach.

3. Brain atlases from MR images

Atlases of the human brain can be considered as maps that associate a spatial position within the brain with structural or functional information such as a specific anatomical structure. Early brain atlases were created based on a single subject or a small number of selected subjects (Mazziotta et al., 1995). While these early atlases provided great insight into brain anatomy they were limited in their ability to represent the large intersubject variability of the human brain. To address this problem, probabilistic population atlases were created from a population of brain images that were aligned with and segmented in a common reference or so-called stereotaxic space,

e.g. in [Shattuck et al., 2008].

Popular stereotaxic spaces are given by the MNI-305¹ [Evans et al., 1993] and the MNI-152² [Mazziotta et al., 1995]) brain templates, which are available through the Montreal Neurological Institute (<http://www.mni.mcgill.ca/>) (MNI). Both templates were created by averaging the contributing MR intensity images (305, and 152 respectively) in a common space after linear image registration [Evans et al., 1993; Mazziotta et al., 1995]. The MNI-152 template was built by the International Consortium for Brain Mapping (ICBM) from images acquired with better contrast and at a higher resolution than the images on which the MNI-305 template is based². The stereotaxic spaces described by these template images can serve as target to align individual atlases of an atlas database [Hammers et al., 2003; Shattuck et al., 2008], such as the ones described in this section. The propagated atlas labels can then be averaged using label fusion techniques (*cf.* Section 2.3) to create a probabilistic brain atlas. It must be noted that the characteristics of the resulting probabilistic atlas, such as spatial variability of individual structures, directly depends on the chosen template space (*e.g.* MNI-305, MNI-152) and the registration approach (*e.g.* affine or nonrigid transformation model). As a consequence, based on atlas databases such as the ones described in the following a variety of probabilistic atlases can be created that can be tailored to a particular application.

A broader description and history of atlases of the human brain can be found in [Mazziotta et al., 1995]; [Shattuck et al., 2008].

In this chapter the focus is, however, on non-probabilistic brain atlases that can be aligned with a specific target image. As many individual brain atlases need to be aligned this approach is computationally more expensive than the alignment with a single atlas only. However, accuracy of the estimated target segmentation can be substantially improved based on *subject-specific* atlases created using label fusion (*cf.* Section 2.3).

In the following, five publicly available non-probabilistic brain atlases based on T1-weighted MR images are introduced. The reviewed brain atlases were selected because they are widely used, publicly available, and contain manually annotated anatomical labels for both cortical and non-cortical structures. An overview over key characteristics of these atlases is given in Table 2.

¹<http://www.bic.mni.mcgill.ca/ServicesAtlases/MNI305>

²<http://www.bic.mni.mcgill.ca/ServicesAtlases/ICBM152Lin>

However, next to this selection of brain atlases there are alternatives available. For example, Klein et al. (2009) employed next to the IBSR18 (*cf.* Section 3.2) and LBPA40 (*cf.* Section 3.4) atlases, the CUMC12³ and MGH10³ brain atlas to evaluate 14 established registration algorithms. Furthermore, Klein and Tourville (2012) recently created a detailed protocol to label human cortices. Based on this protocol MR images from 101 healthy participants, with origins in nine different publicly available data sets, were segmented. The expert labels for all images³ (page 25) were obtained by manually editing label maps that were calculated using FreeSurfer (<http://surfer.nmr.mgh.harvard.edu/>, Fischl and Dale (2000); Fischl et al. (2002, 2004)) (FreeSurfer) to correspond with the protocol.

3.1. AAL atlas

The *Automated Anatomical Labeling (AAL) brain atlas*⁴ (Tzourio-Mazoyer et al., 2002) distinguishes 116 ROIs and is based on a single MR template image with a resolution of 1 mm × 1 mm × 1 mm. The template⁵ was obtained by averaging 27 T1w MR scans of a male individual to increase the signal to noise ratio (SNR) (Holmes et al., 1998; Tzourio-Mazoyer et al., 2002). The MR template was then segmented manually into distinct ROIs based on two-dimensional (2D) axial slices. The ROIs were outlined only on every second axial slice (Tzourio-Mazoyer et al., 2002). The atlas segmentations available for download⁴ are, compared to the MR template, of a lower resolution of 2 mm × 2 mm × 2 mm. The AAL atlas was constructed with a focus on measuring activation patterns within the defined ROIs in functional imaging. To avoid missing some structure-related activity due to a too conservative definition of ROIs and to account for inter-subject variability, structures were outlined quite generously, even beyond the GM boundary (Tzourio-Mazoyer et al., 2002).

In total the template is divided into 116 anatomical structures including the cortical gyri but also the hippocampus, sub-cortical structures such as amygdala, caudate nucleus, putamen, pallidum, thalamus and 26 labels subdividing the cerebellum (Tzourio-Mazoyer et al., 2002). Structures in the left and right hemisphere are treated separately. Ventricles were not segmented. The visualisation of the atlas in Figure 7 clearly shows the generous

³available at <http://mindboggle.info/>

⁴<http://www.gin.cnrs.fr/spip.php?article217>

⁵<http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27Highres>

Table 2: Overview of key characteristics of the five selected atlas databases.

Atlas	N	male/ female	fe- age	(mean [min; max])	ROIs	C1	C2	Literature	URL
AAL	1	1/0	young adult	116	no	no		Tzourio-Mazoyer et al. (2002)	AAL-1/2
IBSR18	18	4/14	38 [7; 71]	34	yes	yes	-		IBSR-1
Ham- mers	30	15/15	31* [20; 54]	83	no	no		Hammers et al. (2003, 2007); Gousias et al. (2008)	Hammers-1
LPBA40	40	20/20	29.2 [19; 39]	56	no	no		Shattuck et al. (2008)	LPBA-1
NMM	30 ⁺	10/20	34.3 [15; 90]	134	yes	yes		Marcus et al. (2007); and Tourville (2013)	NMM-1/2/3

*: median; ⁺: atlas consists of 35 images as five subjects were scanned twice

C1: cortical ROIs contain only grey matter tissue.

C2: all brain tissue including ventricles is annotated. *E.g.* AAL, Hammers or LPBA atlases contain not annotated WM.

AAL-1:	http://www.gin.cnrs.fr/spip.php?article217	
AAL-2:	http://www.bic.mni.mcgill.ca/ServicesAtlases/Colin27Highres	
IBSR-1:	http://www.nitrc.org/projects/ibsr	
Hammers-1:	http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Adult	
LPBA-1:	http://www.loni.usc.edu/atlas/atlases/Atlas_Detail.php?atlas_id=12	
NMM-1:	http://Neuromorphometrics.com/	
NMM-2:	http://www.cma.mgh.harvard.edu/manuals/segmentation/	
NMM-3:	http://www.braincolor.org	

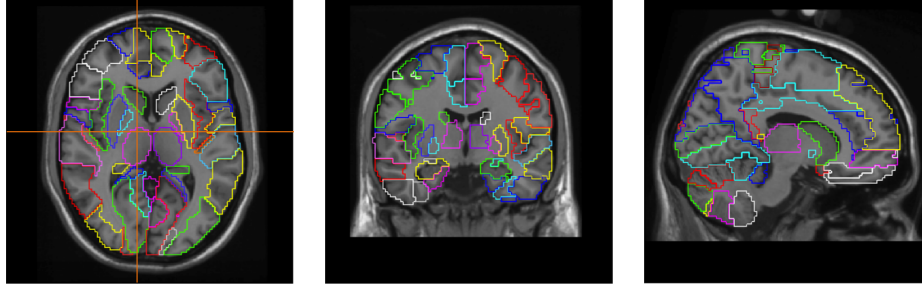


Figure 7: The AAL atlas in axial (left), coronal (middle) and sagittal (right) view view planeplane. T1w MR template with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.

definition of ROIs including more than one tissue type.

3.2. IBSR18 atlas

The *Internet Brain Segmentation Repository (IBSR)* V2.0⁶ atlas consists of T1w MR images acquired from 18 subjects (14 male / 4 female, mean age [min; max] of 14 subjects 38 years [7; 71]⁷). The slice thickness of all images is 1.5 mm. However, the in-plane resolution varies from 0.8371 mm \times 0.8371 mm to 1 mm \times 1 mm. All images were positionally normalised by rotating the images into the Talairach orientation. Expert segmentations of 34 anatomical structures⁸ were created manually.

The annotated structures include for example hippocampus, amygdala, caudate, pallidum, putamen, thalamus, inferior lateral ventricle, lateral ventricle. The cerebellum is split into cerebellum cortex and cerebellum white matter (WM). Structures in the left and right hemisphere are discriminated. Cortical annotations include cortical GM only. An example of a subject of the IBSR atlas is illustrated in Figure 8.

⁶<http://www.nitrc.org/projects/ibsr>

⁷No exact age provided for four subjects.

⁸The readme file describes 39 anatomical structures. In contrast to this description labels for cerebral exterior (left/right), cerebellum exterior (left/right) and amygdala anterior (left/right) are missing while CSF is an additional label not mentioned in the description ($39 - 6 + 1 = 34$).

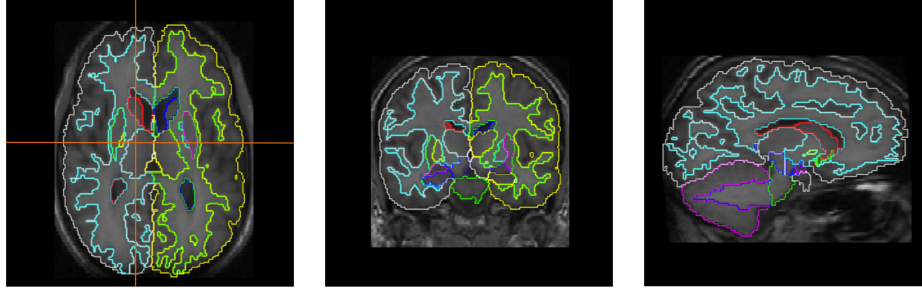


Figure 8: An example of a subject (IBSR01, male, 37 years) of the IBSR atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.

3.3. Hammers atlas

The *Hammers brain atlas*⁹ (Hammers et al., 2003, 2007; Gousias et al., 2008) consists of 30 T1w MR images with corresponding segmentations distinguishing 83 ROIs. The images were acquired from young healthy adults (15 male/female, median age [min; max] of all subjects 31 years [20; 54]) and resliced to an isotropic resolution of $0.9375 \text{ mm} \times 0.9375 \text{ mm} \times 0.9375 \text{ mm}$ (Hammers et al., 2003; Gousias et al., 2008). All images were acquired at 1.5Tesla (T) with the same scanner (Gousias et al., 2008). Originally, Hammers et al. (2003) annotated 49 distinct anatomical regions of 20 subjects. Delineations were done in native space. This work was subsequently extended to 30 subjects and by further subdividing the 49 regions into a total of 83 ROIs (Hammers et al., 2007; Gousias et al., 2008). Following a well defined protocol (Hammers et al., 2003), the MRIs were manually annotated on 2D slices in structure-specific orientations. Subsequent control of difficult cases was conducted by a trained specialist (Hammers et al., 2003). An intra-rater reliability study indicates good reliability (Hammers et al., 2007). Hammers et al. (2003) further created a probabilistic brain atlas by aligning the individual segmentations to the MNI-152² (page 24) template space using Statistical Parametric Mapping (<http://www.fil.ion.ucl.ac.uk/spm/>, Ashburner and Friston (1997, 1999, 2005)) (SPM).

⁹<http://biomedic.doc.ic.ac.uk/brain-development/index.php?n=Main.Adult>

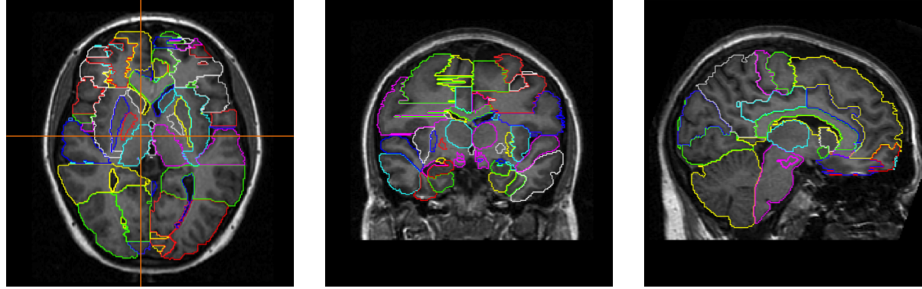


Figure 9: An example of a subject (a01) of the Hammers atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.

Each of the 30 MRIs is divided into 83 ROIs subdividing the cortex, ventricles and non-cortical structures such as the hippocampus, amygdala, caudate nucleus, pallidum, putamen, thalamus. There is no finer separation within the cerebellum. Structures in the left and right hemisphere form separate ROIs. grey matter (GM) and white matter (WM) are combined in one common label for most cortical ROIs (Hammers et al., 2003). An example of a subject of the Hammers atlas is illustrated in Figure 9.

3.4. LPBA40 atlas

The *LONI Probabilistic Brain Atlas (LPBA)*¹⁰ consists of T1w MR images acquired from 40 healthy volunteers (20 male/female, mean age [min; max] of all subjects 29.2 years [19; 39]) (Shattuck et al., 2008). All images were acquired at the same scanner at 1.5T and subsequently *rigidly* aligned to the MNI-305¹ (page 24) (Evans et al., 1993) template space. Images were further resampled to an isotropic resolution of 1 mm × 1 mm × 1 mm (Shattuck et al., 2008). Manual delineation of the images into 56 distinct ROIs was then performed in template space following a detailed protocol (Shattuck et al., 2008). Rater-reliability was assessed. Several versions of a probabilistic atlas were produced by employing different techniques (*e.g.* SPM) to *nonrigidly* align the individual segmentations to the MNI-305 template space (Shattuck et al., 2008). However, the individual 40 annotated non-probabilistic atlases

¹⁰http://www.loni.usc.edu/atlas/Atlas_Detail.php?atlas_id=12

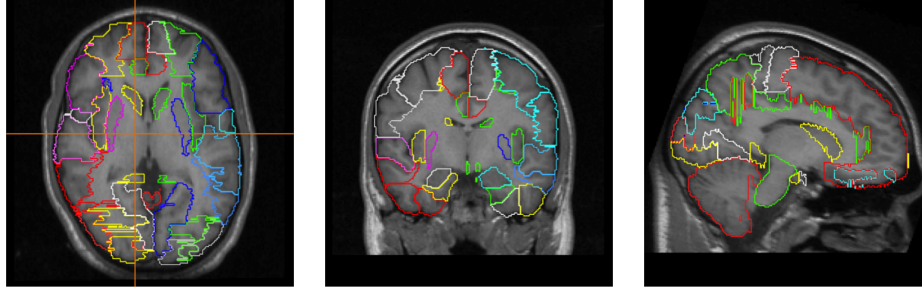


Figure 10: An example of a subject (S01) of the LPBA atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR template with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.

are available.

The atlas distinguishes hippocampus, caudate, putamen, cerebellum and brainstem but mainly parcellates cortical ROIs (Shattuck et al., 2008). Amygdala, pallidum and thalamus are absent. There is no finer separation within the cerebellum. Structures in the left and right hemisphere form individual ROIs. Cortical structures potentially contain WM that lies between sulci or in the vicinity of cortical GM (Shattuck et al., 2008). An example of a subject of the LPBA atlas is illustrated in Figure 10.

3.5. NMM atlas

The *Neuromorphometrics (NMM) brain atlas*¹¹ consists of a continuously growing number of manually annotated T1w MR brain images (Worth and Tourville, 2013). The currently available annotated images can be obtained with an academic license for an annual fee. Recently, a subset of 35 images was used as ‘gold standard’ and was made freely available in the course of the “MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling” (Landman and Warfield, 2012). In the following a description of these 35 images is provided.

The 35 images were taken from the Open Access Series of Imaging Studies (OASIS) database and acquired from 30 individuals (Marcus et al., 2007).

¹¹<http://www.neuromorphometrics.com/>

Repeat scans were acquired of five of the subjects in a second session within 90 days of the original scan (Marcus et al., 2007). All images were acquired at 1.5T at the same scanner (Marcus et al., 2007). Images from the OASIS database were corrected for bias field inhomogeneities and positionally normalised¹². The MR images with corresponding segmentations are resampled to a resolution of 1 mm \times 1 mm \times 1 mm. A description of relevant subsets including information on gender and age is provided in Table 3.

Table 3: Overview of relevant NMM subsets with respective age and gender information. Definition of TRAINING and TEST set as used in the MICCAI 2012 Grand Challenge.

subset	# of images / subjects	gender (# male / # female)	age (mean [min; max])
COMPLETE	35 / 30	10 / 20	34.3 [18; 90]
ATLAS	30 / 30	10 / 20	34.3 [18; 90]
REPEAT	10 / 5	2 / 3	24.6 [20; 29]
TRAINING	15 / 15	5 / 10	23 [19; 34]
TEST	20 / 15	5 / 10	45.7 [18; 90]

The manual segmentation into 138 anatomical structures has been carried out by experts¹³ according to publicly available protocols¹⁴. All manual segmentations were quality controlled by another expert¹² (page 31). As suggested by Landman and Warfield (2012), the small regions “vessel” and “cerebral exterior” were excluded in both the left and the right hemisphere. Thus the atlas comprises effectively 134 ROIs of which 36 are non-cortical and 98 cortical. The non-cortical structures comprise several ROIs such as amygdala, caudate nucleus, hippocampus, pallidum, putamen, thalamus. The cerebellum is subdivided into WM, cerebellum exterior and the cerebellar vermal lobules. Cortical ROIs contain a single tissue type (cortical GM) only. Without any further subdivision, cerebral WM is pooled together in one ROI for the left and right brain hemisphere respectively. The 134 ROIs contain 63 distinct anatomical structures which have symmetric counterparts in their opposite hemisphere, in total 126 paired ROIs. The remaining eight unpaired structures are: 3rd ventricle, 4th ventricle, brain stem, CSF, optic chiasm, cerebellar vermal lobules I-V, cerebellar vermal lobules VI-VII, cerebellar vermal lobules VIII-X. Examples of brain MR images of a young male subject (20 years) and a rather old female subject (90 years) is

¹²<http://www.neuromorphometrics.com/wp-content/uploads/2013/07/DescriptionofLabeledScans.pdf>

¹³provided by Neuromorphometrics, Inc. (<http://Neuromorphometrics.com/>) under academic subscription.

¹⁴<http://www.cma.mgh.harvard.edu/manuals/segmentation/> and <http://www.braincolor.org>

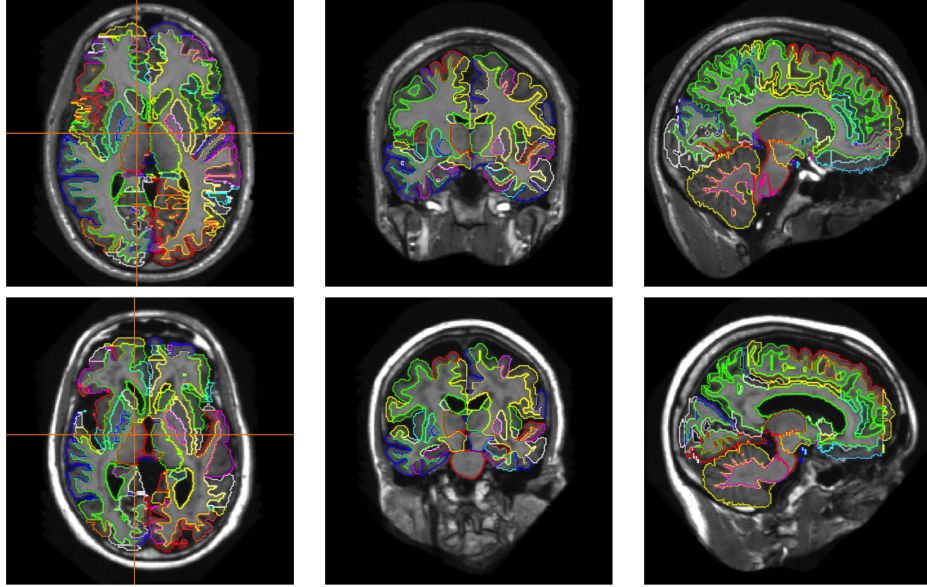


Figure 11: Two examples of subjects of the NMM atlas in axial (left), coronal (middle) and sagittal (right) view plane. T1w MR images of a male subject aged 20 years (OAS1_0285, top) and a female subject aged 90 years (OAS1_0083, bottom) with overlaid segmentation contours in a colour scheme that provides a good contrast between adjacent labels. The cross-hair in the axial view indicates the position of the illustrated coronal and sagittal slice.

illustrated with overlaid manual segmentation outlines in Figure [11](#).

4. Conclusions

In this chapter we have reviewed several approaches for the semantic segmentation of brain MR images. These approaches are nearly always atlas-based segmentation techniques that can fully make use of a-priori knowledge that is encoded in the atlases. The large number of different techniques in this area demonstrates that this is still an active area of research. The performance of current state-of-the-art techniques is starting to approach that of human observers in terms of accuracy. However, the robustness of current approaches is not yet comparable to human observers. This is especially true in cases where pathologies are present in the MRIs that are not present in the atlases.

References

- Akhondi-Asl, A., Warfield, S. K., 2013. Simultaneous truth and performance level estimation through fusion of probabilistic segmentations. *Medical Imaging, IEEE Transactions on* 32 (10), 1840–1852.
- Aljabar, P., Heckemann, R. A., Hammers, A., Hajnal, J. V., Rueckert, D., 2009. Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy. *NeuroImage* 46 (3), 726–738.
- Andersson, J., Jenkinson, M., Smith, S., 2007. Non-linear registration, aka spatial normalisation.
- Artachevarria, X., Munoz Barrutia, A., Ortiz, C. d. S., 2009. Combination strategies in multi-atlas image segmentation: Application to brain MR data. *IEEE Transactions on Medical Imaging* 28 (8), 1266–1277.
- Ashburner, J., Friston, K., 1997. Multimodal Image Coregistration and PartitioningA Unified Framework. *NeuroImage* 6 (3), 209–217.
- Ashburner, J., Friston, K., 1999. Nonlinear spatial normalization using basis functions. *Human Brain Mapping* 7 (4), 254–266.
- Ashburner, J., Friston, K. J., 2005. Unified segmentation. *NeuroImage* 26 (3), 839–851.
- Asman, A., Akhondi-Asl, A., Wang, H., Tustison, N. J., Avants, B. B., Warfield, S. K., Landman, B. A., 2013. MICCAI 2013 Segmentation Algorithms, Theory and Applications (SATA) Challenge Results Summary. URL https://masi.vuse.vanderbilt.edu/workshop2013/images/1/1b/SATA_2013_Proceedings.pdf
- Asman, A. J., Landman, B. A., 2012. Formulating spatially varying performance in the statistical fusion framework. *Medical Imaging, IEEE Transactions on* 31 (6), 1326–1336.
- Asman, A. J., Landman, B. A., 2013. Non-local statistical label fusion for multi-atlas segmentation. *Medical Image Analysis* 17 (2), 194–208.
- Avants, B. B., Epstein, C. L., Grossman, M., Gee, J. C., 2008. Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. *Medical Image Analysis* 12 (1), 26–41.

- Bai, W., Shi, W., Ledig, C., Rueckert, D., 2015. Multi-atlas segmentation with augmented features for cardiac MR images. *Medical Image Analysis* 19 (1), 98–109.
- Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D. B., 2009. Patch-Match: A Randomized Correspondence Algorithm for Structural Image Editing. *ACM Trans. Graph.* 28 (3), 24:1–24:11.
- Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 26 (9), 1124–1137.
- Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 23 (11), 1222–1239.
- Brodmann, K., 1909. *Vergleichende Lokalisationslehre der Großhirnrinde: in ihren Prinzipien dargestellt auf Grund des Zellenbaues*. Verlag von Johann Ambrosius Barth.
- Buades, A., Coll, B., Morel, J., 2005. A Review of Image Denoising Algorithms, with a New One. *Multiscale Modeling & Simulation* 4 (2), 490–530.
- Cardoso, M. J., Clarkson, M. J., Ridgway, G. R., Modat, M., Fox, N. C., Ourselin, S., 2011. LoAd: A locally adaptive cortical segmentation algorithm. *NeuroImage* 56 (3), 1386–1397.
- Cardoso, M. J., Leung, K., Modat, M., Keihaninejad, S., Cash, D., Barnes, J., Fox, N. C., Ourselin, S., 2013. STEPS: Similarity and Truth Estimation for Propagated Segmentations and its application to hippocampal segmentation and brain parcellation. *Medical Image Analysis* 17 (6), 671–684.
- Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P., Marchal, G., 1995. Automated multi-modality image registration based on information theory. *International Conference on Information Processing in Medical Imaging (IPMI)* 3, 263–274.
- Commowick, O., Akhondi-Asl, A., Warfield, S. K., 2012. Estimating A Reference Standard Segmentation With Spatially Varying Performance Parameters: Local MAP STAPLE. *Medical Imaging, IEEE Transactions on* 31 (8), 1593–1606.

- Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., Collins, D. L., 2012. Simultaneous segmentation and grading of anatomical structures for patient’s classification: Application to Alzheimer’s disease. *NeuroImage* 59 (4), 3736–3747.
- Coupé, P., Eskildsen, S. F., Manjón, J. V., Fonov, V. S., Pruessner, J. C., Allard, M., Collins, D. L., 2012. Scoring by nonlocal image patch estimator for early detection of Alzheimer’s disease. *NeuroImage: Clinical* 1 (1), 141–152.
- Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., Collins, D. L., 2010. Nonlocal Patch-Based Label Fusion for Hippocampus Segmentation. *Lecture Notes in Computer Science, MICCAI 2010*, pp. 129–136.
- Coupé, P., Manjón, J. V., Fonov, V., Pruessner, J., Robles, M., Collins, D. L., 2011. Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation. *NeuroImage* 54 (2), 940–954.
- Coupé, P., Yger, P., Prima, S., Hellier, P., Kervrann, C., Barillot, C., 2008. An Optimized Blockwise Nonlocal Means Denoising Filter for 3-D Magnetic Resonance Images. *Medical Imaging, IEEE Transactions on* 27 (4), 425–441.
- Dice, L. R., 1945. Measures of the Amount of Ecologic Association Between Species. *Ecology* 26 (3), 297–302.
- Eritaia, J., Wood, S. J., Stuart, G. W., Bridle, N., Dudgeon, P., Maruff, P., Velakoulis, D., Pantelis, C., 2000. An optimized method for estimating intracranial volume from magnetic resonance images. *Magnetic Resonance in Medicine* 44 (6), 973–977.
- Eskildsen, S. F., Coupé, P., Fonov, V., Manjón, J. V., Leung, K. K., Guizard, N., Wassef, S. N., Østergaard, L. R., Collins, D. L., 2012. BEaST: Brain extraction based on nonlocal segmentation technique. *NeuroImage* 59 (3), 2362–2373.
- Evans, A. C., Collins, D. L., Mills, S. R., Brown, E. D., Kelly, R. L., Peters, T. M., 1993. 3D statistical neuroanatomical models from 305 MRI volumes 3, 1813–1817.
- Fennema-Notestine, C., Ozyurt, I. B., Clark, C. P., Morris, S., Bischoff-Grethe, A., Bondi, M. W., Jernigan, T. L., Fischl, B., Segonne, F., Shattuck, D. W., Leahy, R. M., Rex, D. E., Toga, A. W., Zou, K. H., Brown,

- G. G., 2006. Quantitative evaluation of automated skull-stripping methods applied to contemporary and legacy images: Effects of diagnosis, bias correction, and slice location. *Human Brain Mapping* 27 (2), 99–113.
- Fischl, B., Dale, A. M., 2000. Measuring the thickness of the human cerebral cortex from magnetic resonance images. *Proceedings of the National Academy of Sciences* 97 (20), 11050–11055.
- Fischl, B., Salat, D. H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A. M., 2002. Whole brain segmentation: Automated labeling of neuroanatomical structures in the human brain. *Neuron* 33 (3), 341–355.
- Fischl, B., van der Kouwe, A., Destrieux, C., Halgren, E., Ségonne, F., Salat, D. H., Busa, E., Seidman, L. J., Goldstein, J., Kennedy, D., Caviness, V., Makris, N., Rosen, B., Dale, A. M., 2004. Automatically Parcellating the Human Cerebral Cortex. *Cerebral Cortex* 14 (1), 11–22.
- Freeborough, P. A., Fox, N. C., Kitney, R. I., 1997. Interactive algorithms for the segmentation and quantitation of 3-D MRI brain scans. *Computer Methods and Programs in Biomedicine* 53 (1), 15–25.
- Freund, Y., Schapire, R. E., 1995. A decision-theoretic generalization of on-line learning and an application to boosting. In: *Computational Learning Theory*. Vol. 904 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, pp. 23–37.
- Gerber, S., Tasdizen, T., Fletcher, P. T., Joshi, S., Whitaker, R., 2010. Manifold modeling for brain population analysis. *Medical Image Analysis* 14 (5), 643–653, special Issue on the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009.
- Gousias, I. S., Rueckert, D., Heckemann, R. A., Dyet, L. E., Boardman, J. P., Edwards, A. D., Hammers, A., 2008. Automatic segmentation of brain MRIs of 2-year-olds into 83 regions of interest. *NeuroImage* 40 (2), 672–684.
- Greig, D. M., Porteous, B. T., Seheult, A. H., 1989. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society. Series B (Methodological)* 51 (2), 271–279.

- Hamm, J., Ye, D. H., Verma, R., Davatzikos, C., 2010. Gram: A framework for geodesic registration on anatomical manifolds. *Medical Image Analysis* 14 (5), 633–642, special Issue on the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI) 2009.
- Hammers, A., Allom, R., Koepp, M., Free, S. L., Myers, R., Lemieux, L., Mitchell, T. N., Brooks, D. J., Duncan, J. S., 2003. Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. *Human Brain Mapping* 19 (4), 224–247.
- Hammers, A., Chen, C.-H., Lemieux, L., Allom, R., Vossos, S., Free, S. L., Myers, R., Brooks, D. J., Duncan, J. S., Koepp, M. J., 2007. Statistical neuroanatomy of the human inferior frontal gyrus and probabilistic atlas in a standard stereotaxic space. *Human Brain Mapping* 28 (1), 34–48.
- Han, X., Hoogeman, M. S., Levendag, P. C., Hibbard, L. S., Teguh, D. N., Voet, P., Cowen, A. C., Wolf, T. K., 2008. Atlas-Based Auto-segmentation of Head and Neck CT Images 5242, 434–441.
- Heckemann, R. A., Hajnal, J. V., Aljabar, P., Rueckert, D., Hammers, A., 2006. Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. *NeuroImage* 33 (1), 115–126.
- Heckemann, R. A., Keihaninejad, S., Aljabar, P., Gray, K. R., Nielsen, C., Rueckert, D., Hajnal, J. V., Hammers, A., The AD Neuroimaging Initiative, 2011. Automatic morphometry in Alzheimer’s disease and mild cognitive impairment. *NeuroImage* 56 (4), 2024–2037.
- Heckemann, R. A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J. V., Hammers, A., 2010. Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. *NeuroImage* 51 (1), 221–227.
- Holmes, C. J., Hoge, R., Collins, L., Woods, R., Toga, A. W., Evans, A. C., 1998. Enhancement of MR Images Using Registration for Signal Averaging. *Journal of Computer Assisted Tomography* 22 (2), 324–333.
- Kass, M., Witkin, A., Terzopoulos, D., 1988. Snakes: Active contour models. *International Journal of Computer Vision* 1 (4), 321–331.
- Katkovnik, V., Foi, A., Egiazarian, K., Astola, J., 2010. From Local Kernel to Nonlocal Multiple-Model Image Denoising. *International Journal of Computer Vision* 86 (1), 1–32.

- Kittler, J., Hatef, M., Duin, R. P. W., Matas, J., 1998. On combining classifiers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 20 (3), 226–239.
- Klein, A., Andersson, J., Ardekani, B. A., Ashburner, J., Avants, B., Chiang, M.-C., Christensen, G. E., Collins, D. L., Gee, J., Hellier, P., Song, J. H., Jenkinson, M., Lepage, C., Rueckert, D., Thompson, P., Vercauteren, T., Woods, R. P., Mann, J. J., Parsey, R. V., 2009. Evaluation of 14 non-linear deformation algorithms applied to human brain MRI registration. *NeuroImage* 46 (3), 786–802.
- Klein, A., Tourville, J., 2012. 101 labeled brain images and a consistent human cortical labeling protocol. *Frontiers in Neuroscience* 6 (171), 1–12.
- Klein, S., van der Heide, U. A., Raaymakers, B. W., Kotte, A. N. T. J., Staring, M., Pluim, J. P. W., 2007. Segmentation of the prostate in mr images by atlas matching. *Biomedical Imaging: From Nano to Macro (ISBI). IEEE International Symposium on*, 1300–1303.
- Kuncheva, L. I., 2004. *Combining Pattern Classifiers: Methods and Algorithms*. John Wiley & Sons, Inc.
- Landman, B., Warfield, S. K., 2012. MICCAI 2012 Grand Challenge and Workshop on Multi-Atlas Labeling.
URL https://masi.vuse.vanderbilt.edu/workshop2012/index.php/Main_Page
- Landman, B. A., Asman, A. J., Scoggins, A. G., Bogovic, J. A., Xing, F., Prince, J. L., 2012. Robust statistical fusion of image labels. *IEEE Transactions on Medical Imaging* 31 (2), 512–522.
- Ledig, C., Heckemann, R. A., Makropoulos, A., Hammers, A., Lötjönen, J., Menon, D., Rueckert, D., 2015. Robust whole-brain segmentation: Application to traumatic brain injury. *Medical Image Analysis* 21 (1), 40–58.
- Leung, K. K., Barnes, J., Modat, M., Ridgway, G. R., Bartlett, J. W., Fox, N. C., Ourselin, S., 2011. Brain maps: An automated, accurate and robust brain extraction technique using a template library. *NeuroImage* 55 (3), 1091–1108.
- Lötjönen, J. M., Wolz, R., Koikkalainen, J. R., Thurfjell, L., Waldemar, G., Soininen, H., Rueckert, D., 2010. Fast and robust multi-atlas segmentation of brain magnetic resonance images. *NeuroImage* 49 (3), 2352–2365.

- Manjón, J. V., Eskildsen, S. F., Coupé, P., Romero, J. E., Collins, D. L., 2014. Nonlocal Intracranial Cavity Extraction. *International Journal of Biomedical Imaging* 2014 (Article ID: 820205), 11 pages.
- Marcus, D. S., Wang, T. H., Parker, J., Csernansky, J. G., Morris, J. C., Buckner, R. L., 2007. Open Access Series of Imaging Studies (OASIS): Cross-sectional MRI Data in Young, Middle Aged, Nondemented, and Demented Older Adults. *Journal of Cognitive Neuroscience* 19 (9), 1498–1507.
- Mazziotta, J. C., Toga, A. W., Evans, A., Fox, P., Lancaster, J., 1995. A Probabilistic Atlas of the Human Brain: Theory and Rationale for Its Development: The International Consortium for Brain Mapping (ICBM). *NeuroImage* 2 (2, Part A), 89–101.
- Modat, M., Ridgway, G. R., Taylor, Z. A., Lehmann, M., Barnes, J., Hawkes, D. J., Fox, N. C., Ourselin, S., 2010. Fast free-form deformation using graphics processing units. *Computer Methods and Programs in Biomedicine* 98 (3), 278–284.
- Roche, A., Malandain, G., Ayache, N., 1999. Unifying Maximum Likelihood Approaches in Medical Image Registration. Tech. Rep. RR-3741.
- Rohlfing, T., Brandt, R., Menzel, R., Maurer Jr., C. R., 2004. Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains. *NeuroImage* 21 (4), 1428–1442.
- Rousseau, F., Habas, P. A., Studholme, C., 2011. A supervised patch-based approach for human brain labeling. *IEEE Transactions on Medical Imaging* 30 (10), 1852–1862.
- Rueckert, D., Sonoda, L. I., Hayes, C., Hill, D. L. G., Leach, M. O. and Hawkes, D. J., 1999. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging* 18 (8), 712–721.
- Sabuncu, M. R., Yeo, B. T. T., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging* 29 (10), 1714–1729.
- Sandor, S., Leahy, R., 1997. Surface-based labeling of cortical anatomy using a deformable atlas. *Medical Imaging, IEEE Transactions on* 16 (1), 41–54.

- Ségonne, F., Dale, A. M., Busa, E., Glessner, M., Salat, D., Hahn, H. K., Fischl, B., 2004. A hybrid approach to the skull stripping problem in MRI. *NeuroImage* 22 (3), 1060–1075.
- Shattuck, D. W., Mirza, M., Adisetiyo, V., Hojatkashani, C., Salamon, G., Narr, K. L., Poldrack, R. A., Bilder, R. M., Toga, A. W., 2008. Construction of a 3D probabilistic atlas of human cortical structures. *NeuroImage* 39 (3), 1064–1080.
- Shi, W., Lombaert, H., Bai, W., Ledig, C., Zhuang, X., Marvao, A., Dawes, T., O’Regan, D., Rueckert, D., 2014. Multi-atlas Spectral PatchMatch: Application to cardiac image segmentation. accepted at MICCAI.
- Smith, S. M., 2002. Fast robust automated brain extraction. *Human Brain Mapping* 17 (3), 143–155.
- Song, Z., Tustison, N., Avants, B., Gee, J. C., 2006. Integrated Graph Cuts for Brain MRI Segmentation 4191, 831–838.
- Sotiras, A., Davatzikos, C., Paragios, N., 2013. Deformable Medical Image Registration: A Survey. *Medical Imaging, IEEE Transactions on* 32 (7), 1153–1190.
- Studholme, C., Hill, D. L. G., Hawkes, D. J., 1999. An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition* 32 (1), 71–86.
- Ta, V.-T., Giraud, R., Collins, D. L., Coupé, P., 2014. Optimized Patch-Match for Near Real Time and Accurate Label Fusion 8675, 398–406.
- Thirion, J.-P., 1998. Image matching as a diffusion process: an analogy with maxwell’s demons. *Medical Image Analysis* 2 (3), 243–260.
- Tong, T., Wolz, R., Coupé, P., Hajnal, J. V., Rueckert, D., 2013. Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling. *NeuroImage* 76 (0), 11–23.
- Tzourio-Mazoyer, N., Landeau, B., Papathanassiou, D., Crivello, F., Etard, O., Delcroix, N., Mazoyer, B., Joliot, M., 2002. Automated Anatomical Labeling of Activations in SPM Using a Macroscopic Anatomical Parcelation of the MNI MRI Single-Subject Brain. *NeuroImage* 15 (1), 273–289.
- van der Lijn, F., den Heijer, T., Breteler, M. M. B., Niessen, W. J., 2008. Hippocampus segmentation in MR images using atlas registration, voxel classification, and graph cuts. *NeuroImage* 43 (4), 708 – 720.

- Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P., 1999. Automated model-based tissue classification of MR images of the brain. *IEEE Transactions on Medical Imaging* 18 (10), 897–908.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2009. Diffeomorphic demons: Efficient non-parametric image registration. *NeuroImage* 45 (1, Supplement 1), S61–S72.
- Viola, P., Wells III, W. M., 1997. Alignment by maximization of mutual information. *International Journal of Computer Vision* 24 (2), 137–154.
- Vovk, U., Pernuš, F., Likar, B., 2007. A Review of Methods for Correction of Intensity Inhomogeneity in MRI. *Medical Imaging, IEEE Transactions on* 26 (3), 405–421.
- Wang, H., Das, S. R., Suh, J. W., Altinay, M., Pluta, J., Craige, C., Avants, B. B., Yushkevich, P. A., 2011. A Learning-Based Wrapper Method to Correct Systematic Errors in Automatic Image Segmentation: Consistently Improved Performance in Hippocampus, Cortex and Brain. *NeuroImage* 55 (3), 968–985.
- Wang, H., Suh, J. W., Das, S. R., Pluta, J., Craige, C., Yushkevich, P. A., 2013. Multi-atlas segmentation with joint label fusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (3), 611–623.
- Warfield, S. K., Zou, K. H., Wells, W. M., 2004. Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. *IEEE Transactions on Medical Imaging* 23 (7), 903–921.
- Wells III, W. M., Grimson, W. E. L., Kikinis, R., Jolesz, F. A., 1996. Adaptive segmentation of MRI data. *IEEE Transactions on Medical Imaging* 15 (4), 429–442.
- Wolz, R., Aljabar, P., Hajnal, J. V., Hammers, A., Rueckert, D., 2010a. LEAP: Learning embeddings for atlas propagation. *NeuroImage* 49 (2), 1316–1325.
- Wolz, R., Aljabar, P., Rueckert, D., Heckemann, R. A., Hammers, A., 2009. Segmentation of subcortical structures and the hippocampus in brain MRI using graph-cuts and subject-specific a-priori information. *Biomedical Imaging: From Nano to Macro (ISBI). IEEE International Symposium on*, 470–473.

- Wolz, R., Heckemann, R. A., Aljabar, P., Hajnal, J. V., Hammers, A., Lötjönen, J., Rueckert, D., 2010b. Measurement of hippocampal atrophy using 4D graph-cut segmentation: Application to ADNI. *NeuroImage* 52 (1), 109–118.
- Worth, A., Tourville, J., 2013. An ever-improving model of the structure of the living human. *Proceedings of 10th Annual World Congress of Society for Brain Mapping and Therapeutics (SBMT)*, 20.
 URL <http://www.worldbrainmapping.org/docs/2013-Oral%20Poster%20Abstracts.pdf>
- Wu, M., Rosano, C., Lopez-Garcia, P., Carter, C. S., Aizenstein, H. J., 2007. Optimum template selection for atlas-based segmentation. *NeuroImage* 34 (4), 1612–1618.
- Xu, L., Krzyzak, A., Suen, C. Y., 1992. Methods of combining multiple classifiers and their applications to handwriting recognition. *Systems, Man and Cybernetics, IEEE Transactions on* 22 (3), 418–435.
- Zhang, J., 1992. The Mean Field Theory in EM Procedures for Markov Random Fields. *IEEE Transactions on Signal Processing* 40 (10), 2570–2583.
- Zhang, Y., Brady, M., Smith, S., 2001. Segmentation of brain MR images through a hidden Markov random field model and the expectation maximization algorithm. *IEEE Transactions on Medical Imaging* 20 (1), 45–57.
- Zitová, B., Flusser, J., 2003. Image registration methods: a survey. *Image and Vision Computing* 21 (11), 977–1000.