HIPS: A new hippocampus subfield segmentation method

José E. Romero, Pierrick Coupé, José V. Manjón

HIPS: A new hippocampus subfield segmentation method

José E. Romero[1], Pierrick Coupe[2,3], José V. Manjón[1]

[1]Instituto de Aplicaciones de las Tecnologías de la Información y de las Comunicaciones Avanzadas (ITACA), Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, España.
[2] Univ. Bordeaux, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.
[3] CNRS, LaBRI, UMR 5800, PICTURA, F-33400 Talence, France.

Abstract. The importance of the hippocampus in the study of several neurodegenerative diseases such as Alzheimer's disease makes it a structure of great interest in neuroimaging. However, few segmentation methods have been proposed to measure its subfields due to its complex structure and the lack of high resolution magnetic resonance (MR) data. In this work, we present a new pipeline for automatic hippocampus subfield segmentation using two available hippocampus subfield delineation protocols that can work with both high and standard resolution data. The proposed method is based on multi-atlas label fusion technology that benefits from a novel multi-contrast patch match search process (using high resolution T1-weighted and T2-weighted images). The proposed method also includes as post-processing a new neural network-based error correction step to minimize systematic segmentation errors. The method has been evaluated on both high and standard resolution images and compared to other state-of-the-art methods showing better results in terms of accuracy and execution time.

1 Introduction

The hippocampus (HC) is a small bilateral brain structure located in the medial temporal lobe at both sides of the brainstem near to the cerebellum. Its name comes from its similarity to the sea-horse. Starting from the upper end at the hippocampal sulcus we find the dentate gyrus (DG) followed by the Cornu Ammonis (CA) which is subdivided in four consecutive parts (CA4 to CA1) and the Subiculum at the bottom end. The CA is also structured in six layers called strata. These layers are the Stratum oriens (SO), Stratum pyramidale (SP), Stratum lucidum (SLU), Stratum radiatum (SR), Stratum lacunosum (SL) and the Stratum molecuare (SM).

HC is involved in many brain functions such as memory and spatial reasoning (Milner et al., 1958; Schmajuk 1990; Hafting et al., 2005). Several studies showed that it has an important role in many neurodegenerative diseases such as Alzheimer's disease (AD) (Braak et al., 1991) or schizophrenia (Altshuler et al., 1998). The study of the hippocampus volume is of great interest as it is a valuable tool for follow-up and treatment adjustment (Jack et al., 2000; Jack et al., 2005; Dickerson and Sperling, 2005). However, the HC anatomy is complex and variable, and the limits between different subfields have been described in the neuroanatomy literature using cytoarchitectonic features that require histological staining and microscopic resolution to visualize (Insausti and Amaral, 2004).

Due to the key importance of this structure, several segmentation methods and protocols have been developed (Barnes et al., 2008; Collins el al., 2010; Coupe et al., 2011). However, one of the main problems to advance in this field was the disparity of HC definitions and the lack of manually labelled cases. Recently, a harmonized full hippocampus protocol has been proposed (jointly with 120 1 mm^3 resolution manually segmented examples) which will become the common reference for the development and comparison of new segmentation methods (Boccardi et al., 2015). Classically, due to the limitations in MR image resolution, the studies were restricted to consider the hippocampus as a single structure (Chupin et al., 2009). Even though the analysis of the whole hippocampus has been shown to be a good approach to study AD, some ex-vivo studies revealed that normal aging and AD affects the subfields differently during the lifespan (Braak et al., 1991).

Currently, many HC subfield segmentation protocols have been developed as a response to the advances in MR sequences that allow acquiring high resolution images making possible to divide the hippocampus into its constituent parts. However, there is still little consensus between the different HC subfield protocols as shown in (Yushkevich et al., 2015a) where 21 delineation protocols were compared. Some of these protocols have been used to create anatomically labeled MRI datasets which are a fundamental resource to develop new segmentation methods. For example, 9.4 T ultra-high resolution ex-vivo images were used to create an anatomical atlas (Yushkevich et al., 2009) including the CA1, CA2-3, the DG and the vestigial hippocampal sulcus obtained by manual delineation. In 2013, Winterburn presented a new in-vivo high resolution atlas (Winterburn et al., 2013) to divide the hippocampus in five different subregions: CA1, CA2-3, CA4/DG, Stratum and Subiculum (jointly with 5 manually segmented examples, we call this the Winterburn dataset). Later in 2015, Kulaga-Yoskovitz developed another segmentation protocol (Kulaga-Yoskovitz et al., 2015) consisting of three structures: CA1-3, CA4/DG and Subiculum (jointly with 25 manually segmented examples, we call this the Kulaga-Yoskovitz dataset).

To conduct volumetric studies and apply these delineation protocols, automatic segmentation tools are necessary. It is well known that manual delineation of a new case represents an issue in terms of reproducibility. It is also extremely time consuming as well as it has a high economic cost (it can take from 10 to 20 hours of an expert rater time per subject to manually segment the hippocampus subfields (Iglesias et al., 2015)). Since manual segmentation is not an affordable option, several automatic methods have been developed in the last years. One of the first HC subfield segmentation methods was proposed by Van Leemput (Van Leemput et al., 2009) using a generative model of the hippocampus region. This model is produced using a mesh-based probabilistic atlas containing information about where the anatomical labels are most likely to occur. The probabilistic atlas is learned from a set of ultra high resolution training images. Recently, Iglesias (Inglesias et al., 2015) continued this work and improved the model by using a more accurate atlas generated from ultra-high resolution ex-vivo MR images and also using multi-contrast data. Pipitone proposed a multi-atlas-based method (Pipitone et al., 2014) using T2w

images intended to segment a considerable large dataset (targets) using a few manually labeled cases (atlases). However, this method, named MAGeT (Chakravrty et al., 2013) has a high temporal cost. In 2015, Yushkevich proposed another method (Yushkevich et al., 2015b) using T2w images where a multi-atlas approach is combined with a similarity-weighted voting and a boosting based error correction. Unfortunately, this method takes hours to produce a segmentation due to the exhaustive use of non-linear registrations as in the case of MAGeT. Recently, in 2016, Caldairou presented a new hybrid method (Caldairou et al., 2016) where a set of training subjects are non-linearly registered to the test case. Then, using patch-correspondences, a surface mesh is generated from the manual labels. These patch correspondences are re-computed for each mesh vertex minimizing the error to adjust a deformable model to the case to be segmented.

In this paper, we propose a new patch-based segmentation method which has been validated using two hippocampus subfield segmentation protocols with publically available datasets. Our method uses an adaptation of MOPAL (Romero et al., 2016), a multi-contrast version of a patch matching segmentation method OPAL (Giraud et al., 2016) to produce fast and accurate segmentations. The presented method works using high resolution (0.5x0.5x0.5 mm^3) T1w and T2w images. It also works on standard resolution images as well as single T1w or single T2w images. During our validation, we show that the proposed approach performs well also on mono-contrast T1w and T2w images as well as when using standard resolution images upsampled using the LASR (Manjón et al., 2010a; Coupe et al., 2013) superresolution method. Our method also includes a new error corrector post processing step based on the use of a boosted ensemble of neural networks is proposed to minimize systematic segmentation errors at post-processing.

2 Material and methods

2.1 Image data

In this work, we have used two different datasets corresponding to two manual labeling hippocampus subfield segmentation protocols, both with high resolution (HR) T1w and T2w MR images. An example of these images and their manual labels can be seen in Figure 1.

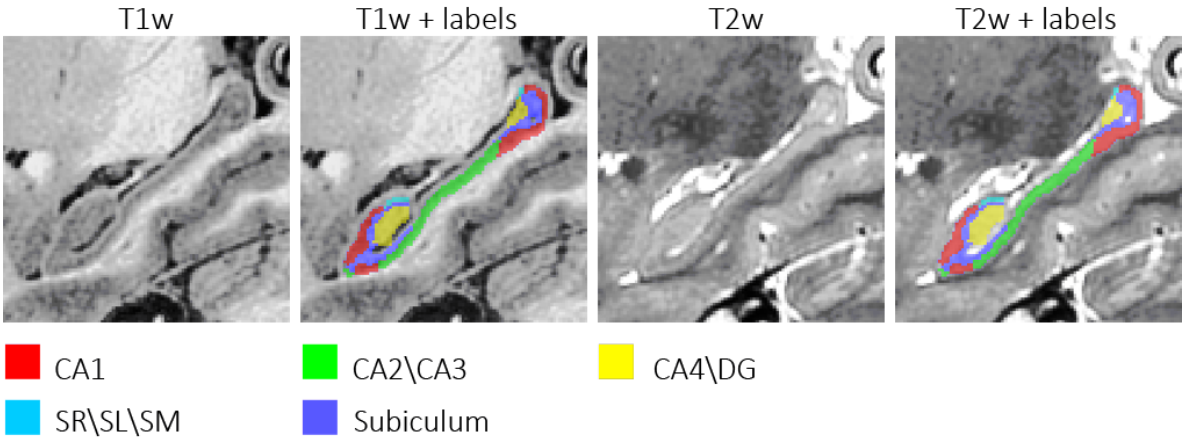
Kulaga-Yoskovitz dataset

This dataset includes 25 subjects from a public repository (http://www.nitrc.org/projects/mni-hisub25) (31 ± 7 yrs, 12 males, 13 females) with manually-drawn labels dividing the HC in three parts (CA1-3, DG-CA4 and Subiculum). MR data from each subject consist of an isotropic 3D-MPRAGE T1-weighted (0.6 mm^3) and anisotropic 2D T2-weighted TSE images (0.4×0.4×2 mm^3). Images underwent automated correction for intensity non-uniformity, intensity standardization and were linearly registered to the MNI152 space. T1w and T2w images were resampled to a resolution of 0.4 mm^3. To reduce interpolation artifacts, the T2w data was upsampled using a non-local superresolution method (Manjón et al., 2010a). For more details about the labeling protocol see the original paper (Kulaga-Yoskovitz et al., 2015).

Winterburn dataset

This dataset contains 5 subjects with 0.3x0.3x0.3 mm^3 high resolution T1-weighted and T2-weighted images obtained by 2x interpolation of 0.6x0.6x0.6 mm^3 acquisitions. and their corresponding manual segmentations. The HR images are publicly available at the CoBrALab website (http://cobralab.ca/atlases). These MR images were taken from 5 healthy volunteers (2 males, 3 females, aged 29–57). High-resolution T1-weighted images were acquired using the 3D inversion-prepared fast spoiled gradient-recalled echo acquisition (TE/TR=4.3 ms/9.2 ms, TI=650 ms, α=8°, 2-NEX and isotropic resolution of 0.6 mm^3). High-resolution T2-weighted images were acquired using the 3D fast spin echo acquisition, FSE-CUBE (TE/TR=95.3 ms/2500 ms, ETL=100 ms, 2NEX, and isotropic resolution of 0.6 mm^3). Reconstruction filters, ZIPX2 and ZIP512, were also used resulting in a final isotropic 0.3 mm^3 dimension voxels. The hippocampi and each of their subfields were segmented manually by an expert rater including 5 labels (CA1, CA2/3, CA4/DG, (SR/SL/SM), and subiculum). For more details about the labeling protocol see the original paper (Winterburn et al., 2013).

Winterburn



Kulaga-Yoskovitz

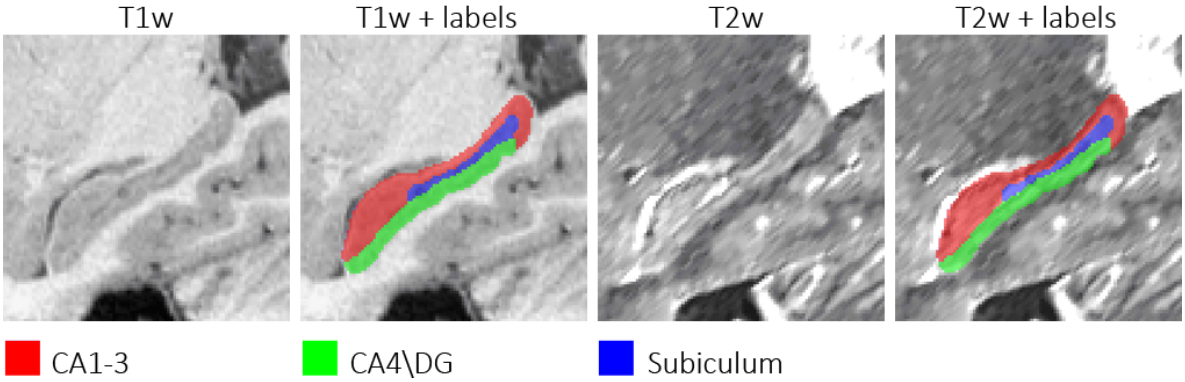


Figure 1: Examples from Winterburn and Kulaga-Yoskovitz datasets showing T1w, T2w and manual segmentations.

2.2 Proposed method

2.2.1 Preprocessing

The images used for this work were preprocessed to locate them in a common intensity and coordinate space. For this we applied the following steps: Denoising using the Spatially Adaptive Non-local Means Filter(Manjón et al., 2010b). This filter is able to automatically deal with both stationary and spatially varying noise levels. Intensity inhomogeneity correction using the N4 bias field correction (Tustison et al., 2010).

The images were moved to a common coordinate space to better match the anatomy between library subjects and the case to be segmented. To achieve this, the images were first linearly registered to the Montreal Neurological Institute (MNI) space by applying the Advanced Normalization Tools (ANTs) (Avants et al., 2009). This registration was estimated using the T1w MNI152 template and the T1w images, and applied to both T1w and T2w images (a rigid transformation from T2w to T1w was first estimated and later concatenated with T1w transformation to perform a single interpolation step when registering both T1w and T2w images). Note that when processing HR images (voxel size smaller than 1x1x1 mm^3) a HR MNI152 template version was used (0.5x0.5x0.5 mm^3) instead of the classical one (1x1x1 mm^3 resolution). As we will describe later, the segmentation is performed always at 0.5x0.5x0.5 mm^3 resolution.

The images were intensity normalized so brain tissues have similar intensity levels across all the subjects of the library. For this purpose, we applied a histogram matching method (Nyúl and Udupa, 1999). Then, to reduce the memory requirements and the computational cost, the images were cropped around HC area. For this procedure a bounding box surrounding the hippocampus was calculated (using a margin of 5 voxels in each direction) in the MNI space from the manual segmentations to ensure that all the manual segmentations were included in this bounding box.

If the resolution of a new case to be segmented is lower than 0.5x0.5x0.5 mm^3, the cropped data must be first upsampled to produce HR 0.5x0.5x0.5 mm^3 resolution data. This is performed using a patch-based super-resolution technique called LASR (Coupé et al., 2013). If HR data is used as input this step is skipped.

To achieve a better match between the different subjects anatomy, a non-linear deformation was estimated between the cropped regions of every subject and the HC cropped MNI125 template. For this, we used a multi-contrast registration framework using T1w and T2w images having both equal weights. The non-linear deformation is estimated using the Advanced Normalization Tools (ANTs) (Avants et al., 2009) using cross correlation metric and pyramidal framework at 8x 4x 2x 1x

scales and 200, 200, 200 and 0 iterations at each scale. This registration process introduces blurring due to the interpolation used to apply the transformations. This has a negative impact in the segmentation step. For this reason, to enhance the images, we sharpened both T1w and T2w images by adding the laplacian of each image.

2.2.2 Library construction

The proposed method requires the construction of a training library of manually annotated images located in the same intensity and geometrical space that new case to be segmented. To this end, we constructed a training atlas library by preprocessing each atlas as previously described. Additionally, after the cropping step, the images were left-right flipped to double up the number of cases. To segment a new case, it is preprocessed in the same way than the library cases. Once the preprocessing is done, we have a set of cropped images (and their corresponding manual segmentations) and their non-linear transformations to the cropped MNI space and a cropped case to be segmented and its non-linear transformation to the cropped MNI space. Then, we generate a subject specific library by concatenating the direct non-linear transformation of every library case with the inverse non-linear transformations of the case to be segmented. This way we move the entire library to the new case space (note that previous to the non-linear registration, an affine registration to the MNI space was done so the segmentations is performed in the linear MNI space) as done in a previous work (Romero et al., 2017).

2.2.3 Labeling

Multi-contrast Optimized PatchMatch (MOPAL)

Our segmentation method is based on the idea of non-local patch-based label fusion technique (Coupe et al., 2011) were patches of the subject to be segmented are compared with patches of the training library to look for similar patterns within a defined search volume to assign the proper label *v* as can be seen in equations 1.

$$v(x_i) = \frac{\sum_{s=1}^{N} \sum_{j \in V_i} w(x_i, x_{s,j}) y_{s,j}}{\sum_{s=1}^{N} \sum_{j \in V_i} w(x_i, x_{s,j})} \quad (1)$$

where V_i corresponds to the search area, N is the number of subjects in the template library, $y_{s,j}$ is a possible label from the voxel $x_{s,j}$ at the position j in the library subject s and $w(x_i, x_{s,j})$ is the patch similarity defined as:

$$w\left(x_i, x_{s,j}\right) = exp^{\frac{-D_{i,j,s}}{h^2}} \quad (2)$$

$$D_{i,j,s} = \left\| P(x_i) - P(x_{s,j}) \right\|_2^2 \quad (3)$$

where $P(x_i)$ is the patch centered at x_i, $P(x_{s,j})$ the patch centered at x_j in the templates and $||.||_2$ is the normalized L2 norm (normalized by the number of elements) calculated from the distance

between each pair of voxels from both patches $P(x_i)$ and $P(x_{s,j})$. h is a normalization parameter that is estimated from the minimum of all patch distances within the search area.

However, exhaustive patch comparison process is very time consuming (even in reduced neighborhoods). To reduce the computational burden of this process, we have used an adaptation of the OPAL method (Giraud et al., 2016) that is a 3D adaptation of the PatchMatch technique (Barnes et al., 2009). This technique is an efficient algorithm to finds correspondences between patches of two images (in our case a target image A and a library of image subjects B) using the concept of Approximated Nearest Neighbor Field (ANNF). It consists of three steps: First, an initialization is made and random correspondences are assigned to each patch A using patches randomly selected from the library B. Then, a propagation is done based in the hypothesis that if a patch x from A has a good match with a patch y of B, then adjacent patches of x will probably have good matches in adjacent patches of y. Finally a restricted local random search is also done to avoid local minima. The second and third steps are repeated iteratively to improve the matches. We refer the reader to the original paper for more details (Giraud et al., 2016).

Multi-scale label fusion

In the original OPAL method, label probability maps are estimated using two independent processes with two different patch sizes to account for multi-scale features. This maps are uniformly averaged (process called late fusion) to obtain the final probability map. In this new variant of the algorithm, that we called MOPAL, we use label dependent multi-scale mixing coefficients α to balance the different scale contributions per label (eq. 4).

$$p(l) = \alpha(l)p_1\,(l) + (1 - \alpha(l))p_2(l) \tag{4}$$

where $p_1(l)$ is the probability map corresponding to scale 1 for label *l*, $p_2(l)$ the probability map corresponding to scale 2 for label l, $p(l)$ is the resultant combined probability map for label l and $\alpha(l) \in [0,1]$ is the mixing coefficient for label l. These coefficients are optimized using a gradient descent technique. For this optimization, we minimized the segmentation error calculated as 1 - DICE coefficient as done in (Romero et al., 2017).

Multi-contrast patch similarity

The patch matching process can benefit from the use of multiple contrast data (Xiao et al., 2015; Fisher and Oliver, 1995). As we work with T1w and T2w images we have improved the matching process by using a multi-contrast similarity. This modified similarity measure takes into account information derived from two channels, T1w and T2w images, in order to compute patch correspondences in a robust manner. OPAL estimates the quality of a match by computing a distance as the sum of squared differences (SSD) (eq. 3). Our proposed multi-contrast similarity consists on a SSD-based semi-norm (one SSD per channel) that takes into account the discriminative power of each channel locally. We called this multi-contrast semi norm (MSN):

$$MSN_{i,j,s} = \frac{\left\| P(A_i) - P(A'_{s,j}) \right\|_2^2 \cdot \left\| P(B_i) - P(B'_{s,j}) \right\|_2^2}{M\left(\left\| P(A_i) - P(A'_{s,j}) \right\|_2^2 + \left\| P(B_i) - P(B'_{s,j}) \right\|_2^2 \right)} \quad (5)$$

Where A and B represent the target image for T1w and T2w, A' and B' represent the libraries for T1w and T2w respectively, P(Ai) ∈ A is a patch form image A centered on the coordinates i, P(Bj) ∈ B is a patch from image B centered on the coordinates j, s represents the subject number and M is the number of voxels per patch.

2.2.4 Systematic error correction

Any segmentation method is subject to random and systematic errors. The former can be typically minimized using aggregation techniques leading to the reduction of classification error standard deviation. This is the case of MOPAL where a high number of votes per voxel are used to reduce the classification error. Unfortunately, systematic errors cannot be reduced using this strategy since they are not random. However, this systematic bias can be learned to correct/calibrate the segmentation output. In 2011, Wang et al., realized about this issue and proposed his well-known SegAdapter method. This method is based on the use of an Adaboost classifier which locally learns and corrects systematic errors using spatial (coordinates) and intensity information (patches).

Inspired by the pioneer work of Wang et al., we propose an error corrector method based on a patch-based ensemble of neural networks (PEC for Patch-based Ensemble Corrector) to increase the segmentation accuracy by reducing the systematic errors produced by our segmentation method. The neural network ensemble is trained with image patches of sizes 3x3x3 voxels (fully sampled) and 7x7x7 voxels (subsampled by skipping two voxels at each dimension) from T1w, T2w images, the automatic segmentations, a Euclidean distance value, and their x,y and z coordinates in MNI space. The Euclidean distance map was calculated for the whole hippocampus as the lower distance in voxels from each point to the hippocampus edge. This results in a feature vector of 166 features that are mapped to a patch of manual segmentations of size 3x3x3 voxels. We used an overcomplete patch-based classification as proposed in a previous work (Manjón et al., 2016). We also used a multilayer perceptron with two hidden layers of size 83 and 55 neurons resulting in a network with a topology of 166x83x55x27 neurons. An ensemble of 10 neural networks was trained using a boosting strategy. Each new network was trained with a different subset of data which was selected by giving a higher probability of appearance to the samples that were misclassified in the previous ensemble.

Differently from SegAdapter method, we used 2 patches per location which allows us to be point specific while having also context information and topological information from the geodesic map. Our patch-based overcomplete correction scheme also increases the number of votes making the estimation more robust. Finally, although the number of networks used in the ensemble (M=10) may seem low compared to the 500 trees used in SegAdapter method, it has to be noted that neural networks are much stronger classifiers than trees. Figure 2 shows an example of the

proposed error correction output. We have named our proposed multi-contrast segmentation method (including PEC) as HIPS (for HIPpocampus subfield Segmentation).

3 Experiments and results

In this section, the parameters of the proposed method and its results are presented. The method parameters has been adjusted independently to work with the two different segmentation protocols/datasets and it has been compared to other related state-of-the-art methods. To evaluate the segmentation accuracy, we have used the DICE coefficient (Zijdenbos et al., 1994) measured in the linear MNI space. In order to evaluate the significance of the results we applied a Kruskal-Wallis test to find out if any of the configurations present significant differences. Finally a pair-wise Wilcoxon test was applied to find the specific differences.

3.1 MOPAL parameters

In all the experiments, we used patch sizes of 3x3x3 and 7x7x7 voxels, for each scale respectively. The search volume was set to 7x7x7 voxels. We used 64 independent Patch Matches with 4 iterations each. All these parameters were optimized for both datasets in as similar manner as done in (Romero et al., 2017).

Winterburn dataset

In all the experiments using this dataset we used the following multi-scale mixing coefficients (5 structures + background) being α=[0.4711, 0.3443, 0.3826, 0.3900, 0.8439, 0.7715]. These coefficients have been optimized doing a Leave-Two-Out Cross-Validation consisting of 5 rounds of optimization leaving the case under study and its flipped version out (5 rounds of 8 subjects for optimizations and 1 subject for validation). The result were 5 sets of coefficients that we used in the following experiments. For the sake of simplicity, the α values provided correspond to the mean of the 5 optimization rounds as done in (Romero et al., 2017).

In table 1, it can be seen the results (measured through a LOOCV)comparing both versions of the mono-contrast method (T1w and T2w) and the multi-contrast version based on SSD and MSN. We found that for this dataset, T1w MR has a low contribution in the segmentation process as no significant differences were found between T2w mono-contrast and multi-contrast. This makes sense since T2w images from this dataset have better contrast than T1w. In fact, manual delineation was performed over the T2w images only.

*Table 1: Mean DICE in the MNI space and standard deviation for each structure segmentation using high resolution T1w, T2w and Multi-contrast respectively over the Winterburn dataset. Best results in bold. Kurskal-Wallis test revealed that for each row (Average, structures and Hippocampus) there exist differences. Significant differences are marked with * for T1w and T2w, ɬ for T1w and T1w+T2w MSN, ǂ for T2w and T1w+T2w and φ for T1w+T2w SSD and T1w+T2w MSN (p< 0.05).*

Structure	***T1w HR***	***T2w HR***	***T1w+T2w HR SSD***	***T1w+T2w HR MSN***
Average	0.6222 ± 0.0946	0.6830 ± 0.0727*	**0.6803** ± 0.0711	**0.6943 ± 0.0689** ɬ
CA1	0.6633 ± 0.0455	0.7394 ± 0.0287 *	0.7321 ± 0.0270	**0.7468 ± 0.0285** ɬ
CA2\CA3	0.5186 ± 0.0788	0.5916 ± 0.0511 *	0.5893 ± 0.0494	**0.5965 ± 0.0483** ɬ
CA4\DG	0.7242 ± 0.0254	**0.7727 ± 0.0277** *	0.7542 ± 0.0282	0.7686 ± 0.0294 ɬ
SR\SL\SM	0.5245 ± 0.0566	0.6604 ± 0.0389 *	0.6229 ± 0.0378	**0.6604 ± 0.0373** ɬ ϕ
Subiculum	0.6805 ± 0.0439	0.6510 ± 0.0629	0.7032 ± 0.0427	**0.6992 ± 0.0412** ǂ
Hippocampus	0.8717 ± 0.0284	0.8925 ± 0.0105 *	**0.9019** ± 0.0133	**0.9056 ± 0.0114** ɬǂ

Kulaga-Yoskovitz dataset

For this dataset, we estimated again the optimal value for the 4 multi-scale mixing coefficients (3 structures + background) being α=[0.4, 0.5, 0.5, 0.9]. These coefficients have been optimized in the same way that Winterburn ones doing a Leave-Ten-Out Cross-Validation consisting of 5rounds of optimization leaving five cases (and its flipped version) out (5 rounds of 40 subjects for optimization and 10 subjects for test). The result were 5 sets of coefficients that we used in the following experiments. For the sake of simplicity, the α values provided correspond to the mean of the 10 optimization rounds as done in (Romero et al., 2017).

In table 2, it can be seen the results (measures using a using also a LFOCV) comparing both versions of the mono-contrast method (T1w and T2w) and the multi-contrast version based on SSD and MSN. We found that T2w presents little contribution to the segmentation process as no significant differences were found between T1w mono-contrast and multi-contrast. Again, this makes sense since T2w images from this dataset present artifacts from the acquisition process. We assume that this is the reason why manual delineation was performed using T1w images only in this dataset.

*Table 2: Mean DICE in the MNI space and standard deviation for each structure segmentation using high resolution T1, T2 and Multi-contrast respectively over the Kulaga-Yoskovitz dataset. Best results in bold. Kurskal-Wallis test revealed that for each row (Average, structures and Hippocampus) there exist differences. Significant differences are marked with * for T1w and T2w, ɬ for T1w and T1w+T2w MSN, ǂ for T2w and T1w+T2w and φ for T1w+T2w SSD and T1w+T2w MSN (p< 0.05).*

Structure	*T1w HR*	*T2w HR*	*T1w+T2w HR SSD*	*T1w+T2w HR MSN*
Average	0.8797 ± 0.0265	0.8426 ± 0.0304 *	**0.8753 ± 0.0228**	**0.8826 ± 0.0259** ǂ
CA1-3	0.9088 ± 0.0153	0.8727 ± 0.0208 *	0.9015 ± 0.0144	**0.9115 ± 0.0151** ǂ ϕ
CA4\DG	0.8571 ± 0.0321	0.8429 ± 0.0476	0.8600 ± 0.0349	**0.8616 ± 0.0339** ŧǂ
Subiculum	0.8733 ± 0.0209	0.8120 ± 0.0381 *	0.8645 ± 0.0238	**0.8746 ± 0.0236** ǂ ϕ
Hippocampus	**0.9583 ± 0.0073**	0.9202 ± 0.0152 *	0.9507 ± 0.0075	0.9581 ± 0.0067 ǂ ϕ

3.2 Systematic error corrector

Finally, we evaluated our proposed error corrector and compared it with the SegAdapter method (Wang et al., 2011). For this comparison, we used the 1.9 version of the SegAdapter method with optimal parameters empirically obtained. For the Winterburn dataset, we used a dilation radius of 1 to obtain the uncertainty ROI, a sampling rate of 0.15, a patch size of 7x7x7 voxels and T1w and T2w images as features. For Kulaga-Yoskovitz dataset, we used also a dilation radius of 1, a sampling rate of 0.1, a patch size of 7x7x7 voxels and T1w images as features (T2w images did not help the correction in this dataset).

For the Winterburn dataset, we trained both SegAdapter and PEC method 5 times in a leave-two-out cross validation (L2OCV) strategy by removing each pair of hippocampus (left and right) from each case being evaluated. For the Kulaga-Yoskovitz dataset, a L2OCV would result in 5 labels x 25 subjects = 125 training rounds which supposes several weeks of processing so we performed only two training rounds splitting the dataset in two groups with 15 and 10 subjects and cross validated them.

Tables 3 and 4 show the DICE coefficient achieved by the correction methods over both datasets. Note that PEC performed well for both datasets. However, the increment obtained for Winterburn dataset was higher. This makes sense since Winterburn results had more room for improvement than Kulaga-Yoskovitz (0.6943 against 0.8826) and the library size was quite small. Also, Winterburn protocol structures are smaller in its definition than Kulaga-Yoskovitz ones so small changes have a greater impact in DICE coefficient. The improvement provided by segAdapter was quite small for the Kulaga-Yoskovitz dataset while it had a negative impact with the Winterburn dataset. This was counterintuitive so extensive experiments were performed to assess the validity of the conclusions. We can only suppose that maybe SegAdapter method is not well adapted to HR data as it was developed for standard resolution data and also that the small number of training data (Winterburn dataset mainly) may introduce some overfitting problems. We also evaluated the methods excluding a subset of labels from the correction process. We found that PEC performs better over the Kulaga-Yoskovitz dataset by excluding the CA4\DG so the correction is not applied

to this structure. We didn´t apply this selection on SegAdapter method since this method requires all labels to be corrected at the same time (no partial correction is allowed).

*Table 3: Mean DICE in the MNI space for HIPSS segmentation after applying SegAdapter and PEC over the Winterburn. Kurskal-Wallis test revealed that there exist differences for CA1 and SR\SL\SM structures. Significant differences are marked with * for HIPSS and HIPSS + PEC, ł between HIPSS + SegAdapter and HIPSS + PEC. No significant differences were found between HIPSS and HIPSS + SegAdapter.*

Structure	***HIPS(no correction)***	***HIPS (SegAdapter)***	***HIPS (PEC)***
Average	0.6943 ± 0.0689	0.6822 ± 0.0786	**0.7158 ± 0.0652ł**
CA1	0.7468 ± 0.0285	0.7470 ± 0.0226	**0.7762 ± 0.0251*ł**
CA2\CA3	0.5965 ± 0.0483	0.5683 ± 0.0512	**0.6179 ± 0.0630**
CA4\DG	0.7686 ± 0.0294	0.7622 ± 0.0317	**0.7750 ± 0.0307**
SR\SL\SM	0.6604 ± 0.0373	0.6489 ± 0.0274	**0.7018 ± 0.0191*ł**
Subiculum	0.6992 ± 0.0412	0.6844 ± 0.0418	**0.7082 ± 0.0597**
Hippocampus	0.9056 ± 0.0114	0.9003 ± 0.0117	**0.9111 ± 0.0098ł**

*Table 4: Mean DICE in the MNI space for HIPSS segmentation after applying SegAdapter and PEC over the Kulaga-Yoskovitz. Kurskal-Wallis test revealed that there exist differences for Subiculum. Significant differences are marked as * between HIPSS and HIPSS + PEC and ł between HIPSS + SegAdapter and HIPSS + PEC. No significant differences were found between HIPSS and HIPSS + SegAdapter.*

Structure	***HIPS(no correction)***	***HIPS (SegAdapter)***	***HIPS (PEC)***
Average	0.8826 ± 0.0259	0.8833 ± 0.0247	**0.8879 ± 0.0271**
CA1-3	0.9115 ± 0.0151	0.9115 ± 0.0126	**0.9158 ± 0.0145**
CA4\DG	0.8616 ± 0.0339	**0.8656 ± 0.0286**	0.8616 ± 0.0339
Subiculum	0.8746 ± 0.0236	0.8727 ± 0.0226	**0.8863 ± 0.0206*ł**
Hippocampus	0.9581 ± 0.0067	0.9573 ± 0.0061	**0.9595 ± 0.0064**

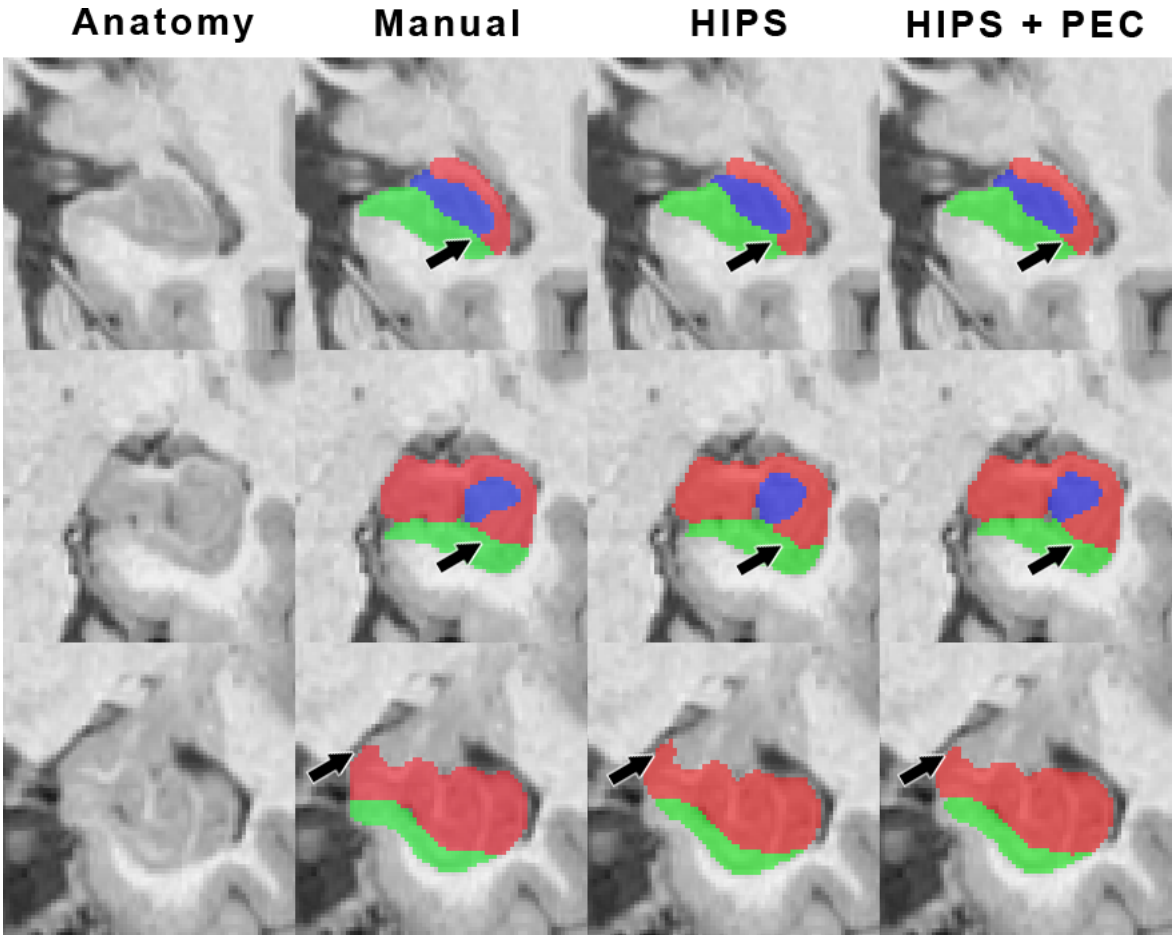


Figure 2: Error correction example from Kulaga-Yoskovitz dataset showing the anatomical T1w, the automatic segmentation and the PEC correction. Arrows pointing to areas where PEC made changes in the segmentation.

3.3 Standard resolution vs. High resolution

The proposed method works with high resolution MR images but these sequences are not always available either in research or in clinical environments. However, it would be desirable to be able to analyze legacy data. For these reasons, we have evaluated the method over standard resolution (1x1x1 mm^3) images upsampled to 0.5x0.5x0.5 mm^3 using B-spline interpolation and a recent super-resolution technique called LASR (Coupé et al., 2013). To do this, we reduced the resolution of the HR images by a factor 2 and later upsampled them using the described methods. We performed this experiment using two configurations: the proposed multi-contrast method (tables 5 and 7) and also a mono-contrast (using T1w) version (tables 6 and 8) as it better represents the application to legacy data which usually consist of T1w sequences.

Tables 5, 6, 7 and 8 show the results for both datasets. The results confirm that HIPS can produce competitive results when using standard resolution images. Note that the results using LASR are better than using B-spline interpolation for the Kulaga-Yoskovitz dataset. However, this improvement is small despite the statistical significance. Moreover, the segmentation accuracy obtained using images up-sampled with LASR is close to the accuracy obtained using HR data as no significant differences were found. This important result shows that the proposed framework can efficiently process usual 1x1x1 mm^3 MR data. On the other hand, one may think that the method is using heavy prior information instead of following the anatomical references contained in the image. Our analysis of the results suggest that this is not the case because just looking at the SR data, image features such as the hippocampus "dark-band" (T1) are visible after the up-sampling process (see figures 3 and 4). Besides, differences between SR and B-spline interpolated data can be only explained due to the effect of the SR process. Figures 3 and 4 show an example of HR and SR based segmentation and how anatomy is partially recovered.

Table 5: Mean DICE in the MNI space and standard deviation for each structure segmentation using the high resolution library applying B-spline interpolation and LASR to the previously downsampled image to be segmented. Segmentation produced using the multi-contrast version of the method over the Winterburn dataset. No significant differences were found between B-spline and LARS, B-spline and HR and LASR and HR. Best results in bold..Results using the HR images are also provided for comparison.

Structure	***B-spline T1w + T2w***	***LASR T1w + T2w***	***HR T1w + T2w***
Average	0.7078 ± 0.0659	**0.7108 ± 0.0647**	0.7158 ± 0.0652
CA1	0.7690 ± 0.0267	**0.7707 ± 0.0267**	0.7762 ± 0.0251
CA2\CA3	0.6108 ± 0.0741	**0.6170 ± 0.0655**	0.6179 ± 0.0630
CA4\DG	0.7690 ± 0.0306	**0.7732 ± 0.0305**	0.7750 ± 0.0307
SR\SL\SM	0.6871 ± 0.0230	**0.6903 ± 0.0216**	0.7018 ± 0.0191
Subiculum	**0.7030 ± 0.0668**	0.7025 ± 0.0614	0.7082 ± 0.0597
Hippocampus	0.9080 ± 0.0089	**0.9119 ± 0.0130**	0.9111 ± 0.0098

Table 6: Mean DICE in the MNI space and standard deviation for each structure segmentation using the high resolution library applying B-spline interpolation and LASR to the previously downsampled image to be segmented. Segmentation produced using the mono-contrast (T1w) version of the method over the Winterburn dataset. No significant differences were found between B-spline and LARS, B-spline and HR and LASR and HR. Best results in bold. Results using the HR images are also provided for comparison.

Structure	***BSpline T1w***	***LASR T1w***	***HR T1w***
Average	0.6082 ± 0.0986	**0.6176 ± 0.0953**	0.6222 ± 0.0946
CA1	0.6590 ± 0.0504	**0.6638 ± 0.0478**	0.6633 ± 0.0455
CA2\CA3	0.5011 ± 0.0823	**0.5154 ± 0.0787**	0.5186 ± 0.0788
CA4\DG	0.7139 ± 0.0278	**0.7166 ± 0.0236**	0.7242 ± 0.0254
SR\SL\SM	0.5046 ± 0.0531	**0.5154 ± 0.0521**	0.5245 ± 0.0566
Subiculum	0.6626 ± 0.0472	**0.6769 ± 0.0437**	0.6805 ± 0.0439
Hippocampus	0.8741 ± 0.0186	**0.8765 ± 0.0205**	0.8717 ± 0.0284

*Table 7: Mean DICE in the MNI space and standard deviation for each structure segmentation using the high resolution library applying B-spline interpolation and LASR to the previously downsampled image to be segmented. Segmentation produced using the multi-contrast version of the method over de Kulaga-Yoskovitz dataset. Kurskal-Wallis test revealed tha there exist differences for the average DICE, the CA1-3 and the whole hippocampus. Significant differences are marked with * for B-spline and LASR, ł for LASR and HR and ǂ for B-spline and HR. Best results in bold. Results using the HR images are also provided for comparison.*

Structure	***BSpline T1w + T2w***	***LASR T1w + T2w***	***HR T1w + T2w***
Average	0.8803 ± 0.0288	**0.8828 ± 0.0280***	0.8879 ± 0.0271łǂ
CA1-3	0.9100 ± 0.0146	**0.9120 ± 0.0137***	0.9158 ± 0.0145łǂ
CA4/DG	0.8525 ± 0.0331	**0.8563 ± 0.0325**	0.8616 ± 0.0339
Subiculum	0.8783 ± 0.0226	**0.8800 ± 0.0220**	0.8863 ± 0.0206
Hippocampus	0.9552 ± 0.0070	**0.9566 ± 0.0065***	0.9595 ± 0.0064łǂ

*Table 8: Mean DICE in the MNI space and standard deviation for each structure segmentation using the high resolution library applying B-spline interpolation and LASR to the previously downsampled image to be segmented. Segmentation produced using the mono-contrast (T1w) version of the method over de Kulaga-Yoskovitz dataset. Kurskal-Wallis test revealed that there exist differences for the average DICE and the CA4/DG. Significant differences are marked with * for B-spline and LASR, ł for LASR and HR and ǂ for B-spline and HR. Best results in bold. Results using the HR images are also provided for comparison*

Structure	***BSpline T1w***	***LASR T1w***	***HR T1w***
Average	0.8709 ± 0.0314	**0.8732 ± 0.0307**	0.8797 ± 0.0265ǂ
CA1-3	0.9030 ± 0.0159	**0.9052 ± 0.0152**	0.9088 ± 0.0153
CA4/DG	0.8403 ± 0.0326	**0.8439 ± 0.0326**	0.8571 ± 0.0321ǂ
Subiculum	0.8693 ± 0.0218	**0.8704 ± 0.0214**	0.8733 ± 0.0209
Hippocampus	0.9546 ± 0.0080	**0.9566 ± 0.0077**	0.9583 ± 0.0073

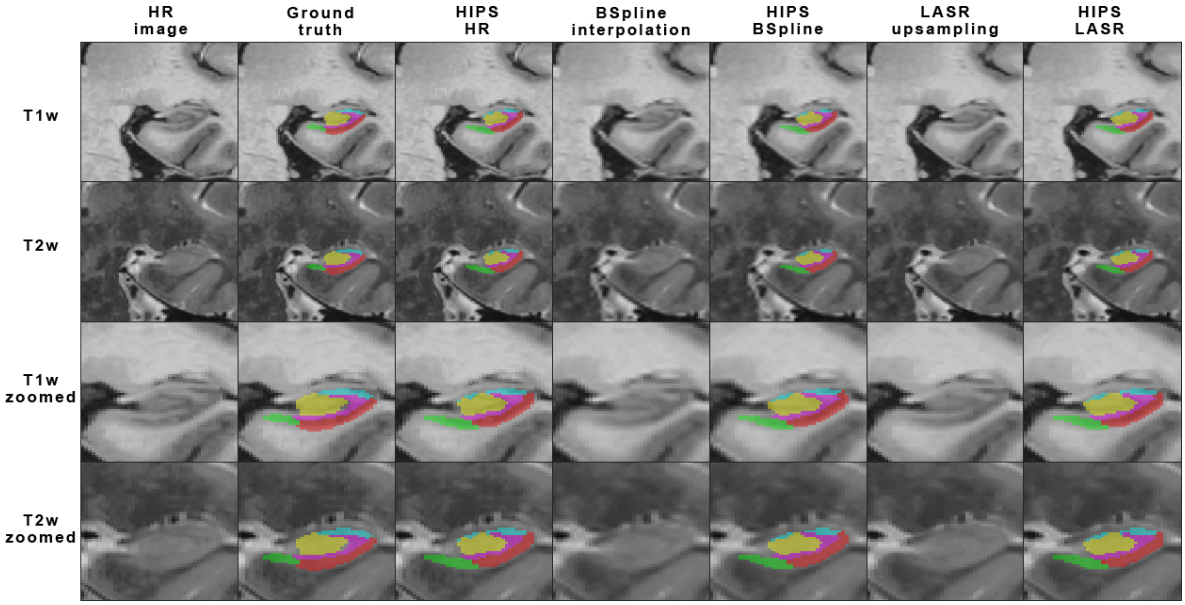


Figure 3: Example result of multi-contrast HIPS segmentation on HR, B-spline interpolations and LASR upsampling (Winterburn dataset).

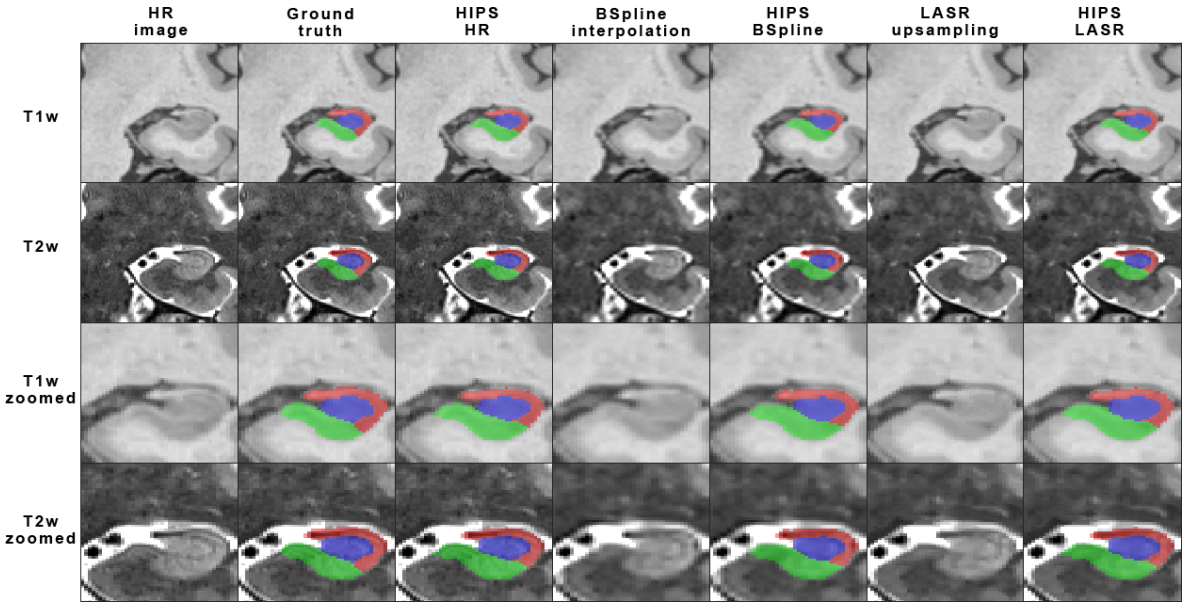


Figure 4: Example result of multi-contrast HIPS segmentation on HR, B-spline interpolation and LASR upsampling (Kulaga-Yoshcovitz dataset).

3.4 Methods comparison

We compared the proposed method HIPS with other recent methods applied to hippocampus subfield segmentation over both datasets. We compared HIPS results on the Winterburn dataset with MAGeT results (Pipitone et al., in 2014) as reported by the authors. We also compared HIPS results on Kulaga-Yoskovitz dataset with ASHS (Yushkevich et al., 2015b) and SurfPatch (Caldairou et al., 2016). We used the published results of ASHS and SurfPatch as provided in (Caldairou et al., 2016) in the comparison.

We also included human rater information from Winterburn dataset original paper (Winterburn et al., 2013) and Kulaga-Yoskovitz dataset original paper (Kulaga-Yoskovtiz et al., 2015) as a reference. Table 7 shows results for MAGeT and HIPS on the Winterburn dataset while table 8 shows results for ASHS, SurfPatch and HIPS on the Kulaga-Yoskovitz dataset. For a fair comparison between considered methods, all the DICE coefficients for HIPS have been calculated using the segmentations in native space (using the corresponding inverse affine registration).

In case of comparison with MAGeT, Winterburn images were at a 0.3x0.3x0.3 mm^3 voxel resolution but MAGeT provided segmentations at 0.9x0.9x0.9 mm^3 resolution for efficiency. For this reason, we downsampled the Winterburn images to 0.9x0.9x0.9 mm^3 in the native space to be able to make a fair comparison. HIPS showed an overall improvement of 2.6 % in comparison reaching an overall DICE of 0.661.

In the case of Kulaga-Yoskovtiz our proposed method improved clearly all the structures even surpassing inter-rater agreement by a 3 % for the CA1-3 and staying only a 1 % below the overall with an average DICE of 0.8744. Regarding to the execution time, the whole HIPS pipeline takes less than 20 minutes while the other compared methods have a computational burden of several hours per case.

Table 7: Mean DICE in the native space for each structure. Segmentation performed by MAGeT and HIPS at 0.9x0.9x0.9 mm^3 over the Winterburn dataset. Best results (for automatic segmentation) in bold. Human rater mean DICE at 0.3x0.3x0.3 mm^3 is also provided as reference.

Structure	***MAGeT (T1 0.9 mm)***	***HIPS (T1+T2 0.9 mm)***	***Intra-rater (T1 0.3 mm)***
Average	0.526	**0.661**	0.742
CA1	0.563	**0.670**	0.780
CA2\CA3	0.412	**0.522**	0.640
CA4\DG	0.647	**0.763**	0.830
SR\SL\SM	0.428	**0.599**	0.710
Subiculum	0.580	**0.722**	0.750
Hippocampus	0.816	**0.876**	0.910

Table 8: Mean DICE in the native space for each structure. Segmentation performed by ASHS, SurfPatch, HIPS and human rater (intra-rater and inter-rater) over de Kulaga-Yoskovitz dataset. Best results (for automatic segmentation) in bold.

Structure	***ASHS***	***SurfPatch***	***HIPS***	***Inter-rater***	***Intra-rater***
Average	0.8513	0.8503	**0.8744**	0.8833	0.9113
CA1-3	0.8736 ± 0.0197	0.8743 ± 0.0247	**0.9030 ± 0.0138**	0.8760 ± 0.048	0.9290 ± 0.010
CA4\DG	0.8254 ± 0.0345	0.8271 ± 0.0285	**0.8497 ± 0.0332**	0.9030 ± 0.036	0.9000 ± 0.019
Subiculum	0.8548 ± 0.0243	0.8495 ± 0.0245	**0.8705 ± 0.0212**	0.8710 ± 0.053	0.9050 ± 0.016

4 Discussion

One of the contributions of this work is a new multi-contrast patch similarity consisting on a multi-contrast SSD-based semi-norm (MSN). Introducing this similarity measure in OPAL (now MOPAL), we achieved good segmentation results using T1w+T2w images. By using the semi-norm to combine distances we obtain a robust and self-balanced similarity that takes benefit from information coming from both channels. This means image corruption or low image quality in one of the channels can be overcome using the proposed similarity properties. This contribution makes the method more robust especially when applied to different datasets/conditions.

In addition, we proposed a new systematic error correction method using an ensemble of patch-based neural networks (PEC). The use of this error corrector significantly improves the results over both datasets having an execution time overload of just a few seconds. Even though both are ensemble methods, PEC has shown to perform better that SegAdapter when using a significant lower number of base classifiers. Both methods use a boosting technique to learn the misclassified patterns. The main difference of PEC is the use of patch-based strong classifiers instead of weak classifiers as done by SegAdapter jointly with a richer feature descriptor. We chose a neural network base classifier because its versatility and availability to perform structured prediction (patch correction vs voxel correction) enhancing label regularity. Furthermore, the chosen classifier strength allowed to converge rapidly needing only 10 networks to reach the maximum accuracy. It is worth to note that using this correction, the proposed method almost reaches human rater accuracy for the Kulaga-Yoskovitz dataset where obtained a higher DICE than the inter-rater for the CA1-3 (0.9030 obtained by HIPS versus 0.8760 obtained by the inter-rater), almost the same accuracy for the Subiculum (0.8705 obtained by HIPS versus 0.8710 obtained by the inter-rater) and presents an overall dice of 0.8744 which is considerably close to the 0.8833 obtained by the inter-rater.

Comparing the results obtained in both datasets, it is expectable to see an improvement over the Winterburn dataset if more manually delineated cases were released. Even although this is heavily

dependent on the application, we think (based in our previous works using multi-atlas segmentation) that optimal results can be obtained using at least 20 reference atlases. We plan to add new manually labeled cases to the library to further leverage the method results. HIPS pipeline will be made publically available to the scientific community as part of our online volBrain platform (http://volbrain.upv.es).

5 Conclusion

In this work, we have presented a new method for HR hippocampus subfield segmentation called HIPS. It uses two publically available segmentation protocols and datasets (Winterburn and Kulaga-Yoskovitz). Our method is based on MOPAL, a multi-contrast extension of the OPAL patch-based label fusion segmentation method and a novel neural network based error corrector. HIPS works in a fully automated manner providing accurate results in less than 20 minutes thanks to MOPAL that performs fast segmentation as well as to the subject specific library registration that only requires estimating one non-linear registration over small-region to translate the whole library to the case to be segmented.

Furthermore, as it has been shown, our proposed method is able to produce competitive results on standard resolution images. This is an important feature as it makes the method a suitable tool for standard resolution data analysis and opens the door to analyze large legacy databases. This contributes to the method scalability as well as the use of a library of manually labeled images as knowledge base. The system can learn new anatomy patterns just by extending this library with new segmented cases. This way the method can be either extended or adapted to a specific dataset or pathology.

From the robustness point of view, the registration is a key step. For this reason we used a multi-contrast registration as well as multi-contrast segmentation. This way we covered more variability and reduced the results dispersion in terms of accuracy especially over the Kulaga-Yoskovitz dataset.

We showed that HIPS outperforms other state-of-the-art methods in term of segmentation accuracy achieving an overall DICE of 0.661 for the Winterburn dataset while MAGeT (Pipitone et al., 2014) obtains a DICE of 0.5260 and an overall DICE of 0.8744 for Kulaga-Yoskovitz while ASHS (Yushkevich et al., 2015b) obtains 0.8513 and SurfPatch (Caldairou et al., 2016) obtains 0.8503. HIPS is also faster than the other methods taking an average execution time under 20 minutes compared to several hours required by the other methods. It is also important to note that HIPS performance, for the Kulaga-Yoskovitz dataset, is close to the human rater reaching better results than the inter-rater segmentation for the CA1-3 sctructure. This does not happen for the Winterburn dataset which can be explained by the low number of manually labeled cases (only 5 subjects) as well as the higher structural complexity of the segmentations.

Acknowledgements

This research was supported by Spanish UPV2016-0099 and TIN2013-43457-R grants from UPV and the Ministerio de Economia y competitividad. This study has been carried out with financial support from the French State, managed by the French National Research Agency (ANR) in the frame of the Investments for the future Program IdEx Bordeaux (ANR-10-IDEX-03-02, HL-MRI Project), Cluster of excellence CPU and TRAIL (HR-DTI ANR-10-LABX-57) and the CNRS multidisciplinary project "Défi imag'In".

We also want to thank Javier Juan Albarracín for his valuable contribution to the development of this method.

References

Altshuler, L.L., Bartzokis, G., Grieder, T., Curran, J., Mintz, J., 1998. Amygdala enlargement in bipolar disorder and hippocampal reduction in schizophrenia: an MRI study demonstrating neuroanatomic specificity. Arch. Gen. Psychiatry 55 (7), 663 - 664.

Avants, B.B., Tustison, N., Song, G., 2009. Advanced normalization tools (ANTS). Insight Journal.

Barnes, C., Shechtman, E., Finkelstein, A., Goldman, D.B., 2009. PatchMatch: a randomized correspondence algorithm for structural image editing. ACM Trans Graph 28(3).

Barnes, J., Foster, J., Boyes, R.G., Pepple, T., Moore, E.K., Schott, J.M., Frost, C., Scahill, R.I., Fox, N.C. 2008. A comparison of methods for the automated calculation of volumes and atrophy rates in the hippocampus. Neuroimage 40, 1655 - 1671.

Boccardi, M., Bocchetta, M., Apostolova, L.G., Barnes, J., Bartzokis, G., Corbetta, G., DeCarli, C., deToledo-Morrell, L., Firbank, M., Ganzola, R., Gerritsen, L., Henneman, W., Killiany, R.J., Malykhin, N., Pasqualetti, P., Pruessner, J.C., Redolfi, A., Robitaille, N., Soininen, H., Tolomeo, D., Wang, L., Watson, C., Wolf, H., Duvernoy, H., Duchesne, S., Jack Jr, C.R., Frisoni, G.B., for the EADC-ADNI Working Group on the Harmonized Protocol for Manual Hippocampal Segmentation., 2015. Delphi Definition of the EADC-ADNI Harmonized Protocol for Hippocampal Segmentation on Magnetic Resonance. Alzheimer's & Dementia 11(2), 126 - 138.

Braak, H., Braak, E., 1991. Neuropathological stageing of Alzheimer-related changes. Acta Neuropathol. 82 (4), 239 - 259.

Caldairou, B., Bernhardt, B.C., Kulaga-Yoskovitz, J., Kim, H., Bernasconi, N., Bernasconi, A., 2016. A Surface Patch-Based Segmentation Method for Hippocampal Subfields. MICCAI, Part II, LNCS 9901, 379 - 387.

Chakravarty M, Steadman P, Eede M, Calcott R, Gu V, Shaw P, Raznahan A, Collins D.L., and Lerch J.P. 2013. Performing label-fusion-based segmentation using multiple automatically generated templates. *Hum Brain Mapp*, 34(10):2635–54.

Chupin, M., Gérardin, E., Cuingnet, R., Boutet, C., Lemieux, L., Lehéricy, S., Benali, H., Garnero, L., Colliot, O., Alzheimer's Disease Neuroimaging Initiative., 2009. Fully automatic hippocampus segmentation and classification in Alzheimer's disease and mild cognitive impairment applied on data from ADNI. Hippocampus 19(6), 579 - 587.

Collins, D.L., Pruessner, J.C., 2010. Towards accurate, automatic segmentation of the hippocampus and amygdala from MRI by augmenting ANIMAL with a template library and label fusion. Neuroimage 52(4), 1355 - 66.

Coupé, P., Manjón, J.V., Chamberland, M., Descoteaux, M., Hiba, B., 2013. Collaborative patch-based super-resolution for diffusion-weighted images. NeuroImage 83, 245 - 261.

Coupé, P., Manjón, J.V., Chamberland, M., Descoteaux, M., Hiba, B., 2013. Collaborative patch-based super-resolution for diffusion-weighted images. Neuroimage. 83, 245 - 261.

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 - 54.

Dickerson, B.C., Sperling, R.A., 2005. Neuroimaging biomarkers for clinical trials of disease-modifying therapies in Alzheimer's disease. NeuroRx 2, 348 - 360.

Fisher, R.B., Oliver, P., 1995. Multi-variate cross-correlation and image matching. Proceedings of the 6th British machine vision conference 1, 623 - 632.

Giraud, R., Ta, V.T., Papadakis, N., Manjón, J.V., Collins, D.L., Coupé, P., Alzheimer's Disease Neuroimaging Initiative, 2016. An Optimized PatchMatch for Multi-scale and Multi-feature Label Fusion. NeuroImage 124, 770 - 782.

Hafting, T., Fyhn, M., Molden, S., Moser, M., Moser, E.I., 2005. Microstructure of a spatial map in the entorhinal cortex. Nature 436, 801-806.

Iglesias, J.E., Augustinack, J.C., Nguyen, K., Player, C.M., Player, A., Wright, M., Roy, N., Frosch, M.P., McKee, A.C., Wald, L.L., Fischl, B., Van Leemput, K., Alzheimer's Disease Neuroimaging Initiative, 2015. A computational atlas of the hippocampal formation using ex vivo, ultra-high resolution MRI: Application to adaptive segmentation of in vivo MRI. NeuroImage 115(15), 117 - 137.

Insausti, R., Amaral, D.G. 2004. Hippocampal formation. The Human Nervous System Pages, 871-914

Jack, C.R., Petersen, R.C., Xu, Y., O'Brien, P.C., Smith, G.E., Ivnik, R.J., Boeve, B.F., Tangalos, E.G., Kokmen, E., 2000. Rates of hippocampal atrophy correlate with change in clinical status in aging and AD. Neurology 55, 484 - 489.

Jack, C.R., Shiung, M.M., Weigand, S.D., O'Brien, P.C., Gunter, J.L., Boeve, B.F., Knopman, D.S., Smith, G.E., Ivnik, R.J., Tangalos, E.G., Petersen, R.C., 2005. Brain atrophy rates predict subsequent clinical conversion in normal elderly and amnestic MCI. Neurology 65, 1227 - 1231.

Kulaga-Yoskovitz, J., Bernhardt, B.C., Hong, S., Mansi, T., Liang, K.E., van der Kouwe, A.J.W., Smallwood, J., Bernasconi, A., Bernasconi, N., 2015 Multi-contrast submillimetric 3Tesla hippocampal subfield segmentation protocol and dataset. Sci Data. 2, 150059.

Manjón, J.V., Coupé, P., Buades, A., Fonov, V., Collins, D.L., Robles, M., 2010a. Non-Local MRI Upsampling. Medical Image Analysis 14(6), 784 - 792.

Manjón, J.V., Coupé, P., Martí-Bonmatí, L., Collins, D.L., Robles, M., 2010b. Adaptive non-local means denoising of MR images with spatially varying noise levels. J Magn Reson Imaging 31, 192 - 203.

Manjón J.V., Coupe P, Raniga P, Xia Y, Fripp J, and Salvado O. 2016. HIST: HyperIntensity Segmentation Tool. International Workshop on Patch-based Techniques in Medical Imaging, 92-99, MICCAI2016.

Milner, B., 1958. Psychological defects produced by temporal lobe excision. Res. Publ. Assoc. Res. Nerv. Ment. Dis. 36, 244 - 257.

Nyúl, L.G., Udupa, J.K., 1999. On standardizing the MR image intensity scale. Magn Reson Med. 42(6), 1072 - 81.

Pipitone, J., Park, M.T.M., Winterburn, J., Lett, T.A., Lerch, J.P., Pruessner, J.C., Lepage, M., Voineskos A.N., Chakravarty, M.M., the Alzheimer's Disease Neuroimaging Initiative, 2014. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. NeuroImage 101, 494 - 512.

Romero, J.E., Coupé, P., Giraud, R., Ta, V.T., Fonov, V., Park, M.T.M., Chakravartyg, M.M., Voineskos, A.N., Manjón, J.V., 2017. CERES: A new cerebellum lobule segmentation method. NeuroImage 147, 916 - 924.

Romero, J.E., Coupe, P., Manjón, J.V., 2016. High Resolution Hippocampus Subfield Segmentation Using Multispectral Multi-atlas Patch-Based Label Fusion. Patch-MI, LNCS 9993, 117 - 124.

Schmajuk N. A., 1990. Role of the hippocampus in temporal and spatial navigation. Behavioural Brain Research, 39(3), 205-229.

Tustison, N.J., Avants, B.B., Cook, P.A., Zheng, Y., Egan, A., Yushkevich, P.A., Gee, J.C., 2010. N4ITK: improved N3 bias correction. IEEE Trans Med Imaging 29(6), 1310 - 1320.

Van Leemput, K., Bakkour, A., Benner, T., Wiggins, G., Wald, L.L., Augustinack, J., Dickerson, B.C., Golland, P., Fischl, B., 2009. Automated segmentation of hippocampal subfields from ultra-high resolution in vivo MRI. Hippocampus 19(6), 549 - 557.

Wang, H., Das, S.R., Suh, J.W., Altinay, M., Pluta, J., Craige, C., Avants, .B.B., Yushkevich, P.A., Alzheimer's Disease Neuroimaging Initiative, 2011. A Learning-Based Wrapper Method to Correct Systematic Errors in Automatic Image Segmentation: Consistently Improved Performance in Hippocampus, Cortex and Brain Segmentation. Neuroimage 55(3), 968 - 985.

Winterburn, J.L., Pruessner, J.C., Chavez, S., Schira, M.M., Lobaugh, N.J., Voineskos, A.N., Chakravarty, M.M., 2013. A novel in vivo atlas of human hippocampal subfields using high-resolution 3 T magnetic resonance imaging. NeuroImage 74, 254 - 265.

Xiao, Y., Fonov, V.S., Beriault, S., Gerard, I., Sadikot, A.F., Pike, G.B., Collins, D.L., 2015., Patch-based label fusion segmentation of brainstem structures with dual-contrast MRI for Parkinson's disease. Int J Comput Ass Rad 10(7), 1029 - 1041.

Yushkevich, P.A., Avants, B.B., Pluta, J., Das, S., Minkoff, D., Mechanic-Hamilton, D., Glynn, S., Pickup, S., Liu, W., Gee, J.C., Grossman, M., Detre, J.A., 2009. A high-resolution computational atlas of the human hippocampus from postmortem magnetic resonance imaging at 9.4 T. Neuroimage 44(2), 385 - 98.

Yushkevich, P.A., Amaral, R.S., Augustinack, J.C., Bender, et al. 2015a. Quantitative comparison of 21 protocols for labeling hippocampal subfields and parahippocampal subregions in in vivo MRI: Towards a harmonized segmentation protocol. NeuroImage 111, 526 - 541.

Yushkevich, P.A., Pluta, J.B., Wang, H., Xie, L., Ding, S.L., Gertje, E,C., Mancuso, L., Kliot, D., Das, S.R., Wolk, D.A., 2015b. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Hum. Brain Mapp. 36(1), 258 - 287.

Zijdenbos, A.P., Dawant, B.M., Margolin, R.A., Palme, A.C., 1994. Morphometric analysis of white matter lesions in MR images: method and validation. IEE Trans Med Imaging 13, 716 - 724.