Surfing the Hippocampus Wave

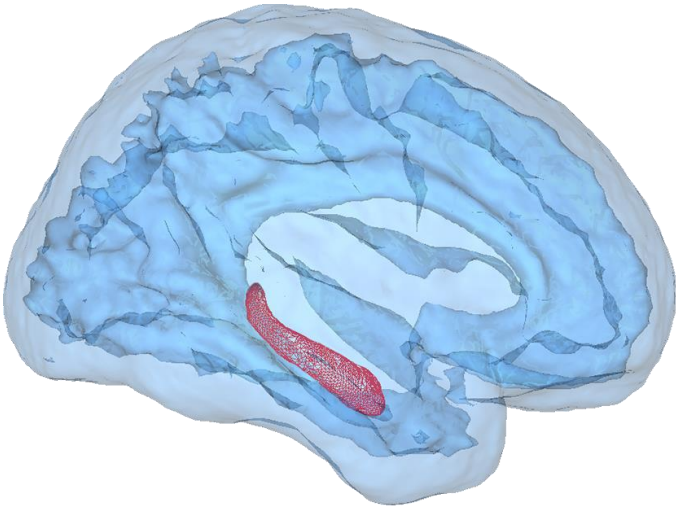
Bartel, F.

2019

document version
Publisher's PDF, also known as Version of record

Chapter 7

Discussion, future perspectives and conclusions



7.1. General objective

The general objective of this thesis was to develop, test and compare methods for hippocampus segmentation and hippocampal atrophy quantification. For that purpose, commonly used available automatic segmentation methods were compared and regionally analysed. A semi-automatic hippocampus segmentation method was developed and tested to outline subcortical brain structures. Finally, a hippocampal atrophy quantification analysis was performed and methods with high statistical power to measure hippocampal atrophy rates were identified.

In the present chapter, the main findings of these studies are summarized and discussed in the context of previous research studies. With the perspective on clinical applications, developments and findings of this work are combined and possible clinical work flows in neuroradiology and radiotherapy are introduced. Finally, future perspectives are presented, and recommendations are given.

7.2. Summary of findings

The aim of this thesis, to optimize methods for hippocampus segmentation and atrophy measurements in the field of neuroradiology and radiotherapy, was addressed as follows:

The first topic presented in this thesis was the quantification of reproducibility and differences between manual and automatic segmentations (chapter 2). The most important finding of this study was that manual segmentations had poorer reproducibility for within-session rescans than one of the automatic segmentation methods (FSL-FIRST). Also, systematic regional differences between methods were detected. These findings reflected the different underlying definitions of the hippocampal structure used by the different methods, thus emphasizing the need for a harmonized hippocampal outlining protocol [1–3]. Furthermore, FSL-FIRST's higher outline reproducibility indicated that automatic methods can have reduced segmentation variabilities. However, the benefits of outline reproducibility should be treated with care, because good reproducibility in itself does not necessarily imply that the hippocampus is outlined accurately or that the method is sensitive to disease-related change.

In chapter 3 hippocampus segmentation agreement was determined between multiple observers using a radiotherapy delineation protocol (RTOG), for the purpose of planning hippocampus sparing radiotherapy. The ICCs of 0.56 and 0.69 for the left and right hippocampus respectively, were relatively low compared to segmentations performed in neuroradiology, which usually exhibit inter-rater ICCs higher than 0.85 [4]. Also, the average overlap index of all observers (0.62) was

much lower than the average overlap index of the study from chapter 2 (0.79). Segmentations from the study of chapter 2 were performed by a single radiology technician with multiple years of hippocampal delineation experience. In chapter 3, the participating observers had different specialisms (radiotherapy technician, oncologist, neuroradiologist), and used a hippocampus delineation protocol which was relatively new to them. This might explain the difference between the overlap indices of these studies and it also leads to the conclusion of chapter 3 that conformity of delineations could for instance be improved by using a better delineation protocol and providing more training. Nevertheless, this variability had limited effect on the ability to reach the clinical goal of reducing the average radiation dose to the hippocampus because in almost all cases, all dose constraints were met.

In chapter 4, a novel semi-automatic delineation tool, FASTSURF, was presented. In this proof of concept study, sparse delineations were simulated and FASTSURF's performance was evaluated in comparison with manual and automatic segmentations. It was shown that FASTSURF segmentations have higher agreement with manual hippocampus segmentation than both tested automatic segmentation methods, even when only six to seven contours were used as input. Possible bias in favour of FASTSURF was limited by comparing the semi-automatic reconstruction to the manual delineation of a different back-to-back scan. The same conclusion, that five to eight input contours suffice, was reached for different outlining protocols and datasets where no back-to-back scans were available. It was also found that with only five input contours, FASTSURF was already better reproducible than the automatic methods FreeSurfer and FSL-FIRST.

The work described in chapter 5 evaluated whether FASTSURF can be used to segment subcortical structures such as the putamen, caudate nucleus and thalamus as a key ingredient of a novel protocol to make a standard segmentation reference set in an optimized labour-saving manner. Similar to chapter 4, FASTSURF segmentations had excellent overlap with complete manual segmentations, concluding that FASTSURF is indeed a time-saving alternative to fully manual segmentations.

Hippocampal atrophy rates as measured on structural MRI are an important biomarker in clinical drug trials. Chapter 6 compared eight different methods to determine one-year hippocampal atrophy rates on structural MRI in presence of AD and MCI and evaluated plausibility, reproducibility and sensitivity of those methods. Results showed that registration-based methods outperformed manual segmentations regarding all three aspects. The registration methods with best performance, in descending order, were ANTs, Elastix and NiftyReg. This study

specifically addressed the feasibility of using these methods to detect radiation-induced hippocampal volume loss. By consulting recent radiotherapy literature to obtain a rough estimate of the expected effect size, this chapter predicts that radiation-induced hippocampal volume loss should be detectable even with a fairly small sample size (N≈20 per group) when one of these automated methods is used. Remarkably, with fully manual delineation with the outlining protocol used in [5] such detection would require much larger sample sizes (N≈90).

7.3. General discussion

Hippocampus segmentation

Manual segmentations are usually considered as the “gold standard” method and the base for all semi-automatic or automatic segmentation methods. Every method is validated with manual segmentations as gold standard by using overlap indices, volume correlations or other similar measures [6]. Generally, a method with high agreement to manual segmentations is considered a good segmentation method. Manual segmentation accuracy determined by the ICC (volume comparison) or overlap indices is relatively high as shown in chapter 2 or in [4] in which multiple studies are listed showing ICCs higher than 0.85 for manual hippocampus segmentation. Higher manual delineation disagreement was presented in chapter 3, which was most likely due to a delineation protocol which was new to the observers and which had little hippocampus segmentation training.

Similar reliability scores were obtained by FSL-FIRST and FreeSurfer in chapter 2 and for instance in [7]. Furthermore, good agreement with manual segmentations were observed in multiple studies for FSL-FIRST and FreeSurfer in [8–13] and for different multi-atlas segmentation methods in [14–16]. These findings suggest that automatic segmentation methods can replace trained observers. However, chapters 2 and 5 also showed that automatic segmentation methods can fail or produce erroneous segmentations. Imaging limitations such as reduced signal-to-noise ratio (SNR), contrast-to-noise ratio (CNR), imaging artefacts and intensity non-uniformity highly affect the segmentation task and especially intensity-based automatic MRI analysis decreases in performance [17]. Moreover, unusual pathology and anatomy are difficult to account for and can lead into large errors for automatic segmentation methods.

This thesis made use of the fact that outlining the hippocampus on every slice of an MRI is inefficient and a lot of redundant information is provided, which makes it possible to reduce the amount of observer input. The same is true for the segmentation of deep grey nuclei. These considerations show the need for the development of semi-automatic methods and render the proposed FASTSURF

method highly relevant. FASTSURF is novel because it is completely based on mesh processing procedures without taking into account any (intensity) information from the image. It should be noted that before the development of FASTSURF, another method was developed and investigated using a priori shape information and sparse contours, but the results of that method were limited.

Usually, semi-automatic methods are based on deformable models such as active contour or shape models (ACM and ASM), or atlas-based registration, or classification methods. Semi-automatic hippocampus segmentation using ACM or ASM have been proposed almost three decades ago, in which either initial lines [18], initial contours [19] or initial points [20,21] need to be drawn to create a model for deformation. Chupin et al. developed a region growing based method in which initial points are placed in the amygdala and hippocampus to minimize the error in detecting the border between those structures [22]. In atlas-based semi-automatic segmentation methods either landmarks need to be selected or key regions need to be segmented to improve atlas registration [23–25]. Alejo et al. proposed a combined supervised and unsupervised classification method in which a contour has to be outlined for the supervised classification [26].

Compared to those methods, FASTSURF has several advantages:

1. FASTSURF does not require intensity information to complete the hippocampus from partial contours, therefore it is less prone to errors arising from noise, artefacts, anatomical abnormalities or gross pathological changes in the image.
2. A (trained) observer can outline the initial sparse contours for FASTSURF using conventional segmentation protocols. For methods in which landmarks need to be selected or specific regions need to be outlined extra training for the observer is required.
3. The computation to complete the hippocampus is extremely fast compared to for instance atlas-based semi-automatic segmentation methods.
4. With simulated sparse contours, FASTSURF has been validated with large datasets. Dill et al. lists semi-automatic segmentation methods with corresponding dataset sizes on which these methods were validated. For all of these methods, the dataset sizes were much smaller than the number of scans on which FASTSURF was tested [16].

In comparison to the deformable model-based method FSL-FIRST and atlas-based method FreeSurfer, FASTSURF with only five input contours was in better agreement with manual segmentations (chapter 4). These promising results suggest that it may be advantageous to use FASTSURF instead of FSL-FIRST or FreeSurfer for cross-sectional segmentation. Furthermore, chapter 5 showed that for deep grey nuclei structures FASTSURF segmentations were almost identical to

manual segmentations, emphasizing that FASTSURF has the potential to be used to create standard reference sets and completely replace fully manual segmentations. Such reference sets are very important to create atlases and train and validate automatic segmentation methods, especially in diseases with evident concomitant pathological changes such as MS.

Atrophy measurement

The most straightforward way to quantify atrophy is by segmenting brain structures on both baseline and follow up images. Alternatively, one can use non-linear registration and measure atrophy rates from the deformation fields or by using a method called boundary shift integral (BSI, [27]). The BSI aligns baseline and repeat images with rigid registration and determines the shift in boundary location using differences in voxel intensities. BSI was not investigated in this dissertation, because Boyes et al. showed that determining hippocampal atrophy using deformation fields resulting from non-linear registration (Jacobian integration) was more accurate than BSI [28]. However, hippocampal atrophy determined by BSI was similar to manually measured atrophy rates [29] and since Boyes's article from 2006 the BSI has been further developed and improved [30,31] and is therefore still a method of choice in ongoing clinical drug trials (NCT03446001). Furthermore, in [32] a multi-atlas segmentation method (MAPS) has been combined with BSI and showed better atrophy reproducibility in a longitudinal single-session scan-rescan setup than FreeSurfer, FSL-FIRST, AdaBoost (machine learning based) and MAPS independently.

In concordance with literature [33–35], it has been shown in chapter 6 that registration-based methods outperformed manual segmentations for determining longitudinal hippocampal atrophy. Also, Yushkevich et al. [36] compared their non-linear registration-based hippocampal atrophy estimations with a study in which a semi-automatic method was used to determine hippocampal atrophy [37], and showed that non-linear registration is more sensitive in measuring atrophy rates. FASTSURF is based on manually segmented sparse segmentation. This implies that FASTSURF might be equally poor as fully manual segmentation in measuring atrophy rates when using independent FASTSURF segmentations on all time points. On the other hand, FASTSURF results in intrinsically smooth surfaces, implying that, compared to manual segmentations, the subtraction of uncorrelated errors is avoided. In manual segmentations, that is one of the causes of its poor performance in atrophy determination. Alternatively, single-timepoint FASTSURF segmentations could be combined with a Jacobian integration-based quantification of volume change between scans.

The results described in this thesis highlight the fact that hippocampus segmentation and hippocampal atrophy determination are complementary image processing tasks, with very different benefits from automatic data analysis methods. Whereas atrophy measurement seems hardly possible with manual expert segmentation, probably due to the extremely high accuracy requirements (sub-millimetre, see Chapter 1 *Segmentation sensitivity*) and uncorrelated noise in baseline and follow-up delineation, highly reproducible atrophy determinations are possible with adequately chosen automated methods. Accurate hippocampus segmentation still requires expert input, using either fully manual or semi-automatic protocols.

Advantage in using mesh-based methods

Throughout this thesis it has been shown that mesh-based methods allow an accurate analysis of reproducibility in a scan-rescan setting, which is an important validation of different segmentation methods. In this context, by using registration parameters obtained from linear registration, meshes are very useful to bring shapes to the same position in space without information loss and shape deformation. With binary image volumes, due to interpolation and thresholding, computing such overlap indices leads to substantial errors as shown in chapter 1. Furthermore, with meshes it is possible to perform other precise measures such as surface distance measures to determine regional atrophy patterns.

In combination with deformation fields obtained from a non-linear registration process, hippocampal meshes can be deformed and atrophy rates can be measured as accurately as with Jacobian integration (chapter 6). The advantage of meshes compared to Jacobian integration is that shape deformation can be immediately visually inspected in 3D.

Finally, this thesis shows that meshes provide a completely novel way to implement semi-automatic delineation methods, such as the FASTSURF technique described in chapter 2.

Limitations of the work in this thesis

In three chapters of this dissertation manual hippocampus segmentation were compared to automatic segmentation methods or FASTSURF (chapter 2, 4 and 6). The longitudinal scan-rescan MRI dataset on which 360 hippocampi were manually segmented is an excellent dataset to analyze segmentation and atrophy performance. The minor limitation for this segmentation dataset is that a single observer segmented all hippocampi, potentially biasing the segmentation results, while improving consistency.

Chapter 4 and 5 showed that FASTSURF is a promising time-saving method for the segmentation of the hippocampus and deep grey nuclei. However, FASTSURF is limited in a few aspects which need further attention:

- In chapter 4 and 5 FASTSURF was only validated with simulated sparse contours and not with independently created manual sparse contours.
- FASTSURF can only be used for relatively smooth structures, not for delineation of irregular structures such as tumours.
- So far, for FASTSURF only one contour on each slice can be outlined, which might not always be sufficient. However, this is not a fundamental limitation because it is conceptually straightforward to connect a single contour on one slice to two or more contours on the next, by defining an appropriate connecting mesh.
- If a structure contains cavities that should be excluded from the volume, special precautions in the outlining software need to be implemented to account for this.

The harmonized hippocampus outlining protocol (HarP) is the most modern and most widely accepted outlining protocol. Therefore, in the ideal case longitudinal hippocampus segmentation in a scan-rescan MRI setting outlined using HarP would complete a validation for FASTSURF or automatic segmentation methods.

In chapter 6, non-linear registration was performed on 3D T1 weighted MPRAGE MRI scans acquired at various locations from different vendors (GE, Philips and Siemens). Image post-processing procedures such as bias field correction or intensity normalization might improve registration performance and make the registration more robust across images from different vendors. Furthermore, for images acquired with different sequences parameters of non-linear registration methods most probably need to be tuned to those images. However, even without such procedures, the results in chapter 6 are very promising.

As mentioned in the introduction, imaging limitations such as resolution, noise and artefacts affect the investigation of segmentation and hippocampal atrophy. In this thesis, only segmentations of T1-weighted images were investigated. In the future perspective and development section of this chapter it is discussed why high field MRI T2-weighted images might be advantageous.

Clinical workflow

Considering the need for fast and accurate hippocampus segmentation and reliable automatic atrophy measurement methods in clinical neuroradiology and radiotherapy, two possible clinical workflows are presented resulting from the work of this dissertation.

In subjects with neurodegenerative diseases in which hippocampal atrophy is a secondary outcome measure in clinical trials the subject must undergo longitudinal MRI scan sessions with the same MRI protocol. To obtain the hippocampus segmentation of the baseline MRI, a trained operator can delineate approximately five contours of the hippocampus and use FASTSURF to complete the segmentation. Then, with a symmetric non-linear registration method (for instance ANTs, Elastix or NiftyReg) the baseline MRI can be mapped to the follow up MRI and atrophy rates can be determined by either measuring the longitudinal volume change of the baseline and the mapped hippocampus using mesh-based methods or by using Jacobian integration. The same workflow is applicable for clinical trials in which the assessment of radiation induced brain atrophy is of interest.

In subjects that receive HA-PCI, a subject must obtain a brain MRI and CT simulation scan, which is used for HA-PCI treatment planning. The MRI must be mapped to the CT scan prior hippocampus delineation. Using FASTSURF with approximately five input contours left and right hippocampus delineations can be obtained, and hippocampal avoidance region can be generated by expanding hippocampal contours by 5mm. To create the HA-PCI treatment plan, also the whole brain and lenses must be contoured, and the planning CT and accompanying contours must be transferred to a treatment planning system. Finally, a radiotherapy treatment plan can be determined which minimizes the dose to the lenses and the hippocampi.

7.4. Future perspectives and developments

FASTSURF

In the future, FASTSURF might find its role in creating standard reference segmentation sets and, as described in the clinical workflow section, it could be used in clinical trials in which changes of anatomical structures need to be determined. For these applications, the FASTSURF algorithm needs to be implemented in a delineation tool for ease of use. Furthermore, FASTSURF should be validated against fully manual and automatic segmentation methods using independently obtained manual sparse contours. For an even more complete validation, a multiple observer analysis should be performed, in which experienced and inexperienced/untrained observers use FASTSURF with a well-defined segmentation protocol.

FASTSURF should be further improved and developed to increase its performance. A possible improvement of FASTSURF might be the inclusion of drawing sparse contours in multiple directions (coronal/axial/sagittal) to increase segmentation accuracy and ease of use.

T2-weighted MRI and hippocampal subfield analysis

Almost a decade ago, special oblique coronal T2-weighted high field strength (3T-7T) MRI sequences have been developed to increase in-plane resolution of images (~0.5x0.5mm^2) and to provide better intensity contrast in the hippocampal region [38–43]. The slice thickness for those methods is usually thicker (~2.0mm), nevertheless the voxel volume can be reduced [38]. It has been shown that whole hippocampal volumes measured on T1- and T2-weighted MRI did not differ [44], but because T2-weighted images provide better intensity contrast in the hippocampal region and higher in-plane resolution, the study concludes that measurements on T2-weighted images are more sensitive [44]. Furthermore, visual evaluation on T2-weighted images seems to provide a better discrimination power between AD and MCI groups [45]. Finally, on T2-weighted MRI manually segmenting hippocampal subfields is possible [38–43,46,47]. Hippocampal subfield analysis was shown to be more sensitive to disease effects, as for instance in AD related neuronal tangles tend to happen most in the CA1 region of the hippocampus [48–50]. Also, annual atrophy rates in the CA regions and dentate gyrus have been shown to be significantly larger (~1% to 1.5% larger) in MCI than in controls [44]. However, manually segmenting hippocampal subfield is even more labour intensive then measuring whole hippocampal volumes, but atlas-based (semi-)automatic segmentation of these subfields seems to be possible [47,51,52]. These findings suggest that measuring hippocampal (subfield) volumes on high field strength T2

weighted MRI is more sensitive than measuring volume on T1 weighted MRI with lower (1.5T) field strength. But it should be noted that determining 1-3% atrophy rates in smaller regions might be more challenging than measuring whole hippocampus atrophy rates on clinical T1-weighted MRI.

Deep learning algorithms

Due to its popularity in many areas of medical image analysis in the last few years [53], the topic *deep learning* in the field of brain image analysis cannot be overlooked and will be briefly discussed.

Deep learning is a machine learning technique, which computes predictions using imaging features from large image databases. The difference between classical machine learning techniques and deep learning is that features are learned from data and are not designed by human engineers [54]. It is appearing that such self-learned imaging features may be useful for quantitative brain image analysis.

Deep learning consists of neural networks with multiple layers, usually more than five [53]. Neural networks are decision trees with features as inputs and an output prediction. Due to the large quantity of necessary training data, deep learning is computationally very expensive. However, advances in graphics processing units (GPU) massively accelerate deep learning training and makes deep learning algorithms more accessible [54].

Akkus et al. give an excellent overview of deep learning algorithms used for brain segmentation [53] and it is shown that convolutional neural networks (CNNs) are the most commonly used for image classification and segmentation. CNNs are a type of multi-layer models consisting of trainable filters and pooling operations which are alternatingly applied on the input images [55,56]. Akkus et al. also show that CNN-based deep learning approaches achieved high segmentation accuracy results [53]. For instance, dice overlap indices higher than 0.8 for sub-cortical structures were obtained in [57]. Similarly, for the hippocampus, WM, GM and CSF Dice overlap values of 80.45%, 86.15%, 89.46% and 84.25% were obtained with only five training MRIs, respectively [58]. The later article also shows that the dice overlap for the hippocampus increased to 84.91% using 25 training MRIs, concluding that segmentation accuracy can be increased using more training data. In [59] a CNN model was trained with a large dataset of automated FreeSurfer hippocampus segmentation and simulated data. CNN based segmentations correlated well with FreeSurfer segmentations, but less erroneous segmentations were observed, and the trained model was much faster in segmenting one hippocampus (<30s) compared to FSL-FIRST (~10min) or FreeSurfer (~4h).

The segmentation results of deep learning algorithms seem very promising. However, a large amount of training data is necessary for deep neural networks to make them robust for different types of scanners, image inhomogeneities or anatomical variation. FASTSURF could be used for to create such training datasets.

7.5. Conclusions

In this thesis it has been shown that testing and developing methods to analyse anatomical structures is challenging. The following main conclusions can be drawn from the work presented:

- For manual segmentations a well described outlining protocol and good guidelines are important (chapter 3). However, manually segmenting anatomical structures is subjective and segmentations can only be reproduced to some extent (chapter 2 and chapter 6). Furthermore, outlining brain structures on every slice of T1-weighted 1.5T or 3T MRI is very inefficient and a lot of redundant information is provided.
- FASTSURF is a novel segmentation method which reduces segmentation time. It can be used to create standard segmentation reference sets or to outline brain structures to create avoidance treatment plans in radiotherapy (chapter 4 and chapter 5).
- Mesh representations of 3D objects are recommended for accurate segmentation comparison in which segmentations need to be mapped to the same space.
- Non-linear registration methods are recommended to determine hippocampal atrophy rates (chapter 6). FASTSURF could be used to obtain a baseline segmentation for such a registration-based setting.
- Generally, the quality of segmentations should be visually inspected (preferably in 3D) to avoid and correct for segmentation errors.

References

[1] Boccardi M, Bocchetta M, Apostolova LG, Barnes J, Bartzokis G, Corbetta G, et al. Delphi definition of the EADC-ADNI Harmonized Protocol for hippocampal segmentation on magnetic resonance. Alzheimer's Dement 2015;11:126–38.

[2] Frisoni GB, Jack CR, Bocchetta M, Bauer C, Frederiksen KS, Liu Y, et al. The EADC-ADNI Harmonized Protocol for manual hippocampal segmentation on magnetic resonance: evidence of validity. Alzheimers Dement 2015;11:111–25.

[3] Apostolova LG, Zarow C, Biado K, Hurtz S, Boccardi M, Somme J, et al. Relationship between hippocampal atrophy and neuropathology markers: a 7T MRI validation study of the EADC-ADNI Harmonized Hippocampal Segmentation Protocol. Alzheimers Dement 2015;11:139–50.

[4] Boccardi M, Ganzola R, Bocchetta M, Pievani M, Redolfi A, Bartzokis G, et al. Survey of protocols for the manual segmentation of the hippocampus: Preparatory steps towards a joint EADC-ADNI harmonized protocol. Adv Alzheimer's Dis 2011;2:111–25.

[5] Jack CR. MRI-Based Hippocampal Volume Measurements in Epilepsy. Epilepsia 1994;35:S21–9.

[6] Frisoni GB, Jack CR. Harmonization of magnetic resonance-based manual hippocampal segmentation: A mandatory step for wide clinical use. Alzheimer's Dement 2011;7:171–4.

[7] Morey RA, Selgrade ES, Wagner HR, Huettel SA, Wang L, McCarthy G. Scan-rescan reliability of subcortical brain volumes derived from automated segmentation. Hum Brain Mapp 2010;31:1751–62.

[8] Morey R a., Petty CM, Xu Y, Pannu Hayes J, Wagner HR, Lewis D V., et al. A comparison of automated segmentation and manual tracing for quantifying hippocampal and amygdala volumes. Neuroimage 2009;45:855–66.

[9] Sánchez-Benavides G, Gómez-Ansón B, Sainz A, Vives Y, Delfino M, Peña-Casanova J. Manual validation of FreeSurfer's automated hippocampal segmentation in normal aging, mild cognitive impairment, and Alzheimer Disease subjects. Psychiatry Res - Neuroimaging 2010;181:219–25.

[10] Dewey J, Hana G, Russell T, Price J, McCaffrey D, Harezlak J, et al. Reliability and validity of MRI-based automated volumetry software relative to auto-assisted manual measurement of subcortical structures in HIV-infected patients from a multisite study. Neuroimage 2010;51:1334–44.

[11] Lehmann M, Douiri A, Kim LG, Modat M, Chan D, Ourselin S, et al. Atrophy patterns in Alzheimer's disease and semantic dementia: A comparison of FreeSurfer and manual volumetric measurements. Neuroimage 2010;49:2264–74.

[12] Patenaude B, Smith SM, Kennedy DN, Jenkinson M. A Bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage 2011;56:907–22.

[13] Grimm O, Pohlack S, Cacciaglia R, Plichta M, Demirakca T, Flor H. Amygdala and hippocampal volume: A comparison between manual segmentation, Freesurfer and VBM. J Neurosci Methods 2015;253:254–61.

[14] Pipitone J, Park MTM, Winterburn J, Lett TA, Lerch JP, Pruessner JC, et al. Multi-atlas segmentation of the whole hippocampus and subfields using multiple automatically generated templates. Neuroimage 2014;101:494–512.

[15] Zhu H, Cheng H, Yang X, Fan Y, Alzheimer's Disease Neuroimaging Initiative. Metric Learning

for Multi-atlas based Segmentation of Hippocampus. Neuroinformatics 2017;15:41–50.

[16] Dill V, Franco AR, Pinho MS. Automated methods for hippocampus segmentation: the evolution and a review of the state of the art. Neuroinformatics 2015;13:133–50.

[17] Despotović I, Goossens B, Philips W. MRI segmentation of the human brain: challenges, methods, and applications. Comput Math Methods Med 2015;2015:450341.

[18] Ashton E a, Parker KJ, Berg MJ, Chen CWCCW. A novel volumetric feature extraction technique, with applications to MR images. Proc Int Conf Image Process 1997;16:365–71.

[19] Ghanei A, Soltanian-Zadeh H, Windham JP. A 3D deformable surface model for segmentation of objects from volumetric data in medical images. Comput Biol Med 1998;28:239–53.

[20] Kelemen A, Székely G, Gerig G. Elastic model-based segmentation of 3-D neuroradiological data sets. IEEE Trans Med Imaging 1999;18:828–39.

[21] Shen D, Moffat S, Resnick SM, Davatzikos C. Measuring size and shape of the hippocampus in MR images using a deformable shape model. Neuroimage 2002;15:422–34.

[22] Chupin M, Mukuna-Bantumbakulu AR, Hasboun D, Bardinet E, Baillet S, Kinkingnéhun S, et al. Anatomically constrained region deformation for the automated segmentation of the hippocampus and the amygdala: Method and validation on controls and patients with Alzheimer's disease. Neuroimage 2007;34:996–1019.

[23] Christensen GE, Joshi SC, Miller MI. Volumetric transformation of brain anatomy. IEEE Trans Med Imaging 1997;16:864–77.

[24] Haller JW, Banerjee A, Christensen GE, Gado M, Joshi S, Miller MI, et al. Three-dimensional hippocampal MR morphometry with high-dimensional transformation of a neuroanatomic atlas. Radiology 1997;202:504–10.

[25] Pluta J, Avants BB, Glynn S, Awate S, Gee JC, Detre JA. Appearance and incomplete label matching for diffeomorphic template based hippocampus segmentation. Hippocampus 2009;19:565–71.

[26] Pérez de Alejo R, Ruiz-Cabello J, Cortijo M, Rodriguez I, Echave I, Regadera J, et al. Computer-assisted enhanced volumetric segmentation magnetic resonance imaging data using a mixture of artificial neural networks. Magn Reson Imaging 2003;21:901–12.

[27] Freeborough PA, Fox NC. The boundary shift integral: an accurate and robust measure of cerebral volume changes from registered repeat MRI. IEEE Trans Med Imaging 1997;16:623–9.

[28] Boyes RG, Rueckert D, Aljabar P, Whitwell J, Schott JM, Hill DLG, et al. Cerebral atrophy measurements using Jacobian integration: Comparison with the boundary shift integral. Neuroimage 2006;32:159–69.

[29] Barnes J, Boyes RG, Lewis EB, Schott JM, Frost C, Scahill RI, et al. Automatic calculation of hippocampal atrophy rates using a hippocampal template and the boundary shift integral. Neurobiol Aging 2007;28:1657–63.

[30] Prados F, Cardoso MJ, Leung KK, Cash DM, Modat M, Fox NC, et al. Measuring brain atrophy with a generalized formulation of the boundary shift integral. Neurobiol Aging 2015;36 Suppl 1:S81-90.

[31] Leung KK, Clarkson MJ, Bartlett JW, Clegg S, Jack CR, Weiner MW, et al. Robust atrophy rate measurement in Alzheimer's disease using multi-site serial MRI: Tissue-specific intensity normalization and parameter selection. Neuroimage 2010;50:516–23.

[32] Cover KS, van Schijndel RA, Versteeg A, Leung KK, Mulder ER, Jong RA, et al. Reproducibility of hippocampal atrophy rates measured with manual, FreeSurfer, AdaBoost, FSL/FIRST and the MAPS-HBSI methods in Alzheimer's disease. Psychiatry Res Neuroimaging 2016;252:26–35.

[33] Crum WR, Scahill RI, Fox NC. Automated hippocampal segmentation by regional fluid registration of serial MRI: validation and application in Alzheimer's disease. Neuroimage 2001;13:847–55.

[34] van de Pol LA, Barnes J, Scahill RI, Frost C, Lewis EB, Boyes RG, et al. Improved reliability of hippocampal atrophy rate measurement in mild cognitive impairment using fluid registration. Neuroimage 2007;34:1036–41.

[35] Henneman WJP, Sluimer JD, Barnes J, Van Der Flier WM, Sluimer IC, Fox NC, et al. Hippocampal atrophy rates in Alzheimer disease: Added value over whole brain volume measures. Neurology 2009;72:999–1007.

[36] Yushkevich P a., Avants BB, Das SR, Pluta J, Altinay M, Craige C. Bias in estimation of hippocampal atrophy using deformation-based morphometry arises from asymmetric global normalization: An illustration in ADNI 3 T MRI data. Neuroimage 2010;50:434–45.

[37] Schuff N, Woerner N, Boreta L, Kornfield T, Shaw LM, Trojanowski JQ, et al. MRI of hippocampal volume loss in early Alzheimer's disease in relation to ApoE genotype and biomarkers. Brain 2009;132:1067–77.

[38] Malykhin N V., Lebel RM, Coupland NJ, Wilman AH, Carter R. In vivo quantification of hippocampal subfields using 4.7 T fast spin echo imaging. Neuroimage 2010;49:1224–30.

[39] La Joie R, Fouquet M, Mézenge F, Landeau B, Villain N, Mevel K, et al. Differential effect of age on hippocampal subfields assessed using a new high-resolution 3T MR sequence. Neuroimage 2010;53:506–14.

[40] Mueller SG, Stables L, Du AT, Schuff N, Truran D, Cashdollar N, et al. Measurement of hippocampal subfields and age-related changes with high resolution MRI at 4 T. Neurobiol Aging 2007;28:719–26.

[41] Mueller SG, Schuff N, Raptentsetsang S, Elman J, Weiner MW. Selective effect of Apo e4 on CA3 and dentate in normal aging and Alzheimer's disease using high resolution MRI at 4 T. Neuroimage 2008;42:42–8.

[42] Mueller SG, Schuff N, Yaffe K, Madison C, Miller B, Weiner MW. Hippocampal atrophy patterns in mild cognitive impairment and alzheimer's disease. Hum Brain Mapp 2010;31:1339–47.

[43] Mueller SG, Weiner MW. Selective effect of age, Apo e4, and Alzheimer's disease on hippocampal subfields. Hippocampus 2009;19:558–64.

[44] Das SR, Avants BB, Pluta J, Wang H, Suh JW, Weiner MW, et al. Measuring longitudinal change in the hippocampal formation from in vivo high-resolution T2-weighted MRI. Neuroimage 2012;60:1266–79.

[45] Fischbach-Boulanger C, Fitsiori A, Noblet V, Baloglu S, Oesterle H, Draghici S, et al. T1 or T2 weighted MRI: What is the best choice to evaluate the atrophy of the hippocampus? Eur J Neurol 2018.

[46] Yushkevich PA, Wang H, Pluta J, Das SR, Craige C, Avants BB, et al. Nearly automatic segmentation of hippocampal subfields in in vivo focal T2-weighted MRI. Neuroimage 2010;53:1208–24.

[47] Pluta J, Yushkevich P, Das S, Wolk D. In vivo analysis of hippocampal subfield atrophy in mild

cognitive impairment via semi-automatic segmentation of T2-weighted MRI. J Alzheimer's Dis 2012;31:85–99.

[48] Bobinski M, Wegiel J, Tarnawski M, Bobinski M, Reisberg B, de Leon MJ, et al. Relationships between regional neuronal loss and neurofibrillary changes in the hippocampal formation and duration and severity of Alzheimer disease. J Neuropathol Exp Neurol 1997;56:414–20.

[49] Bobinski M, De Leon MJ, Tarnawski M, Wegiel J, Bobinski M, Reisberg B, et al. Neuronal and volume loss in CA1 of the hippocampal formation uniquely predicts duration and severity of Alzheimer disease. Brain Res 1998;805:267–9.

[50] West MJ, Coleman PD, Flood DG, Troncoso JC. Differences in the pattern of hippocampal neuronal loss in normal ageing and Alzheimer's disease. Lancet 1994;344:769–72.

[51] Yushkevich PA, Pluta JB, Wang H, Xie L, Ding SL, Gertje EC, et al. Automated volumetry and regional thickness analysis of hippocampal subfields and medial temporal cortical structures in mild cognitive impairment. Hum Brain Mapp 2015;36:258–87.

[52] Wisse LEM, Kuijf HJ, Honingh AM, Wang H, Pluta JB, Das SR, et al. Automated hippocampal subfield segmentation at 7T MRI. Am J Neuroradiol 2016;37:1050–7.

[53] Akkus Z, Galimzianova A, Hoogi A, Rubin DL, Erickson BJ. Deep Learning for Brain MRI Segmentation: State of the Art and Future Directions. J Digit Imaging 2017;30:449–59.

[54] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521:436–44.

[55] LeCun Y, Boser B, Denker JS, Henderson D, Howard RE, Hubbard W, et al. Backpropagation Applied to Handwritten Zip Code Recognition. Neural Comput 1989;1:541–51.

[56] Krizhevsky A, Sutskever I, Hinton GE. ImageNet classification with deep convolutional neural networks. Proc. 25th Int. Conf. Neural Inf. Process. Syst. - Vol. 1, Curran Associates Inc.; 2012, p. 1097–105.

[57] Bao S, Chung ACS. Multi-scale structured CNN with label consistency for brain MR image segmentation. Comput Methods Biomech Biomed Eng Imaging Vis 2018;6:113–7.

[58] Chen H, Dou Q, Yu L, Qin J, Heng PA. VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images. Neuroimage 2018;170:446–55.

[59] Thyreau B, Sato K, Fukuda H, Taki Y. Segmentation of the hippocampus by transferring algorithmic knowledge for large cohort processing. Med Image Anal 2018;43:214–28.