Multi Atlas Segmentation applied to in vivo mouse brain MRI

Ma Da[1,2], M. Jorge Cardoso[1], Marc Modat[1], Nick Powell[1,2], Holly Holmes[2], Mark Lythgoe[2], Sébastien Ourselin[1,3]

[1] Centre for Medical Imaging Computing, Department of Medical Physics and Bioengineering, University College London, UK,
[2] Centre for Advanced Biomedical Imaging, University College London, UK,
[3] Dementia Research Centre, Institute of Neurology, WC1N 3BG, University College London, UK.

Abstract. Multi-atlas segmentation propagation has evolved quickly in recent years, becoming a state-of-the-art method for automatic structural parcellation for brain MRI. However, few studies have applied these methods to preclinical research. In this study, we present a fully automatic multi-atlas segmentation pipeline for mouse brain MRI tissue parcellation. The pipeline adopts the Multi-STEPS multi-atlas segmentation algorithm, which utilises a locally normalised cross correlation (LNCC) similarity metric for atlas selection and an extended STAPLE framework for multi-label fusion. The segmentation accuracy of the pipeline was evaluated using an *in vivo* mouse brain atlas with pre-segmented manual labels as gold standard, and optimised parameters were obtained. Results show a mean Dice similarity coefficient of 0.839 over all the structures and for all the samples in the database, significantly higher than in a single atlas propagation strategy, and also generally higher than STAPLE strategy, although the improvement is not significant.

1 Introduction

Mice share more than 80% of genomes with human, making it a good animal for preclinical study of human brain diseases, such as Alzheimers disease and Downs syndrome. Preclinical studies normally require relatively large sample size, thus an accurate, robust and reproducible method for quantitative analysis of preclinical MRI images is necessary for high-throughput studies. More specifically, structural parcellation enables the study of volume, shape and morphological characteristics of key brain structures. Despite the labour intensive and expert-dependent nature of the task, manual labelling of anatomical structures is still standard practice in mouse brain MRI studies [1, 2]. Various automated labelling algorithms have thus been developed to address these limitations [3, 4]. Segmentation propagation is a method where a template, i.e. an accurate manual delineation of anatomical structures that follows a well-defined segmentation protocol, is propagated to a query image through the process of image registration. Although the accuracy and efficiency of image registration algorithms has been constantly improving, the segmentation performance is still

limited by inaccuracies in the registration process, especially between morphologically dissimilar subjects. This problem can be greatly ameliorated by propagating multiple image templates and subsequently fusing them into a consensus segmentation through a process known as label fusion [5–7]. A great deal of effort has been put in exploring the structural parcellation of human brain MRI [8–11]. However, in preclinical research (e.g. mouse model), a study about structural parcellation techniques is still lacking. Maheswaran *et al.* compared a single atlas segmentation propagation with deformation based morphormetry (DBM) [12], and concluded that atlas-base method can identify longitudinal and cross-sectional group difference, but is less sensitive to much smaller regional changes compared to DBM. Artaechevarria *et al.* [13] adopted a weighted majority voting label fusion using an ex vivo mouse brain MRI atlas containing 10 individual samples with 20 manually labelled structures [14]. In cases where only one single template is available, Chakravarty *et al.* [15] proposed to first propagate the template to a set of unlabelled images with traditional single-atlas segmentation propagation and then propagate the resultant set of segmentations to another new image using majority voting label fusion, which resulted in an increase in performance. However, with the improvement of hardware and scanning protocols, mouse studies are moving from *ex vivo* to *in vivo* imaging, leading to much lower contrast to noise ratios (CNR) and signal to noise ratios (SNR). Bai *et al.* [3] have recently published a study using majority voting and STAPLE (Simultaneous Truth and Performance Level Estimation) multi-atlas label fusion to in vivo mouse brain MRI, and compare the improvement gained with that of the non-rigid image registration. In this paper, we use a new multi-atlas based structural parcellation method, Multi-STEPS (Multi-label Similarity and Truth Estimation for Propagated Segmentations) [16], on *in vivo* mouse brain MRI images. We developed a fully automated pipeline for brain parcellation and optimised the fusion strategy parameters using a leave-one-out cross validation.

2 Methods

In this section we will describe the steps used for multi-atlas structural parcellation. A schematic diagram of the pipeline is shown in Figure 1.

2.1 Brain extraction

Starting from a set of template images with associated tissue parcellations and brain masks, the first step of the pipeline was to create a brain mask for the query image. This goal was achieved by propagating the brain masks defined on the template images using the Multi-Atlas Propagation and Segmentation (MAPS) strategy developed by Leung *et al.* [17].

2.2 Bias field correction

MR images are corrupted by intensity non-uniformity, or bias, cause by the inhomogeneity of the RF excitation field, the spatially nonuniform receiver coil

sensitivity profiles, the induced currents and standing wave affects [18]. Intensity non-uniformity leads to misalignment in the registration process due to corrupted intensity profile. We thus used the N3 intensity non-uniformity correction algorithm developed by Sled *et al.* [18] to correct the bias field.

2.3 Template registration

After correction of the intensity non-uniformity, we first globally and then non-linearly registered the masked template images to the query image. The global registration was performed using a block-matching approach [19]. A parametric approach based on a cubic B-Spline parameterisation [20] was used for non-linear registration. We used the efficient implementation proposed by Modat *et al.* [21]. The resulting deformation fields obtained from the registrations were then used to resample the labels from the template image spaces to the query image. Nearest-neighbour interpolation was used to preserve the integer nature of the original labels.

2.4 Multi-atlas label fusion

The label fusion was conducted using the Multi-STEPS algorithm developed by Cardoso *et al.* [16]. Multi-STEPS is an extension of the original STAPLE algorithm [5, 6] with several improvements. Firstly, it includes a Markov Random Field (MRF) used in an iterative manner to maintain spatial consistency. It also incorporates a template selection step using a ranking strategy based on the locally normalised cross correlation (LNCC) over a local Gaussian window. This fusion strategy has two main user-defined parameters: the width of the Gaussian kernel for image comparison and the number of labels to fuse after ranking. The optimisation of these parameters is described in section 3.2.
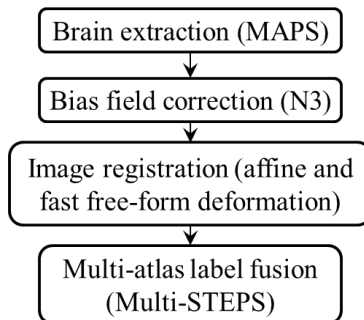
```
┌────────────────────────────┐
│   Brain extraction (MAPS)  │
└────────────────────────────┘
              ↓
┌────────────────────────────┐
│  Bias field correction (N3)│
└────────────────────────────┘
              ↓
┌────────────────────────────┐
│  Image registration (affine and
│   fast free-form deformation)
└────────────────────────────┘
              ↓
┌────────────────────────────┐
│   Multi-atlas label fusion │
│       (Multi-STEPS)        │
└────────────────────────────┘
```

Fig. 1. Proposed multi-atlas segmentation propagation pipeline.

3 Validation and results

This section will present the optimisation of the fusion strategy parameters as well as the segmentation performance evaluation.

3.1 Data

To evaluate the performance of the method and optimise the parameters of the proposed pipeline, we used a previously described *in vivo* mouse brain MRI database containing 12 individual brain T2* MRI of 12-14 weeks old C57BL/6J mouse. Each MRI brain image is associated with 20 manually delineated structures. Detailed scanning parameters are described in [1]. Due to missing labels in 3 of the 12 available templates, only 9 images and associated parcellations were included in this study.

3.2 Parameter optimisation

The Multi-STEPS label fusion performance depends mostly on the width of the Gaussian kernel and the number of top ranked templates used for fusion. We will thus focus on optimising these parameters.

A leave-one-out cross validation was performed to assess the segmentation accuracy as well as to optimise the parameters of Multi-STEPS. For each of the 9 samples, the remaining 8 samples were used as template sets for multi-atlas segmentation. The average Dice similarity coefficient between the automatic segmentation and the manual segmentation of all the structures for all the individual sample images was calculated. We ran the leave-one-out validation for each combination of parameters, with the Gaussian kernel standard deviation varying from 1 to 6 (step of 0.5) and the number of templates used from 3 to 8. The parameter combination that gave the highest Dice similarity coefficient was selected and regarded as the optimal combination.

The results of the Multi-STEPS parameter optimisation are shown in Figure 2. The best combination of parameters was: number of local templates used equal to 6, with a Gaussian kernel with a standard deviation of 4. The corresponding average Dice similarity coefficient between automatic and manual segmentation, for all structures and templates, was 0.839 with standard deviation of 0.025.

Figure 2 shows that there is a large plateau zone (i.e. a small variation in Dice similarity coefficient) close to the optimal model parameters, indicating high stability of the pipeline with regards to the selection of parameters. Another possible explanation for the segmentation stability could be the smaller inter-template morphological variation for mice when compared to humans, thus making the fusion less dependent on the parameter selection. Example images of segmentation results and the correspond-ing manual labelling are presented in Figure 3.

We also compared the average Dice similarity coefficient of our pipeline result with the single template-based segmentation propagation and STAPLE. For single template-based segmentation, we propagated all templates and averaged the
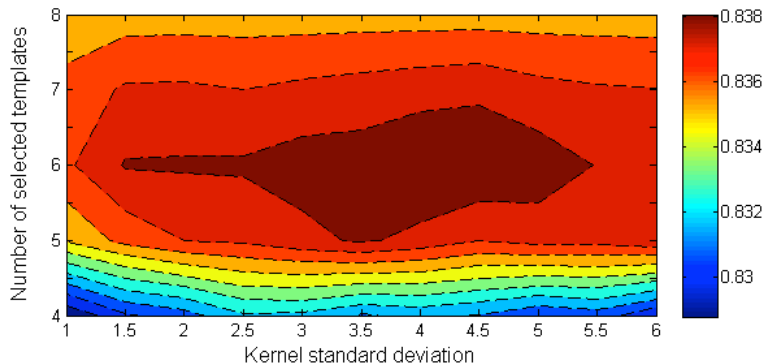
Fig. 2. Average Dice similarity coefficients for different combinations of Gaussian kernel standard deviation and number of selected templates in Multi-STEPS algorithm. The optimal parameter were found to be: number of templates = 6 and Gaussian kernel standard deviation = 4 (Dice = 0.839).

Dice similarity coefficients. The results are shown in Figure 4. The average Dice similarity coefficient of our pipeline was in general higher than both the single-atlas method and STAPLE for most of the structures. When compared with single-atlas method, significant improvements were found in External Capsule, Ant Commissure, Internal Capsule, Ventricles and Fimbria. The improvements compared to STAPLE were not statistically significant. We believe the low samples number of the manual segmentations is the main cause of this effect. Also, the fact that all the manual labels come from the same atlas in leave-one-out cross validation may effectively result in relatively high Dice similarity coefficient for STAPLE, which may not be the case for newly acquired images. Further research will explore and characterise these limitations.

3.3 Pipeline robustness testing

In order to test the ability to segment new unseen datasets, we acquired *in vivo* images of mouse brains and applied the pipeline to obtain the corresponding anatomical labels. These scans were obtained using a Varian VNMRS 9.4 Tesla MRI system (Agilent Technologies Inc. Palo Alto CA, USA). A 72-mm volume coil (RAPID Biomedical GmbH, Würzburg, Germany) was used for excitation and a quadrature mouse brain surface coil (Bruker Biospin GmbH, Ettlingen, Germany) was used for signal detection. T1 weighted contrast was achieved using an efficient fast spin echo (FSE) sequence. The data was acquired with the parameters $TR/TE_{eff} = 2500/12.7$ ms, ETl = 4, 1 average, field of view $= 24.6 \times 16.8 \times 12.0$, with spatial resolution of $150 \times 150 \times 150$ isotropic. The total *in vivo* imaging time was approximately 1 hour and 30 minutes. An example of the segmentation results in one of the scanned subjects is shown in Figure 5. Due to the lack of available manual segmentations, quantitative analysis could
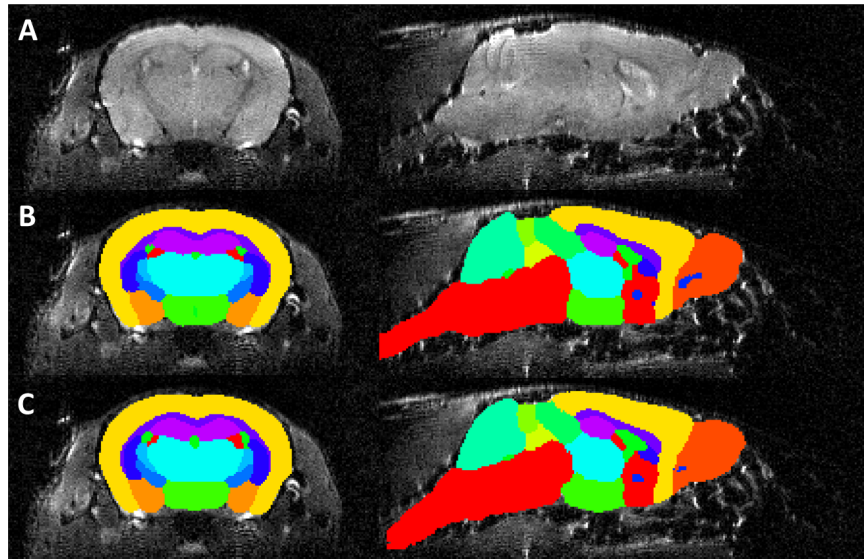
Fig. 3. Sample images showing the coronal view (left) and sagittal view (right) of the template image (A), overlaid with the multi-atlas segmentation results (B) and the manual labels (C).

not be performed on these new datasets. However, visual inspection has shown good segmentation accuracy.

4 Discussion and conclusion

The proposed work utilises the state-of-the-art multi-atlas segmentation propagation method Multi-STEPS, along with the fast free-form deformation registration algorithm and other pre-processing techniques such as brain extraction (MAPS) and intensity non-uniformity correction (N3), to create an integrated and fully automated pipeline for brain segmentation. This paper presents the successful application of advanced multi-atlas segmentation techniques for *in vivo* mouse brain parcellation.

The optimised Multi-STEPS parameters were chosen based on the average Dice similarity coefficient over all the structures and for all samples in the template data-base. However, one should note that the Dice similarity coefficient intrinsically favours large structures. For example, small structures (e.g. external capsule, anterior commissure) show much worse performance than larger structures (e.g. hippocampus, neocortex) due to local registration errors, inter-template morphological variability and human segmentation consistency. Contrast between structures can also have a detrimental effect on segmentation performance. The lack of contrast between some neighbouring anatomical regions can lead to decreased performance as the registration algorithm will have
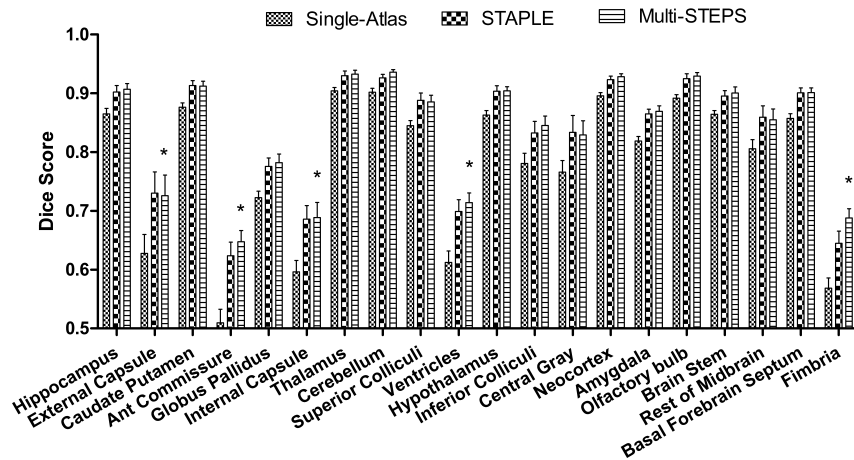
Fig. 4. Average Dice similarity coefficient comparison between traditional single-atlas segmentation propagation, STAPLE and the Multi-STEPS method. Two-way ANOVA statistical test was performed. Significant differences were found for some structures (*) between single-atlas segmentation propagation and Multi-STEPS method (*: p < 0.001). The improvement of STEPS compared to STAPLE does not reach statistically significance.

to rely on the regularisation term (rather than on image features) for accurate structural matching. Conversely, the nature of most measure of similarity will also lead to a registration algorithm that is governed by high contrast edges, possibly reducing the propagation accuracy in low-contrast areas.

Compared to human brain MRI segmentation studies, the availability of mouse templates and the amount of information about the segmentation protocols is very limited. Subsequently, label fusion techniques are limited in performance by several different factors. First, the templates used for the presented work are limited in number and are defined only on T2* images, impeding their direct application to other imaging modalities. While certain similarity measures for image registration can deal with multi-modal images, the lack of contrast between certain anatomical structures in other modalities will reduce the accuracy of the parcellation algorithm. Second, the lack of anatomical standardisation and vague definition of the segmentation protocol reduces the consistency between human raters. Finally, intra- and inter-rater labelling variability has not been assessed in mice. Since the manual segmentations are used for comparison, as they are considered as gold standard, the information about intra- and inter-rater labelling variability is of critical importance because it represents the theoretical performance upper limit for automated methods.

Lastly, the estimated optimal parameters and segmentation performance were only assessed within the template database. Although the application to new testing data has good visual assessed segmentation accuracy, further validation is still necessary in order to enable the unsupervised use of this algorithm
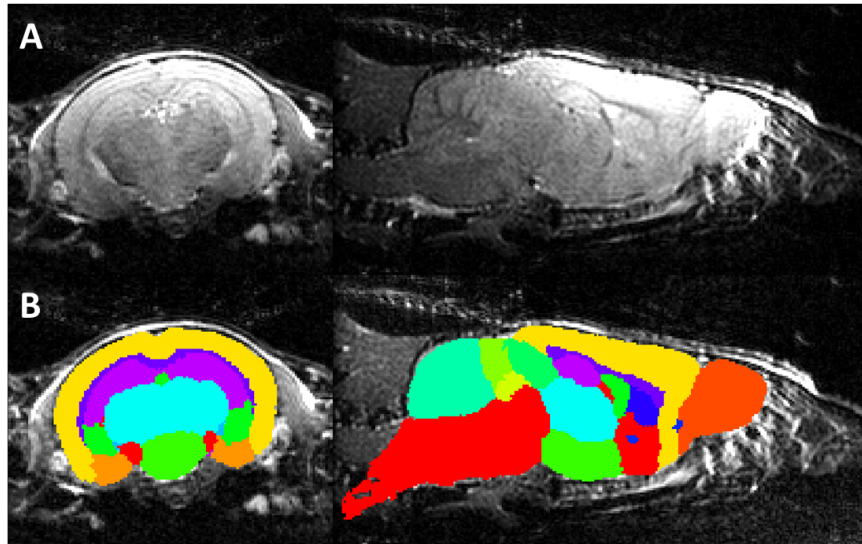
Fig. 5. Sample images showing the coronal view (left) and sagittal view (right) of the test image data (A), overlaid with the multi-atlas segmentation results (B).

in a pre-clinical setting and for different mouse models. Future work will also include the optimisation of the registration parameters, which are here considered as fixed.

Acknowledgement

References

1. Ma, Y., Smith, D., Hof, P.R., Foerster, B., Hamilton, S., Blackband, S.J., Yu, M., Benveniste, H.: In vivo 3D digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. Frontier Neuroanatomy **2** (Apr 2008) 1–10
2. Richards, K., Watson, C., Buckley, R.F., Kurniawan, N.D., Yang, Z., Keller, M.D., Beare, R., Bartlett, P.F., Egan, G.F., Galloway, G.J., Paxinos, G., Petrou, S., Reutens, D.C.: Segmentation of the mouse hippocampal formation in magnetic resonance images. NeuroImage **58**(3) (Oct 2011) 732–740

3. Bai, J., Trinh, T.L.H., Chuang, K.H., Qiu, A.: Atlas-based automatic mouse brain image segmentation revisited: model complexity vs. image registration. Magnetic Resonance Imaging **30**(6) (Jul 2012) 789–798

4. Lee, J., Jomier, J., Aylward, S., Tyszka, M., Moy, S., Lauder, J., Styner, M.: Evaluation of atlas based mouse brain segmentation. In: Proceedings of SPIE. (2009) 725943–725949

5. Rohlfing, T., Russakoff, D.B., Maurer, C.R.: Performance-based classifier combination in atlas-based image segmentation using expectation-maximization parameter estimation. IEEE Transactions on Medical Imaging **23**(8) (Aug 2004) 983–994

6. Warfield, S.K., Zou, K.H., Wells III, W.M.: Simultaneous truth and performance level estimation (STAPLE): an algorithm for the validation of image segmentation. IEEE Transactions on Medical Imaging **23**(7) (Jul 2004) 903–921

7. Aljabar, P., Heckemann, R., Hammers, A., Hajnal, J.V., Rueckert, D.: Classifier selection strategies for label fusion using large atlas databases. In: International Conference on Medical Image Computing and Computer Assisted Intervention. Volume 10. (2007) 523–531

8. Heckemann, R.A., Keihaninejad, S., Aljabar, P., Rueckert, D., Hajnal, J.V., Hammers, A., Initiative, A.D.N.: Improving intersubject image registration using tissue-class information benefits robustness and accuracy of multi-atlas based anatomical segmentation. NeuroImage **51**(1) (May 2010) 221–227

9. Hammers, A., Allom, R., Koepp, M.J., Free, S.L., Myers, R., Lemieux, L., Mitchell, T.N., Brooks, D.J., Duncan, J.S.: Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe. Human Brain Mapping **19**(4) (Aug 2003) 224–247

10. Heckemann, R.A., Hajnal, J.V., Aljabar, P., Rueckert, D., Hammers, A.: Automatic anatomical brain MRI segmentation combining label propagation and decision fusion. NeuroImage **33**(1) (Oct 2006) 115–126

11. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., Montillo, A., Makris, N., Rosen, B., Dale, A.M.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron **33**(3) (Jan 2002) 341–355

12. Maheswaran, S., Barjat, H., Bate, S.T., Aljabar, P., Hill, D.L.G., Tilling, L., Upton, N., James, M.F., Hajnal, J.V., Rueckert, D.: Analysis of serial magnetic resonance images of mouse brains using image registration. NeuroImage **44**(3) (Feb 2009) 692–700

13. Artaechevarria, X., Munoz-Barrutia, A., Ortiz-de Solorzano, C.: Combination strategies in multi-atlas image segmentation: application to brain MR data. IEEE Transactions on Medical Imaging **28**(8) (Aug 2009) 1266–1277

14. Ma, Y., Hof, P.R., Grant, S.C., Blackband, S.J., Bennett, R., Slatest, L., McGuigan, M.D., Benveniste, H.: A three-dimensional digital atlas database of the adult C57BL/6J mouse brain by magnetic resonance microscopy. Neuroscience **135**(4) (Sep 2005) 1203–1215

15. Chakravarty, M.M., van Eede, M.C., Lerch, J.P.: Improved segmentation of mouse MRI data using multiple automatically generated templates. In: International Society for Magnetic Resonance in Medicine. Volume 15. (2011) 134

16. Cardoso, M.J., Modat, M., Keihaninejad, S., Cash, D., Ourselin, S.: Multi-STEPS: Multi-label similarity and truth estimation for propagated segmentations. In: Mathematical Methods in Biomedical Image Analysis, 2012 IEEE Workshop on. (2012) 153–158

17. Leung, K.K., Barnes, J., Modat, M., Ridgway, G.R., Bartlett, J.W., Fox, N.C., Ourselin, S., Initiative, A.D.N.: Brain MAPS: an automated, accurate and robust brain extraction technique using a template library. NeuroImage **55**(3) (Apr 2011) 1091–1108

18. Sled, J., Zijdenbos, A., Evans, A.: A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Transactions on Medical Imaging **17**(1) (Feb 1998) 87–97

19. Ourselin, S., Roche, A., Subsol, G., Pennec, X., Ayache, N.: Reconstructing a 3D structure from serial histological sections. Image and Vision Computing **19**(1-2) (Jan 2001) 25–31

20. Rueckert, D., Sonoda, L., Hayes, C., Hill, D., Leach, M., Hawkes, D.: Nonrigid registration using free-form deformations: Application to breast MR images. IEEE Transactions on Medical Imaging **18**(8) (Aug 1999) 712–721

21. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine **98**(3) (Jun 2010) 278–84