



Unsupervised Deep Learning for Bayesian Brain MRI Segmentation

Adrian V. Dalca^{1,2}(✉), Evan Yu³, Polina Golland², Bruce Fischl¹,
Mert R. Sabuncu^{3,4}, and Juan Eugenio Iglesias^{1,2,5}

¹ Martinos Center for Biomedical Imaging, MGH, Harvard Medical School,
Boston, USA

² Computer Science and Artificial Intelligence Laboratory (CSAIL),
MIT, Cambridge, USA
adalca@mit.edu

³ Meinig School of Biomedical Engineering, Cornell University, Ithaca, USA

⁴ School of Electrical and Computer Engineering, Cornell University, Ithaca, USA

⁵ Centre for Medical Image Computing (CMIC),
University College London, London, UK

Abstract. Probabilistic atlas priors have been commonly used to derive adaptive and robust brain MRI segmentation algorithms. Widely-used neuroimage analysis pipelines rely heavily on these techniques, which are often computationally expensive. In contrast, there has been a recent surge of approaches that leverage deep learning segmentation tools that are computationally efficient at test time. However, most of these strategies rely on supervised learning from manually annotated images and are therefore sensitive to the intensity profiles in the training dataset. A deep learning-based segmentation model for a new image dataset (e.g., of different contrast), usually requires a new labeled training dataset, which can be prohibitively expensive, or suboptimal *ad hoc* adaptation or augmentation approaches. In this paper, we propose an alternative strategy that combines conventional probabilistic atlas-based segmentation with deep learning, enabling training of a segmentation model for new MRI scans without the need for any manually segmented images. Our experiments include thousands of brain MRI scans and demonstrate that the proposed method achieves good accuracy for a brain MRI segmentation task for different MRI contrasts, requiring only approximately 15 s at test time on a GPU. The code is freely available at <http://voxelmorph.mit.edu>.

Keywords: Unsupervised learning · Segmentation · Brain MRI

1 Introduction

Bayesian segmentation of medical images, particularly in the context of brain MRI, is a well-studied problem. Probabilistic models for image segmentation frequently exploit atlas priors, and account for variations in contrast and imaging

artifacts such as MR inhomogeneity [19,21]. Most of the popular neuroimage processing pipelines rely on segmentation algorithms based on these ideas [2,8,15]. While these tools achieve high robustness to changes in MRI contrast of the input scan, they are computationally demanding (e.g., 23 min per image using a multi-threaded setup [16]), which limits their deployment at scale and in time-sensitive applications. Therefore, there is a need for computationally efficient methods that are contrast-adaptive, requiring no additional labeled training images to segment a new dataset.

Recently, there has been a surge in the application of deep learning (DL) techniques to medical image segmentation, often based on convolutional neural networks (CNN) that excel at learning contextually important multi-scale features. An advantage of these methods is their computational efficiency at test (segmentation) time, offering the potential to use automatic segmentation in new application areas, and large datasets [11,17]. Moreover, these algorithms can be combined with atlas priors for increased robustness [5,12,13]. However, DL based techniques are notoriously sensitive to changes in the image intensity data distribution. For example, upgrades to MRI scanners or changes in pulse sequence, field strength, or RF coils, can alter contrast properties and dramatically reduce the performance of a CNN-based segmentation model [10]. This issue can be alleviated via domain adaptation or data augmentation, which requires simulating expected variations. However, even with additional data, these methods only partially close the gap with the fully supervised setting [14]. Furthermore, the dependency on manually annotated datasets means that existing DL approaches are only applicable if enough resources are available to compile the required training data. This is often infeasible, for example in the context of continuously upgrading imaging technologies.

In this paper, we consider the scenario in which we have a general probabilistic atlas prior and a collection of images with *no manual delineations*. The probabilistic atlas is a volume where each voxel has an associated vector with the prior probabilities of observing the various segmentation labels at that location. Our approach assumes the availability of such an atlas (in brain imaging, they are readily available), and is independent of how it was created. For example, it could have been obtained by averaging a collection of manually annotated volumes of a different imaging modality, or derived from an anatomical template after applying spatial blurring to account for spatial variability.

The main contribution of this paper is the integration of mathematical ideas from the Bayesian segmentation literature with an unsupervised deep learning framework, to achieve fast, contrast adaptive brain MRI segmentation. Specifically, we assume a probabilistic model, which requires estimation of parameters comprising an atlas deformation and image intensity statistics. The estimation of the atlas warp has traditionally relied on classic deformable registration algorithms [18], which are based on iterative optimization, and are therefore computationally expensive. Instead, we leverage recent advances in learning-based registration [3,4,20] to efficiently estimate the warp jointly with the intensity parameters. We use a novel interpretable loss function from the probabilistic

model via Bayesian inference. Integrating DL with Bayesian segmentation, we attain two highly desirable features. First, given a probabilistic atlas, the method is unsupervised and contrast adaptive. Training it does not require any ground truth segmentations. Second, the segmentation is efficient, requiring approximately 15s on a GPU.

2 Method

2.1 Segmentation as Bayesian Inference

Let \mathbf{I} represent the intensities of a 3D brain MRI scan, defined over a discrete domain $\Omega \subset \mathbb{R}^3$. Let \mathbf{S} be a corresponding discrete segmentation into L neuroanatomical labels. Bayesian segmentation relies on Bayes' rule to derive the posterior probability distribution of the segmentation given the input image. Then, the segmentation $\hat{\mathbf{S}}$ is estimated as the mode of this posterior:

$$\hat{\mathbf{S}} = \arg \max_{\mathbf{S}} p(\mathbf{S}|\mathbf{I}) = \arg \max_{\mathbf{S}} p(\mathbf{I}|\mathbf{S})p(\mathbf{S}). \quad (1)$$

The posterior distribution $p(\mathbf{S}|\mathbf{I})$ depends on two terms: a prior $p(\mathbf{S})$ and a likelihood $p(\mathbf{I}|\mathbf{S})$, in contrast to discriminative approaches which model $p(\mathbf{S}|\mathbf{I})$ directly. The prior represents knowledge about the spatial distribution of segmentation labels (i.e., underlying anatomy), and typically has the form of a probabilistic atlas endowed with a deformation model. The likelihood models the relationship between the segmentation and image intensities, including image artifacts such as noise and bias field. Both the prior and likelihood may have associated parameters, which we define as θ_S and θ_I , respectively. The former describes attributes such as label probabilities and an atlas deformation, while the latter typically includes image intensity statistics as a function of label.

The likelihood parameters may be learned from a training dataset, or estimated specifically for each test scan. We build on Bayesian segmentation models that follow the latter approach [2, 16, 19, 21, 22], enabling models to *adapt* to the intensity characteristics of input scans, making them robust to changes in MRI contrast. Expanding Eq. (1) to include model parameters, which we treat as random variables, yields:

$$\hat{\mathbf{S}} = \arg \max_{\mathbf{S}} \int_{\theta_S} \int_{\theta_I} p(\mathbf{S}|\theta_S, \theta_I, \mathbf{I})p(\theta_S, \theta_I|\mathbf{I})d\theta_S d\theta_I, \quad (2)$$

which is intractable. A standard approximation uses point estimates for the parameters. First, one estimates the mode of the parameter posterior distribution:

$$\{\hat{\theta}_S, \hat{\theta}_I\} = \arg \max_{\{\theta_S, \theta_I\}} p(\theta_S, \theta_I|\mathbf{I}) = \arg \max_{\{\theta_S, \theta_I\}} p(\theta_S)p(\theta_I) \sum_{\mathbf{S}} p(\mathbf{I}|\mathbf{S}, \theta_I)p(\mathbf{S}|\theta_S), \quad (3)$$

where we assume independence between the parameters of the prior and of the likelihood. The computation often requires estimating an atlas deformation in θ_S and intensity parameters in θ_I and is typically achieved with a combination of numerical optimization and the Expectation Maximization (EM) algorithm [6]. Given point estimates, the final segmentation is computed as:

$$\hat{S} = \arg \max_S p(S|\hat{\theta}_S, \hat{\theta}_I, \mathbf{I}) = \arg \max_S p(\mathbf{I}|S, \hat{\theta}_I)p(S|\hat{\theta}_S). \quad (4)$$

2.2 Proposed Model

Our model instantiation builds on existing work [2, 16, 19]. The prior is defined by a given probabilistic atlas \mathbf{A} , where $A(l, \mathbf{x})$ provides the probability of observing each neuroanatomical label $l = 1, \dots, L$ at each location $\mathbf{x} \in \Omega$. The atlas is deformed by a transform ϕ , parameterized by a stationary velocity field \mathbf{v} , (i.e., $\phi_v = \exp[\mathbf{v}]$, which guarantees diffeomorphic ϕ [1]). Therefore, the prior is parametrized by $\theta_S = \mathbf{v}$. Assuming independence over voxels:

$$p(S|\theta_S; \mathbf{A}) = p(S|\mathbf{v}; \mathbf{A}) = \prod_{j \in \Omega} A(S_j, \phi_v(\mathbf{x}_j)), \quad (5)$$

where S_j is the segmentation at voxel j , and \mathbf{x}_j is its spatial location. We discourage strongly varying deformations by penalizing the spatial gradient $\nabla \mathbf{u}_v$ of displacement \mathbf{u}_v , where $\phi_v = Id + \mathbf{u}_v$, i.e., $p(\theta_S; \lambda) = p(\mathbf{v}; \lambda) \propto \exp[-\lambda \|\nabla \mathbf{u}_v\|^2]$. The hyperparameter λ controls the strength of the penalty.

Conditioned on a segmentation, we assume that the observed intensities at different voxel locations are independent samples of Gaussian distributions:

$$p(\mathbf{I}|S, \theta_I) = p(\mathbf{I}|S, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{j \in \Omega} \mathcal{N}(I_j; \mu_{S_j}, \sigma_{S_j}^2), \quad (6)$$

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ is the Gaussian distribution, I_j is the image intensity at voxel j , and the likelihood parameters $\theta_I = \{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$ are L means μ_l and variances σ_l^2 , each associated with a different label l . We complete the model with a flat prior for these parameters: $p(\theta_I) \propto 1$. The model can be easily extended to the multi-spectral case (i.e., inputs with multiple MRI contrasts) by replacing means and variances by mean vectors and covariance matrices, respectively.

Learning. To avoid the computationally expensive optimization typically required for maximum *a posteriori* (MAP) estimation in Eq. (3), we propose to train a CNN to estimate model parameters directly from an input scan. Specifically, we design a CNN $g_{\theta_C}(\mathbf{I}, \mathbf{A}) = (\theta_S, \theta_I) = (\mathbf{v}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ with global convolutional parameters θ_C that takes as input a scan \mathbf{I} and probabilistic atlas \mathbf{A} , and outputs model parameters $\mathbf{v}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2$ for that scan. To learn the network parameters θ_C , we use a pool of N unlabeled scans $\{\mathbf{I}^n\}_{n=1}^N$ to minimize the

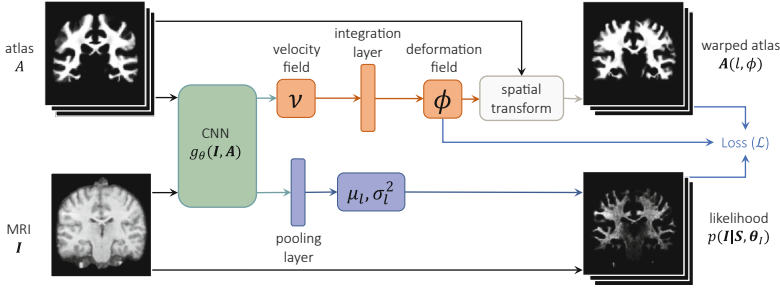


Fig. 1. Method overview. The network block $g_{\psi}(\cdot, \cdot)$ outputs a stationary velocity field \mathbf{v} , enabling alignment of the probabilistic atlas to the input volume, and likelihood Gaussian parameters $\boldsymbol{\mu}, \boldsymbol{\sigma}^2$, which yield likelihood maps for each label.

negative log posterior distribution of the image-specific parameters given the training images:

$$\begin{aligned}
 & - \sum_{n=1}^N \log p(\mathbf{v}^n, \boldsymbol{\mu}^n, [\boldsymbol{\sigma}^2]^n | \mathbf{I}^n; \mathbf{A}, \lambda) \tag{7} \\
 & = - \sum_{n=1}^N \sum_{j \in \Omega} \log \left[\sum_{l=1}^L \mathcal{N}(I_j^n; \mu_l^n, [\sigma_l^2]^n) A(l, \phi_{v_m}(\mathbf{x}_j)) \right] + \lambda \|\nabla \mathbf{u}^n\|^2 + \text{const}
 \end{aligned}$$

The network outputs different parameters \mathbf{v} , $\boldsymbol{\mu}$, and $\boldsymbol{\sigma}$ for each test image \mathbf{I} .

We design the CNN $g_{\theta_C}(\cdot, \cdot)$ based on a 3D UNet-style architecture [17] and the VoxelMorph implementation [3]. The network consists of downsampling convolutional layers with 32 filters, 3×3 kernels, stride of 2, and LeakyReLU activations, followed by mirror upsampling layers and skip-connections. An additional convolutional layer is used to output \mathbf{v} , a dense 3D velocity field defined over Ω ; and an additional pair of convolutional layers followed by global max pooling yield the Gaussian parameters $\boldsymbol{\mu}, \boldsymbol{\sigma}^2$. We compute $\phi_v = \exp(\mathbf{v})$ using a *scaling and squaring* integration layer [1, 4], enabling the computation of the loss regularization term. Combining the Gaussian parameters with the input image yields likelihood maps. These maps, together with a warped probabilistic atlas \mathbf{A} via a spatial transform layer, enable computation of the first term in Eq. (7) (Fig. 1).

Efficient Segmentation. Given a test scan, the trained network efficiently provides the image-specific parameter point estimates $\hat{\mathbf{v}}$ and $\hat{\boldsymbol{\theta}}_I$ in a single forward pass. The optimal segmentation is efficiently computed for each voxel j :

$$\hat{S}_j = \arg \max_l \mathcal{N}(I_j; \hat{\mu}_l, \hat{\sigma}_l^2) A(l, \phi_{\hat{\mathbf{v}}}(\mathbf{x}_j)). \tag{8}$$

3 Experiments and Results

Data. We evaluate our approach on three different image sets. The first (“multi-site”) includes 8,332 T1-weighted scans from several public datasets.¹ We randomly selected 7,332 scans to train and validate, and the remaining 1,000 were held out for testing. Manual delineations are not available for these scans, but we used automated segmentations produced by FreeSurfer [8] as a silver standard. The second dataset (“T1”) consist of 38 T1-weighted scans, used only for testing, each with 36 manually delineated brain structures [8]. The third dataset (“PD”) consists of eight proton density-weighted (PD) scans, manually segmented with the same protocol [9]. All scans were preprocessed with FreeSurfer, including skull stripping, bias field correction, intensity normalization, affine registration to Talairach space, and resampling to 1 mm^3 isotropic resolution [7].

Experimental Setup. We perform three experiments, one for each dataset. In the first, we fit our network to the T1-weighted training scans of the multi-site dataset, and use the resulting model to segment the 1,000 test scans. Despite the lack of manual gold standard, this experiment enables assessment of performance on a large, heterogeneous dataset. In a second experiment, we use the model already trained in the first to segment test scans from the separate T1 dataset. This enables evaluation with manual ground truth on scans from a scanner and pulse sequence not observed during training. In the third experiment, we train a network on the PD dataset, and then use it to segment the 8 PD scans. This is a different scenario than the first two experiments, since we learn to segment the test dataset directly. This experiment enables us to assess the ability of our algorithm to segment a substantially different MRI contrast, and fit datasets of reduced size. In all experiments, we use our method with the publicly available atlas from [16]. We emphasize that all networks are trained in an *unsupervised* fashion, and segmentation maps are only used for evaluation.

Baseline. We compare our method to a reimplementaion of [19], which solves Eq. (7) with no deformation (i.e., $\mathbf{u} = \mathbf{0}$, $\phi_v = Id$) via the EM algorithm. Since the model does not include deformation, using the nonrigid version of the atlas would yield low performance, and instead it relies on an affine version of the aforementioned atlas and Gaussian likelihood functions.

Evaluation. We used Dice scores for a subset of structures of interest (Fig. 2), and also focus on deep structures such as the hippocampus, which is the target of many neuroimaging studies due to its significance in dementia.

Implementation. We implement our method using Keras with a Tensorflow backend and the ADAM optimizer. We predict the velocity field \mathbf{v} and resulting deformation field ϕ at every second voxel in each dimension, due to memory constraints. We obtain a final dense deformation field via linear interpolation.

¹ Obtained online via: OASIS (oasis-brains.org), MCIC (central.xnat.org/data/projects/MCIC), PPMI (ppmi-info.org), HABS (nmr.mgh.harvard.edu/lab/harvardagin_gbrain), GSP (neuroinfo.org/gsp), ABIDE (fcon.1000.projects.nitrc.org/indi/abide/), and ADHD200 (fcon.1000.projects.nitrc.org/indi).

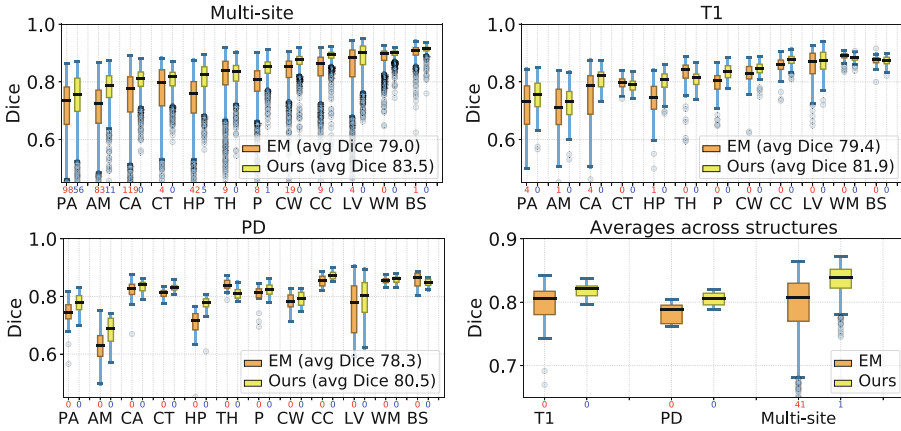


Fig. 2. Segmentation Statistics. Dice scores for: cerebral cortex (CT) and white matter (WM); lateral ventricle (LV); cerebellar cortex (CC) and white matter (CW); thalamus (TH); caudate (CA); putamen (P); pallidum (PA); brainstem (BS); hippocampus (HP); and amygdala (AM). Scores of contralateral structures are averaged. The number of outliers under the x axis is shown in red (baseline) and blue (ours). (Color figure online)

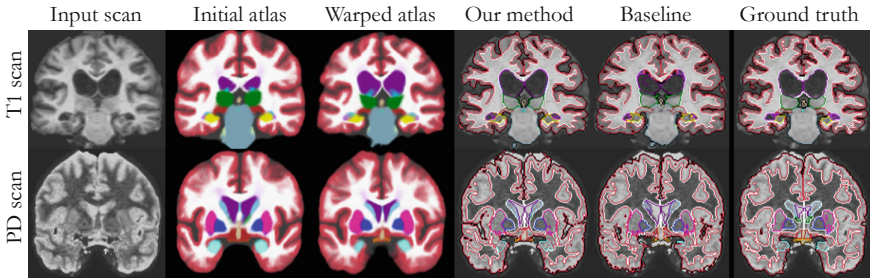


Fig. 3. Example Results. Coronal slices of two scans (one from each of the T1 and PD datasets), along with the initial and deformed probabilistic atlas, and corresponding segmentations. In the atlas, the color of each pixel is a combination of the colors of different labels, weighted by their probabilities. In the segmentations, we show the contour of the labels in the corresponding colors. We use the FreeSurfer color map [7]. (Color figure online)

For all experiments, we set $\lambda = 10$, the only free parameter, having visually evaluated segmentation results for several validation subjects (held out from training). We also group anatomical labels with similar intensity properties into eleven merged labels, forcing groups of original labels to share Gaussian parameters, increasing robustness [16]. Specifically, we group: contralateral structures (in general), gray matter structures (cerebral gray matter, hippocampus, amygdala, caudate, accumbens), and cerebrospinal fluid structures.

Results. Our method requires only 15 s per scan on an NVIDIA Titan Xp GPU. Figure 2 reports segmentation statistics for all experiments. Our method achieves considerably higher Dice scores than the baseline on the multi-site dataset (average over all structures 83.5% *vs.* 79.0%), especially in deep brain structures such as the hippocampi (81.1% *vs.* 73.1%). Moreover, it largely reduces the number of outliers with very poor segmentation (e.g., there are over 100 cases with Dice < 50% in the caudate for the baseline approach, and none for our method). In the T1 dataset, the test intensity distribution is slightly different than that of the training dataset. However, our approach successfully generalizes and outperforms the baseline (average 81.9% *vs.* 79.4%, hippocampi 79.9% *vs.* 73.5%). The results of the third experiment illustrate the ability of our method to adapt to contrasts other than T1, even when the data are limited, and outperform the baseline (average 80.5% *vs.* 78.3%, hippocampi 76.6% *vs.* 69.8%).

Figure 3 shows two segmentations from the T1 and PD datasets. In the T1 scan, the atlas successfully deforms to match the large ventricles of the subject, producing more accurate segmentations than the baseline – not only for the ventricles (purple), but also for surrounding structures, e.g., thalami (green). In the PD scan, our method manages to segment all structures including the amygdalae (light blue), which are missed by the baseline.

4 Conclusion

We propose a principled approach for unsupervised segmentation, which enables training a CNN for a dataset without the need for *any* manually annotated images. The likelihood model may be extended to incorporate more complex functions (such as mixtures of Gaussians) and artifacts such as partial voluming and bias fields. In addition to segmentations, the method produces a dense non-linear deformation field that is a useful output by itself, e.g., for tensor-based morphometry. Using a large dataset, we demonstrate that the proposed approach achieves state-of-the-art accuracy for *unsupervised* brain MRI segmentation in different MRI contrasts. Our method runs in under 15 s on a GPU, facilitating deployment on large studies and in time-sensitive applications.

Acknowledgements. Supported by ERC (Starting Grant 677697); NIH (U01-MH117023, R21-AG050122, P41-EB015902, 5U01-MH093765, R01-LM012719, R01AG053949); NIBIB (P41-EB015896, 1R01-EB023281, R01-EB006758, R21-EB018907, R01-EB019956), NIDDK (1R21-DK-108277-01), NINDS (R01-NS05-25851, R21-NS072652, R01-NS070963, R01-NS083534, 5U01-NS086625, 5U24-NS10059103); NSF CAREER 1748377, NIA (5R01-AG008122, R01-AG016495), SIG 1S10RR023401, 1S10RR019307, 1S10RR023043; Wistron; SIP; and AWS. BF has financial interest in CorticoMetrics LLC, managed by MGH.

References

1. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 924–931. Springer, Heidelberg (2006). https://doi.org/10.1007/11866565_113
2. Ashburner, J., Friston, K.: Unified segmentation. *Neuroimage* **26**, 839–851 (2005)
3. Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A.V.: Voxelmorph: A learning framework for deformable medical image registration, *IEEE Trans. Med. Imaging* (accepted)
4. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.R.: Unsupervised learning for fast probabilistic diffeomorphic registration. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11070, pp. 729–738. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00928-1_82
5. Dalca, A.V., Guttag, J., Sabuncu, M.R.: Anatomical priors in convolutional networks for unsupervised biomedical segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 9290–9299 (2018)
6. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. *J. R. Stat. Soc. Ser. B* **39**(1), 1–22 (1977)
7. Fischl, B.: Freesurfer. *Neuroimage* **62**(2), 774–781 (2012)
8. Fischl, B., Salat, D.H., Busa, E., Albert, M., Dieterich, M., et al.: Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**(3), 341–355 (2002)
9. Fischl, B., et al.: Sequence-independent segmentation of magnetic resonance images. *Neuroimage* **23**, S69–S84 (2004)
10. Jog, A., Fischl, B.: Pulse sequence resilient fast brain segmentation. In: Frangi, A.F., Schnabel, J.A., Davatzikos, C., Alberola-López, C., Fichtinger, G. (eds.) MICCAI 2018. LNCS, vol. 11072, pp. 654–662. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-00931-1_75
11. Kamnitsas, K., et al.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Med. Image Anal.* **36**, 61–78 (2017)
12. Lee, M.C.H., Petersen, K., Pawlowski, N., Glocker, B., Schaap, M.: TETRIS: template transformer networks for image segmentation with shape priors. *IEEE Trans. Med. Imaging* (Accepted)
13. Oktay, O., et al.: Anatomically constrained neural networks (ACNNs): application to cardiac image enhancement and segmentation. *IEEE Trans. Med. Imaging* **37**(2), 384–395 (2017)
14. Pan, S.J., Yang, Q.: A survey on transfer learning. *IEEE Trans. Knowl. Data Eng.* **22**(10), 1345–1359 (2009)
15. Patenaude, B., Smith, S., Kennedy, D., Jenkinson, M.: A bayesian model of shape and appearance for subcortical brain segmentation. *Neuroimage* **56**, 907–922 (2011)
16. Puonti, O., Iglesias, J., Van Leemput, K.: Fast sequence-adaptive whole-brain segmentation using parametric bayesian modeling. *NeuroImage* **143**, 235–249 (2016)
17. Ronneberger, O., Fischer, P., Brox, T.: U-net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
18. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: a survey. *IEEE Trans. Med. Imaging* **32**(7), 1153–1190 (2013)

19. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. *IEEE Trans. Med. Imaging* **18**(10), 897–908 (1999)
20. de Vos, B.D., Berendsen, F.F., Viergever, M.A., Sokooti, H., Staring, M., Išgum, I.: A deep learning framework for unsupervised affine and deformable image registration. *Med. Image Anal.* **52**, 128–143 (2019)
21. Wells, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. *IEEE Trans. Med. Imaging* **15**(4), 429–442 (1996)
22. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm. *IEEE Trans. Med. Imaging* **20**(1), 45–57 (2001)