Unsupervised deep learning for Bayesian brain MRI segmentation

Adrian V. Dalca[1,2], Evan Yu[3], Polina Golland[2], Bruce Fischl[1],
Mert R. Sabuncu[3,4], and Juan Eugenio Iglesias[1,2,5]

[1] Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Harvard Medical School
[2] Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology
[3] Meinig School of Biomedical Engineering, Cornell University
[4] School of Electrical and Computer Engineering, Cornell University
[5] Centre for Medical Image Computing (CMIC), University College London

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 to implement segmentation tools that are computationally efficient at test time. However, most of these strategies rely on learning from manually annotated images. These supervised deep learning methods are therefore sensitive to the intensity profiles in the training dataset. To develop a deep learning-based segmentation model for a new image dataset (e.g., of different contrast), one usually needs to create a new labeled training dataset, which can be prohibitively expensive, or rely on suboptimal *ad hoc* adaptation or augmentation approaches. In this paper, we propose an alternative strategy that combines a conventional probabilistic atlas-based segmentation with deep learning, enabling one to train 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 seconds 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 scans, is a well-studied problem. Most probabilistic models for image segmentation exploit atlas priors, and account for variations in contrast and imaging artifacts such as MR inhomogeneity [30,32]. Most of the popular neuroimage processing pipelines rely on segmentation algorithms based on these ideas, including FreeSurfer [13], SPM [3], and FSL [26]. While these tools achieve high robustness to changes in MRI contrast of the input scan, a significant drawback is that they are computationally demanding (e.g., 23 minutes using a multithreaded setup, in a recent study [27]), 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 network (CNN) architectures that excel at learning contextually important multiscale 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, such as those involving very large test datasets [19,28]. However, DL based techniques are notoriously sensitive to changes in the image intensity data distribution. For example, an upgrade to the MRI scanner or a change in the pulse sequence might alter contrast properties that can dramatically reduce the performance of a CNN-based segmentation model [17]. This issue can be alleviated via domain adaptation, which usually requires some amount of labeled training data for the new conditions, or data augmentation, which requires the user to simulate expected variations. However, even with additional data, these methods only partially close the gap with the fully supervised setting [25]. 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 different 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. Alternatively, it could have been derived from an anatomical template, after applying spatial blurring to account for variability in location.

Several recent methods tackle segmentation tasks in the presence of small training datasets. Most assume at least one manually segmented image from the same modality as the main task, and leverage data augmentations techniques and exploit priors to enable the use of supervised methods [5,34]. Other methods require no labelled examples from the target modality, but leverage a large collection of segmentation maps from other datasets [9,18].

The main contribution of this paper is the integration of mathematical ideas from the Bayesian segmentation literature with an unsupervised deep learning framework. Specifically, we assume a probabilistic model, which requires estimation of scan-specific parameters comprising an atlas deformation and image intensity statistics. The estimation of the atlas warp has traditionally relied on classic deformable registration algorithms [29], which are based on iterative, numerical optimization, and are therefore computationally expensive. Instead, we leverage recent advances in learning-based registration [4,8,21,31] to efficiently estimate the warp jointly with the intensity parameters. We use a novel loss function, which is derived from the probabilistic model with Bayesian inference, and is thus principled and interpretable. Integrating DL with Bayesian segmentation, we attain two highly desirable features. First, given a probabilistic atlas, the method is unsupervised, and hence contrast adaptive: given a new dataset with a previously unobserved MRI contrast (e.g., a change of pulse sequence in MRI acquisition), we train our network without the need to label any MRI scans. Second, the segmentation is efficient and runs in approximately 15 seconds on a GPU.

2 Method

2.1 Segmentation as Bayesian inference

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

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

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

The likelihood parameters may be global for a training dataset, or estimated specifically for each test scan. Here we are interested in a subset of Bayesian segmentation models that follow the latter approach [3,27,30,32,33], which enables these models to *adapt* to the intensity characteristics of the input scans, making them robust to changes in MRI contrast. Expanding Eq. (1) to include model parameters, which we treat as random variables, yields:

$$\hat{\boldsymbol{S}} = \arg\max_{\boldsymbol{S}} \int_{\boldsymbol{\theta}_S} \int_{\boldsymbol{\theta}_I} p(\boldsymbol{S}|\boldsymbol{\theta}_S, \boldsymbol{\theta}_I, \boldsymbol{I}) p(\boldsymbol{\theta}_S, \boldsymbol{\theta}_I|\boldsymbol{I}) d\boldsymbol{\theta}_{\boldsymbol{S}} d\boldsymbol{\theta}_{\boldsymbol{I}}, \tag{2}$$

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

$$\begin{aligned}\{\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_I\} &= \arg\max_{\{\boldsymbol{\theta}_S, \boldsymbol{\theta}_I\}} p(\boldsymbol{\theta}_S, \boldsymbol{\theta}_I | \boldsymbol{I}) \\ &= \arg\max_{\{\boldsymbol{\theta}_S, \boldsymbol{\theta}_I\}} p(\boldsymbol{\theta}_S) p(\boldsymbol{\theta}_I) \sum_{\boldsymbol{S}} p(\boldsymbol{I}|\boldsymbol{S}, \boldsymbol{\theta}_I) p(\boldsymbol{S}|\boldsymbol{\theta}_S), \end{aligned} \tag{3}$$

where we have assumed independence between the parameters of the prior and likelihood. The computation often requires estimating an atlas deformation in $\boldsymbol{\theta}_S$ and intensity parameters in $\boldsymbol{\theta}_I$, and is typically achieved with a combination of numerical optimization and the Expectation Maximization (EM) algorithm [10]. Given point estimates, the final segmentation is computed efficiently as:

$$\hat{\boldsymbol{S}} = \arg\max_{\boldsymbol{S}} p(\boldsymbol{S}|\hat{\boldsymbol{\theta}}_S, \hat{\boldsymbol{\theta}}_I, \boldsymbol{I}) = \arg\max_{\boldsymbol{S}} p(\boldsymbol{I}|\boldsymbol{S}, \hat{\boldsymbol{\theta}}_I) p(\boldsymbol{S}|\hat{\boldsymbol{\theta}}_S), \tag{4}$$

and is often produced directly by the same EM algorithm.

2.2 Model

Our model instantiation builds on existing work [3,27,30]. The prior is defined by a given probabilistic atlas $\boldsymbol{A}$, such that $A(l, \boldsymbol{x})$ provides the probability of observing each neuroanatomical label $l = 1, \dots, L$ at each location $\boldsymbol{x} \in \Omega$. The atlas is deformed by a diffeomorphic transform $\boldsymbol{\phi}$, parameterized by a stationary velocity field $\boldsymbol{v}$, (i.e., $\boldsymbol{\phi}_v = \exp[\boldsymbol{v}]$, see [2]) that parametrizes the prior such that $\boldsymbol{\theta}_S = \boldsymbol{v}$. Assuming independence over voxels:

$$p(\boldsymbol{S}|\boldsymbol{\theta}_S; \boldsymbol{A}) = p(\boldsymbol{S}|\boldsymbol{v}; \boldsymbol{A}) = \prod_{j \in \Omega} A\Big(S_j, \boldsymbol{\phi}_v(\boldsymbol{x}_j)\Big), \tag{5}$$

where S_j is the segmentation at voxel j, and $\boldsymbol{x}_j$ is its spatial location.

We discourage strong deformations by penalizing the spatial gradient $\nabla \boldsymbol{u}_v$ of displacement $\boldsymbol{u}_v$, where $\boldsymbol{\phi}_v = Id + \boldsymbol{u}_v$:

$$p(\boldsymbol{\theta}_S; \lambda) = p(\boldsymbol{v}; \lambda) \propto \exp[-\lambda \|\nabla \boldsymbol{u}_v\|^2]. \tag{6}$$

The hyperparameter λ controls the weight for the atlas deformation penalty.

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

$$p(\boldsymbol{I}|\boldsymbol{S}, \boldsymbol{\theta}_I) = p(\boldsymbol{I}|\boldsymbol{S}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \prod_{j \in \Omega} \mathcal{N}(I_j; \mu_{S_j}, \sigma^2_{S_j}), \tag{7}$$

where $\mathcal{N}(\cdot; \mu, \sigma^2)$ is the Gaussian distribution, I_j is the image intensity at voxel j, and the likelihood parameters $\boldsymbol{\theta}_I = \{\boldsymbol{\mu}, \boldsymbol{\sigma}^2\}$ 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: $p(\boldsymbol{\theta}_I) \propto 1$. The model can be easily extended to the multispectral case (i.e., inputs with multiple MRI contrasts) by replacing means and variances by mean vectors and covariance matrices, respectively.

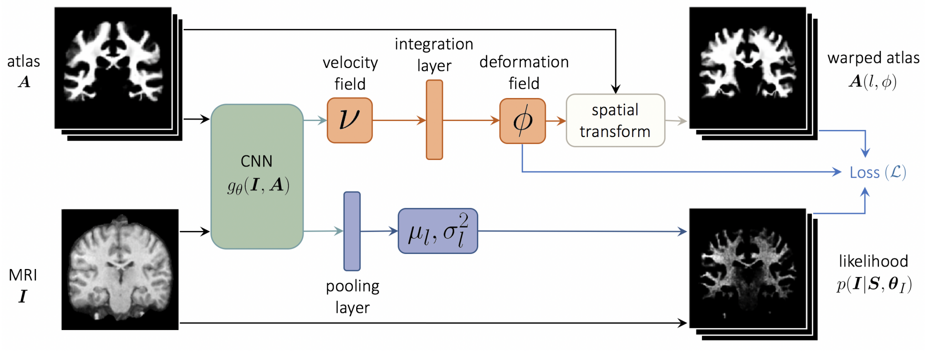

Fig. 1. Method overview. The network block $g_\psi(\cdot,\cdot)$ outputs a stationary velocity field $\boldsymbol{v}$, enabling alignment of the probabilistic atlas to the input volume, and likelihood Gaussian parameters $\boldsymbol{\mu},\boldsymbol{\sigma}^2$, which yield likelihood maps for each label.

2.3 Learning

To avoid computationally expensive optimization typically required for maximum *a posteriori* (MAP) estimation in Eq. (3), we propose to train a CNN to estimate these parameters directly from an input scan. Specifically, we design a CNN $g_{\boldsymbol{\theta}_C}(\boldsymbol{I},\boldsymbol{A}) = (\boldsymbol{\theta}_S,\boldsymbol{\theta}_I) = (\boldsymbol{v},\boldsymbol{\mu},\boldsymbol{\sigma}^2)$ with convolutional parameters $\boldsymbol{\theta}_C$ that takes as input a scan $\boldsymbol{I}$ and the probabilistic atlas $\boldsymbol{A}$, and outputs the model parameters $\boldsymbol{v},\boldsymbol{\mu},\boldsymbol{\sigma}^2$ for that scan. To learn the neural network, we use a pool of N *unlabeled* scans $\{I^n\}_{n=1}^N$ to minimize the negative log posterior distribution of the image-specific parameters given the training images:

$$
\begin{aligned}
&-\sum_{n=1}^{N}\log p(\boldsymbol{v}^n,\boldsymbol{\mu}^n,[\boldsymbol{\sigma}^2]^n|\boldsymbol{I}^n;\boldsymbol{A},\lambda) \qquad (8)\\
&=-\sum_{n=1}^{N}\sum_{j\in\Omega}\log\left[\sum_{l=1}^{L}\mathcal{N}(I_j^n;\mu_l^n,[\sigma_l^2]^n)A\Big(l,\boldsymbol{\phi}_{v_m}(\boldsymbol{x}_j)\Big)\right]+\lambda\|\nabla\boldsymbol{u}_v^n\|^2-K(\lambda),
\end{aligned}
$$

where $K(\lambda)$ is a log-partition function that depends on the hyperparameter λ, and which does not affect the optimization. We emphasize that the network outputs different parameters $\boldsymbol{\mu},\boldsymbol{\sigma}$, and $\boldsymbol{v}$ for each test image $\boldsymbol{I}$.

We design the neural network $g_{\boldsymbol{\theta}_C}(\cdot,\cdot)$ based on a 3D UNet-style architecture [28] and the public VoxelMorph implementation [4]. The network consists of downsampling convolutional layers with 32 filters with 3x3 kernels, stride of 2, and LeakyReLu activations, followed by mirror upsampling layers and skip-connections. From this point, an additional convolutional layer is used to output $\boldsymbol{v}$, a dense 3D velocity field defined over Ω; and an additional pair of convolutional layers followed by a global max pooling operation to output the Gaussian parameters $\boldsymbol{\mu},\boldsymbol{\sigma}^2$. We compute $\boldsymbol{\phi} = \exp(\boldsymbol{v})$ using a network integration layer that implements *scaling and squaring* [2,8,21], enabling the computation of the loss regularization term. We warp the probabilistic atlas $\boldsymbol{A}$ with a spatial transform layer. Combining the Gaussian parameters with the input image yields likelihood maps, which together with the warped atlas enable computation of the first term of the loss function.

2.4 Efficient segmentation

Given a trained network and a new test subject, the network efficiently provides the image-specific parameter point estimates $\hat{\boldsymbol{v}}$, and $\hat{\boldsymbol{\theta}}_I$ via a single forward pass. The optimal segmentation can be efficiently computed for each voxel:

$$\hat{S}_j = \arg\max_l \mathcal{N}(I_j; \hat{\mu}_l, \hat{\sigma}_l^2) A\Big(l, \phi_{\hat{v}}(\boldsymbol{x}_j)\Big). \tag{9}$$

Both terms in Eq. (9) are computed inside our GPU implementation (Fig. 1).

3 Experiments and results

3.1 Data

We evaluate our approach on three different image sets. The first dataset ("multi-site") includes 8,332 T1-weighted scans from several public datasets: OASIS [22], ABIDE [11], ADHD200 [24], MCIC [15], PPMI [23], HABS [7], and Harvard GSP [16]. 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 [13] as a silver standard, for evaluation only. The second dataset ("T1") consist of 38 T1-weighted scans, used only for testing, each with 36 manually delineated brain structures [13]. The third dataset ("PD") consists of eight proton density-weighted (PD) scans, manually segmented with the same protocol [14]. 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 [12].

3.2 Experimental setup

We perform three experiments, one for each dataset. In the first experiment, we fit our network to the 7,332 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 a 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 experiment (i.e., on the 7,332 T1 scans) to segment scans from the separate T1 dataset. This experiment enables evaluation with manual ground truth on scans from a scanner and pulse sequence that were not observed by the neural network during training. In the third experiment, we train a network on the PD dataset, and then use it to segment those 8 PD

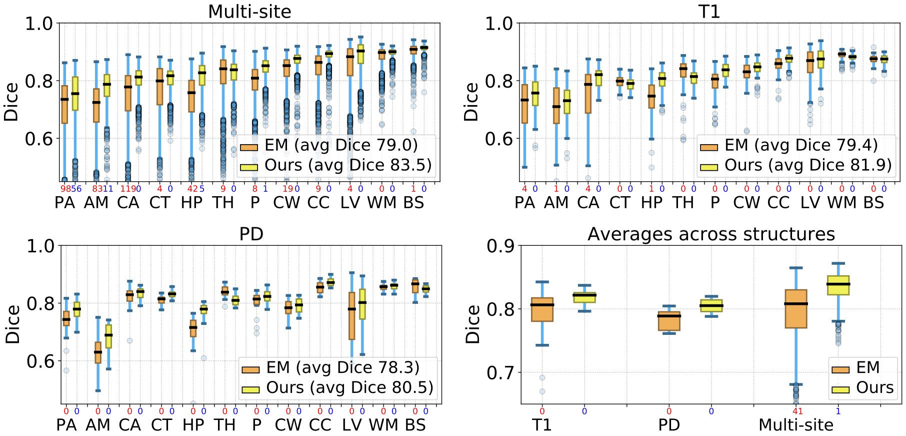

Fig. 2. Segmentation Statistics. Dice scores for: cerebral cortex (CT) and white matter (WM); lateral ventricle (LV); cerebellar cortex (CC) and white matter (CW); thalamus (TH); caudate (CA); putamen (P); pallidum (PA); brainstem (BS); hippocampus (HP); and amygdala (AM). Scores of contralateral structures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

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 to fit datasets of reduced size. In all experiments, we use our method with the publicly available atlas from [27]. We emphasize that all networks are trained in an unsupervised fashion, and segmentation maps are only used for evaluation.

3.3 Baseline

We compare our method to a reimplementation of [30], which relies on an affine version of the aforementioned atlas and Gaussian likelihood functions. Specifically, the baseline method solves Eq. (8), but with no deformation (i.e., $\boldsymbol{v} = \boldsymbol{u} = \mathbf{0}$, and $\phi_v = Id$), and the model parameters are estimated with the EM algorithm. Since the model does not include deformation, using the nonrigid version of the atlas would yield very low performance.

3.4 Implementation Details

We group anatomical labels with similar intensity properties into eleven merged labels to force groups of original labels to share Gaussian parameters, increasing robustness [27]. Specifically, we group: contralateral structures (in general), gray matter structures (cerebral gray matter, hippocampus, amygdala, caudate, accumbens), and cerebrospinal fluid structures. For evaluation, we used Dice scores for a subset of structures of interest (Fig. 2). We quantify the results on these

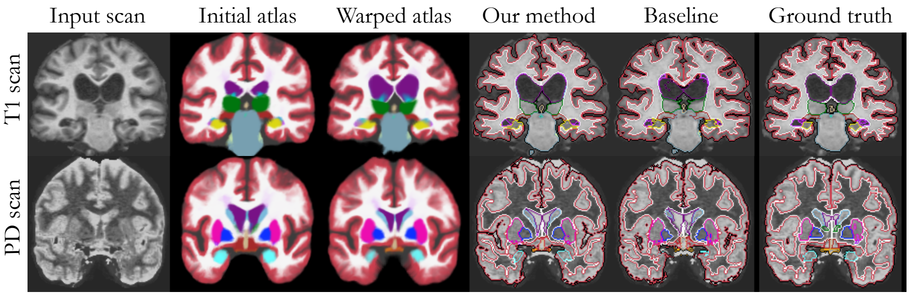

Fig. 3. Example Results. Coronal slices of two scans (one from each of the T1 and PD datasets), along with the initial and deformed probabilistic atlas, and corresponding segmentations. In the atlas, the color of each pixel is a combination of the colors of different labels, weighted by their probabilities. In the segmentations, we show the contour of the labels in the corresponding colors. We use the FreeSurfer color map [12].

structures, and also focus on deep structures such as the hippocampus, which is the target of many neuroimaging studies due to its significance in dementia.

We implement our method using Keras [6] with a Tensorflow [1] backend and the ADAM optimizer [20]. We predict the velocity field $\boldsymbol{v}$ and resulting deformation field $\boldsymbol{\phi}$ at every second voxel in each dimension, due to memory constraints. We linearly interpolate to obtain a final dense deformation field. To set λ, the only free parameter of our framework, we visually evaluated segmentation results for several validation subjects (held out from the training dataset), and set $\lambda = 10$ in all experiments.

3.5 Results

Our method requires only 15 seconds per scan on an NVIDIA Titan Xp GPU. Fig. 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%), particularly 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 lower than 50% in the caudate for the baseline approach, and none for our method). In the T1 dataset, the test intensity distribution is slightly different 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., the thalami (green).

In the PD scan, despite the small dataset, 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 field. In addition to segmentations, the method produces a dense nonlinear 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 seconds on a GPU, facilitating deployment on large studies and in time-sensitive applications.

Acknowledgement

This work was supported by the European Research Council (Starting Grant 677697, project "BUNGEE-TOOLS", awarded to JEI); the BRAIN Initiative Cell Census Network (U01-MH117023); NIH (R21-AG050122, P41-EB015902, 5U01-MH093765, R01LM012719, R01AG053949, and 1R21AG050122); NIBIB (P41-EB015896, 1R01-EB023281, R01-EB006758, R21-EB018907, R01-EB019956), NIA (5R01-AG008122, R01-AG016495), NIDDK (1R21-DK-108277-01), NINDS (R01-NS0525851, R21-NS072652, R01-NS070963, R01-NS083534, 5U01-NS086625, 5U24-NS10059103); NSF (NeuroNex Grant grant 1707312); Wistron Corp.; SIP; and AWS; and was made possible by the resources provided by Shared Instrumentation Grants 1S10RR023401, 1S10RR019307, and 1S10RR023043. In addition, BF has a financial interest in CorticoMetrics, a company whose medical pursuits focus on brain imaging and measurement technologies. BF's interests were reviewed and are managed by Massachusetts General Hospital and Partners HealthCare in accordance with their conflict of interest policies.

References

1. Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., et al.: Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} Symposium on Operating Systems Design and Implementation ({OSDI} 16). pp. 265–283 (2016)
2. Arsigny, V., Commowick, O., Pennec, X., Ayache, N.: A log-euclidean framework for statistics on diffeomorphisms. In: MICCAI. pp. 924–931. Springer (2006)
3. Ashburner, J., Friston, K.: Unified segmentation. Neuroimage **26**, 839–851 (2005)
4. Balakrishnan, G., Zhao, A., Sabuncu, M., Guttag, J., Dalca, A.V.: Voxelmorph: A learning framework for deformable medical image registration. IEEE:TMI (2019)

5. Chaitanya, K., Karani, N., Baumgartner, C., Konukoglu, E.: Semi-supervised and task-driven data augmentation. arXiv preprint arXiv:1902.05396 (2019)
6. Chollet, F.: Keras. `https://github.com/fchollet/keras` (2015)
7. Dagley, A., LaPoint, M., Huijbers, W., Hedden, T., McLaren, D.G., Chatwal, J.P., Papp, K.V., Amariglio, R.E., Blacker, D., Rentz, D.M., et al.: Harvard aging brain study: dataset and accessibility. NeuroImage (2015)
8. Dalca, A.V., Balakrishnan, G., Guttag, J., Sabuncu, M.: Unsupervised learning for fast probabilistic diffeomorphic registration. MICCAI **11070**, 729–738. (2018)
9. 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)
10. Dempster, A.P., Laird, N.M., Rubin, D.B.: Maximum likelihood from incomplete data via the EM algorithm. J. Royal Statistical Society: Series B **39**(1), 1–22 (1977)
11. Di Martino, A., et al.: The autism brain imaging data exchange: towards a large-scale evaluation of the intrinsic brain architecture in autism. Molecular psychiatry **19**(6), 659–667 (2014)
12. Fischl, B.: Freesurfer. Neuroimage **62**(2), 774–781 (2012)
13. 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)
14. Fischl, B., Salat, D.H., Van Der Kouwe, A.J., Makris, N., Ségonne, F., Quinn, B.T., Dale, A.M.: Sequence-independent segmentation of magnetic resonance images. Neuroimage **23**, S69–S84 (2004)
15. Gollub, R.L., Shoemaker, J.M., King, M.D., White, T., Ehrlich, S., Sponheim, S.R., Clark, V.P., Turner, J.A., Mueller, B.A., Magnotta, V., et al.: The MCIC collection: a shared repository of multi-modal, multi-site brain image data from a clinical investigation of schizophrenia. Neuroinformatics **11**(3), 367–388 (2013)
16. Holmes, A.J., Hollinshead, M.O., OKeefe, T.M., Petrov, V.I., Fariello, G.R., Wald, L.L., Fischl, B., Rosen, B.R., Mair, R.W., Roffman, J.L., et al.: Brain genomics superstruct project initial data release with structural, functional, and behavioral measures. Scientific data **2** (2015)
17. Jog, A., Fischl, B.: Pulse sequence resilient fast brain segmentation. In: MICCAI. pp. 654–662. Springer (2018)
18. Joyce, T., Chartsias, A., Tsaftaris, S.A.: Deep multi-class segmentation without ground-truth labels. MIDL (2018)
19. Kamnitsas, K., Ledig, C., Newcombe, V., Simpson, J., Kane, A.D., Menon, D., Rueckert, D., Glocker, B.: Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. Medical image analysis **36**, 61–78 (2017)
20. Kingma, D.P., Ba, J.: ADAM: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
21. Krebs, J., e Delingette, H., Mailhé, B., Ayache, N., Mansi, T.: Learning a probabilistic model for diffeomorphic registration. IEEE transactions on medical imaging (2019)
22. Marcus, D.S., Wang, T.H., Parker, J., Csernansky, J.G., Morris, J.C., Buckner, R.L.: Open access series of imaging studies (oasis): cross-sectional mri data in young, middle aged, nondemented, and demented older adults. Journal of cognitive neuroscience **19**(9), 1498–1507 (2007)
23. Marek, K., Jennings, D., Lasch, S., Siderowf, A., Tanner, C., Simuni, T., Coffey, C., Kieburtz, K., Flagg, E., Chowdhury, S., et al.: The parkinson progression marker initiative (ppmi). Progress in neurobiology **95**(4), 629–635 (2011)

24. Milham, M.P., Fair, D., Mennes, M., Mostofsky, S.H., et al.: The ADHD-200 consortium: a model to advance the translational potential of neuroimaging in clinical neuroscience. Frontiers in systems neuroscience **6**, 62 (2012)
25. Pan, S.J., Yang, Q.: A survey on transfer learning. IEEE:TKDE **22**, 1345–59 (2010)
26. Patenaude, B., Smith, S., Kennedy, D., Jenkinson, M.: A bayesian model of shape and appearance for subcortical brain segmentation. Neuroimage **56**, 907–922 (2011)
27. Puonti, O., Iglesias, J., Van Leemput, K.: Fast sequence-adaptive whole-brain segmentation using parametric bayesian modeling. NeuroImage **143**, 235–249 (2016)
28. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: MICCAI. pp. 234–241. Springer (2015)
29. Sotiras, A., Davatzikos, C., Paragios, N.: Deformable medical image registration: A survey. IEEE:TMI **32**(7), 1153 (2013)
30. Van Leemput, K., Maes, F., Vandermeulen, D., Suetens, P.: Automated model-based tissue classification of MR images of the brain. IEEE:TMI **18**, 897–908 (1999)
31. de Vos, B., Berendsen, F., Viergever, M., Staring, M., Išgum, I.: End-to-end unsupervised deformable image registration with a CNN. DLMIA, pp. 204–212 (2017)
32. Wells, W.M., Grimson, W.E.L., Kikinis, R., Jolesz, F.A.: Adaptive segmentation of MRI data. IEEE:TMI **15**(4), 429–442 (1996)
33. Zhang, Y., Brady, M., Smith, S.: Segmentation of brain MR images through a hidden markov random field model and the expectation-maximization algorithm. IEEE:TMI **20**(1), 45–57 (2001)
34. Zhao, A., Balakrishnan, G., Durand, F., Guttag, J.V., Dalca, A.V.: Data augmentation using learned transforms for one-shot medical image segmentation. arXiv preprint arXiv:1902.09383 (2019)