TOWARDS A DEEP LEARNING APPROACH TO BRAIN PARCELLATION

Noah Lee[⋆†] *Andrew F. Laine*[⋆] *Arno Klein*[†]

[⋆] Columbia University, Department of Biomedical Engineering,
351 Engineering Terrace MC-8904, 1210 Amsterdam Avenue, New York, NY 10027
[†]Columbia University, New York State Psychiatric Institute,
1051 Riverside Drive, Unit 42, New York, NY 10032

ABSTRACT

Establishing correspondences across structural and functional brain images via labeling, or parcellation, is an important and challenging task for clinical neuroscience and cognitive psychology. A limitation with existing approaches is that they i) possess shallow architectures, ii) are based on heuristic manual feature engineering, and iii) assume the validity of the designed feature model. In contrast, we advocate a deep learning approach to automate brain parcellation. We present a novel application of convolutional networks to build discriminative features for brain parcellation, which are automatically learned from labels provided by human experts. Initial validation experiments show promising results for automatic brain parcellation, suggesting that the proposed approach has potential to be an alternative to template or atlas-based parcellation approaches.

Index Terms— Brain Parcellation, Deep Learning, Convolutional Networks, Feature Learning.

1. INTRODUCTION

Delineation of structural and functional regions ("parcellation") of the human brain is an important and challenging task for clinical neuroscience and cognitive psychology. Accurate and precise parcellation enables quantification of normal and abnormal changes in the brain as well as analysis of relationships between brain function and structural appearance. Such information is crucial for clinical diagnosis and prediction of treatment outcome in neurodegenerative and pychiatric disorders. However, there still does not exist a widely accepted standard (protocol) for brain image parcellation [1]. The choice of parcellation units is usually dictated by software packages that make use of a labeled atlas brain image, in which a parcellation protocol has been applied to a single individual. Only recently have large-scale efforts come about to establish and manually apply a standard brain parcellation protocol to many brain images (http://www.braincolor.org/protocols). However, because manual parcellation is a tedious, time-consuming, and inconsistent endeavor that requires expertise, many researchers rely on automatic brain parcellation methods. The challenge for both humans and computers is the intrinsic variability of the human brain, which makes it extremely difficult to define consistent correspondences across brains.

To establish correspondences, researchers ubiquitously co-register brain images to each other, commonly with a template or labeled atlas brain of the same imaging modality [2]. However, such registration methods typically assume image similarity as a surrogate for anatomical similarity, continuous mapping between corresponding features, and representativeness of the template or atlas. On a lower level a main drawback with existing automatic brain parcellation approaches is that they i) employ algorithms with shallow architectures, ii) are based on heuristic manual feature engineering, and iii) assume the validity of the underlying feature engineered model. In [3] the authors have demonstrated that shallow architectures are limited and non-optimal when learning complex high-dimensional functions. Examples of learning algorithms with shallow architectures are kernel machines or single-layer neural networks. In comparison to deep architectures shallow learning algorithms are limited in efficiently representing complex function families to learn high-level learning tasks. Many learning algorithms rely on human input to handcraft features, which requires a complete understanding of the problem domain. Such feature engineering approaches limit the generalizability of the model, which may lead to feature redesign and validation, a costly, error prone, and impractical process.

In contrast to existing methods, we would like to advocate a deep learning [3] approach to automate brain image parcellation. We are motivated by models from biologically inspired artificial intelligence, in particular cortical network models such as deep convolutional networks [4] (CNs). Our research is driven by two main questions. First, can we truly automate brain parcellation in a realistic clinical setting? Second, can cortical network models, which possess deep learning architectures, provide the computational intelligence for this challenging task? In this paper we report on a novel application of convolutional networks to build discriminative features for brain parcellation, which are automatically

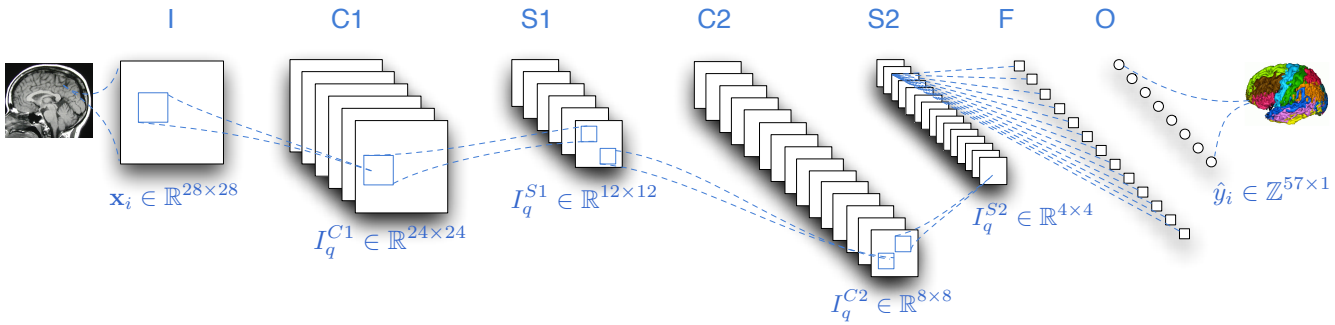

Fig. 1. **Deep learning approach to automate brain image parcellation using a convolutional network model.** *From left to right, the deep architecture consists of several layers starting with the input layer (I). In an alternating manner the CN consists of a hierarchical architecture of convolutional (C1, C2) and subsampling (S1, S2) layers followed by a full-connection layer (F), and finally the output layer (O).*

learned from labels provided by human experts. The idea we would like to pursue is a structured hierarchical approach using context-aware feature learning to perform parcellation without resorting to an atlas or a template-based registration approach. Moreover, our approach does not require the engineering design of handcrafted features, reducing human expert intervention and the need for prior knowledge. Initial validation experiments show promising results for automatic multi-class brain parcellation, suggesting that the proposed approach has potential as an alternative to existing template or atlas-based parcellation approaches.

2. METHODS

2.1. Problem formulation

Consider the problem of finding a function $f : X \rightarrow Y$ that maps an input space to an output space. Here X refers to the brain image data and Y to a delineated label space of a multi-class brain parcellation. We are given a dataset $\mathcal{D}$ as a collection of N images $\{\mathcal{I}\} = \{\mathcal{I}_1, \mathcal{I}_2, ..., \mathcal{I}_N\}$. The dataset is further partitioned into $\mathcal{D} = \{\mathcal{D}_l, \mathcal{D}_u\}$, where $\mathcal{D}_l = \{s_i, y_i\}_{i=1}^{l}$ denotes the labeled training set and $\mathcal{D}_u = \{s_i, \hat{y}_i\}_{i=l+1}^{n}$ the unlabeled test set. Each pair consists of an image site s_i (e.g., voxel) and a label y_i, which assumes values in a finite set $y = \{0, ..., C\}$. The index n refers to the number of sites within each image. For each site in the training set we form a d-dimensional patch $\mathbf{x}_i \in \mathbb{R}^d$. A detailed description of $\mathbf{x}_i$ can be found in section 2.3. The input-output pairs in $\mathcal{D}_l$ are drawn in an independent and identically distributed manner from some unknown probability distribution $\mathbb{P}(X, Y)$ defined jointly over X and Y. Our goal is, given $\mathcal{D}_l$, to predict $\hat{y}$ for the unlabeled test set $\mathcal{D}_u$ such that the learned approximation to f has low probability of error $\mathbb{P}(f(X) \neq Y)$.

2.2. The convolutional network architecture

Convolutional networks (CN) belong to the class of cortical network models and are an extension to the classical multi-layer perceptrons (MLPs) model. They consist of a multi-layer hierarchical architecture of feature maps as depicted in Fig. 1. The learned model $\Phi = \{\mathbf{w}, b\}$ includes convolutional operators $\mathbf{w}$ and bias term b, which in combination with a nonlinear activation function γ (e.g. sigmoid or hyperbolic tangent), form so-called "activity feature maps" $\mathbf{I}_q^k$, where k indexes a CN layer and q a particular feature map of layer k. The first step is to perform a forward propagation of an input patch through the CN architecture shown in Fig. 1. The feature maps in each convolutional layer (C1, C2) are computed through a recursive forward dynamic of the form

$$\mathbf{I}_q^k = \gamma(\mathbf{u}_q^k) \tag{2.1}$$

$$\mathbf{u}_q^k = b_q^k + (\sum_p \mathbf{w}_{q,p}^k \otimes \mathbf{I}_p^{k-1}), \tag{2.2}$$

where γ denotes a smooth differentiable nonlinearity to ensure differentiability, $\mathbf{u}_q^k$ a pre-activation image, $\mathbf{I}_p^{k-1}$ the feature image at layer $k-1$, $\mathbf{w}_{q,p}^k$ a directed convolution kernel from map p to q, and b_q^k a bias term. The layers (S1, S2) are simple subsampling layers to reduce the computational load of the model and to introduce scale invariance of the learned features. The full-connnection layer (F) is a hidden layer as in MLPs with a nonlinear activation function. The output nodes of layer (O) represent the class labels of the network. Forward propagated class labels are compared with ground truth labels and the errors are then back-propagated through the network to refine the CN model in an iterative fashion.

Given our deep CN model Φ, the forward dynamics in Equations (2.1, 2.2), and some training data $(\mathbf{x}_i, y_i)$, our algorithm automatically learned discriminative features of parcellation units by solving an optimization problem via recursive error back-propagation

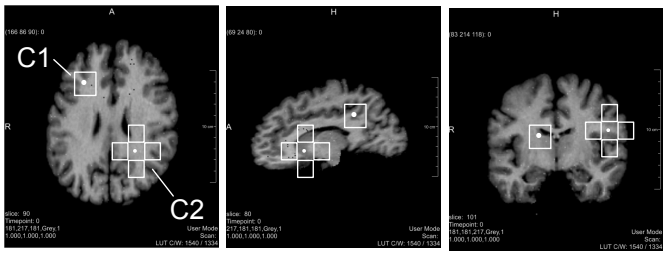

Fig. 2. **Context-aware feature configurations.** *Shown are orthogonal slices of subject one from the LBPA40 dataset.*

$$\arg\min_{\Phi} \mathcal{J}(\mathbf{x}_i, y_i, \Phi) \tag{2.3}$$

$$\Phi_{t+1} \leftarrow \Phi_t - \eta \nabla_{\Phi} \mathcal{J}(\mathbf{x}_i, y_i, \Phi). \tag{2.4}$$

To solve Equation (2.3) we employed an online learning strategy using stochastic gradient descent by minimizing the negative log-likelihood

$$\mathcal{J}(\mathbf{x}, y, \Phi) = \sum_i \log(\mathbb{P}(Y = y_i | \mathbf{x}_i, \Phi)). \tag{2.5}$$

2.3. Context-aware feature learning

We built two different kinds of context-aware feature configurations. For each image site s_i we considered a large neighborhood of surrounding image data, providing discriminative contextual information to determine the class label of the site. Fig. 2 shows examples of feature configuration C1 and C2 that were used to assess the performance of our approach. For testing, we performed linear patch sampling in order to learn a site-wise class probability. Invalid patch samples near the border were ignored. Parcellation labels were then obtained by choosing the class label with the highest probability given $\mathbf{x}_{C_1,C_2}$ and the learned CN model Φ

$$\hat{y}_{C_1,C_2} = \arg\max_i \mathbb{P}(Y = i | \mathbf{x}_{C_1,C_2}, \Phi). \tag{2.6}$$

C1 consisted of a single patch centered around s_i, whereas C2 consisted of a cross configuration of four patches (i.e., north, south, west, east) around s_i. Both were obtained through randomized sampling to build our training and validation set $(\mathbf{x}_i, y_i)$. To perform cortical and subcortical parcellation, we constrained the context area to a 28 x 28 dimensional patch. For C2, the four-element patches had dimensions 14 x 14, which were then concatenated to the final 28 x 28 patch dimension.

3. EXPERIMENTS AND RESULTS

We used 40 brain images and their labels (56 structures + background) from the LONI Probabilistic Brain Atlas (LPBA40) at the Laboratory of Neuro Imaging (LONI) at UCLA. We performed two sets of experiments on the LPBA40 dataset to assess the performance of our approach using feature configuration C1 and C2. For both configurations we have used the following settings (η = 0.1, batch size = 60, number of training epochs = 50, number of randomized patches = 25000, number of feature maps in each layer $(C1, S1) = 6$, $(C2, S2) = 12$). Training and validation (50:50 split) was performed by random patch samples from a single central slice of subject 1. The validation set of subject 1 was used to determine the best performing model during online stochastic gradient descent learning. After the best model was obtained test performance was assessed on single central slices of all other LPBA40 subjects. For quantitative validation we computed the Dice coefficient D_c for the overall cortical structure and for individual subcortical structures

$$D_c = \frac{2|A \cap B|}{|A| + |B|}. \tag{3.1}$$

Here D_c measures the set agreement between the ground truth labels and our computed brain parcels. The D_c score ranges from (0-1), where 1 means perfect agreement. For experiment C1, the complete cortical structure had a mean D_c of 0.85 ($\pm$ 0.04), whereas for C2, the same structure had a lower mean D_c of 0.73 ($\pm$ 0.04). The superior and middle frontal gyrus for C2 had a higher D_c and lower variance than for corresponding parcels in experiment C1. The parcellation performance for the superior and middle temporal gyrus for C1 and C2 could not be computed since they were not part of the central slice during testing. In general the D_c was low for small brain parcels in comparison to larger sub-cortical structures. Overall parcellation performance showed high variability and no significant difference between C1 and C2 could be found. We have used the convolutional network implementation provided by Theano [5].

4. DISCUSSION AND CONCLUSION

In this paper we have presented a novel application of biologically inspired cortical network models to automate brain image parcellation using a deep convolutional network architecture. We were able to demonstrate parcellation of the complete cerebral cortex, without human intervention to build handcrafted features or to provide other prior knowledge. The feature configurations were able to correctly reject the detection of the main white matter regions. We attribute the low parcellation performance and high inter-subject variability to the very limited training set that we used. Another factor that affected the performance was the crud registration of the dataset causing the central slices that were used for training and testing to be misaligned. Misalignment caused by registration errors however can be accounted for by enriching the training set samples from a slap of slices. In future work

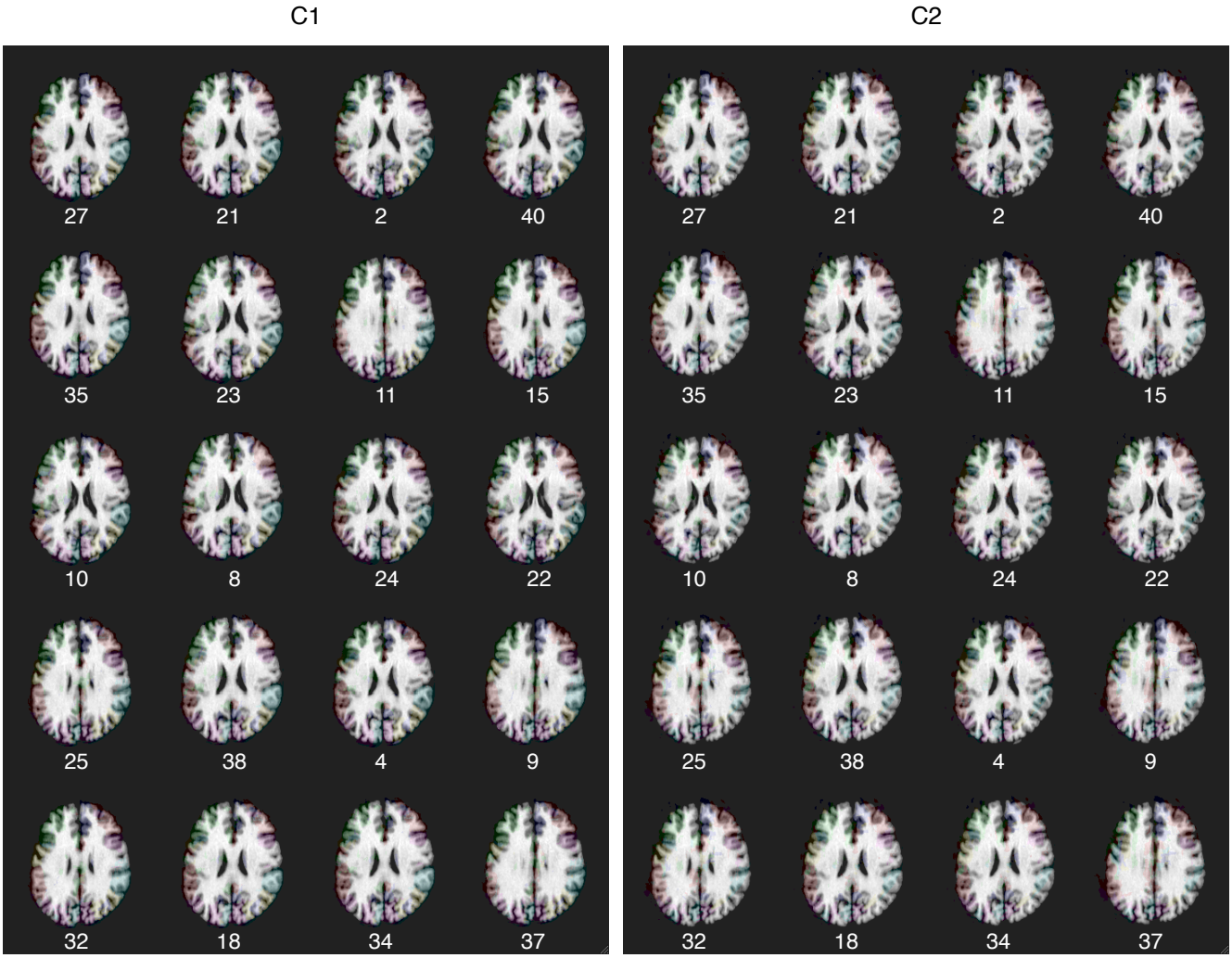

Fig. 3. **Qualitative performance results on automatic brain parcellation.** *Shown are central slices from 20 randomly selected subjects from the LPBA40 dataset. Left: computed brain parcels using feature configuration C1 as translucent color overlays. Right: computed brain parcels using feature configuration C2 as translucent color overlays. LPBA40 subject IDs are shown in white below each slice.*

we plan to improve upon the results obtained by these initial experiments and to extend our current approach to three-dimensional, context-aware feature learning and in-depth validation of the model in a clinical setting.

5. ACKNOWLEDGEMENTS

This work was supported by grant NIMH R01 MH084029-02.

6. REFERENCES

[1] J. W. Bohland, Hemant Bokil, Cara B Allen, and Partha P Mitra, "The brain atlas concordance problem: quantitative comparison of anatomical parcellations.," *PloS one*, vol. 4, no. 9, pp. e7200, Jan. 2009.

[2] A. Gholipour, N. Kehtarnavaz, S. Member, R. Briggs, and M. Devous, "Brain Functional Localization : A Survey of Image Registration Techniques," vol. 26, no. 4, pp. 427–451, 2007.

[3] Y. Bengio and Y. LeCun, "Scaling learning algorithms towards AI," in *Large Scale Kernel Machines*, Léon Bottou, Olivier Chapelle, D. DeCoste, and J. Weston, Eds. MIT Press, 2007.

[4] Y. Lecun, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.

[5] J. Bergstra, O. Breuleux, F. Bastien, P. Lamblin, R. Pascanu, G. Desjardins, J. Turian, and Y. Bengio, "Theano: a CPU and GPU math expression compiler," June 2010.