A Convolutional Autoencoder for Multi-Subject fMRI Data Aggregation

Po-Hsuan Chen[1], Xia Zhu[2], Hejia Zhang[1], Javier S. Turek[2], Janice Chen[3], Theodore L. Willke[2], Uri Hasson[3], Peter J. Ramadge[1]
[1]Department of Electrical Engineering, Princeton University, [2]Intel Labs,
[3]Princeton Neuroscience Institute and Department of Psychology, Princeton University

Abstract [1]

Finding the most effective way to aggregate multi-subject fMRI data is a long-standing and challenging problem. It is of increasing interest in contemporary fMRI studies of human cognition due to the scarcity of data per subject and the variability of brain anatomy and functional response across subjects. The standard method for aggregating such data uses anatomical registration across subjects [2, 3, 4]. Since this does not adequately align subjects' functional responses [4, 5, 6, 7, 8], there is a growing body of recent research exploring more direct approaches to functional registration. This includes cortical warping to align functional time series [6] and functional connectivity across subjects [7, 9], and the application of factor methods[10, 11, 12, 13, 14, 15, 8]. The most recent work in this vein, called the shared response model (SRM) [16], has focused on learning probabilistic latent factors that jointly model subject specific functional topographies and a shared temporal response.

These works show promising results but generally do not preserve spatial locality in the brain. By preserving spatial locality we mean that information is only aggregated in a small region (e.g. ball) about each voxel. We focus on the preservation of spatial locality during whole brain multi-subject data aggregation with the aim of improving anatomical and functional interpretability of the analysis results. Preserving spatial locality helps with interpretability because its straight forward to pin point the searchlight within the brain of which the signal comes from. A natural approach is that can satisfy this constraint to combine factor models and searchlight based analysis [17, 18]. Searchlight analysis uses a small window of contiguous voxels around a known location to conduct a spatially local analysis. This analysis is performed at all locations in the volume, thus generalizing an ROI approach to multiple (overlapping) spatially local "searchlights" across the brain. We focus on this approach with an aim of making a connection between searchlight analysis and convolution neural networks. Other approaches that aim to ensure spatial locality are also possible. For example, a data-driven approach that learns "soft" boundaries of local activated areas.

Our goal then is to design a multi-layer convolutional autoencoder for multi-subject, whole brain, spatially local, fMRI data aggregation. To understand the relevance of a convolutional autoencoder we note that a two layer fully connected autoencoder can replicate the performance of SRM on multi-subject fMRI data. But like the SRM, this autoencoder does not have spatial locality. We then argue that we can add spatial locality by transitioning from a fully connected to a convolutional autoencoder. To do so we create a network structure, Fig. 1, that matches the inherent multi-dataset nature of the problem and address some computational challenges arising from dealing with large-scale, multi-subject fMRI data. Our key contribution is to show that a suitably designed convolutional autoencoder can provide data aggregation that is competitive with methods based on whole brain searchlight analysis using latent factor methods, Fig. 2. We also examine approaches to address the computational challenges of training a convolutional autoencoder using multi-subject fMRI data. To our knowledge the application of a convolutional autoencoder to this task is novel and moves away from factor model approaches which appear to be hitting a performance ceiling. With further refinement, a well-trained convolution autoencoder may lead to a more powerful means of accomplishing the fMRI data aggregation task.

[1]Full paper [1]

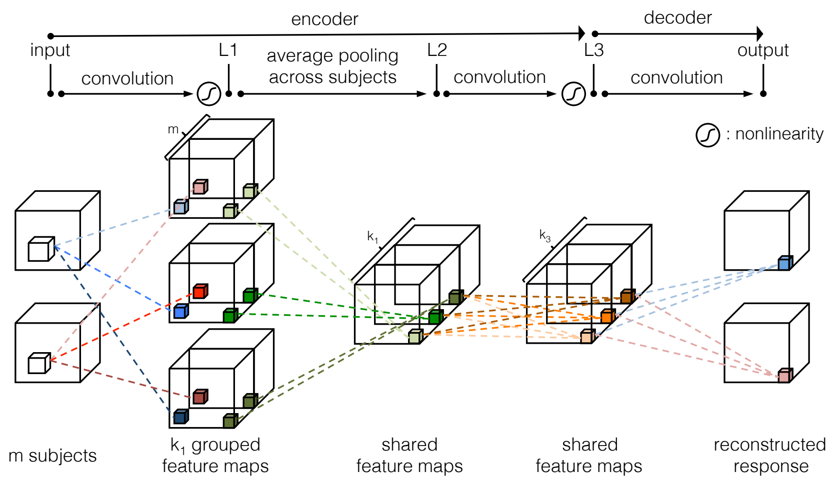

Figure 1: Proposed 4D Convolutional Autoencoder.

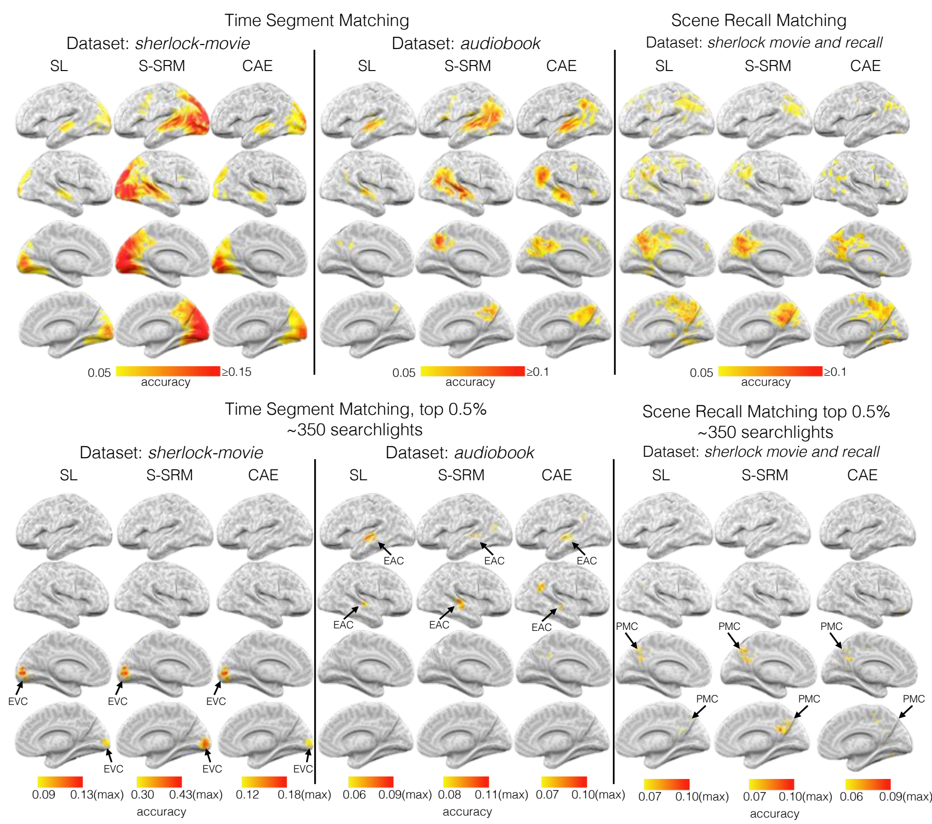

Figure 2: Comparison of searchlight (SL), searchlight-SRM (S-SRM), and proposed convolutional autoencoder (CAE). Experiment details in [1]. Top Left: Accuracy maps for time segment matching experiment using *sherlock-movie* and *audiobook*; Top Right: Accuracy maps for scene recall matching experiment using *sherlock-movie* and *sherlock-recall*. Top figures are thresholded at corresponding scales for visualization clarity purpose. Please refer to bottom row figures for high end of the range. Bottom Left: Accuracy maps for top 0.5% searchlights for time segment matching experiment; Bottom Right: Accuracy maps of top 0.5% searchlights for scene recall matching experiment. Early Visual Cortex (EVC), Early Auditory Cortex (EAC), Posterior Medial Cortex (PMC).

References

[1] P.-H. Chen, X. Zhu, H. Zhang, J. S. Turek, J. Chen, T. L. Willke, U. Hasson, and P. J. Ramadge. A convolutional autoencoder for multi-subject fmri data aggregation. *arXiv preprint arXiv:1608.04846*, 2016.

[2] J. Talairach and P. Tournoux. *Co-planar stereotaxic atlas of the human brain. 3-Dimensional proportional system: an approach to cerebral imaging*. Thieme, 1988.

[3] B. Fischl et al. High-resolution intersubject averaging and a coordinate system for the cortical surface. *Human brain mapping*, 1999.

[4] J. Mazziotta et al. A probabilistic atlas and reference system for the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 2001.

[5] M. Brett and et al. The problem of functional localization in the human brain. *Nat Rev Neurosci*, 2002.

[6] M. Sabuncu et al. Function-based intersubject alignment of human cortical anatomy. *Cereb. Cortex*, 2010.

[7] B. R. Conroy et al. fMRI-based inter-subject cortical alignment using functional connectivity. In *Adv. in Neu. Inf. Proc. Sys.*, 2009.

[8] J. V. Haxby et al. A common, high-dimensional model of the representational space in human ventral temporal cortex. *Neuron*, 2011.

[9] B. Conroy et al. Inter-subject alignment of human cortical anatomy using functional connectivity. *NeuroImage*, 2013.

[10] A. Hyvärinen et al. *Independent component analysis*. John Wiley & Sons, 2004.

[11] V.D. Calhoun et al. A method for making group inferences from functional MRI data using independent component analysis. *Human brain mapping*, 2001.

[12] J. Lee et al. Independent vector analysis: multivariate approach for fMRI group study. *Neuroimage*, 2008.

[13] Y.-O. Li et al. Joint blind source separation by multiset canonical correlation analysis. *IEEE Transactions on Signal Processing*, 2009.

[14] E. B. Erhardt et al. Comparison of multi-subject ica methods for analysis of fMRI data. *Human brain mapping*, 2011.

[15] A. M. Michael et al. Preserving subject variability in group fMRI analysis: performance evaluation of GICA vs. IVA. *Frontiers in systems neuroscience*, 2014.

[16] P.-H. Chen et al. A reduced-dimension fMRI shared response model. In *Adv. in Neu. Inf. Proc. Sys.*, 2015.

[17] N. Kriegeskorte et al. Information-based functional brain mapping. *Proc. of the Nat. Acad. of Sci.*, 2006.

[18] J. A. Etzel et al. Searchlight analysis: promise, pitfalls, and potential. *Neuroimage*, 2013.