MS-FSS Project Report: Analytical Stability in fMRI Statistical Maps

In recent years there has been a growing focus on evaluating the reproducibility of scientific findings. Meta-analyses have been performed across several domains of science [1]–[3] which have shown that a significant proportion of claims fail to replicate, posing a significant problem for the scientific community. In neuroimaging this issue has been explored with respect to the impact that analysis tools themselves have on claims and their replicability, and in functional Magnetic Resonance Imaging (fMRI) selection of software tool has been shown to produce different results [4]. Currently, there lacks efforts to harmonize derivatives produced by these various tools, and there is no evaluation of their relative stability and the impact this has on inferences. We attempt to address both of these in particular with respect to the generation of statistical parametric maps (SPMs) across various software packages and noise settings.

Project Accomplishments

This harmonization required the development of a three-dimensional deep convolutional neural network autoencoder (3D-CNN-AE) to learn a common encoding for the SPMs, and then reconstruct the maps from this common representation. The necessity of using a deep architecture for this task, as opposed to a much simpler averaging scheme, was driven by the desired use case of providing a single arbitrarily-generated map and regenerating it without the tool-inherent bias. The architecture and code developed are publicly available at https://github.com/gkiar/3d-cnn-ae/. The design consists of a symmetric set of three convolutional and pooling encoding layers with matching deconvolution and unpooling decoding layers, and was based initially upon the architecture demonstrated in [5]. Several parameters of the architecture were modified or inferred where incompletely specified, such as the width of several convolution kernels or the upsampling strategy, respectively. The resulting architecture was ultimately modified to achieve sufficient performance, as determined both by visual inspection and sufficient minimization of our loss function that we believe differences to be negligible between the original and reconstructed maps. This optimization was performed on our dedicated training set, discussed next, and verified using both a test set and finally the single-application upon our held-out validation set.

Since deep learning techniques typically require large datasets for training, and this is especially true here given the high dimensional data being processed (approximately 50x50x50 voxels per map), the network training needed to be split into two distinct phases: **weight initialization**, and **one-shot siamese optimization**. The goal of weight initialization was to establish the network with the ability to encode and decode SPMs, regardless of their tool or dataset of origin. Because of the limited metadata required for this reconstruction (i.e. no tool or source-data information is being used), we leveraged the NeuroVault [6] repository of publicly available SPMs produced and published by researchers. After performing basic quality control, such as ensuring maps existed within the same physical space as one another and that we knew the statistic encoded in each, this database provided a cohort of 52,398 maps when

accessed on April 8th, 2019. The dataset was split into training, testing, and validation sets consisting of 90%, 9%, and 1% of the samples, respectively. The samples were randomly generated into these classes reproducibly across runs by fixing a random seed, and were stratified across the statistic encoded (Z- or T-statistic) and the analysis level which produced the map (group or subject). The training set was used for the initial construction, testing, and tuning of the network. Once satisfied, the test set was used to verify out-of-sample performance. The validation set has not yet been used and will only be used upon completion of the downstream analytical experiments to verify true out-of-sample performance, ensuring that the model was not overfit to both the training and test sets during its iterative development.

Ongoing Activities

The project started by this Michael Smith Foregin Study Supplement will continue beyond the exchange and become a piece of one chapter of my Ph.D. thesis. While the current state of the project is the refinement of the weight initialization of our network, once that is completed the siamese training will take place using statistic maps generated using FSL [7], AFNI [8], and SPM [9] (the three most popular libraries for generating functional maps) and the Human Connectome Project Dataset [10]. The work completed as a result of this project has the potential to significantly reduce tool-introduced bias and numerical instabilities in functional neuroimage processing, and will be submitted to MICCAI 2019, a world-class conference in medical image analysis and machine learning.

References

[1] J. P. A. Ioannidis, "Why Most Published Research Findings Are False," *PLOS Med.*, vol. 2, no. 8, p. e124, Aug. 2005.
[2] K. A. Baggerly and K. R. Coombes, "Deriving chemosensitivity from cell lines: Forensic bioinformatics and reproducible research in high-throughput biology," *Ann. Appl. Stat.*, vol. 3, no. 4, pp. 1309–1334, Dec. 2009.
[3] O. S. Collaboration, "Estimating the reproducibility of psychological science," *Science*, vol. 349, no. 6251, p. aac4716, Aug. 2015.
[4] A. Bowring, C. Maumet, and T. E. Nichols, "Exploring the impact of analysis software on task fMRI results," *Hum. Brain Mapp.*, vol. 0, no. 0, May 2019.
[5] Y. Zhao *et al.*, "Constructing fine-granularity functional brain network atlases via deep convolutional autoencoder," *Med. Image Anal.*, vol. 42, pp. 200–211, Dec. 2017.
[6] K. J. Gorgolewski *et al.*, "NeuroVault.org: a web-based repository for collecting and sharing unthresholded statistical maps of the human brain," *Front. Neuroinformatics*, vol. 9, 2015.
[7] M. Jenkinson, C. F. Beckmann, T. E. J. Behrens, M. W. Woolrich, and S. M. Smith, "FSL," *NeuroImage*, vol. 62, no. 2, pp. 782–790, Aug. 2012.
[8] R. W. Cox, "AFNI: Software for Analysis and Visualization of Functional Magnetic Resonance Neuroimages," *Comput. Biomed. Res.*, vol. 29, no. 3, pp. 162–173, Jun. 1996.
[9] K. J. Friston, P. Jezzard, and R. Turner, "Analysis of functional MRI time-series," *Hum. Brain Mapp.*, vol. 1, no. 2, pp. 153–171, 1994.
[10] D. C. Van Essen, S. M. Smith, D. M. Barch, T. E. J. Behrens, E. Yacoub, and K. Ugurbil, "The WU-Minn Human Connectome Project: An overview," *NeuroImage*, vol. 80, pp. 62–79, Oct. 2013.

Program Evaluation

This process of accepting and completing a study exchange through the MS-FSS program was simple and without problem. The only recommendation I'd make for the program is working with applicants on scoping their project based on the duration of their stay, as in this case I found that we planned far greater than was feasible in the three-month stay.