Functional Annotation of Human Cognitive States using Graph Convolution Networks

Anonymous Author(s)
Affiliation
Address
email

Abstract

A key goal in neuroscience is to understand brain mechanisms of cognitive functions. An emerging approach is to study "brain states" dynamics using functional magnetic resonance imaging (fMRI). So far in the literature, brain states have typically been studied using 30 seconds of fMRI data or more, and it is unclear to which extent brain states can be reliably identified from very short time series. In this project, we applied graph convolutional networks (GCN) to decode brain activity over short time windows in a task fMRI dataset, i.e. associate a given window of fMRI time series with the task used. Starting with a populational brain graph with nodes defined by a parcellation of cerebral cortex and the adjacent matrix extracted from functional connectome, GCN takes a short series of fMRI volumes as input, generates high-level domain-specific graph representations, and then predicts the corresponding cognitive state. We investigated the performance of this GCN "cognitive state annotation" in the Human Connectome Project (HCP) database, which features 21 different experimental conditions spanning seven major cognitive domains, and high temporal resolution task fMRI data. Using a 10-second window, the 21 cognitive states were identified with an excellent average test accuracy of 89% (chance level 4.8%). As the HCP task battery was designed to selectively activate a wide range of specialized functional networks, we anticipate the GCN annotation to be applicable as a base model for other transfer learning applications, for instance, adapting to new task domains.

1 Introduction

Identifying brain networks involved in human cognition has been one of the main goals of neuroscience research. Modern imaging techniques, such as functional magnetic resonance imaging (fMRI), provide an opportunity to accurately map the neural substrates of human cognition. An emerging topic in the literature is the identification of "brain states", characterized by a canonical spatial pattern of functional activity, which were found to associate with specific cognitive states. A popular approach to identify these brain states, called multi-voxel pattern analysis (MVPA), uses machine learning tools to decode which task a subject performed based on recordings of brain activity in task fMRI (17). But the algorithm is usually limited to specific cognitive domains and relies on long acquisition of brain activity with repeated blocks to accurately decode a brain state.

In this project, we proposed a GCN architecture for annotating human brain activity on a cognitive battery of 21 task states. Instead of using the averaging BOLD signals of each task trial or even the activation maps resulting from GLM analysis, GCN takes a short series of fMRI volumes as input, generates task-specific graph representations, and then predicts the corresponding cognitive labels. Comparing to the multi-class support vector machines classification, GCN achieved much higher performance in identifying a variety of cognitive states.

Submitted to 33rd Conference on Neural Information Processing Systems (NeurIPS 2019). Do not distribute.

2 Datasets and Preprocessing

In this project, we are using block-design task fMRI data from the Human Connectome Project (HCP) S1200 release. The minimal preprocessed fMRI data of the CIFTI format were used, which maps individual fMRI time series onto the standard surface template with 32k vertices per hemisphere. Further details on fMRI data acquisition, task design and preprocessing can be found in (3) and (9).

The task fMRI data includes seven cognitive tasks, which are emotion, gambling, language, motor, relational, social, and working memory. In total, there are 23 different cognitive states. Considering the short event design nature of the gambling trials (1.5s for button press, 1s for feedback and 1s for ITI), we evaluated the decoding models with and without the two gambling conditions and found a much lower precision and recall scores for gambling task (average f1-score = 61%) than other cognitive domains (average f1-score $> 91\%$). In the following experiments, we excluded the two gambling conditions and only reported results on the remaining 21 cognitive states. The detailed description of the tasks can be found in (3). A summary table is also shown in Tab. 1.

Table 1: Parameters of task-designs from HCP dataset.

Task domain	**Subjects**	**Runs**	**Cond**	**Volumes per run**	**Trials per run**	**Minimal Dura**(sec)
Working Memory	1085	2	8	405	10	25
Motor	1083	2	5	284	8	12
Language	1051	2	2	316	8	12
Social Cognition	1051	2	2	274	5	23
Relational Processing	1043	2	2	232	6	16
Emotion Processing	1047	2	2	176	6	18

3 Graph Signal Processing and Graph Convolution Network

Graph signal processing (GSP) provides a non-linear embedding tool to project brain activity to graph Laplacian modes associated with different frequencies (19). This method has been widely used in neuroscience studies, including parcellating brain areas (7), identifying functional areas and networks, (2) and generating connectivity gradients (15). Recently, studies have found that, by decomposing the fMRI signals during cognitive task using GSP, the graph modes of different frequencies strongly associated with cognitive performance and learning (16). These findings brought new opportunities for the application of GSP on neuroimaging analysis (11). Moreover, with the aid of deep learning techniques, the optimal graph representation of brain activity, i.e. a combination of graph modes in the Laplacian spectrum, could be automatically learned for a particular task.

Starting with a brain signal x, GSP first maps the signal onto a weighted graph $\mathcal{G} = (\mathcal{V}, \mathcal{E}, W)$ that defines a network structure among a set of brain regions. The set $\mathcal{V}$ is a parcellation of cerebral cortex into N regions, and $\mathcal{E}$ is a set of connections between each pair of brain regions, with its weights defined as $W_{i,j}$. Many alternative approaches can be used to build a brain graph $\mathcal{G}$, including different brain parcellation schemes and various types of brain connectomes (for a review, see (5)). Here we used the multimodal cortical parcellation of the cerebral cortex (8), which delineates 180 functional areas per hemispheree bounded by sharp changes in cortical architecture, function, connectivity, and topography. These brain regions were defined on the 32k surface template. The connections between brain areas were estimated by calculating the group averaged resting-state functional connectivity (RSFC) based on 1080 minimal prepossessed resting-state fMRI data (9). Additional preprocessing steps were applied before the calculation of RSFC, including regressing out the signals from white matter and csf, and bandpass temporal filtering on frequencies between 0.01 to 0.1 HZ. Functional connectivity was first calculated on individual brain using Pearson correlation and then normalized using Fisher z-transform before averaging among subjects. After that, the z-score connectivity values were weighted by a Gaussian kernel with the width equals to the mean, and a KNN-graph was built with each brain region only connected to its 8 strongest connected neighbours.

The spectral analysis of the graph signal relies on the graph Laplacian, which characterizes the smoothness of graph signals among inter-connected brain regions. The normalized graph Laplacian is defined as:

$$L = I - D^{-1/2} W D^{-1/2}, \tag{1}$$

where D is a diagonal matrix of node degrees and I is the identity matrix. As we assume the weights to be undirected and symmetric, the matrix L can be factored as $U\Delta U^T$, where $U = (u_0, \ldots, u_{N-1})$ is the matrix of Laplacian eigenvectors, also called graph Fourier modes, and $\Delta = \text{diag}(\lambda_0, \ldots, \lambda_{N-1})$ is a diagonal matrix with the corresponding eigenvalues, specifying the frequency of the modes. By decomposing task-evoked BOLD signals onto different graph frequencies, studies have found that the low frequency mode associated with motor learning (12), while the high frequency mode correlated with cognitive flexibility (16).

Merging the spectral graph theory with deep learning techniques, (4) first proposed a new architecture of graph convolutional neural networks (GCN). Instead of calculating a weighted sum of the inputs as in a feedforward neural network, GCN generates a linear combination of the graph modes across different frequencies in the Laplacian spectrum. To accomplish that, GCN uses the graph filter, which generalizes the classical convolution operations on regular grids to the graph domain. Specifically, for a signal x defined on the brain graph, i.e. assigning a feature vector to each brain region, its graph Fourier transform is defined as $\mathcal{L}x = U^T x$ and its inverse as $x = U(\mathcal{L}x)$. Based on this, the convolution operation between the graph signal x and a filter g_θ based on graph $\mathcal{G}$, is defined as their element-wise Hadamard product in the graph Fourier domain, i.e.:

$$x * g_\theta = U(U^T g_\theta) \odot (U^T x) = U G_\theta U^T x \tag{2}$$

where $G_\theta = \text{diag}(U^T g_\theta)$ and θ indicates a parametric model for g_θ, and $U^T x$ is actually to project the graph signal onto the full spectrum of graph modes. With different choices of θ, we can construct graph convolutional networks with different types of graph filters, which generate various combination of graph modes of the signal x and find the optimal graph representations for a particular prediction task.

To avoid calculating the eigen-decomposition of the graph Laplacian especially for large-scale graphs, ChebNet convolution (6) used a truncated expansion of the Chebychev polynomials, which are defined recursively by $T_k(x) = 2T_{k-1}(x) - T_{k-2}(x)$ with $T_0(x) = 1, T_1(x) = x$:

$$x * g_\theta = \sum_{k=0}^{K} \theta_k T_k(\widetilde{L}) x \tag{3}$$

where $\widetilde{L}$ is a normalized version of graph Laplacian, equals to $2L/\lambda_{\text{max}} - I$, with λ_{max} being the largest eigenvalue. Kipf and colleagues (13) introduced a simplified version of GCN by taking a first-order approximation of the above Chebychev polynomial expansion and $\lambda_{\text{max}} \approx 2$:

$$x * g_\theta = \theta(I + D^{-1/2} W D^{-1/2}) x, \tag{4}$$

where θ is a single parameter to be learned. Thus, the output of a graph convolution layer is defined as:

$$X^{l+1} = \sigma(\widetilde{W} X^l \Theta^l), \quad \widetilde{W} = I + D^{-1/2} W D^{-1/2} \tag{5}$$

where $X^l \in \mathbb{R}^{N \times F}$ denotes the matrix of input graph signals on layer l, with N brain regions and F graph filters. $\Theta^l \in \mathbb{R}^{F_{in} \times F_{out}}$ is the parameters to be learned on layer l with F_{in} income channels/filters and F_{out} outcome filters. These parameters are shared among all nodes on layer l. $\sigma(.)$ denotes an activation function, such as the $\text{ReLU}(x) = \max(0, x)$. It's worth noted that this type of graph convolution only takes into account the direct neighbours in the graph for each brain region. By stacking multiple GCN layers, we could propagate brain activity among k_{th}-order neighbourhood, with k is the number of convolution layers. Specifically, in the first GCN layer, we treat the short series of fMRI volumes as multiple channels, with $X^1 \in \mathbb{R}^{N \times F}$ being a 2D matrix of N brain regions and F time windows. During model training, the first GCN layer learns various versions of the spatio-temporal convolution kernel (spatially only consider the direct neighbours, and temporally use the same size as the time window) for fMRI time-series, as a replacement of the canonical hemodynamic response function (HRF). The derived activation maps are then inputted to the following convolutional layers to generate high-level graph representations.

To summarize, a brain state annotation model was proposed based on the GCN architecture, consisting of 6 graph convolutional layers, followed by a global average pooling layer and 2 fully connected layers. The model takes short series of fMRI data as input, propagates information among interconnected brain regions and networks, generates a high order graph-level representations and finally predicts the corresponding cognitive labels as a multi-class classification problem. The overview of the fMRI decoding model was illustrated in Figure 1.

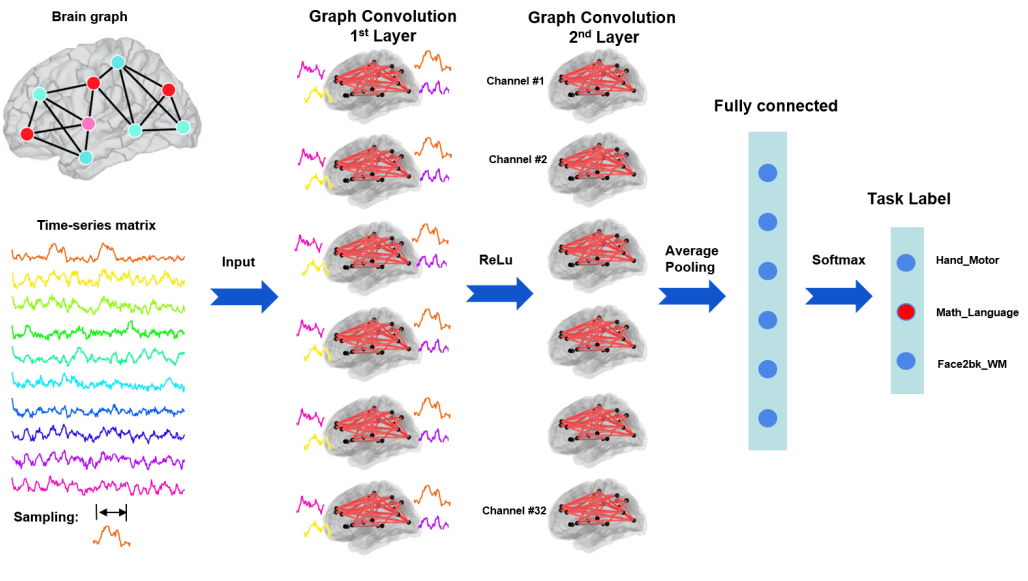

Figure 1: Pipeline of functional brain decoding using graph convolutions

The implementation of the GCN model was based on Tensorflow 1.12.0, which includes 32 graph filters for each of the 6 graph convolutional layers. The rectified linear unit (ReLU) function (14) was used as the activation function for all layers except the last layer where the softmax function was used to predict the cognitive labels. The network was trained for 100 epochs with the batch size set to 128. We used Adam as the optimizer with the initial learning rate as 0.001. Additional l2 regularization of 0.0005 on weights was used to control model overfitting and the noise effect of fMRI signals. The entire dataset was split into training (70%), validation (10%), test (20%) sets using a subject-specific split scheme, which ensures that all fMRI data from the same subject was assigned to one set. The best model with the highest prediction score on the validation set was saved and then evaluated separately on the test set.

3.1 Saliency map of graph convolutions

There are several ways to visualize a deep neural network, including visualizing layer activation (22) and filters (18), and heatmaps of class activation (21). Here, we chose the first method due to its easy implementation and generalization to graph convolutions. The basic idea is that if an input is relevant, a little variation on it will cause high change in the layer activation. This can be characterized by the gradient of the output given the input, with the positive gradients indicate that a small change to the input signals increase the output value. To visualize the gradients, we could simply use a backward pass of the activation of a single unit through the network. However, this type of maps are usually very noisy, and uninversible pooling operations and nonlinear activation function can bias the gradient. To alleviate these problems, (22) proposed to suppress the flow of gradients through neurons wherein either of input or incoming gradients were negative. Specifically, for the graph signal X^l of layer l and its gradient R^l, the overwritten gradient $\nabla_{X^l} R^l$ can be calculated as follows:

$$\nabla_{X^l} R^l = (X^l > 0) \odot (\nabla_{X^{l+1}} R^{l+1} > 0) \odot \nabla_{X^{l+1}} R^{l+1} \tag{6}$$

To calculate the saliency map based on the trained model, we started from the output layer and used the above chain rule to propagate the gradients at each layer until reaching the input layer. This guided-backpropagation approach can provide high resolution saliency map with the same dimension as the input data. Since we have used multiple time channels in the first layer of GCN model, the approach also provides one saliency map per timestamp. Considering the fluctuation of task-evoked BOLD signals, we generalized the calculation of heatmaps by taking the variance across time steps and scaling to the range [0,1].

4 Results

4.1 Cognitive states can be decoded with high accuracy from 10s of fMRI data

The GCN state annotation model was evaluated using the cognitive battery of HCP task-fMRI dataset. Here we mainly focused on the 21 task conditions spanning over 6 cognitive domains, namely: emotion, language, motor, relational, social, and working memory. Using a 10 second window of fMRI time series, the 21 conditions can be identified with an average test accuracy of 89.83%, significantly different from the chance level of 4.8%. The confusion matrix showed a nice block diagonal architecture (Figure 2 A). After summarizing the confusion matrix according to the 6 cognitive domains, each cognitive domain could be identified with a recall accuracy >91% (Figure 2 B). Among the cognitive domains, the language tasks (story vs math) and motor tasks (left/right hand, left/right foot and tongue) were the most recognizable conditions, which showed the highest precision and recall scores (average f1-score = 95% and 94% respectively for language and motor conditions). Meanwhile, the relational processing and working memory task conditions were identified with the lowest performance (average f1-score = 81.6% and 84.2% respectively), with some misclassifications between emotion/relational processing and working memory tasks.

4.2 Classification errors due to high similarity in task stimuli

Misclassifications of cognitive states were found both between and within cognitive domains. For instance, the emotion and relational processing conditions were misidentified as working memory task, probably due to the high similarity in task stimuli. Specifically, the emotion task involves the matching of faces, overlapping with face encoding and retrieval in working memory tasks. The relational processing requires matching of shapes and textures of drawn objects, somewhat similar to encoding and retrieval of bodies and tools in working memory task. Also, some trials were misclassified within the same cognitive domain. For instance, most misclassifications within working memory task were found between 0-back and 2-back conditions, which were still observed even when the decoding model was trained for the single cognitive domain. (Figure 3). By contrast, for face and place working memory stimuli, brain decoding reached high accuracy, regardless of using a multi-domain or single-domain classifier (misclassification less than 0.2%).

4.3 Performance of GCN annotation was associated with in-scanner behaviors

We found strong association between prediction accuracy of GCN state annotation and the median reaction time within scanner (Figure 4). For instance, during relational processing task which consists of two conditions, i.e. relational processing and feature matching, participants reacted faster to the matching condition than relational processing. Similarly, GCN also achieved higher prediction for matching (f1=0.96) than relational processing (f1=0.91). Moreover, within each task condition, GCN achieved higher accuracy on trials when participants were more engaged and responded faster. The analysis was performed on 200 subjects from the test set.

4.4 GCN achieved much better performance than SVC

We also wanted to establish if the performance of deep GCN represented a substantial improvement over more traditional machine learning tools. We thus also evaluated the same brain decoding tasks using a multi-class support vector machines classification (SVC) with a linear kernel, as our baseline model. The results showed that, using 10s of fMRI data, SVC-linear achieved much lower prediction accuracy in classifying the 21 states (89% vs 63% respectively for GCN and SVC-linear) and took a longer time for training (560s vs 9518s). Even when only focusing on a single cognitive domain, SVC-linear still showed much lower performance (96% vs 87% for motor task; 86% vs 70% for working memory conditions).

4.5 Saliency map demonstrate biologically meaningful features learned by GCN

We investigated whether a set of biologically meaningful features were learned by the GCN. For this purpose, we generated saliency maps on the model trained on the corresponding cognitive domain, by propagating the non-negative gradients backwards till the input layer. An input feature is salient or

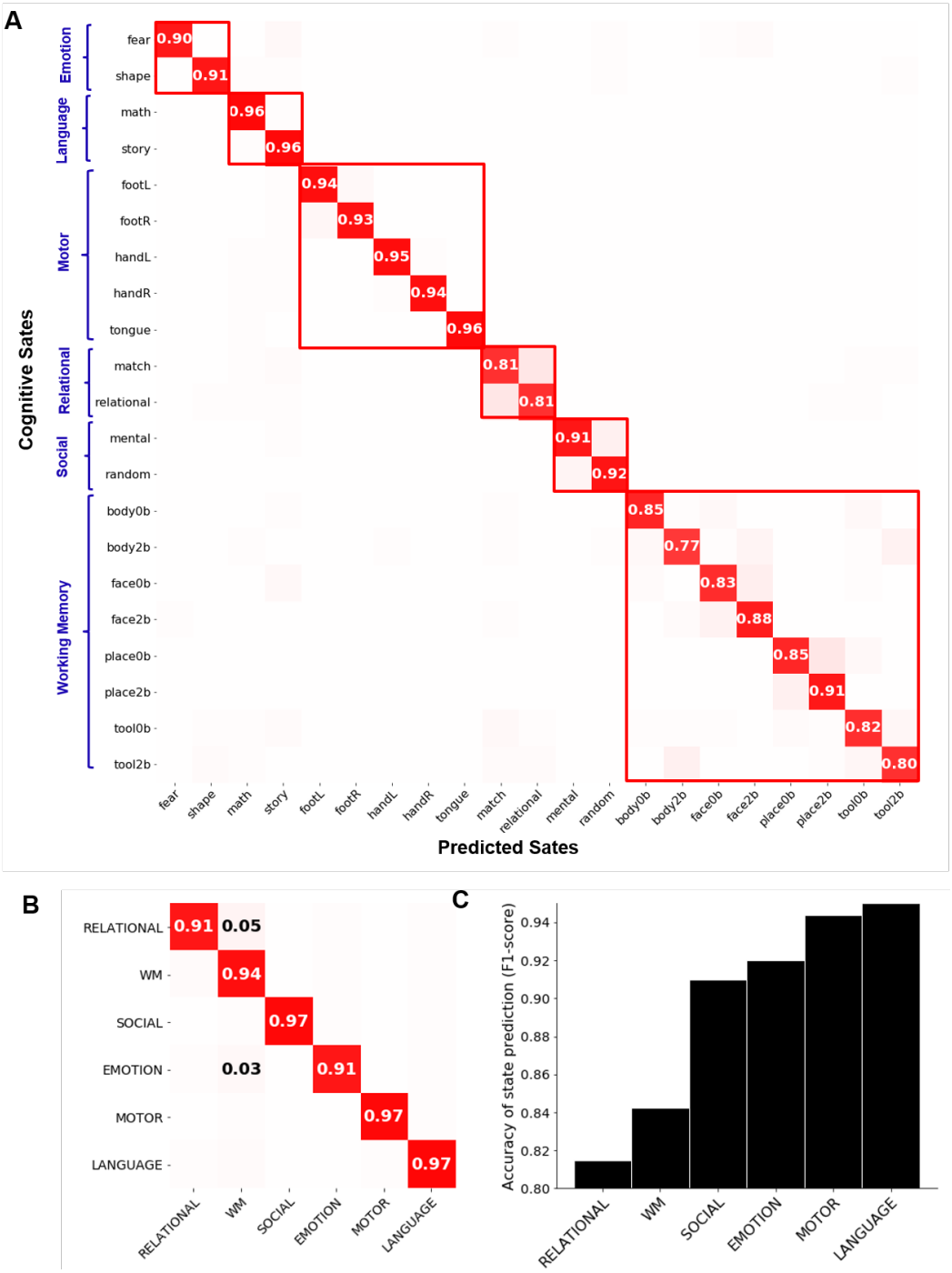

Figure 2: Confusion matrix of 21 cognitive states using 10s of fMRI data in GCN annotation (A). Confusion matrix (B) and f1-scores (C) were then summarized according to the six cognitive domains.

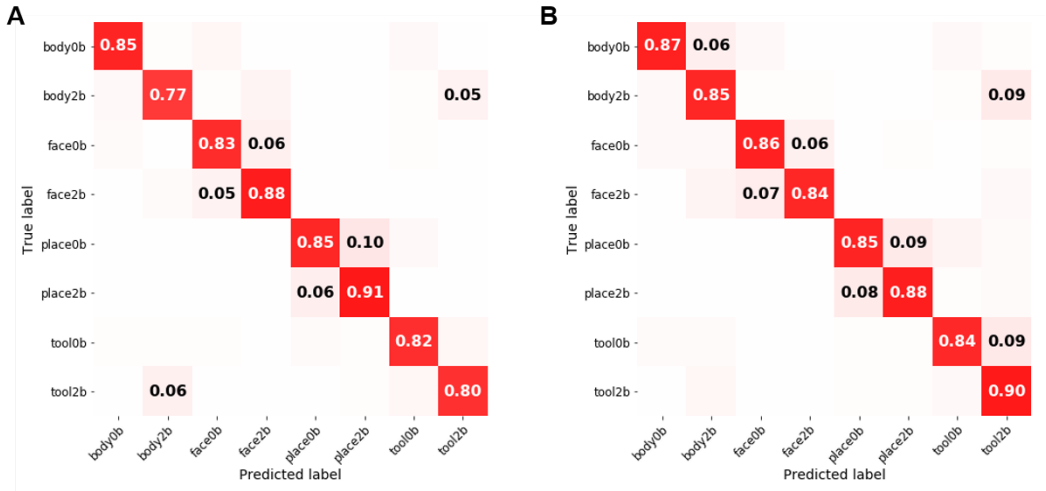

Figure 3: Misclassification table of the conditions within working memory task. The confusion matrix was either extracted from the full model of 21 cognitive conditions (A) or calculated separately for the single cognitive domain (B). Both models were trained using 10s of fMRI time series.

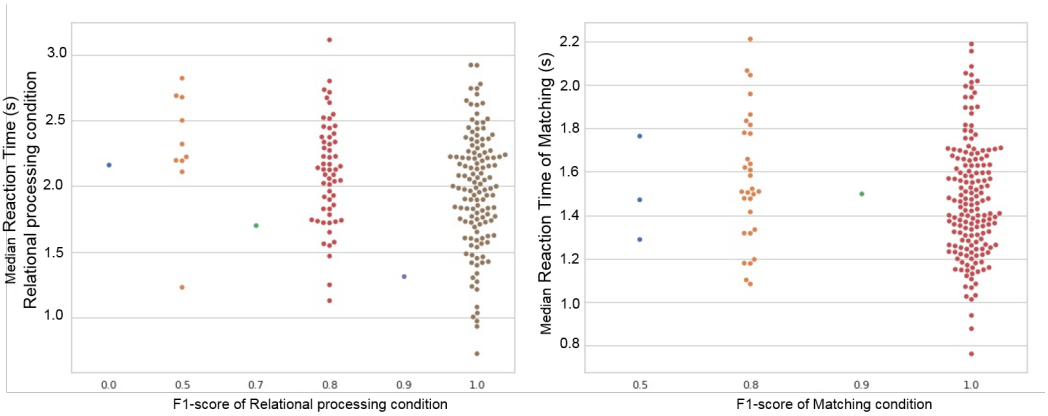

Figure 4: Association between GCN performance and in-scanner reaction time for relational processing task

important only if its little variation causes big changes in the decoding output. Thus, high values in the saliency map indicate large contributions during prediction of cognitive states.

The two language conditions, story and mathematics, shared a number of salient features (Figure 5 A), likely related to shared cognitive processes. Both conditions involved the processing of auditory statements, which may explain high salience in the primary auditory cortex and perisylvian language-related brain regions, consisting of inferior frontal gyrus (IFG), supramarginal gyrus/angular gyrus, and superior temporal gyrus (STG). There were also some salient features found only for either mathematics or story. For instance, the story condition involved salient features in more anterior part of left IFG, including pars triangularis and orbitalis. By contrast, mathematical statements involved more posterior part, including pars opercularis of IFG and precentral sulcus. Additional inferior temporal regions were salient for mathematics only, which have been shown to be more involved in mathematical than non-mathematical judgment tasks (1).

As expected, no salient features in the perisylvian language-related brain regions were found for the motor task (Figure 5 B). Different types of movements were associated with high salience in the primary motor and somatosensory cortex, which has been shown as the main territories engaged during movements of the human body (20). No clear somatotopic organization among different types of movements were identified here, which was somewhat expected because the primary motor and

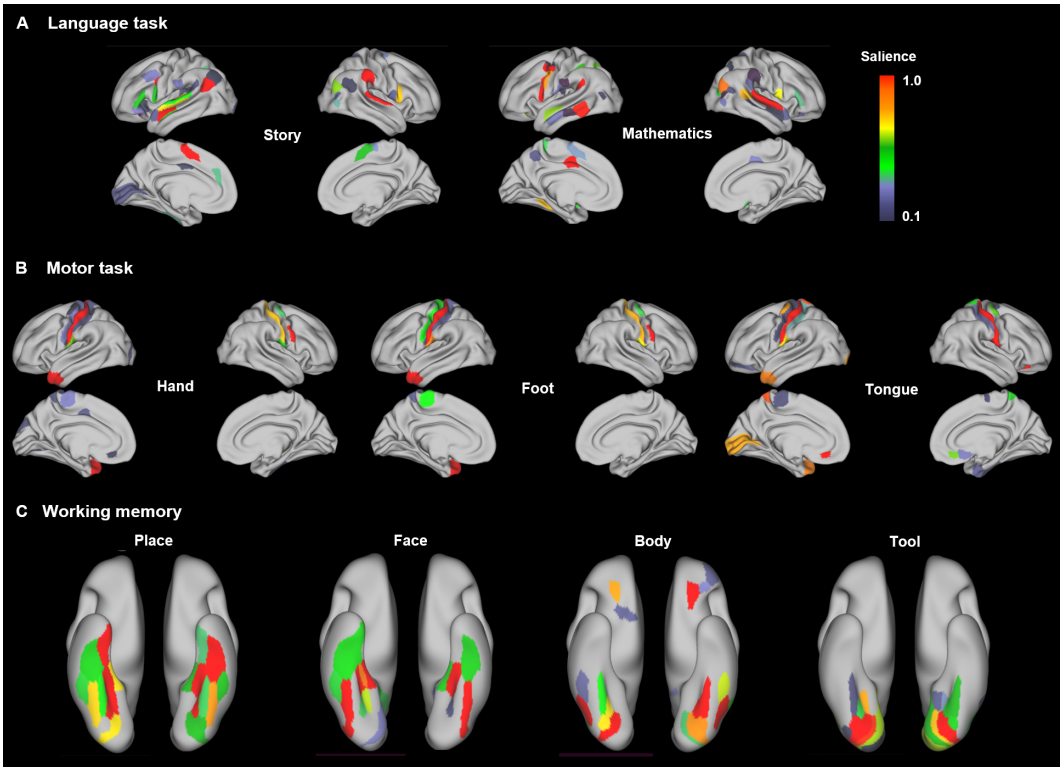

Figure 5: Saliency maps of language,motor and working memory tasks

somatosensory cortex were parcellated into single strips in the Glasser's atlas (8). Some category-specific salient features were still identified, for instance in medial primary motor cortex for foot movement and in lateral orbitofrontal cortex for tongue movement.

Moreover, salient features in the ventral visual stream were identified for image recognition in working memory task (Figure 5 C). Specifically, the place stimuli activated more medial areas in the temporal cortex including parahippocampal gyrus; while the face stimuli activated more lateral temporal regions including fusiform gyrus. This observation is consistent with the known, strong spatial segregation of the neural representation for face vs place image, in fusiform face area (FFA) and parahippocampal place area (PPA) respectively (10).

In summary, the regions highlighted by the saliency maps are consistent with prior knowledge from the literature, and suggest that the GCN model has learned biologically meaningful features, rather than confounding effects for example motion artifacts.

5 Conclusion

We propose a new graph convolution based model to annotate human cognitive states using a short series of fMRI signals. This model annotates human brain activity with fine temporal resolution and fine cognitive granularity. Using a 10s window of fMRI signals, our model identified 21 different task conditions with a test accuracy of 89%. Besides, the GCN performance was associated with participants' in-scanner behaviors. The saliency maps of GCN also demonstrated that biologically meaningful and domain-specific features have been learned during model training. Together, our project provides an automated tool to annotate the dynamics of human cognition in real time.

References

[1] Amalric, M., Dehaene, S.: Origins of the brain networks for advanced mathematics in expert mathematicians. Proceedings of the National Academy of Sciences **113**(18), 4909–4917 (2016)

[2] Atasoy, S., Donnelly, I., Pearson, J.: Human brain networks function in connectome-specific harmonic waves. Nature communications **7**, 10340 (2016)

[3] Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., et al.: Function in the human connectome: task-fmri and individual differences in behavior. Neuroimage **80**, 169–189 (2013)

[4] Bruna, J., Zaremba, W., Szlam, A., LeCun, Y.: Spectral networks and locally connected networks on graphs. arXiv preprint arXiv:1312.6203 (2013)

[5] Bullmore, E., Sporns, O.: Complex brain networks: graph theoretical analysis of structural and functional systems. Nature reviews neuroscience **10**(3), 186 (2009)

[6] Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. In: Advances in neural information processing systems. pp. 3844–3852 (2016)

[7] Fan, L., Li, H., Zhuo, J., Zhang, Y., Wang, J., Chen, L., Yang, Z., Chu, C., Xie, S., Laird, A.R., et al.: The human brainnetome atlas: a new brain atlas based on connectional architecture. Cerebral cortex **26**(8), 3508–3526 (2016)

[8] Glasser, M.F., Coalson, T.S., Robinson, E.C., Hacker, C.D., Harwell, J., Yacoub, E., Ugurbil, K., Andersson, J., Beckmann, C.F., Jenkinson, M., et al.: A multi-modal parcellation of human cerebral cortex. Nature **536**(7615), 171 (2016)

[9] Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al.: The minimal preprocessing pipelines for the human connectome project. Neuroimage **80**, 105–124 (2013)

[10] Golarai, G., Ghahremani, D.G., Whitfield-Gabrieli, S., Reiss, A., Eberhardt, J.L., Gabrieli, J.D., Grill-Spector, K.: Differential development of high-level visual cortex correlates with category-specific recognition memory. Nature neuroscience **10**(4), 512 (2007)

[11] Huang, W., Bolton, T.A., Medaglia, J.D., Bassett, D.S., Ribeiro, A., Van De Ville, D.: A graph signal processing perspective on functional brain imaging. Proceedings of the IEEE **106**(5), 868–885 (2018)

[12] Huang, W., Goldsberry, L., Wymbs, N.F., Grafton, S.T., Bassett, D.S., Ribeiro, A.: Graph frequency analysis of brain signals. IEEE journal of selected topics in signal processing **10**(7), 1189–1203 (2016)

[13] Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)

[14] Maas, A.L., Hannun, A.Y., Ng, A.Y.: Rectifier nonlinearities improve neural network acoustic models. In: Proc. icml. vol. 30, p. 3 (2013)

[15] Margulies, D.S., Ghosh, S.S., Goulas, A., Falkiewicz, M., Huntenburg, J.M., Langs, G., Bezgin, G., Eickhoff, S.B., Castellanos, F.X., Petrides, M., et al.: Situating the default-mode network along a principal gradient of macroscale cortical organization. Proceedings of the National Academy of Sciences **113**(44), 12574–12579 (2016)

[16] Medaglia, J.D., Huang, W., Karuza, E.A., Kelkar, A., Thompson-Schill, S.L., Ribeiro, A., Bassett, D.S.: Functional alignment with anatomical networks is associated with cognitive flexibility. Nature human behaviour **2**(2), 156 (2018)

[17] Norman, K.A., Polyn, S.M., Detre, G.J., Haxby, J.V.: Beyond mind-reading: multi-voxel pattern analysis of fmri data. Trends in cognitive sciences **10**(9), 424–430 (2006)

[18] Olah, C., Mordvintsev, A., Schubert, L.: Feature visualization. Distill **2**(11), e7 (2017)

[19] Ortega, A., Frossard, P., Kovačević, J., Moura, J.M., Vandergheynst, P.: Graph signal processing: Overview, challenges, and applications. Proceedings of the IEEE **106**(5), 808–828 (2018)

[20] Penfield, W., Boldrey, E.: Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain **60**(4), 389–443 (1937)

[21] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 618–626 (2017)

[22] Springenberg, J.T., Dosovitskiy, A., Brox, T., Riedmiller, M.: Striving for simplicity: The all convolutional net. arXiv preprint arXiv:1412.6806 (2014)