

Learning Human Cognition via fMRI Analysis Using 3D CNN and Graph Neural Network

Xiuyan Ni¹, Tian Gao², Tingting Wu³, Jin Fan³, and Chao Chen^{1,4}

¹ Department of Computer Science, The Graduate Center, City University of New York (CUNY), New York, NY, USA

² IBM Thomas J. Watson Research Center, Yorktown Heights, NY, USA

³ Department of Psychology, CUNY Queens College, Flushing, NY, USA

⁴ Department of Biomedical Informatics, Stony Brook University, NY, USA

Abstract. Human cognitive control involves how mental resources are allocated when the brain processes various information. The study of such complex brain functionality is essential in understanding different neurological disorders. To investigate cognition control, various cognitive tasks have been designed and functional MRI data have been collected. In this paper, we study uncertainty representation, an important problem in human cognition study, with task-evoked fMRI data. Our goals are to learn how brain region of interests (ROIs) are activated under tasks with different uncertainty levels and how they interact with each other. We propose a novel neural network architecture to achieve the two goals simultaneously. Our architecture uses a 3D convolutional neural network (CNN) to extract a high-level representation for each ROI, and uses a graph neural network module to capture the interactions between ROIs. Empirical evaluations reveal that our method significantly outperforms the existing methods, and the derived brain network is consistent with domain knowledge.

Keywords: Graph Neural Network, Brain Network Learning

1 Introduction

Cognitive control study learns how are mental resources allocated when the human brain processes various information. It involves how the brain selects and prioritizes different information processing tasks, which is crucial in understanding the mechanisms of different neurological disorders [17]. We study a particular perspective of cognitive control, the uncertainty representation. We design a task called Choice Reaction Time task to study the uncertainty representation. In our experiments, the human subjects are displayed with arrows of random directions and colors and instructed to press buttons accordingly. Depending on the number of possible directions and colors, the task may have different levels of uncertainty, measured as Shannon entropy [18].

Since our cognitive control task spans a very short time, each fMRI image is only a single 3D image (See Fig. 1 for more details), unlike resting-state fMRI or fMRI taken under tasks with long duration which can be considered 4D data with

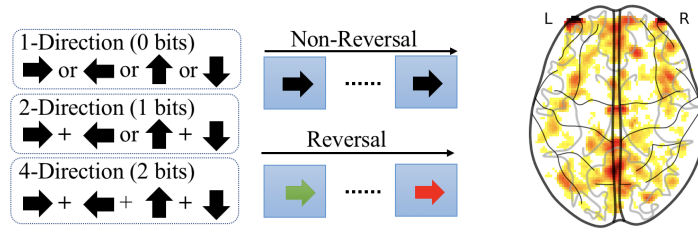


Fig. 1: **Left:** Experiment setting. Human subjects are shown images of arrows and click keys accordingly. The arrows may have one, two, or four possible directions, corresponding to uncertainty levels of zero, one, and two bits measured in Shannon’s entropy [18]. **Middle:** More uncertainty is introduced by additional colors and corresponding actions. Green arrows require buttons with corresponding directions. Red buttons require buttons with the opposite directions. In total there are six different tasks, with various uncertainty levels. **Right:** An example of fMRI data.

one additional dimension for time sequences. The traditional brain functional connectivity analysis methods [5, 1] designed for fMRIs with time-sequential information can not be applied to our data, which leaves us low flexibility for analyzing the networks. Recently, deep neural networks have been proved to be very efficient in learning high-level representations for various types of data. Directly applying neural networks to our data could improve the prediction performance. For our setting, however, we are not only interested in predicting uncertainty levels, but also want to learn how different brain regions interact with each other at different uncertainty levels.

In this paper, we propose a new neural network architecture to solve the two tasks jointly: one as a classification task, the other as a graphical model learning task. Our architecture uses a 3D convolutional neural network (CNN). It extracts high-level feature representation for the classification of fMRIs for different uncertainties, and a graph neural network (GNN) layer for extracting the edge representations. The edge representations are further converted into a Markov Random Field (MRF) and learned through a loss derived from the likelihood of the MRF. Our architecture learns both ROI representations for tasks of different uncertainty levels and a graphical model to encode how different ROIs interacts. We test our method on a task-evoked fMRI dataset, and found that our model outperforms the existing state-of-the-art classifiers. In the meantime, our model generates a brain network for uncertainty representation.

Related Work. The fMRI data are widely used for analyzing neurological disorders and locating task-related key regions [7, 14]. Recently, neural network models have been broadly applied to fMRI data based study [4, 2, 24], and also for learning brain networks [20, 2, 24]. Belilovsky et al. [2] use convolutional neural networks for learning brain networks. Their model takes pre-computed covariance matrices as inputs and outputs the same dimensional matrices. The outputs can be viewed as adjacency matrices of the brain network. Nonetheless, their method requires each region to be represented by one single value, which

might lose a lot of information. Instead of using pre-calculated covariance matrix, we take all the information from brain regions as inputs for our architecture and extract high-level features for each region. In our architecture, valuable information is leveraged by the neural networks and can be used for multitask learning. Zhang et al. [24] apply Graph Convolutional Network to brain image analysis for the prediction of Parkinson’s disease. However, their graph convolutional network is not used for learning brain networks.

2 Method

2.1 Network Architecture

We propose a novel neural network architecture to classify fMRI images, and learn a brain network at the same time (Fig. 2). For our input, each fMRI image is a 3D volumetric image with uncertainty level ranging between 1 and 6. Our architecture uses a 3D convolutional layer for extracting high-level feature embeddings for brain ROIs, and a GNN layer with MRF for structure learning of human brain networks. We use traditional cross entropy loss for the classifying uncertainty levels and meanwhile a maximum-likelihood-loss for MRF learning. For each fMRI image, we extract R ROIs. The size of each ROI is $7 \times 7 \times 7$. The size of the input for the model is $R \times 1 \times 7 \times 7 \times 7$, which can be viewed as R single-channel 3D images. Let the number of inputs be N , and the number of ROIs be R . For each image $k \in \{1, \dots, N\}$, we denote each ROI r as $x_r^k, r \in \{1, \dots, R\}$. Each image is labeled with an uncertainty level $y^k \in \{1, \dots, 6\}$.

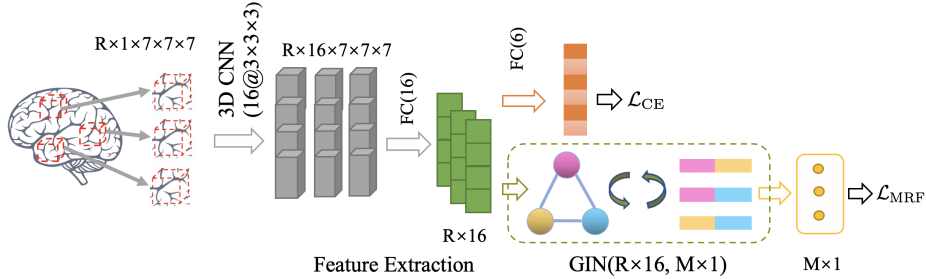


Fig. 2: The architecture of our model.

Our model starts with feature extraction for each ROI. This is done with a 3D CNN layer of 16 channels with kernel size $3 \times 3 \times 3$ (CNN(16@ $3 \times 3 \times 3$)) followed by a fully connected layer ($FC(16)$). The output feature is a vector of length 16 for each ROI. We perform two tasks on the output features:

- For the uncertainty classification task, we map the concatenation of the output of all R ROIs to a 1D vector of length 6 with a fully connected layer ($FC(6)$). The 1D vector is then fed to a softmax layer to calculate the label probability. The loss from this task is the cross entropy loss $\mathcal{L}_{CE}$.
- For brain network structure learning, we feed the R output features to the graph neural network layer and then use MRF to estimate the connection

between each ROI. More details will be discussed in Section 2.2. The loss from this task is $\mathcal{L}_{\text{MRF}}$.

Our overall loss to minimize is $\mathcal{L} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{MRF}}$.

2.2 Graph Neural Network Layer for Brain Network Learning

Graph neural network (GNN) has been proposed for learning the efficient representation for many graph-structured data [9]. GNN usually employs a message-passing schema, which means each node aggregates the information of its neighbors and transforms into new representation using the information. The generated new representation of the nodes captures the structured information of the inherited graph [12, 21].

In this work, we build a brain network learning layer based on an existing graph neural network called Graph Isomorphism Network (GIN) [23]. The input of the layer is a set of 16-dimensional vectors, each of which represents one ROI in the human brain. The output is edge potentials for each possible edge between the ROIs. GIN is used for helping with the message passing between nodes and edges in the graph. We use multi-layer perceptrons (MLPs) as our message-passing function in our model. We assume the initial graph is a fully connected graph with all possible edges. Thus, we have $M = R \times (R - 1)/2$ edges. We use the initial graph as input for GNN layer. Let $\mathcal{N}(v)$ be the neighbours of node v , $h_{t,v}$ be the representation of node v at t^{th} iteration, and ϵ_t be an arbitrarily small number, the GIN updates the node representation as

$$h_{t,v} = \text{MLP}_t \left((1 + \epsilon_t) \cdot h_{t-1,v} + \sum_{u \in \mathcal{N}(v)} h_{t-1,u} \right) \quad (1)$$

where $h_{0,v}$ is the high-level representation of v^{th} ROI after feature extraction. We denote the final output of GIN layer as $h_v = h_{T,v}$.

After the GIN layer, we construct the edge representation as

$$\phi_{i,j} = [h_i; h_j] \mathbf{W}_{\text{edge}} + \mathbf{b}_{\text{edge}},$$

where $\mathbf{W}_{\text{edge}} \in \mathbb{R}^{16 \times 32}$ and $\mathbf{b}_{\text{edge}} \in \mathbb{R}^{16}$ are the weights for the linear layer, and $[\cdot; \cdot]$ is the concatenation operation.

For image k , the energy function $E(x_1^k, \dots, x_R^k)$ can be calculated as the sum of potentials for each edge $f_{i,j}$. Thus, the MRF probability of data x^k can be calculated as:

$$\begin{aligned} P(x^k) &= \frac{1}{Z} \exp(-E(x^k)) = \frac{1}{Z} \exp \left(- \sum_{(i,j) \in \mathcal{E}} f_{i,j}(h_i^k, h_j^k) \right) \\ &= \frac{1}{Z} \exp \left(- \sum_{(i,j) \in \mathcal{E}} w_{i,j} \cdot \langle \phi_{i,j}^k, \theta_{i,j} \rangle \right), \end{aligned}$$

where $\theta_{i,j} \in \mathbb{R}^{16}$ is used to map the edge representation to a real value potential, $w_{i,j} \in \mathbb{R}$ is used to re-weight each edge, and Z is the partition function.

In theory, Z should be calculated over all distribution space, which is impractical. To alleviate this problem, Z can be approximated using all samples of the data, which is still very expensive to evaluate. In this paper, we approximate Z with all examples in the same training batch. $Z \approx \sum_{k=1}^B \exp(-E(x^k))$, in which B is the batch size. Our MRF learning module uses the negative log-likelihood as the loss, for data k , the MRF learning loss is:

$$\mathcal{L}_{\text{MRF}} = -\log \prod_{k=1}^N P(x^k) = \sum_{k=1}^N [E(x^k) + \log Z]. \quad (2)$$

3 Experiments and Discussions

We apply our method to task-evoked fMRI images. Our dataset is collected when the subjects are instructed to perform Choice Reaction Time (CRT) tasks [22]. At each CRT task, the subjects are presented with an arrow and instructed to press the corresponding buttons. The details can be found in Fig. 1. There are 6 uncertainty conditions manipulated by the directions and colors of the arrows, corresponding to 6 labels for the classification task. Our data contains 16 subjects. Each subject is asked to perform around 1000 trials. Each CRT task trial only lasts for two seconds, thus only one 3D image is collected for each trial. We collect 17226 fMRI 3D images in total from the 16 subjects. The images are preprocessed using SPM8. Each gradient-echo planar imaging (EPI) image volume was realigned to the first volume, registered with structural MRI, and normalized to the Montreal Neurological Institute (MNI) ICBM152 space. Then all the images are resampled to a voxel size of $2 \times 2 \times 2$ mm, and spatially smoothed. The dimension of the processed images is $79 \times 95 \times 68$.

Neuroscience studies state that cognitive control network (CCN) and default mode network (DMN) [11, 19, 15, 22] are two major networks in the human brain that are related to uncertainty tasks. In this work, we focus on $R = 19$ regions of interests (ROIs) from the two brain networks. Control network (CCN) [6] is composed of anterior cingulate cortex (ACC), anterior insula (AI), and frontal eye field (FEF), etc. Default mode network (DMN) [8] consists of domain-specific networks such as visual, auditory, etc.. Each ROI is with dimension $7 \times 7 \times 7$.

Our method is compared with Random Forest Classifier (RF), Linear Regression Classifier (LG), and SVM with Linear kernel (SVM) [3]. To prove the effectiveness of our graph neural network layer, we also conduct an ablation experiment which uses a simple 3D CNN model. The 3D CNN model is similar to our model, but did not use the GNN layer, and only trained on cross-entropy loss. For each subject, we test all methods. For all methods, we reserve the same 80% of subject's data as the training set, 10% as the validation set and 10% as the testing set. The 3D-CNN model and our model are trained using stochastic gradient descent optimizer (SGD) with learning rate = 0.001, $\epsilon = 0$ for 1000 epochs. The classification results are summarized in Table 1. We can find from Table 1 that 3D CNN model outperforms the traditional classifiers, and our GNN layer further improves the classification performance.

For brain network learning, we visualize the results in Fig. 3. In Neuroscience studies, the CCN is known as a network responsible for cognitive processes [10,

Table 1: The average classification accuracy (%) for 16 subjects.

	RF	LG	SVM	3D-CNN	Our Model
Accuracy	33.54	61.34	56.68	71.03	75.59

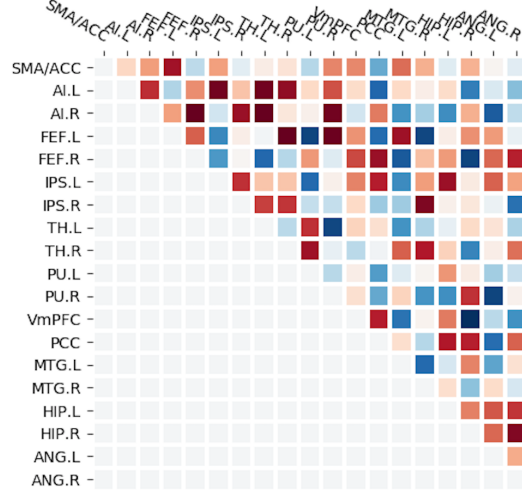


Fig. 3: The brain network learned among the 19 ROIs. SMA/ACC: supplementary motor area extending to anterior cingulate cortex. AI: anterior insular cortex. FEF: frontal eye find. IPS: area around and along the intraparietal sulcus. TH: thalamus. vmPFC: ventral medial prefrontal cortex. PCC: posterior cingulate cortex. MTG: middle temporal gyrus. ANG: angular gyrus. L: ROI located in left hemisphere of the brain. R: ROI located in the right hemisphere. The red box indicates positive connection and the blue box indicates the negative connection. Darker color means stronger connection.

8]. The DMN is considered as the network handling human’s self-related activities such as emotion and autobiographical memory [16]. The two networks work together for processing uncertainty tasks [13, 8]. From Fig. 3, we can find quantitative proof for the above statements. That is, we observe both intra-network and inter-network edges. We further visualize the network in CCN and DMN in brain templates respectively as in Fig. 4. The positive and negative edges are colored in warm and cool colors respectively. We can find from Fig. 4 clearly that the more connections can be found from CCN than DMN. The stronger intra-network connection in CCN means that CCN is more crucial in uncertainty processing, which is consistent with neuroscience knowledge about uncertainty representation [10, 22]. Overall, the GNN not only helps with the classification task but also generates meaningful brain network for uncertainty processing.

4 Conclusions

In this paper, we propose a novel neural network framework to classify the CRT task-evoked fMRI data, and learn the brain network. Our framework integrates a

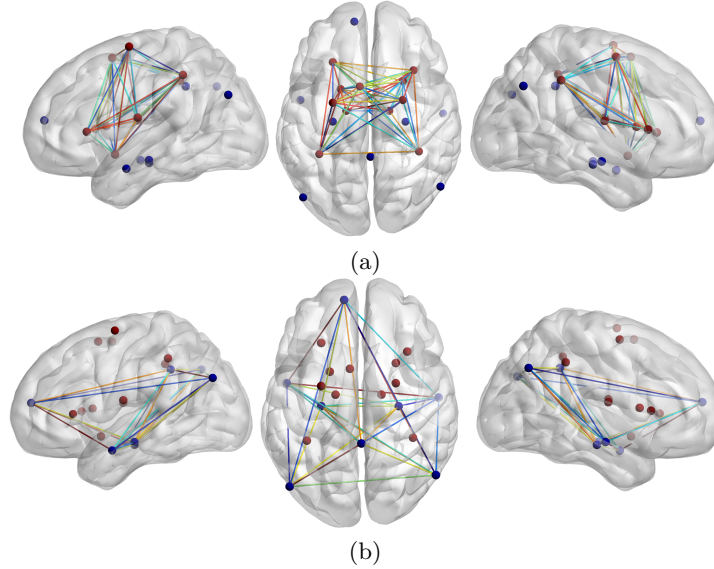


Fig. 4: The network in CCN (a) and DMN (b) in three views of brain templates. The nodes in red are the ROIs in CCN and blue nodes are the ROIs in DMN. The positive edges are in warm color, while the negative edges are in cool colors.

3D CNN for extracting high-level features for brain ROIs, a traditional softmax block trained on cross-entropy loss for classifying the fMRI images at different uncertainty levels, and a graph neural network layer trained on MRF loss for learning the structures of the brain network. Our method outperforms the traditional classifiers and the plain 3D CNN model. Besides, our model also learns the structure of the brain network, which is, how the brain ROIs interact with each other during the uncertainty tasks. Our model provides a quantitative assessment for cognitive control study and has the potential to be applied to any other labeled data with underlying graph structures.

Acknowledgement. This work was partially supported by NSF IIS-1855759 and CCF-1855760.

References

1. Bassett, D.S., Yang, M., Wymbs, N.F., Grafton, S.T.: Learning-induced autonomy of sensorimotor systems. *Nature neuroscience* **18**(5), 744 (2015)
2. Belilovsky, E., Kastner, K., Varoquaux, G., Blaschko, M.B.: Learning to discover sparse graphical models. In: *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. pp. 440–448. JMLR. org (2017)
3. Bishop, C.M.: *Pattern recognition and machine learning*. springer (2006)
4. Chen, P.H., Zhu, X., Zhang, H., Turek, J.S., Chen, J., Willke, T.L., Hasson, U., Ramadge, P.J.: A convolutional autoencoder for multi-subject fmri data aggregation. *arXiv preprint arXiv:1608.04846* (2016)

5. Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., Petersen, S.E.: Intrinsic and task-evoked network architectures of the human brain. *Neuron* **83**(1), 238–251 (2014)
6. Cole, M.W., Schneider, W.: The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage* **37**(1), 343–360 (2007)
7. Diamond, A., Barnett, W.S., Thomas, J., Munro, S.: Preschool program improves cognitive control. *Science* (New York, NY) **318**(5855), 1387 (2007)
8. Elton, A., Gao, W.: Task-positive functional connectivity of the default mode network transcends task domain. *Journal of cognitive neuroscience* **27**(12), 2369–2381 (2015)
9. Hamilton, W.L., Ying, R., Leskovec, J.: Representation learning on graphs: Methods and applications. *IEEE Data Engineering Bulletin* **40**(3), 52–74 (2017)
10. Hellyer, P.J., Shanahan, M., Scott, G., Wise, R.J., Sharp, D.J., Leech, R.: The control of global brain dynamics: opposing actions of frontoparietal control and default mode networks on attention. *Journal of Neuroscience* **34**(2), 451–461 (2014)
11. Kelly, A.C., Uddin, L.Q., Biswal, B.B., Castellanos, F.X., Milham, M.P.: Competition between functional brain networks mediates behavioral variability. *Neuroimage* **39**(1), 527–537 (2008)
12. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
13. Leech, R., Kamourieh, S., Beckmann, C.F., Sharp, D.J.: Fractionating the default mode network: distinct contributions of the ventral and dorsal posterior cingulate cortex to cognitive control. *Journal of Neuroscience* **31**(9), 3217–3224 (2011)
14. Ni, X., Yan, Z., Wu, T., Fan, J., Chen, C.: A region-of-interest-reweight 3D convolutional neural network for the analytics of brain information processing. In: *Medical Image Computing and Computer Assisted Intervention – MICCAI*. pp. 302–310 (2018)
15. Power, J.D., Schlaggar, B.L., Lessov-Schlaggar, C.N., Petersen, S.E.: Evidence for hubs in human functional brain networks. *Neuron* **79**(4), 798–813 (2013)
16. Raichle, M.E., MacLeod, A.M., Snyder, A.Z., Powers, W.J., Gusnard, D.A., Shulman, G.L.: A default mode of brain function. *Proceedings of the National Academy of Sciences* **98**(2), 676–682 (2001)
17. Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., Nieuwenhuis, S.: The role of the medial frontal cortex in cognitive control. *science* **306**(5695), 443–447 (2004)
18. Shannon, C.E., Weaver, W.: *The mathematical theory of communication*. University of Illinois press (1949)
19. Spreng, R.N., Sepulcre, J., Turner, G.R., Stevens, W.D., Schacter, D.L.: Intrinsic architecture underlying the relations among the default, dorsal attention, and frontoparietal control networks of the human brain. *Journal of cognitive neuroscience* **25**(1), 74–86 (2013)
20. Van Den Heuvel, M.P., Mandl, R.C., Kahn, R.S., Hulshoff Pol, H.E.: Functionally linked resting-state networks reflect the underlying structural connectivity architecture of the human brain. *Human brain mapping* **30**(10), 3127–3141 (2009)
21. Veličković, P., Cucurull, G., Casanova, A., Romero, A., Lio, P., Bengio, Y.: Graph attention networks. *International Conference on Learning Representations* (2018)
22. Wu, T., Dufford, A.J., Egan, L.J., Mackie, M.A., Chen, C., Yuan, C., Chen, C., Li, X., Liu, X., Hof, P.R., et al.: Hick-hyman law is mediated by the cognitive control network in the brain. *Cerebral Cortex* pp. 1–16 (2017)
23. Xu, K., Hu, W., Leskovec, J., Jegelka, S.: How powerful are graph neural networks? In: *International Conference on Learning Representations* (2019)

24. Zhang, X., He, L., Chen, K., Luo, Y., Zhou, J., Wang, F.: Multi-view graph convolutional network and its applications on neuroimage analysis for parkinson's disease. arXiv preprint arXiv:1805.08801 (2018)