

A Region-of-Interest-Reweight 3D Convolutional Neural Network for the Analytics of Brain Information Processing

Xiuyan Ni¹, Zhennan Yan², Tingting Wu³, Jin Fan³, and Chao Chen^{1,4}

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

² Siemens Healthineers, Malvern, PA, USA

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

⁴ Department of Computer Science, CUNY Queens College, Flushing, NY, USA

Abstract. We study how human brains activate to process input information and execute necessary cognitive tasks. Understanding the process is crucial in improving our diagnostic and treatment of different neurological disorders. Given functional MRI images recorded when human subjects execute tasks with different levels of information uncertainty, we need to identify the similarity and difference between brain activities at different regions of interest (ROIs), and thus gain insights into the underlying mechanism. To achieve this goal, we propose a new ROI-reweight 3D convolutional neural network (CNN). Our CNN not only learns to classify the task-evoked fMRIs with a high accuracy, but also locates crucial ROIs based on a reweight layer. Our findings reveal several brain regions to be crucial in differentiating brain activity patterns facing tasks of different uncertainty levels.

Keywords: uncertainty representation, task-evoked fMRI, CNN

1 Introduction

Cognitive control is a high-order information processing system in human brain which selects appropriate information, inhibits inappropriate response, and coordinates with actions under guidance of context-specific goals and intention [14, 5]. It is implemented by a set of brain regions called the cognitive control network (CCN) (Figure 1(Left)) and also by its interactions with other domain-specific networks (visual, auditory, somatosensory, motor) and the default mode network (DMN) [17, 2, 20]. Characterizing how regions of CCN activate, collaborate, and interact with other networks will improve the theoretical understanding of human brains. Furthermore, knowledge in cognitive control will empower us with advanced diagnosis and treatment of various neurological disorders [6, 1, 4, 19].

In this paper, we investigate the uncertainty representation in cognitive control, i.e., how brains process different levels of uncertainty in cognitive tasks. Early observations, summarized by the Hick-Hyman law [8, 10], state that the brain reaction time grows linearly to the uncertainty level, measured in Shannon entropy [18]. In other words, human brains need to work more to execute

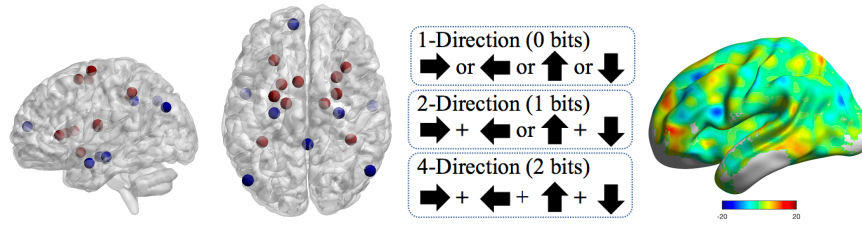


Fig. 1. Left and Middle-Left: The cognitive control network (CCN, red) and the default mode network (DMN, blue). **Middle-Right:** Experiment setting. Human subjects are shown images of arrows and click keys accordingly. The arrows may have one, two, or four possible directions, corresponding to uncertainty levels of zero, one, and two bits measured in Shannon’s entropy [18]. More uncertainty is introduced by additional colors and corresponding actions. In total there are six different tasks, with various uncertainty levels. **Right:** An example of fMRI image mapped from volume to surface.

uncertain tasks. Recently, it has been confirmed that the overall brain activation is linearly correlated to the uncertainty level [5, 22]. However, it remains unknown how different regions coordinate and activate to process uncertainty in tasks. To gain region-specific insights, we analyze task-evoked functional MRIs, in particular, fMRI images collected while a human subject is executing choice reaction time (CRT) tasks. These CRT tasks are the same but with different uncertainty levels, corresponding to different numbers of possible choices. See Figure 1(Middle-Right) and Section 3 for more details.

Many learning methods have been leveraged to analyze task-evoked fMRI images. Examples include analysis of variance (ANOVA) [21], general linear model (GLM) [12], and support vector machine (SVM) [3]. Early studies focus on the association of specific regions of interest and the stimuli, and thus miss the information carried by fine scale brain activity patterns. The multi-voxel pattern analysis (MVPA) [7, 16] approach was the first to take the whole fMRI image as a multivariable input, and use various classifiers to discover basis patterns crucially related to different stimuli.

In this paper, we use a 3D convolutional neural network (CNN) to capture fine scale activity patterns in the task-evoked fMRI. Recent years have witnessed the success of CNN in computer vision and medical image analysis. In particular, 3D CNN has been used in action recognition [11], object recognition [13], etc. In Neuroimage study, 3D CNN has been used to diagnose Alzheimer’s disease (AD) and mild cognitive impairment (MCI) [9], to predict the survival time of brain tumor patients [15], and to reconstruct functional connectivity networks [23]. However, despite its success in achieving the prediction goal, CNN lacks the crucial explainability, namely, the ability to explain the underlying rationale of a prediction. In our problem setting, a standard 3D CNN does not help identify the regions crucial for the differentiation of activity patterns from different cognitive tasks. Unfortunately, this is indeed the primary goal of neuroscientists.

Our contributions. To deliver both prediction power and explainability, we propose a new ROI-reweight CNN. Our method not only classifies fMRIs based

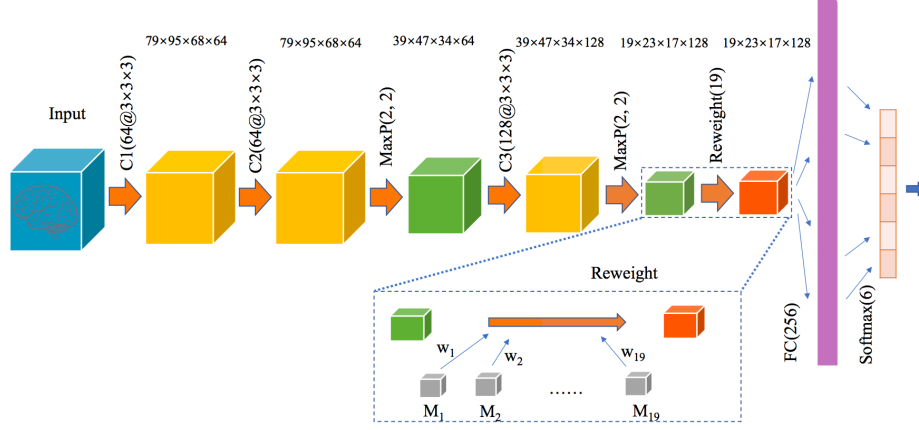


Fig. 2. The architecture of the ROI-reweight 3D CNN model.

CRT tasks, but also learns weights of different ROIs in making the prediction. The idea is to add a novel reweight layer to adjust the high-level representation, i.e., the representation after all convolution and pooling layers. The weight of each element in the high-level representation is determined by nearby ROIs. The ROI reweight layer is detailed in Section 2.2. After training, the learned weights of different ROIs measure how important their nearby patterns are in the classification. To the best of our knowledge, our method is the first CNN-based approach to achieve both classification power and explainability in functional MRI study. Our method identifies several regions in both CCN and DMN, i.e., anterior insula (AI), thalamus (TH), posterior cingulate cortex (PCC), etc., as keys to the differentiation of brain activities for different uncertainty levels.

2 Method

Our method classifies fMRI images into six different classes, corresponding to six different uncertainty levels. Meanwhile, the method learns weights on ROIs, measuring the significance of each ROI in the classifier. We consider 19 ROIs that are considered significant in cognitive control. These ROIs constitute the cognitive control network and the default mode network (Figure 1).

We propose a 3D CNN with a reweight layer before the fully connected layers. The reweight layer adjusts the significance of different elements in the high-level representation. But the reweighting tensor is parametrized by 19 weights, associated to 19 ROIs. During the training, these 19 weights are learned. After training, these weights can be used to measure the significance of ROIs in classification. In Section 2.1, we explain the CNN architecture. Detailed explanation of the reweight layer will be given in Section 2.2.

2. METHOD

2.1 3D CNN Architecture

Our network has three convolutional layers, two max-pooling layers, two fully connected layers, and one reweight layer. See Fig. 2 for an overview of the architecture. The input is a single channel fMRI image of size $75 \times 95 \times 68$. The first two convolutional layers, $C1(64@3 \times 3 \times 3)$ and $C2(64@3 \times 3 \times 3)$, both have 64 channels and kernel size $3 \times 3 \times 3$. The third one, $C3(128@3 \times 3 \times 3)$, has 128 channels and kernel size $3 \times 3 \times 3$. The convolution stride is fixed to 1 voxel. The spatial padding of each convolutional layer input is set to 1, such that the spatial size is preserved after convolution. Two max-pooling layers are applied after the second and the third convolutional layers respectively. Max-pooling is performed over a $2 \times 2 \times 2$ voxel window, with stride 2.

The crucial contribution of our paper is a reweight layer following the last max-pooling layer. This layer has the same input and output size; it reweights all input values using a weight tensor controlled by 19 weight parameters, corresponding to 19 brain ROIs. More details will be given in Section 2.2. The reweight layer is followed by two fully connected layers, with 256 and 6 channels, respectively. The final layer is the soft-max layer. All hidden layers use the rectified linear unit (ReLU) for non-linearity. The model is trained on cross-entropy loss.

2.2 ROI-reweight Layer

The input, I , and output, O , of the reweight layer have the same size, i.e., $19 \times 23 \times 17 \times 128$. The input, I , is the high-level representation produced by the convolutional and max-pooling layers. We can view it as 128-channel signals at $19 \times 23 \times 17$ voxels. The reweight layer uses a $19 \times 23 \times 17$ weight tensor, $\widehat{M}$, to reweight the input. All channels of the (i, j, k) -th voxel of I are multiplied by $\widehat{M}(i, j, k)$. Formally, $O(i, j, k, \ell) = I(i, j, k, \ell) \cdot \widehat{M}(i, j, k)$, $\forall \ell = 1, \dots, 128$.

We parametrize entries of the weight tensor $\widehat{M}$ by weights associated to the ROIs. For generality, we assume R many ROIs. Their weights constitute an R dimensional vector, $\mathbf{w} = [w_1, \dots, w_R]^T$.⁵ The relationship between the $\widehat{M}(i, j, k)$ and the ROI weight w_r depends on how much the r -th ROI affects the (i, j, k) -th entry in the high-level representation. To determine such relationship, we first map the (i, j, k) -th entry back to the original image domain by reversing the convolution and max-pooling operation. Each entry (i, j, k) corresponds to a cube in the original image domain. The coordinates of cube center are denoted by $\mathbf{p}_{(i,j,k)} \in \mathbb{R}^3$. Let $\mathbf{q}_r$ be the center coordinates of the r -th ROI. We use 3D radial basis function (RBF) kernel to define the relationship between the two. Formally,

$$\widehat{M}(i, j, k) = \sum_{r=1}^R w_r \cdot \exp\left(-\frac{\|\mathbf{p}_{(i,j,k)} - \mathbf{q}_r\|^2}{2\sigma^2}\right) = \sum_{r=1}^R w_r \cdot M_r(i, j, k),$$

⁵ We use boldface font for vectors, but normal font for matrices and tensors.

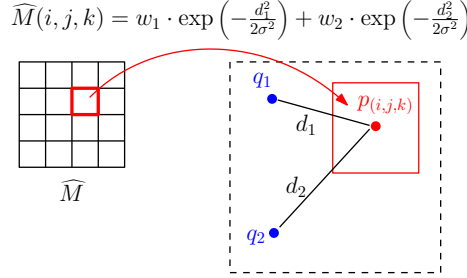


Fig. 3. An entry in $\widehat{M}$ is determined by weights of ROIs by mapping it to the original image domain. d_1 and d_2 are the Euclidean distance between its center and the two ROI centers, q_1 and q_2 .

in which σ is tuned on validation set. In other words, the entry $\widehat{M}(i, j, k)$ is the RBF kernel representation of kernels centered at all ROIs and weighted by $\mathbf{w}$. See Figure 3 for an illustration.

Note that the entries of $\widehat{M}$ depend on the weight vector $\mathbf{w}$ linearly. Formally, we have $\widehat{M} = \sum_{r=1}^R w_r M_r$, in which M_r is a constant tensor of the same size as $\widehat{M}$. It is straightforward to see that if we vectorize the tensor $\widehat{M}$, we have the linear relationship $\widehat{M} = \mathcal{M}\mathbf{w}$, in which $\mathcal{M} = [M_1, \dots, M_R]$ is a matrix with R columns, each corresponds to one vectorized tensor M_r . It is easy to see that the partial derivative of the loss, $\mathcal{L}$, w.r.t. the weight vector $\mathbf{w}$,

$$\frac{\partial \mathcal{L}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial \widehat{M}} \cdot \frac{\partial \widehat{M}}{\partial \mathbf{w}} = \frac{\partial \mathcal{L}}{\partial \widehat{M}} \cdot \mathcal{M}^T$$

in which $\partial \widehat{M} / \partial \mathbf{w} = \mathcal{M}^T$ is the Jacobian matrix. To train this layer, we just need to update $\mathbf{w}$ accordingly at each iteration.

3 Experiments and Discussions

We apply our method to task-evoked fMRI images. These images were collected when human subjects executed choice reaction time (CRT) tasks [22]. Each subject performed around 1100 trials. In each trial, the subjects were presented with an arrow and were supposed to respond accordingly. Depending on the possible directions and colors of the arrows, there are six different tasks with different levels of uncertainty. See Figure 1 for more information. The data was preprocessed using SPM 8. Each gradient-echo planar imaging (EPI) image volume was realigned to the first volume, registered with structural MRI, normalized to the Montreal Neurological Institute (MNI) ICBM152 space, resampled to a voxel size of $2 \times 2 \times 2$ mm, and spatially smoothed. The dimension of each final fMRI image is $79 \times 95 \times 68$. In this study, we focus on 19 ROIs that are considered the most important in cognitive control. These ROIs constitute the two brain networks CCN and DMN (Figure 1).

We train and test our method and other baselines on data from five different human subjects, and report the average accuracy. We randomly reserve 10% of

3. EXPERIMENTS AND DISCUSSIONS

Table 1. The classification results.

Classifier	RF	LG	SVM	3D-CNN	ROI-CNN
Accuracy(%)	34.18	79.68	83.53	87.42	89.04

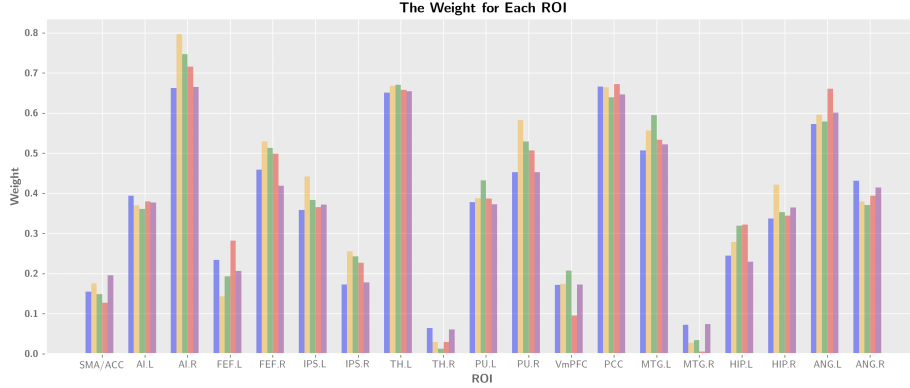


Fig. 4. The learned weights of the 19 ROIs. Bars of 5 different colors correspond to 5 different human subjects. SMA/ACC: supplementary motor area extending to anterior cingulate cortex. AI: anterior insular cortex. FEF: frontal eye find. IPS: area around and along the intraparietal sulcus. TH: thalamus. 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 trials for each subject as the validation set and 10% as the testing set. For each subject individually, all classifiers are trained on the same training set. The best models are selected based on the performances on the validation set.

Random forest (RF), logistic regression (LR), and support vector machine (SVM) are used as baselines. To investigate the effect of the reweight layer, we also apply a 3D CNN without reweight layer (3D-CNN), i.e., our proposed network without the reweight layer, as an ablation. Among all methods, our ROI-reweight CNN (ROI-CNN) achieves the best accuracy. To our surprise, ROI-CNN even outperforms 3D-CNN. We believe the reason is that focusing on the patterns near the 19 crucial ROIs improves the overall performance of the neural network. All results are reported in Table. 1.

Our model learns weights on 19 ROIs. We show all learned weights of the five subjects in Figure 4. We observe high weights on anterior insula (AI), thalamus (TH), posterior cingulate cortex (PCC), etc. This is consistent with the domain knowledge theory in cognitive control [5]. These weights can be used to quantitatively analyze the roles of these ROIs in dealing with uncertain tasks. In Figure 5, we visualize these ROIs in their spatial location, with ball radii proportional to their weights.

We observe that the learned weights are highly consistent across different human subjects even though we trained the models separately. Note that due to cross-subject variation, our model trained on one subject usually overfits

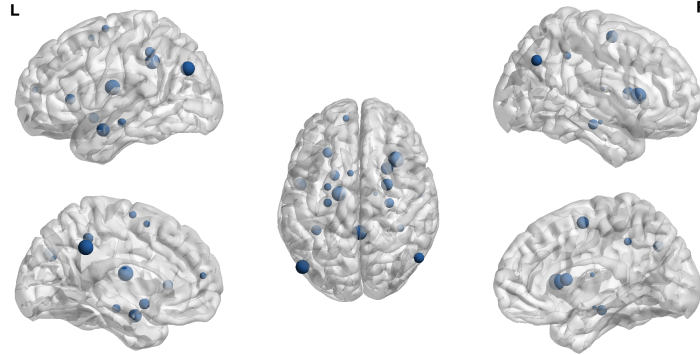


Fig. 5. The 19 ROIs drawn in a brain template. The size of the balls is proportional to the learned weight, which represents the importance of an ROI in the tasks.

and cannot perform well on other subjects (the average accuracy is 19.53%, only slightly better than chance level). In other words, despite the fact that the models overfit on individual subjects and cannot generalize to others, they all learn very similar ROI weights. This shows that our model successfully locates the key regions for uncertainty representation. But the activity patterns within each ROI may vary over different human subjects.

4 Conclusions and Future Work

In this paper, we propose an ROI-reweight 3D convolutional neural network framework to classify the CRT task-evoked fMRI data, and locate key ROIs. Our framework achieves 89.04% average accuracy in the experiments, and outperforms the existing state-of-the-art linear classifiers and the traditional 3D CNN. In the meantime, it also provides quantitative assessment of the significance of the key ROIs in the brain for uncertainty representation, which could benefit cognitive control study. In the future, we plan to extend our framework to be more robust to cross-subject variation.

Acknowledgement. This research is partially supported by the NSF grants IIS 1718802 and CCF 1733866.

References

1. Castellanos, F.X., Sonuga-Barke, E.J., Milham, M.P., Tannock, R.: Characterizing cognition in ADHD: beyond executive dysfunction. *Trends in cognitive sciences* **10**(3), 117–123 (2006)
2. Cole, M.W., Schneider, W.: The cognitive control network: integrated cortical regions with dissociable functions. *Neuroimage* **37**(1), 343–360 (2007)
3. Cox, D.D., Savoy, R.L.: Functional magnetic resonance imaging (fMRI) brain reading: detecting and classifying distributed patterns of fMRI activity in human visual cortex. *Neuroimage* **19**(2), 261–270 (2003)
4. Diamond, A., Barnett, W.S., Thomas, J., Munro, S.: Preschool program improves cognitive control. *Science (New York, NY)* **318**(5855), 1387 (2007)

4. CONCLUSIONS AND FUTURE WORK

5. Fan, J.: An information theory account of cognitive control. *Frontiers in human neuroscience* **8**, 680–680 (2014)
6. Happé, F.: Autism: cognitive deficit or cognitive style? *Trends in cognitive sciences* **3**(6), 216–222 (1999)
7. Haxby, J.V., Gobbini, M.I., Furey, M.L., Ishai, A., Schouten, J.L., Pietrini, P.: Distributed and overlapping representations of faces and objects in ventral temporal cortex. *Science* **293**(5539), 2425–2430 (2001)
8. Hick, W.E.: On the rate of gain of information. *Quarterly Journal of Experimental Psychology* **4**(1), 11–26 (1952)
9. Hosseini-Asl, E., Keynton, R., El-Baz, A.: Alzheimer’s disease diagnostics by adaptation of 3D convolutional network. In: *Image Processing (ICIP), 2016 IEEE International Conference on*. pp. 126–130. IEEE (2016)
10. Hyman, R.: Stimulus information as a determinant of reaction time. *Journal of experimental psychology* **45**(3), 188 (1953)
11. Ji, S., Xu, W., Yang, M., Yu, K.: 3D convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence* **35**(1), 221–231 (2013)
12. Josephs, O., Turner, R., Friston, K.: Event-related fMRI. *Human brain mapping* **5**(4), 243–248 (1997)
13. Maturana, D., Scherer, S.: Voxnet: A 3D convolutional neural network for real-time object recognition. In: *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*. pp. 922–928. IEEE (2015)
14. Miller, E.K.: The prefrontal cortex and cognitive control. *Nature reviews neuroscience* **1**(1), 59–65 (2000)
15. Nie, D., Zhang, H., Adeli, E., Liu, L., Shen, D.: 3D deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 212–220. Springer (2016)
16. 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)
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. Solomon, M., Ozonoff, S.J., Ursu, S., Ravizza, S., Cummings, N., Ly, S., Carter, C.S.: The neural substrates of cognitive control deficits in autism spectrum disorders. *Neuropsychologia* **47**(12), 2515–2526 (2009)
20. Sridharan, D., Levitin, D.J., Menon, V.: A critical role for the right fronto-insular cortex in switching between central-executive and default-mode networks. *Proceedings of the National Academy of Sciences* **105**(34), 12569–12574 (2008)
21. Vuilleumier, P., Armony, J.L., Driver, J., Dolan, R.J.: Effects of attention and emotion on face processing in the human brain: an event-related fMRI study. *Neuron* **30**(3), 829–841 (2001)
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. Zhao, Y., Dong, Q., Zhang, S., Zhang, W., Chen, H., Jiang, X., Guo, L., Hu, X., Han, J., Liu, T.: Automatic recognition of fMRI-derived functional networks using 3D convolutional neural networks. *IEEE Transactions on Biomedical Engineering* (2017)