Published in final edited form as:

Med Image Anal. 2017 December ; 42: 200–211. doi:10.1016/j.media.2017.08.005.

Constructing Fine-granularity Functional Brain Network Atlases via Deep Convolutional Autoencoder

Yu Zhao¹, Qinglin Dong¹, Hanbo Chen¹, Armin Iraj², Yujie Li¹, Milad Makkie¹, Zhifeng Kou², and Tianming Liu¹

¹Cortical Architecture Imaging and Discovery Lab, Department of Computer Science and Bioimaging Research Center, The University of Georgia, Athens, GA

²Department of Biomedical Engineering, Wayne State University, Detroit, MI

Abstract

State-of-the-art functional brain network reconstruction methods such as independent component analysis (ICA) or sparse coding of whole-brain fMRI data can effectively infer many thousands of volumetric brain network maps from a large number of human brains. However, due to the variability of individual brain networks and the large scale of such networks needed for statistically meaningful group-level analysis, it is still a challenging and open problem to derive group-wise common networks as network atlases. Inspired by the superior spatial pattern description ability of the deep convolutional neural networks (CNNs), a novel deep 3D convolutional autoencoder (CAE) network is designed here to extract spatial brain network features effectively, based on which an Apache Spark enabled computational framework is developed for fast clustering of larger number of network maps into fine-granularity atlases. To evaluate this framework, 10 resting state networks (RSNs) were manually labeled from the sparsely decomposed networks of Human Connectome Project (HCP) fMRI data and 5275 network training samples were obtained, in total. Then the deep CAE models are trained by these functional networks' spatial maps, and the learned features are used to refine the original 10 RSNs into 17 network atlases that possess fine-granularity functional network patterns. Interestingly, it turned out that some manually mislabeled outliers in training networks can be corrected by the deep CAE derived features. More importantly, fine granularities of networks can be identified and they reveal unique network patterns specific to different brain task states. By further applying this method to a dataset of mild traumatic brain injury study, it shows that the technique can effectively identify abnormal small networks in brain injury patients in comparison with controls. In general, our work presents a promising deep learning and big data analysis solution for modeling functional connectomes, with fine granularities, based on fMRI data.

Graphical abstract

The proposed computational framework. (a) Sparse representation of fMRI data for generation of functional network maps in each individual. Then 10 RSNs (resting state networks) were manually

---

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

labeled for these networks via sparse coding methods for each task/resting state fMRI scan of each subject. (b) Training 3D CAE (convolutional autoencoder) using all the labeled networks, and extracting features using the trained encoder. (c) Network clustering using the extracted features and generation of fine-granularity ICNs (intrinsic connectivity networks).

Keywords

fMRI; functional brain networks; deep learning

1. INTRODUCTION

Understanding the organizational architecture of human brain function has been of intense interest since the inception of human neuroscience. After decades of active research using in-vivo functional neuroimaging techniques such as fMRI (Heeger and Ress, 2002), there has been mounting evidence (Dosenbach et al., 2006; Duncan, 2010; Fedorenko et al., 2013; Fox et al., 2005; Pessoa et al., 2012) that the human brain function emerges from and is realized by the interaction of multiple concurrent neural processes or networks, each of which is spatially distributed across specific structural substrate of neuroanatomical areas (Bullmore and Sporns, 2009; Huettel, Scott A., Allen W. Song, 2004). Thus, faithful reconstruction and quantitative modeling of those concurrent neural networks from noisy fMRI data has been of a major neuroscientific research topic for years (Bullmore and Sporns, 2009; Dosenbach et al., 2006; Duncan, 2010; Fedorenko et al., 2013; Fox et al., 2005; Huettel, Scott A., Allen W. Song, 2004; Pessoa et al., 2012). Popular brain network reconstruction techniques based on fMRI data include general linear model (GLM) (Friston et al., 1994; Worsley, 1997) for task-based fMRI (tfMRI), independent component analysis (ICA) (Beckmann et al., 2005; Calhoun et al., 2004) for resting state fMRI (rsfMRI), and dictionary learning/sparse representation (Ge et al., 2016; Jiang et al., 2015; Li et al., 2016; Lv et al., 2015a, 2015b, 2015c, 2015d; Xintao Hu et al., 2015; Zhang et al., 2016; Zhao et al., 2016, 2015) for both tfMRI and rsfMRI, all of which can effectively reconstruct concurrent network maps from whole brain fMRI data. For instance, by using the dictionary learning and sparse coding algorithms (Ge et al., 2016; Jiang et al., 2015; Li et al., 2016; Lv et al., 2015a, 2015b, 2015c, 2015d; Mairal et al., 2010; Xintao Hu et al., 2015; Zhang et al., 2016; Zhao et al., 2015), several hundred of concurrent functional brain networks, characterized by both spatial maps and associated temporal time series, can be effectively decomposed from either tfMRI or rsfMRI data of an individual brain. Pooling and

integrating the spatial maps of those functional networks from many brains such as those of Human Connectome Project (HCP) subjects can significantly advance our understanding of the regularity and variability of brain functions across individuals and populations (Lv et al., 2015a, 2015b; Zhao et al., 2016)(Zhao et al., 2016)(Zhao et al., 2016)(Zhao et al., 2016)(Zhao et al., 2016)(Zhao et al., 2016).

Traditionally, there are two mainstream approaches to constructing functional network templates. The first category of approaches used group-level GLM (for tfMRI), group-wise ICA (for rsfMRI), or group-wise sparse representation (for either tfMRI or rsfMRI) to infer group-wise consistent brain networks across individuals as the network templates. The advantage of this category of approaches is that they can effectively map brain networks with good regularity or commonality, but they are short of the description capability of representing the variability of brain networks. The second category of approaches use GLM, ICA or dictionary learning/sparse representation to reconstruct functional networks in individual brains first, followed by spatial warping and averaging steps to finally construct the network templates. The advantage of this category of approaches is that they are straightforward and can account for individual variability, but they heavily depend on the representation features used to identify corresponding components from the individual brain network maps in individuals and to group them into templates. For example, spatial overlap similarity rates are widely used for modeling spatial pattern similarities between different brain networks. This approach provided a straightforward and intuitive solution for small-scale studies. However, the metric of spatial overlap similarity has limited spatial description power and limited robustness to noises and image registration errors. Also, spatial overlap of functional networks is an intrinsic property of functional brain organization (Xu et al., 2016), which is supported by the neuroscientific point of view that cortical microcircuits are rather interdigitated with each other instead of being independent, and it complicates the application of spatial overlap similarity for overlapping networks. Thus, in literature studies, inter-expert visual validation after the spatial overlap similarity measurement is needed to confirm and establish spatial brain networks' correspondences across individual subjects, which renders that the overlap similarity metric is both prone to inter-expert variability and time-consuming. Furthermore, studies with a large dataset and a large quantity of voxels within each network based on spatial overlap similarity metric are very computationally expensive. Though in the literature various efforts have already been made to address this problem, e.g., in (Zhao et al., 2016), a volumetric connectivity map was proposed for feature reduction to represent the spatial pattern of a brain network. However, as mentioned in (Zhao et al., 2016), the volumetric connectivity map has limitations in its spatial pattern description capabilities. Instead of extracting high level features of the spatial distributions of the volumetric maps, the method is based on the shallow model directly projecting the spatial patterns into a uniform sphere, which is sensitive to noisy patterns and incurs loss of spatial information. The 144 generated templates can be found at <http://hafni.cs.uga.edu/autism/> for comparisons with the templates generated using the proposed method in this paper. In short, a major challenge in network template construction for the purpose of modeling the regularity and variability of brain networks is the critical lack of the ability of network volume map descriptors to effectively model the spatial patterns of remarkably variable functional networks across individuals. An effective framework for extracting the

3D volumetric spatial pattern distribution in a higher or deeper level space and in a relatively low dimensionality is much needed, especially for large scale dataset.

Recently, convolutional neural network (CNN) (He et al., 2015; Karpathy et al., 2014; Krizhevsky et al., 2012; Lecun et al., 1998; Nian Liu et al., 2015; Simonyan and Zisserman, 2014) has gained great popularity in the deep learning field. Overall, CNN stemmed from the traditional feed-forward artificial neural networks, and its connections between layers were inspired by the visual cortex of animal brains (De Valois et al., 1982). CNN can preserve the inputs' neighborhood relations and spatial locality in their latent high-level feature representations. Inspired by the successes of CNNs in spatial pattern recognition and representation, we adopt and design 3D deep convolutional structures to deal with the above-mentioned spatial pattern variability of brain networks with a large amount of noises and artifacts. Presumably, these 3D deep convolutional structures can also serve as an effective spatial pattern descriptors of functional brain networks and thus can achieve significant dimension reductions. In this work, network templates generation is a data-driven process of an unsupervised learning by autoencoder, which is one of the most popular unsupervised methods to extract generally useful hierarchical features (Masci et al., 2011). Briefly, in order to extract hierarchical spatial patterns, a 14 layered deep convolutional autoencoder (CAE) is designed and trained to extract features upon manually labeled resting state networks (RSNs) (10 labels for 10 RSNs for each subject's fMRI scan session) derived from our previous HAFNI (holistic atlases of functional networks and interactions) project (Lv et al., 2015a, 2015b). Then an Apache Spark enabled fast parallel computational pipeline is designed to perform fast clustering on the extracted features from the deep CAE to obtain common ICNs (intrinsic connectivity networks) for all, which are then treated as the fine-granularity brain network atlases. In this paper, fine-granularity means that network templates have substantially improved differentiation capability than existing network templates (Lv et al., 2015a, 2015b; Zhao et al., 2016), e.g., one type of brain network template might have multiple variants that are dominant in different cognitive states. Our experimental results showed that the highly abstract features learned by the deep CAE are very effective in spatial pattern description and the heavily dimension-reduced features make clustering on large-scale brain networks very efficient (180 times faster). Intriguingly, our results show that the refined network granularities by deep CAE are associated with different brain task states, thus revealing hierarchical functional organization of the spatial patterns of brain networks. Finally, the application of the proposed CAE framework on mild traumatic brain injury (mTBI) fMRI data showed the discriminative ability of the CAE extracted abstract features among mTBI patients and healthy controls.

2. METHODS AND MATERIAL

2.1 Experimental datasets

The Human Connectome Project (HCP) dataset provided a systematic and comprehensive mapping of connectome-scale functional networks and core nodes over a large population in the literature (Barch et al., 2013). In the Q1 release, 68 subjects were scanned with 7 tasks and 1 resting state for fMRI data acquisition. The imaging parameters are referred to (Barch et al., 2013). FSL FEAT software was adopted for preprocessing steps including skull

removal, motion correction, slice time correction, spatial smoothing, global drift removal (high-pass filtering). Resting state fMRI (rsfMRI) data acquisition and preprocessing steps were similar to tfMRI (Smith et al., 2013). Based on HCP Q1 released tfMRI and rsfMRI datasets, many robust and consistent task-evoked and resting state networks have been generated in our prior project (Lv et al., 2015a, 2015b) via effective matrix decomposition techniques. Thus, our experimental datasets are based on the 10 commonly reconstructed RSNs (Smith et al., 2009) manually labeled in our previous HAFNI project on HCP dataset (Lv et al., 2015a, 2015b). After discarding some problematic cases from the HCP's Q1 release of 68 subjects with 7 tasks and 1 resting state fMRI data, a total number of 5275 manually labeled brain networks with 10 common RSN networks were used for this study. Fig. 1 shows examples of such manually labeled brain networks. Due to the spatial resolution of 2 mm, the initial voxel dimension of one volume map is $91 \times 109 \times 91$. To reduce computation, all functional RSNs maps were downsampled from the resolution of $91 \times 109 \times 91$ to $45 \times 54 \times 45$.

2.2 Computational framework

Our 3D CAE structure is designed for efficient feature extraction and dimension reduction of spatial brain networks. The extracted features by 3D CAE are then utilized in a clustering pipeline to find the corresponding networks and generate fine-granularity ICN atlases. Note that we use the term “RSN” to denote 10 common resting state networks from (Smith et al., 2009), while the term “ICN” denotes the networks obtained using our CAE framework. The complete computational framework is summarized in Fig. 2. Subfigure (a) describes the sparse representation of 5275 manually labeled brain networks with 10 common RSNs as their labels (Lv et al., 2015b). Subfigure (b) illustrates the 3D CAE structure used for feature extraction. The brain networks obtained from (a) are used to train the CAE, comprising an encoder and a decoder. The features are extracted after training the autoencoder, and only the encoder part is used for feature extraction. Subfigure (c) represents the last step to generate the fine-granularity ICNs atlases from the brain networks generated in (a). An Apache Spark based parallel computing platform is developed for a high-performance pairwise feature similarity matrix generation, upon which a spectral clustering (Chen et al., 2013) is performed to derive 17 clusters representing subject-level ICNs. Then 17 ICN atlases are averaged across each ICN cluster's samples shown in subfigure (c). The core components of the 3D CAE model are denoted in the red (b) and yellow (c) subfigures. The model learning is an unsupervised data-driven process. The major reason that we used labeled networks in subfigure (a) is that the initial 10-classes clustered dataset that resembles the 10 well-known RSNs provides a good experimental comparison reference to demonstrate the effectiveness of our framework.

2.3 3D deep CAE structure

CAE is a special type of auto-encoders (AEs) (Bengio et al., 2007). In the traditional fully-connected AEs for 2D image feature extraction tasks, the 2D image spatial structure was typically ignored, thus resulting in not only problems like dealing with realistically varied sized inputs, but also redundancy in the parameters to be trained. AEs also make each extracted feature to be global. In order to discover localized features, the convolutional layer, through which weights are shared among all locations in the previously layers, was

introduced into the AEs such as the 2D CAEs (Masci et al., 2011). 2D CAEs have achieved great successes in computer vision and object recognition fields (Lowe, 1999; Serre et al., 2005). Since the brain network maps are 3D images, we aim to achieve localized hierarchical spatial pattern features, and thus a 3D CAE structure is proposed in our feature extraction framework.

Spatial brain network maps like those Fig. 1 and Fig. 2 in covering distinct cognitive domains can well elucidate neural substrates of brain functions and mental disorders (Sporns, 2013). In this work, a 3D convolutional structure is thus designed to model the complex intrinsic spatial structures of brain networks, and specifically an autoencoder deep learning network structure (Hinton et al., 2006) is utilized to perform the unsupervised learning process. The autoencoder comprises two parts: an encoder and a decoder. The output of the autoencoder is trained to be as similar as its input so that it ensures the feature (output of the encoder part) is well modeled and can thus reconstruct its original input. The detailed 3D CAE structure is illustrated in Fig. 3. We padded zero values at borders to let 3 dimensions of the input to be even ($48 \times 56 \times 48$), such that it can be reconstructed exactly the same size after operations of convolutions. To ensure that the outputs of the encoder, namely features, have the dimension of $6 \times 7 \times 6$ from the input dimension $48 \times 56 \times 48$, which is a significant dimension reduction compared to the input volume, and at the same time to achieve highly abstract hierarchical feature extractions, the 3D CAE is designed as convolutional-pooling/up-sampling alternated 14 layers structure.

In the 3D CAE, the loss functional for optimization is binary cross entropy between predictions and targets:

$$L(\theta) = \sum_{i=1}^m (-\theta^T x_i \cdot \log(x_i) - (1 - \theta^T x_i) \log(1 - x_i)) \quad (1)$$

where m is the number of samples in one batch, θ represents all of the parameters in the network to be estimated, x and is the input volume. The reason that we chose the binary cross entropy instead of mean squared error (MSE) is that MSE would emphasize too much on the incorrect outputs and in each input network, there are quite large background noises. Besides, in the gradient descent training process, binary cross entropy loss function doesn't contain the $\theta^T x (1 - \theta^T x)$ term, which usually result in gradient saturation and slow down the training process.

For training the 3D deep CAE network we used a novel method for gradient descent called ADADELTA (Zeiler, 2012). This method can determine a good learning rate during different training stages by introducing a new dynamic learning rate that is computed on a per-dimension basis using only the first order information. This training method requires no manual tuning of a learning rate and robust to noisy gradient information, model architecture choices, various data modalities, and selection of hyperparameters. Besides, using this method will also benefit from its minimal computation over gradient descent and robustness for large gradient and different architecture choices. The training process for the total number of brain networks (5,275 in this study) took 30 epochs. The experiment was

performed on an 8GB memory GPU (NVidia Quadro M4000), which took 7 hours in total. The reconstructed brain networks well resemble the input spatial network maps, demonstrating the effectiveness of encoded features, which will be explained in detail in the results section.

2.4 Apache Spark based pair-wise similarity parallel computing platform

A polynomial time complexity $O(n^2)$ is required for building a pair-wise similarity matrix for clustering, where the n is the total number of brain networks (5275 in this study). For this experimental work, 5275 is a quite large number for 3D brain networks, letting alone thousands of subjects in future studies. The Apache Spark (<https://spark.apache.org>) leverages the computing resources and memory to employ an efficient parallel/distributed computing, and thus it is employed here for speedup. The objective of this step is to build a similarity matrix upon extracted features by 3D CAE representing each brain network. The similarity of the features is defined below using weighted overlap rate between each pair:

$$\text{overlap rate} = \sum_{k=1}^{|V|} \frac{\min(V_k, W_k)}{(V_k + W_k)/2} \quad (2)$$

where V_k and W_k are the activation scores of voxel k in the feature maps V and W , respectively. $|V|$ is the total number of the voxels in the feature maps, e.g., $6 \times 7 \times 6 = 252$. Accelerated by the Apache Spark properties of large dataset management and low dimensional features extracted, the total time of computing time is only 58.45 seconds in our study. Compared to using the brain network volumes as the input features with our CAE extracted features, it takes 10,707 seconds to finish, which is 180 times slower. Afterwards, a spectral clustering (Luxburg, 2007) procedure is performed on the obtained similarity matrix to iteratively bi-partition the feature clusters. Each cluster represents one ICN cluster, and the fine-granularity ICN atlases are generated as the averaged brain network volume of the samples in each cluster.

3. RESULTS

3.1 CAE derived features and reconstructed brain networks

The CAE model training is an unsupervised process, which will finish successfully when the reconstructed outputs achieve minimum reconstruction loss against the inputs. The loss function between the outputs and inputs are chosen to be binary cross entropy. Meanwhile, minimum loss reconstructions mean well-encoded features. We chose 10 input networks corresponding to 10 common ICNs to visualize the learned encoded features and reconstructed outputs in Fig. 4. As shown in Fig. 4, each pair of the input network and reconstructed output network matched well in spatial patterns. Moreover, each encoded feature of the 10 different input ICNs has distinctive patterns. This result suggests the good feature extraction capability of the proposed 3D CAE.

3.2. 17 fine-granularity ICN atlases via encoded feature clustering

All the encoded features for 5275 brain networks after CAE training have been pooled as inputs to the Apache Spark pair-wise similarity parallel computing framework to generate similarity matrix. After spectral clustering was performed on the similarity matrix, 17 fine-granularity ICN clusters (shown in Fig. 5) were generated from the original 10 RSN clusters. As all 17 ICNs are derived from 10 resting state networks (Fig. 1), the correctness of our results can be more conveniently validated, as follows.

Across clusters, ICNs show different spatial patterns as shown in Fig. 5. Even though some pairs have largely overlapped regions, e.g. ICN 3 and ICN 12, ICN 4 and ICN 5, ICN 10 and ICN 11, ICN 14 and ICN 16, they tend to exhibit different activation patterns, which further confirms that spatial overlap of different functional networks is an intrinsic functional organization property of the human brain (Xu et al., 2016).

Within each cluster, brain networks from each subject appear to be consistent with each other. As an example, ICN 3, 10, and 12 are shown here to demonstrate within-cluster consistency (Fig. 6). This result shows the robustness of our framework in modeling consistent spatial patterns across individual networks with high variability. The within-cluster consistency and between-cluster distinction validated the effectiveness of our deep 3D CAE in extracting useful features and our clustering framework in generating fine-granularity network atlases.

3.3 Manual label correction and fine-granularity networks identification

Our unsupervised data-driven framework generated 17 fine-granularity ICNs. Thereafter, the previous 10 RSNs' manual labels of the input brain networks can be further evaluated by comparison with the 17 ICNs. Except that ICN 15 (shown in Fig. 5) seems to be less interpretable, we believe all other 16 ICNs are functionally meaningful.

Specifically, the ICN 15 only has a cluster size of 22, among which 20 have labels of RSN 9 (refer to Fig. 2) and 2 have labels of RSN 7 (refer to Fig. 2). Indeed, the ICN 15 has relatively high spatial overlap rates with the RSN 9 and RSN 7, and due to the small size of this ICN 15 cluster (22 out of 5275), we can consider that these 22 networks are wrong labels during manual labeling process. Interestingly, our proposed fully-automated CAE framework has effectively modeled them into an independent cluster that is far away from other clusters with large numbers of samples. This ICN 15 cluster demonstrated the promising description capability of our framework.

For the other 16 ICNs, we did not identify any new ICN spatial pattern other than the 10 common RSN templates, which also validated the good performance of the manual labeling process when generating the input networks. However, intriguingly, the 6 extra ICNs generated by our methods showed certain hierarchical sub-patterns of the 10 common RSNs, as shown in Fig. 7. The pair-wise overlap rate similarity of the encoded features of all brain networks and the pair-wise overlap rate similarity of the input networks calculated in equation (2) are shown in Fig. 8(a) and Fig. 8(b), respectively, which are all rearranged in the cluster order from 1 to 17. As we can see from Fig. 8 (b), the overlap rates between ICN 3 and ICN 12, ICN 11 and ICN 10, ICN 14 and ICN 16 are quite high, while their spatial

patterns are still differentiable as shown in Fig. 8 (a). By using feature similarity, we can still cluster them as independent clusters, and indeed they exhibit different functional activation patterns as shown in Fig. 5. These results demonstrated the superior spatial pattern modeling ability of our CAE framework to identify subtle but unneglectable network differences.

3.4 Refined granularities of default mode network in different brain task states

As abovementioned, subtle network variations have been detected, such as for RSN 2, 4, 5, 6. Here, we focus on the analysis on RSN 4, namely, the default mode network (DMN). According to Fig. 9, we can clearly see that the DMN tends to have 3 different variants shown by ICN 3, 10 and 12, and they have quite different count percentages in seven tasks and resting state. Within each DMN variant's ICN cluster, individual brain networks from different tasks contribute differently. For ICN 12, it contains more RSN brain networks than other task networks, which suggests the complete presences of all main 6 regions of DMN are more associated resting states. In contrast, DMN networks during working memory (WM) task tend to have a different DMN variant as shown in ICN 3, with some of the typical DMN regions absent. This result potentially reveals the underlying hierarchical functional network organization of human brains, e.g., DMN regions are hierarchically organized in to sub-variants.

3.5 Application of CAE on mild traumatic brain injury (mTBI) dataset

In this section, application of our 3D CAE on mild traumatic brain injury (mTBI) fMRI dataset was presented to demonstrate CAE's capability of differentiating mTBI patients from healthy controls. There are 24 healthy controls with DTI and rsfMRI scans and 16 mTBI subjects with DTI and rsfMRI scans at both acute and subacute stages, and the detailed data descriptions are as follows. This study was approved by the Wayne State University Human Investigation Committee the data collecting institution. Written informed consent was obtained from each subject before enrollment. Data were collected on a 3-Tesla Siemens Verio scanner with a 32-channel radiofrequency head-only coil. Imaging protocol includes baseline T1, T2, and FLAIR sequences and other advanced MRI sequences including susceptibility weighted imaging (SWI), diffusion imaging and resting state fMRI. Specifically, resting state functional imaging was performed by a gradient echo EPI sequence with the following imaging parameters: TR/TE = 2000/30 ms, slice thickness = 3.5 mm, slice gap = 0.595 mm, pixel spacing size = 3.125×3.125 mm, matrix size = 64×64 , flip angle = 90° , 240 volumes for whole-brain coverage, NEX = 1. During resting state scans, subjects were instructed to keep their eyes closed.

First, all the brain networks (8,000 in total) reconstructed by dictionary learning and sparse coding methods (Lv et al., 2015a, 2015b) from all patients and controls were pooled as inputs into the 3D CAE model that was trained using 5275 brain networks from the HCP dataset as described in Fig. 2. With the well trained ability of the CAE model to describe 3D spatial patterns of brain networks (Fig. 4), the CAE derived features and their reconstructed networks were obtained for all the networks in mTBI and controls. The similarity of the features defined in equation (2) were calculated between 17 ICN templates features and all the mTBI brain networks features to identify 17 ICN corresponding networks using the maximum similarity values with each ICN within each subject. All the 17 ICNs among

randomly selected 4 subjects together with their CAE extracted features and reconstructed networks are shown in Fig. 10 and Fig. 11. It can be seen from Figs. 10–11 that the trained CAE models by HCP data can well represent and reconstruct the spatial features for a new testing dataset of mTBI in this section.

For each common ICN, two-sample t-tests were performed on the feature similarity value vectors between patients both at acute and subacute stages and controls to discover the significant differences with false discovery rate control for multiple comparison correction. Since ICN 15 appears to be less interpretable, here we remove ICN 15 from further discussion and focus on other 16 ICNs. In total, 5 ICNs, including ICN# 1, 7, 8, 13 and 17 with significantly higher (p-value: 0.05) feature similarities in healthy controls than mTBI patients at acute stage were detected (Fig. 12). Notably, ICN 1 is brain stem network, ICN 6 and 7 are visual networks, and ICN 17 is auditory network, which are all reported in the literature with decreased activation among mTBI patients. As in (Gilmore et al., 2016; Iraj et al., 2015; Stevens et al., 2012a), the visual processing circuit, including bilateral visual cortex (ICN 7), showed decreased functional connectivity; in (Iraj et al., 2015; Stevens et al., 2012b), diminished functional connectivity was also found for the motor systems (ICN 13); in (Schairer et al., 2012; Vander Werff, 2012), the mechanisms of injury to the auditory system in the mTBI was introduced and dysfunctions on auditory network were reported (ICN 17). In particular, the brainstem (Kushner et al., 1998), which has cranial nerves for eye movements and the auditory pathway, also showed decreased functional connectivity in mTBI patient group than the control group, relative to the visual and auditory networks. With the recovery of the patient group in the subacute stage, we can see the increased functional connectivity of the mTBI patient group at the subacute stage reflected by increased feature similarities with these 5 corresponding ICNs, but most of them are still not as strong as control group, e.g., only ICN 8 left after the t test showing significant difference with control group. The consistencies of our findings using CAE extracted features upon mTBI dataset with the current literatures cross-validated the effectiveness of our proposed framework.

4. DISCUSSION AND CONCLUSIONS

Reconstructions of connectome-scale concurrent functional brain networks have been enabled by public datasets such as HCP and powerful techniques such as sparse representation. However, the major challenge lies in the lack of effective and efficient brain network descriptors that can quantitatively model such large-scale brain networks, while accommodating the individual variabilities, noises, and image registration errors. In this work, a fully unsupervised 3D deep CAE framework was proposed to address this problem, and extensive experimental results have demonstrated the effectiveness and efficiency of our framework. Further application into a dataset of mTBI study demonstrates its capability to differentiate the abnormal brain networks in mTBI patients after injury. In general, this promising framework provides a novel approach for fMRI big data analysis and a better understanding of human brain functions. Our framework can be improved in the following aspects in the future. First, although our experiments used 5275 manually labeled functional networks from 68 subjects of HCP Q1 release, which is reasonably large at this moment, we will use all of the recently released HCP 900 subjects' data to train and evaluate our models

in the near future. Second, in addition to the used 10 RSNs in this paper, we will consider extending our 3D CAE models to include many other possible RSNs and task-evoked networks reconstructed from resting state fMRI data and task-based fMRI data. In this way, our 3D CAE model will be able to quantitatively describe a majority of connectome-scale functional network maps, which can be further applied for a variety of cognitive and clinical neuroscience problems based on fMRI data.

Acknowledgments

T Liu was partially supported by National Institutes of Health (DA033393, AG042599) and National Science Foundation (IIS-1149260, CBET-1302089, BCS-1439051 and DBI-1564736). Z Kou was partially supported by National Institutes of Health (R21NS090153). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

References

- Barch DM, Burgess GC, Harms MP, Petersen SE, Schlaggar BL, Corbetta M, Glasser MF, Curtiss S, Dixit S, Feldt C, Nolan D, Bryant E, Hartley T, Footer O, Bjork JM, Poldrack R, Smith S, Johansen-Berg H, Snyder AZ, Van Essen DC, WU-Minn HCP Consortium. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage*. 2013; 80:169–89. DOI: 10.1016/j.neuroimage.2013.05.033 [PubMed: 23684877]
- Beckmann CF, DeLuca M, Devlin JT, Smith SM. Investigations into resting-state connectivity using independent component analysis. *Philos Trans R Soc Lond B Biol Sci*. 2005; 360:1001–13. DOI: 10.1098/rstb.2005.1634 [PubMed: 16087444]
- Bengio Y, Bengio Y, Lamblin P, Popovici D, Larochelle H. Greedy layer-wise training of deep networks. *NIPS*. 2007
- Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci*. 2009; 10:186–98. DOI: 10.1038/nrn2575 [PubMed: 19190637]
- Calhoun VD, Pekar JJ, Pearson GD. Alcohol Intoxication Effects on Simulated Driving: Exploring Alcohol-Dose Effects on Brain Activation Using Functional MRI. *Neuropsychopharmacology*. 2004; 29:2097–2017. DOI: 10.1038/sj.npp.1300543 [PubMed: 15316570]
- Chen H, Li K, Zhu D, Jiang X, Yuan Y, Lv P, Zhang T, Guo L, Shen D, Liu T. Inferring group-wise consistent multimodal brain networks via multi-view spectral clustering. *IEEE Trans Med Imaging*. 2013; 32:1576–86. DOI: 10.1109/TMI.2013.2259248 [PubMed: 23661312]
- De Valois RL, William Yund E, Hepler N. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res*. 1982; 22:531–544. DOI: 10.1016/0042-6989(82)90112-2 [PubMed: 7112953]
- Dosenbach NUF, Visscher KM, Palmer ED, Miezin FM, Wenger KK, Kang HC, Burgund ED, Grimes AL, Schlaggar BL, Petersen SE. A core system for the implementation of task sets. *Neuron*. 2006; 50:799–812. DOI: 10.1016/j.neuron.2006.04.031 [PubMed: 16731517]
- Duncan J. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn Sci*. 2010; 14:172–9. DOI: 10.1016/j.tics.2010.01.004 [PubMed: 20171926]
- Fedorenko E, Duncan J, Kanwisher N. Broad domain generality in focal regions of frontal and parietal cortex. *Proc Natl Acad Sci U S A*. 2013; 110:16616–21. DOI: 10.1073/pnas.1315235110 [PubMed: 24062451]
- Fox MD, Snyder AZ, Vincent JL, Corbetta M, Van Essen DC, Raichle ME. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc Natl Acad Sci U S A*. 2005; 102:9673–8. DOI: 10.1073/pnas.0504136102 [PubMed: 15976020]
- Friston KJ, Holmes AP, Worsley KJ, Poline JP, Frith CD, Frackowiak RSJ. Statistical parametric maps in functional imaging: A general linear approach. *Hum Brain Mapp*. 1994; 2:189–210. DOI: 10.1002/hbm.460020402

- Ge B, Makkie M, Wang J, Zhao S, Jiang X, Li X, Lv J, Zhang S, Zhang W, Han J, Guo L, Liu T. Signal sampling for efficient sparse representation of resting state fMRI data. *Brain Imaging Behav.* 2016; 10:1206–1222. DOI: 10.1007/s11682-015-9487-0 [PubMed: 26646924]
- Gilmore CS, Camchong J, Davenport ND, Nelson NW, Kardon RH, Lim KO, Sponheim SR. Deficits in Visual System Functional Connectivity after Blast-Related Mild TBI are Associated with Injury Severity and Executive Dysfunction. *Brain Behav.* 2016; 6:e00454.doi: 10.1002/brb3.454 [PubMed: 27257516]
- He K, Zhang X, Ren S, Sun J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. 2015
- Heeger DJ, Ress D. What does fMRI tell us about neuronal activity? *Nat Rev Neurosci.* 2002; 3:142–51. DOI: 10.1038/nrn730 [PubMed: 11836522]
- Hinton GE, Osindero S, Teh YW. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* 2006; 18:1527–1554. DOI: 10.1162/neco.2006.18.7.1527 [PubMed: 16764513]
- Huettel, Scott A., Song, Allen W., M, G. Functional magnetic resonance imaging. Sunderland: Sinauer Associates; 2004.
- Iraji A, Benson RR, Welch RD, O'Neil BJ, Woodard JL, Ayaz SI, Kulek A, Mika V, Medado P, Soltanian-Zadeh H, Liu T, Haacke EM, Kou Z. Resting State Functional Connectivity in Mild Traumatic Brain Injury at the Acute Stage: Independent Component and Seed-Based Analyses. *J Neurotrauma.* 2015; 32:1031–1045. DOI: 10.1089/neu.2014.3610 [PubMed: 25285363]
- Jiang X, Li X, Lv J, Zhang T, Zhang S, Guo L, Liu T. Sparse representation of HC *P* grayordinate data reveals novel functional architecture of cerebral cortex. *Hum Brain Mapp.* 2015; 36:5301–5319. DOI: 10.1002/hbm.23013 [PubMed: 26466353]
- Karpathy A, Toderici G, Shetty S, Leung T, Sukthankar R, Fei-Fei L. Large-scale Video Classification with Convolutional Neural Networks. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2014:1725–1732.
- Krizhevsky A, Sutskever I, Hinton GE. ImageNet Classification with Deep Convolutional Neural Networks. *NIPS.* 2012:1097–1105.
- Kushner D, et al. Mild Traumatic Brain Injury. *Arch Intern Med.* 1998; 158:1617.doi: 10.1001/archinte.158.15.1617 [PubMed: 9701095]
- Lecun Y, Eon Bottou L, Bengio Y, Haaner P. Gradient-Based Learning Applied to Document Recognition RS-SVM Reduced-set support vector method. SDNN Space displacement neural network. SVM Support vector method. TDNN Time delay neural network. V-SVM Virtual support vector method. *PROC IEEE.* 1998
- Li, X., Makkie, M., Lin, B., Sedigh Fazli, M., Davidson, I., Ye, J., Liu, T., Quinn, S. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD '16. ACM Press; New York, New York, USA: 2016. Scalable Fast Rank-1 Dictionary Learning for fMRI Big Data Analysis; p. 511-519.
- Lowe, DG. Proceedings of the Seventh IEEE International Conference on Computer Vision. Vol. 2. IEEE; 1999. Object recognition from local scale-invariant features; p. 1150-1157.
- Luxburg U. A tutorial on spectral clustering. *Stat Comput.* 2007; 17:395–416. DOI: 10.1007/s11222-007-9033-z
- Lv J, Jiang X, Li X, Zhu D, Chen H, Zhang T, Zhang S, Hu X, Han J, Huang H, Zhang J, Guo L, Liu T. Sparse representation of whole-brain fMRI signals for identification of functional networks. *Med Image Anal.* 2015a; 20:112–34. DOI: 10.1016/j.media.2014.10.011 [PubMed: 25476415]
- Lv J, Jiang X, Li X, Zhu D, Zhang S, Zhao S, Chen H, Zhang T, Hu X, Han J, Ye J, Guo L, Liu T. Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. *IEEE Trans Biomed Eng.* 2015b; 62:1120–31. DOI: 10.1109/TBME.2014.2369495 [PubMed: 25420254]
- Lv J, Jiang X, Li X, Zhu D, Zhao S, Zhang T, Hu X, Han J, Guo L, Li Z, Coles C, Hu X, Liu T. Assessing effects of prenatal alcohol exposure using group-wise sparse representation of fMRI data. *Psychiatry Res.* 2015c; 233:254–68. DOI: 10.1016/j.psychres.2015.07.012 [PubMed: 26195294]

- Lv, J., Lin, B., Zhang, W., Jiang, X., Hu, X., Han, J., Guo, L., Ye, J., Liu, T. Modeling Task FMRI Data via Supervised Stochastic Coordinate Coding. Springer International Publishing; 2015d. p. 239-246.
- Mairal J, Bach F, Ponce J, Sapiro G. Online Learning for Matrix Factorization and Sparse Coding. *J Mach Learn Res*. 2010; 11:19–60.
- Masci, J., Meier, U., Cirean, D., Schmidhuber, J. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. Springer; Berlin, Heidelberg: 2011. p. 52-59.
- Nian, Liu, Han, J., Zhang, D., Shifeng, Wen, Liu, T. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE; 2015. Predicting eye fixations using convolutional neural networks; p. 362-370.
- Pessoa L, Amaral D, Behnia H, Kelly J, Averbeck BB, Seo M, Bressler SL, Menon V, Damaraju E, Huang YM, Barrett LF, Pessoa L, Honey CJ, Kotter R, Breakspear M, Sporns O, Modha DS, Singh R, Newman M, Padmala S, Pessoa L, Pessoa L, Pessoa L, Adolphs R, Pessoa L, Petrovich GD, Canteras NS, Swanson LW, Swanson LW, Thompson E, Varela FJ. Beyond brain regions: Network perspective of cognition–emotion interactions. *Behav Brain Sci*. 2012; 35:158–159. DOI: 10.1017/S0140525X11001567 [PubMed: 22617666]
- Schairer KS, et al. Mild Traumatic Brain Injury and Associated Effects on the Auditory System. *Perspect Hear Hear Disord Res Diagnostics*. 2012; 16:18.doi: 10.1044/hhd16.1.18
- Serre, T., Wolf, L., Poggio, T. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE; 2005. Object Recognition with Features Inspired by Visual Cortex; p. 994-1000.
- Simonyan K, Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition. 2014
- Smith SM, Beckmann CF, Andersson J, Auerbach EJ, Bijsterbosch J, Douaud G, Duff E, Feinberg DA, Griffanti L, Harms MP, Kelly M, Laumann T, Miller KL, Moeller S, Petersen S, Power J, Salimi-Khorshidi G, Snyder AZ, Vu AT, Woolrich MW, Xu J, Yacoub E, Urbil K, Van Essen DC, Glasser MF. Resting-state fMRI in the Human Connectome Project. *Neuroimage*. 2013; 80:144–168. DOI: 10.1016/j.neuroimage.2013.05.039 [PubMed: 23702415]
- Smith SM, Fox PT, Miller KL, Glahn DC, Fox PM, Mackay CE, Filippini N, Watkins KE, Toro R, Laird AR, Beckmann CF. Correspondence of the brain's functional architecture during activation and rest. *Proc Natl Acad Sci U S A*. 2009; 106:13040–5. DOI: 10.1073/pnas.0905267106 [PubMed: 19620724]
- Sporns O. Structure and function of complex brain networks. *Dialogues Clin Neurosci*. 2013; 15:247–62. [PubMed: 24174898]
- Stevens MC, Lovejoy D, Kim J, Oakes H, Kureshi I, Witt ST. Multiple resting state network functional connectivity abnormalities in mild traumatic brain injury. *Brain Imaging Behav*. 2012a; 6:293–318. DOI: 10.1007/s11682-012-9157-4 [PubMed: 22555821]
- Stevens MC, Lovejoy D, Kim J, Oakes H, Kureshi I, Witt ST. Multiple resting state network functional connectivity abnormalities in mild traumatic brain injury. *Brain Imaging Behav*. 2012b; 6:293–318. DOI: 10.1007/s11682-012-9157-4 [PubMed: 22555821]
- Vander Werff KR. Auditory Dysfunction Among Long-Term Consequences of Mild Traumatic Brain Injury (mTBI). *Perspect Hear Hear Disord Res Diagnostics*. 2012; 16:3.doi: 10.1044/hhd16.1.3
- Worsley KJ. An overview and some new developments in the statistical analysis of PET and fMRI data. *Hum Brain Mapp*. 1997; 5:254–258. DOI: 10.1002/(SICI)1097-0193(1997)5:4<254::AID-HBM9>3.0.CO;2-2 [PubMed: 20408225]
- Hu, Xintao, Lv, Cheng, Cheng, Gong, Lv, Jinglei, Guo, Lei, Han, Junwei, Liu, Tianming. Sparsity-Constrained fMRI Decoding of Visual Saliency in Naturalistic Video Streams. *IEEE Trans Auton Ment Dev*. 2015; 7:65–75. DOI: 10.1109/TAMD.2015.2409835
- Xu J, Potenza MN, Calhoun VD, Zhang R, Yip SW, Wall JT, Pearson GD, Worhunsky PD, Garrison KA, Moran JM. Large-scale functional network overlap is a general property of brain functional organization: Reconciling inconsistent fMRI findings from general-linear-model-based analyses. *Neurosci Biobehav Rev*. 2016; 71:83–100. DOI: 10.1016/j.neubiorev.2016.08.035 [PubMed: 27592153]
- Zeiler MD. ADADELTA: An Adaptive Learning Rate Method. 2012

- Zhang S, Li X, Lv J, Jiang X, Guo L, Liu T. Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations. *Brain Imaging Behav.* 2016; 10:21–32. DOI: 10.1007/s11682-015-9359-7 [PubMed: 25732072]
- Zhao S, Han J, Lv J, Jiang X, Hu X, Zhao Y, Ge B, Guo L, Liu T. Supervised Dictionary Learning for Inferring Concurrent Brain Networks. *IEEE Trans Med Imaging.* 2015; 34:2036–2045. DOI: 10.1109/TMI.2015.2418734 [PubMed: 25838519]
- Zhao Y, Chen H, Li Y, Lv J, Jiang X, Ge F, Zhang T, Zhang S, Ge B, Lyu C, Zhao S, Han J, Guo L, Liu T. Connectome-scale group-wise consistent resting-state network analysis in autism spectrum disorder. *NeuroImage Clin.* 2016; 12:23–33. DOI: 10.1016/j.nicl.2016.06.004 [PubMed: 27358766]

Highlights

- A new deep 3D convolutional autoencoder to model brain network maps.
- Derived fine-granularity functional brain network atlases.
- Revealed unique network patterns specific to different brain task states using HCP fMRI data.
- Effectively identified abnormal small networks in brain injury patients in comparison with controls.

Fig. 1. Visualization of 10 RSNs templates (top panel) and examples of manually labeled networks (bottom panels, each row representing an individual brain).

Fig. 2.

The proposed computational framework. (a) Sparse representation of fMRI data for generation of functional network maps in each individual. Then 10 RSNs were manually labeled for these networks via the methods in (Lv et al., 2015b) for each task/resting state fMRI scan of each subject. (b) Training 3D CAE using all the labeled networks, and extracting features using the trained encoder. (c) Network clustering using the extracted features and generation of fine-granularity ICNs (17 in total).

Fig. 3.

3D CAE structure with detailed dimensions of each layer (channels, x, y, z). We padded zeros at borders of the input volume, so the $45 \times 54 \times 45$ volume is turned into $48 \times 56 \times 48$. The obtained feature is a $6 \times 7 \times 6$ cube.

Fig. 4.
Extracted features and reconstructed networks of 10 input networks corresponding to common RSNs. Each panel shows one RSN.

Fig. 5.

17 ICN atlases, together of their corresponding CAE extracted features, generated as the averages of the individual brain network maps within each cluster. The numbers of brain network maps in each cluster are also denoted at the top of each panel of ICN.

Fig. 6.
Spatial pattern consistency within clusters. Arrows with the same color denote consistent spatial activation patterns.

Fig. 7.
Hierarchical sub-patterns of the 10 common RSN networks identified. The bottom panel showed individual network consistencies within each ICN cluster using DMN sub pattern ICNs.

Fig. 8.

(a) Clustered brain network feature similarity matrix; (b) clustered brain network overlap rate similarity matrix.

Fig. 9. DMN tends to have different sub-variants during different brain task states including the seven HCP tasks and resting state as shown in the figure. The count percentage distributions of ICN3, ICN10, and ICN12 in seven tasks and the resting state are provided. Significant differences in the network distributions are highlighted in red. WM: Working Memory.

Fig. 10.

Extracted features and reconstructed networks of mTBI data brain networks corresponding to common ICNs. Each panel shows one ICN on 4 individual brains as examples. Only ICN 3, 7 and 17 are shown here, please refer to Fig. 11 for all 17 ICNs.

Fig. 11.

CAE extracted features and reconstructed networks of mTBI data's brain networks corresponding to 17 ICNs. 4 individual networks were randomly selected for visualizations for each ICN.

Fig. 12.

Significant differences of average feature similarity values with ICNs between control and both patient acute (ICN 1, 7, 8, 13, 17) and patient subacute (ICN 8) groups. Error bar is one standard deviation.