

DEEP LEARNING FRAMEWORKS FOR FUNCTIONAL AND STRUCTURAL
MEDICAL IMAGE ANALYSIS

by

YU ZHAO

(Under the Direction of Tianming Liu)

ABSTRACT

Medical image analysis plays an important role for understanding both human psychology and physiology. Human body functional and structural information can be recorded using different medical imaging techniques. As for functional image analysis, understanding the organizational architecture of human brain function has been of intense interest since the inception of human neuroscience. In-vivo Functional Magnetic Resonance Imaging (fMRI) technology enabled the investigation of the human brain functional mechanism and the decomposition of the brain functional network components. However, it is a great challenge to conduct simultaneous spatial-temporal decomposition analysis due to the 4D nature of the fMRI data. Existing methods, i.e., general linear model, independent component analysis, sparse representation, have been proposed for purely functional network decomposition on either spatial or temporal feature of fMRI data. Spatial functional network analysis after fMRI decomposition is still an open question. The major challenge is the lack of ability to effectively describe spatial volume maps of brain networks exposed to extensive individual variability. Besides, the 4D nature of fMRI data has not been fully investigated due to the traditional methodology limitation. To address

the abovementioned challenges, we proposed a series of deep learning frameworks for spatial functional network map descriptor and 4D spatial temporal analysis of fMRI data. Applications include both task-evoked fMRI, resting state fMRI and also brain disease fMRI data. As for structural image analysis, magnetic resonance (MR) images and computed tomography (CT) are both primary structural imaging modalities for solving and analyzing various medical problems. Understanding of human body structure is advanced by medical imaging technologies, while medical images analysis is advanced by the development of new digital analysis techniques from computer science community. To provide better solutions for some existing challenging medical problems, we proposed deep learning frameworks to facilitate cross modality synthesis and landmark detection tasks, both are widely used and eagerly needed in healthcare industry.

INDEX WORDS: functional magnetic resonance imaging analysis, structural medical imaging analysis, spatial temporal analysis, cross modality analysis, deep learning

DEEP LEARNING FRAMEWORKS FOR FUNCTIONAL AND STRUCTURAL
MEDICAL IMAGE ANALYSIS

by

YU ZHAO

B.E., Huazhong University of Science and Technology, China, 2013

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2018

© 2018

YU ZHAO

All Rights Reserved

DEEP LEARNING FRAMEWORKS FOR FUNCTIONAL AND STRUTURAL
MEDICAL IMAGE ANALYSIS

by

YU ZHAO

Major Professor:	Tianming Liu
Committee:	Suchendra M. Bhandarkar Shannon Quinn

Electronic Version Approved:

Suzanne Barbour
Dean of the Graduate School
The University of Georgia
December 2018

DEDICATION

This dissertation is dedicated to my loving parents for their great spiritual and financial support. This dissertation is also dedicated to my girlfriend Jiangyi Chen for her encouragement and company at my darkest time during my Ph. D. study.

ACKNOWLEDGEMENTS

This dissertation represents not only the Ph. D. pursuing in cortical architecture and imaging discovery (CAID) laboratory at UGA computer science department, but also the composition of every detail of the wonderful life experience here. Everything started with meeting with Dr. Tianming Liu, my academic supervisor, who offered me the great opportunity to get started with the Ph. D. life at UGA, and all the following academic achievements and excitements step by step. It has truly been a treasure for me to be part of the enthusiastic and supportive research team under the mentorship of Dr. Liu over years. This dissertation is the recognition and witness of the remark guidance from Dr. Liu and the result of 5 and half year's endurance endorsed by Dr. Liu. Therefore, I would like to express my most sincere gratitude to my academic and life mentor, Dr. Liu, for all his efforts on me to finally bring me here, completing the Ph. D. study with this dissertation. Moreover, I cannot make all the achievements without collaborations and supports from my colleagues who also always stay as the front-line researchers in CAID laboratory. Lab life with all my colleagues has been nothing but wonderful.

I would like to express my special thanks to my Ph. D. committee: Dr. Suchendra M. Bhandarkar, and Dr. Shannon Quinn. Their expertise in computer vision, big data analysis has always been an invaluable contribution to my achievements in combining deep learning with medical image analysis. Their inspiring and insightful advice to my projects truly helped me get rid of the difficulties met during research. Keeping advising me in patient and innovative way for years really inspired and moved me a lot.

Specifically, I would like to express my gratitude to my mentors during internship at Siemens Healthineers. They truly transferred their expertises in both academic and industrial field to me selflessly. They are Dr. Yiqiang Zhan, Dr. Xiang Sean Zhou, Dr. Shu Liao, Dr. Yimo Guo, Dr. Zhennan Yan, Dr. Liang Zhao, Dr. Gerardo Hermosillo. My internships with my mentors truly brought me insights into how industrial field are solving medical image analysis problems to benefit patients. I believe that their invaluable mentorship will finally lead me to invaluable contribution to the society.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	x
LIST OF FIGURES	xi
CHAPTER	
1 INTRODUCTION AND LITERATURE REVIEW	1
1.1 Dissertation Statement	3
1.2 Dissertation Outline	4
1.3 Contributions.....	5
2 SPATIAL FUNCTIONAL ANALYSES FOR BRAIN FMRI IMAGES	8
2.1 Abstract.....	9
2.2 Connectivity map: Connectome-scale Group-wise Consistent Resting- state Network Analysis in Autism Spectrum Disorder	10
2.3 A Novel Framework for Groupwise Registration of fMRI Images based on Common Functional Networks	33
2.4 Constructing Fine-granularity Functional Brain Network Atlases via Deep Convolutional Autoencoder.....	42
2.5 Automatic Recognition of fMRI-derived Functional Networks using 3D Convolutional Neural Networks	61

2.6	Automatic Recognition of Holistic Functional Brain Networks Using Iteratively Optimized Convolutional Neural Networks (IO-CNN) with Weak Label Initialization.....	82
2.7	3D Deep Convolutional Neural Network Revealed the Value of Brain Network Overlap in Differentiating Autism Spectrum Disorder from Healthy Controls	109
3	TEMPORAL FUNCTIONAL ANALYSES FOR BRAIN FMRI IMAGES	122
3.1	Abstract	123
3.2	Template-guided Functional Network Identification via Supervised Dictionary Learning	124
3.3	Modeling Task fMRI Data via Deep Convolutional Autoencoder	135
4	SPATIO-TEMPORAL FUNCTIONAL ANALYSIS FOR BRAIN FMRI IMAGES	122
4.1	Modeling 4D fMRI Data via Spatio-Temporal Convolutional Neural Networks (ST-CNN).....	144
5	STRUCTURAL ANALYSES: CROSS IMAGING MODALITY SYNTHESIS FOR MR-ONLY RADIOTHERAPY TREATMENT PLANNING	156
5.1	Towards MR-Only Radiotherapy Treatment Planning: Synthetic CT Generation Using Multi-view Deep Convolutional Neural Networks	157
6	STRUCTURAL ANALYSES: LANDMARK DETECTION USING REINFORCEMENT LEARNING FRAMEWORK.....	169
6.1	Introduction of Reinforcement Learning (RL)	169
6.2	MR Landmark Detection using RL.....	172

7 DISCUSSION AND CONCLUSION	176
REFERENCES	179

LIST OF TABLES

	Page
Table 2.1. RSNS NUMBERS OF TRAINING SETS AND TESTING SETS FOR 3D CNN TRAINING AND TESTING.....	66
Table 2.2. CNN corrections of original wrong labels for each of the 10 RSN types.	73
Table 2.3. Algorithm for IO-CNN	92
Table 2.4. Accuracy, sensitivity and specificity in ASD/control classification.....	119
Table 3.1. Number of subjects (out of 78) in each of the normal control (NC) and ASD patient group with the identified functional networks based on the 5 templates. Results from S-r1DL are listed in the top two rows, results from unsupervised r1DL are listed in the bottom two rows.....	133
Table 4.1. Performance of ST-CNN measured by spatial overlap rate	152
Table 5.1. MAE values comparison for 7 purely testing subjects	167

LIST OF FIGURES

	Page
Fig. 1.1. Dissertation outline.....	4
Fig. 1.2. Functional analysis. fMRI data have been investigated at individual, group level in a hierarchical way. Deep learning and big data frameworks have been deployed for both spatial, temporal and spatio-temporal analysis on both healthy control and brain disease data.	5
Fig. 1.3. Towards MR-only radiotherapy treatment planning: cross modality synthesis from MR to CT using multi-view multi-channel deep learning framework.	6
Fig. 1.4. MR landmark detection using deep reinforcement learning.	6
Fig. 2.1. Flow chart of the proposed computational pipeline.	16
Fig. 2.2. Illustration of network components generated by sparse representation.....	18
Fig. 2.3. Examples of similarity measurement. <i>Connectivity map</i> similarity (CMS) and overlap rate similarity (ORS) between each pair of selected SR components were listed below each subfigure. The similar parts were highlighted by red circles and dissimilar parts were highlighted by green arrows.	19
Fig. 2.4. Illustration of connectivity map for volume image. (a) 2-D spatial map; (b) projection of spatial map in (a) to unit circle; (c) 3 selected projection centers represented by red dots; (d) 3-D spatial map of a functional network; (e) projection of spatial map in (c) to unit sphere; (f) connectivity map for (c).....	22

Fig. 2.5. 8 of 128 common ICNs consistently identified in both autism patients and healthy controls.....26

Fig. 2.6. Templates of ICNs with significantly higher (p-value: 0.025) ORS in ASD patients. All the networks were sorted by p-values. Templates of ICNs with significant higher (pFDR-value) ORS in controls left after multiple comparison correction are listed inside the red box.26

Fig. 2.7. Templates of ICNs with significantly higher (p-value: 0.025) ORS in healthy controls. All the networks were sorted by p-values. Templates of ICNs with significant higher (pFDR-value) ORS in patients left after multiple comparison correction are listed inside the red box.27

Fig. 2.8. (a) Spatial map of network 7. (b) ORS between the template of network 7 and the corresponding SR components in each individual’s brain. (c) ICNs with decreased temporal interactions to network 7. (d) ICNs with increased temporal interactions to network 7.29

Fig. 2.9. Group-wise comparisons of inter-network interactions. (a) Differences in average temporal interactions between ASD patients and healthy controls (TCs). (b) Differences in average spatial interactions between ASD patients and TCs. (c) Significantly increased or decreased interactions identified by one-tailed t-test (p-value: 0.025). Color bars were listed at the bottom of subfigures. The matrices in the second and third rows were arranged by clusters. In each subfigure, starting from top left, the 1st magenta box is the ICNs with weak connections. Following it, the rest of the boxes are clusters from #1 to #14 respectively.30

Fig. 2.10. Label map stacking procedure.....37

Fig. 2.11. The functional ICNs guided groupwise registration framework. (1). Initial registration from subject space to a standard space using traditional strategy; (2). Congealing groupwise registration of label map images; (3). Warp fMRI images from individual space to common space; (4). Group ICA: obtain common ICNs; (5). Obtain new individualized label map images based on common ICNs.37

Fig. 2.12. Averaged ICN templates using proposed method and traditional method.39

Fig. 2.13. hippocampal ICN templates for NC and AD groups.....39

Fig. 2.14. Averaged DMN templates of NC and AD groups, using proposed framework and two traditional strategies41

Fig. 2.15. Decreased DMN activity in AD. Red circles showed the significant decreases in bilateral inferior–lateral–parietal cortex; Yellow circles denotes more significantly decreased connectivity in hypo-area of precuneus and posterior cingulate; Green arrows showed more significantly decreased activity in hippocampal area; Blue arrows denotes more significantly decreased activity in ventromedial frontal cortex.

Fig. 2.16. 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).49

Fig. 2.17. 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.....	51
Fig. 2.18. Extracted features and reconstructed networks of 10 input networks corresponding to common RSNs. Each panel shows one RSN.....	54
Fig. 2.19. 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.	55
Fig. 2.20. Spatial pattern consistency within clusters. Arrows with the same color denote consistent spatial activation patterns.....	56
Fig. 2.21. 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.	58
Fig. 2.22. (a) Clustered brain network feature similarity matrix; (b) clustered brain network overlap rate similarity matrix.....	58
Fig. 2.23. 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.....	60
Fig. 2.24. Overview of the computational steps in the 3D CNN framework.	66

Fig. 2.25. Overview of the 3D CNN structure. Input map and output of each feature extraction layer are visualized. C1 represents convolutional layer 1, which contains 32 types of kernels or filters of size 5 with a stride step 2; C2 represents convolutional layer 2, which contains 32 types of kernels of filters of size 3 with a stride step 1; P represents pooling layer with pooling kernel size of 2; FC represents fully connected layer, with 128 nodes in this layer; Output layer contains 10 nodes representing each class of the 10 RSNs labels.67

Fig. 2.26. Visualization of the trained RSNs-specific filters for convolutional layer 1. This layer contains 32 different types of filters, each of which has a size of 5*5*5*32, with entire filters and clipped filters separated by a dashed line in the middle of each panel.....69

Fig. 2.27. Visualization of the trained RSNs-specific filters for convolutional layer 2. This layer also contains 32 different types of filters, each of which has a size of 3*3*3*32 (32 input channels for this layers). Here, only the filters for the first channel of all the 32 are shown, with entire filters and clipped filters separated by a dashed line in the middle in each panel.69

Fig. 2.28. Extracted feature maps using trained CNN layers with DMN as an input example.

70

Fig. 2.29. Visualization of 10 common RSNs templates derived in the HAFNI project. For each RSN, 9 most informative slices are displayed.73

Fig. 2.30. Representative cases of CNN classification errors for 10 RSNs. CNN’s predicted labels (denoted as CNN label in the figure) appear to be more reasonable than the original manual labels by experts.75

Fig. 2.31. 3D visualizations of input map 1 in figure 8 with RSN templates overlaid.

Subfigure (1a), (1b) and (1c) are input maps (blue regions) with RSN1 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Subfigure (2a), (2b) and (2c) are input maps (blue regions) with RSN2 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively.75

Fig. 2.32. Representative cases of overlap-based classification results for 10 RSNs.....77

Fig. 2.33. 3D visualizations of input map 2 in figure 10 with RSN templates overlaid.

Subfigure (1a), (1b) and (1c) are input maps (blue regions) with overlap-based classification of RSN2 template overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Subfigure (2a), (2b) and (2c) are input maps (blue regions) with RSN3 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. RSN3 should be the category for this input map, but overlap-based method gave the wrong result as RSN2.77

Fig. 2.34. 3D visualizations of input map 4 in figure 10 with RSN templates overlaid.

Subfigure (1a), (1b) and (1c) are input maps (blue regions) with overlap-based classification of RSN4 template overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Subfigure (2a), (2b) and (2c) are input maps (blue regions) with RSN8 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Due to that the major activation regions reside in the prefrontal lobes, RSN8 should be the category for this input map, but overlap-based method gave the wrong result.....77

Fig. 2.35. 3 RSN classification discrepancies between CNN and overlap-based method in common prediction disagreement with the original manual labels.79

Fig. 2.36. Illustration of CNN and overlap-based method’s performance when differentiating input RSN maps of high spatial overlaps.....80

Fig. 2.37. 3D CNN structure used for training and classification. Bold numbers indicated the feature map sizes (e.g. 48, 64, 48); Red numbers indicated convolutional kernel size (e.g. $3 \times 3 \times 3$) and nodes number of fully connected layers (e.g. 128, 144); Green numbers indicated the channel size of each feature map.89

Fig. 2.38. Logic connection between the atlases generation based on ABIDE I data in (Zhao et al., 2016) and the functional network recognition work based on these atlases using the ABIDE II data in this paper. Both ADIBE I and II fMRI datasets were decomposed to functional networks using sparse coding, where networks from ABIDE I were utilized for generating atlases, while the functional networks from ABIDE II were decomposed for atlases-based recognition via a weak labeling process based on spatial overlap rate.90

Fig. 2.39. IO-CNN framework. (a). weak label initialization process based on spatial overlap rate; (b). iteratively optimized CNN (IO-CNN) training process. The training process will iterate over the 219,800 input 3D maps starting with the initial weak labels. Only the 3D maps with nonzero labels will be taken as input for the current training iteration. After each training iteration, the trained model will predict on all the 219,800 training samples for new labels, which will be used to replace and update the previous training labels for the next training iteration.

Meanwhile, the label differences between the previous training labels and the new predicted labels will be recorded for iteration termination condition check.91

Fig. 2.40. Label change procedure during each iteration. During each iteration, initially assigned labels of some networks will be optimized to other labels during the CNN training iteration.....94

Fig. 2.41. Initial training set’s 144 functional network label distributions.....95

Fig. 2.42. IO-CNN training accuracy and loss curves from iteration 0 to 13. Iteration 0 is the weak label initialization process, where IO-CNN is not trained yet.....96

Fig. 2.43. Sample size’s dynamic changes and label change percentages during IO-CNN training. Iteration 0 is the weak label initialization process, where IO-CNN is not trained yet. Initial labels (80,293 initial training samples) were first used in iteration 1 for IO-CNN training.96

Fig. 2.44. Initial label distribution and final label distribution after iterative IO-CNN optimization.97

Fig. 2.45. (c). 3 label predictions with the least data distribution using 5 instances each. Each row contains the atlas picture at the first column, with the rest 5 columns as prediction instances from individual networks.100

Fig. 2.46. Previously unlabeled network assigned with label 7 and 14. The overlap rate with those networks are all below 0.2 yielding no label initially.101

Fig. 2.47. (a). New instances with initial labels 3, 47 assigned with final label 7; (b). New instances with initial labels 4, 24 assigned with final label 14. First row of pictures indicates input networks; second row indicates final predicted atlases; third row indicates initial overlap rate assigned labels.102

Fig. 2.48. (a). 3 axis 3D visualization for new instances with initial labels 3, 47 assigned with final label 7; (b). 3 axis 3D visualization for new instances with initial labels 4, 24 assigned with final label 14.....103

Fig. 2.49. (a). Label shifted from label 7 to label 5, 8; (b). Label shifted from label 14 to label 4, 57. First row of pictures indicates input networks; second row indicates final predicted atlases; third row indicates initial overlap rate assigned104

Fig. 2.50. (a). 3 axis 3D visualization for instances with label shifted from label 7 to label 5, 8; (b). 3 axis 3D visualization for instances with label shifted from label 14 to label 4, 57.....105

Fig. 2.51. Proposed 135-class predictions on previous 10-class labelled HCP testing set networks reveals fine-granularities of the variants of 10 RSNs. The radar charts show the 135 templates granular percentages out of the 10 RSN atlases.107

Fig. 2.52. Fine-granularities among 135 atlases and previous ambiguous labels using RSN 1 as an example.....108

Fig. 2.53. Overlap map illustration using one subject’s ICNs. Blue color and yellow color denote ICN 1 and ICN 2 volume map respectively, while red denotes the overlap map.114

Fig. 2.54. Overlap map feature selections. Blue and yellow maps in each panel are the ICN maps, while red maps are the overlap maps of each pair of ICN maps. (a). Common overlap maps indicate both ASD patient and control groups have these types of overlaps among top 50 patterns. (b). Control-dominant overlap maps (c). ASD-dominant overlap maps.115

Fig. 2.55. Computational framework for ASD classification.116

Fig. 2.56. Deep 3D CNN structure for feature extraction and classification. Dimensions of each layer using different colors (green: channel, black: 3D dimension).....117

Fig. 2.57. Losses and accuracies during 10-fold training process.118

Fig. 2.58. Control-dominant network overlap patterns involving default mode network (ICN 7) and fusiform gyrus (ICN 144).120

Fig. 3.1. (a) Illustration of the algorithm pipeline of the S-r1DL model with the visualization of the spatial maps of one sample template network (marked as “TN”) and the result network defined in v1. (b) Illustration of the alternative strategy based on the unsupervised r1DL model. The same template network is visualized as “TN”. The identified network based on spatial overlap method is defined in vn and also visualized.131

Fig. 3.2. (a) Visualization of the 5 template networks (from top to bottom): Fusiform, IFG, ACC, PCC and DMN. (b) Visualization of the functional networks identified by the S-r1DL method based on the 5 templates.132

Fig. 3.3. Visualization of the (a) 23 functional networks identified from normal control, and (b) 34 functional networks from ASD patient subjects based on the Fusiform Gyrus template.133

Fig. 3.4. Illustration of the DCAE structure with its two validation studies. (a) DCAE structure, the hidden states, and the feature maps, as well as their relationship with the validation studies. (b) The detailed processes of max pooling and unpooling.

138

Fig. 3.5. Comparison of the performance in reconstructing the fMRI signals by different methods. (a) The Pearson correlation coefficients between the reconstructed signals

by DCAE and the original signals; (b) The Pearson correlation coefficients between the reconstructed signals by SL-SDL and the original signals.	139
Fig. 3.6. The comparison of HL-SDL and SLSDL in detecting task related patterns. GLM results are shown here as the reference of comparison.....	141
Fig. 3.7. Exemplar patterns that are only found in HL-SDL. Avg_pc denote the average correlation between the group of hypothetical models and time series patterns.	141
Fig. 4.1. (a). Algorithmic pipeline of ST-CNN; (b). Spatial network structure, temporal network structure, and the combination of the spatial and temporal domain.	148
Fig. 4.2. Training losses (y-axis) versus training steps (x-axis). (a). first stage spatial network training loss; (b). second stage temporal network training loss; (c). fine-tuning training loss.....	150
Fig. 4.3. Examples of comparisons between ST-CNN outputs and ground-truth from sparse representation. Here we showed 2 subjects' comparison results from two different datasets (1 HCP Q1 subjects and 1 HCP S900 subjects). Spatial maps are very similar and time series have Pearson correlation coefficient values 0.878 in HCP Q1 data, and 0.744 in HCP S900 data. Red curves are ground-truth. Blue curves are ST-CNN temporal outputs.	151
Fig. 4.4. Example of the better DMN identification of ST-CNN than sparse representation (denoted by red arrows). The temporal dynamics of the two networks are also different, where output from sST-CNN (blue) are more reasonable.	152
Fig. 4.5. Example of ST-CNN outputs and ground-truth (sparse representation) for EMOTION task. Spatial maps are very similar and time series have Pearson	

correlation 0.754. Red curve is ground-truth, blue curve is the temporal output by ST-CNN.153

Fig. 4.6. Comparison of the spatial maps between ST-CNN results and supervised sparse representation which takes temporal output of ST-CNN as pre-defined atoms.....154

Fig. 5.1. Multi-channel multi-view U-Net based deep fully convolutional network framework.160

Fig. 5.2. Single view 2D U-Net (sagittal view and axial view). Multi-channel 2D MR slices (In-phase (IP) and Out-of-phase(OOP)) are network inputs. Loss is designed as maxpooling and hinge-like Huber loss.161

Fig. 5.3. (a) Maxpooling hinge-like Huber loss function for U-Net structure training. (b) Hinge-like Huber loss function.163

Fig. 5.4. A schematic explanation of the impact of mis-registration to intensity transformation. (a) Image1 (Modality 1), (b) Perfectly registered Image modality 2, (c) Image 2 with rigid mis-alignment, (d) Image 2 with non-rigid mis-alignment. Triangles in (a)-(d) represent the same object. Dashed lines in (c) and (d) denote the locations of the perfectly registered Image2.163

Fig. 5.5. Combination Network. Single view networks include 2D U-Net structures for both axial and sagittal view. 3D combination network takes 32 channels output 3D features from 2 single view networks as input and output a 3D synthetic CT volume.

164

Fig. 5.6. Improved quality of the bone area synthesis compared to benchmark U-Net schemes. (a) Ground-truth sCT used for training; (b) Bench mark result using

original U-Net; (c) Result using benchmark U-Net with maxpooling function; (d)	
Proposed result.....	166
Fig. 5.7. Removed 2D slice stitching blurring effects (red arrow) by combining multi-view	
U-Net and improved image quality compared to results from a shrunk size 3D	
benchmark network.....	166
Fig. 5.8. Improved synthetic quality compared to the ground-truth CTs in 3 different views.	
Each column is a comparison among the input MR, ground-truth image and	
proposed predictions.	167
Fig. 6.1. The typical framing of a Reinforcement Learning (RL) scenario: an agent takes	
actions in an environment, which is interpreted into a reward and a representation	
of the state, which are fed back into the agent. (Figure source: Wikipedia).....	170
Fig. 6.2. Reinforcement learning workflow.....	173
Fig. 6.3. Agent: deep learning network.....	175
Fig. 6.4. Exemplar output coordinate of patella.....	175

CHAPTER 1

INTRODUCTION AND LITERATURE REVIEW

Medical images are essential to both theoretical researches and practical diagnoses & treatments advanced by computer science technology for computing analysis. Typically, medical images are measured from human body on different scales and different imaging modalities (e.g. magnetic resonance imaging (MR) and computed tomography (CT)) for different physical properties (e.g. brain functional activity, body structures and radiodensities). Therefore, medical images play an important role in initiatives to improve understanding of human functional and structural organization and diagnostic treatments for public health. With considerably increasing availability of medical equipment and number of imaging-based procedure (both research and practical), facilitated by advancing computing technologies, especially artificial intelligence guided (AI-guided) smart analysis, medical image analysis is desperately demanding and awarding for both researches and clinical practices.

Functional images are mainly generated from functional imaging technologies based on time-of-flight methods, such as Electroencephalography (EEG), Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET) and so on. Functional imaging is mostly referred as the study of human brain function using the abovementioned imaging techniques. There are diverse and multitudinous applications of functional imaging. For example, it can be used to analyze basic brain anatomy and physiology. All systems of the human brain have

been investigated applying functional imaging techniques (Frackowiak, 2004; Gazzaniga, 2004). A lot of brain diseases (e.g. Parkinson's Diseases (PD), Alzheimer's Disease(AD), Autism Spectral Disorder(ASD), Movement Disorders (MD)) and injuries (Traumatic Brain Injury (TBI)) causing brain functional abnormalities have also been widely investigated (Assaf et al., 2010; McAllister et al., 2001; M. J. McKeown et al., 1998; Politis et al., 2010; Ward et al., 2003). Moreover, functional imaging can also serve as intervention for clinical use. For example, pre-surgical mapping (Matthews et al., 2006) using fMRI to localize cerebral functions in tissues is a great application established using functional imaging.

As opposed to functional imaging, structural imaging centers on recording static images at a specific time point. Structure medical images can be obtained from much wider imaging modalities, and thus widely used for clinically diagnostic and treatment practices. Conventional medical image analysis is mainly referred as structural analysis, whose methods can be grouped into image segmentation (Pham et al., 2000), image registration (Maintz and Viergever, 1998), image reconstruction (Zeng, 2010), image detection and so on (Litjens et al., 2017; Shen et al., 2017). In terms of these interdisciplinary category, computer vision, information engineering, math and medicine areas are intersected to develop computational and mathematical solutions pertaining to medical images for medical researches and clinical practices (Duncan and Ayache, 2000). Especially due to the fast-emerging artificial intelligence techniques enabled by deep learning which intrinsically fit the conventional medical image analysis, advanced massive industrial medical solutions based on structural medical images are developed recently for computer aided diagnosis and treatment.

1.1 Dissertation Statement

During my doctoral research program in Cortical Architecture and Imaging Discovery (CAID) laboratory, together with my collaborators, a series of projects have been conducted regarding to discoveries in human brain functional and structural organizations. Most methodologies developed from CAID lab are based on MR images, including structural MR and fMRI. Our designed frameworks cover human brain imaging data including both healthy and patient subjects, as well as animal brain imaging data. Most of my contributions focus on functional brain imaging analysis.

Part of my dissertation is fulfilled during my internships at Siemens Healthineers. Siemens Healthineers contributed large amount and high quality of medical solutions for imaging diagnostics and treatments. Combined technologies from areas of computer vision, statistical analysis, mathematics and medicine were experimented and deployed in their medical solution products. As part of the team serving as research student, most of my structural medical image analysis researches are carried out here. Experience here is more industrial related, while experience in CAID is more academic research oriented. However, similar methodologies are applied in both industrial and academic, both functional and structural medical image researches, which provides me with deeper insight into medical image analysis field.

As artificial intelligence (AI) is greatly promoted by the success of deep learning technologies, AI healthcare is attracting tremendous interests. Deep learning technologies are promisingly applied to different fields of medical image analysis. Inspired by the success of deep learning, my doctoral dissertation contributes to both functional and structural medical image analysis by developing deep-learning-based frameworks. My

works, represented in the dissertation, mainly focus on human brain fMRI data (including both healthy and patient subjects), and human body structural MR data.

1.2 Dissertation Outline

This dissertation contains 7 chapters as shown in Fig. 1.1, which can be split into 2 categories: functional medical image analysis and structural medical image analysis. Chapter 2, 3, 4 covers functional analysis topics done during studies within CAID lab including spatial fMRI analysis, temporal fMRI analysis and spatio-temporal analysis respectively. Chapter 5, 6 covers structural analysis topics done during internships with Siemens Healthineers including cross modality synthesis (MR to CT) and MR image landmark detection respectively. The whole dissertation concludes with chapter 7.

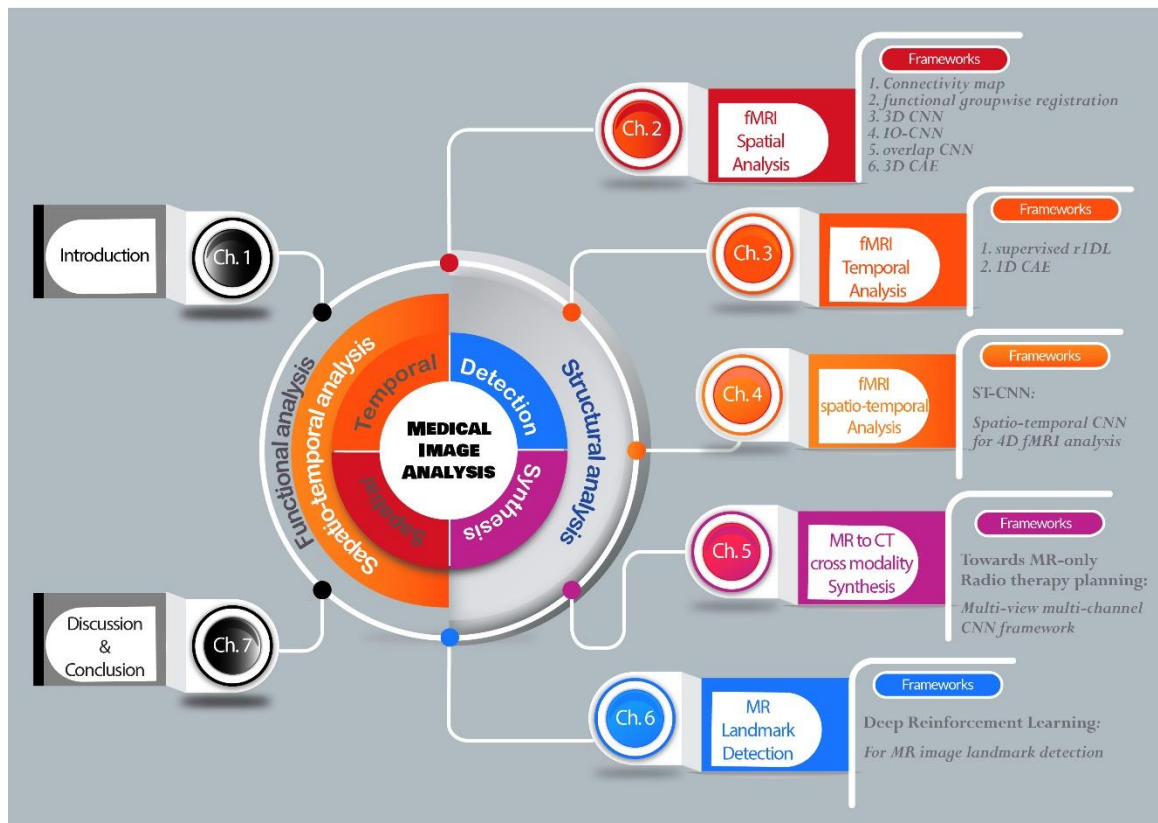


Fig. 1.1. Dissertation outline.

1.3 Contributions

This dissertation contributes to both functional and structural medical image analysis fields. The major image modality researched is MR, including fMRI and structural MR.

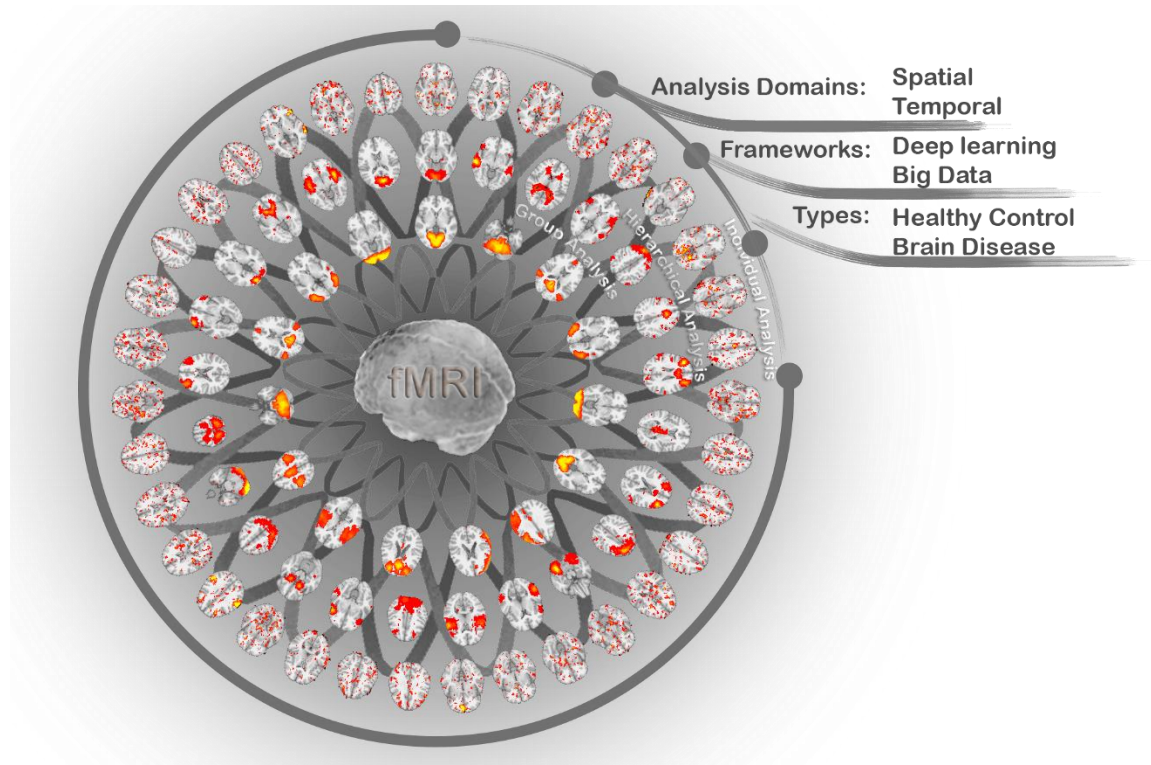


Fig. 1.2. Functional analysis. fMRI data have been investigated at individual, group level in a hierarchical way. Deep learning and big data frameworks have been deployed for both spatial, temporal and spatio-temporal analysis on both healthy control and brain disease data.

For functional part, the major contribution is combined spatial (Ren et al., 2017; Y. Zhao et al., 2018a, 2018b, Zhao et al., 2017d, 2017b, 2017a, 2016) and temporal analysis (Huang et al., 2017; Zhao et al., 2017c) for 4D fMRI data to a spatio-temporal analysis (Y. Zhao et al., 2018c) by years of researches on both directions (Fig. 1.2). This research

progress witnessed how we converted from using traditional frameworks to deep learning enabled frameworks (Zhao et al., 2016). As for group level analysis (Zhao et al., 2017d), the increasing amount of available data becomes bottlenecks of various researches by challenging the ability of existing frameworks to process big data (Zhao et al., 2016). Thus, big data analysis is also investigated as part of the proposed frameworks. In order to understand the cognitive processes within human brain, both healthy control and brain disease data have been cross-investigated. Significant findings on the human brain functions have been achieved.

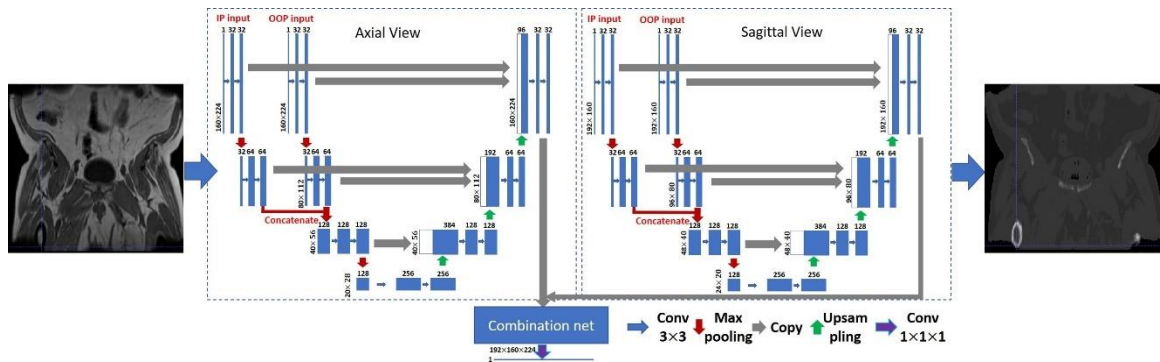


Fig. 1.3. Towards MR-only radiotherapy treatment planning: cross modality synthesis from MR to CT using multi-view multi-channel deep learning framework.

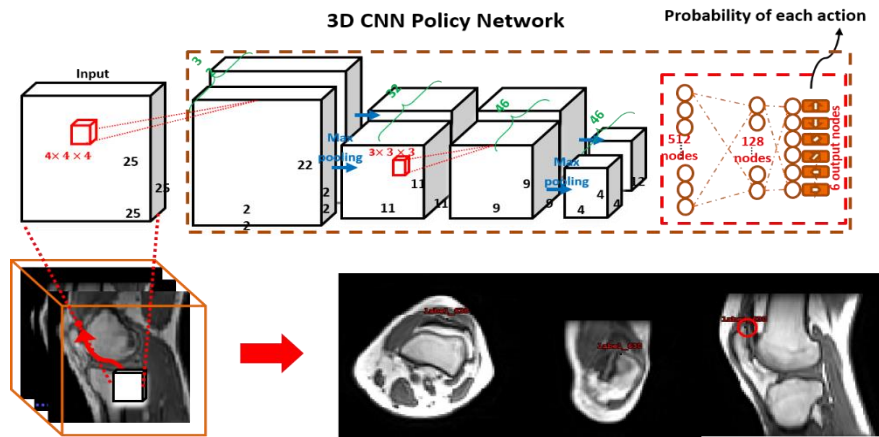


Fig. 1.4. MR landmark detection using deep reinforcement learning.

For structural part, the projects are more industrial field related for healthcare. Two major contributions have been achieved. The first one aims to solve the MR-only radiotherapy treatment planning problem with a multi-view multi-channel deep learning framework (Fig. 1.3) (Y. Zhao et al., 2018d). As radiotherapy treatment planning involves both MR and CT images for target delineation and electron density info for dosage calculation, a cross modality deformable registration process is needed between MR and CT, which is very time consuming and inaccurate. Besides, the economical and time costs are much higher as for both MR and CT scans. MR-only radiotherapy will greatly advance the whole workflow by providing effective, efficient and economical solutions. The second one aims to solve the landmark detection problem on MR images using a deep reinforcement learning framework (Fig. 1.4). Detected landmarks are desired for numerous diagnostic and clinical tasks. A fast and accurate landmark detection framework was proposed to better fulfill landmark detection tasks.

CHAPTER 2

SPATIAL FUNCTIONAL ANALYSES FOR BRAIN FMRI IMAGES^{1,2,3,4,5,6,7}

¹Connectome-scale Group-wise Consistent Resting-state Network Analysis in Autism Spectrum Disorder
Yu Zhao*, Hanbo Chen*, Yujie Li, Jinglei Lv, Xi Jiang, Fangfei Ge, Tuo Zhang, Shu Zhang, Bao Ge,
Cheng Lyu, Shijie Zhao, Junwei Han, Lei Guo, Tianming Liu. *NeuroImage: Clinical*, 2016. vol.12, pp 23-
33.

*Co-first author.

Reprinted here with the permission of publisher.

² 3D Deep Convolutional Neural Network Revealed the Value of Brain Network Overlap in Differentiating
Autism Spectrum Disorder from Healthy Controls

Yu Zhao, Fangfei Ge, Tianming Liu. Accepted by MICCAI, 2018.

Reprinted here with the permission of publisher.

³ Constructing fine-granularity functional brain network atlases via deep convolutional autoencoder

Yu Zhao, Qinglin Dong, Hanbo Chen, Armin Irabi, Yujie Li, Milad Makkie, Zhifeng Kou, Tianming Liu.
Medical Image Analysis, 2017. vol. 42, pp 200-211

Reprinted here with the permission of publisher.

⁴ 3D Functional Brain Network Classification using Convolutional Neural Networks

Yu Zhao*, Dehua Ren*, Hanbo Chen, Qinglin Dong, Jinglei Lv, Tianming Liu. Accepted by ISBI, 2017.

*Co-first author.

Reprinted here with the permission of publisher.

⁵ Automatic Recognition of fMRI-derived Functional Networks using 3D Convolutional Neural Networks

Yu Zhao, Qinglin Dong, Shu Zhang, Wei Zhang, Hanbo Chen, Xi Jiang, Lei Guo, Xintao Hu, Junwei Han,
Tianming Liu. *IEEE TBME*, 2017. vol.pp(99), pp 1-1.

Reprinted here with the permission of publisher.

⁶ Automatic Recognition of Holistic Functional Brain Networks Using Iteratively Optimized Convolutional
Neural Networks (IO-CNN) with Weak Label Initialization

Yu Zhao, Fangfei Ge, Tianming Liu. *Med Image Anal.* 2018 Jul; 47:111-126.

Reprinted here with the permission of publisher.

⁷ 3D Deep Convolutional Neural Network Revealed the Value of Brain Network Overlap in Differentiating
Autism Spectrum Disorder from Healthy Controls

Yu Zhao, Fangfei Ge, Tianming Liu. Accepted by MICCAI, 2018.

Reprinted here with the permission of publisher.

2.1 Abstract

Current fMRI data modeling techniques such as Independent Component Analysis (ICA) (Beckmann and Smith, 2004; De Martino et al., 2007; Martin J McKeown et al., 1998; McKeown et al., 2003; Rolfe et al., 2009) and Sparse Coding methods (Jiang et al., 2015; Lv et al., 2015a, 2015d, 2015c, 2013a; Zhang et al., 2016; Zhao et al., 2015) can effectively reconstruct dozens or hundreds of concurrent interacting functional brain networks simultaneously from the whole brain fMRI signals. Due to significant individual variabilities, brain diseases, imaging protocols and noise sources, the functional brain networks tend to be very different across individuals among populations. Pooling and integrating the spatial maps of those functional networks from many brains can significantly advance our understanding of the regularity and variability of brain functions across individuals and populations (Zhao et al., 2016). Accurate registration also plays a critical role in group-wise functional Magnetic Resonance Imaging (fMRI) image analysis as spatial correspondence among different brain images is a prerequisite for inferring meaningful patterns. After several years of attempts at dealing with abovementioned challenges when integrating, pooling and comparing spatial network maps across individuals and populations, we realized that the major challenge is the lack of ability to effectively describe spatial volume maps of brain networks. We have proposed frameworks for functional network registration (section 2.3), functional network spatial map descriptors for the analyses (section 2.2, 2.4, 2.5, 2.6, 2.7) correspondingly. From traditional frameworks, the limitations have been recognized for describing spatial maps. Thus in the afterward proposed frameworks, deep learning based frameworks were utilized to tackle the challenges, which achieved great success. All proposed frameworks together not only

deepened the understanding of human brain functional organizations, but also discovered some important finding on understanding brain diseases such as Alzheimer’s Disease (AD) and Autism Spectrum Disorder (ASD) (section 2.2, 2.3, 2.7).

2.2 Connectivity map: Connectome-scale Group-wise Consistent Resting-state Network Analysis in Autism Spectrum Disorder

Abstract

This section discusses our first attempt to propose an effective and efficient spatial brain map descriptor, namely, connectivity map, before the massive applications of the deep learning frameworks, for integrating, pooling and comparing spatial network maps across individuals and populations. By applying connectivity map to Autism Spectrum Disorder (ASD) brain functional map analysis, we examined and evaluated the efficiency and effectiveness of this proposed spatial map descriptor. New findings and discoveries using the descriptor motivated us to integrate the spatial analysis system of brain fMRI data. Limitations of this descriptor also inspired us to dive deeply into “descriptors with deep structures” after this research.

Understanding the organizational architecture of human brain function and its alteration patterns in diseased brains such as ASD patients are of great interests. In-vivo Functional Magnetic Resonance Imaging (fMRI) offers a unique window to investigate the mechanism of brain function and to identify functional network components of the human brain. Previously, we have noticed that multiple concurrent functional networks can be derived from fMRI signals using whole-brain sparse representation. Yet it is still an open question to derive group-wise consistent networks featured in ASD patients and controls.

Here we proposed an effective volumetric network descriptor, named connectivity map, to compactly describe spatial patterns of brain network maps and implemented a fast framework in Apache Spark environment that can effectively identify group-wise consistent networks in big fMRI dataset. Our experiment results identified 144 group-wisely common Intrinsic connectivity networks (ICNs) shared between ASD patients and healthy control subjects, where some ICNs are substantially different between the two groups. Moreover, further analysis on the functional connectivity and spatial overlap between these 144 common ICNs reveals connectomics signatures characterizing ASD patients and controls. In particular, the computing time of our Spark-enabled functional connectomics framework is significantly reduced from 240 hours (C++ code, single core) to 20 hours, exhibiting a great potential to handle fMRI big data in the future.

Introduction

Seven decades after initially discovered and reported by Kanner (Kanner, 1943) in the United States and Asperger (Asperger, 1944) in Austria, Autism Spectrum Disorder (ASD) has drawn enormous attention from society and the research community due to the large number of affected children and the complicated pathology of the disease. ASD generally develops at an early age. Children diagnosed with ASD usually have difficulty in social interaction and communication, as well as repetitive patterns of behavior, interests, or activities. According to a recent report, ASD affects 1 out of 68 children aged 8 years in the United States (Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators and Centers for Disease Control and Prevention (CDC), 2014).

While years of research have shown ASD to be a highly genetic disorder, an understanding of the corresponding phenotype, i.e., the organizational architecture of human brain function and its alteration patterns in ASD patients' brains, are of great significance to yield new insights on the causes of the disease and to discover potential treatments. In-vivo Functional Magnetic Resonance Imaging (fMRI) offers a unique window to investigate the mechanism of brain function and to identify functional network components of the human brain (Friston, 2009; Heeger and Ress, 2002; Li et al., 2009; Logothetis, 2008). By analyzing brain activity during different tasks using fMRI, the neurology logics of ASD patients' symptoms can be studied. For instance, atypical activation patterns in Fusiform Face Area during face recognition tasks suggests functional abnormalities during face processing; decreased activation in the left Inferior Frontal Gyrus and increased activation in the Planum and Temporale regions during language tasks indicates that instead of integrating the meanings of individual words into a coherent conceptual structure, ASD subjects pay more attention to the meanings of individual words (Stigler et al., 2011). In addition to the abnormal brain activations during tasks, decreased functional connectivity in the default mode network (DMN) were also identified when the brain is in a resting state (Kennedy and Courchesne, 2008).

However, most of the previous ASD fMRI studies focused on individual or a relatively small number of brain regions or networks. mounting evidence has shown that the human brain functions are realized via the interactions of multiple concurrent neural networks, each of which is spatially distributed across specific structural substrates of neuroanatomical areas (Dosenbach et al., 2006; Duncan, 2010; Fedorenko et al., 2013; Fox et al., 2005; Pessoa, 2012), and even some currently available studies involving whole-

brain scale ASD related functional connectivities are examined (Kana et al., 2014)(Moseley et al., 2015), however, there still seems to lack a group-wise consistently study across whole brain scale regions on large-scale subjects. The major challenge to elucidating the mechanism of how functions are altered in ASD patients across the entire brain lies in the enormous difficulty in determining the corresponding brain regions of interest (ROIs) in different brains given the remarkable variability in cortical structures and functions across populations (Liu, 2011). The added complications are the unclear cytoarchitectural boundaries between cortical regions and the dramatic changes in functional brain connectivity due to non-linear properties of the brain (Zhu et al., 2013), making the correspondence establishment between different brains even more challenging. The traditional clustering methods and Independent Component Analysis (ICA) (Beckmann and Smith, 2004; Martin J McKeown et al., 1998) can well construct anatomically distinct brain networks.

However, the human brain is a highly connected system, thus there is no biological reason for different spatial components to hold independent distribution. A brain area could be involved in multiple functional processes and a functional network could recruit various heterogeneous neuroanatomic areas, even in resting states (Lv et al., 2015b, 2015a).

In our previous studies, we have shown that by using sparse representation (SR) method (Mairal et al., 2010) to decode fMRI signals (Lv et al., 2015a, 2015b), the whole brain can be decomposed into multiple concurrent functional network components. Specifically, by using this method to analyze data from 68 HCP (human connectome project) subjects, 23 task related networks and 9 resting state components were consistently identified from 7 different tasks, namely, Holistic Atlases of Functional Networks and

Interactions (HAFNI). This finding suggests that there exists common concurrent functional components across individuals, and more importantly, these functional components can potentially solve the difficulty posed by the brain variability and be employed in unveiling the differences in diseased brains (Lv et al., 2015c).

To further derive common networks and patient-specific networks requires a relatively large amounts of patients and controls and thereby large amount of fMRI data. Notably, SR method will decompose brain fMRI signals into hundreds of over-complete dictionary components. This big data challenge poses difficulties in making sensible comparisons among healthy controls and diseased brains and then further deriving group-wise consistent networks that are common to patients and controls. In response, we proposed an effective volumetric network descriptor, named connectivity map, to compactly and quantitatively describe spatial patterns of brain network maps and resort to the power of distributed system and implemented a fast, novel framework in the Apache Spark (<http://spark.apache.org/>) environment that can identify group-wise consistent networks featured in ASD patients and controls. The applications of our new methods and systems on the Autism Brain Imaging Data Exchange (ABIDE) dataset achieved promising results.

Methods

Fig. 2.1 summarizes the pipeline of our proposed computational framework. First, by using SR method (Lv et al., 2015a), the whole-brain fMRI data of each subject was decomposed into multiple components. Then, the connectivity map was obtained from the spatial map of each network component. To reduce computational complexity, two similarity measurements of different levels of accuracy were employed to characterize how

similar two components are. The coarser similarity, measured by similarity between connectivity map, is calculated as the preliminary similarity measurement and the pairwise components overlap rate is used as the final similarity measurement. A sparse similarity matrix between all SR components obtained from all subjects was generated using high-speed computing framework implemented on Apache Spark. Then spectral clustering was performed on the similarity matrix to identify group-wise consistent brain resting-state networks (ICNs) across individuals. For those component clusters that can be reproducibly identified in individuals, an ICN template was generated for each of them. Then the interactions between those ICNs were further analyzed in individual brain space. Specifically, a two-sample t-test was performed to identify the inter-network interactions that are significantly different between ASD patients and healthy controls as connectomics signatures. In addition, a support vector machine (SVM) classifier was trained to differentiate ASD patients from controls based on the connectomics signatures. To better understand the details of the networks that show differences between controls and patients, another round of spectral clustering was carried out on the interactions of ICNs.

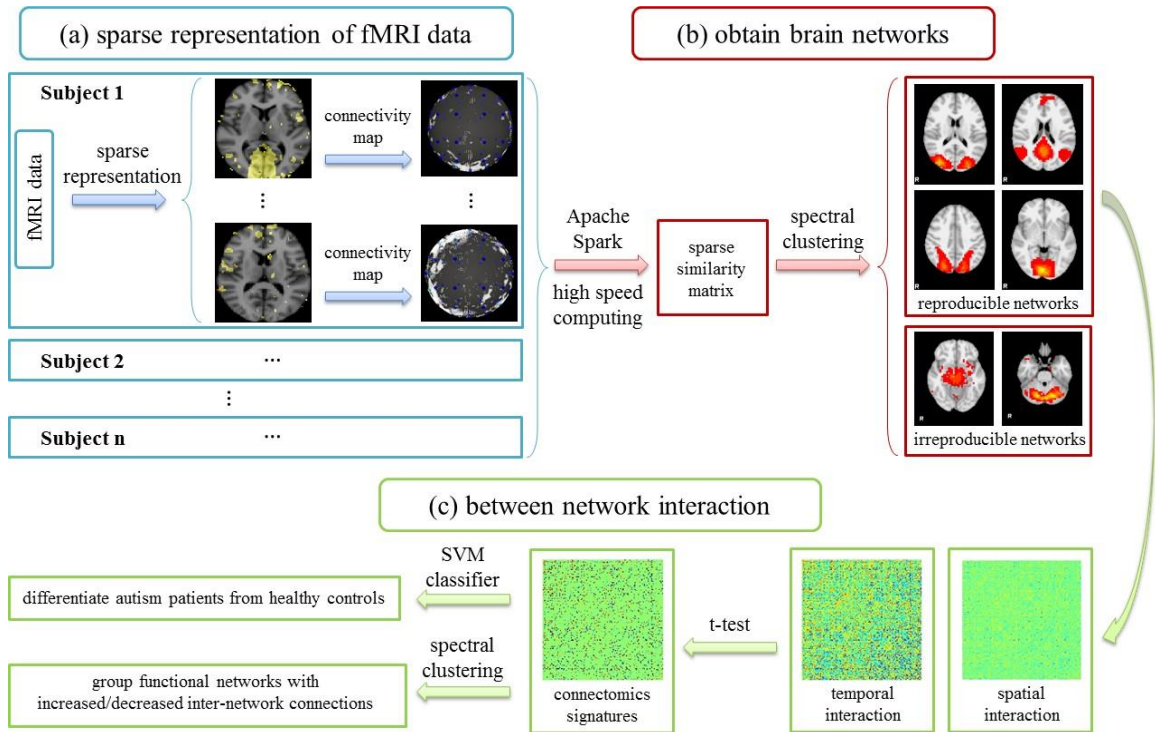


Fig. 2.1. Flow chart of the proposed computational pipeline.

Experimental Data and Preprocessing

Our experimental data was downloaded from the publicly available Autism Brain Imaging Data Exchange (ABIDE) which provides previously collected resting state functional magnetic resonance imaging (rsfMRI) datasets from individuals with ASD and healthy controls (http://fcon_1000.projects.nitrc.org/indi/abide/). In this study, the rsfMRI data obtained from 77 ASD patient individuals and 101 healthy controls from NYU Langone Medical Center were used to develop and test our computational framework. The acquisition parameters were as follows: 240mm FOV, 33 slices, TR=2s, TE=15ms, flip angle = 90°, scan time = 6mins, voxel size = 3 × 3 × 4 mm.

Data preprocessing based on FSL (Jenkinson et al., 2012) was similar to that in (Lv et al., 2015a), which includes skull removal, motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and linear registration

to the MNI space. All of these steps are implemented by FSL FEAT and FLIRT. Then the fMRI data was decomposed into different functional networks using sparse representation (SR) (Lv et al., 2015a) introduced later.

Sparse Representation of Whole-brain fMRI data

Sparse representation is a useful machine learning method that can faithfully reconstruct the signal and achieve a compact representation of signals. Based on sparse representation, the whole-brain fMRI signals of each individual could be decomposed into multiple network components as proposed in (Lv et al., 2015a). The whole process is illustrated in Fig. 2.2. First, the BOLD signal in each voxel of fMRI data was normalized. Then the normalized signals in the whole brain were extracted from fMRI data to form a matrix $X \in \mathfrak{R}^{t \times n}$ with n columns containing normalized BOLD signals from n foreground voxels. By applying the online dictionary learning and sparse coding method (Mairal et al., 2010), each column in X was modeled as a linear combination of atoms of a learned basis dictionary D such that $X = D \times \alpha + \varepsilon$, where $D \in \mathfrak{R}^{t \times m}$ is a dictionary matrix, $\alpha \in \mathfrak{R}^{m \times n}$ is a sparse coefficient matrix, and ε is the error term. Finally, each row in the α matrix was mapped back to the brain volume as a network component for future analysis. In (Lv et al., 2013a), the authors have shown that the meaningful networks decomposed do not change significantly with the alteration of dictionary size. Considering the size of fMRI data matrix and the number of signals, dictionary size was empirically set to 200 and sparsity constraint lambda was set to 0.15 for this study.

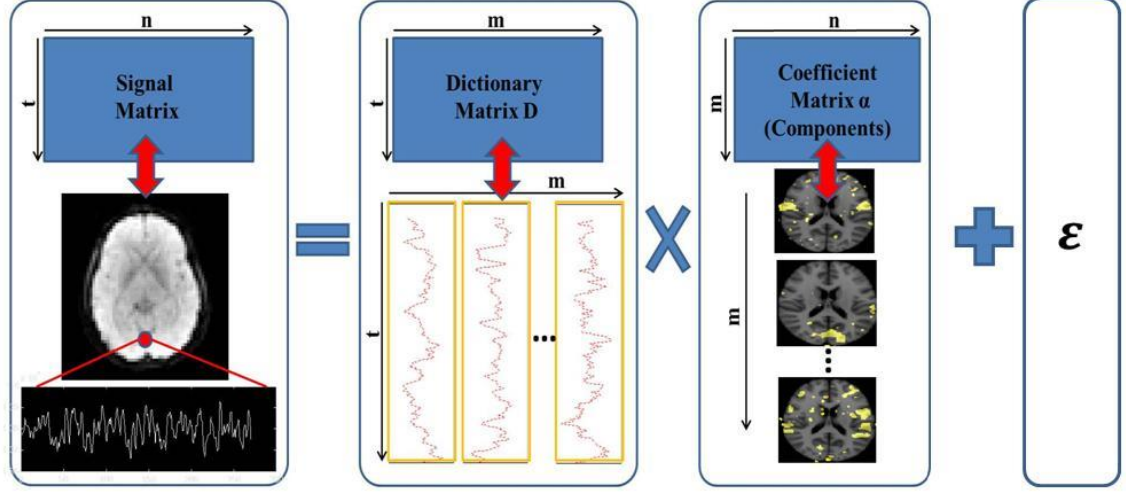


Fig. 2.2. Illustration of network components generated by sparse representation.

Brain Network Similarity Based on Overlap Rate

To identify group-wise common brain networks, it is essential to first define a quantitative measurement of the similarity between brain networks. Since all the subjects have been pre-registered to the MNI space, it is intuitive to use the overlap rate between spatial maps of brain networks as a similarity metric. Specifically, we performed voxel-wise comparisons and define overlap rate similarity (ORS) as the summation of the minimum value in each voxel between two components over the summation of the averaged value in each voxel as the following:

$$S(v_i, v_j) = \frac{\sum_{k=1}^{|V|} \min(v_{i,k}, v_{j,k})}{\sum_{k=1}^{|V|} (v_{i,k} + v_{j,k}) / 2} \quad (2.1)$$

where $v_{i,k}$ is the value in the k^{th} voxel of a brain network component V_i . The larger the overlap rate is, the more similar two components are (Fig. 2.3). The major advantage of this measurement is that it takes into consideration each of the foreground voxels of both components and thereby offers an accurate estimation of the similarity. The drawback,

traded with the high accuracy, is the computational complexity. Since there are a total of 35,600 components from 178 subjects and each component has a feature length of over 100,000 (i.e., the number of foreground voxels in a brain image), pairwise similarity is both very time-consuming and memory-expensive to compute. Two solutions were proposed and explored here. One is to perform feature dimension reduction without compromising too much on accuracy. The other approach is to employ the power of data-intensive computing platform of Apache Spark. Both strategies will be detailed in the following sections.

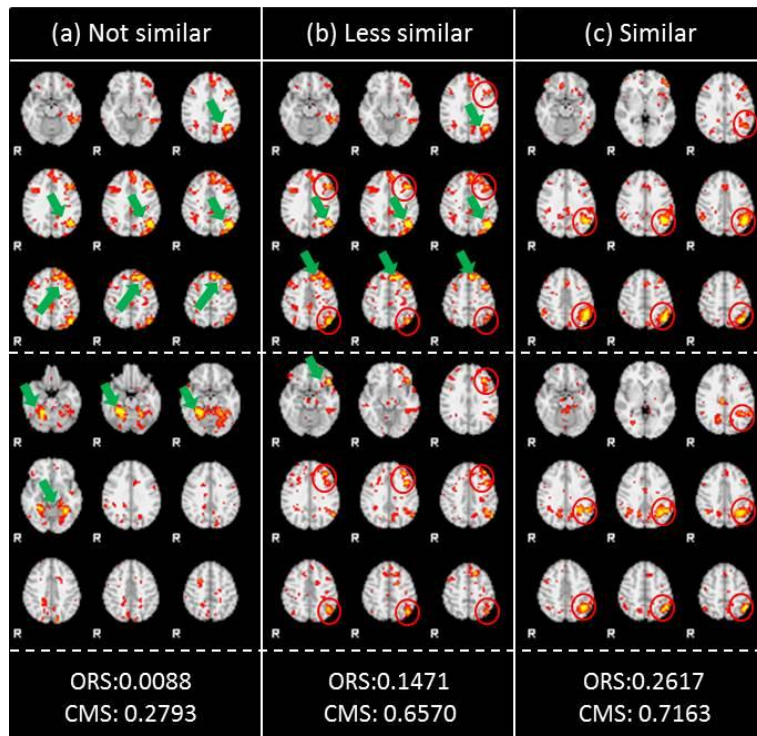


Fig. 2.3. Examples of similarity measurement. *Connectivity map* similarity (CMS) and overlap rate similarity (ORS) between each pair of selected SR components were listed below each subfigure. The similar parts were highlighted by red circles and dissimilar parts were highlighted by green arrows.

Apache Spark Implementation for Speed-up Calculation

Building pair-wise correlation matrices requires a polynomial time algorithm with a time complexity of $O(n^2)$, where n is the total number of components (35,600 in this study). Even if there is enough memory to load the volumes of all the components, it will take approximately 1280 hours by using single core to compute pair-wise ORS (C++ implementation).

To efficiently manage memory and to employ the power of parallel/distributed computing, the whole computation was implemented on Apache Spark (<https://spark.apache.org>). Apache Spark is an open-source cluster computing framework originally developed in the AMPLab at the University of California Berkeley, which provides a fast and general engine for big data processing. The key steps to deploy Spark in this task include loading each volumetric component as a Resilient Distributed Dataset (RDD), which is the basic distributed memory abstraction in Spark and re-organizing the data to perform pairwise comparison using Cartesian products. The output of the processing is a sparse similarity matrix.

Notably, Apache Spark takes care of job scheduling and resource allocation among multiple clusters and cores. It is a very powerful tool for complex, multi-step data pipelines development. Another useful feature is that Spark not only performs in-memory computing, but also allows data exchange between memory chips and hard drives with optimized speed when the memory is not sufficient to store the big data. Due to these properties, Spark is ideal in solving our big data problem of calculating similarity matrices among all of the SR components. With Spark, the original 1280 hours' computation on one core can be significantly shortened to roughly 100 hours running on our 16-core server.

Brain Network Similarity Based on Connectivity Map

To further reduce the computational complexity, dimensionality reduction techniques were exploited. Here we proposed a compact volume shape descriptor named connectivity map. The basic idea of the connectivity map is to unfold the spatial pattern of volumetric voxels by projecting them to points on a unit sphere. Then by sampling the distribution of points on the sphere, a 1-dimensional numerical vector can be obtained to describe the distribution pattern of the spatial map. The idea was inspired by the shape descriptor of streamline bundles (Chen et al., 2013; Zhu et al., 2012). But the novelty here is that it is customized for brain network description in the 3D volumetric space. The procedure of projection is illustrated in 2D space as shown in Fig. 2.4(a)-(b). An example of a 3D volume and obtained connectivity map is shown in Fig. 2.4(d)-(f). The mapping procedure from a network map to a connectivity map is listed as follows.

1. Select a projection center v_0 in the 3D space (red point in Fig. 2.4 (a) and Fig. 2.4 (d));
2. Calculate the vectors from the projection center to each foreground voxel in the network map (red arrows in Fig. 2.4 (a)-(b)) and normalize them such that each vector can be represented by a point on unit sphere (blue arrows in Fig. 2.4 (b)) and its intensity recorded accordingly:

$$W = \{(i_k, \vec{u}_k) \mid \vec{u}_k = (\vec{v}_k - \vec{v}_0) / \|\vec{v}_k - \vec{v}_0\|, k \in V\} \quad (2.2)$$

where V is the set of a brain network's foreground (non-zero value) voxels; W is the set of projected points; v_k and v_0 are the coordinates of voxels and projection centers; i_k is the voxel intensity (the weighted coefficient in the sparse representation).

3. The probabilistic distribution density of P on the unit sphere is applied to describe the connection map of the voxel data. Specifically, the sphere is segmented into 48 equally

sized regions (Chen et al., 2013). Then, the intensities of projected points in each region are accumulated and normalized:

$$P_j = \sum_{k \in W \cap R_j} i_k / \sum_{k \in W} i_k (j=1 \dots 48) \quad (2.3)$$

where R_j is the area covered by region j , P is the connectivity map of a volumetric image.

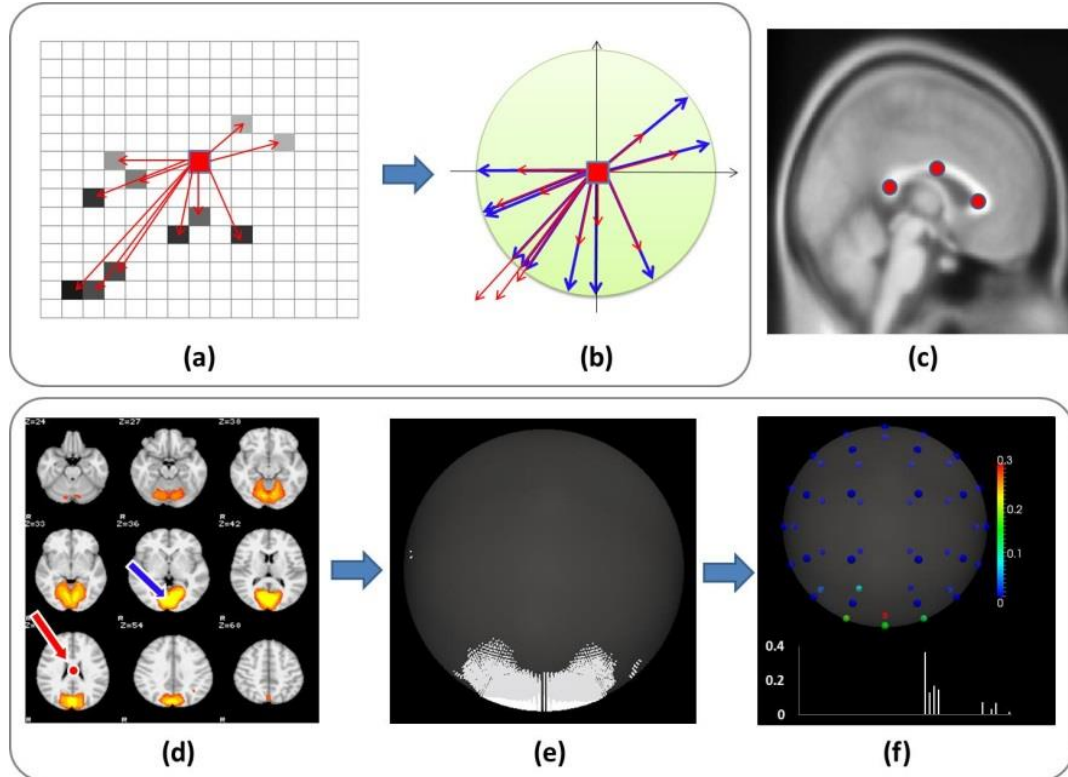


Fig. 2.4. Illustration of connectivity map for volume image. (a) 2-D spatial map; (b) projection of spatial map in (a) to unit circle; (c) 3 selected projection centers represented by red dots; (d) 3-D spatial map of a functional network; (e) projection of spatial map in (c) to unit sphere; (f) connectivity map for (c).

Considering the property of the connectivity map as a probability density vector, the intersection between connectivity maps of two networks V_i and V_j can be applied as a similarity measurement:

$$S(P(V_i), P(V_j)) = \sum_{k=1}^{48} \min(P_k(V_i), P_k(V_j)) \quad (2.4)$$

By this definition, the similarity value will be between 0 and 1, and a higher similarity value indicates networks of higher similarity (Fig. 2.3).

The description power and the accuracy of connectivity maps largely depend on brain orientation and the selection of projection centers. Intuitively, once the brain orientation or the location of projection center changes, the connectivity map for the same spatial map may also vary. Thus, to make connectivity maps of network spatial maps comparable among individuals, we need to 1) align the brains in the same orientation and 2) select projection centers at the same location among individuals' brains. Meanwhile, connectivity map may lose description power if the projection center locates inside the network cluster of a brain network. For instance, if we select the activation center highlighted by the blue arrow in Fig. 2.4(d) as the projection center, the projected points will be evenly distributed on the unit sphere and thus it will be hard to tell the pattern of network spatial map with a connectivity map. Another limitation of the connectivity map is that it cannot identify two different networks when they locate in the same orientation from the perspective of projection center. To solve these issues, as shown by red dots in Fig. 2.3 (c), we selected 3 projection centers along the corpus callosum and generated 3 connectivity maps for each network map analyzed. The final connectivity map similarity (CMS) is defined as the average similarity between 3 pairs of connectivity map. Our rationale is that: 1) by selecting projection centers in white matter regions, the situation that it locates inside network clusters could be eliminated; 2) by applying multiple projection centers to observe each network spatial map from different views, it could be ensured that different network maps will result in different connectivity maps; 3) by selecting projection

centers from consistent and identical anatomical structures in the brain, the errors caused by individual variability can be reduced and the connectivity map's accuracy in comparing brain networks across individuals can be improved.

Our testing results showed that CMS performs quite well in identifying dissimilar SR components (Fig. 2.3 (a)). Yet its performance in discerning similar SR components is less robust. In several cases, dissimilar components may also receive a high CMS (Fig. 2.3 (b)-(c)). Nevertheless, given its precise and fast ability in identifying dissimilar SR components, CMS can be employed initially to quickly filter out evidently dissimilar SR components with high confidence. Later, similarities between the remaining components are measured by the more accurate ORS. Intuitively, by increasing CMS threshold, the amount of similarities to be calculated using ORS will be significantly reduced, while the chance of ignoring similar components will be increased. Thus in our computation, a relatively conservative CMS threshold (0.5) was chosen to minimize accuracy trade off during computational speedup. And only when ORS is larger than 0.2, the two components will be considered as similar and the ORS between them will be recorded. After integrating CMS measurement into our computational framework, further speedup was achieved and the whole computation only took 64 hours to finish by using 16 cores on a single server.

Results

Resting-state Networks in Autism and Healthy Control

For each common ICN, we then performed one-tailed t-test to compare the ORSs between ASD patients and healthy controls. 4 ICNs with significantly higher (p-value: 0.025) ORS in ASD patients were identified (Fig. 2.6). 12 ICNs with significantly higher (p-value: 0.025) ORS in healthy controls were also identified (Fig. 2.7).

For the rest of 128 ICNs, they were commonly distributed in both populations (8 of which are shown in Fig. 2.5). An increased/decreased ORS reflects increased/decreased connections within each ICN. A decreased ORS in DMN (network 7, Fig. 2.8(a)-(b)) agrees with the previous findings that the connection within DMN decreases in ASD patients in comparison with the healthy controls during resting state (Assaf et al., 2010; Kennedy and Courchesne, 2008). In addition, we identified increased connections within Right Frontal Pole (FP) (network 137); Left FP (network 123); Middle Temporal Gyrus (MTG) (network 133); and Precuneous Cortex (PC) (network 45). The functional networks with decreased connections (decreased ORS) we identified includes Occipital Fusiform Gyrus (OFG) / Inferior Lateral Occipital Cortex (ILOC) (network 144, 57, 69); Anterior Cingulate Gyrus (ACG) / Insular (network 30); Right Superior Lateral Occipital Cortex (SLOC) / Right Middle Frontal Gyrus (MFG) (network 38); PC / FP / Frontal Medial Cortex (FMC) / ACG / Posterior Cingulate Gyrus (PCG) (network 3); FP / Superior Frontal Gyrus (SFG) / PC (network 20); Caudate / Putamen / Thalamus (network 32); Right MFG / Right SFG / Right SLOC (network 58). Intriguingly, we also identified decreased functional connections within Cerebral Crus (network 55) and Lateral Ventricle (network 12). We will discuss these findings later.

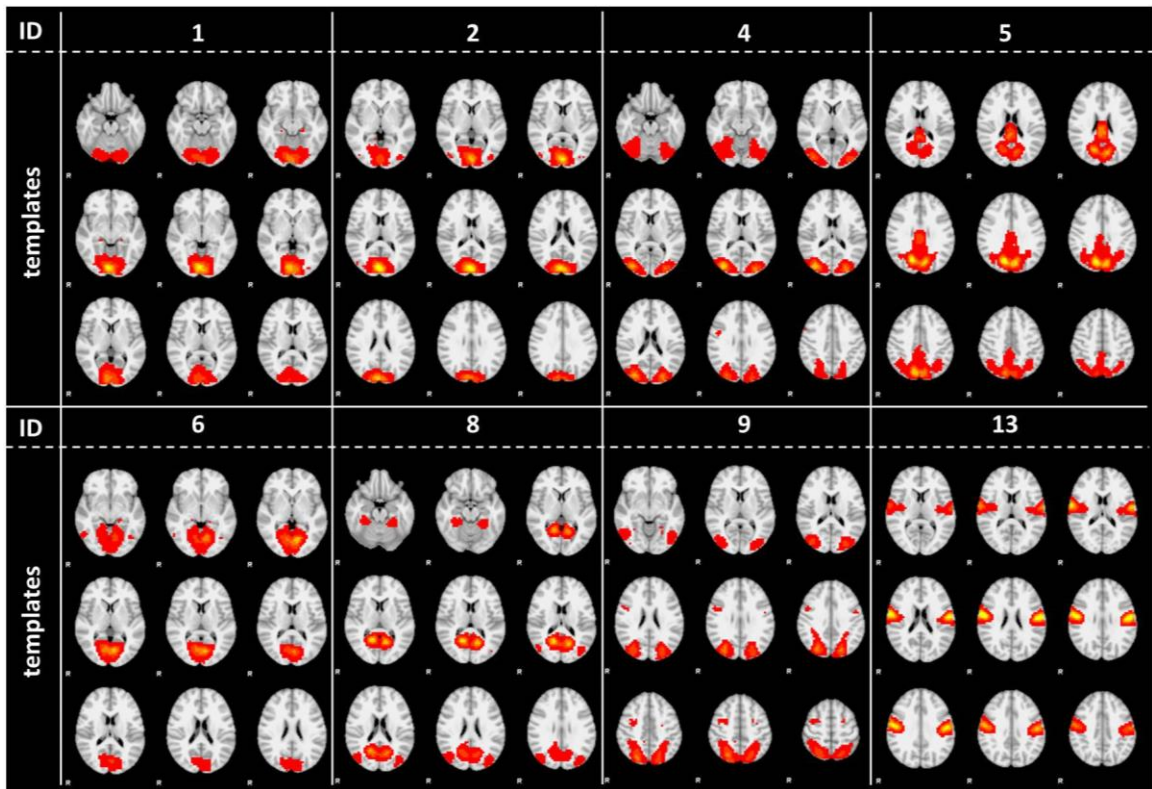


Fig. 2.5. 8 of 128 common ICNs consistently identified in both autism patients and healthy controls.

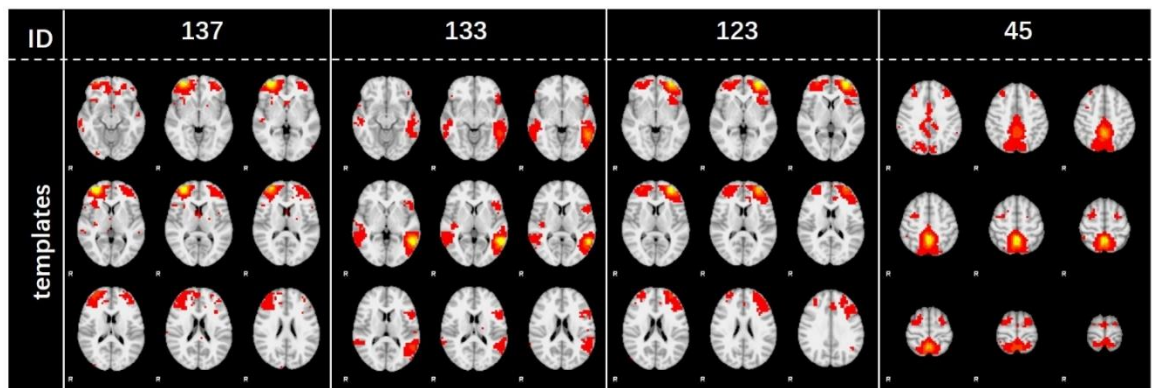


Fig. 2.6. Templates of ICNs with significantly higher (p-value: 0.025) ORS in ASD patients. All the networks were sorted by p-values. Templates of ICNs with significant

higher (pFDR-value) ORS in controls left after multiple comparison correction are listed inside the red box.

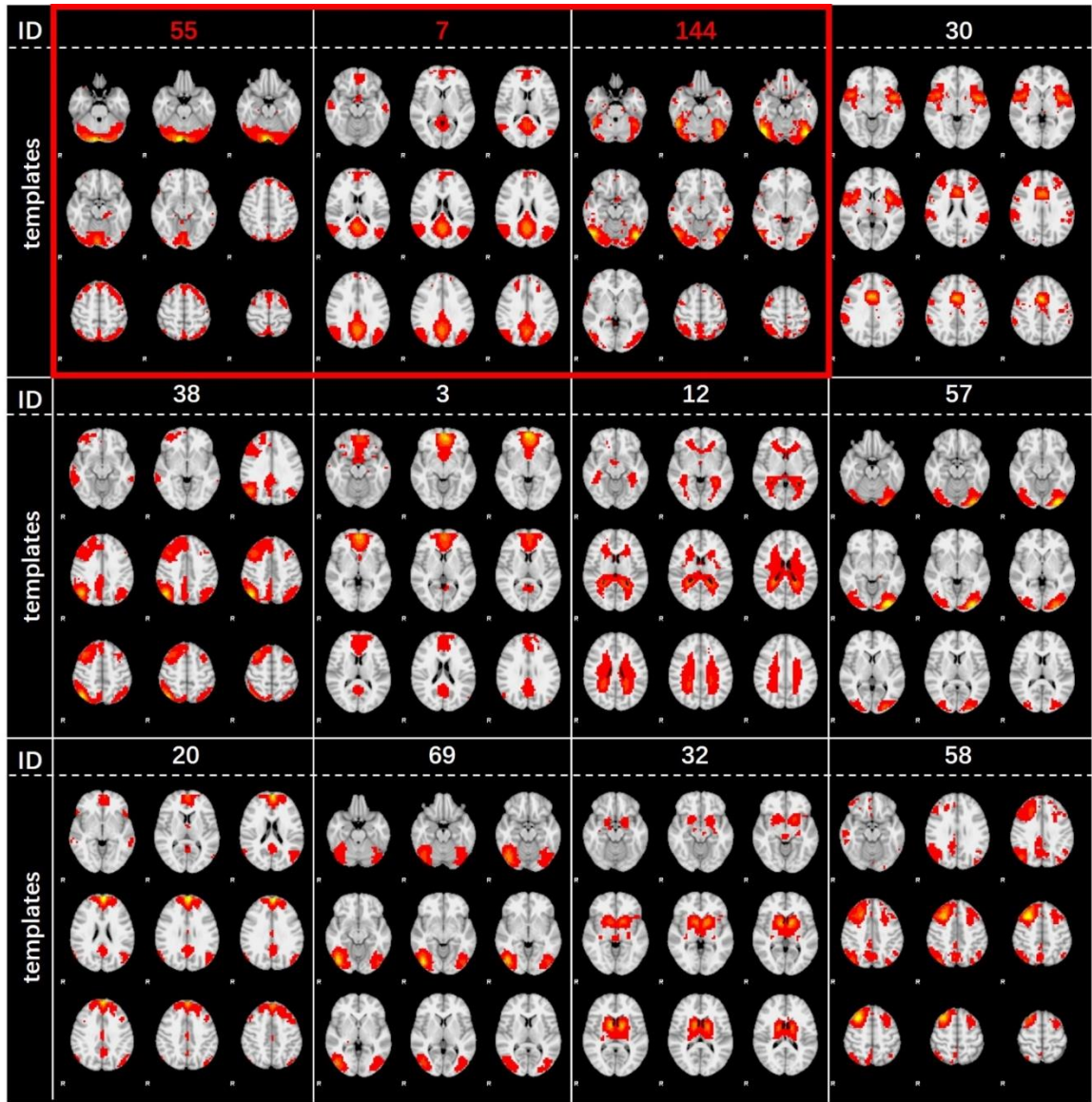


Fig. 2.7. Templates of ICNs with significantly higher (p-value: 0.025) ORS in healthy controls. All the networks were sorted by p-values. Templates of ICNs with significant higher (pFDR-value) ORS in patients left after multiple comparison correction are listed inside the red box.

Between Network Interactions

In order to further investigate the differences between autism patients and healthy

controls, we computed the interactions between all 144 ICNs in each individual’s brain. First, time series were extracted from the dictionary atoms as introduced in section 2.2. The temporal interactions between each pair of networks are measured using the Pearson correlations of the extracted time series. And the spatial interaction is measured by the ORS between corresponding components. Then based on two-sample t-test with 1000 permutations, the interactions that are significantly different (p-value: 0.05) between ASD patients and healthy controls were selected as connectomics signatures. Based on these connectomics signatures, SVM classification was performed and tested in 10 folds cross validation manner. On average, we achieved 93.8% accuracy, 91.1% sensitivity, and 95.3% precision in differentiating ASD subjects from the healthy controls, which is quite promising and encouraging.

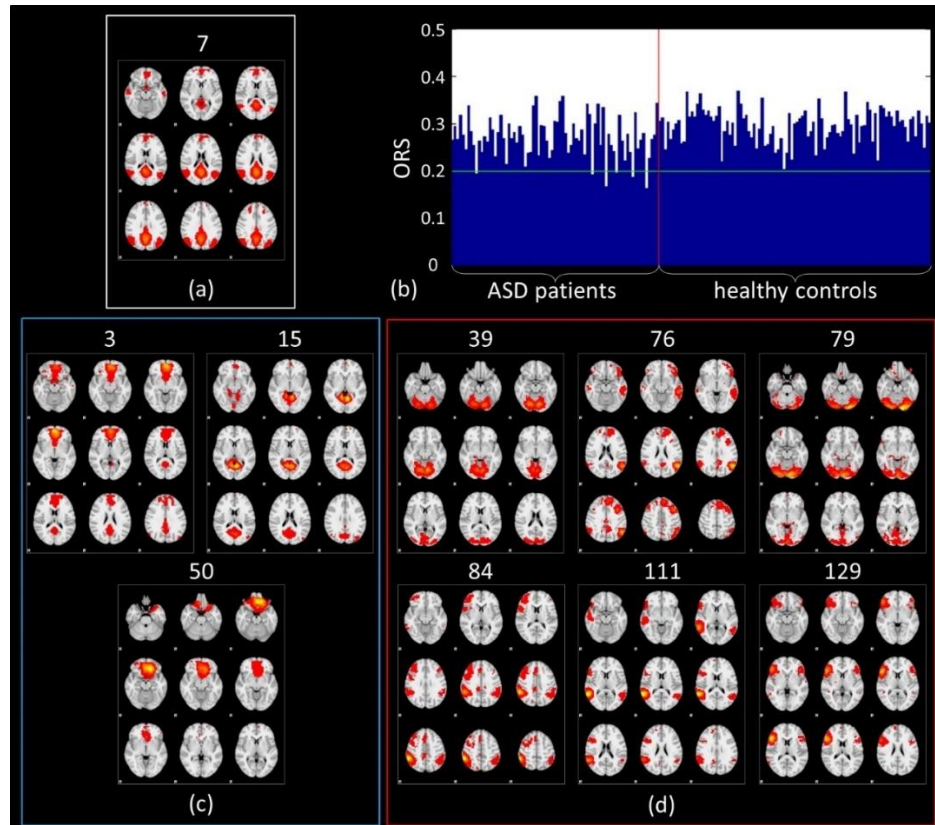


Fig. 2.8. (a) Spatial map of network 7. (b) ORS between the template of network 7 and the corresponding SR components in each individual's brain. (c) ICNs with decreased temporal interactions to network 7. (d) ICNs with increased temporal interactions to network 7.

One-tailed t-test was performed to identify the interactions that are significantly increased or decreased in ASD patients. The findings in both types of interactions were largely consistent with each other (Fig. 2.9). There is no conflict finding between temporal interactions and spatial interactions – if one interaction significantly increased, the other either significantly increased or had no significant differences between ASD patients and healthy controls. All the interactions significantly changed were listed in Fig. 2.9 (c) together with the differences in average temporal interactions (Fig. 2.9 (a)) and average spatial interactions (Fig. 2.9 (b)). Intriguingly, increasing interactions between DMN and multiple language-related ICNs were identified. This may relate to the impairment in social interactions and communication of ASD subjects and will be further discussed.

Based on the results of one-tailed t-test, we further performed clustering to cluster brain ICNs with strong interaction increase or decrease. Specifically, two affinity matrices (increased/decreased) were generated such that if both temporal and spatial interactions are significantly increased or decreased, the corresponding connectivity is defined as 1; and if only one type of interaction significantly increased or decreased, the connectivity was defined as 0.5, otherwise the connectivity was 0. For each affinity matrix, we first eliminated the ICNs with weak interactions, and then performed spectral clustering to cluster ICNs with strong interactions. 14 network clusters were finally obtained for both

increased interactions and decreased interactions, and for convenience, they are referred as increased/decreased network interaction clusters (INICs/DNICs) (Fig. 2.9). Intriguingly, we also found a considerable amount of decreased interactions within INICs as well as increased interactions within DNICs (highlighted by black arrows in Fig. 2.9). For interactions within INIC, 27.9% of within cluster network interactions increased while 3.9 percent decreased. And for interactions within DNIC, 28.4% of within cluster network interactions decreased while 1.7% increased. The complicated changes in functional network interactions might be related to the network alterations and functional compensation.

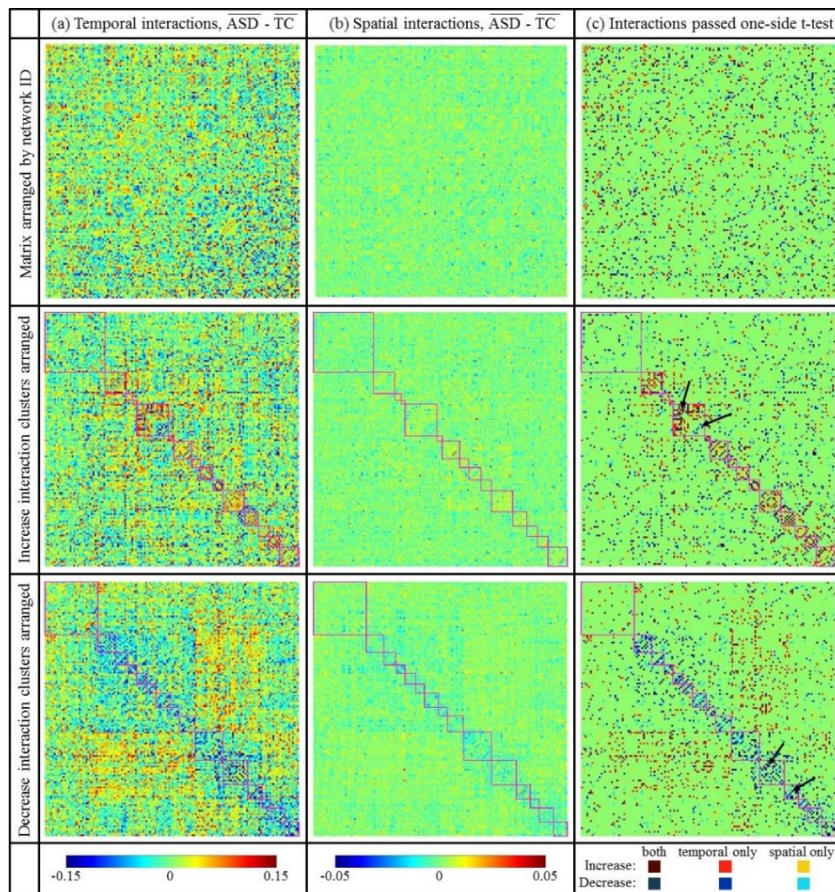


Fig. 2.9. Group-wise comparisons of inter-network interactions. (a) Differences in average temporal interactions between ASD patients and healthy controls (TCs). (b) Differences in

average spatial interactions between ASD patients and TCs. (c) Significantly increased or decreased interactions identified by one-tailed t-test (p-value: 0.025). Color bars were listed at the bottom of subfigures. The matrices in the second and third rows were arranged by clusters. In each subfigure, starting from top left, the 1st magenta box is the ICNs with weak connections. Following it, the rest of the boxes are clusters from #1 to #14 respectively.

Discussion

In this study, we have identified 144 group-wisely consistent ICNs from the rsfMRI data of 77 ASD subjects and 101 healthy controls. To our best knowledge, this is one of the most comprehensive group-wise studies of ICNs among ASD studies at the time of the publication using a non-deep learning framework. Due to the limitations on computational method and analysis power, previous studies usually focused on several functional networks or ICNs (Philip et al., 2012; Stigler et al., 2011). Thanks to the novel and powerful large-scale fMRI data mining framework proposed in this study and the availability of big rsfMRI dataset, it is surprising that the number of common ICNs in the human brain could be as large as 144 (<http://hafni.cs.uga.edu/autism/templates/all.html>), which is significantly larger than the number of networks analyzed in the previous studies (Rosazza and Minati, 2011; van den Heuvel and Hulshoff Pol, 2010). Moreover, our further investigations on these 144 common ICNs between ASD patients and healthy controls identified interesting patterns that may unveil atypical behaviors in ASD.

One major fMRI findings among ASD studies is the atypical patterns of Fusiform Gyrus activation during face processing in ASD (Stigler et al., 2011). These findings are directly related to the explanations of abnormal social behaviors in ASD. In our analysis, decreased connectivity has also been identified in Fusiform Gyrus during resting-state

(network 144, 57, 69). These findings not only add support to the existing studies, but also suggest that the abnormalities in task related functional networks can also be identified during resting-state. In addition to the novel insights on ASD studies, this work also provides a novel and effective solution for big data fMRI studies. Given the computational efficiency of Apache Spark using connectivity map, the method can be easily scaled up to analyze and cluster brain networks in big datasets consisting of thousands of subjects by its integration with large-scale informatics systems such as HAFNI-Enabled Large-scale Platform for Neuroimaging Informatics (HELPNI) (Makkie et al., 2015). In the future, the researchers can apply our proposed framework with connectivity map together with functional brain network decomposition methods such as SR (Lv et al., 2015b, 2015a) to analyze such large scale brain functional datasets and identify abnormal as well as common functional networks in diseased brains in order to better understand the functional mechanisms of brain diseases.

As noticed through this research, an obvious limitation of the proposed connectivity map is the lack of the ability to accurately capture the spatial map distribution, even though a fair accuracy can be achieved with a much faster speed by introducing dimension reduction. Following this achievement and inspired by the success of deep learning frameworks in applications of medical image analysis, section 2.4 introduces an improved deep learning-based framework for better brain functional spatial map description.

Acknowledgement

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) for this research. Z Kou was partially supported by National

Institutes of Health (R21NS090153) for this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

2.3 A Novel Framework for Groupwise Registration of fMRI Images based on Common Functional Networks

Abstract

In parallel with functional spatial map descriptors abovementioned, accurate registration also plays a critical role in group-wise functional Magnetic Resonance Imaging (fMRI) image analysis as spatial correspondence among different brain images is a prerequisite for inferring meaningful patterns. However, the problem is challenging and remains open, and more effort should be made to advance state-of-the-art image registration methods for fMRI images. Inspired by the observation that common functional networks can be reconstructed from fMRI image across individuals, we propose a novel computational framework for simultaneous groupwise fMRI image registration by utilizing those common functional networks as references for spatial alignments. In this framework, firstly, individualized functional networks in each subject are inferred using Independent Component Analysis (ICA); secondly, congealing groupwise registration that takes entropy of stacked independent components (ICs) from all the subjects as objective function is applied to register individual functional maps to be maximally matched. The proposed framework is evaluated by and applied on an Alzheimer's Disease (AD) fMRI dataset and showed reasonably good results. This work, in parallel with connectivity map introduced in section 2.2, is the only non-deep learning frameworks in this dissertation as success of deep learning applications just started around the same work with the publication

of these two researches. However, after investigating fMRI spatial analysis utilizing traditional non-deep learning frameworks by modelling spatial networks, we observed the limitation of the traditional brain spatial networks modeling frameworks, which therefore paved the way for further deep learning based framework.

Introduction

Recent advances in neuroimaging techniques, especially high-resolution functional magnetic resonance imaging (fMRI) of the brain such as the Human Connectome Project (HCP) (Barch et al., 2013; Smith et al., 2013), made it possible for exploring functional human brain organization in unprecedented details. Due to the individual variability of human brain, registering individual's fMRI images to a common space for group-wise or population-scale analysis is a critical step for understanding the human brain structure and function and their relationships. Successes have been reported in many registration algorithms on structural MRI images (Oliveira and Tavares, 2014), including sophisticated pairwise registrations (Jenkinson and Smith, 2001; Shen and Davatzikos, 2002), and image registration methods between different imaging modalities (Luan et al., 2008; Maes et al., 1997). One common registration approach used in group-wise fMRI image analysis is to directly register fMRI images to a template defined by averaged structural MRI images using well established linear registration algorithms originally designed for structural MRI data (Saad et al., 2008). However, typical T2*-weighted volumes of Blood Oxygenation Level Dependent (BOLD) fMRI images are low-resolution and have poor anatomical contrast (Logothetis, 2008). As a result, traditional intensity-based or feature-based structural image registration methods do not necessarily work well for fMRI volume

images. These difficulties motivated us to explore new methods for fMRI image registration in this paper.

Inspired by the observation that multiple functional brain networks can be commonly and robustly identified across individuals based on resting state fMRI data, we propose a novel groupwise fMRI registration framework by using common functional networks as references, which can provide functionally meaningful boundaries and contrast that are missing in raw BOLD fMRI images. This is the major methodological novelty of our image registration framework. In addition, the proposed framework can significantly reduce the biases brought by pairwise registration (Liu and Wang, 2014) through utilizing groupwise registration methodology. To our best knowledge, this is among the earliest works utilizing functional networks information for groupwise fMRI image registration. This fully automated framework was evaluated by and applied on both the Alzheimer’s Disease Neuroimaging Initiative (ADNI) resting state fMRI data, and promising results are achieved.

Material and Method

Experimental Data and Preprocessing

Our experimental data were downloaded from publicly available Alzheimer’s Disease Neuroimaging Initiative (ADNI) (adni.loni.usc.edu), which provides previously collected resting state functional magnetic resonance imaging (rsfMRI) datasets from individuals with AD and healthy controls. Since we are aiming to evaluate groupwise registration framework results by exploring the potential functional abnormalities in the AD patients, only first rsfMRI scans until Sep. 2013 of each 40 NC subjects and 33 AD subjects are

included in this paper. All rsfMRI datasets were acquired on 3.0 Tesla Philips scanners (varied models/systems) at multiple sites. The range for imaging resolution in slice dimension is from 2.29 to 3.31 mm and the slice thickness is 3.31 mm. TE (echo time) for all subjects is 30 msec and TR is from 2.2 to 3.1 sec. For each subject, there are 140 volumes (time points). Preprocessing steps of the rsfMRI data included brain skull removal, motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and band pass filtering (0.01–0.1 Hz) (Zhu et al., 2014).

Functional Intrinsic Connectivity Networks (ICNs) Guided Groupwise Registration

Inspired by the observation that multiple functional brain networks can be commonly identified across individuals via ICA methods (Beckmann and Smith, 2004) or dictionary learning algorithms (Lv et al., 2015a, 2015b), common networks can be utilized to establish correspondence between fMRI images. In this paper, after preprocessing, functional ICNs are decomposed from each fMRI image and a set of z-score maps representing networks' 3D pattern are obtained by FSL MELODIC ICA software (<http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/melodic>). Then label map images of each subject are obtained by “stacking” all the ICNs from the same subject together by choosing the corresponding ICN index number as the new intensity value in the label map, as shown in

Fig. 2.10.

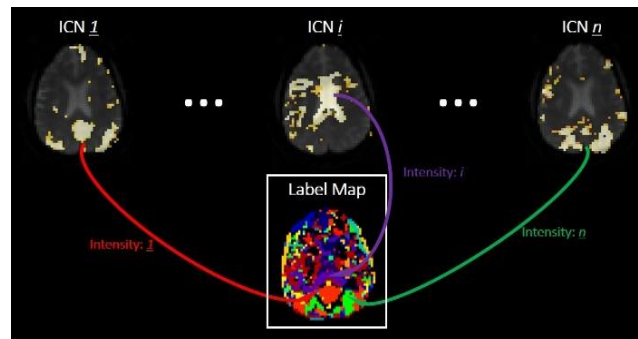


Fig. 2.10. Label map stacking procedure.

This label map “stacking” procedure can be summarized in the following equation:

$$I_i = \mathit{arg}_n \max(I(ICN_n)_i) \quad (2.5)$$

where I_i is the label map intensity at voxel i , $I(ICN_n)_i$ is the intensity value of ICN_n at voxel i .

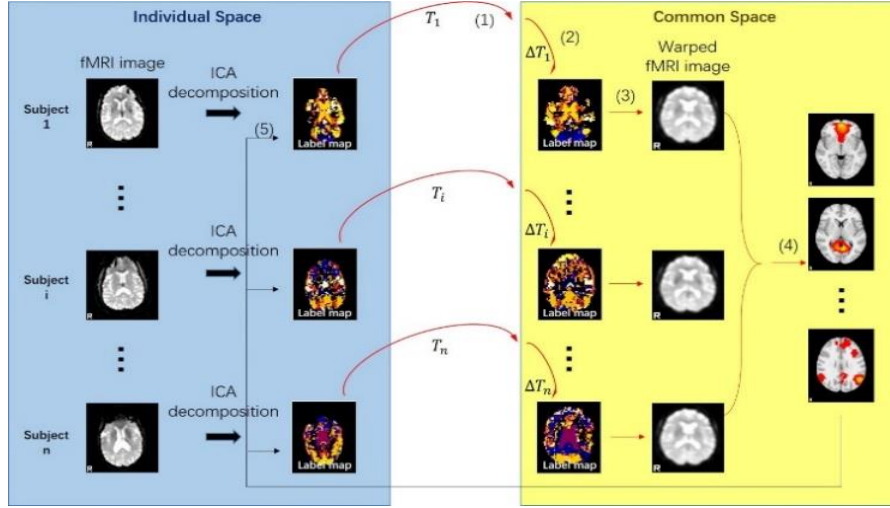


Fig. 2.11. The functional ICNs guided groupwise registration framework. (1). Initial registration from subject space to a standard space using traditional strategy; (2). Congealing groupwise registration of label map images; (3). Warp fMRI images from individual space to common space; (4). Group ICA: obtain common ICNs; (5). Obtain new individualized label map images based on common ICNs.

In general, the work flow of our proposed pipeline is summarized in Fig. 2.11 and detailed explanations are described in the following paragraphs. After the implementation of the whole pipeline, individualized ICNs decomposed using ICA can be mapped into a common space by utilizing the final transformation matrices $T_i + \Delta T_i + \dots$. Groupwise consistent ICNs (e.g. 10 resting-state network (RSN) templates from (Smith et al., 2009))

in the common space from each subject are averaged to be the final ICN templates. The correspondence between the resting-state template (denoted as X in (2)) and the ICNs (denoted as Y in (2.6)) within each subject are established by calculating overlap rate (similar to (Zhao et al., 2016)):

$$overlap(X, Y) = 2 \sum_{i=1}^{\min(|X|, |Y|)} \min(x_i, y_i) / \sum_{i=1}^{\min(|X|, |Y|)} (x_i + y_i) \quad (2.6)$$

where x_i and y_i are the corresponding voxels in the spatial map. The ICN with the highest overlap rate is taken as the correspondence to the template.

Results

Comparisons between Our Method and Traditional Method

In general, in comparison with traditional methods, improvements such as decreased noises, reduced false positive connectivity, and more concentrated average network patterns have been found in the derived ICNs by using our proposed method. As examples, the averaged ICN templates of three RSNs (indexed as RSNs 2, 3 and 4) using ICNs generated from 40 NC subjects from ADNI dataset after the application of our registration method and traditional method are shown in Fig.4. As highlighted in the red circles in Fig.4, the averaged ICNs in RSNs 2-4 by our method have more concentrated functional peaks of less areas, meaning that our image registration method can more accurately warp those 40 ICNs into better spatial alignments than traditional method.

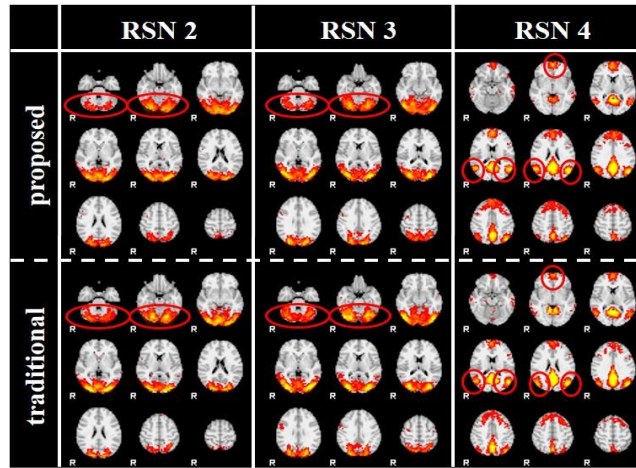


Fig. 2.12. Averaged ICN templates using proposed method and traditional method.

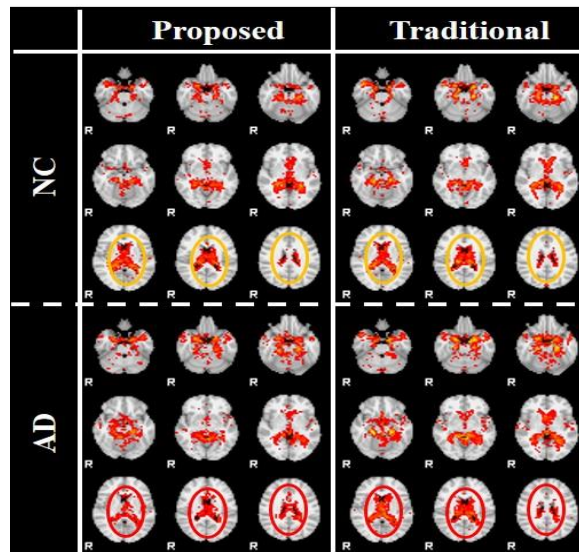


Fig. 2.13. hippocampal ICN templates for NC and AD groups

In addition, for some ICN templates indicating AD group abnormalities, such as hippocampal connectivity ICN, AD group's ICN templates displayed clear changes after applying our proposed method, while NC group's ICN templates remained similar. As shown in the circles in Fig. 2.13, the ventricle regions displayed decreased connectivity after applying our proposed methods (red circles) for AD templates, while no significant

differences (yellow circles) are seen for NC templates. These results suggested better accuracy of our proposed registration method.

Functional Abnormality between NC and AD

Default mode network (DMN) activity is reported to be abnormal in AD (Reiman et al., 1996). In order to further explore the AD's functional connectivity alteration and simultaneously evaluate the ability of our proposed framework for subtle and accurate functional network analysis, the proposed framework was applied on the population of 33 AD patients based on functional network templates.

Fig. 2.14 showed the averaged DMN templates for both NC and AD groups using our proposed framework and two traditional methods. It is clear that AD group exhibits shrunk and altered DMN, as highlighted by the red arrows. Then statistical maps resulting from two-sample t test were generated for quantitative comparisons between NC and AD group (Fig. 2.15). Joint height and extent thresholds of $p < 0.05$ were used to determine significant clusters. The proposed framework showed promising results of detecting AD's decreased DMN activity: more significant deactivation of DMN in most associated brain regions: bilateral inferior-lateral-parietal cortex, precuneus and posterior cingulate, hippocampal area and ventromedial frontal cortex, which are consistent with previously reported literature findings (Greicius et al., 2004). Apparently, our proposed framework showed improvement over all the traditional methods in comprehensively detecting subtle decreased activity in DMN of AD, as shown in Fig. 2.15.

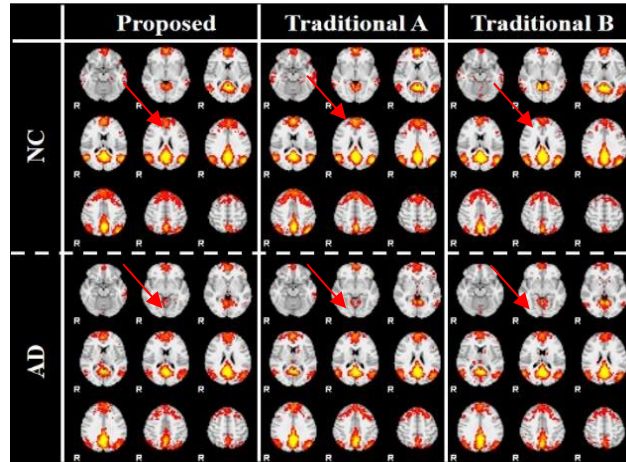


Fig. 2.14. Averaged DMN templates of NC and AD groups, using proposed framework and two traditional strategies

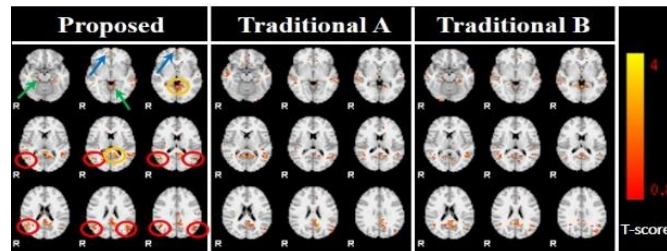


Fig. 2.15. Decreased DMN activity in AD. Red circles showed the significant decreases in bilateral inferior-lateral-parietal cortex; Yellow circles denotes more significantly decreased connectivity in hypo-area of precuneus and posterior cingulate; Green arrows showed more significantly decreased activity in hippocampal area; Blue arrows denotes more significantly decreased activity in ventromedial frontal cortex.

Discussion

In this section, we proposed a new fMRI image registration framework that utilize the individualized common ICNs decomposed by ICA technique from each subject as guidance for groupwise registration. The common functional networks provide functionally meaningful boundaries and contrast that do not exist in raw BOLD fMRI images for better spatial registration. To our best knowledge, this is among the earliest

works that use functional networks for groupwise fMRI image registration. Given the lack of ground truth for validation, the reasonably good experimental results in Section 3 suggest the promise of our framework.

We also observed that the core part for the registration utilizing brain functional spatial maps is still based on the mutual information (MI) of the extracted functional spatial maps. The drawbacks of direct overlap rate or MI metrics on spatial maps are, the overlap rate or MI cannot guarantee the similarities of the spatial maps (as introduced in section 2.4, 2.5, 2.6), as they are not able to model the 3D distribution of the spatial maps. The understandings of the limitations of the traditional frameworks is essential to the proposal of the deep learning frameworks.

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

Abstract

As witnessing the success of the deep learning application in medical image analysis, as well as understanding the limitations of the existing traditional frameworks in spatial analysis, this section introduces a new deep learning framework for a more effective and efficient spatial map modelling for accommodating the variability of individual brain networks and the large scale of such networks needed for statistically meaningful group-level analysis. 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.

Introduction

Before the debut of deep learning application in fMRI analysis, traditional matrix decomposition based 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., 2015c, 2015d, 2015b, 2015a; 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. The fMRI spatial analysis focuses on the several hundreds of concurrent functional brain networks, characterized by both spatial maps and associated temporal time series, effectively decomposed from either tfMRI or rsfMRI data of an individual brain.

As a recapitulation, a brief review of the traditional fMRI decomposition techniques is introduced first. 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. 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.

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 (Zhao et al., 2016), as introduced in section 2.2. As suffering from the limitation of the proposed connectivity map in section 2.2 and inspired from the success of the deep learning application in medical image analysis field, the possibility of utilizing deep learning framework for brain functional network modelling was then explored.

Recently, convolutional neural network (CNN) (He et al., 2015a; 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 section, 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., 2015b, 2015a). 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 section, fine-granularity means that network templates have substantially improved differentiation capability than existing network templates (Lv et al., 2015b, 2015a; 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.

Methods and Materials

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., 2015b, 2015a) 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., 2015b, 2015a). 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. 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 $45 \times 54 \times 45$.

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. 2.16. 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 pair-wise feature similarity matrix generation, upon which a spectral clustering 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.

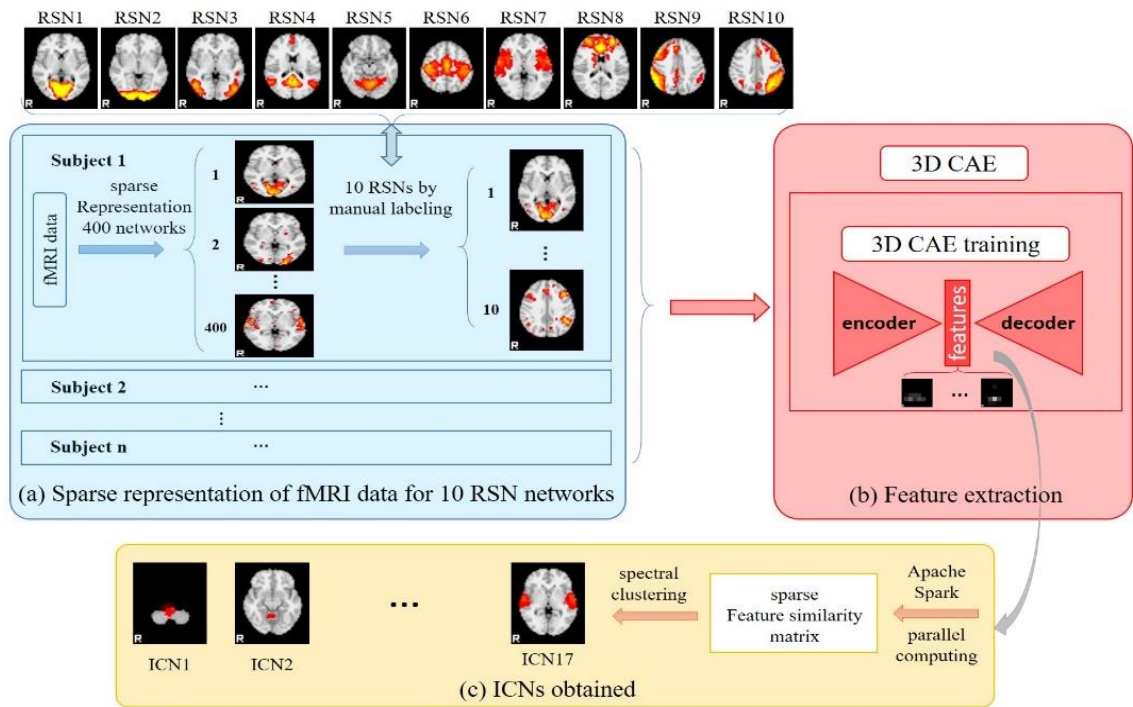


Fig. 2.16. The proposed computational framework. (a) Sparse representation of fMRI data for generation of functional network maps in each individual. Then 10 RSNs were manually labeled for these networks via the methods in (Lv et al., 2015b) for each

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).

3D deep CAE structure

Spatial brain network maps like those Fig. 2.16 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. 2.17. 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.

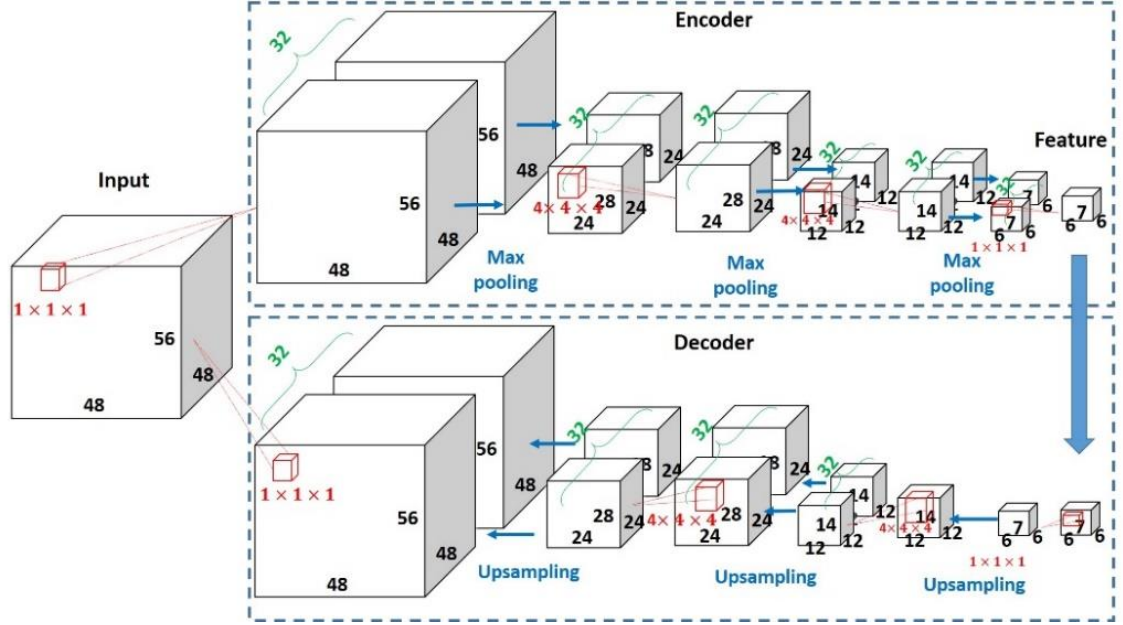


Fig. 2.17. 3D CAE structure with detailed dimensions of each layer (channels, x, y, z). We padded zeros at borders of the input volume, so the $45 \times 54 \times 45$ volume is turned into $48 \times 56 \times 48$. The obtained feature is a $6 \times 7 \times 6$ cube.

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, and x 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 $\theta^T x_i (1 - \theta^T x_i)$ 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 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.

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:

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

where \mathbf{V}_k and \mathbf{W}_k are the activation scores of voxel k in the feature maps \mathbf{V} and \mathbf{W} , respectively. $|\mathbf{V}|$ 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.

Results

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. 2.18. As shown in Fig. 2.18, 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.

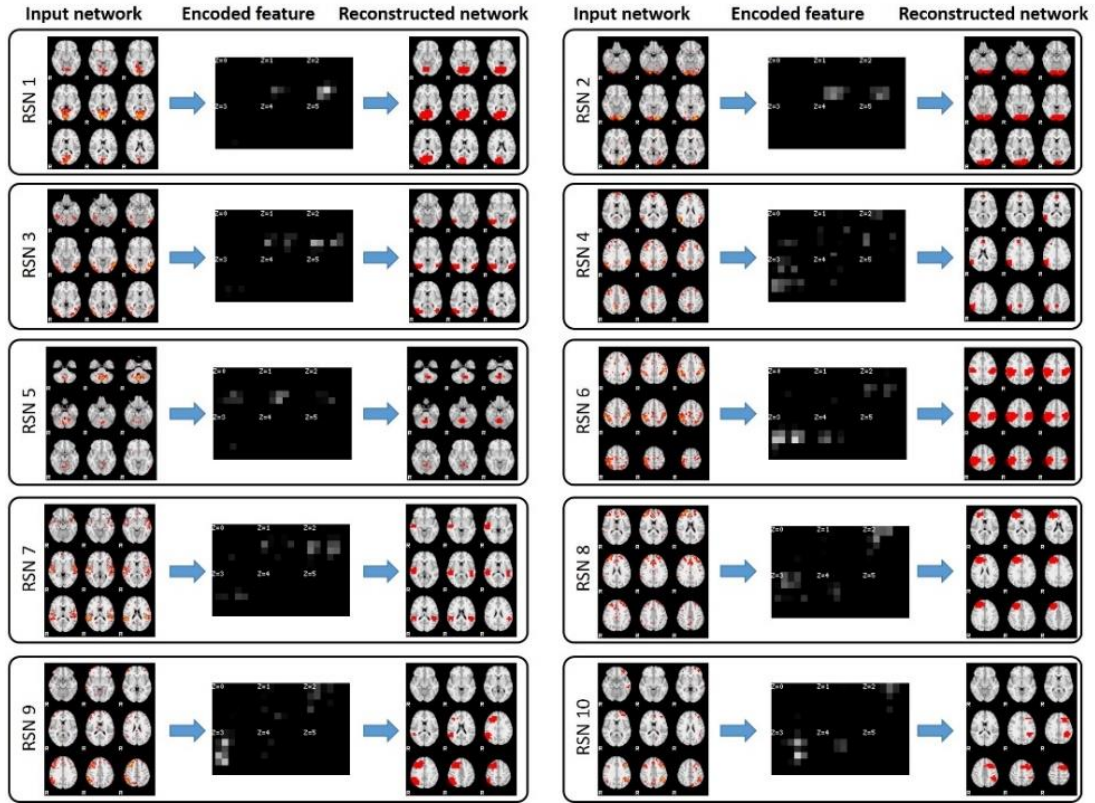


Fig. 2.18. Extracted features and reconstructed networks of 10 input networks corresponding to common RSNs. Each panel shows one RSN.

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. 2.19) were generated from the original 10 RSN clusters. As all 17 ICNs are derived from 10 resting state networks, the correctness of our results can be more conveniently validated, as follows.

Across clusters, ICNs show different spatial patterns as shown in Fig. 2.19. 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).

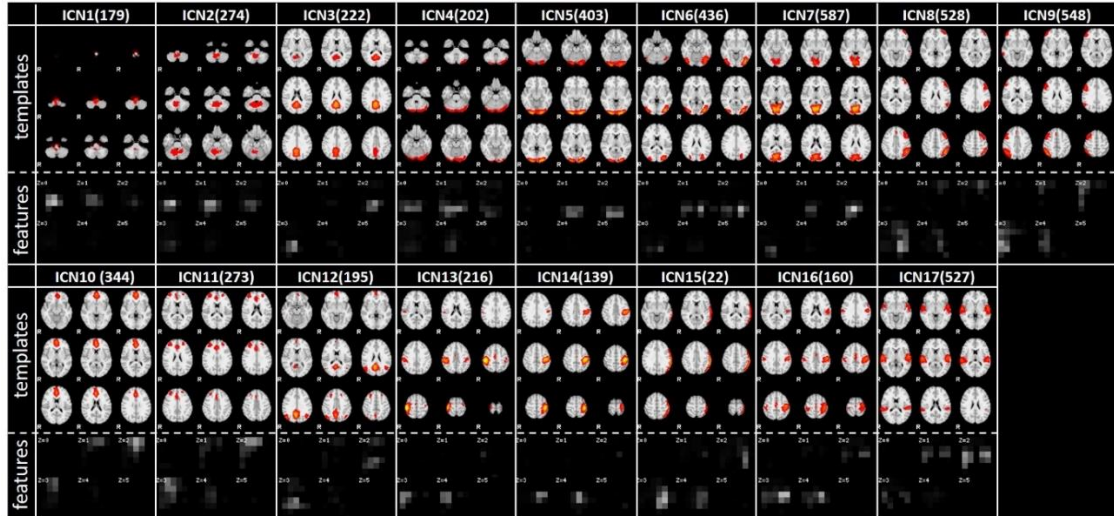


Fig. 2.19. 17 ICN atlases, together of their corresponding CAE extracted features, generated as the averages of the individual brain network maps within each cluster. The numbers of brain network maps in each cluster are also denoted at the top of each panel of ICN.

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. 2.20). 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.

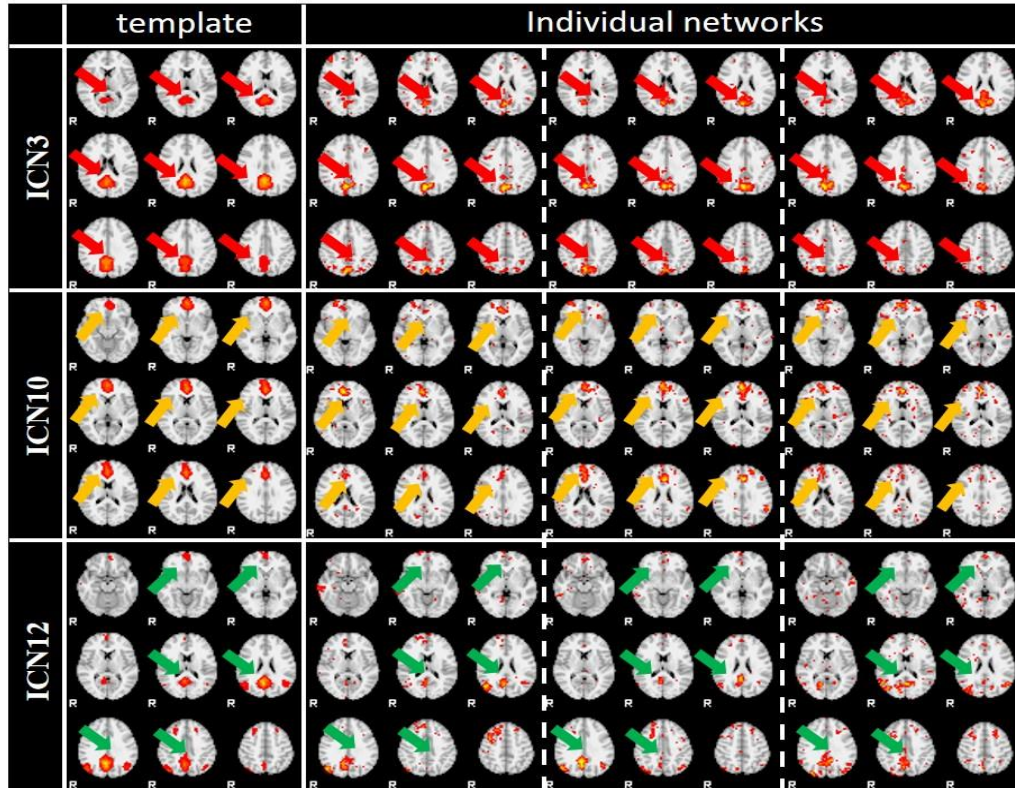


Fig. 2.20. Spatial pattern consistency within clusters. Arrows with the same color denote consistent spatial activation patterns.

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. 2.19) 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. 2.16) and 2 have labels of RSN 7 (refer to Fig. 2.16). 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.

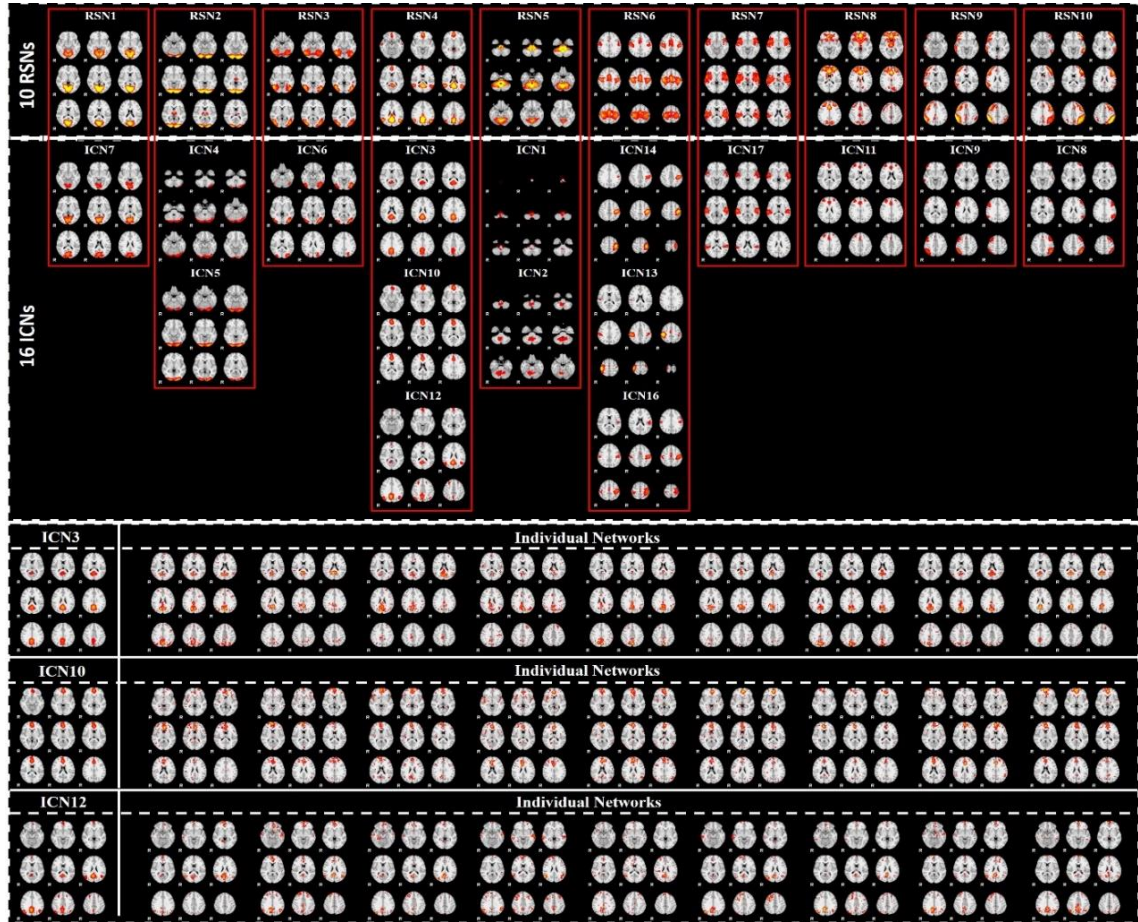


Fig. 2.21. Hierarchical sub-patterns of the 10 common RSN networks identified. The bottom panel showed individual network consistencies within each ICN cluster using DMN sub pattern ICNs.

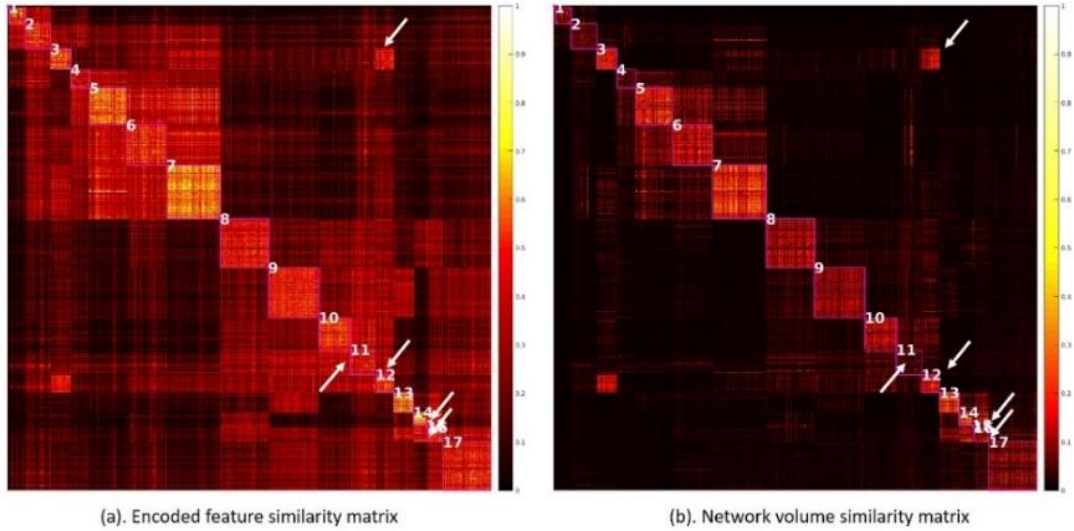


Fig. 2.22. (a) Clustered brain network feature similarity matrix; (b) clustered brain network overlap rate similarity matrix.

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. 2.21. 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. 2.22(a) and Fig. 2.22 (b), respectively, which are all rearranged in the cluster order from 1 to 17. As we can see from Fig. 2.22 (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. 2.22 (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. 2.19. These results demonstrated the superior spatial pattern modeling ability of our CAE framework to identify subtle but unneglectable network differences.

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. 2.23, 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.

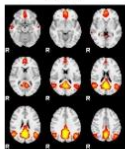
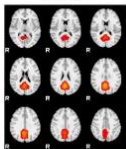
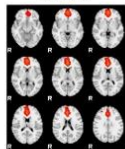
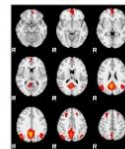
	DMN		ICN3		ICN10		ICN12	
								
Tasks	Count	Percent	Count	Percent	Count	Percent	Count	Percent
WM	38	17.12%	45	13.08%	11	5.64%		
SOCIAL	26	11.71%	47	13.66%	21	10.77%		
RELATIONAL	25	11.26%	51	14.83%	21	10.77%		
MOTOR	35	15.77%	45	13.08%	22	11.28%		
LANGUAGE	28	12.61%	35	10.17%	19	9.74%		
GAMBLING	26	11.71%	47	13.66%	20	10.26%		
EMOTION	32	14.41%	33	9.59%	28	14.36%		
RSN	12	5.41%	41	11.92%	53	27.18%		

Fig. 2.23. 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.

Discussion

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 the recently released HCP 900 subjects' data to train and evaluate our models. 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 the 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.

Inspired by this promising deep learning frameworks, our researches then shifted from traditional spatial map descriptor (e.g., connectivity map, spatial overlap, mutual information) to deep learning based spatial map feature extraction. As a lot of the well-established functional networks are generated using the previous data-driven scheme, we then proposed series of supervised deep learning framework for automatic brain functional network classification and recognition in section 2.6 and 2.6.

Acknowledgement

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) for this research. Z Kou was partially supported by National Institutes of Health (R21NS090153) for this research. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

2.5 Automatic Recognition of fMRI-derived Functional Networks using 3D Convolutional Neural Networks

Abstract

After introducing the data-driven researches in the previous sections, a supervised deep learning framework for automatic recognition of fMRI-derived function networks is proposed here.

Current fMRI data modeling techniques such as Independent Component Analysis (ICA) and Sparse Coding methods can effectively reconstruct dozens or hundreds of concurrent interacting functional brain networks simultaneously from the whole brain fMRI signals. However, such reconstructed networks have no correspondences across different subjects. Thus, automatic, effective and accurate classification and recognition of these large numbers of fMRI-derived functional brain networks are very important for subsequent steps of functional brain analysis in cognitive and clinical neuroscience applications. However, this task is still a challenging and open problem due to the tremendous variability of various types of functional brain networks and the presence of various sources of noises. In recognition of the fact that convolutional neural networks (CNN) has superior capability of representing spatial patterns with huge variability and dealing with large noises, in this section, we design, apply and evaluate a deep 3D CNN framework for automatic, effective and accurate classification and recognition of large number of functional brain networks reconstructed by sparse representation of whole-brain fMRI signals. Our extensive experimental results based on the Human Connectome Project (HCP) fMRI data showed that the proposed deep 3D CNN can effectively and robustly perform functional networks classification and recognition tasks, while maintaining a high tolerance for mistakenly labelled training instances. Our work provides a new deep learning approach for modeling functional connectomes based on fMRI data.

Introduction

Inferring functional brain networks from fMRI data has always been a popular method to better understand human brain functions. As a recapitulation before introducing proposed framework, typically, dozens or hundreds of concurrent functional brain

networks can be effectively and robustly reconstructed from whole-brain functional magnetic resonance imaging (fMRI) data of an individual brain using independent component analysis (ICA) (Beckmann and Smith, 2004; De Martino et al., 2007; Martin J McKeown et al., 1998; McKeown et al., 2003; Rolfe et al., 2009) or sparse representation (Jiang et al., 2015; Lv et al., 2015a, 2015d, 2015c, 2013a; Zhang et al., 2016; Zhao et al., 2015). For instance, by using the online dictionary learning and sparse coding algorithm (Mairal et al., 2010), several hundred of concurrent functional brain networks, characterized by both temporal time series and spatial maps, can be decomposed from either task-based fMRI (tfMRI) or resting-state fMRI (rsfMRI) data of an individual brain (Lv et al., 2015b). Pooling and integrating the spatial maps of those functional networks from many brains can significantly advance our understanding of the regularity and variability of brain functions across individuals and populations (Zhao et al., 2016). Recalling section 2.2, by clustering hundreds of thousands of functional brain networks from Autism Spectrum Disorder (ASD) patients and healthy controls, our recent work identified 144 group-wisely common intrinsic connectivity networks (ICNs) shared between ASD patients and healthy control subjects, where some ICNs are substantially different between the two groups (Zhao et al., 2016). In general, quantitative mapping of spatial maps of functional networks across individuals and populations offers a very powerful way to understand the brain functions in healthy brains and their alterations in brain disorders (Kennedy and Courchesne, 2008; Lv et al., 2015b; Zhu et al., 2014).

After several years of attempts at dealing with challenges when integrating, pooling and comparing spatial network maps across individuals and populations, we realized that the major challenge is the lack of ability to effectively describe spatial volume maps of

brain networks (section 2.2). As a result, developing a descriptive model that can sufficiently deal with spatial pattern variability of brain networks, as well as large noises, is the key towards automatic, effective and accurate classification and recognition of those large numbers of fMRI-derived functional brain networks. Fortunately, plenty of recent studies in the deep learning field have demonstrated that convolutional neural network (CNN) (De Martino et al., 2007; He et al., 2015a; Karpathy et al., 2014; Krizhevsky et al., 2012; Lawrence et al., 1997; Lecun et al., 1998; Nian Liu et al., 2015; Simonyan and Zisserman, 2014) has superior spatial pattern representation ability, e.g., as shown in many visual object recognition tasks (He et al., 2015b; Krizhevsky et al., 2012; Simonyan and Zisserman, 2014). Inspired by the tremendous successes of CNNs in automated, accurate spatial object recognition and their excellent ability of spatial pattern description, we design and employ a 3D CNN (Maturana and Scherer, 2015) framework for functional brain networks identification and recognition in this paper.

Specifically, in this section, an effective 3D CNN framework with two convolutional layers, one pooling layer and one fully connected layer, was designed for functional network map recognition. Then more than 5,000 manually labelled resting state networks (RSNs) (10 labels for 10 RSNs for each subject's fMRI scan session) derived from our HAFNI project (Lv et al., 2015a, 2015b) were utilized for training the deep 3D CNN. Afterwards, a series of experiments were performed to evaluate and compare the proposed 3D CNN framework for automatic recognition of fMRI-derived spatial RSN maps. Extensive experimental results showed that our designed 3D CNN's recognition accuracy is 94.61%, substantially higher than the accuracy achieved by using traditional methods such as the overlap rate. Our work demonstrated the superior capability of 3D

CNNs in dealing with various types of functional RSN maps. It is even surprising that 3D CNN can correct the wrongly labeled RSNs maps by human experts, significantly advancing the state-of-the-art methods and results reported in previous studies. In general, our proposed deep 3D CNN framework exhibited great robustness and effectiveness in functional network identification and recognition, contributing a new deep learning approach for modeling functional connectomes based on fMRI data in cognitive and clinical neuroscience.

Methods

Experimental Dataset

We still used HCP data set as introduced in Chapter 2.4. Similarly, In order to reduce computational burden, all of the functional RSNs maps were down sampled from the resolution of $91 \times 109 \times 91$ to $45 \times 54 \times 45$.

Computational Frameworks

An fMRI-oriented 3D CNN structure was designed for the problem of RSNs identification and recognition. The overall computational framework contains the two key steps, CNN training and testing (Fig. 2.24). Both training set and testing set were selected among the 5070 manually labelled RSNs. Specifically, 80% of the labelled data were randomly selected as the training set, while the remaining 20% were treated as the testing set. Detailed information of the training and testing RSNs is summarized in Table 2.1. In particular, balanced amount of dataset of each label for the training set was maintained to achieve a balanced training performance for each label (Mazurowski et al., 2008).

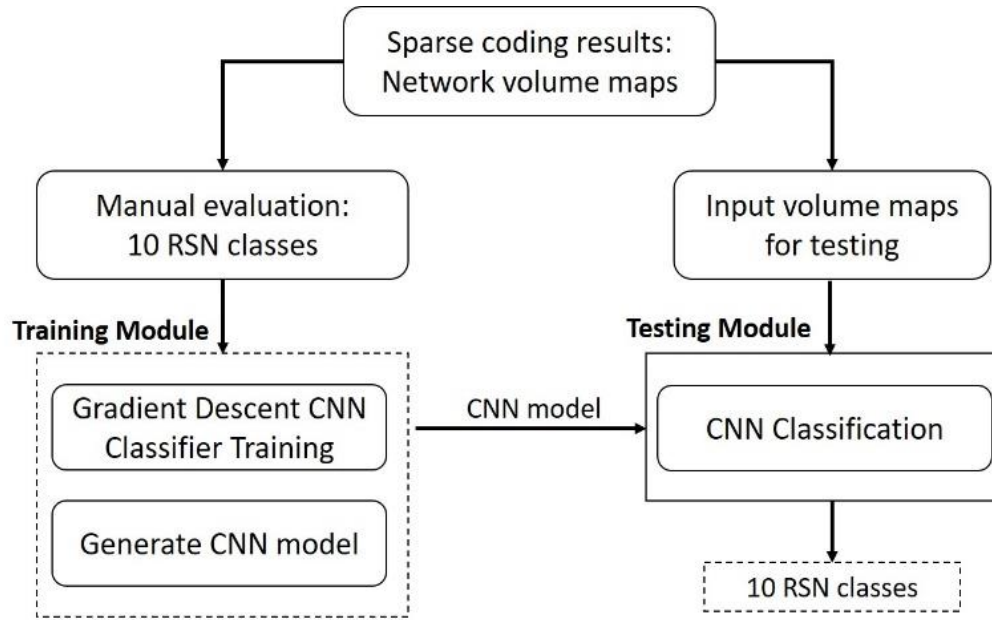


Fig. 2.24. Overview of the computational steps in the 3D CNN framework.

Table 2.1. RSNS NUMBERS OF TRAINING SETS AND TESTING SETS FOR 3D CNN TRAINING AND TESTING.

	RSN1	RSN2	RSN3	RSN4	RSN5	RSN6	RSN7	RSN8	RSN9	RSN10	total
Training	377	373	380	361	384	380	354	373	382	390	3754
Testing	149	154	146	176	146	140	175	155	141	139	1521

3D CNN Structure

Prior studies have shown that a hierarchy of useful features can be learnt from CNN deep learning models. Such learning models can be trained with either supervised or unsupervised approaches. However, many previous CNN-related researches are 2D-centric (Gupta et al., 2014), which might not be optimal for 3D volumetric image representation and could potentially overlook 3D structure information like in our application scenario of 3D RSNs recognition. In this work, we adopt and improve an effective fully 3D CNN framework (Maturana and Scherer, 2015) to train convolutional neural networks that aim

to classify and recognize RSNs reconstructed by dictionary learning and sparse coding methods. This powerful 3D convolutional architecture can well incorporate 3D structure information as intrinsic features, and effectively model the variability of the RSNs volume maps for classification and recognition, as demonstrated in the result section. Besides, the deep-layered nature of CNNs can effectively extract more abstract feature representation of the input RSN maps with deeper layers. These promising characteristics of 3D CNN make it suitable and ideal for automatic, effective and accurate classification and recognition of these large numbers of fMRI-derived functional brain networks. The proposed 3D CNN structure is summarized in Fig. 2.25. The detailed information of each layer and training procedure will be explained in the following sections, respectively.

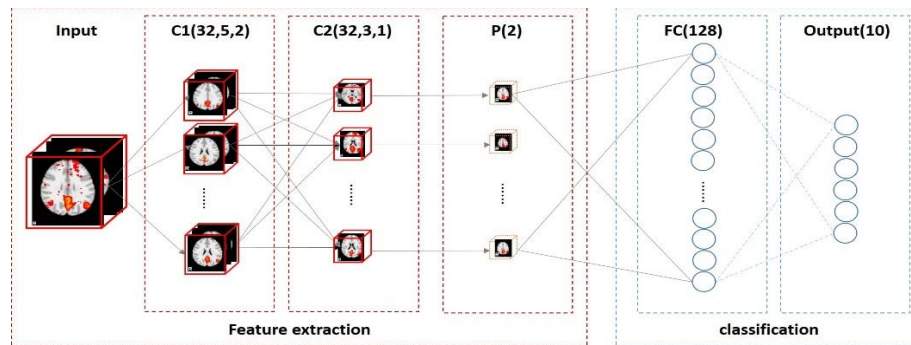


Fig. 2.25. Overview of the 3D CNN structure. Input map and output of each feature extraction layer are visualized. C1 represents convolutional layer 1, which contains 32 types of kernels or filters of size 5 with a stride step 2; C2 represents convolutional layer 2, which contains 32 types of kernels or filters of size 3 with a stride step 1; P represents pooling layer with pooling kernel size of 2; FC represents fully connected layer, with 128

nodes in this layer; Output layer contains 10 nodes representing each class of the 10 RSNs labels.

Convolutional layers: The convolutional layer of the CNN structure is denoted as $C(f,d,s)$, where f is the number of filters or kernels, also the number of feature maps after filtering; d is the size of the 3D filter; s is the stride step. Each convolutional layer is followed by a leaky rectified nonlinearity unit (ReLU) (Maas et al., 2013) with parameter 0.1, which is not shown in Fig. 2.25. for brevity. The initialization scheme of the convolutional layers was adopted from the methods in (He et al., 2015a). After the training stage, RSNs-specific filters were obtained for all the convolutional layers, as shown in

Fig. 2.26 and Fig. 2.27 for the purpose of visualization of filters in convolutional layer 1 and 2.

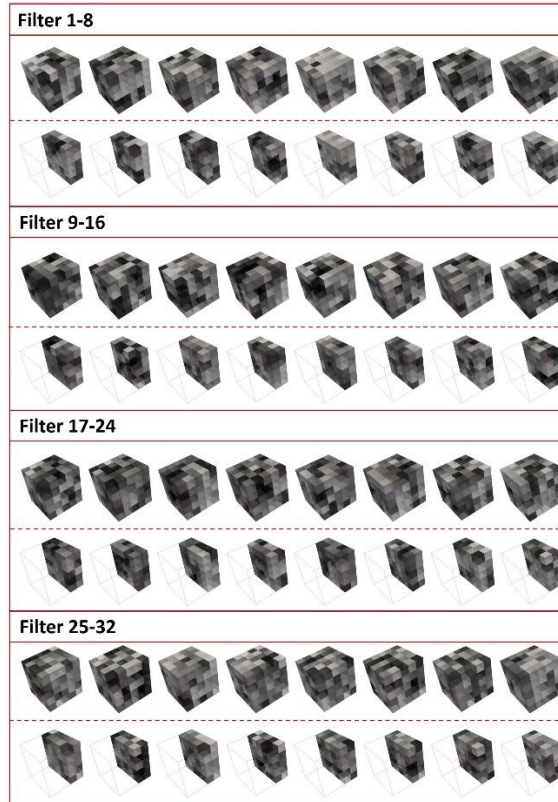


Fig. 2.26. Visualization of the trained RSNs-specific filters for convolutional layer 1. This layer contains 32 different types of filters, each of which has a size of $5*5*5*32$, with entire filters and clipped filters separated by a dashed line in the middle of each panel.

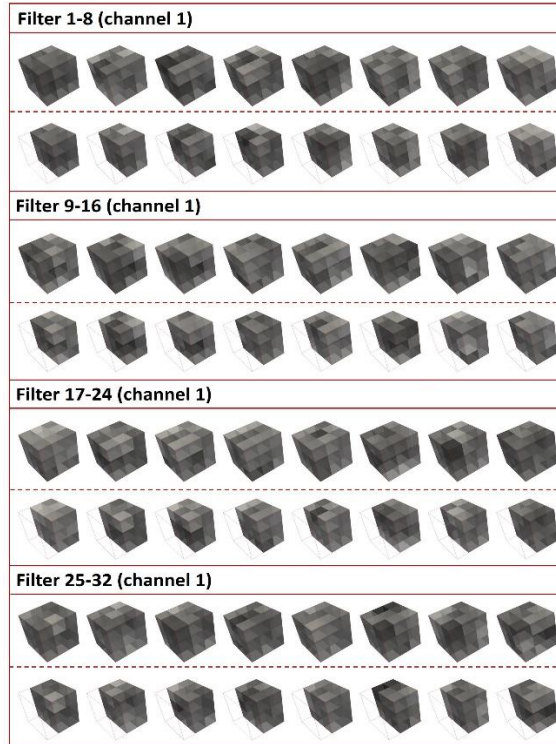


Fig. 2.27. Visualization of the trained RSNs-specific filters for convolutional layer 2. This layer also contains 32 different types of filters, each of which has a size of $3*3*3*32$ (32 input channels for this layers). Here, only the filters for the first channel of all the 32 are shown, with entire filters and clipped filters separated by a dashed line in the middle in each panel.

Notably, the input RSNs can be well represented using feature maps obtained by convolving with the well-trained filters. To demonstrate this point, an example of a default mode network (DMN) (Smith et al., 2009) as input is shown in Fig. 2.28 to illustrate the powerful feature extraction ability of the proposed 2-layered convolutional structure. Typically, DMN has 6 meaningful regions of interests (ROIs), which can be well captured

through different perspectives after convolutional layers (due to limited number of slices selected for visualization, only 4 ROIs are displayed in feature map1 and feature map2 in Fig. 2.28), as shown in Fig. 2.28.

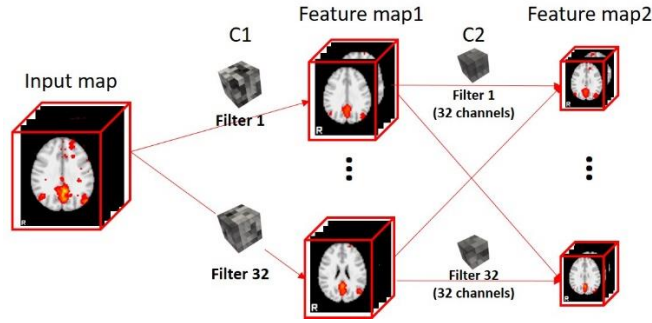


Fig. 2.28. Extracted feature maps using trained CNN layers with DMN as an input example.

Pooling layer: A pooling layer is connected to down sample the convoluted feature maps. This layer reduces the size of the input for the following classification layers, which substantially reduces redundant input information. Also, due to the translation-invariance properties of the pooling layer (Scherer et al., 2010), the global shift resulted by preprocessing steps (such as image registration) and the intrinsic shape and size variability of RSNs from different brains can be significantly alleviated and accounted for. This is one of the major advantages of using 3D CNN for automatic and robust recognition of RSNs, compared to other methods reviewed in the introduction section. In this paper, a max pooling scheme with pooling size of 2 was adopted and it turned out to work well.

Fully connected layer and output layer: These two layers are functioning as the classification/recognition component in the overall 3D CNN framework. With well-extracted features as input, 128 nodes of the fully connected layer can effectively perform the classification task. The output layer contains 10 nodes, each of which predicts the corresponding RSN label probability for each input map by adopting the SoftMax action function.

CNN training: The neural network weights training was performed by the classic Stochastic Gradient Descent (SGD) with momentum. The objective loss function to be optimized is the multinomial negative log-likelihood with a λ (set to 0.01) times the L_2 norm of the network weights as regularization term, as shown in equation 2.7.

$$L(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k \mathbf{1} \cdot \{y^i = j\} \log(\theta^T x^i)_j + \lambda \|\theta\|_2 \quad 2.7$$

where m is the number of samples in one batch (empirically set to 32), and k is the number of the output classes (10 output RSN classes) and $\log(\theta^T x^i)_j$ is the log-likelihood activation value of the j_{th} output node. The momentum parameter was set to 0.9. In this work, the widely-used dropout technique was adopted for each layer during the training process to reduce the overfitting problem that may be caused by large amount of weights to be trained and to reduce testing errors (Bell and Koren, 2007; Krizhevsky et al., 2012). The convolutional layers were initialized using the similar scheme proposed in (He et al., 2015a), and the dense layers were initialized with a Gaussian distribution with $\mu = 0, \sigma = 0.01$. Training was performed by utilizing GPU (Nvidia Quadro M4000) for 80 epochs. For all the 3754 training RSNs samples, the total training time is less than 20 minutes. This scale of training time cost makes the proposed 3D CNN framework very suitable for future cognitive and clinical neuroscience applications.

Results

In this section, a variety of experiments and comparisons are conducted to evaluate the performance of the proposed 3D CNN framework. Traditional automated RSN identification method using overlap rate was performed in comparison to our proposed RSN identification framework using CNN structure. The overlap rate is calculated according to equation 2.8.

$$\text{overlap rate} = \sum_{k=1}^{|V|} \frac{\min(V_k, W_k)}{(V_k + W_k)/2} \quad 2.8$$

where V_k and W_k are the activation score of voxel k in RSN volume maps V and W , respectively.

For 1521 testing RSN samples, based on the originally manually labelled RSNs, 94.61% accuracy was achieved by using the proposed 3D CNN framework. In contrast, only 85.93% accuracy was achieved by overlap rate. Among the 5.39% CNN-based error rate (82 testing errors) and 14.07% overlap-based error rate (214 testing errors), there were 4.4% (67 testing case errors) in common. Overall, CNN classification results significantly outperformed overlap-based results by round 10%. The promising results indicated the powerful spatial description ability of CNN. Through the detailed analysis and visualization of CNN classification error patterns in the following sub-sessions, we will further demonstrate that our designed CNN framework has the ability of accommodating major distributions of the training samples and ignoring outliers in the training samples, thus correcting the wrongly labelled RSNs due to the manual labelling mistakes. For the rest of the sub-sections, the 10 common RSNs templates derived from our HAFNI project (Lv et al., 2015b) are visualized in Fig. 2.29 and will be used as a common spatial reference for evaluations and comparisons.

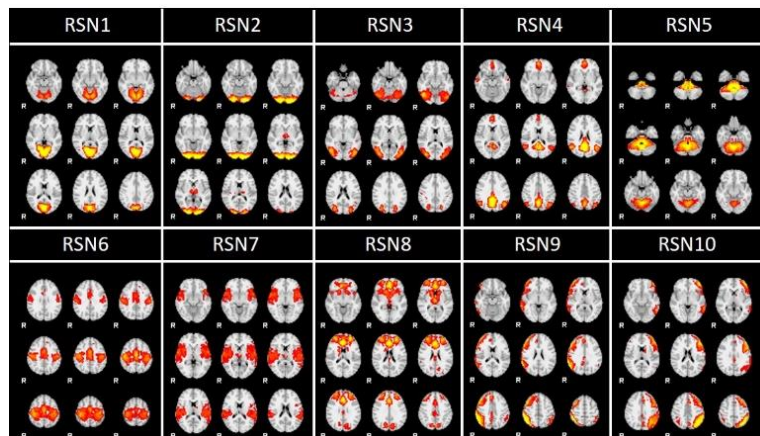


Fig. 2.29. Visualization of 10 common RSNs templates derived in the HAFNI project. For each RSN, 9 most informative slices are displayed.

Correction of Wrongly Manually-Labeled RSNs by 3D CNN

Among a large portion of CNN-based RSN classification errors (82 in total), there are actually testing cases that were originally wrongly labeled by experts. For each of these 10 RSNs we selected one representative example of CNN classification error for visualization in Fig. 2.30 to demonstrate the CNN’s ability of manual label correction. In this case, the real meaning of “wrong” CNN classification is that its prediction does not agree with the expert’s manual labeling. Therefore, if this scenario is double-checked and confirmed, CNN’s prediction can be used to correct the originally wrongly labeled RSNs by expert. As shown in Fig. 2.30, CNN classified labels appear to be more reasonable than the original manual labels, which has been confirmed by separate senior experts other than the original experts. In addition, quantitative measurement of the probability (the softmax values of the output layer) of correct labeling by both CNN and original manual labeling is provided for each representative case on the top of each figure panel in Fig. 2.30. Among all the 82 testing cases with CNN’s “wrong” classifications, 63 of them are considered as CNN’s corrections of original wrongly-labeled RSNs (see Table 2.2. for detailed numbers for each of 10 RSN types), while still 15 of them are remained controversial. For the visualizations of all of 82 CNN prediction errors, please refer to <http://hafni.cs.uga.edu//CNNClassification/errorCheckCNNerrorAll/index1.html>.

Table 2.2. CNN corrections of original wrong labels for each of the 10 RSN types.

	RSN1	RSN2	RSN3	RSN4	RSN5	RSN6	RSN7	RSN8	RSN9	RSN10	total
Wrongly labelled	2	3	9	30	5	6	3	7	1	0	63

It is noted that the original expert manual labelling of 10 RSNs was based on 2D visualization of RSNs' volume slices as shown in Fig. 2.30. In this study, to double-check and confirm the CNN's corrections of those wrongly manually-labelled RSNs, we conducted a more informative 3D visualization of those RSNs using input map 1 in Fig. 2.30 together with RSN 1 and RSN 2 as illustration examples, as shown in Fig. 2.31. It is evident that the CNN's predicted labels truly to be more reasonable than the original manual labels.

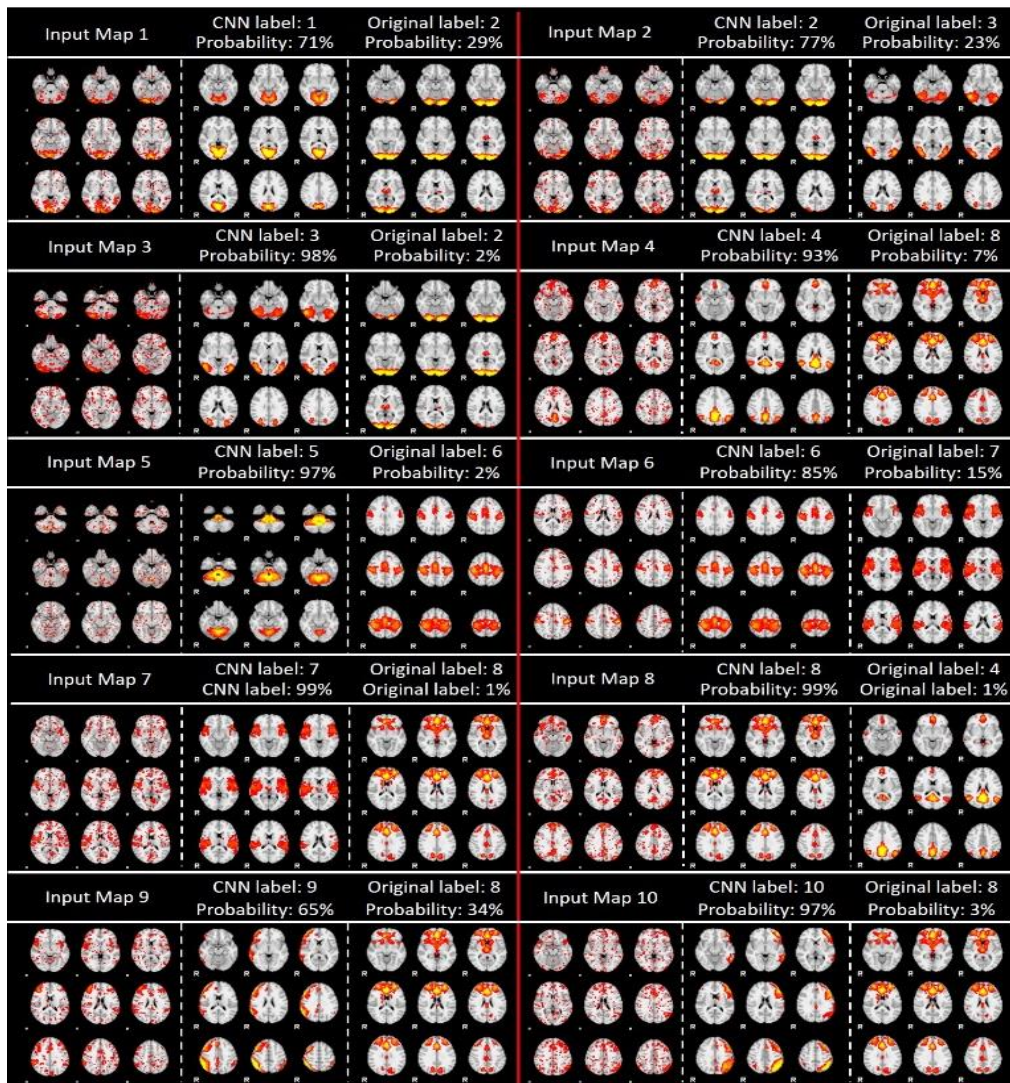


Fig. 2.30. Representative cases of CNN classification errors for 10 RSNs. CNN's predicted labels (denoted as CNN label in the figure) appear to be more reasonable than the original manual labels by experts.

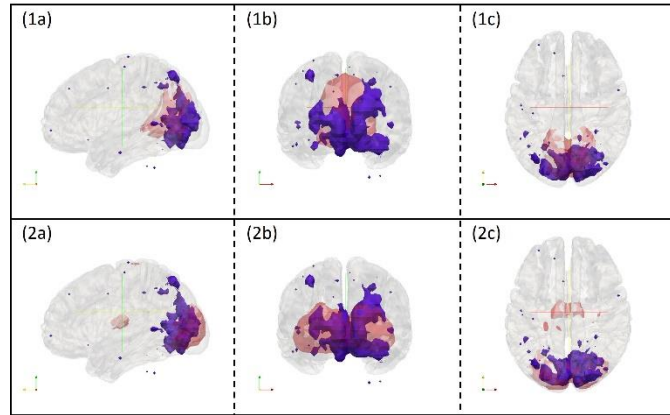


Fig. 2.31. 3D visualizations of input map 1 in figure 8 with RSN templates overlaid. Subfigure (1a), (1b) and (1c) are input maps (blue regions) with RSN1 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Subfigure (2a), (2b) and (2c) are input maps (blue regions) with RSN2 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively.

Overlap-based Classification Error Analysis

In comparison with 82 cases of CNN's disagreements with original manual labels, overlap-based method has 214 disagreements with original manual labels. It turns out that overlap-based method shares common disagreements with original manual labels with the CNN method (to be detailed in section 3.3), and overlap-based method can be more reasonable than the original manual label in some cases. For example, in Fig. 2.32, for RSN map 1, 3, 6, overlap-based prediction labels seem to be more reasonable. However, in many cases, manual labels are more reasonable. As shown in Fig. 2.32, for RSN map 2, 3, 4, 10, overlap-based method had made obviously less reasonable predictions. Among those 214

disagreements with original manual labels by overlap-based methods, 89 of them are believed to be truly wrong classification. In this sense, CNN method certainly significantly outperforms overlap-based method. All the visualizations of the 214 overlap-based predictions can be found on <http://hafni.cs.uga.edu/CNNClassification/errorCheckOverlapAll/index1.html>.

As examples, Fig. 2.33 confirms that for input map 2 in Fig. 2.32, overlap-based method really made wrong classification. This wrong classification might be caused by a variety of reasons, among which spatial registration, alignment error and noise sources could be a major issue; Fig. 2.34 confirms that for input map 4 in Fig. 2.32, overlap-based method made unreasonable classification. This type of wrong classification is due to the intrinsic heterogeneous activities of intermixed neurons in the same brain region or voxel (Harris and Mrsic-Flogel, 2013).

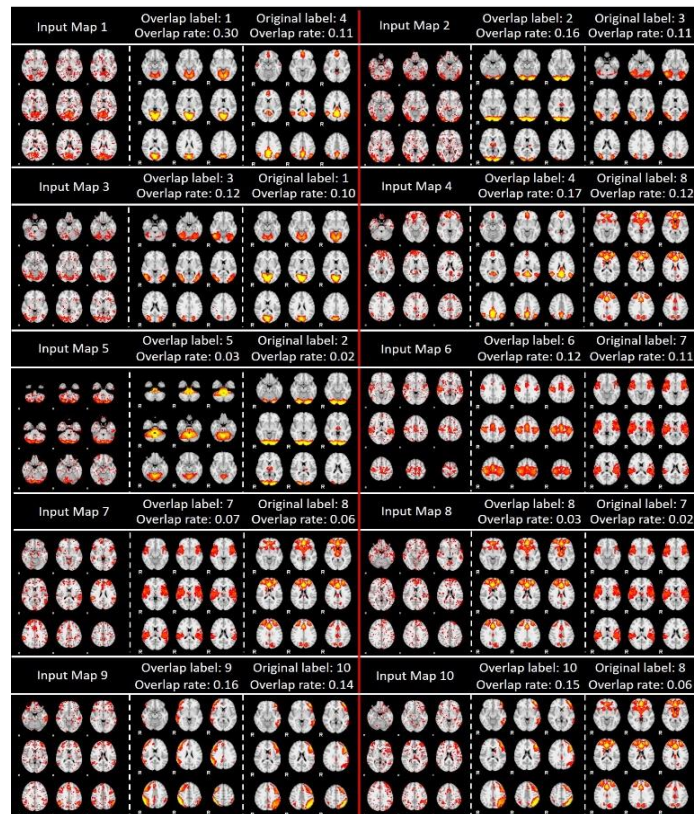


Fig. 2.32. Representative cases of overlap-based classification results for 10 RSNs.

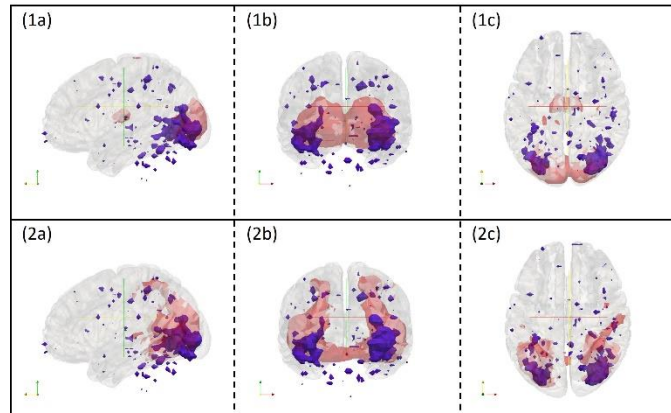


Fig. 2.33. 3D visualizations of input map 2 in figure 10 with RSN templates overlaid. Subfigure (1a), (1b) and (1c) are input maps (blue regions) with overlap-based classification of RSN2 template overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Subfigure (2a), (2b) and (2c) are input maps (blue regions) with RSN3 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. RSN3 should be the category for this input map, but overlap-based method gave the wrong result as RSN2.

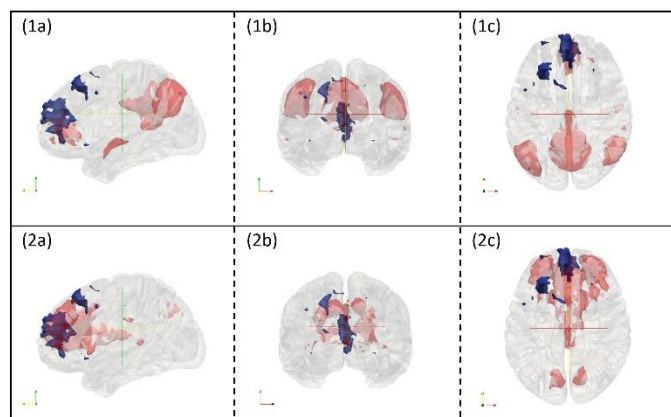


Fig. 2.34. 3D visualizations of input map 4 in figure 10 with RSN templates overlaid. Subfigure (1a), (1b) and (1c) are input maps (blue regions) with overlap-based

classification of RSN4 template overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Subfigure (2a), (2b) and (2c) are input maps (blue regions) with RSN8 overlaid (red regions). They are displayed with cross sections along the x, y and z axes respectively. Due to that the major activation regions reside in the prefrontal lobes, RSN8 should be the category for this input map, but overlap-based method gave the wrong result.

Common Disagreements with Manual Labels by CNN and Overlap-based Method

Our experiment results show that CNN and overlap-based method share 67 common disagreements with the original manual labels. Interestingly, all of these 67 RSN maps tend to be manually assigned with wrong labels. Among the 67 classifications, 64 of them were predicted with the same label using both CNN and overlap-based method. However, there are other 3 RSNs that have different annotations from manual labelling, CNN and overlap-based method, and these 3 RSNs and their different labels by three methods are visualized in Fig. 2.35. As we can see, among the 3 CNN labels, the RSN map 2 has the highest CNN prediction probability (99%), while the other two RSN maps have relatively low probabilities. From visual inspections, we really cannot tell which classes the input RSN map 1 and 3 should belong to. This problem might be caused by the relative low quality of the input RSN maps. Nevertheless, it is certain that input RSN map 2 should be assigned with label 7, which means the high prediction probability provided by CNN is quite reliable. For all the 67 common disagreements with the original manual labels, please refer to <http://hafni.cs.uga.edu/CNNClassification/errorCommon/index1.html>

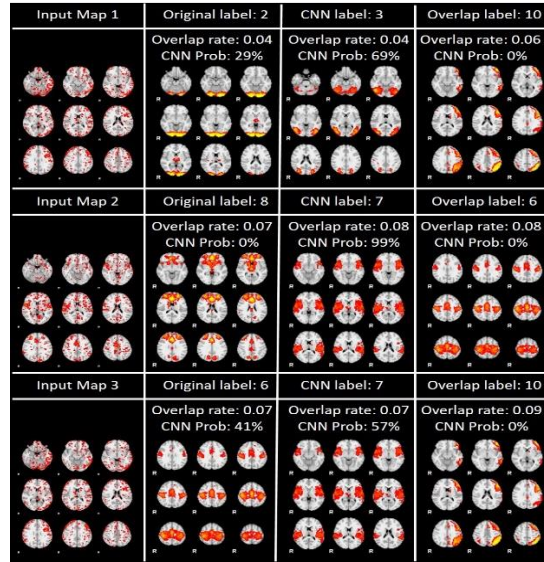


Fig. 2.35. 3 RSN classification discrepancies between CNN and overlap-based method in common prediction disagreement with the original manual labels.

Differentiation between Highly Spatially Overlapped RSNs

fMRI signal from each voxel reflects a highly heterogeneous mixture of functional activities of the entire neuronal assembly of multiple cell types in a voxel. In addition to the heterogeneity of neuronal activities, the convergent and divergent axonal projections in the brain and heterogeneous activities of intermixed neurons in the same brain region or voxel demonstrate that cortical microcircuits are not independent and segregated in space, but they rather overlap and interdigitate with each other (Harris and Mrsic-Flogel, 2013). Thus spatial overlap of functional networks including RSNs is a natural property of functional organization of the human brain (Xu et al., 2016). In this paper, among the 10 RSNs specifically, template RSN2 and RSN3, RSN1 and RSN2, RSN4 and RSN8, RSN1 and RSN4 have relatively high spatial overlap rates (0.1665, 0.1550, 0.1201 and 0.1062, respectively), which made the overlap-based method difficult to differentiate those pairs of highly overlapping RSN patterns. Four examples of such cases have been shown in Fig.

2.36 to demonstrate the advantages of CNN over overlap-based method when differentiating highly overlapped spatial patterns. More such examples can be found on the webpage showing CNN's only disagreements with manual labels (<http://hafni.cs.uga.edu/CNNClassification/errorCheckOnlyCNN/index1.html>) and overlap-based only method's disagreements with manual labels (<http://hafni.cs.uga.edu/CNNClassification/errorCheckOnlyOverlap/index1.html>).

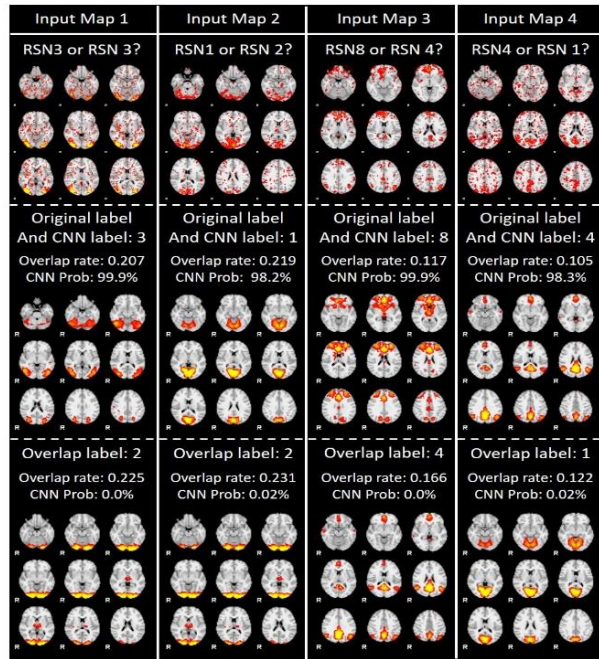


Fig. 2.36. Illustration of CNN and overlap-based method's performance when differentiating input RSN maps of high spatial overlaps.

Discussion and Conclusion

The HAFNI framework has enabled connectome-scale reconstruction of reproducible and meaningful functional brain networks on large-scale populations such as the HCP datasets. However, an unsolved problem in the HAFNI framework is the automatic recognition of HAFNI maps such as RSNs in each individual brain. The major problem in previous traditional methods is that they are not able to deal with the

tremendous variability of various types of functional brain networks (e.g. size, shape and location) and the presence of various sources of noises. In this study, we continued to propose and apply a fully automatic 3D CNN deep learning framework to identify and classify different types of functional brain networks with promising performance in a supervised way. Our experimental results showed a promising classification accuracy of 94.61% by improving approximately 10% compared to overlap-based method. Furthermore, in the result subsections, we conducted comprehensive analysis of the disagreement patterns of CNN labels with manual labels, as well as the overlap-based method's labels. Again, our results demonstrated the superior performance of CNN in recognizing ambiguous RSNs, spatially overlapping RSNs, and misaligned RSNs. In general, our work provides a new deep learning approach for modeling and recognizing functional connectomes based on fMRI data, particularly fMRI big data in the future.

Despite the great promise of the proposed CNN framework, however, there also exist challenges and limitations for the current CNN framework. First, the training sample preparation is a difficult issue for training the CNN networks. As we can see, manually labelled RSNs derived from our previous HAFNI project were used in this study, which entailed huge amount of time devoted to manually labeling dozens of thousands of functional network maps, among which thousands of them are RSNs. Also, manual labeling mistakes and inter-rater variability of labels are inevitable. Though our CNN framework already exhibits the promising property of correcting wrongly manually labeled RSNs, as shown in the result sections, a reliable and fully or semi- automated network labelling method should be explored in the near future to enlarge the training samples and improve the training accuracy. Second, the problem of 10 RSN classifications was

employed in this study for experiment setup, which was just a testbed and showcase for the efficiency of our CNN framework. In the next section, we will develop and use larger scale training sample generation and build a CNN model for classifications and recognitions of many more types of functional networks such as hundreds of networks that were already revealed in our HAFNI project. Third, other advanced or sophisticated CNN structures, e.g., multi-scale CNN (Nian Liu, Junwei Han, Tianming Liu, 2016) or truly deep CNN (He et al., 2015b; Simonyan and Zisserman, 2014; Szegedy et al., 2015), will be explored in the near future. It is expected that these improved CNN structures will possess better ability of spatially representing 3D networks maps and thus will further generate better network classification results. Finally, we plan to adopt and apply these effective CNN frameworks on clinical fMRI datasets for the better understanding of altered brain networks in brain diseases such as Alzheimer's disease and Autism. We envision that 3D CNN model will significantly advance current state-of-the-art fMRI data modeling approaches and pave the way for adopting fMRI into clinical management of brain disorders in the future.

2.6 Automatic Recognition of Holistic Functional Brain Networks Using Iteratively Optimized Convolutional Neural Networks (IO-CNN) with Weak Label Initialization

Abstract

Interpretations of those decomposed networks are still open questions due to the lack of functional brain atlases, as mentioned above in section 2.5. Even though recent studies showed that deep learning, especially deep convolutional neural networks (CNN), has extraordinary ability of accommodating spatial object patterns, e.g., our recent works using

3D CNN for fMRI-derived network classifications achieved high accuracy with a remarkable tolerance for mistakenly labeled training brain networks, the training data preparation is one of the biggest obstacles in these supervised deep learning models for functional brain network map recognitions, since manual labelling requires tedious and time-consuming labors which will sometimes even introduce label mistakes. Especially for mapping functional networks in large scale datasets such as hundreds of thousands of brain networks used in this paper, the manual labelling method will become almost infeasible. In response, in this work, we tackled both the network recognition and training data labelling tasks by proposing a new iteratively optimized deep learning CNN (IO-CNN) framework with an automatic weak label initialization, which enables the functional brain networks recognition task to a fully automatic large-scale classification procedure. Our extensive experiments based on ABIDE-II 1,099 brains' fMRI data showed the great promise of our IO-CNN framework.

Introduction

With the availability of well reconstructed functional brain networks across different individual brains using the current methods, the next step would be to model, interpret and use them in a neuroscientific meaningful context. In this section we are still dealing with the challenges caused by the random initialization nature of the decomposition algorithms in ICA, sparse representation or other methods, together with the variability and heterogeneous characteristics of human brains, incurring no guarantee for the the correspondences of the decomposed networks across different brains, which results in the problems for group level modeling and statistically meaningful analysis for the obtained networks (Lv et al., 2013b; Smith et al., 2009; Zhao et al., 2016). Though the connectivity

map model has achieved promising results (Zhao et al., 2016), it is still noticeable that dealing with the tremendous variability of various types of functional brain networks and the presence of various sources of noises is still very challenging, due to the limited ability of the model itself to describe various spatial pattern distributions.

As discussed in all the previous sections, the lack of effective spatial volume map descriptors has been realized as a major challenge for all functional networks analysis related research studies, such as integrating networks atlases and functional network recognition tasks. The lack of functional brain atlases and no correspondence across decomposed or reconstructed networks across different subjects are considered as a big challenge for brain network classification and recognition. In our previous work (Zhao et al., 2017b) in section 2.5, an effective 3D CNN framework was already designed and applied to achieve effective and robust functional network identification by both accommodating manual label mistakes and outperforming traditional spatial overlap rate based methods.

However, even though our 3D CNN model with two convolutional layers, one pooling layer and two fully connected layers for functional network map recognition has successfully achieved high accuracy in functional networks recognition (section 2.5), the classification category number is only limited to 10 most common networks. More importantly, the around 5,000 training networks preparation went through a very tedious inter-expert manual label work procedure over 210,000 functional networks. As indicated in (Zhao et al., 2017b), due to the highly interdigitated and spatially overlapped nature of the functional brain organization (Xu et al., 2016), the manual labelling process will inevitably introduce some mistakes into the training set labels. As model goes deeper and

deeper, sufficient amount of data is needed for training the model (LeCun et al., 2015). As a result, the accurate labelling and the fast labelling process become the major need for training a deep learning model of functional brain networks.

In response to address the abovementioned large-scale training data preparation problem and extend the previous 10-class CNN-based recognition to large-scale functional network recognition, a novel iteratively optimized CNN (IO-CNN) framework is proposed here, while the initialization network labels were roughly assigned using automatic overlap rate based label. The training dataset is based on 219,800 functional networks decomposed from 1099 subjects' fMRI data provided by publicly available database ABIDE II (http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html). After training, an accurate functional network recognition and identification based on the IO-CNN for 135 functional network atlases will be achieved. Extensive experiments on the ABIDE-II 1,099 brains' fMRI data has demonstrated that the proposed IO-CNN framework has superior spatial pattern modelling capability in dealing with various types of network maps, and the iterative optimization algorithm can gradually accommodate the mistake labels introduced by the fully automatic but rough label initialization, eventually converging to a fine-grained classification accuracy. In general, the automatic rough label initialization together with the iterative optimization framework provides novel and deep insight in training and applying large scale deep learning networks with weak label supervision, contributing to the general field of deep learning for medical imaging.

Materials and Methods

Dataset and preprocessing

Our experimental data were downloaded from the publicly available Autism Brain

http://fcon_1000.projects.nitrc.org/indi/abide/abide_II.html). The ABIDE was established for discovery science on the brain connectome in autism spectrum disorder (ASD). ABIDE I already demonstrated the feasibility and utility of aggregating fMRI data across different sites, while ABIDE II was promoted further for that purpose with larger scale of datasets. To date, ABIDE II involves data from 19 sites with 521 ASD patients and 593 controls (5-64 years old). After manually checking data quality according to preprocessing (e.g. skull removal, registration to standard space) results, only 511 ASD patients and 588 control subjects are selected, which were then used in our following experiments. The acquisition parameters vary across different sites: 190 - 256 mm FOV, 31 - 50 slices, 0.475 – 5.4 s TR, 24 – 86 ms TE, 60 - 90 ° flip angle, (2.5 - 3.8) × (2.5 - 3.8) × (2.5 - 4) mm voxel size. For detailed parameters for each site, please refer to the ABIDE II website.

Preprocessing for the resting state fMRI (rsfMRI) data were performed using the FSL software tools (Jenkinson et al., 2012), including skull removal, motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and linear registration to the Montreal Neurological Institute (MNI) standard brain template space, which were all implemented by the FSL FLIRT and FEAT commands.

After preprocessing, we exploited dictionary learning and sparse coding techniques (Lv et al., 2015a, 2015b) for functional brain networks reconstruction for each subject. The input for dictionary learning is a matrix $X \in \mathfrak{R}^{t \times n}$ with t (length of time points) rows by n columns containing normalized (normalizing signals to 0 mean and standard deviation of 1) fMRI signals from n brain voxels of an individual subject. The output contains one learned dictionary D and a sparse coefficient matrix $\alpha \in \mathfrak{R}^{m \times n}$, w.r.t, $X = D \times \alpha + \varepsilon$, where

ε is the error term and m is the predefined dictionary size. Each row of the output coefficient matrix α was then mapped back to the brain volume space as a 3D spatial map of functional brain network. According to (Lv et al., 2015a; Zhao et al., 2016), dictionary size m was empirically set to 200 for a comprehensive functional brain networks reconstruction, which means each individual fMRI data will have 200 decomposed functional networks.

3D CNN Model

As demonstrated in (Zhao et al., 2017b), 3D convolutional neural network (CNN) has promising performance in modeling 3D spatial distribution patterns and correcting outliers by modeling each class's spatial distribution, especially for functional brain maps. Thus, in this work, we adopted and modified the previous CNN structure (Zhao et al., 2017b) to a new 3D CNN net (shown in Fig. 2.37) with 3 convolutional blocks for feature extraction, and two fully connected layers for classification. Specifically, each convolutional block contains 3 layers: 1) convolutional layer with rectified nonlinearity unit (ReLU) as activation function (Maas et al., 2013), which is not shown in Fig. 2.37 for brevity. The initialization scheme of the convolutional layers was adopted from the methods in (He et al., 2015a); 2) A pooling layer is connected to the convolutional layers. This layer reduces redundant input information and introduces translation-invariance characteristics (Scherer et al., 2010), which is aimed to alleviate the possible global shift resulted by image registration and the intrinsic variability of different individual brains. In this paper, a max pooling scheme with a pooling size of 2 was adopted and it turned out to work quite well (Zhao et al., 2017b); 3) Batch normalization layer is introduced later to

enable faster learning and remove internal covariate shift of each mini-batch (Ioffe and Szegedy, 2015).

The feature extraction part will be accomplished by the 3 convolutional blocks and herein comes with the following fully connected layers for classification purpose. As illustrated in Fig. 2.37, only one hidden layer with 128 nodes were inserted between the output nodes and the extracted feature maps from the convolutional blocks. A softmax function is then applied to obtain the probability of the final predictions.

The loss function used for this canonical multi-class classification problem is the categorical cross entropy 2.9.

$$L(\theta) = -\frac{1}{n} \sum_{i=1}^n \sum_{j=1}^k 1\{y^i = j\} \log(\theta^T x^i)_j \quad 2.9$$

where n is the number of samples in one batch (empirically set to 20), k is the number of the output classes (144 output classes), and $\log(\theta^T x^i)_j$ is the log-likelihood activation value of the j_{th} output node. The optimizer used for the backpropagation is the advanced ADADELTA (Zeiler, 2012). θ is the network structure weights matrix. y^i is the training label.

As the input volumes have dimensions of $48 \times 64 \times 48$ and the full volume with a batch size of 20 was used, 4.2GB memory will be consumed on the GPU card for training the proposed 3D CNN model.

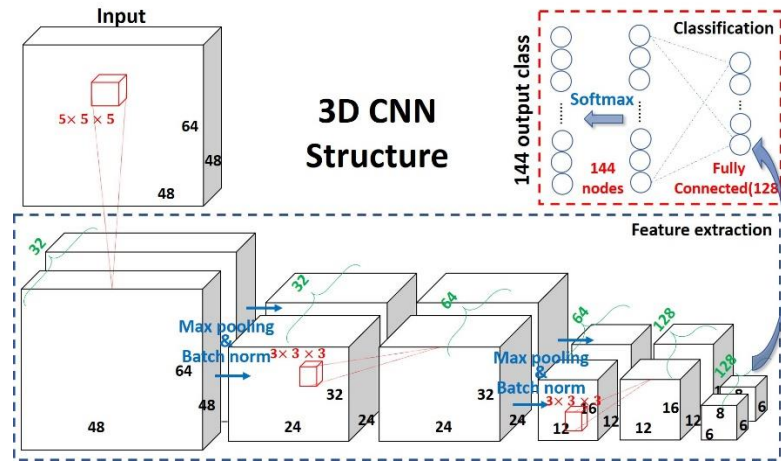


Fig. 2.37. 3D CNN structure used for training and classification. Bold numbers indicated the feature map sizes (e.g. 48, 64, 48); Red numbers indicated convolutional kernel size (e.g. $3 \times 3 \times 3$) and nodes number of fully connected layers (e.g. 128, 144); Green numbers indicated the channel size of each feature map.

Iteratively Optimized CNN (IO-CNN) with weakly label initialization

Unlike the previous project, we do not simply train the inherited and improved CNN mentioned in section B using the same training dataset. Instead, we proposed a new Iteratively Optimized CNN (IO-CNN) to iteratively train on dynamically updated training sets. Specifically, the initial training dataset is generated using a fast and automatic, but weak, labelling process by utilizing the maximal spatial overlap rate scheme, as mentioned in (Zhao et al., 2017b). Notably, this automatic labelling scheme can still achieve around 85% accuracy on prediction, providing a rough but fairly good enough training initialization.

The classification labels are the 144 functional network atlases generated in (Zhao et al., 2016), by using the ASD and Typical Control (TC) populations (from ABIDE I dataset). Briefly, the 144 functional network atlases are generated from the ABIDE I dataset by utilizing a clustering scheme based on spatial overlap rate metrics, which are

assisted and accelerated by the proposed connectivity map in (Zhao et al., 2016). Based on the 144 functional network atlases as classification labels, the larger scale dataset of ABIDE II was then used for the recognition and identification tasks. Fig. 2.38 described the logic relationship and connection between our previous work of generating 144 functional atlases from ABIDE I data in (Zhao et al., 2016) and current work of weakly labeling the reconstructed functional networks from ABIDE II data.

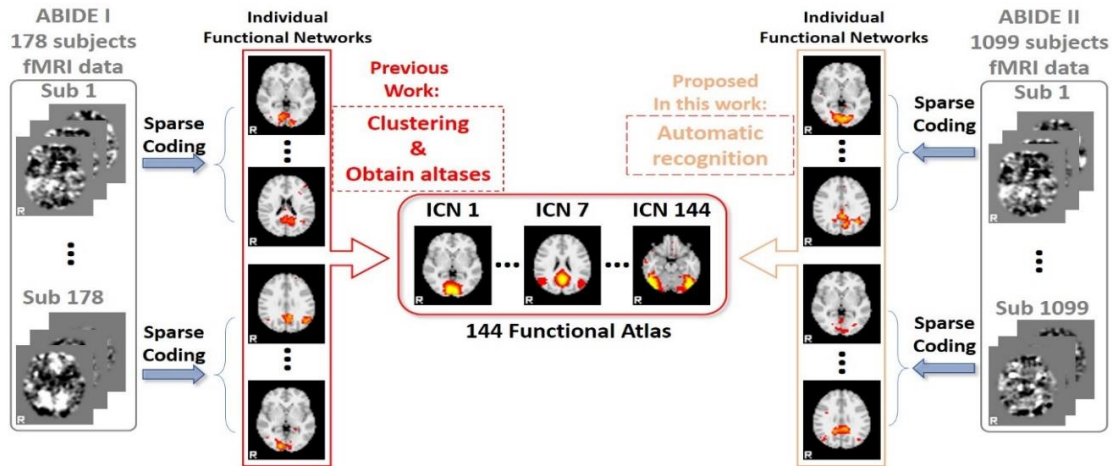


Fig. 2.38. Logic connection between the atlases generation based on ABIDE I data in (Zhao et al., 2016) and the functional network recognition work based on these atlases using the ABIDE II data in this paper. Both ADIBE I and II fMRI datasets were decomposed to functional networks using sparse coding, where networks from ABIDE I were utilized for generating atlases, while the functional networks from ABIDE II were decomposed for atlases-based recognition via a weak labeling process based on spatial overlap rate.

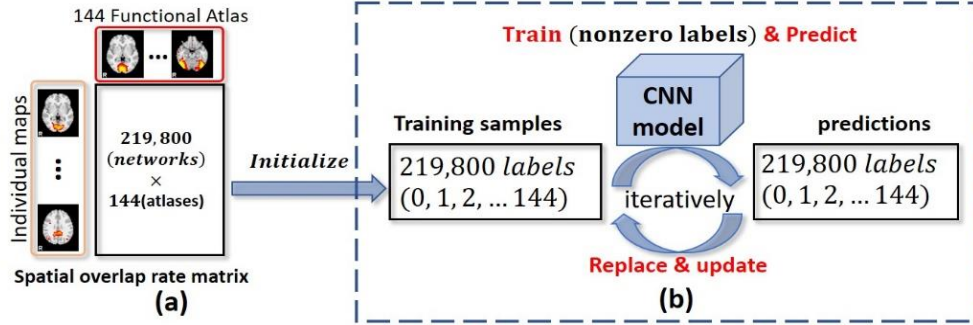


Fig. 2.39. IO-CNN framework. (a). weak label initialization process based on spatial overlap rate; (b). iteratively optimized CNN (IO-CNN) training process. The training process will iterate over the 219,800 input 3D maps starting with the initial weak labels. Only the 3D maps with nonzero labels will be taken as input for the current training iteration. After each training iteration, the trained model will predict on all the 219,800 training samples for new labels, which will be used to replace and update the previous training labels for the next training iteration. Meanwhile, the label differences between the previous training labels and the new predicted labels will be recorded for iteration termination condition check.

Based on the 144 functional atlases from ABIDE I dataset and the individual functional networks derived from ABIDE II dataset, the initial network labels were automatically and roughly assigned to each of the 219,800 networks by calculating the spatial overlap rate similarity matrix as shown in Fig. 2.39 (a). The spatial overlap rate is calculated in 2.10:

$$\text{overlap rate} = \sum_{k=1}^{|V|} \frac{\min(V_k, W_k)}{(V_k + W_k)/2} \quad 2.10$$

where V_k and W_k are the activation scores of voxel k in network volume maps V and W , respectively. To ensure the accuracy of the initial label assigned by this method, the empirical thresholding (threshold 0.2) process was applied on the similarity matrix, as demonstrated in (Zhao et al., 2016). For each individual network map, the label is assigned as atlas number whose spatial overlap rate is the maximum among all 144 atlases. In the case of no similarity value is above 0.2, the corresponding network map will be assigned the label 0, which will not be used in training the 3D CNN. The IO-CNN training process will begin after label initialization Fig. 2.39 (a).

Table 2.3. Algorithm for IO-CNN

Algorithm: deep iterative CNN with weak label initialization

Input: 1). 219,800 individual functional networks (1099 subjects, with 200 functional network each);

2). 144 functional atlases

Initialization: 1). Calculate pairwise overlap rate between individual functional networks and functional atlases $\rightarrow 219,800 \times 144$ similarity matrix S^0

2). Threshold overlap rate value in S below 0.2 to be 0;

3). **For** each individual network row S_i^0 **in** S^0 **do**

If $S_i^0 = \mathbf{0}$ **then**

$label_i = 0$

Else

$label_i = \text{argmax}(S_i^0)$

$\{\text{argmax}(S_i^0) \in N \mid 1 \leq \text{argmax}(S_i^0) \leq 144\}$

End

End

Return $label^0$

Deep iterative training: using none zero labeled individual functional networks and $label^0$ as initial training pairs.

For i **in** $\{0, 1, 2, \dots, \text{maxIter}\}$ **do**

Train CNN on

[none zero labeled individual functional networks, $label^{i-1}$]

```

labelt+1 = CNNmodel predict on
    all functional networks,
label_diff = diff(labelt, labelt+1)
If |label_diff|/219,800 < 0.4%
    Break
End
End
Return CNNmodel

```

The detailed algorithm part of IO-CNN (Fig. 2.39) is elaborated in Table 2.3. Briefly, the IO-CNN training process will iterate over the 219,800 input 3D network maps for a maximum N (e.g., $N=20$) iterations, starting with the initial weak labels based on spatial overlap rate. As mentioned in (Zhao et al., 2017b), spatial overlap based classification can achieve around 85% accuracy on prediction, while the CNN scheme has the superior ability for label correction on that prediction, thus improving the recognition accuracy. This intriguing label correction capability is adopted in the IO-CNN framework to improve the previously assigned training labels during each iteration, thus introducing changes between the labels predictions after training and the labels before training. This improvement is a core methodological contribution of this paper, in addition to the dramatically increased number of used subjects (1,099 subjects in total) and the much larger number of network labels (144 as an initialization). After iterative optimization, a balance will be achieved by the IO-CNN framework when no significant changes (e.g., less than 0.5%) occurs, thus yielding the optimized and well-trained CNN model for functional brain network recognition.

Specifically, the *maxIter* is set to 20 in case of stopping loop for oscillation, and the number of training epoch for each CNN training is empirically set to 5. The actual stop

condition is reached at iteration 13, during which the label change procedure is illustrated in Fig. 2.40 as an example.

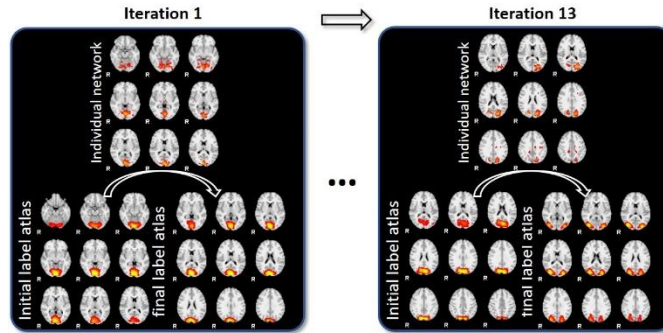


Fig. 2.40. Label change procedure during each iteration. During each iteration, initially assigned labels of some networks will be optimized to other labels during the CNN training iteration.

Results

IO-CNN Training Details

As we have introduced before, the training dataset is based on the decomposed functional networks using dictionary learning and sparse coding methods from ABIDE II dataset. After the rough label initialization process proposed in Table 2.3 as the first step, the label distribution for 144 functional network atlases consisting of 80,293 training samples are shown in Fig. 2.41. As we can see, most of training samples have relatively balanced labels except for a few (9 labels) labels whose training sets only have less than 0.15% samples. This is probably due to the imperfectness of the functional network atlases generated in our previous work [20], in that atlases with only minor individual networks among the population will not be regarded as intrinsic functional networks (Fox et al., 2005). As the IO-CNN training iteration goes on, the minor atlas training samples will be

eliminated or moved to other atlas labels, which reaches a final label number of 135 after the IO-CNN training converges.

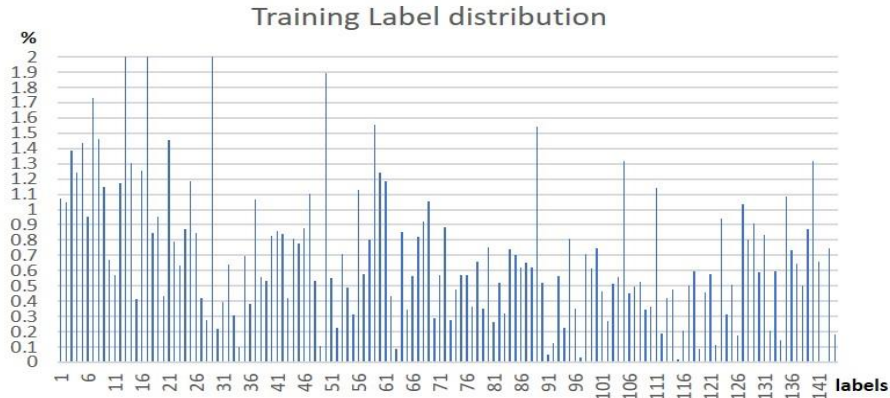


Fig. 2.41. Initial training set’s 144 functional network label distributions.

The training gradually converges to an optimization point as the iteration number increases, as shown in Fig. 2.42. Clearly, we can see that there is a significant accuracy drop during iteration 2, which is probably caused by the significant increase in training sample size from the initial training sample size of 80,293 to the size of 219,787 (Fig. 2.43). Then, the training sample size remains relatively stable, thus following a stable increasing training accuracy and decreasing loss. The convergence is reached at iteration 13, where the training sample label change over the previous iteration is less than 0.4%, which indicates the optimization of the IO-CNN, thus terminating the training iterations.

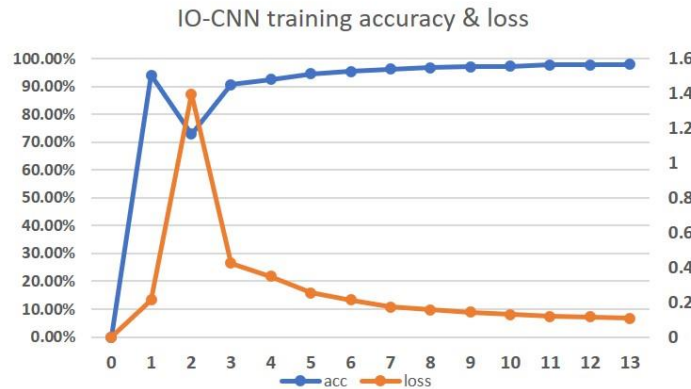


Fig. 2.42. IO-CNN training accuracy and loss curves from iteration 0 to 13. Iteration 0 is the weak label initialization process, where IO-CNN is not trained yet.



Fig. 2.43. Sample size’s dynamic changes and label change percentages during IO-CNN training. Iteration 0 is the weak label initialization process, where IO-CNN is not trained yet. Initial labels (80,293 initial training samples) were first used in iteration 1 for IO-CNN training.

For our IO-CNN training, an 8 GB memory GPU (Nvidia Quadro M4000) was utilized. The training time for each iteration is also dependent on the training sample size (average training samples size is 206,524 per iteration). A total time of 3.8 days (91 hours) was spent for all the 13 iterations of IO-CNN training (7 hours per iteration).

Recognition Results

The optimized IO-CNN model shows 87.5% prediction consistency with the initial overlap based training label, which is consistent with the overlap based recognition accuracy of 85.93% reported in our previous work (Zhao et al., 2017b), further demonstrating the deficiency in functional network recognition task using spatial overlap based methods. It is worth noting that functional network spatial overlap is a natural property of functional organization of the human brain (Xu et al., 2016), in that cortical

microcircuits overlap and interdigitate with each other (Harris and Mrsic-Flogel, 2013), rather than being independent and segregated in space. Thus, it is critically important to develop novel and effective methods to recognize spatially overlapping functional networks. In this paper, the network recognition and classification are based on the previously generated 144 atlases, which can be found and visualized at <http://hafni.cs.uga.edu/autism/templates/all.html>. As we can see, some of the 144 network atlases are highly spatially overlapped but they remain functionally distinct, e.g., Fig. 2.40 shows that the final label atlas and initial label atlas are quite similar due to the high spatial overlap.

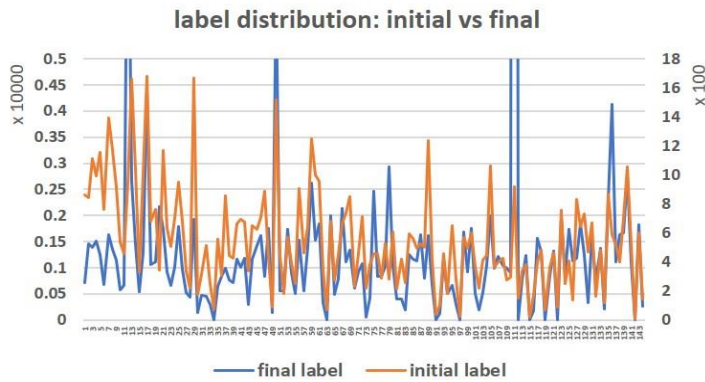


Fig. 2.44. Initial label distribution and final label distribution after iterative IO-CNN optimization.

Here, in this paper, the core idea of weak initialization for the IO-CNN is to use spatial overlap rate to roughly model the training data label distribution, and then optimize the distribution through IO-CNN training. The initial and final 144 label distributions are shown in Fig. 2.44. The Pearson correlation between the 144 initial label distribution and the final label distribution is 0.23, which shows some level of correlation between the initial

and final label distributions. Also, the p-value of the Pearson correlation is 0.0058, which is much smaller than the significance level 1%, indicating that the positive correlation between the initial and final label distributions is confident and significant. The significant correlation between the rough label distribution and the optimized label distribution cross-validated the plausibility of using overlap rate label initialization (~85% accuracy) and the optimization of IO-CNN based on that label initialization.

Due to the limited space, we only showcased 9 networks recognition results here using 3 labels with the most data distributions, 3 labels with medium data distributions, and 3 labels with the smallest data distributions in Fig. 2.45. As showcased in Fig. 2.45, the IO-CNN can really achieve accurate predictions for each network atlas label, which are demonstrated by both visual check and the high overlap rates of the individual networks with the corresponding atlases. The entire predictions for 219,800 networks for all of the 144 network atlas labels are visualized at our website: http://hafni.cs.uga.edu/144templates_CNN/Init-to-itr12/web/predictions.

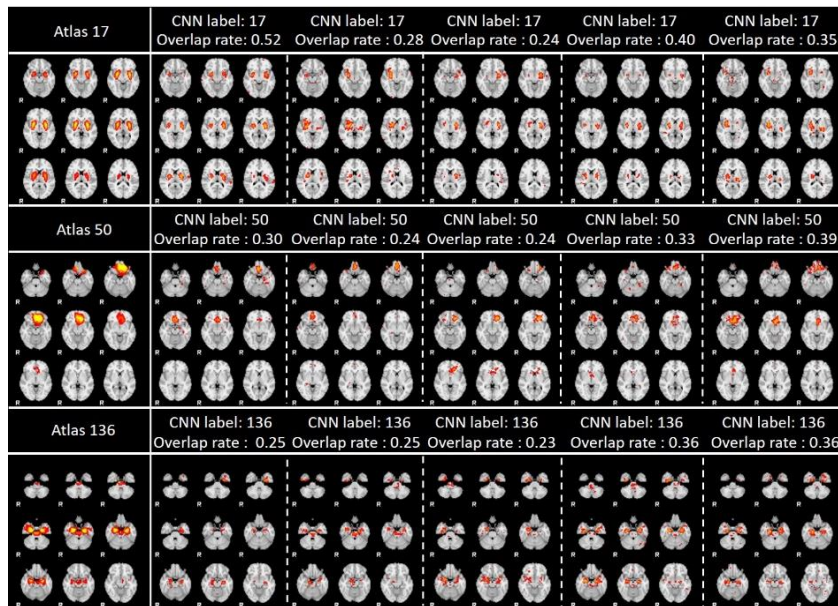


Fig. 2.45. (a). 3 label predictions with the most data distributions using 5 instances each. Each row contains the atlas picture at the first column, with the rest 5 columns as prediction instances from individual networks.

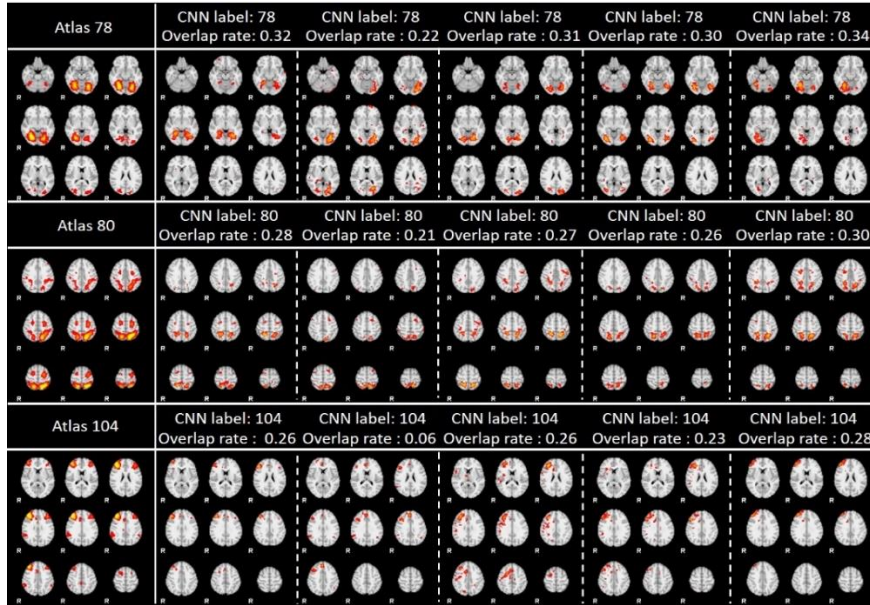


Fig. 2.45. (b). 3 label predictions with the medium data distributions using 5 instances each. Each row contains the atlas picture at the first column, with the rest 5 columns as prediction instances from individual networks.

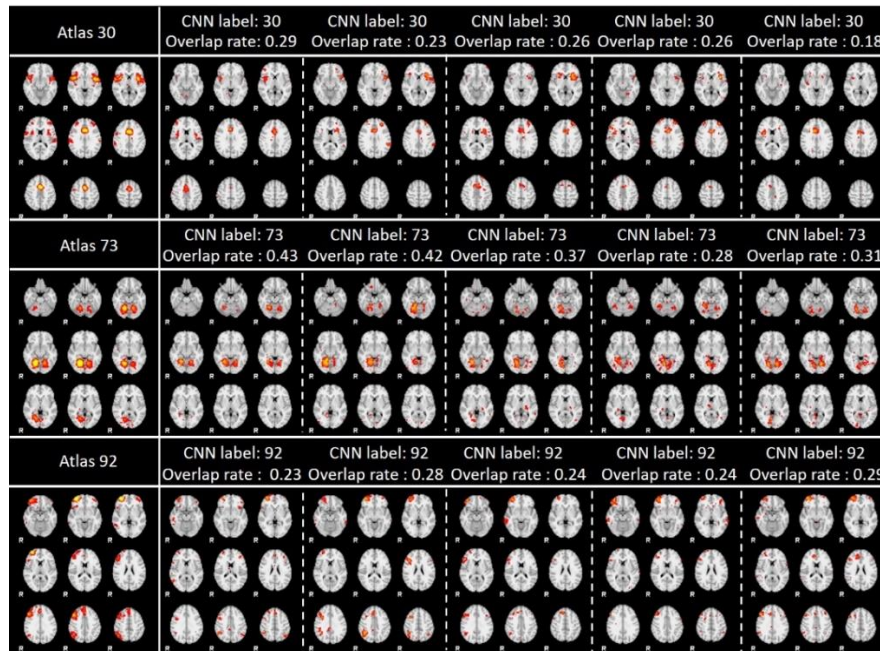


Fig. 2.45. (c). 3 label predictions with the least data distribution using 5 instances each. Each row contains the atlas picture at the first column, with the rest 5 columns as prediction instances from individual networks.

Label Change Analysis and Evaluations

As mentioned in the previous section, an 87.5% consistency was achieved between the initial overlap labels and the final IO-CNN predicted labels. In general, the inconsistent predictions with the initial labels have 4 conditions: 1) Previously unlabeled data is assigned with this label; 2) New instance assigned with this label; 3) Label shifted from this label; 4) Label removed. We will use the two functional network atlases: atlas 7 (default mode network) and atlas 14 (lateral visual area network) for the 4-condition label change analysis in the following sections in details. The comprehensive 144-atlas label changes are visualized at http://hafni.cs.uga.edu/144templates_CNN/Init-to-itr12/web/, where condition 1 is linked to “new label added from no label”, condition 2 is linked to “new labels added for each template from other labels”, condition 3 is linked to “new labels

removed from each template”, and condition 4 is linked to “label removed from each template”. We will explain these 4 conditions one by one in details as follows.

Condition 1). Previously unlabeled data assigned with this label. Intuitively, most label changes come from this condition since the initial labeled size is only 80,293, compared to the final label size of 219,433. Since we set an overlap rate threshold of 0.2 empirically according to (Zhao et al., 2016) to assign initial label, some functionally meaningful networks with lower spatial overlap rate than 0.2 were then discarded. Fortunately, those networks were then assigned back with a reasonable label using our trained IO-CNN model. A few randomly selected showcases using atlas 7 and 14 are illustrated in Fig. 2.46. The comprehensive 144-atlas label changes for this condition are visualized at http://hafni.cs.uga.edu/144templates_CNN/Init-to-itr12/web/added_from_no_label/index.html.

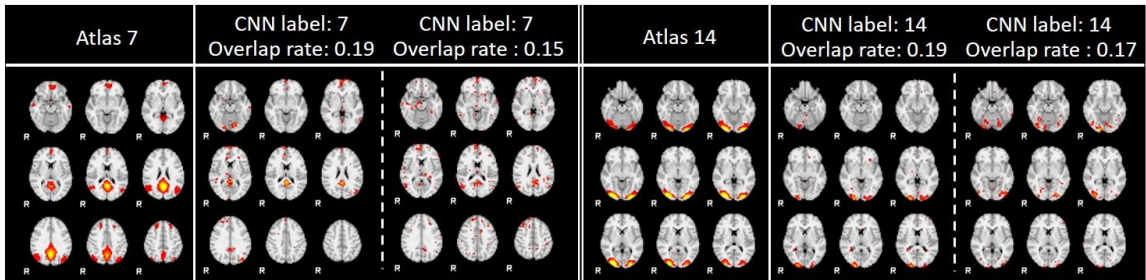


Fig. 2.46. Previously unlabeled network assigned with label 7 and 14. The overlap rate with those networks are all below 0.2 yielding no label initially.

Condition 2). New instance assigned with this label. This condition indicated that the current label is more suitable for the previous label assigned by using the overlap rate scheme.

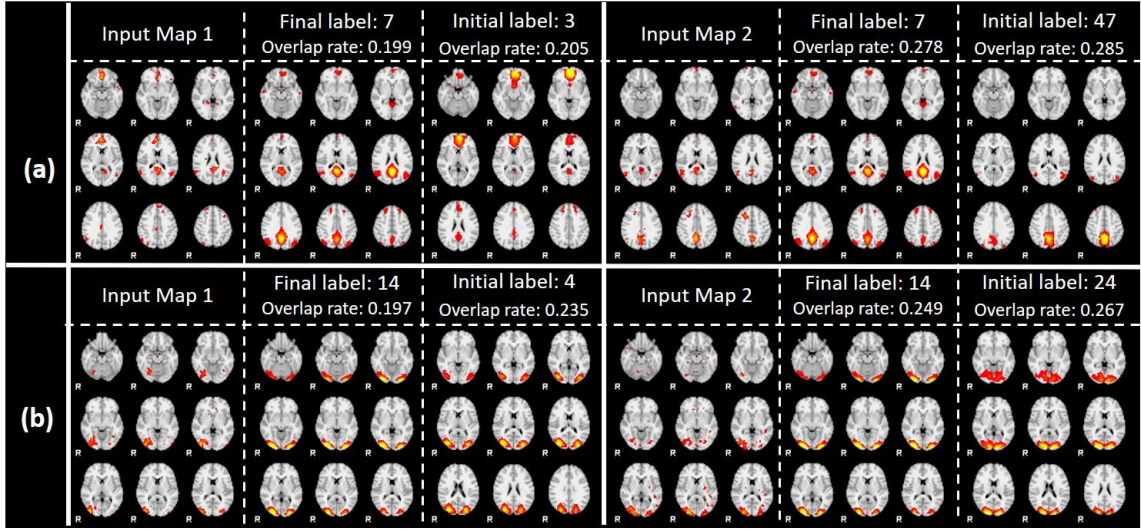


Fig. 2.47. (a). New instances with initial labels 3, 47 assigned with final label 7; (b). New instances with initial labels 4, 24 assigned with final label 14. First row of pictures indicates input networks; second row indicates final predicted atlases; third row indicates initial overlap rate assigned labels.

A few randomly selected showcases still using atlas 7 and 14 are illustrated in Fig. 2.47, and the corresponding 3D visualizations are shown in Fig. 2.48. From the visualization, we can confidently see the improved final label predictions, even though the spatial overlap rate is higher in initial labels than final label, which demonstrated that the spatial overlap rate is incapable of differentiating networks with high overlap rates. The 3D visualizations offer a better visualization of the spatial distribution of the input maps of activation areas, thus confirming that input maps match better with the final labels instead of initial labels. This improvement is the optimization result of our proposed IO-CNN framework and can be extensively illustrated by the comprehensive 144-atlas label change visualization for this condition at http://hafni.cs.uga.edu/144templates_CNN/Init-to-it12/web/added_in_new/index.html.

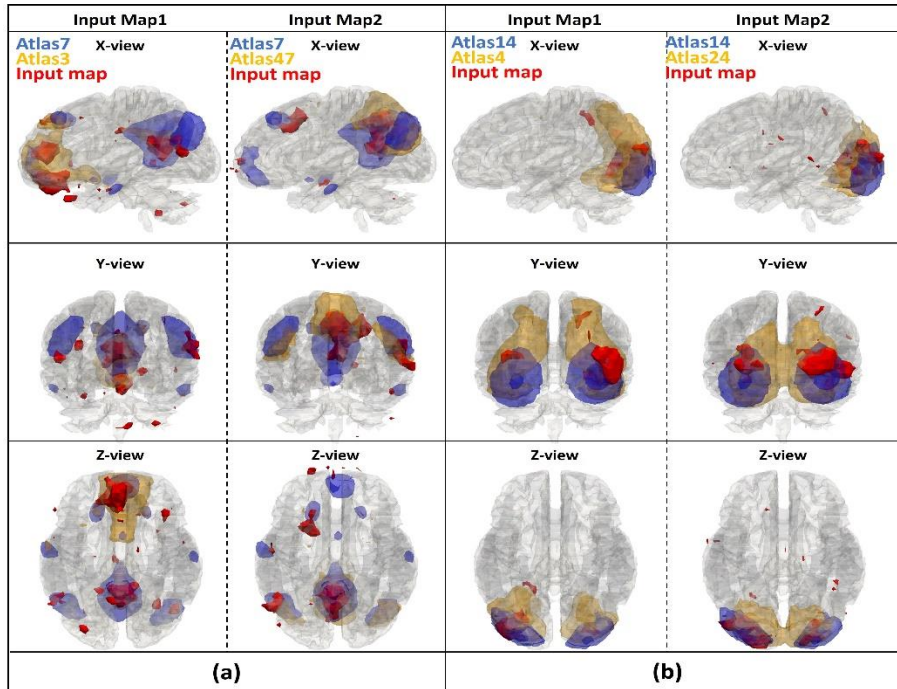


Fig. 2.48. (a). 3 axis 3D visualization for new instances with initial labels 3, 47 assigned with final label 7; (b). 3 axis 3D visualization for new instances with initial labels 4, 24 assigned with final label 14.

Condition 3). Label shifted from this label. This condition also indicated that the current label is more suitable for the previous label assigned using overlap rate scheme. Several randomly selected showcases still using atlas 7 and 14 are illustrated in Fig. 2.49, and the corresponding 3D visualizations are shown in Fig. 2.50. Similar to condition 2), we can clearly see the plausibility of the final predicted labels when the initial label atlas and final label atlas have a high overlap rate. For instance, for input map 2 in Fig. 2.49 (b) and Fig. 2.50 (b), we somehow have difficulties in telling the difference between the highly overlapped atlases 14 and 57. However, if we observe closely, we will find that the spatial distributions of the two high overlapped atlases are different: atlas 57 has a unbalanced lateral activated regions, while atlas 14 has a balanced lateral activated regions. And our

final label for input map 2 is atlas 57 because the input map also has a clearly unbalanced lateral activation pattern, which should be assigned with a label 57. The extensive illustrations of comprehensive 144-atlas label changes for this condition are visualized at http://hafni.cs.uga.edu/144templates_CNN/Init-to-itr12/web/removed_from_old/index.html

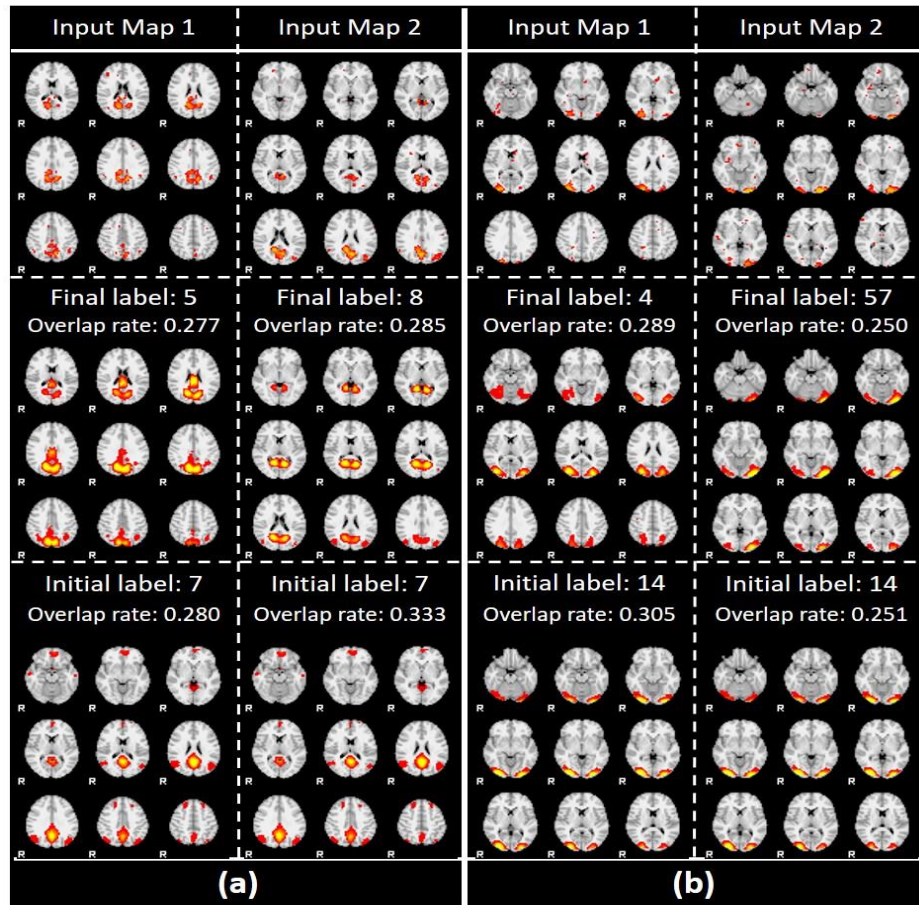


Fig. 2.49. (a). Label shifted from label 7 to label 5, 8; (b). Label shifted from label 14 to label 4, 57. First row of pictures indicates input networks; second row indicates final predicted atlases; third row indicates initial overlap rate assigned

Condition 4). Label removed. This condition indicated that the label was removed (recognized as noisy network) for a network which was initially assigned with a label.

Usually this condition should not happen, since if a network was assigned with an initial label, meaning the overlap rate of this network is at least 0.2 with some atlases, which indicates non-noisy network. As a result, only one initial network in atlas 7 class is removed, and there is no such condition for atlas 14. As expected, this condition does not exist in most of the atlases http://hafni.cs.uga.edu/144templates_CNN/Init-to-itr12/web/removed_in_this_label/index.html.

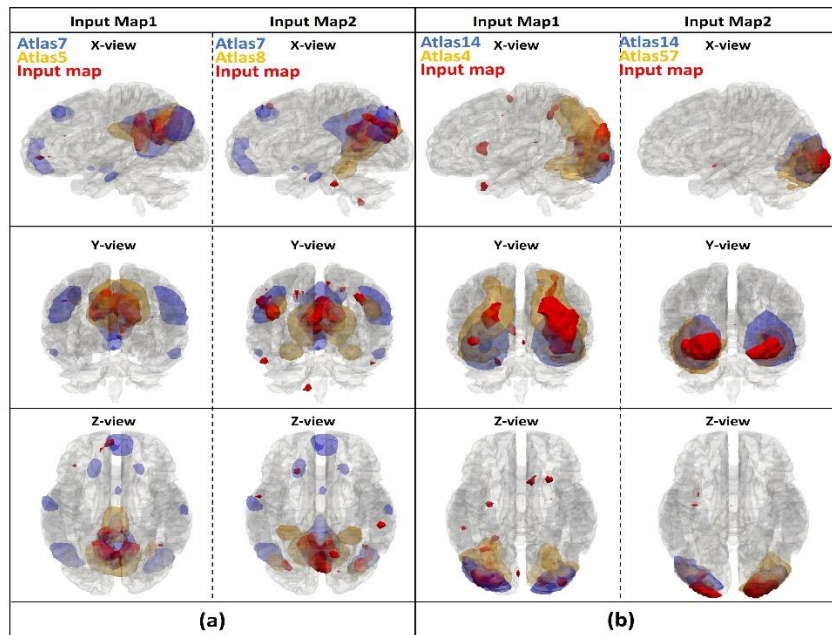


Fig. 2.50. (a). 3 axis 3D visualization for instances with label shifted from label 7 to label 5, 8; (b). 3 axis 3D visualization for instances with label shifted from label 14 to label 4, 57.

Revealing Fine Granularity Networks From 10 RSNs and Ambiguity Removal using HCP Dataset

The Human Connectome Project (HCP) fMRI dataset is considered as a systematic and comprehensive mapping of connectome-scale functional networks and core nodes over

a large population in the literature (Barch et al., 2013), based on which we have a classification framework for previously reported 10 resting state networks (RSN) (Smith et al., 2009) (Zhao et al., 2017b). In this section, we re-classified the 1,521 testing dataset networks from the previous HCP dataset (Van Essen et al., 2013). After the predictions using the proposed 135 (from initial 144) atlas prediction framework compared with the previous 10 RSNs, we found that our 135 atlases are actually variants of the 10 RSNs with fine-granularities. Also, we further confirmed the previous framework using CNN has discovered manual labeling mistakes. As we can see from Fig. 2.51, the most percentage (except noises) of the atlas from 135 atlases resemble the corresponding atlases in 10 RSNs, which demonstrated the robustness of the proposed framework and its applicability on different fMRI data sets. Fig. 2.52 shows an example of the fine granularities among 135 atlases of RSNs and the previous ambiguous labels using RSN 1 as an example. As shown in Fig. 2.52, atlases 4, 24 are not necessarily to be variants of fine granularities of RSN 1, but the reclassification atlases seem to be more reasonable and accurate than the previous RSN 1 using 10 RSN labels. And other atlases are variants of fine granularities of the RSN1. For a comprehensive and detailed reclassification of the HCP dataset, please refer to http://hafni.cs.uga.edu/144templates_CNN/forHCP/web/index.html.

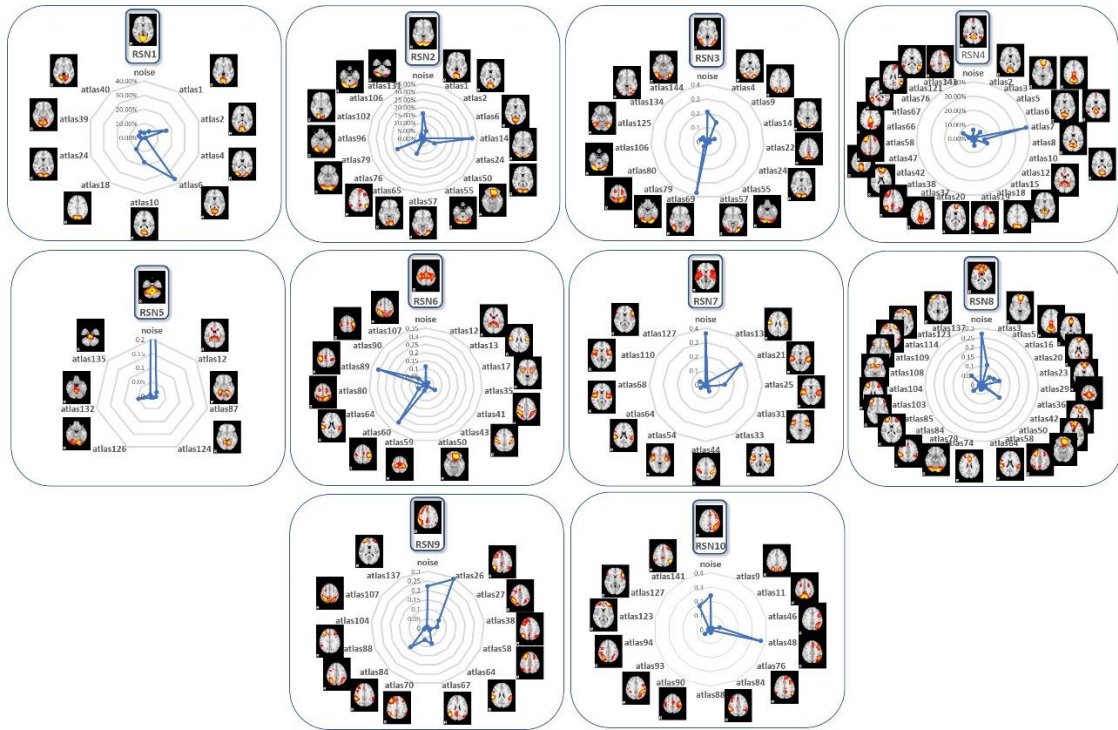


Fig. 2.51. Proposed 135-class predictions on previous 10-class labelled HCP testing set networks reveals fine-granularities of the variants of 10 RSNs. The radar charts show the 135 templates granular percentages out of the 10 RSN atlases.

This result cross-validated the effectiveness and the scalability of our IO-CNN framework to other datasets. And this finding of IO-CNN derived fine granularity of functional brain networks is also consistent with prior results in (Zhao et al., 2017a), providing neuroscientific basis for hierarchical and overlapping architecture of the human brain functions.

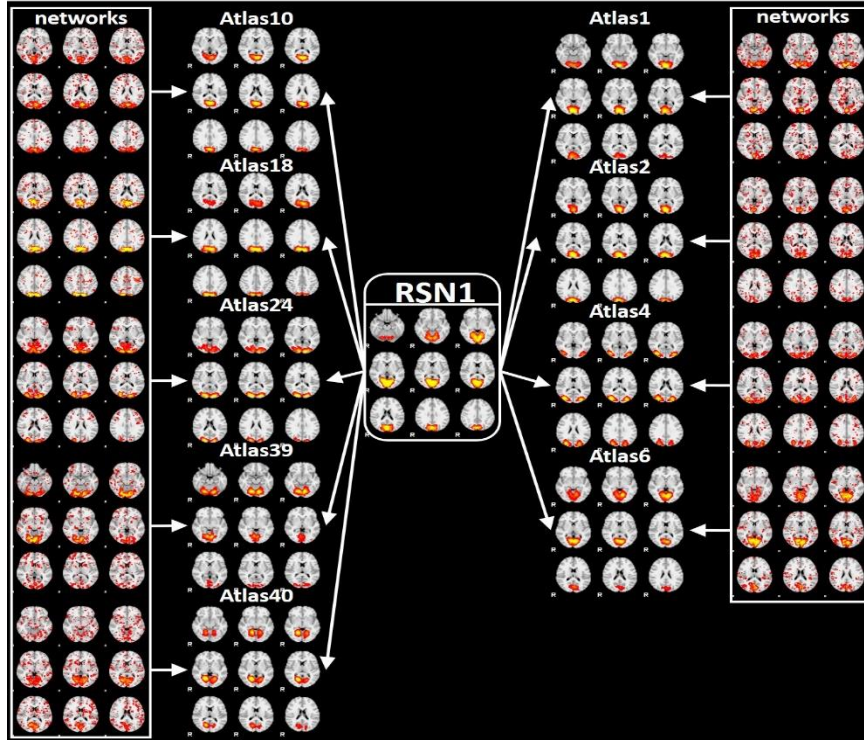


Fig. 2.52. Fine-granularities among 135 atlases and previous ambiguous labels using RSN1 as an example.

DISCUSSION AND CONCLUSIONS

Connectome-scale reconstruction of reproducible and meaningful functional brain networks on large-scale populations based on fMRI data are enabled by using functional brain network decomposition techniques, especially the HAFNI project. However, an unsolved problem in the HAFNI framework is the automatic recognition of hundreds of HAFNI maps such as RSNs in each individual brain, which is challenging due to the lack of functional brain atlases, no correspondence across decomposed or reconstructed networks across different subjects, and the significant individual variabilities. To deal with the tremendous variability of various types of functional brain networks and the presence of various sources of noises, we have proposed a 3D CNN deep learning framework to solve those problems in section 2.5 (Zhao et al., 2017b). However, major challenges like

labelling large scale networks and recognition of them with larger scale labels still remain prominent. The previous 3D CNN framework is extended in this work for much larger networks recognitions (from ~5,000 to ~220,000) with much larger numbers of network atlases (from 10 to 135) with a weak but automatic labeling. The recognition results and the label changes from initial labels by our IO-CNN framework demonstrated the optimization during iterative training. The application of IO-CNN on the separate HCP dataset with previous 10 RSN labels obtained promising results, in which functional network fine-granularities and label ambiguity removal have been achieved, further demonstrating the robustness and efficiency of our proposed IO-CNN framework. This IO-CNN complete the automatic system for holistic brain functional network recognition and classification. Recognizing this differentiability of CNN frameworks, in the next section, we will further refine the applications on clinical fMRI datasets for brain disease modelling, such as Autism Spectrum Disorder (section 2.7).

Acknowledgements: 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) for this research. We thank the ABIDE and HCP projects for sharing their valuable fMRI datasets.

2.7 3D Deep Convolutional Neural Network Revealed the Value of Brain Network Overlap in Differentiating Autism Spectrum Disorder from Healthy Controls

Abstract

Spatial distribution patterns of functional brain networks derived from resting state

fMRI data have been widely examined in both the literature and the abovementioned researches. However, the spatial overlap patterns among those brain networks have been rarely investigated, though spatial overlap is a fundamental principle of functional brain network organization. To bridge this gap, this paper presents an effective 3D convolutional neural network (CNN) framework to derive discriminative and meaningful spatial brain network overlap patterns that can characterize and differentiate Autism Spectrum Disorder (ASD) from healthy controls. Our experimental results demonstrated that the spatial distribution patterns of connectome-scale functional network maps per se have little discrimination power in differentiating ASD from controls via the CNN framework. In contrast, the spatial overlap patterns instead of spatial patterns per se among these connectome-scale networks, learned via the same CNN framework, have remarkable differentiation power in separating ASD from controls. Again, our work suggested the promise of using CNN deep learning methodologies to discover discriminative and meaningful spatial network overlap patterns and their applications in functional connectomics of brain disorders such as ASD.

Introduction

Along with faithful reconstruction and quantitative modeling of brain networks from noisy fMRI data, such as general linear model (GLM) (Friston et al., 1994) for task-based fMRI (tfMRI), independent component analysis (ICA) (Beckmann et al., 2005) for resting state fMRI (rsfMRI), and dictionary learning/sparse representation (Lv et al., 2015b) for both tfMRI and rsfMRI, all of which can reconstruct concurrent network maps from whole brain fMRI data, different spatial analyses on the decomposed functional brain networks have also been investigated in the previous sections in this chapter. However, the spatial

overlap patterns among those hundreds of concurrent brain networks have been rarely investigated, though spatial overlap is a fundamental principle of functional brain network organization (Xu et al., 2016). It is even a bigger mystery whether those spatial network overlap patterns provide valuable information of brain disorders such as Autism Spectrum Disorder (ASD).

To bridge these technical and knowledge gaps and continuing the exploitation of the deep learning frameworks, this section presents an effective 3D convolutional neural network (CNN) (Krizhevsky et al., 2012; Lecun et al., 1998) framework to derive discriminative and meaningful spatial brain network overlap patterns that aim to characterize and differentiate ASD from healthy controls. Faced with the challenges of fractionation of the available data in site-specific studies with relatively small sample sizes, in this section, we utilized multisite (10 sites) rsfMRI datasets from the ABIDE project for the classification problem to improve statistical power, and to accommodate greater variance of ASD and control. In this work, spatial distribution patterns of connectome-scale functional network maps per se were first investigated for classifying ASD patients and controls, and our experimental results demonstrated that they have little discrimination power in differentiating ASD from controls via the CNN framework (less than 50 % average accuracy in testing data). In contrast, the spatial overlap patterns among these connectome-scale networks, learned via the same CNN framework, have remarkably better differentiation power in separating ASD from controls (10-fold accuracy: average at 70.5% and peak at 85%) than using functional network spatial patterns per se. Our work also suggests the promise of using CNN deep learning methodologies to discover discriminative

and meaningful spatial network overlap patterns and their applications in functional connectomics of brain disorders such as ASD.

Method

Experimental data and preprocessing

Our experimental data were downloaded from the publicly available Autism Brain Imaging Data Exchange (ABIDE: http://fcon_1000.projects.nitrc.org/indi/abide/) with rsfMRI datasets collected from ASD patients and healthy controls. In this study, we investigated ASD patient and healthy control' rsfMRI data from 10 different sites. After manually checking data quality according to preprocessing (e.g. skull removal, registration to standard space) results, 100 ASD patients and 100 healthy controls' data were selected for this experiment. The acquisition parameters vary from different sites: 192 - 256 mm FOV, 29 - 47 slices, 1.5 – 3 s TR, 15 – 30 ms TE, 60 - 90 ° flip angle, (3 - 3.4) × (3 - 3.4) × (3 - 4) mm voxel size. For detailed parameters for each site, please refer to the ABIDE website.

Preprocessing of the rsfMRI data was performed using FSL software tools (Jenkinson et al., 2012), including skull removal, motion correction (MCFLIRT command in FSL tools was adopted, where 4 mm (voxel level) motion parameter search was conducted to remove micro head-motions (Jenkinson et al., 2012) and further manual check was performed to remove unsuccessful data), spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and linear registration to the Montreal Neurological Institute (MNI) standard brain template space, which were implemented by FSL FLIRT and FEAT. Next, the preprocessed rsfMRI data were decomposed into

different functional networks using the online dictionary learning algorithm (Mairal et al., 2010).

144 ICNs generation and spatial overlap feature selections

As shown in Fig. 2.55(a), by applying the online dictionary learning and sparse representation techniques in (Mairal et al., 2010), whole brain rsfMRI signals of each individual was decomposed into 200 intrinsic connectivity networks (ICNs). First, whole brain rsfMRI signals were extracted and arranged into a matrix $X \in \mathfrak{R}^{t \times n}$ with n columns representing each voxel and t rows representing each time point of the rsfMRI scan. By using the online dictionary learning and sparse representation technique (Mairal et al., 2010), each column in X was modeled as a linear combination of atoms of a learned basis dictionary D such that $X=D \times \alpha$, where $D \in \mathfrak{R}^{t \times m}$, $\alpha \in \mathfrak{R}^{m \times n}$, and m is set to 200 (as in (Zhao et al., 2016)) in this paper. Finally, each row in the α matrix was mapped back to the brain volume as a network. Then the 144 ICN templates generated in (Zhao et al., 2016) charactering group-wise consistent ASD/control brain networks were used to search for the correspondence in the 200 decomposed networks – the network having the maximum overlap rate (calculated in equation 2.11) with the template was taken as its correspondence:

$$\text{overlap rate} = \sum_{k=1}^{|V|} \frac{\min(V_k, W_k)}{(V_k + W_k)/2} \quad 2.11$$

where V_k and W_k are the activation intensity of voxel k in network volume maps \mathbf{V} and \mathbf{W} , respectively.

The overlap information among all the 144 ICNs are then gathered, by generating overlap maps of each pair of networks among 144 ICNs using equation (2). For 144 ICNs,

a total combination number of 10291 overlap maps were generated for each subject. The illustration for overlapped map generation is shown in Fig. 2.53 by using ICN1 and ICN2 on a randomly selected subject.

$$\text{overlapped map} = V \cap W \quad 2.12$$

where V and W are the binarized ICN volume maps.

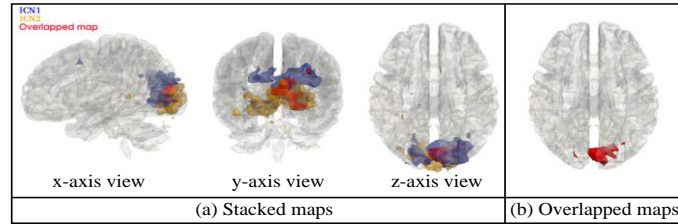


Fig. 2.53. Overlap map illustration using one subject's ICNs. Blue color and yellow color denote ICN 1 and ICN 2 volume map respectively, while red denotes the overlap map.

Among all 10,291 overlap maps, only 150 were selected as representative features for each subject since not all maps are overlapped with each other. In order to select the most overlapped patterns, the 10,291 maps are sorted in a descendant manner using the averaged nonzero voxel numbers in overlapped maps of each group (ASD and control). The first 50 feature maps were selected from maps present in the top of both groups' overlaps (with 1 ASD patient and 1 control subject example shown in Fig. 2.54(a)); the second 50 feature maps were selected from maps only present in top of control group's overlaps (with 1 ASD patient and 1 control subject example shown in Fig. 2.54 (b)), while the last 50 feature maps were selected from maps only present in the top of ASD group's overlaps (with 1 ASD patient and 1 control subject example shown in Fig. 2.54 (c)). The 150 overlap maps were then extracted from all the subjects from the two groups as input features for the CNN training and classification.

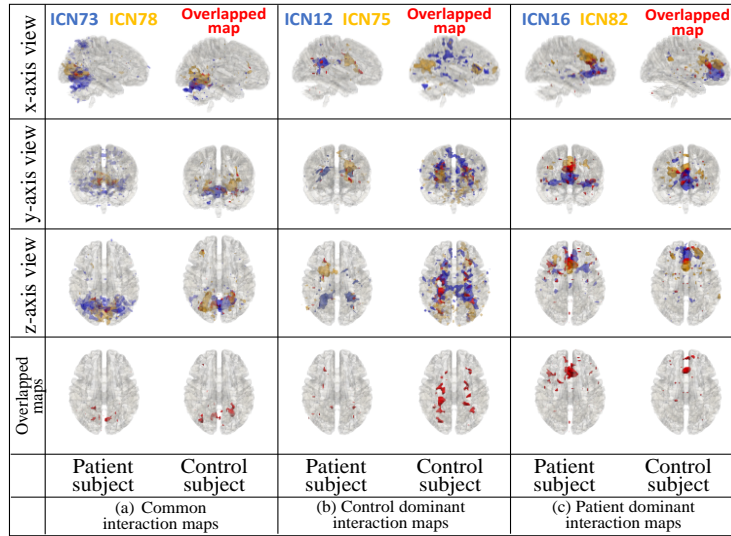


Fig. 2.54. Overlap map feature selections. Blue and yellow maps in each panel are the ICN maps, while red maps are the overlap maps of each pair of ICN maps. (a). Common overlap maps indicate both ASD patient and control groups have these types of overlaps among top 50 patterns. (b). Control-dominant overlap maps (c). ASD-dominant overlap maps.

Computational framework

After obtaining the 144 intrinsic connectivity networks (ICNs) of each subject (to be described in section 2.3), the spatial overlap patterns of 144 ICNs were used as input features and an fMRI-oriented 3D deep convolutional neural network (CNN) was then designed for the problem of ASD/control subjects classification (section 2.4). The whole framework is briefly illustrated in Fig. 2.55.

Specifically, sparse representation was first performed on all the preprocessed data to extract concurrent networks at individual level; then the 144 ICNs templates generated in (Zhao et al., 2016) based on the ABIDE datasets were used to search the corresponding ICNs among 200 concurrent networks within each subject. Even though those ICNs were not directly used as input features for final classifications, they will still be used as input

features for comparison purpose. Instead, the overlapped maps of each ICN pair were generated, and 10291 maps in total for each subject were generated, 150 out of which were selected as the input features for the CNN for each subject (section 2.3). Our core idea here is that we hypothesize the networks' overlap patterns represent their functional interactions due to the highly interdigitated and spatially overlapped nature of the functional brain organization (Xu et al., 2016) and thus the degrees of overlaps carry important functional information of brain networks. Finally, the training and testing phases were done for the proposed deep 3D CNN in a 10-fold cross-validation manner, which will be introduced in section 2.4.

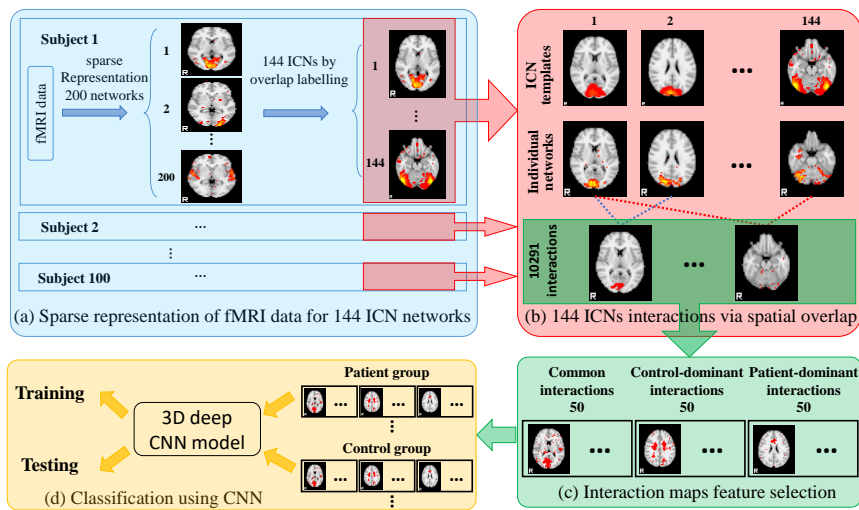


Fig. 2.55. Computational framework for ASD classification.

Deep 3D CNN structure

As mentioned above, the input feature for each subject is a 4D matrix with 150 3D network overlap maps. In order to accommodate the subtle spatial characteristics of the input feature volume and to differentiate them, we adopted and improved an effective 3D CNN framework showing the extraordinary ability in accommodating spatial object pattern representation to extract the hierarchy of useful features from the input volumes. The model

was built using Theano (<http://deeplearning.net/software/theano/>) adopting fully 3D convolutional layer from (Maturana and Scherer, 2015). The structure of the proposed deep 3D CNN is shown in Fig. 2.56. This powerful deep 3D multichannel convolutional architecture can well incorporate 3D structure information as hierarchical intrinsic features and then be used for classification.

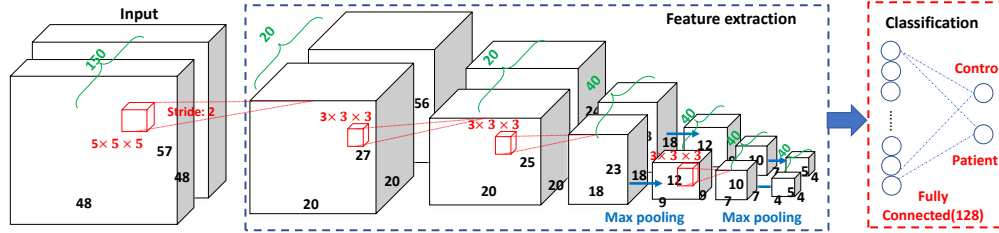


Fig. 2.56. Deep 3D CNN structure for feature extraction and classification. Dimensions of each layer using different colors (green: channel, black: 3D dimension).

The neural network weights training was performed by the classic Stochastic Gradient Descent (SGD) with momentum. The objective loss function to be optimized is the multinomial negative log-likelihood with a λ (set to 0.001) times the L_2 norm of the network weights as regularization term, as shown in Eq.(3).

$$L(\theta) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^k \mathbf{1} \cdot \{y^i = j\} \log(\theta^T x^i)_j + \lambda \|\theta\|_2 \quad 2.13$$

where m is the number of samples in one batch (empirically set to 12 to fit memory), and k is the number of the output classes (2 classes for control and patient) and $\log(\theta^T x^i)_j$ is the log-likelihood activation value of the j_{th} output node. The momentum parameter was set to 0.9. In this work, the widely-used dropout technique (with dropout probability 0.2) was adopted for each layer during the training process to reduce the overfitting problem (Bell and Koren, 2007). The convolutional layers were initialized using the similar scheme proposed in (He et al., 2015a), and the dense layers were initialized with a Gaussian

distribution with $\mu = 0, \sigma = 0.01$. Training was performed by utilizing GPU (NVIDIA Quadro M4000) for 300 epochs. For each training epoch, the total training time is about 2 hours. This scale of training time cost makes the proposed 3D CNN framework very suitable for future cognitive and clinical neuroscience applications.

Results

ASD patient/control classification

Due to the limited sample data, the 3D CNN model may be vulnerable to the overfitting problem. 10-fold cross validation is then operated to make sure the classification accuracy is robust. In total, we have 100 control subjects and 100 ASD subjects, and the 200 subjects were randomly partitioned into 10 equally sized subsamples as testing sets, while the disjoint parts of each testing set are used as the training set. Empirically, the training epochs were set to 300, and learning rate was initially set to 0.01 and then decreased to 0.005 until 66th epoch. It further decreased to 0.001 from 266th epoch to the end. The training accuracy and loss curves for the 10-fold training are shown in Fig. 2.57 with respect to the number of epochs (300 epochs).

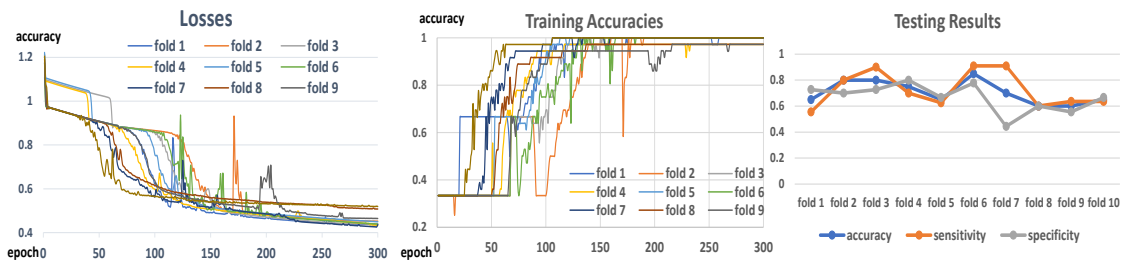


Fig. 2.57. Losses and accuracies during 10-fold training process.

The average, highest and lowest accuracy, sensitivity and specificity are shown in Table 2.4.

Table 2.4. Accuracy, sensitivity and specificity in ASD/control classification.

Averaged 10-fold	Accuracy (%)	Sensitivity (%)	Specificity (%)
Using overlapped	70.5	74	67
Using ICN per se	45.0	51	39

Among previous research studies on ASD classification using rsfMRI data, accuracy at 60% for ABIDE dataset was reported in (Nielsen et al., 2013), and accuracy at ~75% for rsfMRI-based classifiers was reported in (Plitt et al., 2015). Our method using network overlap information of ICN maps alone from rsfMRI, based on deep 3D CNN classifier, achieved the multi-site robust accuracy of 70.5% on average, and the peak accuracy is 85%. We believe this is a reasonably high accuracy, considering that only network overlap information is used here. In comparison, we also tried using only 144 ICNs as input features from each subject without overlap information using the same CNN structure and training schemes. But using the same CNN structure, the averaged testing classification accuracies were less than 50% (Table 2.4), which indicates that the 144 ICNs per se are not discriminating between ASD/control groups if spatial ICN overlap information is not incorporated. These results offer a new insight into the abnormal overlaps of functional networks among ASD patient and control groups. In the following section, the analysis on ICN spatial overlap differences between ASD/control groups will be discussed.

Spatial ICN overlap differences between ASD and controls

The reasonably high accuracy on the 10-fold classification results demonstrated the discriminative features of the ICN's spatial overlap and the ability of spatial pattern description of our proposed deep 3D CNN structure. As previous literature reported (Kennedy et al., 2006; Kleinhans et al., 2008), regions of the default mode network (DMN)

and fusiform gyri are found with decreased functional connectivity among ASD group. It is inspiring that after feature selections by deep 3D CNN described in Fig. 2.55(c), spatial overlaps involving DMN and fusiform gyrus are found to be dominant only in control group. These features are visualized using randomly selected subjects from both groups in Fig. 2.58. As we can see, spatial overlaps are substantially stronger in control subjects than ASD subjects, which provides the neuroanatomic substrates for good classification accuracy by CNN.

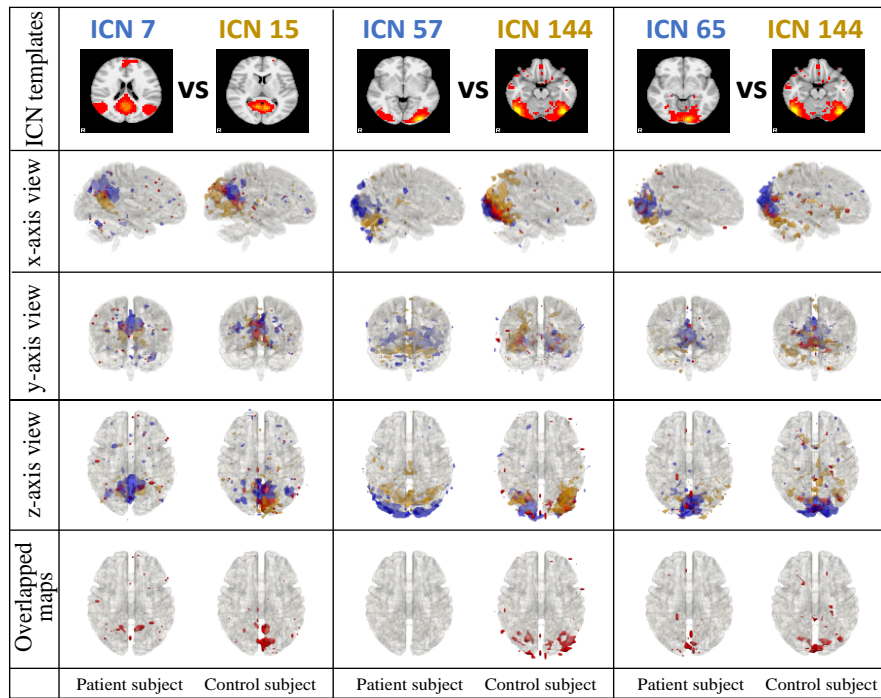


Fig. 2.58. Control-dominant network overlap patterns involving default mode network (ICN 7) and fusiform gyrus (ICN 144).

Discussion and Conclusion

This work used spatial ICN overlap as features in a deep 3D CNN structure for ASD/control classification and achieved substantially better classification results compared to using only spatial network maps themselves. Our experimental results

demonstrated the spatial distribution patterns of connectome-scale functional network maps per se have little discrimination power in differentiating ASD from controls via the CNN framework, while spatial overlap patterns among these connectome-scale networks have remarkable differentiation power. Our work suggested the promise of using CNN deep learning methodologies to discover discriminative and meaningful spatial network overlap patterns and their applications in functional connectomics of brain disorders like ASD in this section. In the future, additional meaningful features such as the temporal functional interaction patterns among ICNs will be utilized and integrated with the spatial ICN overlap patterns for the purpose of defining more comprehensive, multimodal features to be used to improve the CNN classification accuracy. In addition, improved CNN structures and larger scale of training/testing samples will be explored in the future to further improve the prediction accuracy and to discover potential abnormalities of ICNs' spatial overlaps for diagnosis of brain disorders like ASD.

CNN structures have been widely used in the spatial analysis as CNN can well model the distribution of the 3D spatial distribution. Despite the promising applications of CNN in 3D shape analysis, we also notice that the formation of the input to CNN affects the final performance significantly from the research of differentiating ASD from normal controls. Those findings can be instructive for future applications of deep learning frameworks: to devise new deep learning structure and also to devise informative input data representations. The abovementioned researches enclosing the spatial analyses of fMRI in this dissertation but never enclosing new applications of new frameworks in this research field.

CHAPTER 3

TEMPORAL FUNCTIONAL ANALYSES FOR BRAIN FMRI IMAGES^{8,9}

⁸ Modeling Task fMRI Data via Deep Convolutional Autoencoder

Yu Zhao*, Heng Huang, Xintao Hu, Milad Makkie, Qinglin Dong, Shijie Zhao, Lei Guo, Tianming Liu.
IEEE TMI, 2017. Vol 37(7).

*Co-author.

Reprinted here with the permission of publisher.

⁹ Template-guided Functional Network Identification via Supervised Dictionary Learning

Yu Zhao*, Xiang Li*, Milad Makkie, Shannon Quinn, Binbin Lin, Jieping Ye, Tianming Liu. Accepted by
ISBI, 2017.

*Co-first author.

Reprinted here with the permission of publisher.

3.1 Abstract

Spatial analyses rely on the functional brain networks decomposed from fMRI, while temporal analyses focus on how to decompose functional brain networks from fMRI data. In order to decompose fMRI data for both spatial and temporal analysis, different kinds of approaches have been proposed. Conventionally, the independent component analysis (ICA) (Cole et al., 2010; McKeown et al., 2003) is the mainstream, which identifies the temporally independent “signal source” in the 4D fMRI data, then obtains the spatial patterns of those sources through regression. Typically, the number of network components that can be reconstructed from ICA or GLM method is up to several dozens, which are defined by the number of brain sources (Cole et al., 2010) for ICA and the number of linear compounds of parameter estimates (COPE) (Friston et al., 1994) for GLM. In recent several years, a new computational framework of sparse representation (Lv et al., 2015a, 2015b; Mairal et al., 2010) of whole-brain fMRI signals was proposed and used for both resting state and task-evoked fMRI signal decompositions. The neuroscience foundation of this sparse representation framework is that a variety of brain regions and networks exhibit strong functional diversity and heterogeneity (Dosenbach et al., 2006; Duncan, 2010; Fedorenko et al., 2013; Fox et al., 2005; Pessoa, 2012), that is, a brain region could participate in multiple functional processes and a functional network might recruit various heterogeneous neuroanatomic areas. In addition, the number of functional networks that a cortical region is involved in at a specific moment is sparse, typically from several to one or two dozen (Pessoa, 2012). Later on, an advanced method utilizing dictionary learning and sparse representation techniques was proposed to decompose fMRI data for better spatial and temporal analysis. Typically, hundreds of concurrent functional

networks can be reconstructed effectively and robustly by sparse representation methods, thus forming holistic atlases of functional networks and interactions (HAFNI) (Lv et al., 2015b). Concurrent functional networks decomposed by this method have been shown to be superior in revealing the reconstructed task-evoked and/or resting state networks' spatial overlaps and their corresponding functions (Lv et al., 2015b). Recently, deep learning has attracted much attention in the field of machine learning and data mining (Bengio et al., 2013), and deep learning approach is proven to be superb at learning high-level and mid-level features from low-level raw data (Schmidhuber, 2014). Several deep learning based methods have been explored to discover the temporal features from the fMRI data for better decomposition. For instance, Restricted Boltzmann Machine (RBM) and deep convolutional autoencoder (DCAE) have both been leveraged for applications in fMRI signal analysis and modeling (Han et al., 2015; Huang et al., 2017; Zhao et al., 2017a). New methodologies are still emerging for reconstructing brain networks to offer the fundamental understanding of functional brain mechanisms. This chapter will introduce two frameworks that we have been involved into for the temporal analysis of fMRI data. One is traditional shallow model (Zhao et al., 2017c) utilizing dictionary learning and sparse representation in section 3.2, while the other one is deep model utilizing DCAE (Huang et al., 2017) for fMRI temporal feature extraction to better decompose fMRI data in section 3.3.

3.2 Template-guided Functional Network Identification via Supervised Dictionary Learning

Abstract

Functional network analysis based on matrix decomposition/factorization methods

including ICA and dictionary learning models have become a popular approach in fMRI study. Yet it is still a challenging issue in interpreting the result networks because of the inter-subject variability and image noises, thus in many cases, manual inspection on the obtained networks is needed. Aiming to provide a fast and reliable functional network identification tool for both normal and diseased brain fMRI data analysis, in this work, we propose a novel supervised dictionary learning model based on rank-1 matrix decomposition algorithm (S-r1DL) with sparseness constraint. Application on the Autism Brain Imaging Data Exchange (ABIDE) database showed that S-r1DL can fast and accurately identify the functional networks based on the given templates, comparing to unsupervised learning method.

Introduction

In recent functional neuroimaging studies, matrix decomposition-based data driven methods have been widely applied, including Independent Component Analysis (ICA) (Martin J McKeown et al., 1998) and dictionary learning (Lee et al., 2016) (Lv et al., 2015b). The main purpose of these methods is to discover and reconstruct the intrinsic functional organization patterns, modeled by “functional networks”, from fMRI data. Such functional networks are defined by a collection of brain regions with specific temporal/spatial properties (e.g. in dictionary learning, the spatial maps of those regions are constrained by sparseness) and are used to characterizes the brain’s functional aggregation behavior into neuroscientifically meaningful atomic elements. A series of important functional networks have been discovered by these methods in literatures, including the Default Mode Network (DMN) and other well-established resting-state networks (RSNs) (Smith et al., 2009). However, the identification procedure of the

functional networks from a given individual dataset is not trivial. It is still a great challenge to identify a functional network that is corresponding to the pre-defined spatial template (e.g. DMN) from the decomposition results. Faced with the vast diversity within the individual brains and their decomposed functional networks, visual inspection or manual check have been typically adopted for accurate network identification (Kelly et al., 2010), in addition to the template matching based on spatial map similarities. Several automatic approaches were developed incorporating temporal and spatial patterns from paradigm templates to identify corresponding networks (Calhoun et al., 2008, 2005; Perlberg et al., 2007). Advanced techniques employing machine learning algorithms were also applied for network identifications by using representing networks in multi-dimensional space as signatures (De Martino et al., 2007). Yet most of the current works for network identification are training-based, thus the performance is depended on the (usually limited) data availability. In addition, the machine learning methods often add another layer of complexity into the decomposition model, leading to increased time cost and decreased robustness. Therefore, a fast and robust pattern-specific functional network identification modeling is highly demanded for supporting the high-throughput neuroimaging bigdata analytics and aiding the clinical diagnosis of brain disorders based on functional brain imaging data.

Inspired by many flexible frameworks that can supervise the learning procedure with neuroscience knowledge (Lv et al., 2015d; Zhao et al., 2015), and recognizing the needs for a fast functional network identification tool and based on the previous success in using dictionary learning for fMRI analysis (Lv et al., 2015b), in this work, we proposed a supervised dictionary learning model guided by the pre-defined templates. The model

extended the scalable fast rank-1 dictionary Learning (r1DL) (Li et al., 2016) by simply and efficiently initializing rows of coefficient matrix as template networks (TN), aiming to discover the functional network from the massive voxel-wise fMRI data correspondent to the template using a fast rank-1 matrix decomposition algorithm. The model results on the Autism Brain Imaging Data Exchange (ABIDE) database show that S-r1DL can fast and accurately identify the functional networks based on the 5 templates related with ASD dysfunctional abnormalities.

Materials and Methods

Supervised Dictionary Learning based on Rank-1 Decomposition

In this work we used the resting-state fMRI (rsfMRI) data from the Autism Brain Imaging Data Exchange (ABIDE) database (["http://fcon_1000.projects.nitrc.org/indi/abide/,"](http://fcon_1000.projects.nitrc.org/indi/abide/) n.d.). The rsfMRI data were collected from 79 ASD patients and 105 normal controls from NYU Langone Medical Center. The acquisition parameters were as follows: 240mm FOV, 33 slices, TR=2s, TE=15ms, flip angle=90°, scan time=6mins, voxel size=3×3×4mm. In order to have a balanced dataset for the later classification analysis, we used 78 subjects from patients and 78 subjects from normal controls. Data preprocessing includes motion correction, spatial smoothing, temporal pre-whitening, slice time correction, global drift removal, and linear registration to the MNI space.

We used 5 network templates for the supervise dictionary learning which are reported with ASD-related dysfunctional abnormalities in literatures (Stigler et al., 2011). These templates include regions of: 1) Fusiform Gyrus, which has been reported with hypoactive in ASD patients (Stigler et al., 2011); 2) Inferior Frontal Gyrus (IFG), the left

part of IFG was reported to have decreased activation in ASDs compared to controls ; 3) Anterior Cingulate Cortex (ACC), which has been reported to have decreased activation in adults with autism (Showcase et al., 2007); 4) Posterior Cingulate Cortex (PCC), where significantly lower connectivity was reported in ASDs during resting state (Cherkassky et al., 2006). In addition, Default Mode Network (DMN) has also been examined and discussed by enormous ASD researches (Kana et al., 2014). Templates defining the 4 brain regions were extracted from the Harvard-Oxford atlas (“http://neuro.imm.dtu.dk/wiki/Harvard-Oxford_Atlas,” n.d.), templates defining the DMN was obtained by the group-wise Independent Component Analysis (ICA) (Smith et al., 2009).

Data Acquisition and Pre-processing

In this work, the functional network decomposition and identification is achieved by supervised rank-1 dictionary learning method, by extending the data-driven rank-1 dictionary learning algorithm (Li et al., 2016). Given the input data matrix S of dimension $T \times P$, where for fMRI data T is the temporal length measured in volumes and P is the total number of voxels, we aim to learn a series of basis vector pairs $[u, v]$ from S with the following energy function:

$$L(u, v) = \|S - uv^T\|_F, \text{ s. t. } \|u\| = 1, \|v\|_0 \leq r. \quad 3.1$$

Equation 3.1 indicates that u and v are supposed to span (i.e. outer product) a rank-1 matrix which approximates S , while at the same time the total number of non-zero elements in v is smaller or equal to the pre-defined sparsity constraint r . The minimization of Equation 3.1 is followed by the alternative updating of u and v . Given vector v as constant, the estimation of vector u is just a matrix-vector multiplication:

$$u = \underset{u}{\operatorname{argmin}} \|S - uv^T\|_F = \frac{Sv}{\|Sv\|} \quad 3.2$$

While given vector u as constant, the estimation of vector v involves vector-matrix multiplication and a vector partitioning to find the r number of largest elements in v :

$$v = \underset{v}{\operatorname{argmin}} \|S - uv^T\|_F, s. t. \|v\|_0 \leq r. \quad 3.3$$

The alternative updating is finished when the results have converged:

$$\|u^{j+1} - u^j\| < \varepsilon, \varepsilon = 0.01. \quad 3.4$$

In the analysis of fMRI data, u which is a $T \times 1$ vector characterize the temporal pattern of the decomposed functional network over a total of T time points, while v which is a $P \times 1$ vector characterize the spatial pattern of the network over the P number of voxels.

As we are also aiming for identifying the network that resembles the given spatial template defined in the $P \times 1$ vector TN (or the absence of that template network from the data), the estimated v representing the spatial pattern is supposed to be as similar to TN as possible, thus extending the above rank-1 dictionary learning to the supervised rank-1 dictionary learning (S-r1DL) framework. There exist various methods in supervising the learning process and pulling the updating of v towards TN . Based on preliminary studies we found that simply initializing v as TN could work well for the network identification purpose. Also, the initialization strategy has advantages over other supervised learning methods as its results are the most faithful representation of the input data. Specifically, before the updating loop v is set to TN before starting the estimation of u in the updating loop. The result v is compared with TN based on the spatial overlap rate (SOR):

$$S(v, TN) = \frac{\sum_{i=1}^P \min(v_i, TN_i)}{\sum_{i=1}^P (v_i + TN_i)/2} \quad 3.5$$

which is the ratio of the sum of the minimum value in each of the i -th voxel over the summation of the averaged values of each i -th voxel of network defined in v and template defined in TN. We then determine whether the network has been identified based on the SOR using an empirically-determined threshold of 0.2. SOR value smaller than 0.2 indicates that supervised learning result deviates too far from the initialization thus the template network could not be identified from the data S . After identifying the first functional network, the algorithm will deflate the input matrix S to its residual R by subtracting from the rank-1 matrix spanned by $[u, v]$:

$$R^n = R^{n-1} - uv^T, R^0 = S, 1 < n \leq K, \quad 3.6$$

then learn the next pair of dictionary basis following steps described in Equation 3.2, 3.3, 3.4, to the total number of K dictionaries. The learning of the consequent dictionaries can be supervised (using another TN) or unsupervised where the vector u is initialized by randomly selecting one signal from R . In this study, we perform the network identification individually for each template. Thus for one input data S we will perform the learning individually for 5 times (as we are using 5 templates). The algorithm pipeline of the S-r1DL framework using one template is illustrated in Fig. 3.1 (a).

For comparison, we also tested the performance of unsupervised dictionary learning. In order to identify the functional networks correspondent to the template TN from the learning results, we calculated the SOR value between TN and each of the network using Equation 3.6, then selected the network with the maximum SOR value. This method has been used extensively for the network-based analysis both in our previous studies and in various literatures (Wang and Peterson, 2008). The algorithm pipeline of the

unsupervised r1DL framework for functional network identification is illustrated in Fig. 3.1 (b).

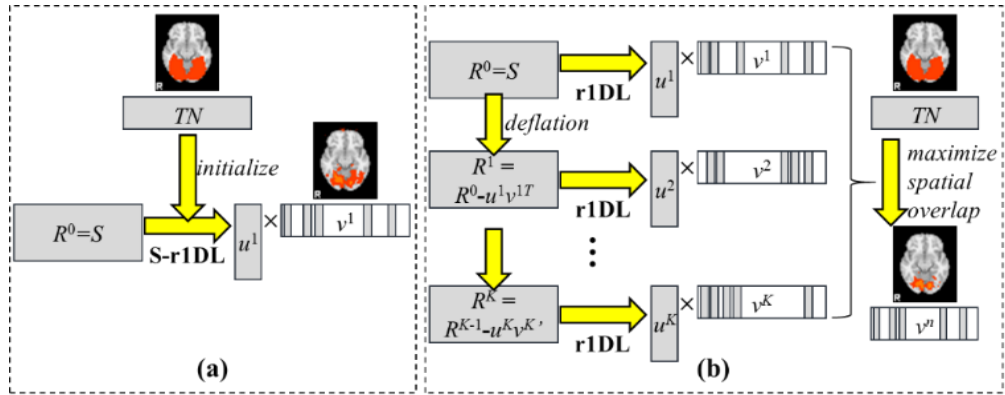


Fig. 3.1. (a) Illustration of the algorithm pipeline of the S-r1DL model with the visualization of the spatial maps of one sample template network (marked as “TN”) and the result network defined in v^1 . (b) Illustration of the alternative strategy based on the unsupervised r1DL model. The same template network is visualized as “TN”. The identified network based on spatial overlap method is defined in v^n and also visualized.

Experimental Results

Functional network identification result based on 5 templates

By performing the S-r1DL on the rsfMRI data of 156 subjects (78 normal controls, 78 patients) from the ABIDE dataset using 5 templates, we identified the corresponding 5 functional networks (or the absence of them) on each subject. An illustration showing the network identification result of one sample subject is shown in Fig. 3.2 below.

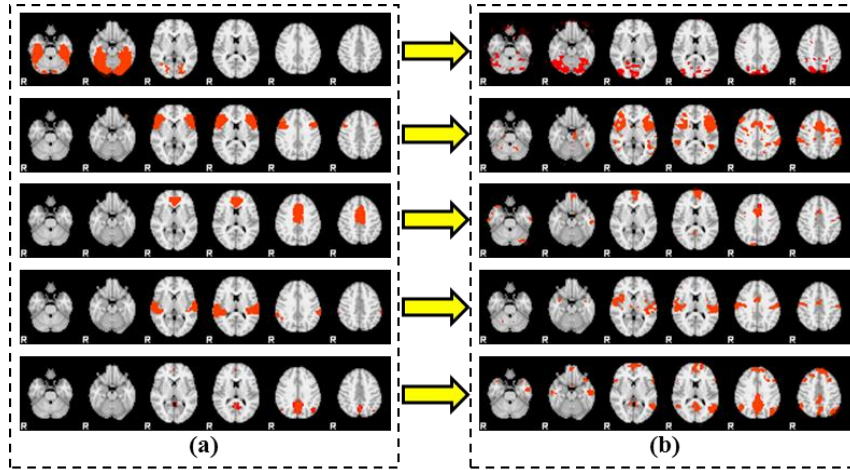


Fig. 3.2. (a) Visualization of the 5 template networks (from top to bottom): Fusiform, IFG, ACC, PCC and DMN. (b) Visualization of the functional networks identified by the S-r1DL method based on the 5 templates.

On the other hand, we have observed substantial variations among the spatial patterns of the identified functional networks across different subjects. Taken Fusiform Gyrus as an example, this network was identified (with SOR value greater than 0.2) from 23 normal control subjects and 34 patient subjects out of the total 156 subjects. The spatial patterns of all the networks are summarized in Fig. 3.3, showing that individual results are highly deviated, but still capable of being identified by the S-r1DL model. The number of subjects with the identified functional networks correspondent to the 5 templates are listed in Table 3.1, showing consistent observation with the literature reports regarding the hyper/lower activities in these regions except the DMN. As in our study it was found that DMN could be recovered as one of the most dominant networks from almost all the subjects regardless of their group.

Table 3.1. Number of subjects (out of 78) in each of the normal control (NC) and ASD patient group with the identified functional networks based on the 5 templates. Results from S-r1DL are listed in the top two rows, results from unsupervised r1DL are listed in the bottom two rows.

Method	Group	Fusiform	IFG	ACC	PCC	DMN
S-r1DL	NC	23	56	64	48	78
	Patient	34	47	63	36	73
r1DL	NC	17	52	53	45	78
	Patient	22	48	60	31	77

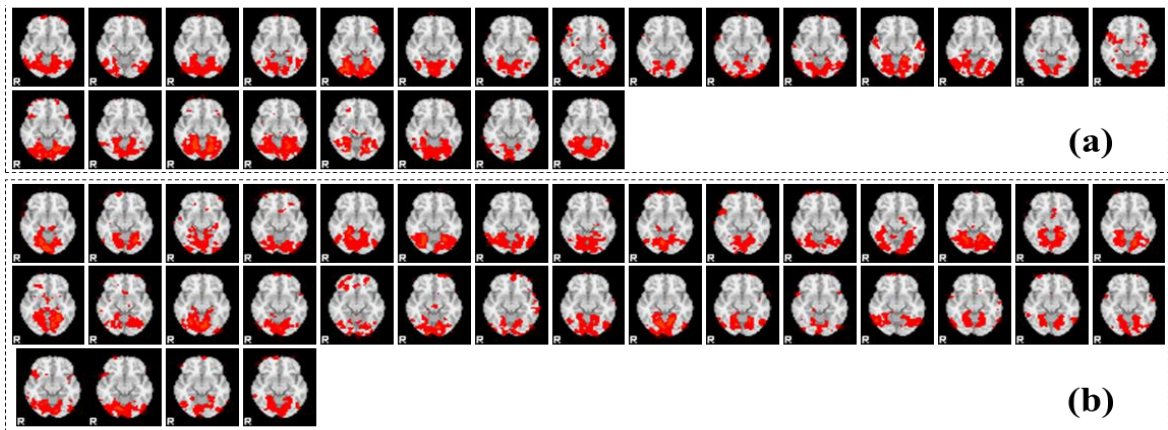


Fig. 3.3. Visualization of the (a) 23 functional networks identified from normal control, and (b) 34 functional networks from ASD patient subjects based on the Fusiform Gyrus template.

Performance comparison between supervised and unsupervised learning

For comparison, we have also performed the unsupervised dictionary learning by r1DL on the same dataset using $K=100$ which supposed to be large enough to cover all the potential functional networks during resting-state. The correspondent functional networks based on the 5 templates were then identified by the SOR maximization method described in 2.1. Results from the unsupervised learning show that most of the functional networks

identified by the S-r1DL can be found in the 100 networks learned by r1DL (with SOR value greater than 0.2), as listed in Table. 1. The spatial patterns of the identified functional networks by the two methods are also very similar, with the average SOR value greater than 0.7. The consistent results from the supervised and unsupervised learning validates the accuracy of the S-r1DL, showing that the supervised learning process was not biased towards the template.

However, it is worth noting that the time cost for analyzing one subject using one template is only around 2 seconds by S-r1DL. Thus for analyzing the whole group of ABIDE dataset, it only took around 6 minutes to finish the network identification of one template. On the other hand, as we need to learn all the 100 functional networks from each individual dataset using the unsupervised learning and calculate their SOR value with the template, the time cost is much greater. It would take more than 10 minutes to analyze one subject using one template, and around 30 hours to finish the analysis on the whole group of data.

Conclusion

In this work we proposed the supervised dictionary learning model based on rank-1 matrix decomposition aiming at identifying functional networks from fMRI data based on pre-defined network templates. While we have observed differences in the networks identified from normal control and patient subjects, we are still investigating the intrinsic alterations in the spatial patterns that causing such difference at individual level. More detailed analysis of the identification process including study on decomposition residuals would be needed before we can perform effective and accurate diagnosis based on the functional network identification results. Another important issue we are trying to address

in the ongoing work is on the large spatial variations of the functional networks across individuals which has been observed in the results. Such large variability will decrease the accuracy of the spatial similarity measurements (e.g. the SOR used in this work). In that sense, traditional network identification methods including manual inspection which is similar to the unsupervised learning scheme used in this work could be suffered from the inaccurate template matching caused by inter-subjects variability and registration. While S-rIDL has been shown to be capable of recovering the highly deviated functional networks, we are developing algorithms to re-align the identification results in order to enable more accurate comparisons.

3.3 Modeling Task fMRI Data via Deep Convolutional Autoencoder

This section is accomplished by my colleague Heng Huang in CAID lab, assisted by me. I will briefly introduce the main methodology and achievements instead of the whole details. This section described a new deep learning based framework for temporal analysis of fMRI data. Compared with traditional sparse dictionary learning framework, deep learning framework exhibit more promising performance.

Abstract

Task based fMRI (tfMRI) has been widely used to study functional brain networks. Modeling tfMRI data is challenging due to two main problems: the lack of the ground truth of underlying neural activity and the intrinsic structure of tfMRI data is highly complex. To deal with the first problem, data-driven approaches were proposed, for instance, Independent Component Analysis (ICA) and Sparse Dictionary Learning (SDL). However, both ICA and SDL only build shallow models and they are under the strong assumption

that original fMRI data could be linearly decomposed into time series components with their corresponding spatial maps. As growing evidences show that the function of human brain is hierarchically organized, new approaches that can infer and model the hierarchical structure of brain networks are widely called for. Recently, deep convolutional neural network (CNN) has drawn much attention, in that deep CNN has been proven to be a powerful method for learning high-level and mid-level abstractions from low-level raw data. Inspired by the powerfulness of deep CNN, in this study, we developed a new neural network structure based on CNN, called Deep Convolutional Auto Encoder (DCAE), in order to take the advantages of both data-driven approach and CNN hierarchical feature abstraction ability for the purpose of learning mid-level and high-level features from complex tfMRI time series in an unsupervised manner. The DCAE has been applied and tested on the publicly available human connectome project (HCP) tfMRI datasets, and promising results are achieved.

Introductions

Task-based functional magnetic resonance imaging (tfMRI) has been a powerful noninvasive tool to study functional networks and cognitive behaviors of the human brain. It has significantly advanced our knowledge of brain areas that are functionally involved in specific tasks. To model the very informative and complex tfMRI time series, a variety of methods have been proposed over the last decades, such as the general linear model (GLM) (Friston et al., 1994) a model-driven approach and independent component analysis (ICA) (McKeown and Sejnowski, 1998) and sparse dictionary learning (Lv et al., 2015b) as data-driven approaches. However, all of these existing methods only build shallow models which can't meet the needs of modeling the reported hierarchical structures in

tfMRI data (Meunier et al., 2009)(Ferrarini et al., 2009), thus overlooking rich information contained in tfMRI.

Among all available deep learning models, convolutional neural network (CNN) is one of the most noteworthy methods. Basically, CNN is a variation of feed-forward artificial neural network, and its connectivity between layers is inspired by animal visual cortex. A recently study on neuroimaging data found similar structures on human brains (De Valois et al., 1982), which suggested CNN is naturally suitable for modeling fMRI data. In this study, we designed a new CNN structure for modeling tfMRI data, called deep convolutional auto-encoder (DCAE). The major advantages and novelties of our DCAE are that the whole training process is on an unsupervised manner and no single label is needed, and meanwhile the DCAE structure inherits the powerful feature abstraction ability of traditional CNNs. To evaluate the effectiveness of our DCAE in modeling fMRI data, we proposed two novel validation approaches. First, we used hypothetical models to verify that the hierarchical feature maps learned by DCAE are related to the underlying neural activities. Second, we performed a comparison study between high-level sparse dictionary learning (HL-SDL) with traditional shallow sparse dictionary learning (SL-SDL) to explore the induced influences when the model depth increases. Experimental results on the human connectome project (HCP) tfMRI datasets demonstrated quite promising results.

Methods

Fig. 3.4 summarizes the structure of the proposed deep convolutional auto encoder (DCAE), as well as the relations with the two validation studies. The DCAE model is consisted of two main components, the encoder and the decoder, to hierarchically model

tfMRI data. Specifically, after data preprocessing, the whole-brain tfMRI signals in each subject are extracted and normalized (with zero mean and standard deviation). As shown in Fig. 3.4, for each voxel's signal, the encoder represents it as mid-level and high-level feature maps as layer goes deeper, and then the decoder decodes these feature maps layer by layer and finally reconstructs the original signal at its top layer. The objective of the DCAE is to minimize the reconstruction errors of the entire training sets (voxels' signals from all subjects), and the whole training process is completely unsupervised.

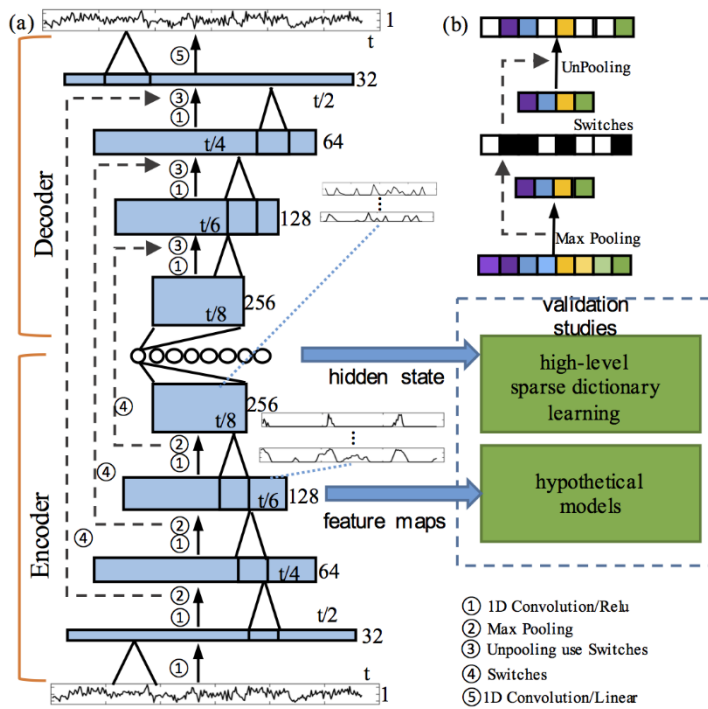


Fig. 3.4. Illustration of the DCAE structure with its two validation studies. (a) DCAE structure, the hidden states, and the feature maps, as well as their relationship with the validation studies. (b) The detailed processes of max pooling and unpooling.

Experimental Results

DCAE Reconstruction Error Analysis

First, to quantitatively examine the signal reconstruction ability of DCAE, the

Pearson correlation coefficients between original fMRI signals and reconstructed signals in the validation data were calculated. Fig. 3.5(a) and Fig. 3.5(b) are the reconstruction results of DCAE and SL-SDL, respectively. Although both the methods reconstructed the original signals quite well (most of the correlations are large than 0.9), we can still see two main differences. First, DCAE has significantly better reconstruction performance (higher average correlations) compared to SL-SDL, as shown in Fig. 3.5. Second, when using DCAE to reconstruct the whole brain signals, there is clear boundary between adjacent brain areas (highlighted with black arrows), while there is less similar observation in SL-SDL.

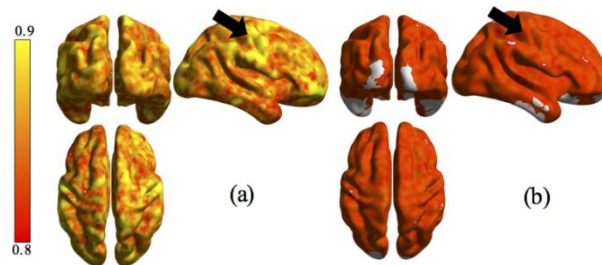


Fig. 3.5. Comparison of the performance in reconstructing the fMRI signals by different methods. (a) The Pearson correlation coefficients between the reconstructed signals by DCAE and the original signals; (b) The Pearson correlation coefficients between the reconstructed signals by SL-SDL and the original signals.

Comparison of HL-SDL and SL-SDL

We also performed a validation study on the hidden state on the top layer of encoder, called HL-SDL. Fig. 3.4 already summarized the computational framework of HL-SDL and SL-SDL. Briefly, except that the input is different, the parameter settings for HL-SDL and SL-SDL are the same. Both models learned 400 dictionaries with 0.7 sparsity regularizer. As the original fMRI signal length is 284, in order to keep the dimensions, we

set the output of the encoder in DCAE as 284. After training, the high-level features are decomposed as high-level dictionaries and corresponding spatial distributions. Then we use decoder to project these high-level dictionaries (time series patterns) back to signal space. Fig. 3.6 shows the results of task-related patterns. Although SL-SDL has learned all the six task patterns, these patterns are mixed with a large number of random noises, and their correlation values with task design are quite small (see PCCs in Fig. 3.6). On the other hand, the HL-SDL contained much less noises in both of the time series patterns and spatial maps. We can conclude that the DCAE model filters noises in each layer, and at the same time it keeps the important information of brain activities. In addition to these basic task related patterns, we also observed more meaningful patterns in HL-SDL compared to SL-SDL, as shown in Fig. 3.7 with examples.

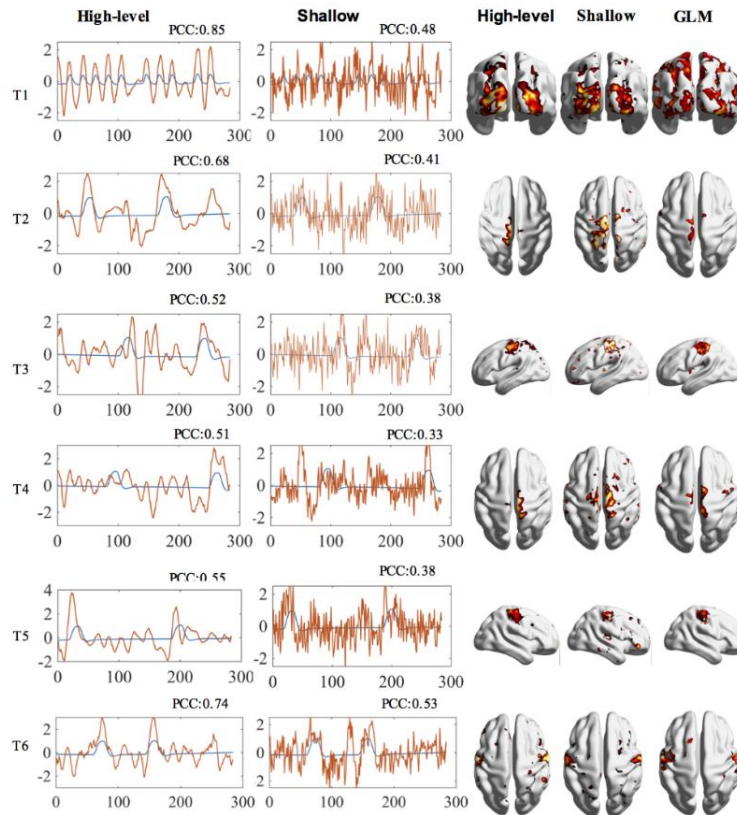


Fig. 3.6. The comparison of HL-SDL and SLSDL in detecting task related patterns. GLM results are shown here as the reference of comparison.

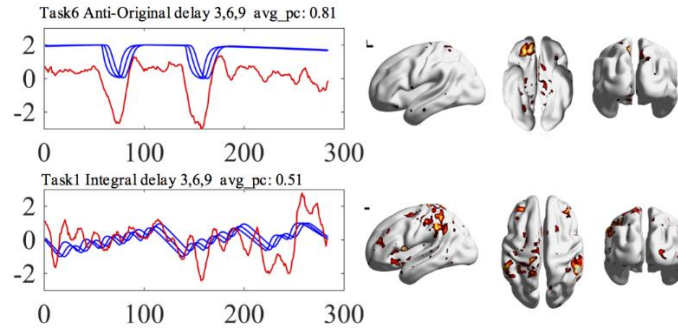


Fig. 3.7. Exemplar patterns that are only found in HL-SDL. Avg_{pc} denote the average correlation between the group of hypothetical models and time series patterns.

Discussion and Conclusion

In this section, we proposed a novel variant of convolutional neural network (CNN) – the deep convolutional auto encoder (DCAE). Our DCAE hierarchically models fMRI data in an unsupervised manner. We used the HCP motor task as a testing bed in this study, and our experiments showed promising results. In order to verify whether the features learned by DCAE are meaningful, we not only visualized the structures of trained DCAE, but also performed two validation studies on feature maps and final output of encoder. We summarize our results as three interesting observations. 1) By visualizing DCAE model, we found the feature maps gain a higher level abstraction of the fMRI signal along with the model depth increasing. In the first layer, the DCAE model tries to learn the shapes of fMRI signals; in the second layer, DCAE begins to have a logical sense of fMRI data. In the third and fourth layers, the patterns become more complex, as it modeled multiple-frequency patterns in one feature map. 2) By using hypothetical models of brain responses, we confirmed that the feature maps are not randomly generated, and the depth greatly

influences the chance to find these hypothetical models. 3) The comparison study of shallow sparse dictionary learning and the high-level dictionary learning showed that the high-level features are superior in task-related regions detection. At the same time, more intrinsic networks could be detected compared to traditional shallow sparse dictionary learning. In our future works, we will conduct more investigations into the feature maps learned by DCAE and develop novel methods to examine spatial distributions of these feature maps on the entire brain, in order to provide more neuroscientific insights of the DCAE.

Since deep learning framework have been successfully applied on both spatial and temporal analysis of fMRI data, a combination of both analyses is desperately needed for better comprehensive analysis of fMRI data. Based on the foundations of both spatial and temporal analyses using deep learning framework and by investigating the intrinsic and intricate correlations of the spatial and temporal features of fMRI, such a comprehensive framework was then proposed and described in chapter 4 concluding both chapter 2 and 3.

CHAPTER 4

SPATIO-TEMPORAL FUNCTIONAL ANALYSIS FOR BRAIN FMRI IMAGES¹⁰

¹⁰ Modeling 4D fMRI Data via Spatio-Temporal Convolutional Neural Networks (ST-CNN)
Yu Zhao, Xiang Li, Wei Zhang, Shijie Zhao, Milad Makkie, Mo Zhang, Quanzheng Li, Tianming Liu.
Accepted by, MICCAI, 2018.
Reprinted here with the permission of publisher.

4.1 Modeling 4D fMRI Data via Spatio-Temporal Convolutional Neural Networks (ST-CNN)

Abstract

Simultaneous modeling of the spatio-temporal variation patterns of brain functional network from 4D fMRI data has been an important yet challenging problem for the field of cognitive neuroscience and medical image analysis. Inspired by the recent success in applying deep learning for functional brain decoding and encoding, in this work we propose a spatio-temporal convolutional neural network (ST-CNN) to jointly learn the spatial and temporal patterns of targeted network from the training data and perform automatic, pin-pointing functional network identification. The proposed ST-CNN is evaluated by the task of identifying the Default Mode Network (DMN) from fMRI data. Results show that while the framework is only trained on one fMRI dataset, it has the sufficient generalizability to identify the DMN from different populations of data as well as different cognitive tasks. Further investigation into the results show that the superior performance of ST-CNN is driven by the jointly-learning scheme, which capture the intrinsic relationship between the spatial and temporal characteristic of DMN and ensures the accurate identification.

Introduction

Recently, analytics of the spatio-temporal variation patterns of functional Magnetic Resonance Imaging fMRI (Heeger and Ress, 2002) has been substantially advanced through machine learning (e.g. independent component analysis (ICA) (Cole et al., 2010; McKeown et al., 2003) or sparse representation (Lv et al., 2015a; S. Zhao et al., 2018)) and deep learning methods (Litjens et al., 2017; Zhao et al., 2017a). As fMRI data are acquired

as series of 3D brain volumes during a span of time to capture functional dynamics of the brain, the spatio-temporal relationships are intrinsically embedded in the acquired 4D data which need to be characterized and recovered.

In literatures, the spatio-temporal analytics methods can be summarized into two groups: the first group performs the analysis on either spatial or temporal domain based on the corresponding priors, then regress out the variation patterns in the other domain. For example, temporal ICA identifies the temporally independent “signal source” in the 4D fMRI data, then obtains the spatial patterns of those sources through regression. Recently proposed deep learning-based Convolutional Auto-Encoder (CAE) model (Huang et al., 2017), temporal time series, and spatial maps are regressed later using resulting temporal features. Sparse representation methods, on the other hand, identify the spatially sparse components of the data, while the temporal dynamics of these components are obtained through regression. Works in (Hjelm et al., 2014) utilizes Restricted Boltzmann Machine (RBM) for spatial feature analysis ignores the temporal feature.

The second group performs the analysis on spatial and temporal domain simultaneously. For example, (Shen et al., 2014) applies Hidden Process Model with spatio-temporal “prototypes” to perform the spatio-temporal modeling. Another effective approach to incorporate temporal dynamics (and relationship between time frames) into the network modeling is through Recurrent Neural Network (Hjelm et al., 2016). Inspired by the superior performance and the better interpretability of the simultaneous spatio-temporal modeling, in this work we proposed a deep spatio-temporal convolutional neural network (ST-CNN) to model the 4D fMRI data. The goal of the model is to pinpoint the targeted functional networks (e.g., Default Mode Network DMN) directly from the 4D

fMRI data. The framework is based on two simultaneous mappings: the first is the mapping between the input 3D spatial image series and the spatial pattern of the targeted network using a 3D U-Net. The second is the mapping between the regressed temporal pattern of the 3D U-Net output and the temporal dynamics of the targeted network, using a 1D CAE. Summed loss from the two mappings are back-propagated to the two networks in an integrated framework, thus achieving simultaneous modeling of the spatial and temporal domain. Experimental results show that both spatial pattern and temporal dynamics of the DMN can be extracted accurately without hyper-parameter tuning, despite remarkable cortical structural and functional variability in different individuals. Further investigation shows that the framework trained from one fMRI dataset (motor task fMRI) can be effectively applied on other datasets, indicating ST-CNN offers sufficient generalizability for the identification task. With the capability of pin-pointed network identification, ST-CNN can serve as a useful tool for cognitive or clinical neuroscience studies. Further, as the spatio-temporal variation patterns of the data are intrinsically intertwined within an integrated framework, ST-CNN can potentially offer new perspectives for modeling the brain functional architecture.

Method and Materials

ST-CNN takes 4D fMRI data as input and generates both spatial map and temporal time series of the targeted brain functional network (DMN) as output. Different from CNNs for image classifications (e.g. (Krizhevsky et al., 2012)), ST-CNN consists of a spatial convolution network and a temporal convolution network, as illustrated in Fig. 4.1(a). The targeted spatial network maps of sparse representation on fMRI data (Lv et al., 2015a) are

used to train the spatial network of ST-CNN, while the corresponding temporal dynamics of the spatial networks are used to train the temporal networks.

Experimental data and preprocessing

We use the Human Connectome Project (HCP) Q1 and S900 release datasets (Barch et al., 2013) for the experiments. Specifically, we use motor task-evoked fMRI (tfMRI) for training the ST-CNN, and test its performance using the motor and emotion tfMRI data from Q1 release and motor task tfMRI data S900 release. The preprocessing pipelines for tfMRI data include skull removal, motion correction, slice time correction, spatial smoothing, global drift removal (high-pass filtering), all implemented by FSL FEAT.

After preprocessing, we apply sparse representation method (Lv et al., 2015a) to decompose tfMRI data into functional networks on both training and testing data sets. The decomposition results consist of both the temporal dynamics (i.e. “dictionary atoms”) and spatial patterns (i.e. “sparse weights”) of the functional networks. The individual targeted DMN is then manually selected based on the spatial patterns of the resulting networks. The selection process is assisted with sorting the resulting network by their spatial overlap rate with the DMN template (from (Smith et al., 2009)), measured by Jaccard similarity (i.e. overlap over union). We use the dictionary (1-D time series) of the selected network as ground-truth time series for training the CAE.

ST-CNN framework

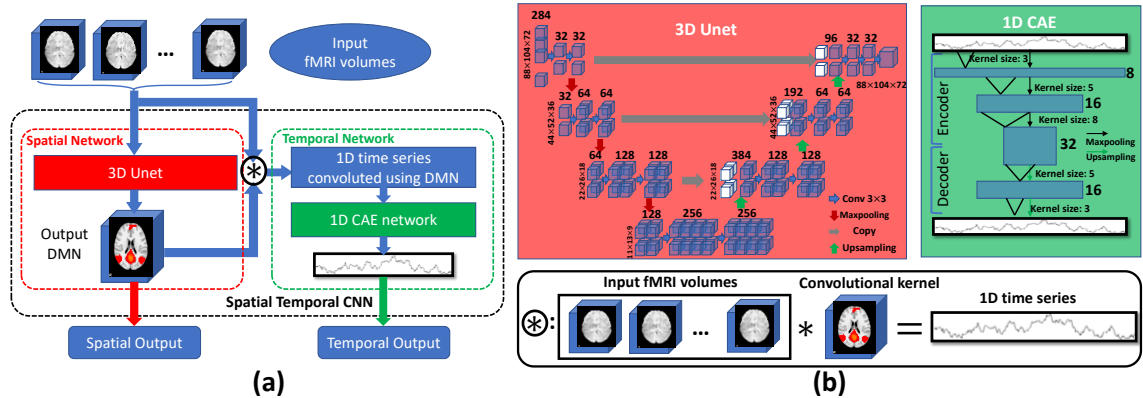


Fig. 4.1. (a). Algorithmic pipeline of ST-CNN; (b). Spatial network structure, temporal network structure, and the combination of the spatial and temporal domain.

— Spatial Network

The spatial network is inspired from the 2D U-Net (Ronneberger et al., 2015) for semantic image segmentation. By extending and adapting the 2D classification U-Net to a 3D regression network (Fig. 4.1 (b)), the spatial network takes 4D fMRI data as input, each 3D brain volume along the time frames is assigned to one independent channel. Basically, this 3D U-Net is constructed by a contracting CNN and an expanding CNN, where the pooling layers (red arrows in Fig. 4.1 (b)) in the contracting CNN are replaced by upsampling layers (green arrows in Fig. 4.1 (b)). This 3D U-shaped CNN structure contains only convolutional layers without fully connected layers. Loss function for training the spatial network is the mean squared error between the network output which is a 3-D image and the targeted DMN.

— Temporal Network

The temporal network (Fig. 4.1 (b)) is inspired by the 1-D Convolutional Auto-Encoder (CAE) for fMRI modeling (Huang et al., 2017). Both the encoder and decoder of

the 1-D CAE have the depth of 3. The encoder starts by taking 1-D signal as input and convolving it with a convolutional kernel size of 3, yielding 8 feature map channels, which are down-sampled using a pooling layer. Then a convolutional layer with kernel size 5 is attached, yielding 16 feature map channels, which are also down-sampled using a pooling layer. The last part of the encoder consists of a convolutional layer with kernel size 8, yielding 32 feature map channels. The decoder takes the output of the encoder as input and symmetrize the encoder as traditional auto-encoder structure. Loss function for training the temporal network is negative Pearson correlation (4.1) between the temporal CAE output time series with the temporal dynamics of the manually-selected DMN.

$$\text{Temporal loss} = -\frac{N \sum_1^N xy - \sum_1^N x \sum_1^N y}{\sqrt{(N \sum_1^N x^2 - (\sum_1^N x)^2)(N \sum_1^N y^2 - (\sum_1^N y)^2)}} \quad 4.1$$

– Combination Joint Operator

This combination (Fig. 4.1 (b)) procedure connects spatial network and temporal network through a convolution operator. Inputs for the combination are the 4-D fMRI data and 3-D output from the spatial network (i.e. spatial pattern of estimated DMN). The 3-D output will be used as a 3-D convolutional kernel to perform a valid no-padding convolution over each 3-D volume across each time frame of the 4-D fMRI data 4.2. Since the convolutional kernel size is the same as each 3D brain volume along the 4th (time) dimension, the no-padding convolution will result in a single value at each time frame, thus forming a time series for the estimated DMN. This output time series ts will be used as the input for temporal 1-D CAE, as described above.

$$ts \in \mathbb{R}^{T \times 1} = \{t_1, t_2, \dots, t_T | t_i = V_i * DMN \in \mathbb{R}\}, \quad 4.2$$

where t_i is the convolution result at each time frame, V_i is the 3-D fMRI volume at time frame i , and DMN is the 3-D spatial network output used as convolution kernel.

Training Process and Model Convergence

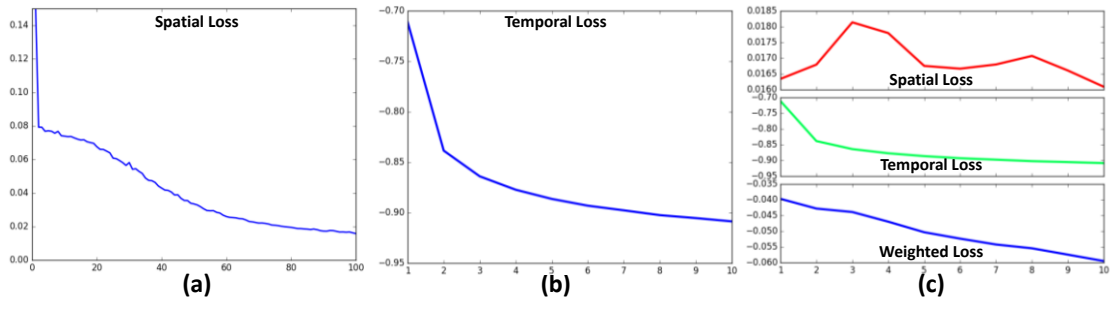


Fig. 4.2. Training losses (y-axis) versus training steps (x-axis). (a). first stage spatial network training loss; (b). second stage temporal network training loss; (c). fine-tuning training loss.

Since the temporal network will rely on the DMN spatial map from the spatial network, we split the training process into 3 stages: at the first stage, only spatial network is trained (Fig. 4.2 (a)); at the second stage, temporal network is trained based on the spatial network results (Fig. 4.2 (b)); and finally, the entire ST-CNN is trained for fine-tuning (Fig. 4.2 (c)). As we can see from Fig. 4.2, the temporal network converges much faster (around 10 times faster) than the spatial network. Thus during the fine-tuning stage, the loss function for ST-CNN is a weighted sum (10:1) of both spatial and temporal loss.

Model Evaluation and Validation

We firstly calculate the spatial overlap rate between the spatial pattern of ST-CNN output and a well-established DMN template to evaluate the performance of spatial network. We then calculate the Pearson correlation of the output time series with ground-truth time series from sparse representation results to evaluate the temporal network. Finally we utilize a supervised dictionary learning method (Zhao et al., 2015) to reconstruct

the spatial patterns of the network based on temporal network result to investigate whether the spatio-temporal relationship is correctly captured by the framework.

Results

We use 52 subjects' motor tfMRI data from HCP Q1 release for training the ST-CNN. We test the same trained network on three datasets: 1) motor tfMRI data from the rest of 13 subjects. 2) motor tfMRI data from 100 randomly -selected subjects in the HCP S900 release. 3) emotion tfMRI data from 67 subjects from HCP Q1 release. Testing results show consistently good performance for DMN identification, demonstrating that trained network is not limited to specific population and specific cognitive tasks.

MOTOR task testing results

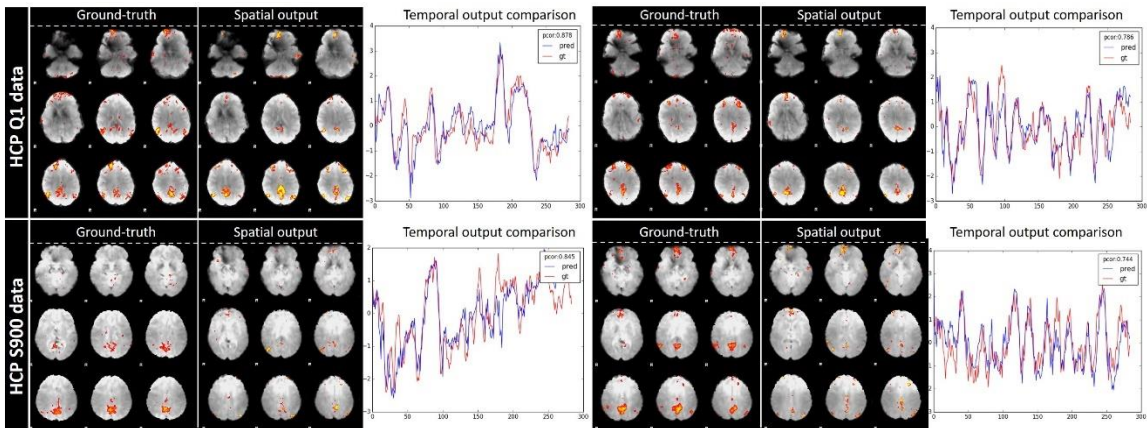


Fig. 4.3. Examples of comparisons between ST-CNN outputs and ground-truth from sparse representation. Here we showed 2 subjects' comparison results from two different datasets (1 HCP Q1 subjects and 1 HCP S900 subjects). Spatial maps are very similar and time series have Pearson correlation coefficient values 0.878 in HCP Q1 data, and 0.744 in HCP S900 data. Red curves are ground-truth. Blue curves are ST-CNN temporal outputs.

The trained ST-CNN is tested on 2 different motor task datasets: 13 subjects from HCP Q1 and 100 subjects from HCP S900, respectively. As shown in Fig. 4.3, the resulting

spatial and temporal patterns are consistent with the ground-truth. Quantitative analyses shown in Table 4.1 demonstrates that the ST-CNN performs better than sparse representation method, although it is trained from the manually-selected results of sparse representation. The rationale is that the ST-CNN can better adapt to the input data by the co-learned spatial and temporal networks, while sparse representation relies on the simple sparsity prior which can be invalid in certain cases. As shown in Fig. 4.4, sparse representation cannot identify DMN from certain subjects while ST-CNN can. In HCP Q1 dataset, we have observed 20% (13 out of 65 subjects) of cases where sparse representation fails while ST-CNN succeeds. Considering the fact that DMN is supposed to be consistently presented in the functioning brain regardless of task, this is an intriguing and desired characteristic of the ST-CNN model.

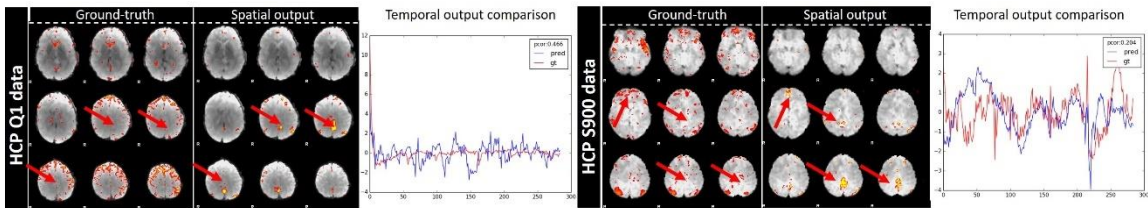


Fig. 4.4. Example of the better DMN identification of ST-CNN than sparse representation (denoted by red arrows). The temporal dynamics of the two networks are also different, where output from sST-CNN (blue) are more reasonable.

Table 4.1. Performance of ST-CNN measured by spatial overlap rate

Datasets	Spatial overlap with DMN template		Temporal similarity (Pearson correlation)
	Sparse Representation	ST-CNN	
HCP Q1 MOTOR (13 subjects)	0.115	0.172	0.55
HCP S900 MOTOR (100 subjects)	0.070	0.066	0.53
HCP Q1 EMOTION (67 subjects)	0.095	0.168	0.51

EMOTION task testing results

The 67 subjects' emotion task-evoked fMRI data (HCP Q1) were further tested to

demonstrate that our trained network based on motor task is not prone to specific cognitive tasks. The ability to extract DMN both spatially and temporally of our framework showed that the intrinsic features of DMN were well captured. As shown in Fig. 4.5 the spatial maps resemble with the ground-truth sparse representation results and so do the temporal outputs. Quantitative analyses in Table 4.1 showed that our outputs also had larger spatial overlap with DMN templates than outputs from sparse representation. The temporal outputs were also shown accurate, with an average Pearson correlation coefficient of 0.51.

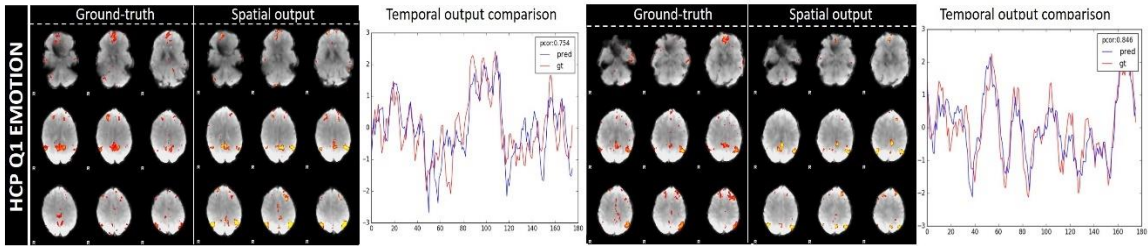


Fig. 4.5. Example of ST-CNN outputs and ground-truth (sparse representation) for EMOTION task. Spatial maps are very similar and time series have Pearson correlation 0.754. Red curve is ground-truth, blue curve is the temporal output by ST-CNN.

Spatial output and temporal output relationship

To perform further validation, supervised sparse representation (Zhao et al., 2015) is applied on 13 testing subjects' HCP Q1 motor task fMRI data. We set the temporal output of ST-CNN as predefined dictionary atoms to obtain the sparse representation on the data by learning the rest of the dictionaries. The resulting network corresponding to the predefined atom, which has the fixed temporal dynamics during the learning, are compared with ST-CNN spatial outputs. We found that the temporal output of ST-CNN can lead to an accurate estimation of the DMN spatial patterns as in Fig. 4.6. The average spatial overlap rate between the supervised results and ST-CNN spatial output is 0.144, suggesting that the spatial output of ST-CNN has close relationship with its temporal output.

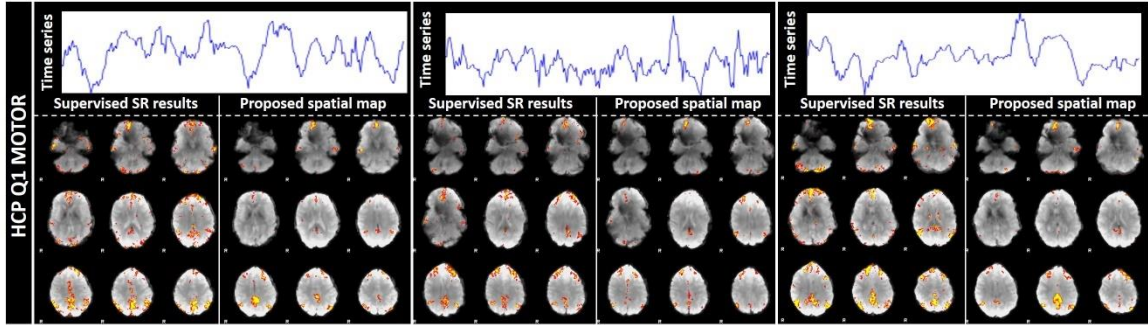


Fig. 4.6. Comparison of the spatial maps between ST-CNN results and supervised sparse representation which takes temporal output of ST-CNN as pre-defined atoms.

Discussion

In this chapter, we proposed a novel spatio-temporal CNN model to identify functional networks (DMN as an example) from 4D fMRI data modelling, concluding the researches in chapter 2 and 3. The effectiveness of ST-CNN is validated by the experimental results on different testing datasets. From an algorithmic perspective, the result shows that ST-CNN embeds the spatial-temporal variation patterns of the 4D fMRI signal into the network, rather than learns the matrix decomposition process by the sparse representation. It is then very important to further refine the framework by training it over DMNs identified by other methods (such as temporal ICA). More importantly, we use DMN as a sample targeted network in the current work, since it should be present in virtually any fMRI data. As detecting the absence/disruption a functional network is as important as identifying it (e.g. for AD/MCI early detection), in the future work we will focus on extending the current framework to pinpoint more functional networks, including task-related networks which should be presented in a limited range of datasets. We will also test ST-CNN on fMRI from abnormal brains for its capability of characterizing the spatio-temporal patterns of the disrupted DMNs. In the extension of this work, we have carried out comprehensive experiments on HCP 900 dataset, and tested the robustness of

model to the noisy inputs. Reproducibility has also been demonstrated on resting state data. The simultaneous and comprehensive 4D spatio-temporal analysis is the future trend for fMRI analysis as the intrinsic features in the raw fMRI data are encoded in a simultaneous and interactive way. Thanks to the deep learning framework, the proposed fully convolutional network is a great fit for 4D fMRI data. The successful application serves as a milestone leading to the future holistic analysis of fMRI.

CHAPTER 5

STRUCTURAL ANALYSES: CROSS IMAGING MODALITY SYNTHESIS FOR MR- ONLY RADIOTHERAPY TREATMENT PLANNING¹¹

¹¹ Towards MR-Only Radiotherapy Treatment Planning: Synthetic CT Generation Using Multi-view Deep Convolutional Neural Networks

Yu Zhao, Shu Liao, Yimo Guo, Liang Zhao, Zhennan Yan, Gerardo Hermosillo, Tianming Liu, Xiang Zhou, Yiqiang Zhan. Accepted by, MICCAI, 2018.

Reprinted here with the permission of publisher.

5.1 Towards MR-Only Radiotherapy Treatment Planning: Synthetic CT Generation Using Multi-view Deep Convolutional Neural Networks

Abstract

Recently, Magnetic Resonance imaging-only (MR-only) radiotherapy treatment planning (RTP) receives growing interests since it is radiation-free and time/cost efficient. A key step in MR-only RTP is the generation of a synthetic CT from MR for dose calculation. Although deep learning approaches have achieved promising results on this topic, they still face two major challenges. First, it is very difficult to get perfectly registered CT-MR pairs to learn the intensity mapping, especially for abdomen and pelvic scans. Slight registration errors may mislead the deep network to converge at a sub-optimal CT-MR intensity matching. Second, training of a standard 3D deep network is very memory-consuming. In practice, one has to either shrink the size of the training network (sacrificing the accuracy) or use a patch-based sliding-window scheme (sacrificing the speed). In this paper, we proposed a novel method to address these two challenges. First, we designed a max-pooled cost function to accommodate imperfect registered CT-MR training pairs. Second, we proposed a network that consists of multiple 2D sub-networks (from different 3D views) followed by a combination sub-network. It reduces the memory consumption without losing the 3D context for high quality CT synthesis. We demonstrated our method can generate high quality synthetic CTs with much higher runtime efficiency compared to the state-of-the-art as well as our own benchmark methods. The proposed solution can potentially enable more effective and efficient MR-only RTPs in clinical settings.

Introduction

Medical imaging plays an important role in radiotherapy treatment planning (RTP) (Edmund and Nyholm, 2017) by providing critical information for organ/tumor localization and dose calculation. Currently computed tomography (CT) is the primary modality, which provides electron density information for dose calculation. Since Magnetic Resonance (MR) imaging is more valuable in organ/tumor localization due to its superior soft tissue contrast, it has received more and more interests in RTP. In traditional workflow, MR will be registered to a principal CT dataset (Edmund et al., 2015; Edmund and Nyholm, 2017) so that its superior soft tissue contrast information can be fused with the CT image. However, due to the imperfectness of the current image registration techniques, registration error will bring systematic spatial uncertainty (Paulson et al., 2014), hence, influencing the accuracy of RTP. Recently MR-only RTP receives growing interests since it is radiation-free and time/cost efficient. A key step in MR-only RTP is the generation of a synthetic CT (sCT) from MR for dose calculation.

The major challenge in CT synthesizing is the intensity ambiguity of different tissues, such as bone and air which both appear dark on MR. Traditional approaches for CT synthesis from MR can be divided into two categories: atlas-based (Sjölund et al., 2015) and segmentation-based (Delpon et al., 2016). For the atlas-based approaches, the focus is to register the MR atlas to the patient MR, and then apply the registration transformation on the corresponding CT atlas to generate the synthetic CT (Dowling et al., 2012). Segmentation-based methods (Delpon et al., 2016) segment different types of tissues from MR. A synthetic CT is then generated by filling a constant CT intensity for each type of tissue. The main obstacles for these approaches are the synthesis speed and registration or

segmentation accuracy. Recently, some context-aware deep learning based models are proposed (Nie et al., 2017, 2016; Wolterink et al., 2017) and they achieved promising results. However, they still face two major challenges. First, standard deep learning requires a set of perfectly registered CT-MR pairs to learn the intensity mapping from MR to CT. However, since MR and CT images are acquired at different time with different patient positionings and table shapes, it is very difficult to perfectly register them, especially for abdomen and pelvic scans (Andreasen et al., 2016). Thus most works (Nie et al., 2017, 2016; Wolterink et al., 2017) focused on brain regions. Slight registration errors may induce large mis-matching in the intensity space, hence, misleading the deep network to converge at a sub-optimal CT-MR intensity matching. Second, training of a standard 3D deep network is very memory-consuming. In practice, even with a high-end deep learning server, one has to simplify the 3D network structure or using a patch-based sliding-window scheme (Nie et al., 2017, 2016) to accommodate large volumes of training data. The simplified network may not model the MR-CT intensity mapping well and sliding-window scheme may sacrifice the speed significantly.

In this work, we proposed a novel method to tackle the aforementioned challenges. First, we designed a maxpooling loss function allowing the network to search optimal intensity matching not only between the corresponding CT-MR patches but across their neighborhood. This kind of “matching freedom” makes the network robust to imperfect CT-MR registration. Second, we proposed a network consisting of multiple 2D sub-networks (from different 3D views) followed by a 3D combination sub-network. It dramatically reduces the memory consumption without losing the 3D context for high

quality CT synthesis. Our method generated high quality sCTs with much higher runtime efficiency compared to the state-of-the-art and our own benchmark method.

Materials and Methods

Overview of multi-view multi-channel U-Net structure

U-Net (Ronneberger et al., 2015) is a deep network originally proposed for image segmentation. It has a symmetric hierarchical structure that enables precise voxel-wise classification by modeling cross-scale anatomical context. In our study, the U-Net is adapted to a regression network, i.e., the output is an image with synthetic CT values. The original U-Net has a 2D fully convolutional structure, which needs to be extended to handle the 3D nature of MR and CT images. In order to train on 3D volumes without reducing network size and speed, we adopt a 2.5D framework (

Fig. 5.1). Our framework consists of two 2D-centric U-Nets (Fig. 5.2) corresponding to sagittal and axial views, respectively. The stacked output 3D features from these two sub-nets are further combined by a 3D combination sub-net (Fig. 5.5). Moreover, to deal with the unpreventable misalignments between MR-CT training pairs for accurate model training, a maxpooling hinge-like Huber function is designed as training loss (Fig. 5.3). Technical details are explained next.

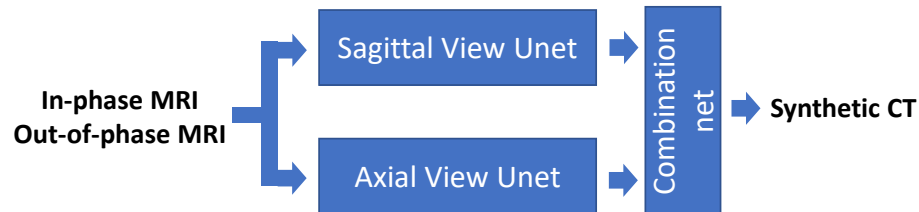


Fig. 5.1. Multi-channel multi-view U-Net based deep fully convolutional network framework.

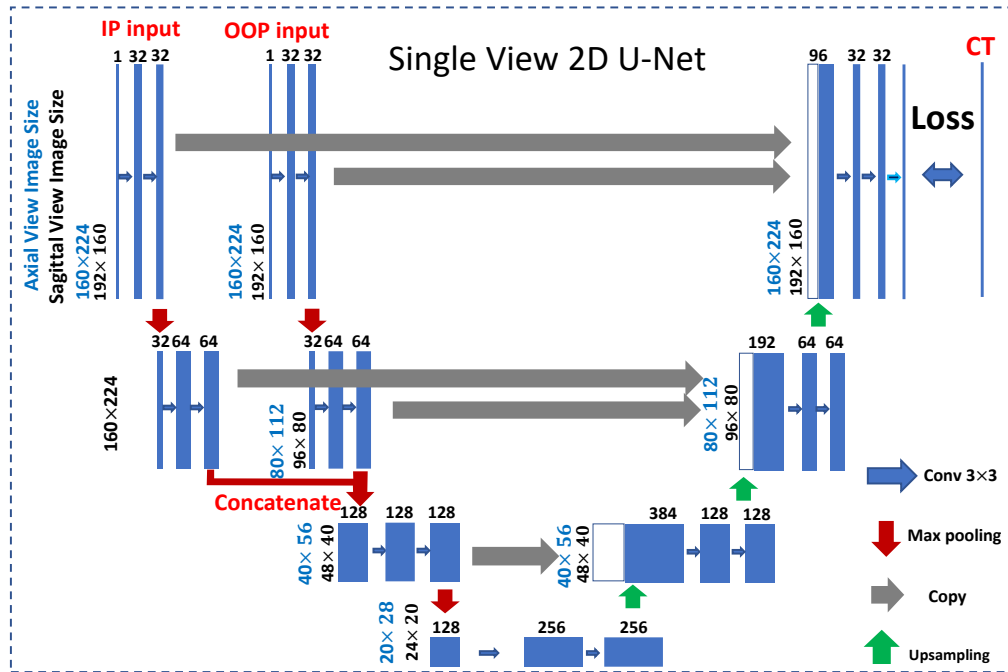


Fig. 5.2. Single view 2D U-Net (sagittal view and axial view). Multi-channel 2D MR slices (In-phase (IP) and Out-of-phase(OOP)) are network inputs. Loss is designed as maxpooling and hinge-like Huber loss.

Multi-channel MR inputs for information enhancement

The input of our method is the In-Phase and Out-of-phase images generated by MR Dixon method. These images capture complementary fat and water information for tissue differentiation. As shown in Fig. 5.2, for single view 2D U-Net inputs, instead of stacking these two images at the input layer, we keep two independent channels for each of them. In this way, the network can capture features from different MR sequences independently for information enhancement.

Maxpooling and hinge-like Huber loss function

In order to learn the intensity transformation from MR to CT, a set of registered MR-CT pairs are needed for training. However, it is very difficult to perfectly register MR and CT due to organ deformations, different table sizes, etc. To address this problem, an

effective loss function is proposed for the network in Fig. 5.2. Instead of calculating the voxel-wise intensity differences directly between the output slice and the ground-truth slice, a maxpooling process (Fig. 5.3 (a)) is applied to accommodate the slight misalignment in a translation invariant fashion (Scherer et al., 2010). A schematic example is presented in Fig. 5.4.

If I_1 and I_2 are perfectly registered, it is easy to learn a consistent mapping function F that maps the intensities of pixel a , b and c to a' , b' and c' , i.e., $F(I_1(a))=I_2(a')$, $F(I_1(b))=I_2(b')$, $F(I_1(c))=I_2(c')$; However, if I_1 and I_2 are not perfectly registered due to deformable or rigid registration errors, it is very difficult to learn a common mapping function that maps the intensities of a , b and c to a' , b' and c' , since the intensity transformation becomes inconsistent. By adding max-pooling in the loss function, we essentially give some spatial freedom to the mapping function, allowing it to map the intensity to its neighborhood, i.e., $F(I_1(a))=I_2(a'+\Delta a')$, $F(I_1(b))=I_2(b'+\Delta b')$, $F(I_1(c))=I_2(c'+\Delta c')$. Thus, a consistent mapping function can be learned. Note that the max-pooling allows different voxels to have different small Δ , which address the non-systematic registration errors. The hinge-like function is also adopted with Huber loss as the final loss function (Fig. 5.3 (b)), also shown in Equation 5.1. It accommodates major loss and ignores minor ones.

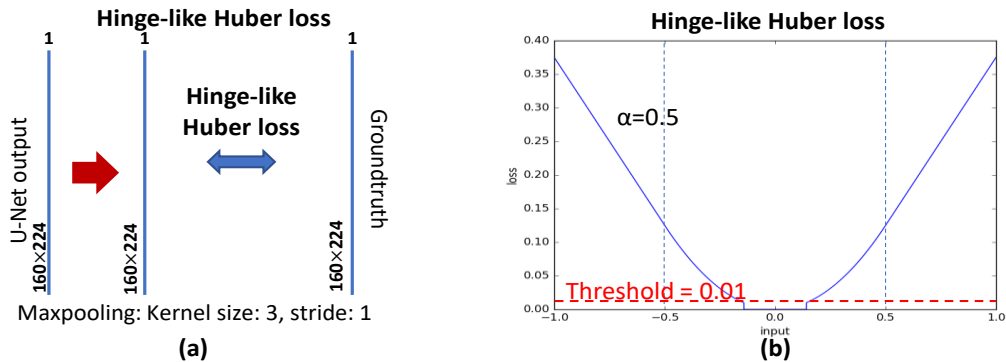


Fig. 5.3. (a) Maxpooling hinge-like Huber loss function for U-Net structure training. (b) Hinge-like Huber loss function.

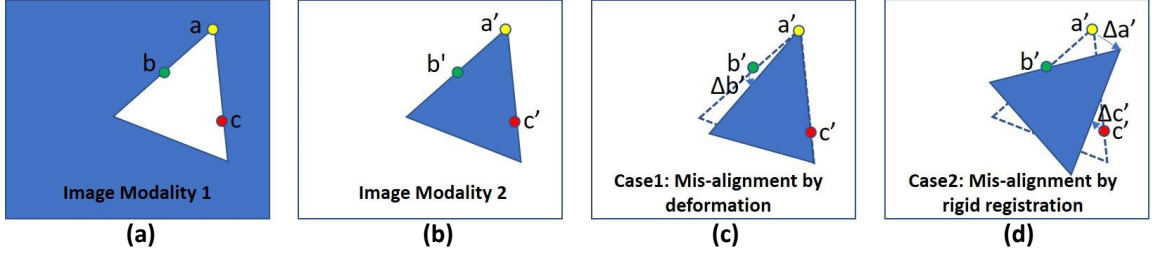


Fig. 5.4. A schematic explanation of the impact of mis-registration to intensity transformation. (a) Image1 (Modality 1), (b) Perfectly registered Image modality 2, (c) Image 2 with rigid mis-alignment, (d) Image 2 with non-rigid mis-alignment. Triangles in (a)-(d) represent the same object. Dashed lines in (c) and (d) denote the locations of the perfectly registered Image2.

$$L(a) = \begin{cases} 0, & |a| < 0.01 \\ \frac{1}{2}a^2, & 0.01 < |a| < \alpha \\ \alpha \left(|a| - \frac{1}{2}\alpha \right), & otherwise \end{cases} \quad 5.1$$

where a is the 2D image slice difference between output and ground-truth CT images.

Multi-view combination of the 2D U-Net like structures

Our network includes two 2D U-Nets followed by a combination network (Fig. 5.5). This design is important to deal with memory limitations. With an 8GB GPU memory, we cannot fit a 3D 192*224*168 volume with a 3D network for training. Therefore, we decompose the 3D volume into 2D axial and sagittal slices, respectively, which can be easily fit into two 2D U-Nets. However, since the 2D U-Net ignores the 3D context across neighboring slides, the output may have stitching blurring effect. (c.f. Fig. 5.7), To remove the blurring effect, output feature maps of 2D U-Nets are stacked into 3D volumes before

feeding into a 3D convolution layer with kernel size $1 \times 1 \times 1$. This 3D convolution layer effectively removes the 2D stitching blurring effect.

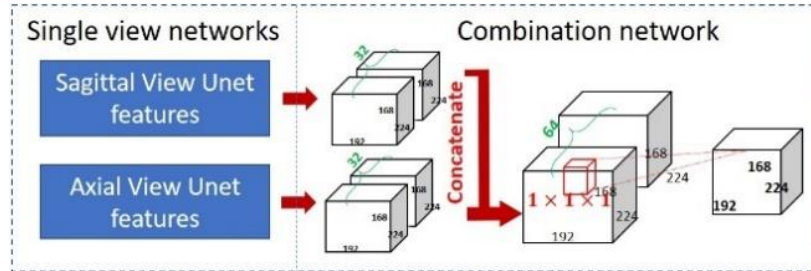


Fig. 5.5. Combination Network. Single view networks include 2D U-Net structures for both axial and sagittal view. 3D combination network takes 32 channels output 3D features from 2 single view networks as input and output a 3D synthetic CT volume.

The overall loss function is a (empirically-set) weighted mean of maxpooling hinge-like Huber loss from two views and a Huber loss of the 3D synthetic volume with the ground-truth volume (5.2).

$$L(v) = 0.6L(v)_{sagittal} + 0.33L(v)_{axial} + 0.07Huber(v) \quad 5.2$$

where v is the 3D volume difference between the final output synthetic CT of the combination net and the ground-truth CT volumes; $L(v)_{sagittal}$ and $L(v)_{axial}$ are the hinge-like Huber loss maxpooled from sagittal view and axial view 2D slices respectively; $Huber(v)$ is the voxel-wise Huber loss of the 3D volume difference.

Network training

Our network training has two stages. First, the two 2D U-Nets of axial and sagittal views are trained independently. Then the feature maps extracted from the second last layers of each 2D U-Net are stacked into 3D volumes and saved as input for further training of the 3D combination network.

Results

Due to the lack of perfectly aligned scanned MR and CT pairs, the ground-truth CTs are generated by a multi-atlas-based regression method (Liao et al., 2016). The quality of the ground-truth synthetic CT image is confirmed and accepted by experienced oncologists. However, since the multi-atlas-based regression method (Liao et al., 2016) takes extensive time (i.e., more than 15 minutes on average) to generate the synthetic CT image, it has limitation in the real world RTP clinical workflow. An Nvidia Quadro M4000 GPU with 8 GB memory was utilized for all the training steps. For the first training stage, training time for each 2D U-Net like structure is dependent on the input size of the images at the corresponding view, 21h and 95h for axial view and sagittal view, respectively. For the second training stage, combination net, 7h was taken. A total time of 123h was used for the 2-stage training procedure. The testing phase only cost less than 8 second for each subject 3D CT volume synthesis.

Effectiveness of the proposed framework

In the experiment, we have 34 MR-CT pairs, where 27 pairs are used for training and the rest 7 pairs for testing. Our proposed method showed significant improvements at 2D slice level compared to the benchmark U-Net structure Fig. 5.6. The multi-view combination of the 2D U-Net structures also showed effectiveness on removing the 2D slice stitching blurs across the 3D volume and avoided sacrificing synthetic image quality by shrinking the size of 3D training network (Fig. 5.7). Comparisons between sCTs generated using our proposed method and ground-truth CTs are discussed in the following sections.

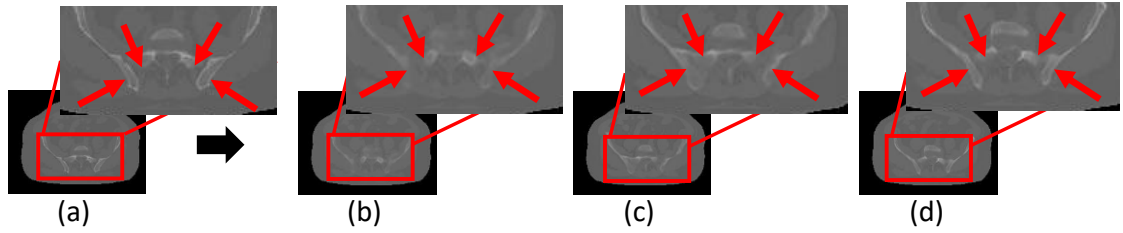


Fig. 5.6. Improved quality of the bone area synthesis compared to benchmark U-Net schemes. (a) Ground-truth sCT used for training; (b) Bench mark result using original U-Net; (c) Result using benchmark U-Net with maxpooling function; (d) Proposed result.

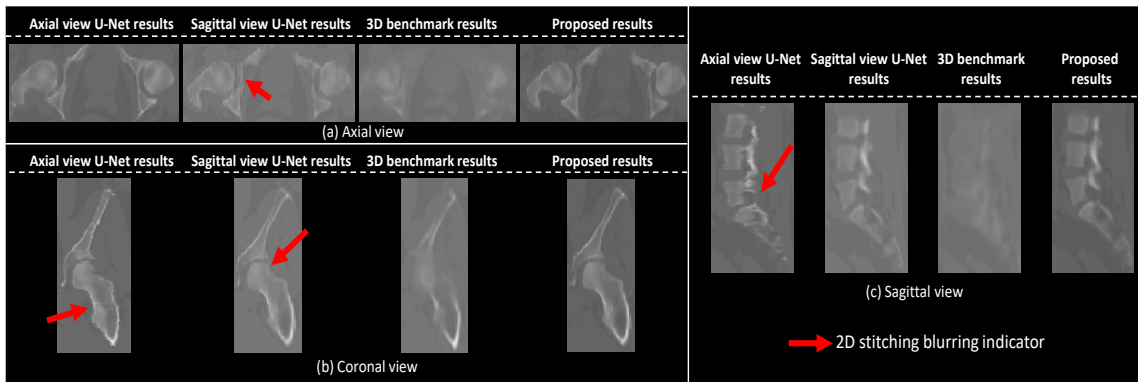


Fig. 5.7. Removed 2D slice stitching blurring effects (red arrow) by combining multi-view U-Net and improved image quality compared to results from a shrunk size 3D benchmark network.

Synthetic CT quality improvement

We can clearly see the small misalignment deficits from the multi-atlases-based sCTs (Liao et al., 2016) used for our training by comparing input MRs in Fig. 5.8. However, our proposed method will compensate these slight misalignments by predicting both the bone edge and soft tissue actual locations, which outperformed the state-of-art multi-atlas-based algorithm.

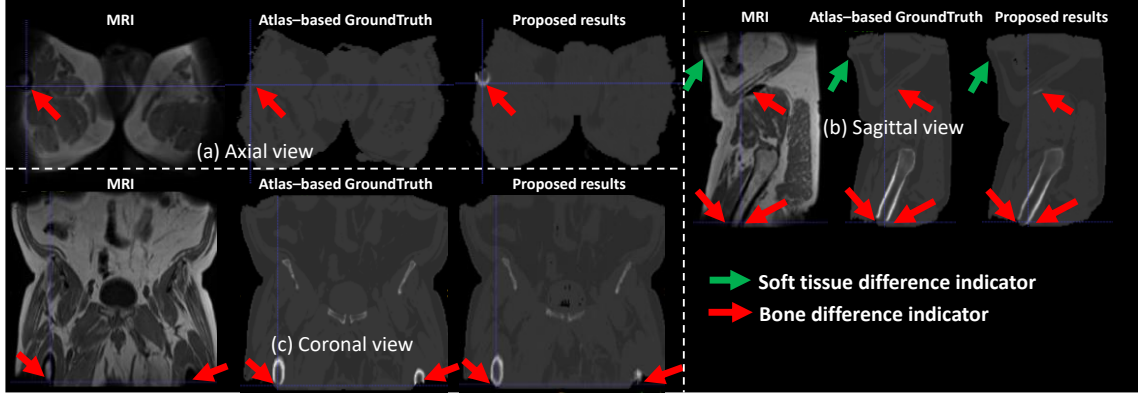


Fig. 5.8. Improved synthetic quality compared to the ground-truth CTs in 3 different views. Each column is a comparison among the input MR, ground-truth image and proposed predictions.

Synthetic CT evaluation

To quantitatively measure the reliability and accuracy of the synthetic CT outputted by our framework and the ground-truth CT images, Mean Absolute Error (MAE) (also used in (Andreasen et al., 2016; Edmund and Nyholm, 2017)) was utilized:

$$MAE = \frac{1}{N} \sum_{x,y,z \in V_1, V_2} |V_1(x, y, z) - V_2(x, y, z)| \quad 5.3$$

where V_1, V_2 represent the synthetic CT and ground-truth volumes, and N represents total number of the voxels.

Table 5.1. MAE values comparison for 7 purely testing subjects

MAE[HU]	Sub 1	Sub 2	Sub 3	Sub 4	Sub 5	Sub 6	Sub 7	Mean
Proposed	<u>14.8</u>	<u>9.0</u>	<u>21.9</u>	<u>16.1</u>	<u>16.4</u>	<u>13.2</u>	27.1	<u>16.9</u>
Sagittal benchmark	51.4	32.1	45.4	43.6	42.4	49.9	53.9	45.5
Axial benchmark	26.0	16.2	28.0	23.6	25.2	25.4	29.7	24.9
Multi-view	18.3	11.7	20.0	17.0	21.8	18.0	<u>21.1</u>	18.3

As for the 7 pure testing cases, the MAE values are very low (average 16.9 HU) (Table 5.1). Compared to the state-of-art results (average 58 HU in (Andreasen et al., 2016))

and around 40 HU in most of works (Edmund and Nyholm, 2017)), our method achieves higher accuracy. Compared with 3 benchmark U-Net-based methods, the proposed scheme achieved the best performance (Table 5.1), demonstrating the effectiveness of our specific design. Besides, the stunning CT synthesis speed (less than 8s) significantly outperformed the state-of-art multi-atlas-based framework used to generate ground-truth synthetic CTs (more than 10min), which paves the way for applying the proposed framework to real clinical settings.

Discussion

In this work, we explored a deep learning framework for CT synthesis from MR. An average MAE of 19.6 HU and ~10s synthesis speed outperform state-of-the-art methods. It shows the potency of the proposed deep learning framework in cross modality synthesis. Compared to other methods, the proposed method also showed significant improvement in sCT quality. In order to evaluate if our method is sufficient for MR-only RTP, it is important to evaluate the dose calculated from sCTs, which is part of our future work. This work gives us a new insight into tackling imperfect training pairs and 3D network training memory efficiency problem and the superior results also gives the promise to our framework for other applications.

CHAPTER 6

STRUCTURAL ANALYSES: LANDMARK DETECTION USING REINFORCEMENT LEARNING FRAMEWORK

6.1 Introduction of Reinforcement Learning (RL)

Reinforcement Learning Concepts

Reinforcement learning (RL) is an area of machine learning inspired by behaviorist psychology, concerned with how software agents ought to take actions in an environment so as to maximize some notion of cumulative reward.

Reinforcement learning can be understood using the concepts of agents, environments, states, actions and rewards. The relationship of these concepts can be illustrated using the following diagram Fig. 6.1. Brief introduction of these concepts are as following.

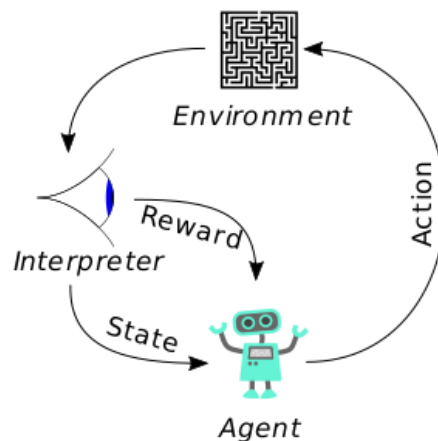


Fig. 6.1. The typical framing of a Reinforcement Learning (RL) scenario: an agent takes actions in an environment, which is interpreted into a reward and a representation of the state, which are fed back into the agent. (Figure source: Wikipedia).

Agent: An agent takes actions; for example, a drone making a delivery, or Super Mario navigating a video game. The algorithm is the agent. In life, the agent is you

Action(A): is the set of all possible moves the agent can make. An action is almost self-explanatory, but it should be noted that agents choose among a list of possible actions. When handling aerial drones, alternatives would include many different velocities and accelerations in 3D space.

Discount factor: The discount factor is multiplied with future rewards as discovered by the agent in order to dampen their effect on the agent's choice of action. It makes future rewards worth less than immediate rewards; i.e. it enforces a kind of short-term hedonism on the agent. Often expressed with the lower-case Greek letter gamma: γ . If γ is .8, and there's a reward of 10 points after 3 time steps, the present value of that reward is $0.8^3 \times 10$. A discount factor of 1 would make future rewards worth just as much as immediate rewards.

Environment: The world through which the agent moves. The environment takes the agent's current state and action as input, and returns as output the agent's reward and next state. If you are the agent, the environment could be the laws of physics and the rules of society that process your actions and determine the consequences of them.

State (S): A **state** is a concrete and immediate situation in which the agent finds itself; i.e. a specific place and moment, an instantaneous configuration that puts the agent in relation to other significant things such as tools, obstacles, enemies or prizes.

Reward (R): A **reward** is the feedback by which we measure the success or failure of an agent's actions. From any given state, an agent sends output in the form of actions to the environment, and the environment returns the agent's new state (which resulted from acting on the previous state) as well as rewards, if there are any. Rewards can be immediate or delayed. They effectively evaluate the agent's action.

Policy (π): The **policy** is the strategy that the agent employs to determine the next action based on the current state. It maps states to actions, the actions that promise the highest reward.

Value (V): The expected long-term return with discount, as opposed to the short-term reward R . $V_{\pi}(s)$ is defined as the expected long-term return of the current state under policy π . We discount rewards, or lower their estimated value, the further into the future they occur. See discount factor.

Q-value (Q): **Q-value** is similar to Value, except that it takes an extra parameter, the current action a . $Q_{\pi}(s, a)$ refers to the long-term return of the current state s , taking action a under policy π . Q maps state-action pairs to rewards.

Trajectory: A sequence of states and actions that influence those states.

Reinforcement Learning Updating Framework

There are two different schemes for updating the agent neural networks in reinforcement learning frameworks: 1. Deep Q learning scheme using Bellman Equation

(Bellman, 1952); 2. REINFORCE scheme (Williams, 1992).

For Bellman Equation scheme, the objective is to minimize the objective function

6.1:

$$\operatorname{argmin} E_{s,a} \left(R(s, a) + \gamma \max_{a'} Q(s', a') \right) - Q(s, a) \quad 6.1$$

where $R(s, a)$ is the instant reward at state s , taking action a , s', a' are the next state and action.

For the REINFORCE scheme, the objective is to get the gradient of the agent neural network using the following rule 6.2:

$$\theta = \theta + E \left(\sum_t G_t \nabla_{\theta} \log(\pi_{\theta}(s_t, a_t)) \right) \quad 6.2$$

where the G_t is the accumulative total rewards, π_{θ} is the policy network with the parameter of θ , s_t is the state at time point t , a_t is the action at time point t .

6.2 MR Landmark Detection using RL

Abstract

A new landmark detection scheme is proposed, which is based on the newly proposed deep reinforcement learning. Specifically, this method combined the Reinforce (Williams 92) algorithm with the popular deep learning networks, namely, deep reinforcement learning. Instead using traditional machine learning or popular end-to-end deep learning networks, the proposed deep reinforcement learning workflow can accomplish more accurate (or at least comparable) results at a much faster speed. For a single 2mm knee MR image, the proposed deep reinforcement learning model can achieve accurate landmark detection at ~0.1 second.

Methods

MR image landmark detection plays an important role in many medical image analysis tasks, for example, image auto-alignment. Precise and doctor-satisfactory auto-alignment rely exclusively on accurate landmark detection. Recently, the landmark detections for CT images have already been applied using reinforcement learning using the Bellman equation framework. However, we also used reinforcement learning methods for MR images utilizing REINFORCE framework. Runtime efficiency has also been achieved to deliver an instant result for clinical purpose.

However, it is still challenging to precisely and instantly detect landmarks for a high resolution MR image. Currently, traditional machine learning methods and end-to-end deep learning based networks have been proposed to address this challenge. However, the traditional methods still suffer from lack of accuracy, robustness and runtime efficiency, while deep learning based methods suffer from insufficient amount of GPU memory size and long training time. All these abovementioned challenges can be promisingly addressed by using deep reinforcement learning. The whole workflow is illustrated in Fig. 6.2.

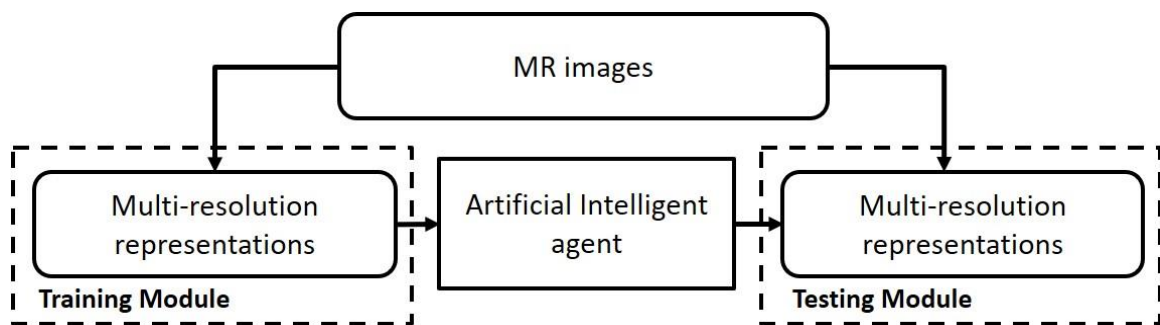


Fig. 6.2. Reinforcement learning workflow.

In this chapter, the proposed deep reinforcement learning workflow used a multi-resolution scheme. 4mm and 2mm resolutions were used. The starting point for both training and testing were predefined using the relative mean location of all the ground truth

landmark coordinates from the training set. The agent is a deep learning network for providing actions (next coordinate to go) from the current coordinate, which is shown in Fig. 6.3. The agent only takes a small image patch (25*25*25) centered around the current location coordinate as input instead of the whole image, which may significantly reduce the memory usage and the runtime efficiency. The agent will finally output a signal coordinate of the desired landmark location (Fig. 6.4).

The work flow is following a training-testing fashion, but only the testing phase is needed during runtime:

Training phase:

- Step 1: Low resolution (4mm) image training

This training procedure includes 3 information: 1. 4mm knee MR image; 2. Starting point coordinate for search in that image; 3. Ground truth landmark coordinate. The total training process took 150 epochs.

- Step 2: High resolution (2mm) image training

This training procedure also includes 3 information: 1. 2mm knee MR image; 2. Starting point coordinate for search in that image, which is actually the output from the low resolution output; 3. Ground truth landmark coordinate. The total training process took 150 epochs.

Results

This stage applies the well-trained agent to the 8 testing subjects.

Testing phase:

This step loads weights trained from step 1 and step 2, and then 4mm and 2mm knee MR images as inputs respectively, the output landmark coordinate from 4mm images

are used as starting point for 2mm images and then output the final landmark coordinates for 2mm images. The whole process took ~0.1 second. And the final detected landmarks are all within 2 mm distance from the ground truth labels, which we assume we achieved 100% accuracy for the 8 pure testing subjects.

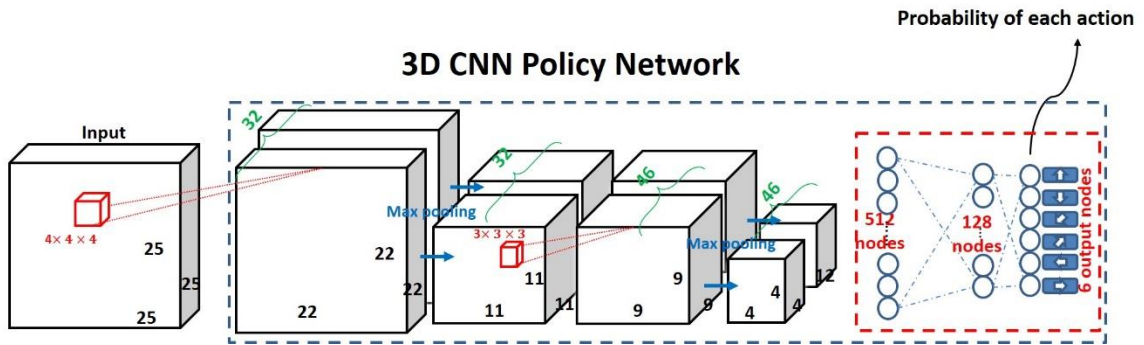


Fig. 6.3. Agent: deep learning network.

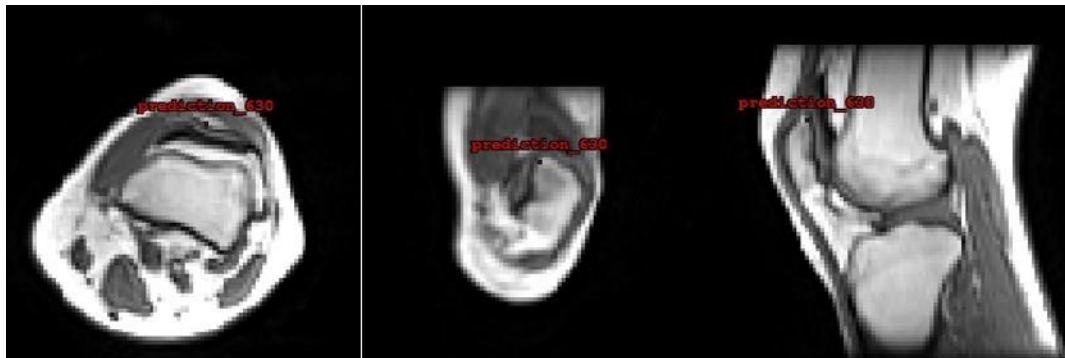


Fig. 6.4. Exemplar output coordinate of patella.

Discussion

This deep reinforcement learning based MR image landmark detection can be superior to the current landmark detection workflow with regarding to the detection accuracy and runtime. Besides this technique is not only limited to any specific image modality and anatomy, but potentially for all kinds of image modality and landmarks. The detected landmarks are essential to the MR auto-align projects. From the promising results,

we expect to see the more contributions of reinforcement learning to medical image analysis field.

CHAPTER 7

DISCUSSION AND CONCLUSION

To conclude this dissertation, the research works during my doctoral study are summarized. This dissertation contributes to medical image analysis in both functional and structural analysis fields. The frameworks proposed for the analyses consist of both shallow models and deep models. Inspired by the difficulties and challenges encountered using shallow models, advanced deep models are then proposed for better analyses.

For the functional medical image analysis part, the frameworks proposed are focusing on fMRI analysis. The findings and discoveries made in these works mainly contribute to academic researches. Starting from independent spatial analysis and temporal analysis of fMRI data using shallow traditional model, comprehensive simultaneous spatio-temporal analysis using deep models are then developed, covering both healthy brains and diseased brain, individual and group-wise analyses.

For the structural medical image analysis part, the frameworks proposed are focusing on MR analysis. The contributions are mainly for industrial and clinical purposes. By developing state-of-art deep learning frameworks and also utilizing deep reinforcement learning frameworks which is gaining increasing popularities, effective and efficient medical solutions are proposed.

Inspired by the promising results combing deep learning frameworks with medical image analysis for both functional and structural, academical research and healthcare industry, the applications of computer vision techniques into medical imaging fields can

be further investigated. The researches introduced in this dissertation are still in preliminary phases. In the future, for fMRI researches, spatio-temporal frameworks shall be extended for larger scale of data as well as for applications in analysis in brain functional disorder data. As spatio-temporal framework can intrinsically capture the nature of the 4D fMRI data, this dissertation can serve as promising milestone for this fMRI research direction. For deep learning applications in healthcare, the success is already tremendously witnessed by recent medical imaging products, which is also evident from the structural analysis part of the dissertation. To conclude at the end of this dissertation, deep learning frameworks will bring medical image analysis into a new next level.

REFERENCES

- Andreasen, D., Edmund, J.M., Zografos, V., Menze, B.H., Van Leemput, K., 2016. Computed tomography synthesis from magnetic resonance images in the pelvis using multiple random forests and auto-context features, in: Styner, M.A., Angelini, E.D. (Eds.), SPIE Medical Imaging. International Society for Optics and Photonics, p. 978417. doi:10.1117/12.2216924
- Asperger, H., 1944. Die autistischen psychopathen in kindersalter. *Eur. Arch. Psychiatry Clin.* 1, 76–136.
- Assaf, M., Jagannathan, K., Calhoun, V.D., Miller, L., Stevens, M.C., Sahl, R., O’Boyle, J.G., Schultz, R.T., Pearlson, G.D., 2010. Abnormal functional connectivity of default mode sub-networks in autism spectrum disorder patients. *Neuroimage* 53, 247–56. doi:10.1016/j.neuroimage.2010.05.067
- Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., Nolan, D., Bryant, E., Hartley, T., Footer, O., Bjork, J.M., Poldrack, R., Smith, S., Johansen-Berg, H., Snyder, A.Z., Van Essen, D.C., WU-Minn HCP Consortium, 2013. Function in the human connectome: task-fMRI and individual differences in behavior. *Neuroimage* 80, 169–89. doi:10.1016/j.neuroimage.2013.05.033
- Beckmann, C.F., DeLuca, M., Devlin, J.T., Smith, S.M., 2005. Investigations into resting-state connectivity using independent component analysis. *Philos. Trans. R.*

- Soc. Lond. B. Biol. Sci. 360, 1001–13. doi:10.1098/rstb.2005.1634
- Beckmann, C.F., Smith, S.M., 2004. Probabilistic independent component analysis for functional magnetic resonance imaging. *IEEE Trans. Med. Imaging* 23, 137–52. doi:10.1109/TMI.2003.822821
- Bell, R.M., Koren, Y., 2007. Lessons from the Netflix prize challenge. *ACM SIGKDD Explor. Newsl.* 9, 75. doi:10.1145/1345448.1345465
- Bellman, R., 1952. On the Theory of Dynamic Programming. *Proc. Natl. Acad. Sci. U. S. A.* 38, 716–9.
- Bengio, Y., Courville, A., Vincent, P., 2013. Representation Learning: A Review and New Perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* 35, 1798–1828. doi:10.1109/TPAMI.2013.50
- Bullmore, E., Sporns, O., 2009. Complex brain networks: graph theoretical analysis of structural and functional systems. *Nat. Rev. Neurosci.* 10, 186–98. doi:10.1038/nrn2575
- Calhoun, V.D., Adali, T., Stevens, M.C., Kiehl, K.A., Pekar, J.J., 2005. Semi-blind ICA of fMRI: A method for utilizing hypothesis-derived time courses in a spatial ICA analysis. *Neuroimage* 25, 527–538. doi:10.1016/j.neuroimage.2004.12.012
- Calhoun, V.D., Maciejewski, P.K., Pearlson, G.D., Kiehl, K.A., 2008. Temporal lobe and “default” hemodynamic brain modes discriminate between schizophrenia and bipolar disorder. *Hum. Brain Mapp.* 29, 1265–75. doi:10.1002/hbm.20463
- Calhoun, V.D., Pekar, J.J., Pearlson, G.D., 2004. Alcohol Intoxication Effects on Simulated Driving: Exploring Alcohol-Dose Effects on Brain Activation Using

- Functional MRI. *Neuropsychopharmacology* 29, 2097–2017.
doi:10.1038/sj.npp.1300543
- Chen, H., Zhang, T., Liu, T., 2013. Identifying group-wise consistent white matter landmarks via novel fiber shape descriptor. *MICCAI, LNCS 8149*, 66–73.
- Cherkassky, V.L., Kana, R.K., Keller, T.A., Just, M.A., 2006. Functional connectivity in a baseline resting-state network in autism. *Neuroreport* 17, 1687–90.
doi:10.1097/01.wnr.0000239956.45448.4c
- Cole, D.M., Smith, S.M., Beckmann, C.F., 2010. Advances and pitfalls in the analysis and interpretation of resting-state FMRI data. *Front. Syst. Neurosci.* 4, 8.
doi:10.3389/fnsys.2010.00008
- De Martino, F., Gentile, F., Esposito, F., Balsi, M., Di Salle, F., Goebel, R., Formisano, E., 2007. Classification of fMRI independent components using IC-fingerprints and support vector machine classifiers. *Neuroimage* 34, 177–194.
doi:10.1016/j.neuroimage.2006.08.041
- De Valois, R.L., William Yund, E., Hepler, N., 1982. The orientation and direction selectivity of cells in macaque visual cortex. *Vision Res.* 22, 531–544.
doi:10.1016/0042-6989(82)90112-2
- Delpon, G., Escande, A., Ruef, T., Darréon, J., Fontaine, J., Noblet, C., Supiot, S., Lacornerie, T., Pasquier, D., 2016. Comparison of Automated Atlas-Based Segmentation Software for Postoperative Prostate Cancer Radiotherapy. *Front. Oncol.* 6, 178. doi:10.3389/fonc.2016.00178
- Developmental Disabilities Monitoring Network Surveillance Year 2010 Principal Investigators, Centers for Disease Control and Prevention (CDC), 2014. Prevalence

- of autism spectrum disorder among children aged 8 years - autism and developmental disabilities monitoring network, 11 sites, United States, 2010. *MMWR. Surveill. Summ.* 63, 1–21.
- Dosenbach, N.U.F., Visscher, K.M., Palmer, E.D., Miezin, F.M., Wenger, K.K., Kang, H.C., Burgund, E.D., Grimes, A.L., Schlaggar, B.L., Petersen, S.E., 2006. A core system for the implementation of task sets. *Neuron* 50, 799–812.
doi:10.1016/j.neuron.2006.04.031
- Dowling, J.A., Lambert, J., Parker, J., Salvado, O., Fripp, J., Capp, A., Wratten, C., Denham, J.W., Greer, P.B., 2012. An Atlas-Based Electron Density Mapping Method for Magnetic Resonance Imaging (MRI)-Alone Treatment Planning and Adaptive MRI-Based Prostate Radiation Therapy. *Int. J. Radiat. Oncol.* 83, e5–e11.
doi:10.1016/j.ijrobp.2011.11.056
- Duncan, J., 2010. The multiple-demand (MD) system of the primate brain: mental programs for intelligent behaviour. *Trends Cogn. Sci.* 14, 172–9.
doi:10.1016/j.tics.2010.01.004
- Duncan, J.S., Ayache, N., 2000. Medical image analysis: progress over two decades and the challenges ahead. *IEEE Trans. Pattern Anal. Mach. Intell.* 22, 85–106.
doi:10.1109/34.824822
- Edmund, J.M., Andreasen, D., Kjer, H.M., Van Leemput, K., 2015. SP-0510: Dose planning based on MRI as the sole modality: Why, how and when? *Radiother. Oncol.* 115, S248–S249. doi:10.1016/S0167-8140(15)40506-7
- Edmund, J.M., Nyholm, T., 2017. A review of substitute CT generation for MRI-only radiation therapy. *Radiat. Oncol.* 12, 28. doi:10.1186/s13014-016-0747-y

- Fedorenko, E., Duncan, J., Kanwisher, N., 2013. Broad domain generality in focal regions of frontal and parietal cortex. *Proc. Natl. Acad. Sci. U. S. A.* 110, 16616–21. doi:10.1073/pnas.1315235110
- Ferrarini, L., Veer, I.M., Baerends, E., Van Tol, M.J., Renken, R.J., Van Der Wee, N.J.A., Veltman, D.J., Aleman, A., Zitman, F.G., Penninx, B.W.J.H., Van Buchem, M.A., Reiber, J.H.C., Rombouts, S.A.R.B., Milles, J., 2009. Hierarchical functional modularity in the resting-state human brain. *Hum. Brain Mapp.* 30, 2220–2231. doi:10.1002/hbm.20663
- Fox, M.D., Snyder, A.Z., Vincent, J.L., Corbetta, M., Van Essen, D.C., Raichle, M.E., 2005. The human brain is intrinsically organized into dynamic, anticorrelated functional networks. *Proc. Natl. Acad. Sci. U. S. A.* 102, 9673–8. doi:10.1073/pnas.0504136102
- Frackowiak, R.S.J., 2004. *Human brain function*. Elsevier Academic Press.
- Friston, K.J., 2009. Modalities, modes, and models in functional neuroimaging. *Science* 326, 399–403. doi:10.1126/science.1174521
- Friston, K.J., Holmes, A.P., Worsley, K.J., Poline, J.-P., Frith, C.D., Frackowiak, R.S.J., 1994. Statistical parametric maps in functional imaging: A general linear approach. *Hum. Brain Mapp.* 2, 189–210. doi:10.1002/hbm.460020402
- Gazzaniga, M.S., 2004. *The cognitive neurosciences*. MIT Press.
- Ge, B., Makkie, M., Wang, J., Zhao, S., Jiang, X., Li, X., Lv, J., Zhang, S., Zhang, W., Han, J., Guo, L., Liu, T., 2016. Signal sampling for efficient sparse representation of resting state fMRI data. *Brain Imaging Behav.* 10, 1206–1222. doi:10.1007/s11682-015-9487-0

- Greicius, M.D., Srivastava, G., Reiss, A.L., Menon, V., 2004. Default-mode network activity distinguishes Alzheimer's disease from healthy aging: evidence from functional MRI. *Proc. Natl. Acad. Sci. U. S. A.* 101, 4637–42. doi:10.1073/pnas.0308627101
- Gupta, S., Girshick, R., Arbeláez, P., Malik, J., 2014. Learning Rich Features from RGB-D Images for Object Detection and Segmentation.
- Han, J., Ji, X., Hu, X., Guo, L., Liu, T., 2015. Arousal Recognition Using Audio-Visual Features and fMRI-Based Brain Response. *IEEE Trans. Affect. Comput.* 6, 337–347. doi:10.1109/TAFFC.2015.2411280
- Harris, K.D., Mrsic-Flogel, T.D., 2013. Cortical connectivity and sensory coding. *Nature* 503, 51–58. doi:10.1038/nature12654
- He, K., Zhang, X., Ren, S., Sun, J., 2015a. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification.
- He, K., Zhang, X., Ren, S., Sun, J., 2015b. Deep Residual Learning for Image Recognition.
- Heeger, D.J., Ress, D., 2002. WHAT DOES fMRI TELL US ABOUT NEURONAL ACTIVITY? *Nat. Rev. Neurosci.* 3, 142–151. doi:10.1038/nrn730
- Hinton, G.E., Osindero, S., Teh, Y.-W., 2006. A Fast Learning Algorithm for Deep Belief Nets. *Neural Comput.* 18, 1527–1554. doi:10.1162/neco.2006.18.7.1527
- Hjelm, R.D., Calhoun, V.D., Salakhutdinov, R., Allen, E.A., Adali, T., Plis, S.M., 2014. Restricted Boltzmann machines for neuroimaging: An application in identifying intrinsic networks. *Neuroimage* 96, 245–260. doi:10.1016/j.neuroimage.2014.03.048
- Hjelm, R.D., Plis, S.M., Calhoun, V., 2016. Recurrent Neural Networks for

- Spatiotemporal Dynamics of Intrinsic Networks from fMRI Data, in: NIPS: Brains and Bits.
- http://fcon_1000.projects.nitrc.org/indi/abide/ [WWW Document], n.d.
- http://neuro.imm.dtu.dk/wiki/Harvard-Oxford_Atlas [WWW Document], n.d.
- Huang, H., Hu, X., Zhao, Y., Makkie, M., Dong, Q., Zhao, S., Guo, L., Liu, T., 2017. Modeling Task fMRI Data via Deep Convolutional Autoencoder. *IEEE Trans. Med. Imaging* 1–1. doi:10.1109/TMI.2017.2715285
- Huettel, Scott A., Allen W. Song, and G.M., 2004. Functional magnetic resonance imaging. Sunderland: Sinauer Associates.
- Ioffe, S., Szegedy, C., 2015. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift.
- Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M., 2012. FSL. *Neuroimage* 62, 782–90.
- Jenkinson, M., Smith, S., 2001. A global optimisation method for robust affine registration of brain images. *Med. Image Anal.* 5, 143–56.
- Jiang, X., Li, X., Lv, J., Zhang, T., Zhang, S., Guo, L., Liu, T., 2015. Sparse representation of HC *P* grayordinate data reveals novel functional architecture of cerebral cortex. *Hum. Brain Mapp.* 36, 5301–5319. doi:10.1002/hbm.23013
- Kana, R.K., Uddin, L.Q., Kenet, T., Chugani, D., Müller, R.-A., 2014. Brain connectivity in autism. *Front. Hum. Neurosci.* 8, 349. doi:10.3389/fnhum.2014.00349
- Kanner, L., 1943. Autistic disturbances of affective contact. *Nerv. Child* 2, 217–250.
- Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., Fei-Fei, L., 2014.

- Large-scale Video Classification with Convolutional Neural Networks. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1725–1732.
- Kelly, R.E., Alexopoulos, G.S., Wang, Z., Gunning, F.M., Murphy, C.F., Morimoto, S.S., Kanellopoulos, D., Jia, Z., Lim, K.O., Hoptman, M.J., 2010. Visual inspection of independent components: defining a procedure for artifact removal from fMRI data. *J. Neurosci. Methods* 189, 233–45. doi:10.1016/j.jneumeth.2010.03.028
- Kennedy, D.P., Courchesne, E., 2008. Functional abnormalities of the default network during self- and other-reflection in autism. *Soc. Cogn. Affect. Neurosci.* 3, 177–90. doi:10.1093/scan/nsn011
- Kennedy, D.P., Redcay, E., Courchesne, E., 2006. Failing to deactivate: resting functional abnormalities in autism. *Proc. Natl. Acad. Sci. U. S. A.* 103, 8275–80. doi:10.1073/pnas.0600674103
- Kleinhans, N.M., Richards, T., Sterling, L., Stegbauer, K.C., Mahurin, R., Johnson, L.C., Greenson, J., Dawson, G., Aylward, E., 2008. Abnormal functional connectivity in autism spectrum disorders during face processing. *Brain* 131, 1000–1012. doi:10.1093/brain/awm334
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. ImageNet Classification with Deep Convolutional Neural Networks, NIPS.
- Lawrence, S., Giles, C.L., Ah Chung Tsoi, Back, A.D., 1997. Face recognition: a convolutional neural-network approach. *IEEE Trans. Neural Networks* 8, 98–113. doi:10.1109/72.554195
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *Nature* 521, 436–444. doi:10.1038/nature14539

- Lecun, Y., Eon Bottou, L., Bengio, Y., Haaner, P., 1998. 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.
- Lee, Y.-B., Lee, J., Tak, S., Lee, K., Na, D.L., Seo, S.W., Jeong, Y., Ye, J.C., 2016. Sparse SPM: Group Sparse-dictionary learning in SPM framework for resting-state functional connectivity MRI analysis. *Neuroimage* 125, 1032–1045.
doi:10.1016/j.neuroimage.2015.10.081
- Li, K., Guo, L., Nie, J., Li, G., Liu, T., 2009. Review of methods for functional brain connectivity detection using fMRI. *Comput. Med. Imaging Graph.* 33, 131–9.
doi:10.1016/j.compmedimag.2008.10.011
- Li, X., Makkie, M., Lin, B., Sedigh Fazli, M., Davidson, I., Ye, J., Liu, T., Quinn, S., 2016. Scalable Fast Rank-1 Dictionary Learning for fMRI Big Data Analysis, in: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16*. ACM Press, New York, New York, USA, pp. 511–519. doi:10.1145/2939672.2939730
- Liao, S., Zhan, Y., Dong, Z., Yan, R., Gong, L., Zhou, X.S., Salganicoff, M., Fei, J., 2016. Automatic Lumbar Spondylolisthesis Measurement in CT Images. *IEEE Trans. Med. Imaging* 35, 1658–1669. doi:10.1109/TMI.2016.2523452
- Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghahfarooian, M., van der Laak, J.A.W.M., van Ginneken, B., Sánchez, C.I., 2017. A survey on deep learning in medical image analysis. *Med. Image Anal.* 42, 60–88.
doi:10.1016/j.media.2017.07.005

- Liu, Q., Wang, Q., 2014. Groupwise registration of brain magnetic resonance images: A review. *J. Shanghai Jiaotong Univ.* 19, 755–762. doi:10.1007/s12204-014-1576-5
- Liu, T., 2011. A few thoughts on brain ROIs. *Brain Imaging Behav.* 5, 189–202. doi:10.1007/s11682-011-9123-6
- Logothetis, N.K., 2008. What we can do and what we cannot do with fMRI. *Nature* 453, 869–78. doi:10.1038/nature06976
- Luan, H., Qi, F., Xue, Z., Chen, L., Shen, D., 2008. Multimodality image registration by maximization of quantitative–qualitative measure of mutual information. *Pattern Recognit.* 41, 285–298. doi:10.1016/j.patcog.2007.04.002
- Luxburg, U., 2007. A tutorial on spectral clustering. *Stat. Comput.* 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., 2015a. Sparse representation of whole-brain fMRI signals for identification of functional networks. *Med. Image Anal.* 20, 112–34. doi:10.1016/j.media.2014.10.011
- 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., 2013a. Identifying functional networks via sparse coding of whole brain FMRI signals, in: 2013 6th International IEEE/EMBS Conference on Neural Engineering (NER). IEEE, pp. 778–781. doi:10.1109/NER.2013.6696050
- 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., 2015b. Holistic atlases of functional networks and interactions reveal reciprocal organizational architecture of cortical function. *IEEE*

- Trans. Biomed. Eng. 62, 1120–31. doi:10.1109/TBME.2014.2369495
- 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., 2015c. Assessing effects of prenatal alcohol exposure using group-wise sparse representation of fMRI data. *Psychiatry Res.* 233, 254–68. doi:10.1016/j.psychres.2015.07.012
- Lv, J., Li, X., Zhu, D., Jiang, X., Zhang, X., Hu, X., Zhang, T., Guo, L., Liu, T., 2013b. Sparse representation of group-wise FMRI signals. *Med. Image Comput. Comput. Assist. Interv.* 16, 608–16.
- Lv, J., Lin, B., Zhang, W., Jiang, X., Hu, X., Han, J., Guo, L., Ye, J., Liu, T., 2015d. Modeling Task FMRI Data via Supervised Stochastic Coordinate Coding. pp. 239–246. doi:10.1007/978-3-319-24553-9_30
- Maas, A.L., Hannun, A.Y., Ng, A.Y., 2013. Rectifier Nonlinearities Improve Neural Network Acoustic Models. *ICML*.
- Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P., 1997. Multimodality image registration by maximization of mutual information. *IEEE Trans. Med. Imaging* 16, 187–198. doi:10.1109/42.563664
- Maintz, J.B.A., Viergever, M.A., 1998. A survey of medical image registration. *Med. Image Anal.* 2, 1–36. doi:10.1016/S1361-8415(01)80026-8
- Mairal, J., Bach, F., Ponce, J., Sapiro, G., 2010. Online Learning for Matrix Factorization and Sparse Coding. *J. Mach. Learn. Res.* 11, 19–60.
- Makkie, M., Zhao, S., Jiang, X., Lv, J., Zhao, Y., Ge, B., Li, X., Han, J., Liu, T., 2015. HAFNI-enabled largescale platform for neuroimaging informatics (HELPNI). *Brain Informatics* 2, 225–238. doi:10.1007/s40708-015-0024-0

- Masci, J., Meier, U., Cireşan, D., Schmidhuber, J., 2011. Stacked Convolutional Auto-Encoders for Hierarchical Feature Extraction. Springer, Berlin, Heidelberg, pp. 52–59. doi:10.1007/978-3-642-21735-7_7
- Matthews, P.M., Honey, G.D., Bullmore, E.T., 2006. Applications of fMRI in translational medicine and clinical practice. *Nat. Rev. Neurosci.* 7, 732–744. doi:10.1038/nrn1929
- Maturana, D., Scherer, S., 2015. VoxNet: A 3D Convolutional Neural Network for real-time object recognition, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, pp. 922–928. doi:10.1109/IROS.2015.7353481
- Mazurowski, M.A., Habas, P.A., Zurada, J.M., Lo, J.Y., Baker, J.A., Tourassi, G.D., 2008. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. *Neural Netw.* 21, 427–36. doi:10.1016/j.neunet.2007.12.031
- McAllister, T.W., Sparling, M.B., Flashman, L.A., Guerin, S.J., Mamourian, A.C., Saykin, A.J., 2001. Differential Working Memory Load Effects after Mild Traumatic Brain Injury. *Neuroimage* 14, 1004–1012. doi:10.1006/NIMG.2001.0899
- McKeown, M.J., Hansen, L.K., Sejnowski, T.J., 2003. Independent component analysis of functional MRI: what is signal and what is noise? *Curr. Opin. Neurobiol.* 13, 620–629. doi:10.1016/j.conb.2003.09.012
- McKeown, M.J., Jung, T.-P., Makeig, S., Brown, G., Kindermann, S.S., Lee, T.-W., Sejnowski, T.J., 1998. Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. *Proc. Natl. Acad. Sci.* 95, 803–810. doi:10.1073/pnas.95.3.803

- McKeown, M.J., Makeig, S., Brown, G.G., Jung, T.P., Kindermann, S.S., Bell, A.J., Sejnowski, T.J., 1998. Analysis of fMRI data by blind separation into independent spatial components. *Hum. Brain Mapp.* 6, 160–188. doi:10.1002/(SICI)1097-0193(1998)6:3<160::AID-HBM5>3.0.CO;2-1
- McKeown, M.J., Sejnowski, T.J., 1998. Independent component analysis of fMRI data: examining the assumptions. *Hum. Brain Mapp.* 6, 368–72. doi:10.1002/(SICI)1097-0193(1998)6:5/6<368::AID-HBM7>3.0.CO;2-E
- Meunier, D., Lambiotte, R., Fornito, A., Ersche, K.D., Bullmore, E.T., 2009. Hierarchical modularity in human brain functional networks. *Front. Hum. Neurosci.* 3, 1–12. doi:10.3389/neuro.11.037
- Moseley, R.L., Ypma, R.J.F., Holt, R.J., Floris, D., Chura, L.R., Spencer, M.D., Baron-Cohen, S., Suckling, J., Bullmore, E., Rubinov, M., 2015. Whole-brain functional hypoconnectivity as an endophenotype of autism in adolescents. *NeuroImage Clin.* 9, 140–152. doi:10.1016/j.nicl.2015.07.015
- Nian Liu, Junwei Han, Tianming Liu, X.L., 2016. Learning to Predict Eye Fixations via Multiresolution Convolutional Neural Networks. *IEEE Trans. Neural Networks Learn. Syst.* Accepted.
- Nian Liu, Han, J., Zhang, D., Shifeng Wen, Liu, T., 2015. Predicting eye fixations using convolutional neural networks, in: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, pp. 362–370. doi:10.1109/CVPR.2015.7298633
- Nie, D., Cao, X., Gao, Y., Wang, L., Shen, D., 2016. Estimating CT Image from MRI Data Using 3D Fully Convolutional Networks. Springer Verlag, pp. 170–178. doi:10.1007/978-3-319-46976-8_18

- Nie, D., Trullo, R., Lian, J., Petitjean, C., Ruan, S., Wang, Q., Shen, D., 2017. Medical Image Synthesis with Context-Aware Generative Adversarial Networks, in: MICCAI. Springer, Cham, pp. 417–425. doi:10.1007/978-3-319-66179-7_48
- Nielsen, J.A., Zielinski, B.A., Fletcher, P.T., Alexander, A.L., Lange, N., Bigler, E.D., Lainhart, J.E., Anderson, J.S., 2013. Multisite functional connectivity MRI classification of autism: ABIDE results. *Front. Hum. Neurosci.* 7, 599. doi:10.3389/fnhum.2013.00599
- Oliveira, F.P.M., Tavares, J.M.R.S., 2014. Medical image registration: a review. *Comput. Methods Biomech. Biomed. Engin.* 17, 73–93. doi:10.1080/10255842.2012.670855
- Paulson, E.S., Erickson, B., Schultz, C., Allen Li, X., 2014. Comprehensive MRI simulation methodology using a dedicated MRI scanner in radiation oncology for external beam radiation treatment planning. *Med. Phys.* 42, 28–39. doi:10.1118/1.4896096
- Perlberg, V., Bellec, P., Anton, J.-L., Péligrini-Issac, M., Doyon, J., Benali, H., 2007. CORSICA: correction of structured noise in fMRI by automatic identification of ICA components. *Magn. Reson. Imaging* 25, 35–46. doi:10.1016/j.mri.2006.09.042
- Pessoa, L., 2012. Beyond brain regions: network perspective of cognition-emotion interactions. *Behav. Brain Sci.* 35, 158–9. doi:10.1017/S0140525X11001567
- Pessoa, L., Amaral, D., Behnia, H., Kelly, J., Averbeck, B.B., Seo, M., Bressler, S.L., Menon, V., Damaraju, E., Huang, Y.-M., Barrett, L.F., Pessoa, L., Honey, C.J., Kotter, R., Breakspear, M., Sporns, O., Modha, D.S., Singh, R., Newman, M., Padmala, S., Pessoa, L., Pessoa, L., Pessoa, L., Pessoa, L., Adolphs, R., Pessoa, L., Petrovich, G.D., Canteras, N.S., Swanson, L.W., Swanson, L.W., Thompson, E.,

- Varela, F.J., 2012. Beyond brain regions: Network perspective of cognition–emotion interactions. *Behav. Brain Sci.* 35, 158–159. doi:10.1017/S0140525X11001567
- Pham, D.L., Xu, C., Prince, J.L., 2000. Current Methods in Medical Image Segmentation. *Annu. Rev. Biomed. Eng.* 2, 315–337. doi:10.1146/annurev.bioeng.2.1.315
- Philip, R.C.M., Dauvermann, M.R., Whalley, H.C., Baynham, K., Lawrie, S.M., Stanfield, A.C., 2012. A systematic review and meta-analysis of the fMRI investigation of autism spectrum disorders. *Neurosci. Biobehav. Rev.* 36, 901–42. doi:10.1016/j.neubiorev.2011.10.008
- Plitt, M., Barnes, K.A., Martin, A., 2015. Functional connectivity classification of autism identifies highly predictive brain features but falls short of biomarker standards. *NeuroImage. Clin.* 7, 359–66. doi:10.1016/j.nicl.2014.12.013
- Politis, M., Wu, K., Loane, C., Quinn, N.P., Brooks, D.J., Rehncrona, S., Bjorklund, A., Lindvall, O., Piccini, P., 2010. Serotonergic neurons mediate dyskinesia side effects in Parkinson’s patients with neural transplants. *Sci. Transl. Med.* 2, 38ra46. doi:10.1126/scitranslmed.3000976
- Reiman, E.M., Caselli, R.J., Yun, L.S., Chen, K., Bandy, D., Minoshima, S., Thibodeau, S.N., Osborne, D., 1996. Preclinical evidence of Alzheimer’s disease in persons homozygous for the epsilon 4 allele for apolipoprotein E. *N. Engl. J. Med.* 334, 752–8. doi:10.1056/NEJM199603213341202
- Ren, D., Zhao, Y., Chen, H., Dong, Q., Lv, J., Liu, T., 2017. 3-D functional brain network classification using Convolutional Neural Networks, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 1217–1221. doi:10.1109/ISBI.2017.7950736

- Rolfe, S.M., Finney, L., Tungaraza, R.F., Guan, J., Shapiro, L.G., Brinkley, J.F.,
Poliakov, A., Kleinhans, N., Alyward, E., 2009. An independent component analysis
based tool for exploring functional connections in the brain, in: Plum, J.P.W.,
Dawant, B.M. (Eds.), . International Society for Optics and Photonics, p. 725921.
doi:10.1117/12.811648
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-Net: Convolutional Networks for
Biomedical Image Segmentation, in: International Conference on Medical Image
Computing and Computer Assisted Intervention.
- Rosazza, C., Minati, L., 2011. Resting-state brain networks: literature review and clinical
applications. *Neurol. Sci.* 32, 773–85. doi:10.1007/s10072-011-0636-y
- Saad, Z.S., Glen, D.R., Chen, G., Beauchamp, M.S., Desai, R., Cox, R.W., 2008. A new
method for improving functional-to-structural MRI alignment using local Pearson
correlation. *Neuroimage* 44, 839–848. doi:10.1016/j.neuroimage.2008.09.037
- Scherer, D., Müller, A., Behnke, S., 2010. Evaluation of Pooling Operations in
Convolutional Architectures for Object Recognition, in: ICANN. Springer, Berlin,
Heidelberg, pp. 92–101. doi:10.1007/978-3-642-15825-4_10
- Schmidhuber, J.J., 2014. Deep learning in neural networks: An overview. *Neural
Networks* 61, 85–117. doi:10.1016/j.neunet.2014.09.003
- Shen, D., Davatzikos, C., 2002. HAMMER: hierarchical attribute matching mechanism
for elastic registration. *IEEE Trans. Med. Imaging* 21, 1421–39.
doi:10.1109/TMI.2002.803111
- Shen, D., Wu, G., Suk, H.-I., 2017. Deep Learning in Medical Image Analysis. *Annu.
Rev. Biomed. Eng.* 19, 221–248. doi:10.1146/annurev-bioeng-071516-044442

- Shen, Y., Mayhew, S.D., Kourtzi, Z., Tiño, P., 2014. Spatial–temporal modelling of fMRI data through spatially regularized mixture of hidden process models. *Neuroimage* 84, 657–671. doi:10.1016/j.neuroimage.2013.09.003
- Showcase, R., Cmu, @, Kana, R.K., Keller, T.A., Minshew, N.J., Just, M.A., 2007. Inhibitory control in high functioning autism: Decreased activation and underconnectivity in inhibition networks Inhibitory Control in High-Functioning Autism: Decreased Activation and Underconnectivity in Inhibition Networks. *Biol. Psychiatry* 198–206.
- Simonyan, K., Zisserman, A., 2014. Very Deep Convolutional Networks for Large-Scale Image Recognition.
- Sjölund, J., Forsberg, D., Andersson, M., Knutsson, H., 2015. Generating patient specific pseudo-CT of the head from MR using atlas-based regression. *Phys. Med. Biol.* 60, 825–839. doi:10.1088/0031-9155/60/2/825
- Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., Kelly, M., Laumann, T., Miller, K.L., Moeller, S., Petersen, S., Power, J., Salimi-Khorshidi, G., Snyder, A.Z., Vu, A.T., Woolrich, M.W., Xu, J., Yacoub, E., Uğurbil, K., Van Essen, D.C., Glasser, M.F., 2013. Resting-state fMRI in the Human Connectome Project. *Neuroimage* 80, 144–168. doi:10.1016/j.neuroimage.2013.05.039
- Smith, S.M., Fox, P.T., Miller, K.L., Glahn, D.C., Fox, P.M., Mackay, C.E., Filippini, N., Watkins, K.E., Toro, R., Laird, A.R., Beckmann, C.F., 2009. Correspondence of the brain’s functional architecture during activation and rest. *Proc. Natl. Acad. Sci. U. S. A.* 106, 13040–5. doi:10.1073/pnas.0905267106

- Sporns, O., 2013. Structure and function of complex brain networks. *Dialogues Clin. Neurosci.* 15, 247–62.
- Stigler, K.A., McDonald, B.C., Anand, A., Saykin, A.J., McDougale, C.J., 2011. Structural and functional magnetic resonance imaging of autism spectrum disorders. *Brain Res.* 1380, 146–61. doi:10.1016/j.brainres.2010.11.076
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going Deeper With Convolutions, in: *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 1–9.
- van den Heuvel, M.P., Hulshoff Pol, H.E., 2010. Exploring the brain network: a review on resting-state fMRI functional connectivity. *Eur. Neuropsychopharmacol.* 20, 519–34. doi:10.1016/j.euroneuro.2010.03.008
- Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., Ugurbil, K., WU-Minn HCP Consortium, for the W.-M.H., 2013. The WU-Minn Human Connectome Project: an overview. *Neuroimage* 80, 62–79. doi:10.1016/j.neuroimage.2013.05.041
- Wang, Z., Peterson, B.S., 2008. Partner-matching for the automated identification of reproducible ICA components from fMRI datasets: algorithm and validation. *Hum. Brain Mapp.* 29, 875–93. doi:10.1002/hbm.20434
- Ward, N.S., Brown, M.M., Thompson, A.J., Frackowiak, R.S.J., 2003. Neural correlates of motor recovery after stroke: a longitudinal fMRI study. *Brain* 126, 2476–2496. doi:10.1093/brain/awg245
- Williams, R.J., 1992. Simple statistical gradient-following algorithms for connectionist reinforcement learning. *Mach. Learn.* 8, 229–256. doi:10.1007/BF00992696

- Wolterink, J.M., Dinkla, A.M., Savenije, M.H.F., Seevinck, P.R., van den Berg, C.A.T., Išgum, I., 2017. Deep MR to CT Synthesis Using Unpaired Data, in: SASHIMI. Springer, Cham, pp. 14–23. doi:10.1007/978-3-319-68127-6_2
- Worsley, K.J., 1997. An overview and some new developments in the statistical analysis of PET and fMRI data. *Hum. Brain Mapp.* 5, 254–258. doi:10.1002/(SICI)1097-0193(1997)5:4<254::AID-HBM9>3.0.CO;2-2
- Xintao Hu, Cheng Lv, Gong Cheng, Jinglei Lv, Lei Guo, Junwei Han, Tianming Liu, 2015. Sparsity-Constrained fMRI Decoding of Visual Saliency in Naturalistic Video Streams. *IEEE Trans. Auton. Ment. Dev.* 7, 65–75. doi:10.1109/TAMD.2015.2409835
- Xu, J., Potenza, M.N., Calhoun, V.D., Zhang, R., Yip, S.W., Wall, J.T., Pearlson, G.D., Worhunsky, P.D., Garrison, K.A., Moran, J.M., 2016. 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.* 71, 83–100. doi:10.1016/j.neubiorev.2016.08.035
- Zeiler, M.D., 2012. ADADELTA: An Adaptive Learning Rate Method.
- Zeng, G.L., 2010. Basic Principles of Tomography, in: *Medical Image Reconstruction*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 1–19. doi:10.1007/978-3-642-05368-9_1
- Zhang, S., Li, X., Lv, J., Jiang, X., Guo, L., Liu, T., 2016. Characterizing and differentiating task-based and resting state fMRI signals via two-stage sparse representations. *Brain Imaging Behav.* 10, 21–32. doi:10.1007/s11682-015-9359-7
- Zhao, S., Han, J., Jiang, X., Huang, H., Liu, H., Lv, J., Guo, L., Liu, T., 2018. Decoding

- Auditory Saliency from Brain Activity Patterns during Free Listening to Naturalistic Audio Excerpts. *Neuroinformatics*. doi:10.1007/s12021-018-9358-0
- Zhao, S., Han, J., Lv, J., Jiang, X., Hu, X., Zhao, Y., Ge, B., Guo, L., Liu, T., 2015. Supervised Dictionary Learning for Inferring Concurrent Brain Networks. *IEEE Trans. Med. Imaging* 34, 2036–2045. doi:10.1109/TMI.2015.2418734
- 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., 2016. Connectome-scale group-wise consistent resting-state network analysis in autism spectrum disorder. *NeuroImage. Clin.* 12, 23–33. doi:10.1016/j.nicl.2016.06.004
- Zhao, Y., Dong, Q., Chen, H., Irajli, A., Li, Y., Makkie, M., Kou, Z., Liu, T., 2017a. Constructing fine-granularity functional brain network atlases via deep convolutional autoencoder. *Med. Image Anal.* 42, 200–211. doi:10.1016/j.media.2017.08.005
- Zhao, Y., Dong, Q., Zhang, S., Zhang, W., Chen, H., Jiang, X., Guo, L., Hu, X., Han, J., Liu, T., 2017b. Automatic Recognition of fMRI-derived Functional Networks using 3D Convolutional Neural Networks. *IEEE Trans. Biomed. Eng.* 1–1. doi:10.1109/TBME.2017.2715281
- Zhao, Y., Ge, F., Liu, T., 2018a. Automatic recognition of holistic functional brain networks using iteratively optimized convolutional neural networks (IO-CNN) with weak label initialization. *Med. Image Anal.* 47, 111–126. doi:10.1016/j.media.2018.04.002
- Zhao, Y., Ge, F., Zhang, S., Liu, T., 2018b. 3D Deep Convolutional Neural Network Revealed the Value of Brain Network Overlap in Differentiating Autism Spectrum

- Disorder from Healthy Controls, in: MICCAI. Springer, Cham, pp. 172–180.
doi:10.1007/978-3-030-00931-1_20
- Zhao, Y., Li, X., Makkie, M., Quinn, S., Lin, B., Ye, J., Liu, T., 2017c. Template-guided Functional Network Identification via Supervised Dictionary Learning, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 72–76. doi:10.1109/ISBI.2017.7950471
- Zhao, Y., Li, X., Zhang, W., Zhao, S., Makkie, M., Zhang, M., Li, Q., Liu, T., 2018c. Modeling 4D fMRI Data via Spatial-temporal Convolutional Neural Networks (ST-CNN), in: Medical Image Computing and Computer Assisted Intervention Society. pp. 181–189. doi:10.1007/978-3-030-00931-1_21
- Zhao, Y., Liao, S., Guo, Y., Zhao, L., Yan, Z., Hong, S., Hermosillo, G., Liu, T., Zhou, X.S., Zhan, Y., 2018d. Towards MR-Only Radiotherapy Treatment Planning: Synthetic CT Generation Using Multi-view Deep Convolutional Neural Networks, in: MICCAI. pp. 286–294. doi:10.1007/978-3-030-00928-1_33
- Zhao, Y., Zhang, S., Chen, H., Zhang, W., Lv, J., Jiang, X., Shen, D., Liu, T., 2017d. A novel framework for groupwise registration of fMRI images based on common functional networks, in: 2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017). IEEE, pp. 485–489. doi:10.1109/ISBI.2017.7950566
- Zhu, D., Li, K., Faraco, C.C., Deng, F., Zhang, D., Guo, L., Miller, L.S., Liu, T., 2012. Optimization of functional brain ROIs via maximization of consistency of structural connectivity profiles. *Neuroimage* 59, 1382–93.
doi:10.1016/j.neuroimage.2011.08.037
- Zhu, D., Li, K., Guo, L., Jiang, X., Zhang, T., Zhang, D., Chen, H., Deng, F., Faraco, C.,

Jin, C., Wee, C.-Y., Yuan, Y., Lv, P., Yin, Y., Hu, X.X., Duan, L., Hu, X.X., Han, J., Wang, L., Shen, D., Miller, L.S., Li, L., Liu, T., 2013. DICCCOL: dense individualized and common connectivity-based cortical landmarks. *Cereb. Cortex* 23, 786–800. doi:10.1093/cercor/bhs072

Zhu, D., Li, K., Terry, D.P., Puente, A.N., Wang, L., Shen, D., Miller, L.S., Liu, T., 2014. Connectome-scale assessments of structural and functional connectivity in MCI. *Hum. Brain Mapp.* 35, 2911–23. doi:10.1002/hbm.22373