Alzheimer's Disease Classification Based on Combination of Multi-model Convolutional Networks

Fan Li, Danni Cheng, Manhua Liu*
Department of Instrument Science and Engineering, School of EIEE, Shanghai Jiao Tong University, Shanghai, China
*mhliu@sjtu.edu.cn

***Abstract*—Alzheimer's disease (AD) is an irreversible neurodegenerative disorder with progressive impairment of memory and cognitive functions. It brings increasingly troubling in today's society. Its early diagnosis and warning are particularly important. Structural magnetic resonance images (MRI) play an important role to help understand the brain anatomical changes related to AD. Traditional methods are usually based on extraction of handcrafted features and training a classifier to distinguish AD from other groups. Motivated by the success of deep learning in image classification, this paper proposes a classification method based on combination of multi-model 3D convolutional networks to learn the various features from MR brain images. First, a deep 3D convolutional neural network (3D CNN) is built to hierarchically transform the MR image into more compact high-level features. Second, the multi-scale 3D convolutional autoencoders (3D CAEs) are constructed to extract features from MR brain images. The features learned by these models are combined with the upper fully connected layers for image classification in AD diagnosis. The proposed method can automatically learn the generic features from the imaging data for classification without segmentations of brain tissues and regions. Our method is evaluated using T1-weighted MR brain images on 428 subjects including 199 AD patients and 229 normal controls (NC) from Alzheimer's Disease Neuroimaging Initiative (ADNI) database. Experimental results show that the proposed method achieves an accuracy of 88.31% for AD classification and an AUC (area under the ROC curve) of 92.73%, demonstrating the promising classification performances.**

Keywords—Alzheimer's disease; convolutional neural network; convolutional autoencoder; feature fusion;

I. INTRODUCTION

Alzheimer's disease (AD) is a progressive brain disorder and the most common case of dementia in the late life. Its early diagnosis is not only challenging but also important for patient care and future treatment. Several popular non-invasive neuroimaging tools, such as structural MRI (sMRI), functional MRI (fMRI), and positron emission tomography (PET), have been investigated for developing such a system [1, 2]. Recently, various pattern recognition methods have been proposed for brain image analysis to identify the patterns related to AD and decode the disease states [1-4].

The raw brain images are too large and noisy to be directly used for classification. Thus, it is necessary to extract representative features for image classification. For morphological analysis of brain images, multiple anatomical regions, i.e., regions of interest (ROIs), were produced by grouping voxels through the warping of a labeled atlas and the regional measurements are computed as the features for image classification [2-4]. Recently, deep learning networks were also used to extract the latent features from measurements of ROIs with different imaging modalities for AD classification [3, 4]. Suk et al. [4] used a stacked autoencoder to separately extract features from MRI, PET, and cerebrospinal fluid (CSF) images and a multi-kernel SVM was used to combine these features to improve the classification performance. Liu et al. [5] designed a novel diagnostic framework with deep learning architecture to aid the diagnosis of AD by using a zero-masking strategy for data fusion to extract complementary information from multiple data modalities.

Although promising results of brain image analysis have been reported, there are still some limitations in the above feature extraction methods. In ROI based method, the definition of ROIs requires the accumulation of long-term experience of researchers. The segmentation of ROIs is also affected by the individual differences and subjective factors of scientific research personnel. The morphological abnormalities caused by neurological disorders do not always occur in pre-defined ROIs, which may involve multiple ROIs or part of the extracted ROI, so the performance of the method in the application is not stable. The voxel-wise features are an objective analysis technique for quantitative measurement of the density or volume of three tissue components (gray matter, white matter and cerebrospinal fluid) based on voxels. The VBM method requires spatial standardization (registration), that is, the individual images of the brain image in the space standardized to a standard three-dimensional space. The standardization process generally includes linear affine transformation and nonlinear deformation registration of two parts. To capture the rich image information, voxel-wise features were extracted after registering all brain image data to associate each voxel with a vector of scalar measurements for AD diagnosis [6,7]. The brain volume is segmented into gray matter (GM), white matter (WM), and CSF parts, and the voxel-wise tissue density maps are computed for classification [6]. Lerch et al. [8] proposed the fully automated measurements of cortical thickness to reproduce the clinical diagnosis in Alzheimer's Disease (AD) and thickness maps were analyzed using three different discriminant techniques to separate patients from controls. Hosseini-Asl et al. [9] proposed to predict AD with a deep 3D convolutional neural network (3D-CNN), which was built upon a 3D convolutional autoencoder and pre-trained to capture anatomical shape

Sponsors: NSFC No. 61375112 & 61773263 and SJTU

978-1-5386-1620-8/17/$31.00 ©2017 IEEE

variations in structural brain MRI scans. This method can learn generic features capturing AD biomarkers and adapt to different domain datasets.

This paper proposes a novel classification method based on combination of multi-model convolutional networks to learn the various features from MR brain images and classify AD and NC subjects. The multi-model convolutional networks are constructed on the whole MR brain images for hierarchical extraction of the compact high-level features. First, a deep 3D convolutional neural networks (3D CNN) is constructed to hierarchically extract more compact high-level features from the MR brain image. In addition, multi-scale 3D convolutional autoencoders (3D CAEs) are constructed to learn various features from MR brain images. Finally, the features learned by these convolutional networks are combined with the upper fully connected layers and softmax layers for image classification in AD diagnosis. The proposed method can automatically learn the generic features from the imaging data without image segmentation. Our experimental results on ADNI database demonstrate the effectiveness of the proposed method for AD diagnosis.

II. PROPOSED METHOD

In this section, we will present the proposed classification framework in detail. Our proposed method makes no assumption on a specific neuroimaging modality. The T1-weighted MR brain images are widely available, non-invasive and often used as the first biomarker in AD diagnostics. Thus, they are used to test the proposed method. The test images were obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (www.loni.ucla.edu/ADNI). In ADNI, the T1-weighted MR images acquired sagittally using volumetric 3D MPRAGE with 1.25×1.25 mm^2 in-plane spatial resolution and 1.2 mm thick sagittal slices. Most of these images were obtained with 1.5T scanners, while a few were acquired using 3T scanners. Detailed information about MR acquisition procedures is available at the ADNI website. Pre-processing of the images was performed before feature extraction. Specifically, all MR images are skull-stripped and cerebellum-removed after a correction of intensity inhomogeneity using a nonparametric nonuniform intensity normalization algorithm [10,11]. The MR images are registered into a template with Demons [13, 14, 15].

Fig. 1 shows the flow chart of the proposed classification method based on combination of multi-model convolutional networks, which consists of two main steps: feature extraction of 3D MR brain image with multi-model convolutional networks, and image classification based on combination of multiple features for AD diagnosis, as detailed below. There are two main advantages to combine multi-model convolutional networks for our task. First, the deep convolutional learning architectures can extract the features from the low-level to high-level with invariance to shift, scale and rotation from the training images. Second, different models of deep convolutional networks can help to learn the complementary features useful for the classification task, which can be combined to capture the rich information of MR brain images and thus improve the final classification.

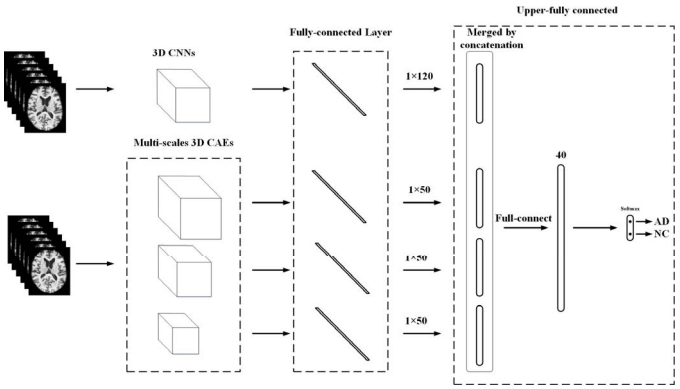

Fig. 1. The flow chart of the proposed classification method based on combination of multi-model convolutional networks.

A. Feature Learning by 3D Convolutional Neural Networks

Deep CNNs have achieved great success in feature learning for image classification. Our MR brain image is 3D data, thus it is very important to learn feature representations from all three dimensions from the volumetric medical data. In this regard, the 3D convolutional kernels are used to encode its richer spatial information of MR brain images. In this work, we construct a deep 3D CNN to extract the features from the MR brain images, which consists of 4 types of layers. The first one is the input layer which accepts a 3D MR image of fixed size (69×59×57 voxels in this work). The second type is the convolutional layer which convolves the learned filters with the input image and produce a feature map for each filter. The third type is the pooling layer which down-samples the input feature map along the spatial dimensions by replacing each non-overlapping block with their maximum. The forth type of layer is the fully connected layer which consists of a number of input and output neurons. Each neuron outputs the learned linear combination of all the inputs from the previous layer and passed through a nonlinearity.

In our implementation, each deep CNN is built with 6 convolutional layers, 3 max pooling layers, and 2 fully connected layer, as illustrated in Table I. The sizes of the first 5 convolution filters are 3×3×3, the last convolution filter is 1×1×1 and the filter numbers are set to 15, 25, 50, 50, 60 and 60 for 6 convolution layers, respectively. Max pooling is applied for each 2×2×2 region, and *Tanh* is adopted as the activation function in these layers because of its good performance for CNNs. During the pre-training period, each deep CNN is optimized individually for the classification task and its output is the class probabilistic score produced by a softmax layer.

After several convolutional and maxpooling layers, the features are flattened to a feature vector, with a softmax top-most output layer for classification. The Adadelta gradient descent is used for back-propagation. Meanwhile, in order to reduce the overfitting problem, dropout strategy is used between each maxpool layer and convolutional layer. Concerning about the size of input, the kernel's size should not be too big, so we set the kernel size to 3×3×3. More

detailed information about the architecture of 3D CNN model is shown in Table I:

TABLE I. THE ARCHITECTURE OF 3D CNN

Layer	Kernel size	Stride	Output size	Feature volumes
Input	-	-	69×59×57	1
C1	3×3×3	1	67×57×55	15
M2	2×2×2	2	33×28×27	15
C2	3×3×3	1	31×26×25	25
M2	2×2×2	2	15×13×12	25
C3	3×3×3	1	13×11×10	50
M3	2×2×2	2	6×5×5	50
C4	3×3×3	1	4×3×3	50
C5	3×3×3	1	2×1×1	60
C6	1×1×1	1	2×1×1	60
FC1	-	-	1×1×1	40
FC2	-	-	1×1×1	2

B. Feature Learning by 3D Convolutional Autoencoder

The 3D Convolutional Autoencoder (CAE) is based on reconstructing the input image for feature extraction of a 3D image. It is composed of two stages: pre-training stage and fine-tuning stage. In the pre-training stage, the data we used have no labels, the optimization of network training is to minimization of reconstruction errors with a set of kernels. In the fine-tuning stage, the data have labels to finetune kernels' weights and output a final network model.

In the first stage, three convolutional layers are used. The convolutional layer is trained one by one, the kernel size and the kernel number are set to 2×2×2 and 16, respectively. Each layer has two steps: encoding and decoding. First, the input image is encoded by mapping each fixed voxel neighborhood to a vectorial feature space in the hidden layer. It is then reconstructed in the output layer to the original space and the reconstruction error is calculated by Euclidean distance between the reconstructed image and the original input. The Adadelta gradient descent is used to minimize the reconstruction error in the training. When one layer is trained, we use the hidden layer after maxpooling as the input of next layer. The maxpooling size is 2×2×2 and the trained weights of the previous convolutional layers are constant. The same steps are iterated to generate all the three layers. *ReLU* is used as activation functions in the convolutional layer.

After three convolutional layers are trained, the output features of the third hidden layer are flattened into a vector as the input of a fully-connected layer. Three convolutional layers are initialized by the trained weights and are finetuned with the classification task by the deep-supervision. The softmax layer is appended as a top-most output layer to the upper fully connected layer, predicting the class probability belonging to the AD or NC group. The Adadelta gradient descent is used to fine-tune the entire stage. The total parameters and structure are shown in Fig 2.

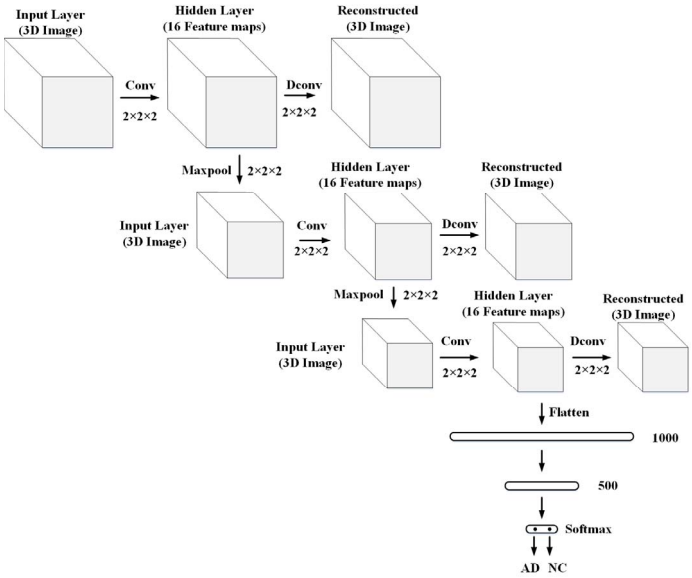

Fig. 2. Structure of the 3D convolutional autoencoder

In addition, since the 3D MR image at multiple scales can capture different image information for classification, we build three 3D CAE models by varying the image sizes. The input 3D MR image is downsampled by 2, 3 and 4, and the multi-scale 3D CAE models are trained to learn the various features from the images. The multi-scale CAE models are built with the same network structure for simplicity.

C. Ensemble Classification of AD

The 3D CNN and 3D CAE are different deep convolutional learning models which can capture the different image features. Thus, we propose to combine these multiple models to capture the rich features of 3D MR image for ensemble classification.

First, the 3D CNN and 3D CAEs are trained individually to extract features of the brain MR Image. For the 3D CNN model, the outputs of FC1 layer are used as the features for ensemble. For the multi-scales CAE models, an additional fully-connected layer is added to reduce the dimension of the output features from 500 to 50. Second, the features of these multi-models are concatenated into a new feature vector of 270, and an upper fully-connected layer is appended to further reduce the feature dimension. Finally, a softmax layer is added to take the reduced features for classification.

To facilitate training of the multi-model convolutional networks, we initially train the CNN and CAE individually to learn the specific features. Then, the trained convolutional and maxpooling layers are fixed, and the parameters of the last convolutional layer and upper fully connected layers are fine-tuned jointly to combine the features with a softmax layer for the task-specific classification. There are two advantages by finetuning the last few layers for final ensemble classification. First, by fixing the first several layers, the information learnt from the 3D CNN and 3D CAEs can be preserved to extract

the fine-level features. By fine-tuning the last few layers, the models can be more adapted to the global classification task. So the information in both imaging variations and classification task can be integrated to help improve the classification accuracy. Second, comparing with training the whole network, fine-tuning the last few layers significantly reduce the computational cost and the overfitting problem.

III. EXPERIMENTAL RESULTS AND DISCUSSIONS

In this section, we will present the experimental results and comparisons as well as the data set used in the experiments.

A. Data set

The data set used to test the proposed method is obtained from the Alzheimer's Disease Neuroimaging Initiative (ADNI) database (adni.loni.usc.edu). We use T1-weighted MR imaging data from the baseline visits of 427 participants including 199 AD and 229 normal controls (NC) for evaluation. The baseline T1-weighted MR brain images are used in our experiments. The details of the dataset are shown in Table II. The preprocessing of the T1-weighted MR images was performed as described in Section 2.

TABLE II. DEMOGRAPHIC DATA FOR 428 SUBJECTS FROM THE ADNI DATABASE (STD – STANDARD DEVIATION)

Diagnosis	AD	NC
Number of subjects	199	229
Male/Female	103/96	119/110
Age(mean±STD)	75.7±7.7	76.0±5.0

To evaluate the classification performance, we use a 5-fold cross-validation strategy to train and test the deep learning models. Each time, one fold of the data set was used for testing, while the other folds were used for training. The training set was split further into training and validation parts for finetuning the numbers of iteration. The classification accuracy (ACC) is computed to measure the proportion of correctly classified subjects among the whole population. In addition, we also compute the sensitivity (SEN), i.e., the proportion of AD patients correctly classified, and the specificity (SPE), i.e., the proportion of NCs correctly classified, for more detailed evaluations.

B. Results

First, we conduct experiments to test the performance improvement of the proposed multi-model combination method when compared to the individual convolutional models. We compared the performance of the proposed method to the individual 3D CNN models and the multi-scale 3D CAE models, which are denoted as "CNN_S3", "CAE_S2", "CAE_S3" and "CAE_S4", respectively. For the 3D CNN model denoted as "CNN_S3", the input MR image is the downsampled image of size 69×59×57 from the original MR image by 3. For the 3D CAE models denoted as "CAE_S2", "CAE_S3" and "CAE_S4", the input MR images are the downsampled images of the original one by 2, 3 and 4 and the image sizes are 102×88×85, 69×59×57 and 54×44×43, respectively. The classification accuracies of the above methods along with the proposed combination method are compared in Table III. Fig. 3 demonstrates comparison of their respective ROC curves. We can see that the proposed combination of multi-model convolutional networks performs better than the individual model methods for AD classification.

TABLE III. COMPARISON OF THE INDIVIDUAL MODEL AND PROPOSED MULTI-MODEL CLASSIFICATION METHODS

Methods	ACC%	SEN%	SPE%	AUC%
CNN_S3	84.12	84.72	83. 42	89.68
CAE_S2	82.24	86.03	77. 89	88.05
CAE_S3	81.19	83.84	77.89	85.74
CAE_S4	76.17	80.79	70.85	80.79
Multi-model	88.31	91.40	84.42	92.73

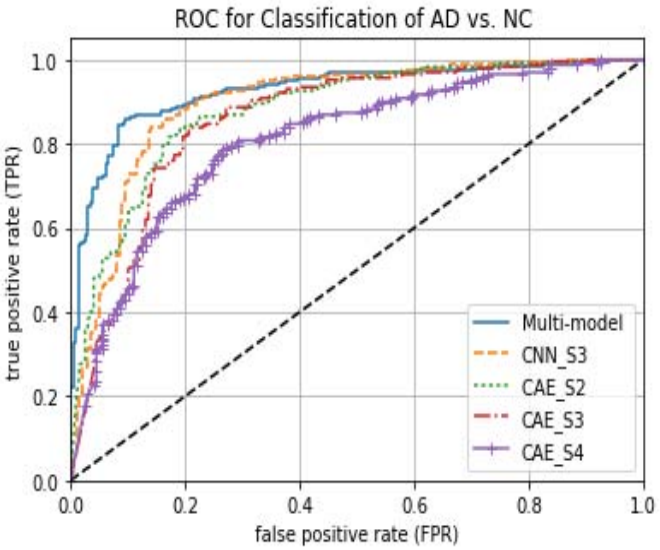

Fig. 3. Comparison of ROCs on the individual model and multi-model classifaction methods

Furthermore, we compared the results of the proposed method with some results reported in the literature that are also based on T1-weighted structural MRI data of ADNI as shown in Table IV. The volumetric features from 93 ROIs were extracted to train a SVM classifier for AD classification [2]. A stacked autoencoder was applied to learn the latent high-level features for classification [4]. A deep 3D convolutional neural network (3D-CNN) was used to learn generic features capturing AD biomarkers and classification [9]. We downloaded the source codes provided in [9] and test it on our dataset. The results are provided in Table IV for comparison. A feature extraction method based on detection of landmark points was proposed for AD classification [12]. It is worth to note that the classification accuracies in Table IV are only based on the structural MRI data. We can see that our proposed method performs better than other methods. These results further validate the efficacy of our proposed method.

All the experiments in this work were conducted in the environment of Ubuntu14.04-x64/ GPU of NVIDIA GeForce

GTX TITAN X. The algorithm is implemented using python 2.7.9 and keras to program in the framework of Theano.

TABLE IV. COMPARISON OF THE CLASSIFICATION ACCURACIES REPORTED IN THE LITERATURE

Methods	Subjects	ACC%	SEN%	SPE%
Zhang et al. 2011[2]	52NC+51AD	86.20	86.00	86.30
Suk HI et al. 2015[4]	52NC+51AD	85.70	-	-
Hosseini et al.2016[9]	229NC+199AD	82.24	86.03	77. 89
Zhang et al. 2016[12]	201NC+159AD	83.10	80.50	85.10
Proposed method	229NC+199AD	88.31	91.40	84.42

C. Discussion

From the experimental results and comparisons, we can see that the proposed ensemble classification of multi-model convolutional neural networks performs better than individual convolutional network. The proposed method has the following advantages:

- No segmentation is required in the preprocessing of MR images, which can reduce the computation cost and the error caused by segmentation.
- By training the multi-model individually and fine-tuning the last few layers for ensemble of different features learned before, the proposed method can be more adapted to the global classification task and achieve better performance.

IV. CONCLUSION

This paper has proposed a classification method based on combination of multi-model convolutional networks, which are 3D CNN and CAEs, for AD diagnosis using MR brain images. The 3D CNNs model and the multi-scales 3D CAEs are constructed to extract the various features from the MR brain images. The features learned by these models are combined with the upper fully connected layers for image classification in AD diagnosis. No segmentation is required in the image preprocessing. Experimental results and comparison demonstrate that the proposed multi-model combination method can achieve higher accuracy to classify the AD from NC on structural brain MRI scans, than the individual convolution networks such as CNN and CAE. Comparison with the existing method shows the promising classification performances for AD diagnosis.

ACKNOWLEDGMENT

This work was supported by National Natural Science Foundation of China (NSFC) under grants No. 61375112 and 61773263 and SJTU Excellent Young Faculty program

REFERENCES

[1] C.R. Jack et al., "Introduction to the recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease," Alzheimers Dement, vol. 7, no. 3, pp. 257–262, 2011.

[2] G.M. McKhann et al., "The diagnosis of dementia due to Alzheimers disease: Recommendations from the National Institute on Aging-Alzheimers Association workgroups on diagnostic guidelines for Alzheimer's disease," Alzheimers Dement, vol. 7, no. 3, pp. 263–269, 2011.

[3] Zhang, D., Wang, Y., Zhou, L., Yuan, H. and Shen, D., "Multimodal classification of Alzheimer's disease and mild cognitive impairment", Neuroimage, 55(3): 856-867 (2011).

[4] Suk H I, Lee S W, Shen D. Latent feature representation with stacked auto-encoder for AD/MCI diagnosis[J]. Brain Structure & Function, 2015, 220(2):841-59.

[5] Liu S, Cai W, Che H, Pujol S, Kikinis R, Feng D, Fulham MJ, "Multimodal neuroimaging feature learning for multiclass diagnosis of alzheimer's disease", IEEE Trans. on Biomedical Engineering, vol. 62, no. 4, pp. 1132–1140, 2015.

[6] Ishii, K., Ka wachi, T., Sasaki, H., Kono, A.K., Fukuda, T., Kojima, Y., Mori, E., "Voxel-based morphometric comparison between early- and late-onset mild Alzheimer's disease and asse ssment of diagnostic performance of z score images", American Journal of Neuroradiology, 26, pp. 333-340, 2005

[7] Li F, Tran L, Thung K H, et al. A Robust Deep Model for Improved Classification of AD/MCI Patients[J]. IEEE Journal of Biomedical & Health Informatics, 2015, 19(5):1610-1616.

[8] J.P. Lerch, J. Pruessner, AP Zijdenbos, DL Collins, et al., "Automated cortical thickness measurements from MRI can accurately separate Alzheimer's patients from normal elderly controls", Neurobiol Aging, 29(1), pp.23–30, 2008.

[9] E. Hosseiniasl, R. Keynto, A. Elbaz, "Alzheimer's Disease Diagnostics by Adaptation of 3D Convolutional Network", IEEE International Conference on Image Processing, Phoenix, Arizona, USA, Sept. 25-28, 2016

[10] Wang, Y., Nie, J., Yap, P.T., Shi, F., Guo, L., Shen, D., "Robust deformable-surface-based skull-stripping for large-scale studies", International Conference on Medical Image Computing and Computer-Assisted Intervention, pp:635-642, 2011

[11] Sled, J.G., Zijdenbos, A.P., Evans, A.C, "A nonparametric method for automatic correction of intensity nonuniformity in MRI data", IEEE Trans Med Imaging, 17, 87-97 (1998)

[12] Zhang J, Gao Y, Gao Y, et al. Detecting Anatomical Landmarks for Fast Alzheimer's Disease Diagnosis.[J]. IEEE Transactions on Medical Imaging, 2016, 35(12):2524-2533.

[13] Kroon D J, Slump C H. MRI modalitiy transformation in demon registration[C]// IEEE International Conference on Symposium on Biomedical Imaging: From Nano To Macro. IEEE Press, 2009:963-966.

[14] Siless V, Glaunès J, Guevara P, et al. Joint T1 and brain fiber log-demons registration using currents to model geometry.[C]// International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer-Verlag, 2012:57-65.

[15] H. Lu, M. Reyes, S. Weber. Multi-modal diffeomorphic demons registration based on point-wise mutual information[C]// IEEE International Conference on Biomedical Imaging: From Nano To Macro. IEEE Press, 2010:372-375.