Multi-Atlas Brain Parcellation Using Squeeze-and-Excitation Fully Convolutional Networks

Zhenyu Tang, Xianli Liu, Yang Li, Pew-Thian Yap, *Senior Member, IEEE*, and Dinggang Shen, *Fellow, IEEE*

***Abstract*— Multi-atlas parcellation (MAP) is carried out on a brain image by propagating and fusing labelled regions from brain atlases. Typical nonlinear registration-based label propagation is time-consuming and sensitive to inter-subject differences. Recently, deep learning parcellation (DLP) has been proposed to avoid nonlinear registration for better efficiency and robustness than MAP. However, most existing DLP methods neglect using brain atlases, which contain high-level information (e.g., manually labelled brain regions), to provide auxiliary features for improving the parcellation accuracy. In this paper, we propose a novel multi-atlas DLP method for brain parcellation. Our method is based on fully convolutional networks (FCN) and squeeze-and-excitation (SE) modules. It can automatically and adaptively select features from the most relevant brain atlases to guide parcellation. Moreover, our method is trained via a generative adversarial network (GAN), where a convolutional neural network (CNN) with multi-scale l_1 loss is used as the discriminator. Benefiting from brain atlases, our method outperforms MAP and state-of-the-art DLP methods on two public image datasets (LPBA40 and NIREP-NA0).**

***Index Terms*— Brain parcellation, fully convolutional networks, squeeze-and-excitation module, brain atlas selection.**

I. Introduction

MULTI-ATLAS parcellation (MAP) has been widely adopted in neuroimage analysis [1]–[9]. It uses nonlinear image registration to propagate manually labelled brain regions from multiple brain atlases to a brain image targeted for parcellation. The parcellation result is derived by fusing the propagated brain region labels [10]–[14]. However, nonlinear image registration methods [15]–[19] are typically time-consuming [20] and sensitive to image noise and inter-subject variation [21].

Manuscript received September 20, 2019; revised February 19, 2020 and March 27, 2020; accepted May 7, 2020. Date of publication May 19, 2020; date of current version July 6, 2020. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Denis Kouame. *(Corresponding author: Dinggang Shen.)*

Zhenyu Tang and Yang Li are with the Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University, Beijing 100191, China.

Xianli Liu is with the School of Computer Science and Technology, Anhui University, Hefei 230601, China.

Pew-Thian Yap is with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA.

Dinggang Shen is with the Department of Radiology and BRIC, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599 USA, and also with the Department of Brain and Cognitive Engineering, Korea University, Seoul 02841, South Korea (e-mail: dgshen@med.unc.edu).

Digital Object Identifier 10.1109/TIP.2020.2994445

Recently, many deep learning methods have been successfully applied for brain image parcellation (DLP) [22]–[25]. Since DLP methods do not rely on nonlinear image registration, they are potentially more efficient and robust than MAP. For example, in [22], a convolutional neural network (CNN) [26] has been adopted to classify each voxel in the target brain image into different brain regions. Specifically, for each voxel, three orthogonal 2D image patches (i.e., axial, coronal and sagittal image patches) and a 3D image patch centered at the voxel are cropped and used as the input data. Each image patch in the input data is proceeded by a pathway containing a series of convolutional and max pooling layers. At last, features learned from each pathway are concatenated and sent to the fully connected (FC) layer to produce the final label for the voxel. In [23], a fully convolutional network (FCN) [27] based method is adopted for brain parcellation of multimodal MR brain images including T1, T2 and FA. Particularly, each image modality is passed through a convolution and pooling pathway to learn parcellation related features. Then features learned from all image modalities are concatenated in a high layer of FCN for the final parcellation result.

Although existing DLP methods are promising, most of them do not consider using brain atlases for producing auxiliary features to guide the parcellation of the tlarget brain image. It is known that the objective of brain image parcellation is to divide brain cortical areas into different brain regions, and the big challenge is that there are often no clear anatomical boundaries between different brain regions. Brain atlases contain high-level information, such as manually labelled brain regions and boundaries between them, and using auxiliary features derived from them could effectively improve the parcellation accuracy. As a result, the performance of existing DLP methods without using auxiliary features from brain atlases is sometimes inferior to MAP.

To solve this issue, a multi-atlas guided FCN (MA-FCN) [25] has been proposed for brain parcellation, where features learned from a number of brain atlases are concatenated with those from the target brain image to predict the final parcellation result. Due to the use of high-level information from brain atlases, MA-FCN achieves accuracy comparable to MAP. However, the brain atlases are manually selected according to hand-crafted features (e.g., image intensity difference), and it is difficult to guarantee that the selected atlases are relevant

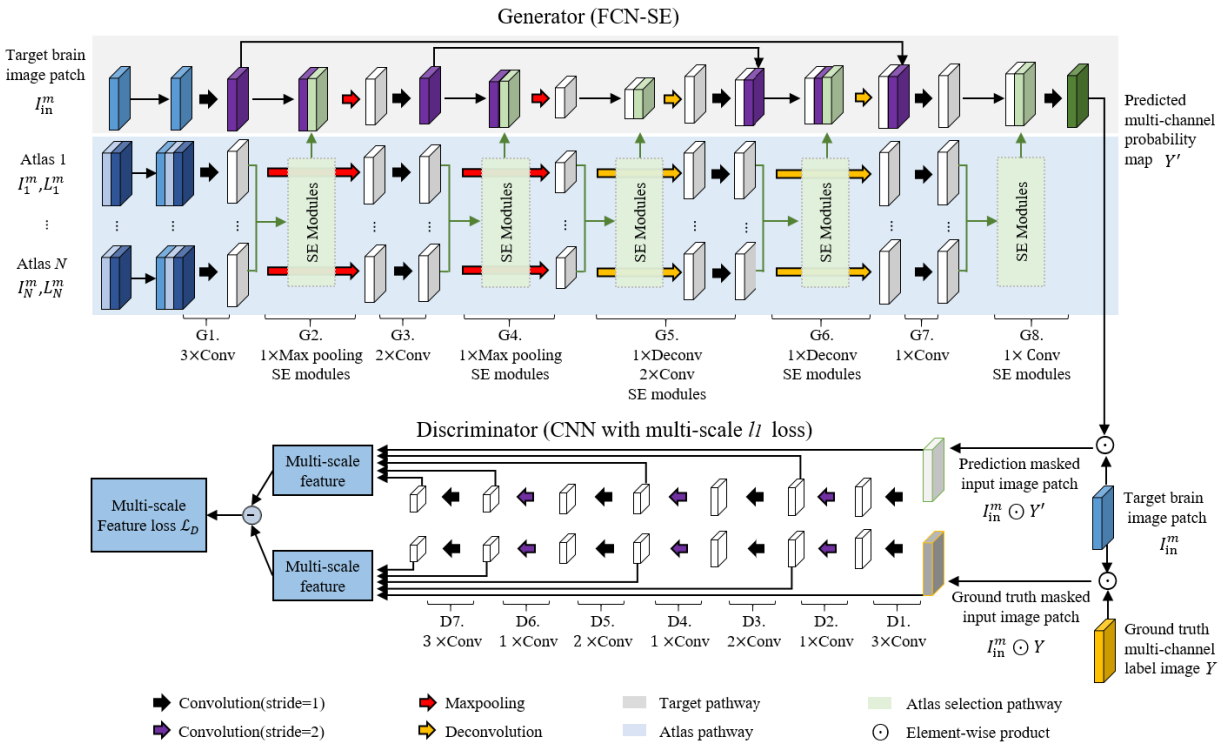


Fig. 1. Structure of FCN-SE-GAN. The Generator is FCN-SE which is composed of three different pathways, i.e., the target pathway (gray band), the atlas pathway (blue band) and the atlas selection pathway (green band). FCN-SE parcellates each target brain image patch in the target pathway with the help of brain atlas features which are learned and selected in the atlas pathway and the atlas selection pathway, respectively. The parcellation result is a multi-channel probability map, where each channel contains the probabilities of a brain region in the target brain image patch. The Discriminator is CNN with multi-scale l_1 loss, and it differentiates the parcellation result of FCN-SE from the ground truth.

to the parcellation, as the hand-crafted features are too simple to cover the complex factors related to the parcellation.

In this paper, we propose a new multi-atlas guided DLP method. Instead of manual brain atlas selection, our method can automatically and adaptively select features from the most relevant brain atlases to guide the parcellation of the target brain image. Our method is composed of a FCN and squeeze-and-excitation (SE) modules [28] (denoted as FCN-SE) and is trained in a generative adversarial network (GAN) [29], where a CNN with multi-scale l_1 loss is used as the discriminator to differentiate the parcellation result produced by FCN-SE from the ground truth. In the experiment, we evaluate our method, MAP, and state-of-the-art DLP methods using two public brain image datasets, i.e., LPBA40 [30] and NIREP-NA0 [31]. Benefiting from automatic and adaptive brain atlas selection, experimental results demonstrate that the parcellation accuracy using our method is improved compared with the other methods.

II. Methods

In the rest of this paper, we use FCN-SE-GAN to denote our method. We denote the input target brain image to be parcellated as I_{in} and the available brain atlases as $A = \{I_1, L_1, \ldots, I_N, L_N\}$, where I_n and L_n, $n = 1, \ldots, N$ are the brain image and the corresponding label image containing manually labelled brain regions, respectively. R is the total number of brain regions to be parcellated. By default, each brain atlas in A has already been aligned to I_{in} by affine transformation in the preprocessing stage.

Fig. 1 shows the structure of FCN-SE-GAN. As aforementioned, FCN-SE-GAN is in the framework of GAN, where the Generator is FCN-SE, and the Discriminator is CNN with multi-scale l_1 loss. It is worth noting that due to the limitation of memory capacity and image number for training, 3D image patches cropped from I_{in} and brain atlases in A are used as the input data of FCN-SE instead of the whole images. For each target brain image patch I_{in}^m, $m = 1, \ldots, M$, where M is the total number of image patches cropped from I_{in}, we have the corresponding N brain atlas patches I_n^m and L_n^m, $n = 1, \ldots, N$ cropped from brain atlases in A at the same location as I_{in}^m. The output of FCN-SE, i.e., the parcellation result Y' of I_{in}^m, is a multi-channel probability map, and each channel contains the probabilities of one brain region. For example, the brain image in LPBA40 is parcellated into 54 brain regions, so Y' produced by FCN-SE is a 55-channel probability map (i.e., 54 brain regions + 1 background). The Discriminator differentiates the multi-channel probability map Y' produced by FCN-SE from the ground truth Y (i.e., multi-channel label image with values of 0 or 1) by multi-scale features learned from $I_{\text{in}}^m \odot Y'$ and $I_{\text{in}}^m \odot Y$, where $\odot$ indicates element-wise product. Details of each component in FCN-SE-GAN are presented in the following subsections.

A. The Generator: FCN-SE

The input data of FCN-SE contains the target brain image patch I_{in}^m and N corresponding brain atlas patches (I_n^m, L_n^m), $n = 1, \ldots, N$. To avoid confusion in feature learning, I_{in}^m and

TABLE I

CONFIGURATION DETAILS OF THE TARGET PATHWAY, THE ATLAS PATHWAY AND THE ATLAS SELECTION PATHWAY IN THE GENERATOR (FCN-SE)

	G1.	G2.	G3.	G4.	G5.	G6.	G7.	G8.
Target pathway	Conv (1,32,3,1)* Conv (32,32,3,1) Conv (32,32,3,1)	Max pooling	Conv (33,64,3,1) Conv (64,64,3,1)	Max pooling	Deconv (66,64,4,2)* Conv (64,64,3,1) Conv (64,64,3,1)	Deconv (129,32,4,2)	Conv (64,32,3,1)	Conv (33,55,3,1)
	(1,24×24×24)	(33,24×24×24)	(33,12×12×12)	(65,12×12×12)	(66,6×6×6)	(129,12×12×12)	(64,24×24×24)	(33,24×24×24)
	(32,24×24×24)	(33,12×12×12)	(64,12×12×12)	(65,6×6×6)	(64,12×12×12)	(32,24×24×24)	(32,24×24×24)	(55,24×24×24)
Atlas pathway	Conv (3,32,3,1) Conv (32,32,3,1) Conv (32,32,3,1)	Max pooling	Conv (32,64,3,1) Conv (64,64,3,1)	Max pooling	Deconv (64,64,4,2) Conv (64,64,3,1) Conv (64,64,3,1)	Deconv (64,32,4,2)	Conv (32,32,3,1)	
	(3,24×24×24)	(32,24×24×24)	(32,12×12×12)	(64,12×12×12)	(64,6×6×6)	(64,12×12×12)	(32,24×24×24)	
	(32,24×24×24)	(32,12×12×12)	(64,12×12×12)	(64,6×6×6)	(64,12×12×12)	(32,24×24×24)	(32,24×24×24)	
Atlas selection pathway	SE modules		SE modules	SE modules	SE modules		SE modules	
	(N×32,24×24×24)		(N×64,12×12×12)	(N×64,6×6×6)	(N×64,12×12×12)		(N×32,24×24×24)	
	(1,24×24×24)		(1,12×12×12)	(1,6×6×6)	(1,12×12×12)		(1,24×24×24)	

*Conv/ Deconv (channelNum_input, channelNum_output, kernel_size, stride); : input image patch size (Channel number, W×H×L); : output image patch size (Channel number, W×H×L);

every corresponding brain atlas patch (I_n^m, L_n^m) are proceeded in their own sub-networks in FCN-SE. Specifically, FCN-SE is composed of three different pathways, i.e., the target pathway, the atlas pathway and the atlas selection pathway, which are marked by gray, blue and green bands in Fig. 1, respectively. Configuration details of the network modules in the three pathways of the Generator (denoted as G1-G8 in Fig. 1) are presented in Table I. The target brain image patch I_{in}^m (with size of $24 \times 24 \times 24$ voxels) cropped from I_{in} is proceeded in the target pathway for parcellation, and the corresponding N brain atlas patches are proceeded in the N atlas pathways for learning features from brain atlas patches. The final parcellation result of I_{in}^m is produced in the target pathway with the help of features learned from the most relevant brain atlases selected by the atlas selection pathways.

1) The Target Pathway: The target pathway is based on U-Net [32], which combines the low-level and high-level features of image patches in the learning process, benefiting both global and local contextual information. The input of the target pathway is the target brain image patch I_{in}^m cropped from I_{in}. At each level of the target pathway, features learned from I_{in}^m are concatenated with features learned by the atlas pathways from the most relevant brain atlases selected by the atlas selection pathways. As benefiting from using high-level information contained in the relevant brain atlases, the parcellation accuracy could be enhanced.

2) The Atlas Pathway: As aforementioned, for each target brain image patch I_{in}^m, we have N corresponding brain atlas patches (I_n^m, L_n^m), $n = 1, \ldots, N$ cropped from all N available brain atlases in the same location as I_{in}^m. Each brain atlas patch (I_n^m, L_n^m) is processed in one atlas pathway, which is based on FCN, for deriving features to guide the parcellation of I_{in}^m in the target pathway. Therefore, the number of atlas pathway is equal to the number of available brain atlases. Since segmentation output is not required in the atlas pathway, no skip connection is used. Considering that, the N corresponding brain atlas patches for I_{in}^m have relatively similar image intensity and anatomical structure, therefore, all atlas pathways in our method are weight-shared. By doing this, the total number of network parameters in FCN-SE is not significantly increased with the number of available brain atlases. Moreover, to make each atlas pathway learn features related to the parcellation of the target brain image patch I_{in}^m, besides the brain atlas patch, I_{in}^m is also used, which is concatenated with (I_n^m, L_n^m) as the input of each atlas pathway.

3) The Atlas Selection Pathway: Clearly, the quality of features learned from each brain atlas patch in the corresponding atlas pathway is different, and it is desired to use features from the most relevant brain atlas patches, by which the parcellation of the target brain image patch I_{in}^m can be effectively guided. To achieve this goal, in our networks, the atlas selection pathway is built, which is based on the squeeze-and-excitation (SE) module proposed for automatic and adaptive feature selection. In order to utilize brain atlas features learned across different levels, we embed the atlas selection pathway at each level of the atlas pathways. In this way, features learned from the most relevant brain atlas patches at each level are selected and fed to the corresponding level of the target pathway, leading to full-level guidance for the parcellation. In our networks shown in Fig. 1, the atlas pathways have 5 levels. Therefore, five atlas selection pathways are embedded at the corresponding levels of the N atlas pathways.

Fig. 2 shows details of the network structure in the atlas selection pathway. Assuming that, the feature learned from the i^{th} level of the n^{th} atlas pathway is $F_{i,n}$, where $i = 1, \ldots, 5$ (i.e., the number of levels in each atlas pathway shown in Fig. 1 is 5), and $n = 1, \ldots, N$. Since $F_{i,n}$ is usually in the form of multi-channel feature map, $F_{i,n,c}$, $c = 1, \ldots, C_i$ represents the feature map at the c^{th} channel in $F_{i,n}$, and C_i is the channel number of $F_{i,n}$. The atlas selection pathway assigns large weights to the features learned from the most relevant brain atlas patches in a bottom up manner. Specifically, the atlas selection pathway contains two steps. The first step is *intra-atlas feature aggregation*, where channels in $F_{i,n}$ are weighted and averaged into a single channel feature map $F'_{i,n}$, which is formulated as

$$F'_{i,n} = \sum_{c=1}^{C_i} w_{i,n,c} \cdot F_{i,n,c}, \tag{1}$$

TABLE II
CONFIGURATION DETAILS OF THE DISCRIMINATOR (CNN)

	D1.	D2.	D3.	D4.	D5.	D6.	D7.
CNN	Conv (55,32,3,1) Conv (32,32,3,1) Conv (32,32,3,1)	Conv (32,32,3,2)	Conv (32,32,3,1) Conv (32,32,3,1)	Conv (32,64,3,2)	Conv (64,64,3,1) Conv (64,64,3,1)	Conv (64,128,3,2)	Conv (128,128,3,1) Conv (128,128,3,1) Conv (128,256,3,1)
	(55,24×24×24)	(32,24×24×24)	(32,12×12×12)	(32,12×12×12)	(64,6×6×6)	(64,6×6×6)	(128,3×3×3)
	(32,24×24×24)	(32,12×12×12)	(32,12×12×12)	(64,6×6×6)	(64,6×6×6)	(128,3×3×3)	(256,3×3×3)

*Conv/ Deconv (channelNum_input, channelNum_output, kernel_size, stride); : input image patch size (Channel number, W×H×L); : output image patch size (Channel number, W×H×L);

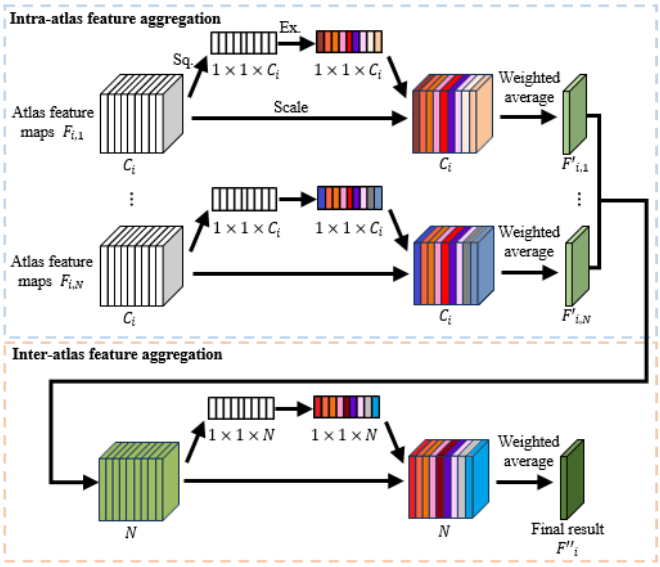


Fig. 2. Structure of the atlas selection pathway. It is composed of two parts: the intra-atlas feature aggregation which weights the features learned from an atlas, and the inter-atlas feature aggregation which weights features from different atlases. Sq. and Ex. stand for squeeze and excitation, respectively.

where $w_{i,n,c}$ is the weighting factor for the c^{th} channel of $F_{i,n}$ assigned by the SE modules in the first step. The second step of the atlas selection pathway is *inter-atlas feature aggregation*, where all the single channel feature maps $F'_{i,n}$, $n = 1, \ldots, N$ produced at the i^{th} level in the first step are weighted and averaged into a single channel feature map F''_i, which is formulated as

$$F''_i = \sum_{n=1}^{N} w'_{i,n} \cdot F'_{i,n}, \tag{2}$$

where $w'_{i,n}$ is the weighting factor for $F'_{i,n}$ assigned by the SE module in the second step. Finally, the resulting F''_i is concatenated with the features learned at the i^{th} level of the target pathway to enhance the accuracy of the final parcellation result.

The loss function of the Generator FCN-SE, which is trained in GAN, is defined as:

$$\mathcal{L}_{\mathrm{G}}(I^m_{\mathrm{in}}, Y', Y) = \mathcal{L}_{\mathrm{CE}}(Y', Y) + \mathcal{L}_{\mathrm{D}}(I^m_{\mathrm{in}}, Y', Y), \tag{3}$$

where $\mathcal{L}_{\mathrm{CE}}$ is the cross entropy between the parcellation result of FCN-SE i.e., Y' and the ground truth Y, $\mathcal{L}_{\mathrm{D}}$ is the loss given by the Discriminator which will be discussed later.

B. The Discriminator: CNN With Multi-Scale l_1 Loss

The role of the Discriminator is to differentiate the parcellation result Y' produced by the Generator from the ground truth Y. In our networks, the Discriminator is based on CNN with multi-scale l_1 loss, which is more advantageous than the 0 (real)/1(fake) output used by conventional GAN in providing stable and sufficient gradient feedback to both the Generator and the Discriminator [33], [34]. Configuration details of the network modules applied in the Generator (denoted as D1-D7 in Fig. 1) are presented in Table II. The input of the Discriminator is the target brain image patch I^m_{in} masked by Y' and Y, respectively. The reasons of using masked target brain image patches are: (1) to preserve texture information in each brain region and facilitate getting effective and rich multi-scale features in the subsequent multi-scale feature learning stage in the discriminator; and (2) to use the masks as attention maps and reduce the possible unbalanced issue (small brain region vs. large brain region) in the target brain image patches. As aforementioned, both Y' and Y are multi-channel probability maps, i.e., $Y' = \{Y'_1, \ldots, Y'_R\}$ and $Y = \{Y_1, \ldots, Y_R\}$. Therefore, the input of the Discriminator is $I^m_{\mathrm{in}} \odot Y' = \{I^m_{\mathrm{in}} \odot Y'_1, \ldots, I^m_{\mathrm{in}} \odot Y'_R\}$ and $I^m_{\mathrm{in}} \odot Y = \{I^m_{\mathrm{in}} \odot Y_1, \ldots, I^m_{\mathrm{in}} \odot Y_R\}$. It is worth noting that the values in Y' is in the range of [0,1], while, the values in Y is 0 or 1.

The Discriminator learns features from $I^m_{\mathrm{in}} \odot Y'$ and $I^m_{\mathrm{in}} \odot Y$ at different scales and makes the decision accordingly. The loss function of the Discriminator is defined as

$$\mathcal{L}_{\mathrm{D}}(I^m_{\mathrm{in}}, Y', Y) = \frac{1}{S} \sum_{s=1}^{S} ||T_s(I^m_{\mathrm{in}} \odot Y') - T_s(I^m_{\mathrm{in}} \odot Y)||_1, \tag{4}$$

where S is the total scale number in the Discriminator, and $T_s(X)$ is the feature learned from X at the s^{th} scale.

In the training stage, the Generator and the Discriminator are alternately optimized in an adversarial manner until convergence. The number of training for the Generator and the Discriminator is set to one in each iteration (i.e., the Generator and the Discriminator are updated once for each batch).

III. RESULTS

In our experiment, two public brain image datasets: LPBA40 [30] and NIREP-NA0 [31] are used to evaluate our method. The LPBA40 contains 40 MR T1-weighted brain images, and each image has a corresponding label image of 54 manually labelled brain regions. The NIREP-NA0 includes 16 MR T1-weighted brain images, each of

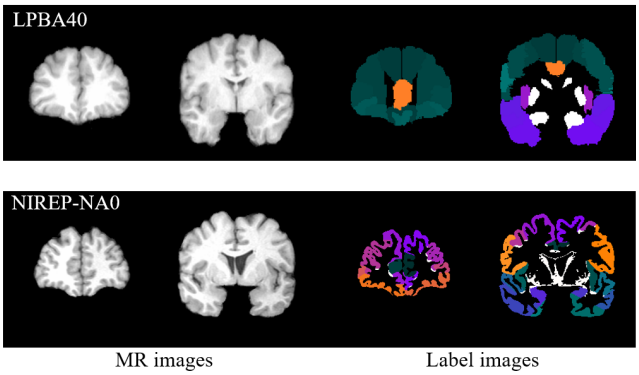


Fig. 3. Examples of MR brain images and the corresponding label images in LBPA40 (top) and NIREP-NA0 (bottom) datasets.

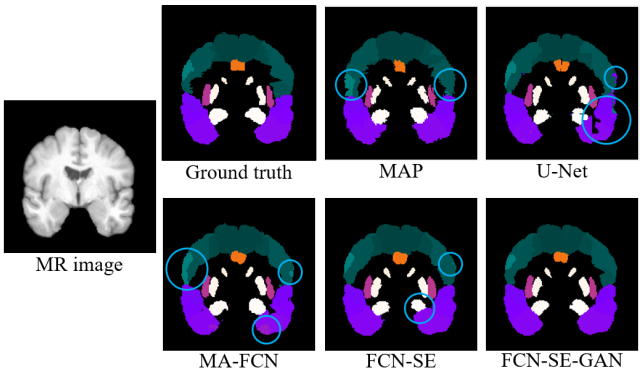


Fig. 4. Examples of a MR brain image in LPBA40 and its parcellation results produced by the methods under evaluation and the corresponding ground truth.

TABLE III
LIST OF ALL METHODS UNDER EVALUATION IN THE EXPERIMENT

Method	Strategy	Using brain atlases	Brain atlas selection	Loss function
MAP	image registration	yes	manual	-
U-Net	deep learning	no	-	cross entropy
MA-FCN	deep learning	yes	manual	cross entropy
FCN-SE	deep learning	yes	automatic	cross entropy
FCN-SE-GAN	deep learning	yes	automatic	cross entropy + adversarial loss

which has a label image of 32 manually labelled brain regions. Fig. 3 shows some examples of MR brain images in LPBA40 and NIREP-NA0. Particularly, two slices extracted from a MR brain image and its corresponding label image in LPBA40 (Fig. 3 top) and NIREP-NA0 (Fig. 3 bottom) are presented, respectively.

Besides our method, MAP and some state-of-the-art DLP methods including U-Net (no use of brain atlas), MA-FCN [25] (FCN with manually-selected brain atlas), and FCN-SE (our method without GAN) are also evaluated. The structure of the U-Net is the same as the target pathway of the Generator (FCN-SE) in FCN-SE-GAN shown in Fig. 1. Table III summarizes all the methods evaluated in our experiment.

Image patches with the size of $24 \times 24 \times 24$ voxels are used for all DLP methods under evaluation including U-Net, MA-FCN, FCN-SE and FCN-SE-GAN. Specifically, 31,780 and 12,712 image patches with overlap of 12 voxels in X, Y and Z directions are cropped from LPBA40 and NIREP-NA0 for training, respectively. For each dataset, four-fold cross validation is performed in all DLP methods under evaluation. Therefore, in LPBA40/NIREP-NA0, image patches cropped from 30/12 brain images are used for training, and the image patches cropped from the rest 10/4 brain images are used for testing.

In MAP, MA-FCN, FCN-SE and FCN-SE-GAN, which use brain atlases, except the target brain image to be parcellated (i.e., I_{in} in MAP) or from which the target brain image patches are cropped (i.e., I_{in}^m in MA-FCN, FCN-SE and FCN-SE-GAN), the rest images are used as the available brain atlases. Therefore, in LPBA40 and NIREP-NA0, 39 and 15 brain atlases are available for I_{in} or I_{in}^m, respectively. For MAP, K brain atlases are manually selected from available brain atlases according to the image intensity difference comparing with I_{in}. For MA-FCN, all possible brain atlas patches are cropped from a $48 \times 48 \times 48$ voxels region centered at the same location as I_{in}^m in each available brain atlas. Then K brain atlas patches, which have the most similar image intensity as I_{in}^m are chosen from these cropped brain atlas patches. In our experiment, K in MAP and MA-FCN is set to 3 as suggested in [25]. In FCN-SE and FCN-SE-GAN, manual brain atlas patch selection is not needed (i.e., setting K is not required), and for I_{in}^m, all brain atlas patches cropped from every available brain atlas at the same location as I_{in}^m are used. Specifically, 39 (LPBA40)/15 (NIREP-NA0) brain atlas patches are used in FCN-SE and FCN-SE-GAN. For all DLP methods to be evaluated, the batch size is set to five, and the maximal number of epoch is 30.

The parcellation accuracy is quantified by Dice index [35]. Specifically, assuming that Ω_{RS} is the parcellation result of the whole brain produced by each parcellation method under evaluation, and Ω_{GT} is the corresponding ground truth. We calculate the Dice indices of each brain region and the whole brain for each brain parcellation result Ω_{RS}. The Dice index of each brain region is calculated by:

$$\text{Dice}_{\text{BR}}(\Omega_{\text{RS}_r}, \Omega_{\text{GT}_r}) = \frac{2 \times |\Omega_{\text{RS}_r} \cap \Omega_{\text{GT}_r}|}{|\Omega_{\text{RS}_r}| + |\Omega_{\text{GT}_r}|}, \tag{5}$$

where $|\Omega_{\text{RS}_r}|$ and $|\Omega_{\text{GT}_r}|$ are the number of voxels contained in r^{th} brain region produced by a method under evaluation and the corresponding ground truth, respectively. The Dice index of the whole brain is obtained by:

$$\begin{aligned}\text{Dice}_{\text{WB}}(\Omega_{\text{RS}}, \Omega_{\text{GT}}) &= \sum_{r=1}^{R} \frac{|\Omega_{\text{GT}_r}|}{|\Omega_{\text{GT}}|} \cdot \frac{2 \times |\Omega_{\text{RS}_r} \cap \Omega_{\text{GT}_r}|}{|\Omega_{\text{RS}_r}| + |\Omega_{\text{GT}_r}|} \\ &= \sum_{r=1}^{R} \frac{|\Omega_{\text{GT}_r}|}{|\Omega_{\text{GT}}|} \cdot \text{Dice}_{\text{BR}}(\Omega_{\text{RS}_r}, \Omega_{\text{GT}_r}),\end{aligned} \tag{6}$$

where R is the number of all brain regions, and $|\Omega_{\text{GT}}|$ is the number of voxels of the whole brain.

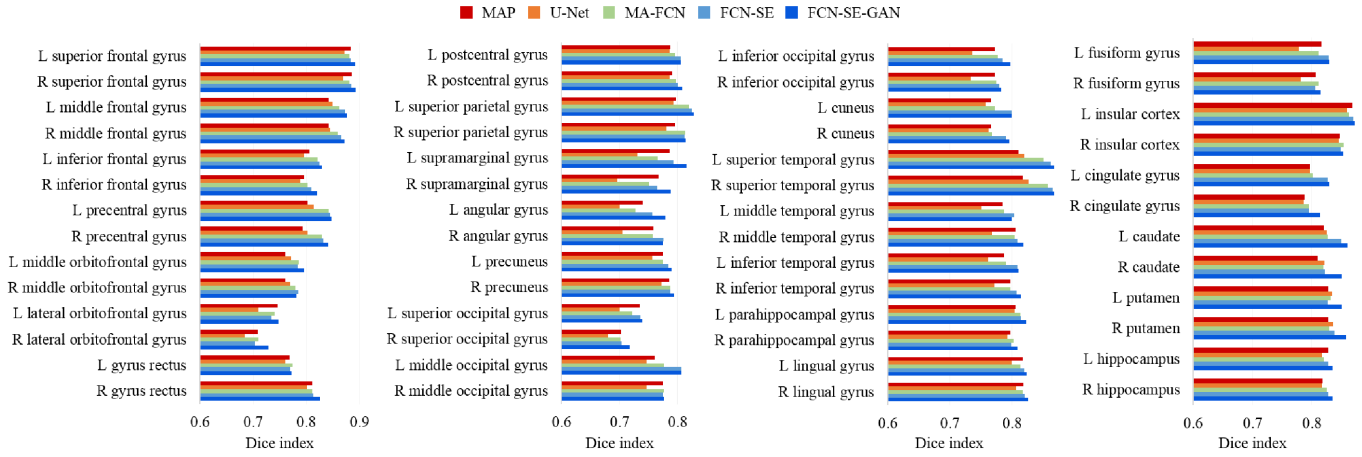


Fig. 5. LPBA40 average Dice indices of brain regions parcellated by each method under evaluation. Our method outperforms MAP and U-Net in all brain regions, MA-FCN in 52 brain regions, and FCN-SE in 50 brain regions, respectively.

A. Evaluation Result of LPBA40

As aforementioned, LPBA40 contains 40 MR brain images, each of which has a manually-segmented label image. At each time of the four-fold cross validation, 30 images are used as the training data, and the rest 10 images are used for testing. For each target brain image I_{in}, from which each image patch I_{in}^m is cropped to be parcellated, the rest 39 images with their label images are regarded as available brain atlases. In this case, the number of atlas pathway in FCN-SE-GAN is 39.

Examples of the parcellation results using each method under evaluation are shown in Fig. 4. For the sake of convenience, 2D slices extracted at the same level of the parcellation results are used in Fig. 4. It is clear that, the parcellation result of our method is more consistent with the ground truth than the rest methods under evaluation, especially in the regions marked by the blue circles.

The average Dice indices of each brain region parcellated from 40 images using each method under evaluation are shown in Fig. 5. FCN-SE-GAN outperforms MAP and U-Net in all 54 brain regions, MA-FCN in 52 brain regions, and FCN-SE in 50 brain regions, respectively. In addition, among these brain regions that have higher average Dice indices using FCN-SE-GAN, 46 (FCN-SE-GAN vs MAP) and 49 (FCN-SE-GAN vs U-Net) out of 54 brain regions, 42 out of 52 brain regions (FCN-SE-GAN vs MA-FCN), and 30 out of 50 brain regions (FCN-SE-GAN vs FCN-SE) are of statistical significance (i.e., $p < 0.05$ in Wilcoxon signed rank test [36]).

The average and standard deviation of the whole-brain Dice indices of the 40 parcellation results using each method under evaluation are 0.814 ± 0.011 (MAP), 0.795 ± 0.013 (U-Net), 0.817 ± 0.011 (MA-FCN), 0.833 ± 0.010 (FCN-SE) and 0.835 ± 0.009 (FCN-SE-GAN), respectively. The p values of the Wilcoxon singed rank test over the 40 whole-brain Dice indices of FCN-SE-GAN and each of the other methods under evaluation are 1.022×10^{-7} (FCN-SE-GAN vs MAP), 3.569×10^{-8} (FCN-SE-GAN vs U-Net), 3.852×10^{-7} (FCN-SE-GAN vs MA-FCN) and 0.022 (FCN-SE-GAN vs FCN-SE), respectively. Therefore, FCN-SE-GAN achieves the highest whole brain parcellation accuracy of all methods under evaluation with statistical significance.

The results above indicate that (1) using brain atlases is a positive factor in brain image parcellation; (2) adaptive brain atlas selection using SE module has advantage over manual selection; and (3) GAN can further improve the parcellation performance.

B. Evaluation Result of NIREP-NA0

NIREP-NA0 includes 16 MR T1-weighted brain images and corresponding label images of 32 manually labelled brain regions. Therefore, at each time of the four-fold cross validation, 12 and 4 images are used as the training and testing data, respectively. For each target brain image patch I_{in}^m cropped from I_{in}, the number of available brain atlases is 15.

Fig. 6 shows some examples of the parcellation results, i.e., slices extracted at the same level of the parcellation results produced by each method under evaluation. FCN-SE-GAN has more consistent parcellation than the rest methods as compared to the ground truth, especially in the regions marked by the blue circles.

The average Dice indices of each brain region using all methods under evaluation are shown in Fig. 7. Among the 32 brain regions, FCN-SE-GAN achieves the higher Dice indices in 29, 32, 30, 26 brain regions with comparison to MAP, U-Net, MA-FCN and FCN-SE, respectively. Among these brain regions which have higher Dice indices using FCN-SE-GAN, 20 out of 29 brain regions (FCN-SE-GAN vs MAP), 25 out of 32 brain regions (FCN-SE-GAN vs U-Net), 25 out of 30 brain regions (FCN-SE-GAN vs MA-FCN), and 18 out of 26 brain regions (FCN-SE-GAN vs FCN-SE) are of statistical significance (i.e., $p < 0.05$ in Wilcoxon signed rank test).

The average and standard deviation of the whole-brain Dice indices of 16 brain images in NIREP-NA0 are 0.782 ± 0.013 (MAP), 0.756 ± 0.019 (U-Net), 0.786 ± 0.016 (MA-FCN), 0.802 ± 0.005 (FCN-SE) and 0.803 ± 0.008 (FCN-SE-GAN), respectively. We can see that the average whole-brain Dice indices of NIREP-NA0 show the same trend as those in

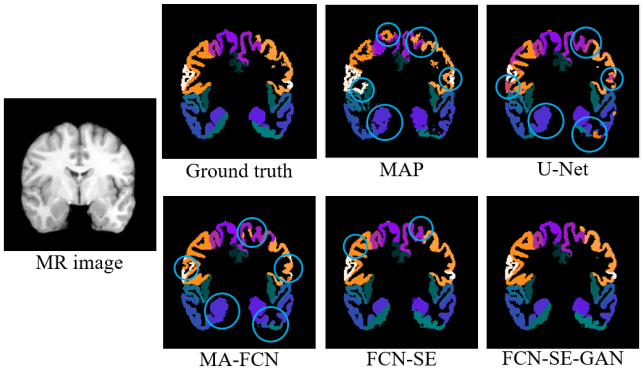


Fig. 6. Examples of a MR brain image in NIREP-NA0 and its parcellation results produced by the methods under evaluation and the corresponding ground truth.

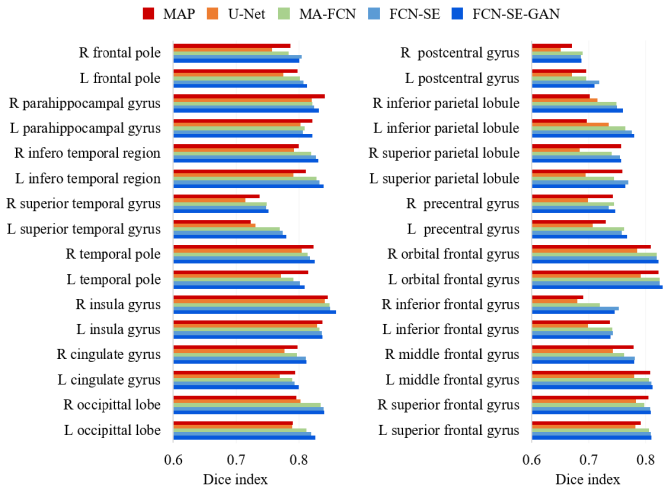


Fig. 7. NIREP-NA0 average Dice indices of brain regions parcellated by each method under evaluation. Our method outperforms MAP in 29 brain regions, U-Net in all brain regions, MA-FCN in 30 brain regions and FCN-SE in 26 brain regions, respectively.

LPBA40 (i.e., U-Net<MAP<MA-FCN<FCN-SE<FCN-SE-GAN). The p values of the Wilcoxon signed rank test over the 16 whole-brain Dice indices of FCN-SE-GAN and each of the other methods under evaluation are 4.378×10^{-4} (FCN-SE-GAN vs MAP), 4.378×10^{-4} (FCN-SE-GAN vs U-Net), 0.001 (FCN-SE-GAN vs MA-FCN) and 0.04 (FCN-SE-GAN vs FCN-SE), respectively.

Fig. 8 shows some examples of the selected top 3 brain atlas patches in MA-FCN for the target brain image patches cropped from LPBA40 (Fig. 8 top) and NIREP-NA0 (Fig. 8 bottom). Particularly, the top 3 brain atlas patches selected in MA-FCN are of the highest image intensity similarity to the corresponding target brain image patch. In addition, for each target brain image patch, three brain atlas patches, which have the highest sum of weights given at the inter-atlas feature selection stage (i.e., second step) in all 5 atlas selection pathways (i.e., $\sum_{i=1}^{5} w'_{i,n}$, n is the ID of each brain atlas candidate) in FCN-SE and FCN-SE-GAN, are also presented. Compared to the patches selected by MA-FCN, the patches selected by FCN-SE and FCN-SE-GAN have labels that are more consistent with the ground truth of the target brain patches. By using more consistent labels as the high-level

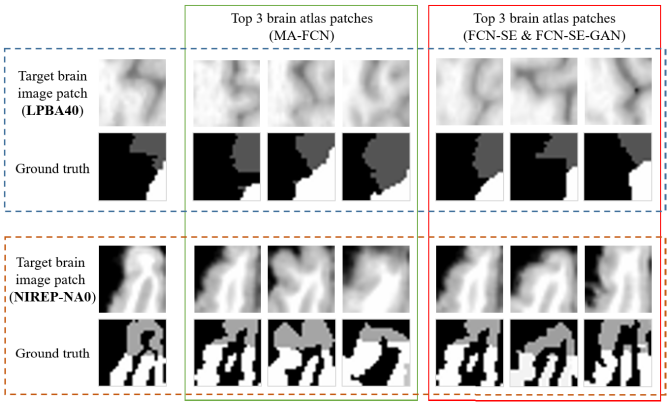


Fig. 8. Examples of the top 3 brain atlas patches selected in MA-FCN for the target brain image patches cropped from LPBA40 (top) and NIREP-NA0 (bottom). In addition, 3 corresponding brain atlas patches which are of the highest sum of weights assigned by the inter-atlas feature selection in all 5 atlas selection pathways in FCN-SE and FCN-SE-GAN are also presented.

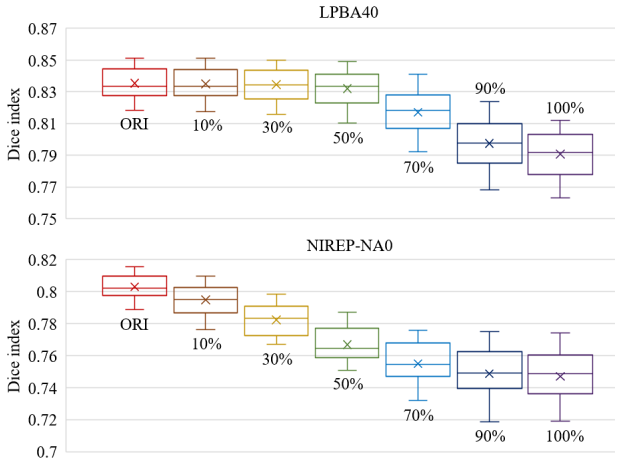


Fig. 9. Boxplot of the whole-brain Dice indices of parcellation results under different proportions of inconsistent brain atlas patches in LPBA40 (top) and NIREP-NA0 (bottom) datasets. ORI means no inconsistent brain atlas patch is included.

information, more accurate parcellation result can therefore be obtained.

C. Evaluation of the Influence of Inconsistent Brain Atlases

The core of FCN-SE-GAN is using brain atlases to provide auxiliary features, guiding the parcellation of the target brain image to a high accuracy level. Therefore, the quality of brain atlases has influence on the performance of FCN-SE-GAN. Specifically, if the majority of brain atlases have consistent labels to the ground truth of the target brain image, the parcellation will be more accurate. We show the consistence of brain atlases used in our experiment (after affine transformation) by calculating the whole-brain Dice indices of all possible brain atlas pairs. The maximum, minimum, mean and standard deviation of these whole-brain Dice indices in LPBA40 and NIREP-NA0 are [0.751, 0.580, 0.700, 0.026] and [0.601, 0.501, 0.553, 0.022], respectively.

In this part, the robustness of FCN-SE-GAN to the inconsistent brain atlases, i.e., brain atlases containing contradicting labels, is evaluated. Specifically, for each input target brain image patch, a certain proportion of the corresponding brain

atlas patches is replaced by inconsistent patches randomly cropped at the position shifting at least 24 voxels from the true position in the brain atlases. For each input target brain image patch, 10%, 30%, 50%, 70%, 90% and 100% of the corresponding brain atlas patches are replaced with these inconsistent patches. Clearly, in each of these inconsistent patches, there are no regions or only a small portion of regions that are consistent to each other. Therefore, increasing the percentage of inconsistent patches implies decreasing the consistency among brain atlases, i.e., decreasing the maximum, minimum and mean of the whole-brain Dice indices between all possible brain atlas pairs. For each percentage of inconsistent patches (i.e., 10%, 30%, 50%, 70%, 90% and 100%), the corresponding parcellation results are evaluated, and details are shown in Fig. 9. It is clear that the performance of brain parcellation (measured by whole-brain Dice index) is degraded as the number of inconsistent patches increased. We can also observe that using more atlases is more robust to the influence of inconsistent patches than that the case of using less atlases. Specifically, for LPBA40 (39 atlases), the performance is relatively stable before 50%, while, for NIREP-NA0 (14 atlases), significantly degraded performance can be observed from 10% on.

IV. Conclusion

We proposed a new deep learning based brain image parcellation method (DLP) denoted as FCN-SE-GAN. Compared with multi-atlas parcellation (MAP), which is computational expensive and sensitive to image quality, our method is more efficient and robust as no nonlinear image registration is needed. In the experiment, the average runtime for processing a target brain image was 311 seconds using MAP (Intel i9-9900, 32GB RAM, 512G SSD, Win 10), while, our method only needed 23 seconds (GTX 1080Ti, 11GB RAM) and produced more accurate parcellation results than MAP. In contrast to existing DLP methods that manually select brain atlases, our method can automatically and adaptively choose features from the most relevant brain atlases for parcellation.

Two public brain image datasets LPBA40 and NIREP-NA0 were used to evaluate our method, MAP, U-Net (DLP without brain atlases), MA-FCN (DLP with manually-selected brain atlases) and FCN-SE (our method without GAN). The experimental results demonstrated that all three factors, i.e., brain atlases, automatic brain atlas selection and GAN, have positive effects on parcellation. Benefiting from these three factors, our method outperformed the other methods in terms of Dice indices for most brain regions.

References

[1] P. Aljabar, R. A. Heckemann, A. Hammers, J. V. Hajnal, and D. Rueckert, "Multi-atlas based segmentation of brain images: Atlas selection and its effect on accuracy," *NeuroImage*, vol. 46, no. 3, pp. 726–738, Jul. 2009.

[2] J. E. Iglesias and M. R. Sabuncu, "Multi-atlas segmentation of biomedical images: A survey," *Med. Image Anal.*, vol. 24, no. 1, pp. 205–219, Aug. 2015.

[3] J. M. Lötjönen *et al.*, "Fast and robust multi-atlas segmentation of brain magnetic resonance images," *NeuroImage*, vol. 49, no. 3, pp. 2352–2365, Feb. 2010.

[4] X. Artaechevarria, A. Munoz-Barrutia, and C. Ortiz-de-Solorzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Trans. Med. Imag.*, vol. 28, no. 8, pp. 1266–1277, Aug. 2009.

[5] C. Mariano, O. Arnau, L. Xavier, F. Jordi, and C. M. Bach, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, p. 158–177, 2011.

[6] Z. Xue, D. Shen, and C. Davatzikos, "CLASSIC: Consistent longitudinal alignment and segmentation for serial image computing," *NeuroImage*, vol. 30, no. 2, pp. 388–399, Apr. 2006.

[7] Y. Fan *et al.*, "Multivariate examination of brain abnormality using both structural and functional MRI," *NeuroImage*, vol. 36, no. 4, pp. 1189–1199, Jul. 2007.

[8] Y. Fan *et al.*, "Unaffected family members and schizophrenia patients share brain structure patterns: A high-dimensional pattern classification study," *Biol. Psychiatry*, vol. 63, no. 1, pp. 118–124, Jan. 2008.

[9] G. Li *et al.*, "Mapping longitudinal hemispheric structural asymmetries of the human cerebral cortex from birth to 2 years of age," *Cerebral Cortex*, vol. 24, no. 5, pp. 1289–1300, May 2014.

[10] A. J. Asman and B. A. Landman, "Non-local statistical label fusion for multi-atlas segmentation," *Med. Image Anal.*, vol. 17, no. 2, pp. 194–208, Feb. 2013.

[11] H. Wang, J. W. Suh, S. R. Das, J. B. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 3, pp. 611–623, Mar. 2013.

[12] T. R. Langerak, U. A. van der Heide, A. N. T. J. Kotte, M. A. Viergever, M. van Vulpen, and J. P. W. Pluim, "Label fusion in atlas-based segmentation using a selective and iterative method for performance level estimation (SIMPLE)," *IEEE Trans. Med. Imag.*, vol. 29, no. 12, pp. 2000–2008, Dec. 2010.

[13] G. Wu, Q. Wang, D. Zhang, F. Nie, H. Huang, and D. Shen, "A generative probability model of joint label fusion for multi-atlas based brain segmentation," *Med. Image Anal.*, vol. 18, no. 6, pp. 881–890, Aug. 2014.

[14] S. K. Warfield, K. H. Zou, and W. M. Wells, "Simultaneous truth and performance level estimation (STAPLE): An algorithm for the validation of image segmentation," *IEEE Trans. Med. Imag.*, vol. 23, no. 7, pp. 903–921, Jul. 2004.

[15] B. Zitová and J. Flusser, "Image registration methods: A survey," *Image Vis. Comput.*, vol. 21, no. 11, pp. 977–1000, Oct. 2003.

[16] B. Avants, C. Epstein, M. Grossman, and J. Gee, "Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, Feb. 2008.

[17] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. S61–S72, Mar. 2009.

[18] J. Ashburner, "A fast diffeomorphic image registration algorithm," *NeuroImage*, vol. 38, no. 1, pp. 95–113, Oct. 2007.

[19] J. Fan, X. Cao, P.-T. Yap, and D. Shen, "BIRNet: Brain image registration using dual-supervised fully convolutional networks," 2018, *arXiv:1802.04692*. [Online]. Available: http://arxiv.org/abs/1802.04692

[20] A. Klein *et al.*, "Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration," *NeuroImage*, vol. 46, no. 3, pp. 786–802, Jul. 2009.

[21] J. Xiong *et al.*, "Intersubject variability in cortical activations during a complex language task," *NeuroImage*, vol. 12, no. 3, pp. 326–339, Sep. 2000.

[22] A. de Brebisson and G. Montana, "Deep neural networks for anatomical brain segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2015, pp. 20–28.

[23] D. Nie, L. Wang, Y. Gao, and D. Shen, "Fully convolutional networks for multi-modality isointense infant brain image segmentation," in *Proc. IEEE 13th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2016, pp. 1342–1345.

[24] L. Fang *et al.*, "Brain image labeling using multi-atlas guided 3d fully convolutional networks," in *Proc. Int. Workshop Patch-Based Techn. Med. Imag.* Cham, Switzerland: Springer, 2017, pp. 12–19.

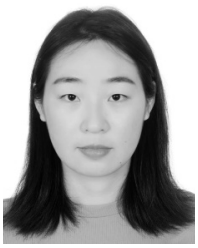
[25] L. Fang *et al.*, "Automatic brain labeling via multi-atlas guided fully convolutional networks," *Med. Image Anal.*, vol. 51, pp. 157–168, Jan. 2019.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 640–651.

[28] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Dec. 2018, pp. 7132–7141.
[29] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. Int. Conf. Neural Inf. Process. Syst.*, 2014, pp. 2672–2680.
[30] D. W. Shattuck *et al.*, "Construction of a 3D probabilistic atlas of human cortical structures," *NeuroImage*, vol. 39, no. 3, pp. 1064–1080, Feb. 2008.
[31] G. E. Christensen *et al.*, "Introduction to the non-rigid image registration evaluation project (nirep)," *IEEE Trans. Magn.*, vol. 30, no. 5, pp. 2972–2975, Oct. 2006.
[32] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2015, pp. 234–241.
[33] Y. Xue, T. Xu, H. Zhang, L. R. Long, and X. Huang, "SegAN: Adversarial network with multi-scale l1 loss for medical image segmentation," *Neuroinformatics*, vol. 16, nos. 3–4, pp. 383–392, Oct. 2018.
[34] X. Liu, H. Zhao, S. Zhang, and Z. Tang, "Brain image parcellation using multi-atlas guided adversarial fully convolutional network," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 723–726.
[35] L. R. Dice, "Measures of the amount of ecologic association between species," *Ecology*, vol. 26, no. 3, pp. 297–302, Jul. 1945.
[36] R. Woolson, *Wilcoxon Signed-Rank Test*. Hoboken, NJ, USA: Wiley, 2007, pp. 1–3.

Zhenyu Tang received the Ph.D. degree in computer engineering from the University of Duisburg-Essen in 2011. He held a postdoctoral position with the University of North Carolina at Chapel Hill and also with the Automation institute of Chinese Academy of Science. He is currently an Associate Professor with the Beijing Advanced Innovation Center for Big Data and Brain Computing, Beihang University. His research interests include medical image analysis, computer vision, and pattern recognition.



Xianli Liu received the master's degree in computer science and technology from Anhui University in 2019. Her main research direction is deep learning-based medical image segmentation.

Yang Li received the Ph.D. degree in automatic control and systems engineering from Sheffield University in 2011. He held a postdoctoral position with the University of North Carolina at Chapel Hill. He is currently a Professor with the Department of Automation Sciences and Electrical Engineering, Beihang University. His research interests include system identification and modeling for complex nonlinear processes, nonlinear and nonstationary signal processing, intelligent computation, data mining, and model optimization.

Pew-Thian Yap (Senior Member, IEEE) is currently an Associate Professor of radiology with the University of North Carolina at Chapel Hill (UNC-CH), Chapel Hill, NC, USA. He is also a Faculty Member with Biomedical Research Imaging Center (BRIC). He has published more than 220 peer-reviewed articles in major international journals and conference proceedings. His research interests include medical image analysis, machine learning, and neuroscience.

Dinggang Shen (Fellow, IEEE) is currently a Professor. He has published more than 1000 articles in the international journals and conference proceedings, with H-index 98. His research interests include medical image analysis, computer vision, and pattern recognition. He is an Fellow of AIMBE and of IAPR. He serves as an editorial board member for eight international journals. He was a General Chair of MICCAI 2019.