Anatomical Attention Guided Deep Networks for ROI Segmentation of Brain MR Images

Liang Sun, Wei Shao, Daoqiang Zhang*, Mingxia Liu*, *Senior Member, IEEE*

***Abstract*—Brain region-of-interest (ROI) segmentation based on structural magnetic resonance imaging (MRI) scans is an essential step for many computer-aid medical image analysis applications. Due to low intensity contrast around ROI boundary and large inter-subject variance, it has been remaining a challenging task to effectively segment brain ROIs from structural MR images. Even though several deep learning methods for brain MR image segmentation have been developed, most of them do not incorporate shape priors to take advantage of the regularity of brain structures, thus leading to sub-optimal performance. To address this issue, we propose an anatomical attention guided deep learning framework for brain ROI segmentation of structural MR images, containing two subnetworks. The first one is a segmentation subnetwork, used to simultaneously extract discriminative image representation and segment ROIs for each input MR image. The second one is an anatomical attention subnetwork, designed to capture the anatomical structure information of the brain from a set of labeled atlases. To utilize the anatomical attention knowledge learned from atlases, we develop an *anatomical gate* architecture to fuse feature maps derived from a set of atlas label maps and those from the to-be-segmented image for brain ROI segmentation. In this way, the anatomical prior learned from atlases can be explicitly employed to guide the segmentation process for performance improvement. Within this framework, we develop two anatomical attention guided segmentation models, denoted as anatomical gated fully convolutional network (AG-FCN) and anatomical gated U-Net (AG-UNet), respectively. Experimental results on both ADNI and LONI-LPBA40 datasets suggest that the proposed AG-FCN and AG-UNet methods achieve superior performance in ROI segmentation of brain MR images, compared with several state-of-the-art methods.**

***Index Terms*—Anatomical Attention, Deep Learning, ROI Segmentation, Brain MR Image**

I. Introduction

BRAIN region-of-interest (ROI) segmentation is an important prerequisite step for many computer-aid medical image analysis tasks [1]–[6]. For instance, in the pipeline of brain network analysis for brain disease diagnosis, brain MR images are usually segmented into multiple ROIs for constructing brain networks, and the constructed brain networks are further used for subsequent analysis and diagnosis. However, manually labeling ROIs of brain MR images is not only time-consuming but also error-prone even for experts. Hence, it is practically useful to develop an effective method to automatically segment ROIs of brain MRIs.

Recent years, deep learning methods achieve great success in medical image segmentation and computer-aided brain disease diagnosis [6]–[12]. Among these methods, several end-to-end networks have been developed for automated image segmentation, which typically include two parts, *i.e.*, 1) encoding parts, and 2) decoding parts. Specifically, the encoding path is employed to extract high-level contextual feature maps from the input image, while the decoding part up-sample these high-level feature maps to predict the dense label map of the to-be-segmented image. Since the human brain has a complicated anatomical structure, brain MR images usually have low intensity contrast around the boundary of ROIs and large variance between different subjects. However, existing deep learning methods generally ignore the anatomical structure information of the brain, thus prone to generating sub-optimal brain ROI segmentation performance.

In the last decade, multi-atlas based segmentation methods have shown their superior performance in ROI segmentation of brain MR images [13]–[16], compared with conventional single-atlas based methods. In multi-atlas based segmentation framework, a set of labeled atlases are firstly registered onto the common space of the to-be-segmented image, and then the labels of multiple atlases are propagated to determine the final label map of the to-be-segmented image. The main advantage of the multi-atlas based segmentation framework is that it can take advantage of the rich anatomical information of the human brain provided by multiple registered atlases (other than one single atlas). However, these methods usually employ handcrafted features (*e.g.*, image intensity) to represent brain MR images, and these handcrafted features may not be well coordinated with subsequent label propagation algorithms, thus negatively affecting the segmentation performance. It's desired to extract task-oriented features of brain MR images for accurate ROI segmentation. Besides, multi-atlas based segmentation methods are generally time consuming, especially when using a large number of atlas images [17]. Note that, for clinical applications, computation time is one of the most important issues that has to be considered.

To address these issues, in this paper, we propose an

L. Sun, W. Shao, and D. Zhang are with the College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, MIIT Key Laboratory of Pattern Analysis and Machine Intelligence, Nanjing 211106, China. M. Liu is with the Department of Information Science and Technology, Taishan University, Taian 271000, China.

*Corresponding authors: D. Zhang (dqzhang@nuaa.edu.cn) and M. Liu (mxliu1226@gmail.com).

This work was supported by the National Key Research and Development Program of China (Nos. 2018YFC2001600, 2018YFC2001602) and the National Natural Science Foundation of China (Nos. 61876082, 61861130366, 61703301, 61732006 and 61902183), the Royal Society-Academy of Medical Sciences Newton Advanced Fellowship (No. NAF\R1\180371), Taishan Scholar Program of Shandong Province in China, Shandong Natural Science Foundation for Distinguished Young Scholar in China (No. ZR2019YQ27), Scientific Research Foundation of Taishan University (No. Y-01-2018019), and China Postdoctoral Science Foundation funded project (No. 2019M661831)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2962792, IEEE Transactions on Medical Imaging

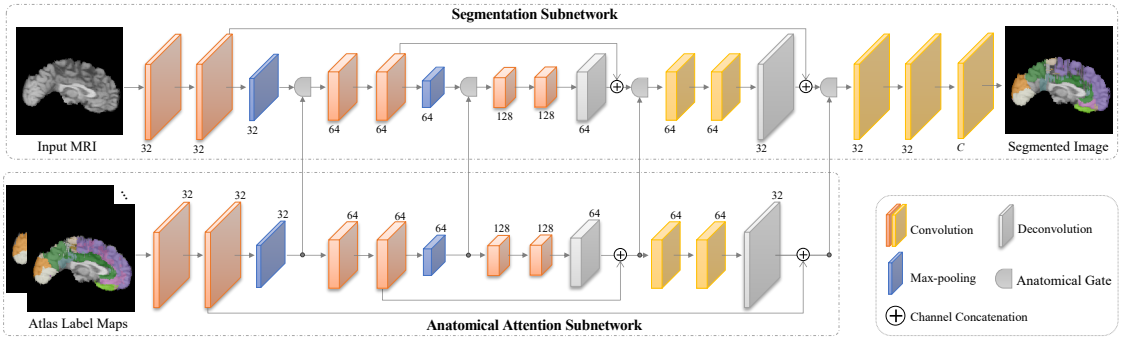

Fig. 1. Illustration of the proposed anatomical attention guided deep learning framework for ROI segmentation of brain MR images. Two major components are included: (1) a segmentation subnetwork for end-to-end brain ROI segmentation (with a U-Net architecture [18], [19] for illustration), and (2) an anatomical attention subnetwork to capture the anatomical structure information of the brain provided by label maps of multiple atlases. The input contains a to-be-segment brain MR image and multiple atlas label maps, while the output is the label map of the input image.

anatomical attention guided framework to segment ROIs of brain MR images. Specifically, as shown in Fig. 1, the proposed framework consists of two subnetworks, *i.e.*, (1) the segmentation subnetwork (with a U-Net architecture [18], [19] as an illustration) and (2) the anatomical attention subnetwork. The segmentation subnetwork follows a conventional convolutional network architecture with a prediction layer for image segmentation, while the anatomical attention subnetwork only contains several convolutional and deconvolutional layers. The input data of this framework include the to-be-segmented image for segmentation subnetwork and the label maps of multiple registered atlases for the anatomical attention subnetwork. Meanwhile, the output is the segmented image with brain ROIs. To effectively incorporate the anatomical structure information of the brain (*i.e.*, features maps learned by the anatomical attention subnetwork) into the segmentation subnetwork, we introduce an *anatomical gate* to integrate these two subnetworks to a unified framework for ROI segmentation. Accordingly, the extracted feature maps that are derived from the segmentation and anatomical attention subnetworks can be adaptively fused in a task-oriented learning manner. Within this framework, we develop two anatomical attention guided segmentation models, which are denoted as anatomical gated fully convolutional network (AG-FCN) and anatomical gated UNet (AG-UNet), respectively. Experimental results on 100 subjects from the ADNI and LONI-LPBA40 datasets suggest that our AG-FCN and AG-UNet methods achieve superior performance in multiple ROI segmentation, compared with several state-of-the-art methods.

The major contributions of this paper can be summarized as follows. *First*, we create an anatomical attention guided deep learning framework to explicitly make use of anatomical prior of brain structures (provided by multiple labeled atlases) for ROI segmentation. *Second*, we develop an anatomical gate architecture to fuse the extracted features from the to-be-segmented image and label maps of registered atlases in a data-driven manner. *Third*, within the proposed framework, we propose two anatomical attention guided deep learning models (*i.e.*, AG-FCN and AG-UNet) using two different network architectures. *Fourth*, we evaluate the proposed methods on ROI segmentation of brain MR images from two public datasets (*i.e.*, ADNI and LONI-LPBA40), with experimental results suggesting the effectiveness of our methods.

The rest of the paper is organized as follows. We first briefly review related studies in Section II. Then, we introduce the proposed anatomical attention guided deep learning framework in Section III. In Section IV, we present materials used in this study, experimental settings, and experimental results. In Section V, we study the influence of parameters in the proposed methods and present the limitations of the current study as well as possible future directions. We finally conclude this paper in Section VI.

II. Related Work

In this section, we first briefly review relevant studies on deep learning based methods for medical image segmentation, and then introduce related studies on multi-atlas based methods for ROI segmentation of brain MR images.

A. Deep Learning for Medical Image Segmentation

In recent years, convolutional neural networks (CNNs) have shown competitive performance in the field of medical image segmentation [7], [8], abnormal region detection [9], [10], and disease diagnosis [6], [11], [12]. Among numerous CNN-based methods, the end-to-end network architecture is commonly used in the task of medical image segmentation, which can directly map the original image from its intensity space to a label space. These end-to-end networks typically consist of an encoding path and a decoding path. The encoding path usually contains several convolutional and pooling operations, which can automatically learn the high-level contextual features of the to-be-segmentation image. The decoding path typically contains several up-sampling/deconvolutional operations, which can decode low-resolution feature maps to high-resolution ones. Due to the up-sampling/deconvolutional operation, the segmented image (*i.e.*, label map) has the same size as the input image. In the early end-to-end network architectures, fully connected layers are usually used to convert the high dimensional feature maps to **1**D feature vector. Since the parameter number of fully connected layers depends on the size of input images, these networks with fully connected layers can only process images with a fixed size.

As one of the state-of-the-art end-to-end architectures, fully convolutional networks (FCN) [20] contain only convolutional

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2962792, IEEE Transactions on Medical Imaging

layers, and hence can perform voxel-wise segmentation for a whole image with an arbitrary size. Moreover, the absence of fully connected layers could reduce the number of parameters, which makes FCN faster than traditional network architectures in both training and test stages. In FCN, the pooling operation is usually used in the encoding path to extract high-level contextual information of the input image, but it also leads to the loss of spatial information of images, thereby reducing the performance of dense prediction.

To effectively combine high-resolution spatial features with high-level contextual features for dense prediction, the U-Net [18], [19] architecture is proposed for biomedical image segmentation. Similar to conventional FCN methods, U-Net also consists of both encoding and decoding paths, with a coarse-fine-connected shortcut architecture. Due to shortcuts for multi-scale feature integration, the high-level contextual feature maps and the high-resolution feature maps can be fused in U-Net to improve the performance of dense prediction. However, existing deep learning methods generally ignore the important anatomical structure information of the brain, thus prone to generating sub-optimal performance in ROI segmentation of brain MR images.

B. Multi-atlas Methods for Brain ROI Segmentation

As a hot topic in the field of medical image analysis, large amounts of neuroimage-based applications often depend on brain ROI segmentation. Among various machine learning methods, multi-atlas based segmentation methods [13]–[16], [21]–[38] have shown their advantages in medical image segmentation in recent years, especially for brain ROI segmentation. Generally, multi-atlas based segmentation methods consist of two key steps, *i.e.*, 1) *image registration* [39]–[42] and 2) *label fusion*. In the image registration step, both affine registration and deformable registration are performed to register multiple atlas images onto a common space of the to-be-segmented image. Then, in the label fusion step, labels of multiple atlases are propagated to the target image by using a specific label fusion strategy. Many recent studies focus on the second step (*i.e.*, label fusion) for multi-atlas based ROI segmentation of brain MR images.

As a commonly used label fusion method, the majority voting (MV) strategy treats each propagated label of atlas equally when determining the final label of to-be-segmented voxels. However, MV-based method ignores the inter-variance between different subjects, by treating all atlases equally. Based on the assumption of voxels should have the same label if they have the similar local appearance, the locally-weighted voting (LWV) method [21] is proposed for label fusion, which considers the pairwise local appearance similarity between the to-be-segmentation voxel and voxels at the same location in each atlas. Hence, the voxel in an atlas with the high local appearance similarity to the to-be-segmented voxel has a large voting weight for label fusion. However, the performances of MV and LWV heavily depend on the results of image registration algorithms, while it is inevitable to produce registration errors in the registration step. To alleviate possible registration errors, several non-local label fusion methods have been proposed. As an example, non-local mean patch-based methods (PBM) [14] propagate the labels not only from the same location in each atlas, but also from a certain local region based on the patch-wise similarity. More recently, some learning-based methods are proposed, such as the joint label fusion method (JLF) [16], [43] that minimizes the expectation of labelling error between the similar patch to jointly learn the voting weights of patches. The JLF method reduces the risk of propogated labelling errors from a similar patch on the atlases. Besides, sparse dictionary learning methods have been also employed for learning the voting weights for label fusion. In the sparse patch-based method (SPBM) [23], a set of patches in a search region of the to-be-segmented voxel on the atlas is firstly selected to construct a region-specific dictionary. Then, this dictionary is used to reconstruct the target patch centered at the to-be-segmented voxel. Due to the use of an l_1-norm constraint, only a small number of patches with high similarity to the target patch are finally selected to determine the label of to-be-segmented images.

Using anatomical prior knowledge provided by multiple atlas images has been demonstrated to be useful in improving the performance of brain ROI segmentation with MR images [21], [25], [27], [28], [34], [35]. Existing multi-atlas based methods usually use handcrafted MRI features (*e.g.*, image intensity), which may degrade segmentation performance due to heterogeneity between features and subsequent label propagation algorithms. Besides, multi-atlas based methods are generally time-consuming since the segmentation is performed in a voxel-by-voxel manner. In contrast, deep networks can extract task-oriented features of brain MR images in a data-driven manner and predict dense label maps of target images at the entire image level (other than for voxel level), so they are usually much faster than multi-atlas based methods. In order to combine the advantages of both deep learning and multi-atlas based methods, we will introduce a unique anatomical attention guided deep learning framework to perform ROI segmentation of brain MR images, where the anatomical prior provided by multiples atlases can be explicitly employed to guide the image segmentation process.

III. Methodology

In this section, we first introduce the notations used in this paper. We then present the architecture of the proposed anatomical attention guided deep network and introduce the proposed anatomical gate in detail. Finally, we introduce the implementation details of brain ROI segmentation using our proposed method.

A. Notations

Given a to-be-segmented brain MR image $\mathbf{I} \in R^{w \times h \times d}$, where w, h and d is the dimension of the input MR image. The aim of brain ROI segmentation is to automatically segment the MR image into multiple ROIs and obtain its label map $\boldsymbol{L}_{\mathbf{I}}$. To capture the complicated anatomical structures of the human brain, we use a set of labeled atlases to guide the network training process. In this study, we use $\mathbf{A}_k$ to denote the k-th atlas, which contains the atlas image $\mathbf{I}_k$ and its corresponding label map $\boldsymbol{L}_k$.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2962792, IEEE Transactions on Medical Imaging

B. Anatomical Attention Guided Deep Learning Framework

Fig. 1 shows the proposed anatomical attention guided deep learning framework implemented for the U-Net architecture [18], [19]. There are two subnetworks in the proposed framework, *i.e.*, (1) the segmentation subnetwork, and (2) the anatomical attention subnetwork to model the anatomical structure information of the brain provided by label maps of multiple atlases.

The segmentation subnetwork has a similar network architecture as U-Net [18], [19], where the left half part is the encoding path and the right half is the decoding path. In the *encoding* path, there are six convolutional layers (size: $3 \times 3 \times 3$) and two max-pooling layers (size: $2 \times 2 \times 2$). The first two convolutional layers and the first max-pooling layer have the same number of channels (*i.e.*, 32), the following two convolutional layers and the second max-pooling layer have 64 channels, and the last two convolutional layers have 128 channels. Each convolution operation is performed using $3 \times 3 \times 3$ kernel, followed by batch normalization (BN) and rectified linear unit (ReLU) activation. Different from conventional U-Net architecture, the proposed anatomical gate follows each max-pooling layer in the encoding path of our network, which is used to fuse the down-sampled feature maps generated by the segmentation subnetwork and the down-sampled features maps of the anatomical attention subnetwork.

In the *decoding* path, a deconvolutional layer (size: $2\times2\times2$; number of channels: 64) is employed to up-sample the feature maps generated by the encoding path. Then, the output of the deconvolutional layer is concatenated with the output of its corresponding convolution layer in the encoding path. An anatomical gate (with details given in Section III-C) follows the concatenated layer, which is used to incorporate the up-sampled features maps of the anatomical attention subnetwork into the segmentation subnetwork. The output of this anatomical gate is further fed into two convolutional layers (size: $3 \times 3 \times 3$; number of channels: 64) with BN and ReLU, followed by a deconvolutional layer (size: $2 \times 2 \times 2$; number of channels: 32) for image up-sampling. Then, an additional anatomical gate is applied to the concatenated layer (of the output of the former deconvolutional layer and that of its corresponding convolutional layer in the encoding path), followed by two 32-channel convolutional layers (size: $3\times3\times3$) and a C-channel convolutional layers (size: $1\times1\times1$). Finally, a softmax non-linear unit is used for predicting the probability map $\mathbf{P} = \{\boldsymbol{p}_i\}_{i=1}^{w\times h\times d}$ for the to-be-segmented image $\mathbf{I}$, where $\boldsymbol{p}_i \in \mathbb{R}^C$ denote the probability of the i-th voxel belonging to the a specific ROI or background, and C is the number of categories (including ROIs and background). Based on the ground-truth label map for the training image, we employ a cross-entropy loss to train the network as follows

$$-\frac{1}{N \times w \times h \times d} \sum_{j=1}^{N} \sum_{i=1}^{w\times h\times d} \sum_{c=1}^{C} \delta(L_{I_{j,i}}, c) \log p_{j,i}^{c}, \quad (1)$$

where N is batch size, and $\delta(L_{I_{j,i}}, c)$ is a Dirac function, which equal to 1 when $L_{I_{j,i}} = c$; and 0, otherwise. Also, the term $p_{j,i}^{c}$ denotes the probability of the i-th voxel of each image in a batch belonging to the c-th category.

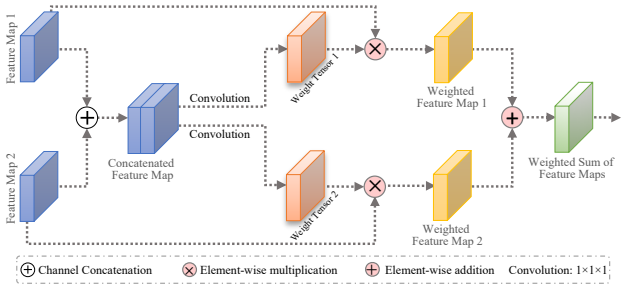

Fig. 2. Overview of the proposed anatomical gate in our anatomical attention guided deep learning framework, used to integrate anatomical prior provided by atlases into the segmentation process. The input of the anatomical gate includes two feature maps generated by two subnetworks, which are first concatenated channel-wisely. Then, the concatenated feature map is fed into two parallel convolutional layers. Each convolutional layer is followed by a sigmoid unit to learn the specific weight tensor for each input feature map. Each weighted tensor is further combined with its original feature map by an element-wise multiplication operation, leading to a weighted feature map. Finally, the weighted feature maps corresponding are fused via an element-wise addition operation to generate a weighted sum feature map.

The proposed anatomical attention subnetwork is a modified version of the segmentation subnetwork, with label maps of multiple atlases as input data. For each atlas, we first register each of multiple atlases onto the to-be-segmented images. Typically, we first use FLIRT in the FSL [40] toolbox for affine registration, and then a deformable registration is performed using the Diffeomorphic Demons method [41]. Each label map of the registered atlas is treated as a channel of the input of our anatomical attention subnetwork. As shown in the bottom part of Fig. 1, in the anatomical attention subnetwork, each block of encoding path consists of two convolutional layers and a max-pooling layer with the same architecture as the segmentation subnetwork, while each block of the decoding path consist of a deconvolution layer and two convolutional layers with the same architecture as the segmentation subnetwork. Among these layers, feature maps of each max-pooling layer or deconvolution layer are integrated into the segmentation subnetwork via the proposed anatomical gate. Hence, the anatomical structure information provided by labeled atlases can be employed as the guidance information to boost the performance of ROI segmentation with brain MR images. In the following, we give the details of the proposed anatomical gate for fusing the extracted feature maps generated from both the segmentation subnetwork and the anatomical attention subnetwork.

C. Anatomical Gate

Because of the complicated human brain structure, low intensity contrast around the boundary of ROIs and large inter-subject variance in brain MR images, only using image intensity information is not enough to perform accurate brain ROI segmentation. In the multi-atlas based segmentation framework, the label fusion step typically utilizes the anatomical prior from multiple atlases for ROI segmentation. Accordingly, we also introduce the anatomical prior to the proposed deep network for ROI segmentation of brain MR images. As mentioned in Section III-B, our proposed network consists of two subnetworks, *i.e.*, the segmentation subnetwork and the anatomical attention subnetwork. Specifically, the

segmentation subnetwork learns feature maps based on image intensity, which only encode the local contextual information of brain MR images. The anatomical attention subnetwork learns the anatomical prior of the brain from a set of labeled atlases, which encodes the local brain structure information from the registered atlases.

To better integrate the local contextual information of MR images and brain anatomical structure information, in this work, we propose an *anatomical gate* to fuse feature maps generated by both the segmentation and anatomical attention subnetworks. As shown in Fig. 2, the output feature map f_i^s of the s-th layer in the segmentation subnetwork and that generated by the anatomical attention subnetwork f_a^s are firstly concatenated as $[f_i^s, f_a^s]$ channel-wisely. Then, the concatenated feature map is fed into two parallel convolutional layers (size: $1\times1\times1$), and each convolutional layer is followed by a sigmoid non-linear unit to learn the specific weight tensor (*e.g.*, o_i^s) for each input feature map. Specifically, these two weight tensors are learned as follows

$$o_i^s = \sigma(W_i^s * [f_i^s, f_a^s] + b), \tag{2}$$

$$o_a^s = \sigma(W_a^s * [f_i^s, f_a^s] + b), \tag{3}$$

where o_i^s and o_a^s denote the weight tensors corresponding to the input feature maps f_i^s and f_a^s, respectively. Using the sigmoid unit, the weight values in o_i^s and o_a^s are constrained within the range of $[0, 1]$. Besides, in Eqs. 2-3, (W_i^s, b) and (W_a^s, b) are parameters corresponding to two convolutional layers, respectively.

Each weight tensor is further combined with its original feature map by an element-wise multiplication operation, leading to a weighted feature map. Finally, the weighted feature maps corresponding to the segmentation and anatomical attention subnetworks are fused by an element-wise addition operation to generate a weighted sum feature map. The output feature map f_o^s of the anatomical gate can be represented as:

$$f_o^s = o_i^s \cdot f_i^s + o_a^s \cdot f_a^s. \tag{4}$$

Based on the proposed anatomical gates at different layers, the proposed deep network in Fig. 1 can *not only* capture the local context information of the to-be-segmented images (via the segmentation subnetwork), *but also* includes the anatomical prior of brain structures provided by multiple atlases (via the anatomical attention subnetwork) at different scales. Notably, using the anatomical gates, our method could automatically learn the optimal weights of feature maps generated by two subnetworks in Fig 1, which is expected to efficiently fuse two subnetworks for accurate ROI segmentation.

D. Implementation

As shown in Fig. 1, in the proposed anatomical attention guided deep learning framework, we employ the U-Net architecture [18], [19] to implement the segmentation subnetwork, and we denote this model as attention gated U-Net (**AG-UNet**) in this work. Using different network architectures, we can derive different models for brain ROI segmentation. Therefore, besides using U-Net, we further employ a fully convolutional network (FCN) [20] to implement the segmentation subnetwork, and denote this model as attention gated FCN (**AG-FCN**). The detailed architecture of AG-FCN can be found in Fig. S1 of the *Supplementary Materials*. Note that in our AG-UNet and AG-FCN models, the anatomical attention subnetworks share the similar network architecture with their corresponding segmentation subnetworks, respectively.

In the *training* stage, we feed label maps of multiple atlases to the attention subnetwork and the training images to the segmentation subnetwork, respectively. These multiple atlases have been aligned to each training MR image via affine and deformable registration. With the ground-truth label map of the training image as the output, we train the proposed anatomical attention guided network in an end-to-end manner. We empirically set the mini-batch size as 1, the number of epochs as $1,000$, and the learning rate as 0.001, respectively. It requires ~ 24 hours to train each of the proposed two models.

In the *test* stage, we first align multiple atlases to the to-be-segmented (*i.e.*, target) image. Then, we feed both label maps of multiple atlases and the target image to the trained network to predict its probability map for brain ROI segmentation. Given C ROIs, for each to-be-segmented voxel v_i, we use the MAP criterion to obtain its label as follows

$$l(v_i) = \arg\max_c \ \{\boldsymbol{p}_i^c\}_{c=1}^C, \tag{5}$$

through which we can generate a label map for the test image.

IV. Experiment

In this section, we first present materials and experimental settings used in our study. We then present experimental results achieved by different methods on two public datasets with brain MR images.

A. Materials

We evaluate the proposed methods on two public datasets, including 1) the Alzheimer's Disease Neuroimaging Initiative (ADNI) dataset [44], and 2) the LONI-LPBA40 dataset [45]. More details can be found as follows.

1) **ADNI [44]:** Following previous studies [24], [29], we employ 60 subjects from ADNI for *hippocampus* segmentation, including 20 Alzheimer's disease (AD) subjects, 20 mild cognitive impairment (MCI) subjects and 20 normal control (NC) subjects. These brain MR images were acquired in the sagittal view, with the in-plane resolution of $1\,mm \times 1\,mm$ and the slice thickness of $1.2\,mm$. All images are resampled to have the resolution of $1\times1\times1\,mm^3$ with trilinear interpolation. The ground-truth label maps were created manually to annotate the right and left *hippocampus* regions in the brain. We perform pre-processing for all MR images via three procedures, including skull removal [46], N4-based bias field correction [47], and intensity standardization [48]. Following [16], [24], we randomly select 20 subjects as atlas images, and the remaining images are randomly split into 2 subsets for 2-fold cross-validation on ADNI.
2) **LONI-LPBA40 [45]:** The LONI-LPBA40 datasets is provided by the Laboratory of Neuro Imaging (LONI)

at UCLA, which contains 40 brain MR images and their corresponding label maps were created manually to annotate the brain structures. High-resolution 3D Spoiled Gradient Echo (SPGR) MRI volumes were acquired on a GE 1.5 Tesla system as 124 contiguous 1.5 mm coronal brain slices (TR: 10.00-12.50 ms; TE: 4.22-4.50 ms; FOV: 220 mm or 200 mm) with in-plane voxel resolution of 0.86 mm or 0.78 mm. All images are resampled to have the resolution of $1 \times 1 \times 1\, mm^3$ with trilinear interpolation. Besides, these MRI volumes are rigidly aligned to the MNI305 template [45]. Following previous studies [24], [49], for this dataset, we randomly select 20 subjects as atlas images, and the remaining 20 subjects are randomly split into 5 subsets for 5-fold cross-validation on LONI-LPBA40.

B. Experimental Settings

For all pre-processed brain MR images, we perform affine registration by FLIRT in the FSL [40] toolbox, using the normalized mutual information as the similarity metric, 12 degrees of freedom and the search range ± 20 in all directions. Then, we further perform a deformable registration using the Diffeomorphic Demons method [41] with default parameters (*i.e.*, smoothing kernel size of 2.0, and iterations in low, middle and high resolutions as $20 \times 10 \times 5$).

By implementing our proposed framework based on two well-known network architectures (*i.e.*, **FCN** and **U-Net**), we have two novel anatomical gated networks *dubbed* as AG-FCN and AG-UNet, respectively. We compare our proposed AG-FCN and AG-UNet methods with their conventional counterparts (*i.e.*, FCN and U-Net) in the experiments. Two evaluation metrics are used to measure the segmentation performance of different methods in the experiments. Specifically, we first use the Dice coefficient (*DC*) as the evaluation metric, defined as

$$DC = \frac{2|R_1 \cap R_2|}{|R_1| + |R_2|}, \tag{6}$$

where the term $\cap$ denotes the overlap between the segmented region R_1 and the ground truth R_2, and $|\cdot|$ denotes the number of voxels belonging to each ROI. Meanwhile, we also use the average surface distance (*ASD*) to measure the performance of different segmentation algorithms, defined as

$$\begin{aligned} ASD = \frac{1}{2}\Big(&\frac{1}{n_1}\sum_{r_1 \in S(R_1)} d(r_1, S(R_2)) \\ &+ \frac{1}{n_2}\sum_{r_2 \in S(R_2)} d(r_2, S(R_1))\Big), \end{aligned} \tag{7}$$

where $d(\cdot,\cdot)$ measures the Euclidean distance, and n_1 and n_2 are the numbers of vertices in the surface $S(R_1)$ and $S(R_2)$, respectively. Also, r_1 and r_2 denotes vertices in the surface $S(R_1)$ and $S(R_2)$, respectively.

C. Results on ADNI

We first perform *hippocampus* segmentation on the ADNI dataset. Table I shows the Dice coefficient (*DC*) and the average surface distance (*ASD*) values achieved by our AG-FCN, AG-UNet and their conventional counterparts (*i.e.*, FCN and U-Net). We further perform the Wilcoxon signed rank test on the Dice coefficient results achieved by different segmentation methods. Our proposed AG-FCN and AG-UNet show the significant improvement ($p < 0.05$) over FCN ($p = 4.4934e-04$) and U-Net ($p = 3.9023e-04$) on *hippocampus* segmentation task. The symbol '$*$' in Table I indicates that our proposed AG-FCN and AG-UNet achieves statistically significant improvement over their conventional counterparts, respectively.

TABLE I
SEGMENTATION RESULTS OF FCN, AG-FCN, U-NET AND AG-UNET ON THE ADNI DATASET FOR THE *hippocampus* SEGMENTATION. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. THE SYMBOL '$*$' INDICATES THAT OUR PROPOSED METHOD CAN SIGNIFICANTLY IMPROVE ITS CONVENTIONAL COUNTERPART BASED ON WILCOXON SIGNED RANK TEST IN TERMS OF *DC*.

Method	*DC*	*ASD* (mm)
FCN	0.8206 ± 0.0278	0.588 ± 0.078
*AG-FCN (Ours)	0.8493 ± 0.0250	0.541 ± 0.075
U-Net	0.8597 ± 0.0159	0.536 ± 0.071
*AG-UNet (Ours)	$\mathbf{0.8864 \pm 0.0212}$	$\mathbf{0.386 \pm 0.058}$

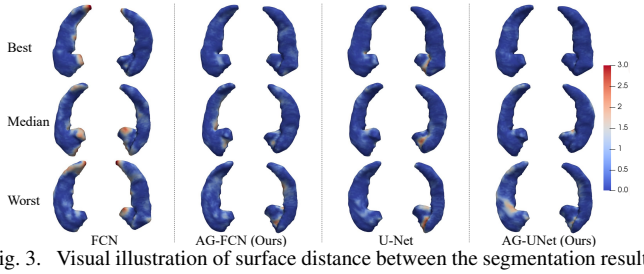

Fig. 3. Visual illustration of surface distance between the segmentation results of different methods and ground truth on the *hippocampus* region. The best, median and worst are the best, media and worst segmented subjects in terms of average surface distance by the proposed AG-UNet method, respectively.

From Table I, we can observe that the proposed AG-UNet achieves the best performance for *hippocampus* segmentation regarding the Dice coefficient metric. For example, our AG-UNet methods achieves the highest Dice coefficient (*i.e.*, 0.8864), which is significantly better than the U-Net method (*i.e.*, 0.8597). Meanwhile, the proposed AG-FCN method also achieves significant improvement over FCN. In general, the proposed AG-FCN and AG-UNet achieve 0.0287 and 0.0267 improvement in terms of Dice coefficient over their counterparts, *i.e.*, FCN and U-Net, respectively. Besides, our proposed methods also achieve better results in terms of *ASD* values, compared with FCN and U-Net. The *ASD* values achieved by AG-FCN and AG-UNet for *hippocampus* segmentation are 0.541 and 0.386, respectively, Which is significantly better than FCN and U-Net. These results demonstrate that incorporating the anatomical structure information (provided by atlases) to the deep learning framework, as we do in AG-FCN and AG-UNet methods, can boost the segmentation performance. On the other hand, compared the FCN and AG-FCN methods, the U-Net and AG-UNet approaches usually achieve better segmentation results. The possible reason is that, with the coarse-fine-connected shortcut architecture, U-Net and AG-UNet can fuse the high-level global contextual feature maps and the high-resolution global feature maps, while FCN

TABLE II
SEGMENTATION RESULTS ACHIEVED BY FCN, AG-FCN, U-NET AND AG-UNET ON THE LONI-LPBA40 DATASET. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. THE SYMBOL '*' INDICATES THAT OUR PROPOSED METHOD CAN SIGNIFICANTLY IMPROVE ITS CONVENTIONAL COUNTERPART BASED ON WILCOXON SIGNED RANK TEST ($p < 0.05$) IN TERMS OF *DC*.

Method	*DC*	*ASD* (mm)
FCN	0.7625 ± 0.0399	1.189 ± 0.078
*AG-FCN (Ours)	0.7826 ± 0.0377	1.099 ± 0.037
U-Net	0.7817 ± 0.0409	1.142 ± 0.220
*AG-UNet (Ours)	$\mathbf{0.8067 \pm 0.0383}$	$\mathbf{1.070 \pm 0.036}$

and AG-FCN has no such shortcut architecture.

In Fig. 3, we plot the best, median and worst automatically segmented subjects by AG-UNet in terms of average surface distance between the segmentation images and ground truth on the left and right *hippocampus* regions. Meanwhile, we also plot surface distance between the automatic segmentation images and ground truth, achieved by FCN, AG-FCN and U-Net, respectively. As shown in Fig. 3, our proposed methods produce the better quality of segmentation results on *hippocampus* when compared with their conventional counterparts, respectively. These results further validate that conventional deep learning methods (*i.e.*, FCN and U-Net) using only the intensity image can't yield accurate segmentation results in brain ROI segmentation. Incorporating the anatomical prior from atlases into deep networks could further improve the performance for brain ROI segmentation, as we do in AG-FCN and AG-UNet. The possible reason for the improvement is that the anatomical prior provide the information of brain structures, which can enhance the segmentation results around the boundary of ROIs with low intensity contrast.

D. Results on LONI-LPBA40

In the second group of experiments, we validate our proposed methods on the LONI-LPBA40 dataset to segment 54 ROIs in each brain MR image. The segmentation results achieved by four different methods are shown in Table II and Fig. S2 in the *Supplementary Materials*. We also perform the Wilcoxon signed rank test on each ROI in terms of Dice coefficient for our methods and their conventional counterparts.

From Table II, we can see that the average Dice coefficient on 54 ROIs are 0.7826 and 0.8067 yielded by AG-FCN and AG-UNet, respectively, which are higher than those achieved by FCN (0.7625) and U-Net (0.7817). In general, the proposed AG-FCN and AG-UNet achieved 0.0201 and 0.0250 improvements over their counterparts, respectively. The achieved average surface distance on 54 ROIs are 1.099 mm and 1.070 mm by our proposed AG-FCN and AG-UNet, respectively, compared with 1.189 mm and 1.142 mm by FCN and U-Net, respectively. Besides, our AG-FCN and AG-UNet methods achieve significant improvement ($p < 0.05$) over FCN ($p = 8.8575e-05$) and U-Net ($p = 1.0335e-04$) in terms of Dice coefficient, respectively. As can be seen from Fig. S2, the Dice coefficient on 54 ROIs achieved by AG-FCN and AG-UNet outperform their conventional counterparts (*i.e.*, FCN and U-Net) in most ROIs.

V. DISCUSSION

In this section, we first compare our proposed deep learning methods with several state-of-the-art multi-atlas based segmentation methods for brain ROI segmentation. Then, we study the influence of the important parameter (*i.e.*, the number of atlases) and deformable registration process on the performance of our methods. Finally, we present the limitations of this work as well as possible future research directions.

A. Comparison with Multi-atlas Segmentation Methods

Since multi-atlas based methods have been widely studied in the field of brain ROI segmentation, we now compare our proposed deep learning methods (*i.e.*, AG-FCN and AG-UNet) with several state-of-the-art multi-atlas segmentation methods on both the ADNI and LONI-LPBA40 datasets. Specifically, the proposed AG-FCN and AG-UNet methods are compared with four well-known multi-atlas segmentation methods, including the locally-weighted weighting (**LWV**) method [21], the patch-based method (**PBM**) [14], the joint label fusion (**JLF**) method [16], [43], and the sparse patch-based method (**SPBM**) [23]. In these four multi-atlas segmentation methods, we utilize the most commonly used parameters in literature, *i.e.*, both the patch size and search region are set as $7 \times 7 \times 7$. For a fair comparison, all six methods (*i.e.*, LWV, PBM, JLF, SPBM, AG-FCN, and AG-UNet) employ the same parameters for affine registration by FLIRT in the FSL [40] toolbox and deformable registration by Diffeomorphic Demons method [41]. In the experiments, the time costs for image registration using the FLIRT and Diffeomorphic Demons algorithms are about 30 seconds and 120 seconds per brain MR image, respectively. Note that all six methods use anatomical structure information of the brain provided by multiple labeled atlases. The difference is that our methods (*i.e.*, AG-FCN and AG-UNet) use end-to-end deep networks to learn image features in a task-oriented manner, while the multi-atlas based approaches (*i.e.*, LWV, PBM, JLF, and SPBM) employ hand-crafted features (*i.e.*, image intensity) to represent brain MR images. The experimental results achieved by six different methods on the ADNI and LONI-LPBA40 datasets are reported in Table III and Table IV, respectively. The symbol '*' in Table III and Table IV indicates that our proposed AG-UNet achieves statistically significant improvement over the multi-atlas segmentation method.

As shown in Table III and Table IV, our proposed AG-UNet method achieves the overall best segmentation results for brain ROI segmentation on both ADNI and LONI-LPBA40 datasets. More specifically, it can be seen from Table III, our proposed AG-UNet method achieves the best segmentation performance on the ADNI dataset for *hippocampus* segmentation. The Dice coefficient and the average surface distance achieved by AG-UNet are 0.8864 and 0.386 mm for *hippocampus* segmentation on ADNI, respectively, which are superior to the best results of multi-atlas based methods (*i.e.*, the Dice coefficient and the average surface distance are 0.8775 and 0.401 mm achieved by SPBM). Also, as can be observed from Table IV that the proposed AG-UNet achieves the best segmentation results on LONI-LPBA40 for segmentation of

TABLE III
SEGMENTATION RESULTS OF FOUR MULTI-ATLAS BASED METHODS (*i.e.*, LWV, PBM, JLF, SPBM) AND OUR AG-FCN AND AG-UNET METHODS ON THE ADNI DATASET FOR *hippocampus* SEGMENTATION. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. THE SYMBOL '*' INDICATES THAT OUR PROPOSED AG-UNET ACHIEVED SIGNIFICANTLY IMPROVEMENT OVER THE MULTI-ATLAS SEGMENTATION METHOD BASED ON WILCOXON SIGNED RANK TEST ($p < 0.05$) IN TERMS OF *DC*.

Method	*DC*	*ASD* (mm)
*LWV	0.8546 ± 0.0144	0.453 ± 0.041
*PBM	0.8697 ± 0.0310	0.456 ± 0.138
*JLF	0.8731 ± 0.0395	0.405 ± 0.074
*SPBM	0.8775 ± 0.0378	0.401 ± 0.096
AG-FCN (Ours)	0.8493 ± 0.0250	0.541 ± 0.075
AG-UNet (Ours)	$\mathbf{0.8864 \pm 0.0212}$	$\mathbf{0.386 \pm 0.058}$

TABLE IV
SEGMENTATION RESULTS OF FOUR MULTI-ATLAS BASED METHODS (*i.e.*, LWV, PBM, JLF, SPBM) AND OUR AG-FCN AND AG-UNET METHODS ON THE LONI-LPBA40 DATASET FOR SEGMENTATION OF MULTIPLE ROIS. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. THE SYMBOL '*' INDICATES THAT OUR PROPOSED AG-UNET ACHIEVED SIGNIFICANTLY IMPROVEMENT OVER THE MULTI-ATLAS SEGMENTATION METHOD BASED ON WILCOXON SIGNED RANK TEST ($p < 0.05$) IN TERMS OF *DC*.

Method	*DC*	*ASD* (mm)
*LWV	0.7822 ± 0.0088	1.234 ± 0.049
*PBM	0.7881 ± 0.0091	1.170 ± 0.056
*JLF	0.7926 ± 0.0107	1.181 ± 0.061
SPBM	0.7965 ± 0.0100	1.196 ± 0.048
AG-FCN (Ours)	0.7826 ± 0.0377	1.099 ± 0.037
AG-UNet (Ours)	$\mathbf{0.8067 \pm 0.0383}$	$\mathbf{1.070 \pm 0.036}$

multiple ROIs. Besides, as shown in Table III, Table IV, the proposed AG-FCN method produces worse results than AG-UNet on both ADNI and LONI-LPBA40 datasets. For instance, in terms of average surface distance, AG-UNet yields a much better result than AG-FCN on the LONI-LPBA40 dataset. The possible reason is that, even though AG-FCN can extract high-level contextual features of brain MR images, it loses local spatial information of brain MR images by using pooling operations. In contrast, the AG-UNet method employs multiple skip connection operations to fuse the high-level contextual features and high-resolution spatial features for the dense prediction, which can boost the segmentation performance. This can also be seen from Fig. 3. That is the proposed AG-FCN with the FCN architecture produces coarse segmentation results for brain MR images with the low intensity contrast around the boundary of ROIs, in comparison to AG-UNet. Besides, the proposed AG-UNet shows significant improvement over the most of multi-atlas segmentation methods based on Wilcoxon signed rank test ($p < 0.05$). For example, the proposed AG-UNet achieves significant improvement over JLF ($p = 1.0335e-04$) and SPBM ($p = 5.9342e-04$) in terms of Dice coefficient on ADNI dataset for *hippocampus* segmentation, respectively. Also, the p-values of AG-UNet over JLF and SPBM are 0.0013 and 0.1014 in terms of Dice coefficient on LONI-LPBA40 for whole brain segmentation.

On the other hand, compared to traditional multi-atlas based methods, the proposed anatomical attention guided deep networks (*i.e.*, AG-FCN and AG-UNet) are more computationally efficient. For example, without considering the time consumption of image registration, the JLF method requires more than 1 hour[1] to segment each MR image into 54 ROIs using 20 atlases on LONI-LPBA40, while our proposed AG-UNet only need 5 seconds per image with 20 atlases. This could be partly because our end-to-end learning strategy used in AG-UNet has high computational efficiency for dense segmentation.

B. Comparison with State-of-the-art Deep Learning Methods

Our proposed method is a general framework and can easily be combined with existing state-of-the-art segmentation network architectures. Within this framework, besides AG-FCN and AG-UNet, we further develop three methods based on three state-of-the-art network architectures, *i.e.*, DeepNAT [50], residual-FCN [51] (R-FCN) and attention-UNet [52] (A-UNet), and denote the corresponding methods as anatomical gated DeepNAT (AG-DeepNAT), anatomical gated A-UNet (AG-AUNet), and anatomical gated R-FCN (AG-RFCN), respectively. We evaluate the proposed three methods (*i.e.*, AG-DeepNAT, AG-RFCN, and AG-AUNet) and their conventional counterparts (*i.e.*, DeepNAT, R-FCN, and A-UNet) on the ADNI and LONI-LPBA40 datasets for brain ROI segmentation, with results are reported in Table V and Table VI, respectively.

As can be seen from Table V and Table VI, the proposed methods consistently outperform their counterparts on two datasets. For example, the AG-AUNet achieved 0.0258 improvement over A-UNet on ADNI dataset for *hippocampus* segmentation. We also perform the Wilcoxon signed rank test on the results achieved by our proposed methods and their counterparts in terms of Dice coefficient, respectively. The symbol '*' in Tables V- VI indicates that our proposed methods achieve statically significant improvement over their counterparts. Our proposed AG-DeepNAT, AG-RFCN and AG-AUNet show significant improvement (p<0.05) over DeepNAT ($p = 1.0335e-04$), R-FCN ($p = 1.0509e-05$) and A-UNet ($p = 5.1004e-05$) on ADNI dataset for *hippocampus* segmentation, respectively. Meanwhile, the proposed AG-DeepNAT, AG-RFCN and AG-AUNet show significant improvement (p<0.05) over DeepNAT ($p = 8.8475e-05$), R-FCN ($p = 8.8575e-05$) and A-UNet ($p = 1.0335e-04$) in terms of Dice coefficient on LONI-LPBA40 dataset for brain ROI segmentation, respectively. Beside, results in Tables I, II, V and VI suggest that R-FCN generally achieves better results than FCN and A-UNet is superior to U-Net. This implies that using self-attention units (as A-UNet) and residual connections (as in R-FCN) could further improve the segmentation performance of deep networks. Also, utilizing anatomical information provided atlases (as we do in AG-DeepNAT) can boost the segmentation performance of DeepNAT. These results further demonstrate the effectiveness of our proposed anatomical attention guided deep learning framework for brain ROI segmentation.

[1] https://www.nitrc.org/projects/picsl_malf

TABLE V
SEGMENTATION RESULTS ACHIEVED BY DEEPNAT, AG-DEEPNAT, R-FCN, AG-RFCN, A-UNET AND AG-AUNET ON THE ADNI DATASET FOR *hippocampus* SEGMENTATION. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. THE SYMBOL '*' INDICATES THAT OUR PROPOSED METHOD CAN SIGNIFICANTLY IMPROVE ITS CONVENTIONAL COUNTERPART BASED ON WILCOXON SIGNED RANK TEST ($p < 0.05$) IN TERMS OF *DC*.

Method	*DC*	*ASD* (mm)
DeepNAT	0.8502 ± 0.0283	0.517 ± 0.084
*AG-DeepNAT(Ours)	0.8695 ± 0.0282	0.448 ± 0.081
R-FCN	0.8331 ± 0.0371	0.574 ± 0.090
*AG-RFCN (Ours)	0.8655 ± 0.0226	0.471 ± 0.064
A-UNet	0.8615 ± 0.0221	0.501 ± 0.073
*AG-AUNet (Ours)	$\mathbf{0.8873 \pm 0.0201}$	$\mathbf{0.379 \pm 0.061}$

TABLE VI
SEGMENTATION RESULTS ACHIEVED BY DEEPNAT, AG-DEEPNAT, R-FCN, AG-RFCN, A-UNET AND AG-AUNET ON THE LONI-LPBA40 DATASET. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. THE SYMBOL '*' INDICATES THAT OUR PROPOSED METHOD CAN SIGNIFICANTLY IMPROVE ITS CONVENTIONAL COUNTERPART BASED ON WILCOXON SIGNED RANK TEST ($p < 0.05$) IN TERMS OF *DC*.

Method	*DC*	*ASD* (mm)
DeepNAT	0.7789 ± 0.0271	1.138 ± 0.136
*AG-DeepNAT (Ours)	0.7987 ± 0.0166	1.085 ± 0.035
R-FCN	0.7705 ± 0.0388	1.139 ± 0.107
*AG-RFCN (Ours)	0.7938 ± 0.0365	1.057 ± 0.065
A-UNet	0.7880 ± 0.0386	1.136 ± 0.186
*AG-AUNet (Ours)	$\mathbf{0.8101 \pm 0.0376}$	$\mathbf{1.046 \pm 0.185}$

C. Influence of Anatomical Gate Architecture

To evaluate the effectiveness of the anatomical gate architecture, we further compare the proposed AG-FCN and AG-UNet with their counterparts using multi-channel and feature concatenation strategies on the ADNI and LONI-LPBA40 datasets. We first compare AG-FCN and AG-UNet with their multi-channel counterparts (called AM-FCN and AM-UNet, respectively). Specifically, AM-FCN and AM-UNet directly use multiple atlases and each input image as multi-channel (*i.e.*, $[\mathbf{I}, \boldsymbol{L}_1, \cdots, \boldsymbol{L}_k]$) input data, while our AG-FCN and AG-UNet feed those K atlases into the proposed Anatomical Attention Subnetwork (as shown in Fig. 1). We then compare AG-FCN and AG-UNet with their feature concatenation counterparts (called AC-FCN and AC-UNet, respectively), where the proposed anatomical gate is replaced by the feature concatenation operation. Specifically, in AC-FCN and AC-UNet, feature maps generated by the segmentation and anatomical attention subnetworks are concatenated channel-wisely.

Table VII and Table VIII report the segmentation results achieved by six different methods on the ADNI and LONI-LPBA40 datasets, respectively. From the Table VII and Table VIII, we can see that the proposed AG-UNet achieves the best performances on the ADNI dataset for *hippocampus* segmentation and the LONI-LPBA40 for whole brain segmentation. For instance, the proposed AG-UNet achieves the best Dice coefficient (0.8864) on ADNI, which is superior to that of AM-UNet (0.8713) and AC-UNet (0.8735). Besides, among three FCN-based methods (*i.e.*, AM-FCN, AC-FCN, and AG-FCN), AG-FCN consistently achieves the best performance

TABLE VII
SEGMENTATION RESULTS ACHIEVED BY AM-FCN, AC-FCN, AG-FCN, AM-UNET, AC-UNET AND AG-UNET ON THE ADNI DATASET FOR *hippocampus* SEGMENTATION. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY.

Method	*DC*	*ASD* (mm)
AM-FCN	0.8380 ± 0.0244	0.557 ± 0.067
AC-FCN	0.8386 ± 0.0279	0.555 ± 0.086
AG-FCN (Ours)	0.8493 ± 0.0250	0.541 ± 0.075
AM-UNet	0.8713 ± 0.0265	0.446 ± 0.085
AC-UNet	0.8735 ± 0.0245	0.441 ± 0.076
AG-UNet (Ours)	$\mathbf{0.8864 \pm 0.0212}$	$\mathbf{0.386 \pm 0.058}$

TABLE VIII
SEGMENTATION RESULTS ACHIEVED BY AM-FCN, AC-FCN, AG-FCN, AM-UNET, AC-UNET AND AG-UNET ON THE LONI-LPBA40 DATASET. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY.

Method	*DC*	*ASD* (mm)
AM-FCN	0.7757 ± 0.0399	1.110 ± 0.045
AC-FCN	0.7783 ± 0.0392	1.109 ± 0.041
AG-FCN (Ours)	0.7826 ± 0.0377	1.099 ± 0.037
AM-UNet	0.7952 ± 0.0445	1.106 ± 0.042
AC-UNet	0.7991 ± 0.0441	1.091 ± 0.041
AG-UNet (Ours)	$\mathbf{0.8067 \pm 0.0383}$	$\mathbf{1.070 \pm 0.036}$

on both datasets. These results suggest that, to take advantage of anatomical prior provided by multiple atlases, our anatomical gate strategy is superior to that the conventional multi-channel strategy and feature concatenation strategy. The possible reason is that, using the proposed anatomical gate architecture to learn task-oriented fusion weights, our AG-FCN and AG-UNet can effectively fuse feature maps learned from the input image and multiple labeled atlases for boosting segmentation performance, compared with their multi-channel variants (*i.e.*, AM-FCN and AM-UNet) and feature concatenation variants (*i.e.*, AC-FCN and AC-UNet). On the other hand, from Tables I-II and Tables VII-VIII, one can observe that AM-FCN, AC-FCN, AG-FCN, AM-UNet, AC-UNet and AG-UNet achieve better results in terms of *DC* and *ASD* over their counterparts, *i.e.*, FCN and U-Net, respectively. It suggests that using anatomical information provided by labeled atlases could boost segmentation performance of FCN and U-Net.

D. Influence of Number of Atlases

The number of atlases is an important parameter in the proposed network. We now study the influence of the number of atlases for our proposed AG-UNet method on the ADNI dataset, by varying the number of atlases within the range of $[5,\ 10,\ 15,\ 20]$. Fig. 4 shows the Dice coefficient and the average surface distance values achieved by our AG-UNet approach using different numbers of atlases.

From Fig. 4, we can see that the best performance is achieved by AG-UNet when the number of atlases is 20 on ADNI for *hippocampus* segmentation. For instance, the best Dice coefficient and average surface distance are 0.8864 and 0.386 mm, respectively, achieved by AG-UNet when the number of atlases is fixed as 20. Also, using only 5 atlases, our AG-UNet method can only yield the Dice coefficient

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TMI.2019.2962792, IEEE Transactions on Medical Imaging

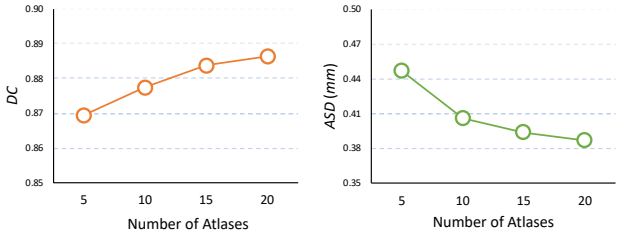

Fig. 4. Dice coefficient (*DC*) and average surface distance (*ASD*) achieved by the proposed AG-UNet method, using different numbers of atlases for *hippocampus* segmentation on the ADNI dataset.

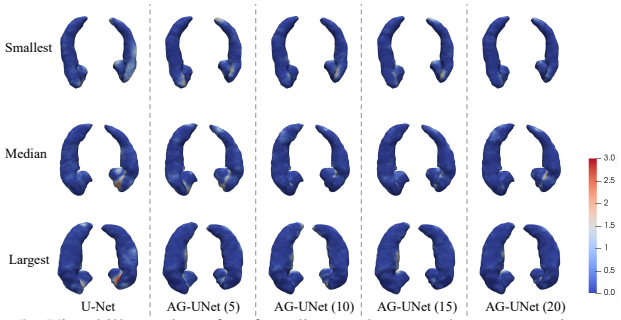

Fig. 5. Visual illustration of surface distance between the segmentation results of U-Net and AG-UNet (with different numbers of atlases) and ground truth on the smallest, median and largest *hippocampus* regions on ADNI dataset, respectively. The number 'a' in "AG-UNet (a)" denote the number of atlases.

of 0.08695 and the average surface distance of $0.447\ mm$. Notably, the ADNI dataset consists of 20 AD subjects, 20 MCI subjects, and 20 NC subjects. Due to the brain of AD patients are extreme atrophy, the volume of *hippocampus* on the ADNI dataset is in a wide range, *i.e.*, $[3000, 5300]$. In Fig. 5, we plot the segmented subject by U-Net and AG-UNet with different number of atlases in terms of average surface distance between the segmentation images and ground truth on the smallest, median and largest *hippocampus* regions. As shown in Fig. 5, our AG-UNet generally outperforms U-Net on the smallest, median and largest *hippocampus* regions. This implies that the proposed AG-UNet can better handle brains with large inter-subject morphological variances compared to its traditional counterpart (*i.e.*, U-Net) that does not use the anatomical information provided by the labeled atlases. From Fig. 5, one may also observe that AG-UNet achieves the best visual quality when the number of atlases is 20 on ADNI for *hippocampus* segmentation. The possible reason is that the multiple atlases could provide more anatomical information of the brain for ROI segmentation. These results indicate that using an appropriate number of atlases will increase the performance of our proposed method.

E. Influence of Deformable Registration

As reported in previous studies [31], [34], [35], the deformable registration could further improve the segmentation performance of multi-atlas based methods. In light of this, we also use the Diffeomorphic Demons method [41] to register atlas images onto the to-be-segmented image after linear registration. To investigate the influence of deformable registration on the performance of our method, we also perform the *hippocampus* segmentation on ADNI, with atlas images only linearly registered onto the to-be-segmented image by using FLIRT algorithm in [40] toolbox. In this work, we denote our AG-UNet approach with only the linearly registered atlases as AG-UNet-L, while AG-UNet employs atlases pre-processed by both linear and deformable registration algorithms. Besides, we also use the ANTs Symmetric Normalization (SyN) algorithm [39] for deformable registration after the linear registration, dubbed AG-UNET-A. Table IX shows the segmentation results achieved by three methods in *hippocampus* segmentation on the ADNI dataset.

TABLE IX
SEGMENTATION RESULTS OF THREE METHODS (*i.e.*, AG-UNET-L, AG-UNET-A, AND AG-UNET) IN *hippocampus* SEGMENTATION ON THE ADNI DATASET. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY.

Method	*DC*	*ASD* (mm)
AG-UNet-L	0.8712 ± 0.0228	0.448 ± 0.079
AG-UNet-A	0.8833 ± 0.0205	0.399 ± 0.059
AG-UNet	$\mathbf{0.8864 \pm 0.0212}$	$\mathbf{0.386 \pm 0.058}$

As can be observed from Table IX, compared to the AG-UNet-L, AG-UNet-A and AG-UNet achieve the improvement of 0.0121, and 0.0152 in terms of Dice coefficient. These results indicate that the deformable registration methods could further boost the segmentation performance of our proposed AG-UNet method. The possible reason for the improvement is that the deformable registration helps reduce the possible registration errors caused by the linear registration. Hence, the proposed network could be able to capture more precise local anatomical prior of brain structures by pooling operations, where the pooling operation is based on local brain patterns on registered atlases. For each brain MR image, the time costs of FLIRT, Diffeomorphic Demons and SyN are about 30 seconds, 120 seconds and 15 seconds, respectively.

F. Influence of Network Parameter

We now study the influence of the number of network parameters, by varying the number of channels in the proposed AG-UNet and its conventional counterpart (*i.e.*, U-Net). The experimental results achieved by these two methods in the task of *hippocampus* segmentation on the ADNI dataset are reported in Table X. It can be seen from Table X that, when using the same number of channels (*e.g.*, 16), the number of parameters in AG-UNet (*i.e.*, $0.77\,M$) is approximately twice that (*i.e.*, $0.39\,M$) of U-Net, and the segmentation results achieved by AG-UNet are better than those of U-Net in terms of both *DC* and *ASD* values. Even though U-Net using a larger number of parameters (*e.g.*, $6.20\,M$ with 64 channels), its performance is still worse than our AG-UNet with 16 channels. We also perform the Wilcoxon signed rank test on the results achieved by our AG-UNet and U-Net in terms of *DC* values. The proposed AG-UNet with 16 and 32 channels at the first layer achieve significant improvement over U-Net with 32 ($p = 3.3845e-04$) and 64 ($p = 0.0040$) channels at the first layer, respectively. These results suggest that the proposed anatomical attention subnetwork provides an efficient and flexible solution to boost the performance of conventional deep networks.

TABLE X
SEGMENTATION RESULTS OF U-NET AND AG-UNET WITH 16, 32, 64 CHANNELS (AT THE FIRST LAYER) ON ADNI DATASET FOR *hippocampus* SEGMENTATION. THE TERMS a AND b IN "$a \pm b$" DENOTE THE MEAN AND STANDARD DEVIATION FOR DIFFERENT SUBJECTS, RESPECTIVELY. M: MILLION.

Channel #	U-Net			AG-UNet (Ours)		
	DC	*ASD* (*mm*)	Parameter #	*DC*	*ASD* (*mm*)	Parameter #
16	0.8522 ± 0.0213	0.549 ± 0.085	$0.39\,M$	0.8818 ± 0.0220	0.405 ± 0.064	$0.77\,M$
32	0.8597 ± 0.0212	0.536 ± 0.071	$1.55\,M$	0.8864 ± 0.0212	0.386 ± 0.058	$3.06\,M$
64	0.8628 ± 0.0325	0.486 ± 0.076	$6.20\,M$	$\mathbf{0.8892 \pm 0.0221}$	$\mathbf{0.371 \pm 0.062}$	$12.24\,M$

G. Limitations and Future Work

There are still several limitations in the current work. *First*, the proposed network architecture consists of two subnetworks, which will increase the memory burden for segmentation. Hence, model compression is an important research direction for practical applications. *Second*, we treat each atlas equally without considering the similarity between the to-be-segmented brain MR image and each atlas image. In the future, we plan to learn anatomical prior knowledge from each atlas based on its similarity with the target image. *Third*, we only use the label maps of atlases for brain ROI segmentation, while the image intensity of multiple atlases may provide complementary information. Thus, as another future work, we plan to employ both the intensity information and label maps of atlases to further boost the segmentation performance. *Third*, we only evaluate our methods in the segmentation of brain MR images in the current work. In the future, we plan to validate our proposed method on other datasets. For example, the coronary artery images usually have a large variance in appearance (with potentially larger deformations) and sparse structures in the images/patches [53]. Hence, incorporating appropriate shape priors and cost-sensitive losses into our method is expected to improve the segmentation performance. *Finally*, several post-processing methods based on anatomical priors [54], [55] have been recently proposed for refining the segmentation results after the initial segmentation has been achieved, where the initial segmented errors can be corrected by using the additional refined segmentation step. Compared with these methods, our proposed framework can provide more accurate segmentation results in the initial step. To take advantage of our methods and post-processing methods, in the future, we plan to extend our proposed method by treating our generated results as the input of those post-processing methods to further boost the results.

VI. CONCLUSION

In this paper, we proposed an anatomical attention guided deep learning framework for ROI segmentation of brain MR images, including a segmentation subnetwork and an anatomical attention subnetwork. Specifically, the segmentation subnetwork is used to extract feature representations of brain MR images, while the anatomical attention subnetwork is employed to learn the anatomical prior from a set of registered atlases. We further introduce an anatomical gate to automatically fuse the feature maps generated by these subnetworks, to include not only the contextual information of to-be-segmented brain MR images, but also the anatomical prior of brain structures provided by multiple atlases. Within this framework, we develop two anatomical attention guided segmentation methods (*i.e.*, AG-FCN and AG-UNet) based on two different network architectures. Experiments on both the ADNI and the LONI-LPBA40 datasets suggest that our proposed AG-FCN and AG-UNet approaches can achieve superior results on ROI segmentation of brain MR images, compared with several state-of-the-art methods.

REFERENCES

[1] D. Zhang, Y. Wang, L. Zhou, H. Yuan, and D. Shen, "Multimodal classification of Alzheimer's disease and mild cognitive impairment," *NeuroImage*, vol. 55, no. 3, pp. 856–867, 2011.
[2] D. Devanand, G. Pradhaban, X. Liu, A. Khandji, S. De Santi, S. Segal, H. Rusinek, G. Pelton, L. Honig, R. Mayeux *et al.*, "Hippocampal and entorhinal atrophy in mild cognitive impairment prediction of Alzheimer disease," *Neurology*, vol. 68, no. 11, pp. 828–836, 2007.
[3] M. Liu, D. Zhang, and D. Shen, "Relationship induced multi-template learning for diagnosis of Alzheimer's disease and mild cognitive impairment," *IEEE Transactions on Medical Imaging*, vol. 35, no. 6, pp. 1463–1474, 2016.
[4] B. Jie, M. Liu, D. Zhang, and D. Shen, "Sub-network kernels for measuring similarity of brain connectivity networks in disease diagnosis," *IEEE Transactions on Image Processing*, vol. 27, no. 5, pp. 2340–2353, 2018.
[5] G. Karanikolas, G. B. Giannakis, K. Slavakis, and R. M. Leahy, "Multi-kernel based nonlinear models for connectivity identification of brain networks," in *Acoustics, Speech and Signal Processing (ICASSP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 6315–6319.
[6] M. Liu, J. Zhang, E. Adeli, and D. Shen, "Landmark-based deep multi-instance learning for brain disease diagnosis," *Medical Image Analysis*, vol. 43, pp. 157–168, 2018.
[7] Y. Chen, H. Gao, L. Cai, M. Shi, D. Shen, and S. Ji, "Voxel deconvolutional networks for 3D brain image labeling," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1226–1234.
[8] C. Lian, J. Zhang, M. Liu, X. Zong, S. C. Hung, W. Lin, and D. Shen, "Multi-channel multi-scale fully convolutional network for 3D perivascular spaces segmentation in 7T MR images," *Medical Image Analysis*, vol. 46, pp. 106–117, 2018.
[9] D. K. Iakovidis, S. V. Georgakopoulos, M. Vasilakakis, A. Koulaouzidis, and V. P. Plagianakos, "Detecting and locating gastrointestinal anomalies using deep learning and iterative cluster unification," *IEEE Transactions on Medical Imaging*, vol. PP, no. 99, pp. 2196–2210, 2018.
[10] H. Lin, H. Chen, S. Graham, Q. Dou, N. Rajpoot, and P.-A. Heng, "Fast ScanNet: Fast and dense analysis of multi-gigapixel whole-slide images for cancer metastasis detection," *IEEE Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1948–1958, 2019.
[11] C. Lian, M. Liu, J. Zhang, and D. Shen, "Hierarchical fully convolutional network for joint atrophy localization and Alzheimer's disease diagnosis using structural MRI," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018.
[12] J. Islam and Y. Zhang, "Brain MRI analysis for Alzheimer's disease diagnosis using an ensemble system of deep convolutional neural networks," *Brain informatics*, vol. 5, no. 2, p. 2, 2018.
[13] X. Artaechevarria, A. Munozbarrutia, and C. Ortizdesolorzano, "Combination strategies in multi-atlas image segmentation: Application to brain MR data," *IEEE Transactions on Medical Imaging*, vol. 28, no. 8, pp. 1266–1277, 2009.
[14] P. Coupe, J. V. Manjon, V. Fonov, J. C. Pruessner, M. Robles, and D. L. Collins, "Patch-based segmentation using expert priors: Application to hippocampus and ventricle segmentation," *NeuroImage*, vol. 54, no. 2, pp. 940–954, 2011.

[15] T. Tong, R. Wolz, P. Coupe, J. V. Hajnal, and D. Rueckert, "Segmentation of MR images via discriminative dictionary learning and sparse coding: Application to hippocampus labeling," *NeuroImage*, vol. 76, pp. 11–23, 2013.

[16] H. Wang, J. W. Suh, S. R. Das, J. Pluta, C. Craige, and P. A. Yushkevich, "Multi-atlas segmentation with joint label fusion," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 3, pp. 611–623, 2013.

[17] H. A. Kirisli, M. Schaap, S. Klein, L. A. Neefjes, A. C. Weustink, T. Van Walsum, and W. J. Niessen, "Fully automatic cardiac segmentation from 3D CTA data: A multi-atlas based approach," in *Medical Imaging 2010: Image Processing*, vol. 7623, 2010, p. 762305.

[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[19] O. Cicek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 424–432.

[20] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2017.

[21] R. A. Heckemann, J. V. Hajnal, P. Aljabar, D. Rueckert, and A. Hammers, "Automatic anatomical brain MRI segmentation combining label propagation and decision fusion," *NeuroImage*, vol. 33, no. 1, pp. 115–126, 2006.

[22] F. Rousseau, P. A. Habas, and C. Studholme, "A supervised patch-based approach for human brain labeling," *IEEE Transactions on Medical Imaging*, vol. 30, no. 10, pp. 1852–1862, 2011.

[23] D. Zhang, Q. Guo, G. Wu, and D. Shen, "Sparse patch-based label fusion for multi-atlas segmentation," in *International Workshop on Multimodal Brain Image Analysis*. Springer, 2012, pp. 94–102.

[24] G. Wu, M. Kim, G. Sanroma, Q. Wang, B. C. Munsell, and D. Shen, "Hierarchical multi-atlas label fusion with multi-scale feature representation and label-specific patch partition," *NeuroImage*, vol. 106, pp. 34–46, 2015.

[25] 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," *Medical Image Analysis*, vol. 18, no. 6, pp. 881–890, 2014.

[26] 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, 2009.

[27] W. Bai, W. Shi, D. Oregan, T. Tong, H. Wang, S. Jamilcopley, N. S. Peters, and D. Rueckert, "A probabilistic patch-based label fusion model for multi-atlas segmentation with registration refinement: Application to cardiac MR images," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1302–1315, 2013.

[28] T. R. Langerak, U. A. V. 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 Transactions on Medical Imaging*, vol. 29, no. 12, pp. 2000–2008, 2010.

[29] Y. Song, G. Wu, K. Bahrami, Q. Sun, and D. Shen, "Progressive multi-atlas label fusion by dictionary evolution," *Medical Image Analysis*, vol. 36, pp. 162–171, 2017.

[30] C. Zu, Z. Wang, D. Zhang, P. Liang, Y. Shi, D. Shen, and G. Wu, "Robust multi-atlas label propagation by deep sparse representation," *Pattern Recognition*, vol. 63, pp. 511–517, 2017.

[31] G. Sanroma, O. M. Benkarim, G. Piella, O. Camara, G. Wu, D. Shen, J. D. Gispert, J. L. Molinuevo, and M. A. G. Ballester, "Learning non-linear patch embeddings with neural networks for label fusion," *Medical Image Analysis*, vol. 44, pp. 143–155, 2018.

[32] L. Sun, W. Shao, M. Wang, D. Zhang, and M. Liu, "High-order feature learning for multi-atlas based label fusion: Application to brain segmentation with MRI," *IEEE Transactions on Image Processing*, pp. 1–1, 2019.

[33] L. Sun, C. Zu, W. Shao, J. Guang, D. Zhang, and M. Liu, "Reliability-based robust multi-atlas label fusion for brain MRI segmentation," *Artificial Intelligence in Medicine*, vol. 96, pp. 12–24, 2019.

[34] M. J. Cardoso, K. Leung, M. Modat, S. Keihaninejad, D. Cash, J. Barnes, N. C. Fox, and S. Ourselin, "STEPS: Similarity and truth estimation for propagated segmentations and its application to hippocampal segmentation and brain parcelation," *Medical Image Analysis*, vol. 17, no. 6, pp. 671–684, 2013.

[35] J. Huo, J. Wu, J. Cao, and G. Wang, "Supervoxel based method for multi-atlas segmentation of brain MR images," *NeuroImage*, vol. 175, pp. 201–214, 2018.

[36] L. Sun, D. Zhang, C. Lian, L. Wang, Z. Wu, W. Shao, W. Lin, D. Shen, G. Li, U. B. C. P. Consortium *et al.*, "Topological correction of infant white matter surfaces using anatomically constrained convolutional neural network," *NeuroImage*, vol. 198, pp. 114–124, 2019.

[37] L. Zhang, Q. Wang, Y. Gao, G. Wu, and D. Shen, "Automatic labeling of mr brain images by hierarchical learning of atlas forests," *Medical Physics*, vol. 43, no. 3, pp. 1175–1186, 2016.

[38] L. Zhang, Q. Wang, Y. Gao, H. Li, G. Wu, and D. Shen, "Concatenated spatially-localized random forests for hippocampus labeling in adult and infant MR brain images," *Neurocomputing*, vol. 229, pp. 3–12, 2017.

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

[40] S. M. Smith, M. Jenkinson, M. W. Woolrich, C. F. Beckmann, T. E. Behrens, H. Johansenberg, P. R. Bannister, M. De Luca, I. Drobnjak, D. E. Flitney *et al.*, "Advances in functional and structural MR image analysis and implementation as FSL," *NeuroImage*, vol. 23, pp. 208–219, 2004.

[41] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache, "Diffeomorphic demons: Efficient non-parametric image registration," *NeuroImage*, vol. 45, no. 1, pp. 61–72, 2009.

[42] A. Sotiras, C. Davatzikos, and N. Paragios, "Deformable medical image registration: A survey," *IEEE Transactions on Medical Imaging*, vol. 32, no. 7, pp. 1153–1190, 2013.

[43] H. Wang and P. Yushkevich, "Multi-atlas segmentation with joint label fusion and corrective learning-an open source implementation," *Frontiers in Neuroinformatics*, vol. 7, p. 27, 2013.

[44] C. R. Jack Jr, M. A. Bernstein, N. C. Fox, P. Thompson, G. Alexander, D. Harvey, B. Borowski, P. J. Britson, J. L. Whitwell, C. Ward *et al.*, "The Alzheimer's disease neuroimaging initiative (ADNI): MRI methods," *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 27, no. 4, pp. 685–691, 2008.

[45] D. W. Shattuck, M. Mirza, V. Adisetiyo, C. Hojatkashani, G. Salamon, K. L. Narr, R. A. Poldrack, R. M. Bilder, and A. W. Toga, "Construction of a 3D probabilistic atlas of human cortical structures," *NeuroImage*, vol. 39, no. 3, pp. 1064–1080, 2008.

[46] F. Shi, L. Wang, Y. Dai, J. H. Gilmore, W. Lin, and D. Shen, "Label: Pediatric brain extraction using learning-based meta-algorithm," *NeuroImage*, vol. 62, no. 3, pp. 1975–1986, 2012.

[47] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved N3 bias correction," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.

[48] A. Madabhushi and J. K. Udupa, "New methods of MR image intensity standardization via generalized scale," *Medical Physics*, vol. 33, no. 9, pp. 3426–3434, 2006.

[49] Y. Chen, H. Gao, L. Cai, M. Shi, D. Shen, and S. Ji, "Voxel deconvolutional networks for 3D brain image labeling," in *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2018, pp. 1226–1234.

[50] C. Wachinger, M. Reuter, and T. Klein, "DeepNAT: Deep convolutional neural network for segmenting neuroanatomy," *NeuroImage*, vol. 170, pp. 434–445, 2018.

[51] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[52] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz *et al.*, "Attention U-Net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.

[53] M. C. H. Lee, K. Petersen, N. Pawlowski, B. Glocker, and M. Schaap, "TETRIS: Template transformer networks for image segmentation with shape priors," *IEEE Transactions on Medical Imaging*, vol. 38, no. 11, pp. 2596–2606, 2019.

[54] N. Painchaud, Y. Skandarani, T. Judge, O. Bernard, A. Lalande, and P.-M. Jodoin, "Cardiac MRI segmentation with strong anatomical guarantees," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 632–640.

[55] A. J. Larrazabal, C. Martinez, and E. Ferrante, "Anatomical priors for image segmentation via post-processing with denoising autoencoders," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2019, pp. 585–593.