IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

SPECIAL SECTION ON DEEP LEARNING FOR COMPUTER-AIDED MEDICAL DIAGNOSIS

Received March 8, 2019, accepted March 17, 2019, date of publication March 26, 2019, date of current version April 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2906890

RP-Net: A 3D Convolutional Neural Network for Brain Segmentation From Magnetic Resonance Imaging

LIANSHENG WANG[1,2], (Member, IEEE), CONG XIE[2], AND NIANYIN ZENG[3]

[1]Fujian Key Laboratory of Sensing and Computing for Smart City, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
[2]Department of Computer Science, School of Information Science and Engineering, Xiamen University, Xiamen 361005, China
[3]Department of Instrumental and Electrical Engineering, Xiamen University, Xiamen 361005, China

Corresponding authors: Liansheng Wang (lswang@xmu.edu.cn) and Nianyin Zeng (zny@xmu.edu.cn)

This work was supported by the National Natural Science Foundation of China, under Grant 61671399.

ABSTRACT Quantitative analysis of brain volume is quite significant for the diagnosis of brain diseases. Accurate segmentation of essential brain tissues from 3D medical images is fundamental to quantitative brain analysis. Since manual segmentation is extremely tedious and time-consuming, there is a growing demand for automated segmentation. In this paper, we propose a 3D convolutional neural network including recursive residual blocks and a pyramid pooling module (RP-Net) for segmenting brain from 3D magnetic resonance (MR) images into white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF). RP-Net is an U-Net like the network that consists of a downsampling path and an upsampling path. Each path consists of four stages with a recursive residual block. All layers in RP-Net are implemented in a 3D manner. The pyramid pooling module is applied before a voxel-wise classification layer for obtaining both local and global context information. The RP-Net has been evaluated on WM, GM, and CSF segmentation with CANDI, IBSR18, and IBSR20 dataset. The experiments show that the RP-Net achieved mean dice similarity coefficients of 90.7% on CANDI, 90.49% on IBSR18 and 84.96% on IBSR20. The results demonstrate that our proposed method has achieved a significant improvement in segmentation accuracy compared to other reported approaches.

INDEX TERMS Brain segmentation, brain tissue, convolutional neural network, deep learning, magnetic resonance imaging.

I. INTRODUCTION

Magnetic resonance imaging (MRI) is a medical imaging technique which forms a vital part in clinical diagnosis. It provides high-resolution images for soft tissues and presents no known health hazards at a certain field strength and radio frequency power. One of the focuses of MRI technology is imaging brain. Brain MRI has been widely used to diagnose degenerative diseases such as Alzheimer's disease and dementia [1]–[3]. Besides, it helps to observe generalized changes of the brain such as specific changes associated with aging [4]. Segmentation of brain tissues has important applications in many brain nerve image analyses, such as analyzing anatomical structures, studying pathological regions, surgical planning, and visualization. Particularly, the accurate segmentation of white matter (WM), gray matter (GM), and cerebrospinal fluid (CSF) is the key to the diagnosis and quantitative measurement of various diseases.

However, accurate segmentation of brain tissues from MR images is a challenging task due to the fuzzy tissue edge, noise, image artifacts and the differences between individuals [5]. For the moment, the gold standard of this task is manual brain segmentation which is a tedious and time-consuming job for radiologists. Manual segmentation is based on vast clinical experience and requires outlining structures slice-by-slice. Therefore, automating the segmentation process is of great value to obtain results faster and reproducibly. Automated segmentation methods have developed rapidly during the recent years. MRI segmentation methods can be divided into the following categories.

Atlas-based segmentation methods. Atlas-based segmentation is used to segment the brain by registering the target image with an atlas [6]. Early approaches segment the brain by propagating the labels of a single manually labelled

The associate editor coordinating the review of this manuscript and approving it for publication was Yudong Zhang.

2169-3536 © 2019 IEEE. Translations and content mining are permitted for academic research only.
Personal use is also permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

atlas [7]. Multi-atlas segmentation methods is a hot topic of current research, which maps all labeled images onto the target image [8]. However, these methods can be computationally expensive due to complex nonrigid registration and atlas construction in which the iterative procedure is combined. Besides, registration errors can reduce the accuracy of label fusion results from multiple atlases.

Intensity-based segmentation methods. This kind of method classifies individual pixels/voxels based on their intensity [9] such as k-Nearest-Neighbors (kNN) classification method, Fuzzy C Means (FCM) clustering method, expectation-maximization (EM) clustering method and region growing method. Automatically trained kNN classification method [10] is achieved by non-rigidly registering the MR image with a tissue probability atlas to automatically select training samples. The EM algorithm [11] is an iterative method for combining statistical parameter estimation with the characteristics of MR images. Many studies [12]–[14] propose methods based on FCM for brain image segmentation. The region growing method [15] is based on similar intensities. However, the main limitation of these methods to become more accurate is that the intensity profiles of more detailed brain tissues overlap.

Deep learning-based segmentation methods. MRI segmentation based on deep learning has been given increasing attention due to its ability for self-learning and generalization. A convolutional neural network (CNN) is one of the most popular deep learning architecture and has shown outstanding results in many tasks, such as alcoholism identification [16], [17], abnormal breast identification [18], multiple sclerosis identification [19], [20], tea-category identification [21], and polarimetric synthetic aperture radar image segmentation [22]. The main advantage compared to classical approaches is that CNN learns abstract feature representations of images without prior knowledge and avoids human effort in feature design. It is a rapidly developing technology that also has found numerous applications in brain segmentation. Nie *et al.* [23] used a 2D fully convolutional network fusing multi-modality information to segment isointense phase brain MR images. Xu *et al.* [24] used a 3D-like fully convolutional network by regarding three successive slices of the volume as input to segment the middle slice. Chen *et al.* [25] proposed a deep voxelwise residual network for volumetric brain segmentation. Dolz *et al.* [26] proposed HyperDenseNet which is a 3D convolutional neural network applying complex combinations between modalities. Recent research shows that deep learning technology has great potential in the field of brain tissue segmentation.

In this paper, we propose a novel 3D Convolutional Neural Network referred as RP-Net for automatically segmenting the brain from 3D MR images into WM, GM, and CSF. All operations in the RP-Net are extended with 3D manner. 3D CNN can fully incorporate volumetric information. The overall framework of RP-Net is based on 3D U-Net architecture, which consists of a downsampling path and an upsampling path. Besides, we introduce the pyramid pooling module and the recursive residual block in our network. The results demonstrate that the method obtains accurate segmentation on the CANDI, IBSR18 and IBSR20 dataset, and hence demonstrates its robustness to differences in acquisition protocol. Our main contributions can be summarized as follows:

1) Recursive residual blocks are employed in our network instead of the widely-used residual blocks. Every residual path in the recursive residual block is directly attached to the first convolutional layer, which helps to learn highly complex features.
2) At the end of our network, we add a pyramid pooling module with 3D operations. The pyramid pooling module can better extract volumetric context information and fuse multi-scale contextual feature representations.
3) In the testing phase, we adopt a two-stage strategy. This strategy allows the model to focus on the brain area and reduce the effect of excessive pixels around.
4) In order to optimize the learning process, we put auxiliary classifiers at each stage in the upsampling path. The auxiliary classifiers make the features of hidden layers more meaningful.

The rest of paper is structured as follows. In Section II, we describe the experimental datasets and data preprocessing, elaborate the detail of the proposed RP-Net. Section III presents the experiments and results followed by discussion in Section IV. Finally in Section V, conclusions are drawn.

II. MATERIAL AND METHODS

A. DATA ACQUISTION

The MR image data is composed by 141 T1-weighted scans from two datasets.

1) CANDI

The CANDI dataset consists of T1-weighted MRI brain scans from 103 subjects and corresponding labels of 39 structures segmented by expert manually. The dataset is publicly available from the Child and Adolescent NeuroDevelopment Initiative (CANDI) at UMass Medical School [27], [28] (https://www.nitrc.org/projects/candi_share). The data originates from four diagnostic groups: Healthy Controls, Schizophrenia Spectrum, Bipolar Disorder with Psychosis, and Bipolar Disorder without Psychosis. All volumes have a size ranging from $256 \times 256 \times 128$ to $256 \times 256 \times 158$ voxels, with voxel spacing of $0.9mm \times 0.9mm \times 1.5mm$. We obtained four-class labeled volumes by assigning each of the 39 structures into GM, WM, CSF, and background (see Table. 1).

2) IBSR

The IBSR dataset consists of two sets of scans. The first set of scans is commonly known in the literature as IBSR20 while the second is known as IBSR18. The dataset is publicly available from the Internet Brain Segmentation Repository (IBSR) [29] (https://www.nitrc.org/projects/ibsr).

IBSR20 consists of 20 low resolution $3.1mm$ T1-weighted scans. All volumes have a size ranging from $256 \times 256 \times 60$ to $256 \times 256 \times 65$ voxels, with voxel spacing of

TABLE 1. Label conversion rules for CANDI dataset.

Original Class Label	Original Category	Conversion Class Label	Conversion Category
2	Left-Cerebral-White-Matter	3	WM
3	Left-Cerebral-Cortex	2	GM
4	Left-Lateral-Ventricle	1	CSF
5	Left-Inf-Lat-Vent	1	CSF
7	Left-Cerebellum-White-Matter	3	WM
8	Left-Cerebellum-Cortex	2	GM
10	Left-Thalamus-Proper	2	GM
11	Left-Caudate	2	GM
12	Left-Putamen	2	GM
13	Left-Pallidum	2	GM
14	3rd-Ventricle	1	CSF
15	4th-Ventricle	1	CSF
16	Brain-Stem	3	WM
17	Left-Hippocampus	2	GM
18	Left-Amygdala	2	GM
24	CSF	1	CSF
26	Left-Accumbens-area	2	GM
28	Left-VentralDC	2	GM
29	Left-undetermined	0	Background
30	Left-vessel	0	Background
41	Right-Cerebral-White-Matter	3	WM
42	Right-Cerebral-Cortex	2	GM
43	Right-Lateral-Ventricle	1	CSF
44	Right-Inf-Lat-Vent	1	CSF
46	Right-Cerebellum-White-Matter	3	WM
47	Right-Cerebellum-Cortex	2	GM
49	Right-Thalamus-Proper	2	GM
50	Right-Caudate	2	GM
51	Right-Putamen	2	GM
52	Right-Pallidum	2	GM
53	Right-Hippocampus	2	GM
54	Right-Amygdala	2	GM
58	Right-Accumbens-area	2	GM
60	Right-VentralDC	2	GM
61	Right-undetermined	0	Background
62	Right-vessel	0	Background
72	5th-Ventricle	0	Background
77	WM-hypointensities	3	WM
85	Optic-Chiasm	0	Background

$1mm \times 1mm \times 3.1mm$. The dataset provides segmentation label with main tissue for evaluation (GM, WM, CSF and background). Some of scans contain important acquisition artifacts and irregularities.

IBSR18 consists of 18 high resolution $1.5mm$ T1-weighted scans. All volumes have a size of $256 \times 256 \times 128$ voxels, with voxel spacing ranging from $0.8mm \times 0.8mm \times 1.5mm$ to$1.0mm \times 1.0mm \times 1.5mm$. We got file named segTRI_ana from IBSR_V2.0 skull-stripped NIFTI folder as ground truth, which segments the brain into four-class tissues, i.e, GM, WM, CSF, and background. IBSR18 scans present higher resolution and image quality than IBSR20, with no apparent acquisition artifacts that can bias the accuracy of some scans.

B. PREPROCESSING AND DATA AUGMENTATION

The voxel spacing of volumes in the above datasets is quite different. In this paper, we fix voxel spacing as $1mm \times 1mm \times 1.5mm$ by using linear interpolation firstly. Data normalization is performed to reduce the variations of input data. The method we used is min-max normalization which rescales the range of features to the range of [0, 1]. Moreover, we apply Contrast-Limited Adaptive Histogram Equalization (CLAHE) to increase the local contrast of all images.

Due to the absence of heavy training dataset, data augmentation is adopted. The augmentation operations include rotation, zoom, and flip. Specifically, we random rotation the original images by $-15, -10, -5, 0, 5, 10, 15$ degrees in the direction of the x and y axis. Besides, we flip the images horizontally and vertically and rescale images with a zoom range [0.5, 1.5] randomly.

C. 3D U-NET

U-Net [30] has extensively been used in medical image segmentation, which combines low-level and high-level information. Taking better advantage of volumetric data, 3D U-Net [31] extended the 2D U-Net architecture by replacing all 2D operations with their 3D counterparts.

The architecture of 3D U-Net in this paper is illustrated in Fig. 1. It consists of two parts, a downsampling path, and an upsampling path. The downsampling path contains four stages. Each stage consists of two $3 \times 3 \times 3$ convolution layer and each followed by a batch normalization (BN) and a rectified liner unit (ReLU). Batch normalization [32] reduces the amount by what the hidden unit values shift around (covariance shift), and it works well to avoid overfitting. At the end of each stage, a $2 \times 2 \times 2$ max pooling with strides of 2 is attached for down sampling. The number of feature channels is doubled after each stage. The upsampling path also contains four stages. Each stage includes an upsampling layer of $2 \times 2 \times 2$ kernel by strides of 2 followed by a convolution layer with BN and ReLU. Next is two $3 \times 3 \times 3$ convolution layer each followed by BN and ReLU. For reducing the impact of information loss, stages in the downsampling path are connected with stages that have the same resolution in the upsampling path. Different from the standard U-Net, we apply an element-wise addition instead of a concatenation operator for connection. In the last layer, a $1 \times 1 \times 1$ convolution with 'softmax' activation is used to produce the output.

D. RECURSIVE RESIDUAL BLOCK

ResNet [33] introduces residual connections to solve the problem of vanishing gradient. Deep neural networks with residual connections perform well on many challenging tasks. The residual connection is defined as

$$x_l = F_l(x_{l-1}) + h(x_{l-1}), \tag{1}$$

where x_{l-1} and x_l are the input and output of the l^{th} unit, $h(x)$ is an identity mapping where $h(x_{l-1}) = x_{l-1}$. $F_l(\cdot)$ denotes the residual function corresponding to l^{th} unit. The function contains two convolutional layers with activation function (BN and ReLU).

In a basic residual block, the activation function (BN and ReLU) is performed after the weight layer. A "pre-activation" structure [34] performs the activation function

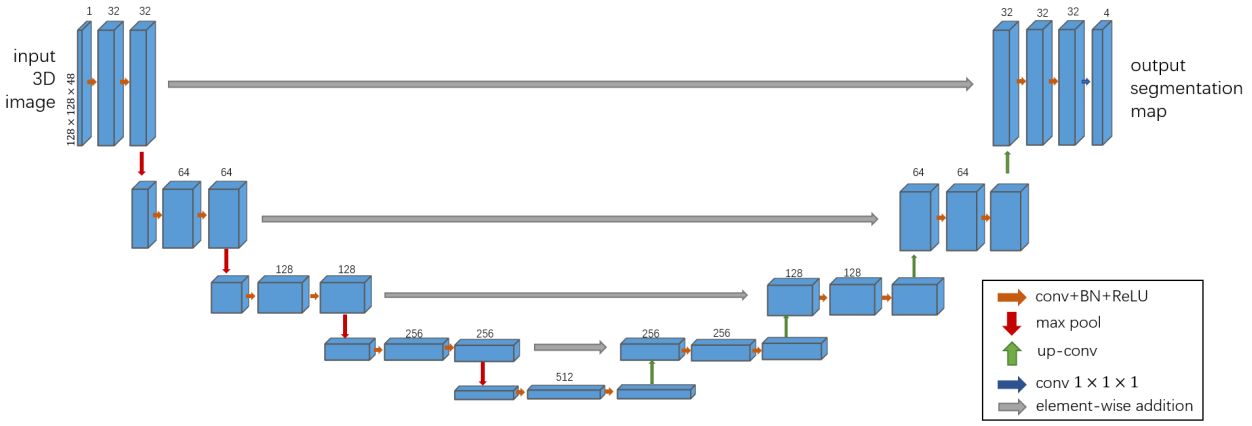


FIGURE 1. 3D U-Net architecture. Each blue box corresponds to a multi-channel feature map. The number of channels is denoted on top of the box. The arrows denote the different operations.

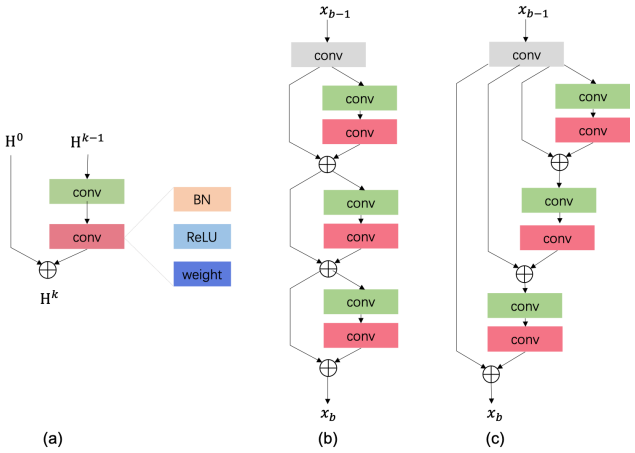


FIGURE 2. Structures of (a) the k^{th} residual unit, (b) a simple residual block with 3 residual units, and (c) a recursive residual block with 3 residual units.

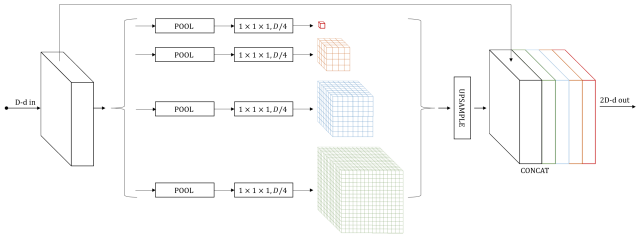


FIGURE 3. Structures of the pyramid pooling module.

before the weight layers and has achieved better results. We use the "pre-activation" structure in our residual unit.

Instead of directly using the residual block mentioned above, we used a recursive residual block that Tai *et al.* [35] proposed. The structure of the recursive residual block is illustrated in Fig. 2(c). The recursive residual block contains several residual units. A residual unit can be seen as Fig. 2(a). The residual paths help to learn highly complex features. The residual unit is formulated as

$$H^k = G(H^{k-1}) = F^k(H^{k-1}) + H^0, \tag{2}$$

where $k = 1, 2 \cdots, K$, K is the number of residual units in a recursive residual block, H^k is the output of the k^{th} residual unit. Especially, H^0 is the result of the first convolutional layer in the recursive residual block.

Thus, the output of the b^{th} recursive block can be define as

$$x_b = H_b^K = G^{(K)}(H_b^0) = G(G(\cdots(G(H_b^0))\cdots)), \tag{3}$$

where K-fold operations of G are performed.

E. PYRAMID POOLING MODULE

Pyramid pooling was first introduced to CNN in SPP-Net [36]. Pyramid pooling is for eliminating the fixed size constraints of convolutional neural networks without losing image location information while it is robust to object deformations. In PSPNet [37], pyramid pooling generates feature maps at different levels. The pyramid pooling module fuses these feature maps for obtaining both local and global context information.

The pyramid pooling module in our network has four pyramid levels as shown in Fig. 3. Instead of only using spatial information, we use volumetric information by performing 3D pooling operators. Bin sizes of pyramid pooling are $1 \times 1 \times 1$, $4 \times 4 \times 4$, $8 \times 8 \times 8$ and $16 \times 16 \times 16$. The different pyramid level divides the feature map into different sub-volumes and gets the pooled feature representation at different locations. A $1 \times 1 \times 1$ convolution layer is followed by each pyramid level to reduce the number of channels to $\frac{1}{4}$. An upsampling layer is used for rescaling the low-dimension feature to original feature map shape. Finally, different levels of features are concatenated as the final pyramid pooling global feature.

F. NETWORK ARCHITECTURE

Here we describe the details of our RP-Net architecture as illustrated in Fig. 4. Like 3D U-Net, the RP-Net has a downsampling path, and an upsampling path and each path contains four stages. In the downsampling path, each stage

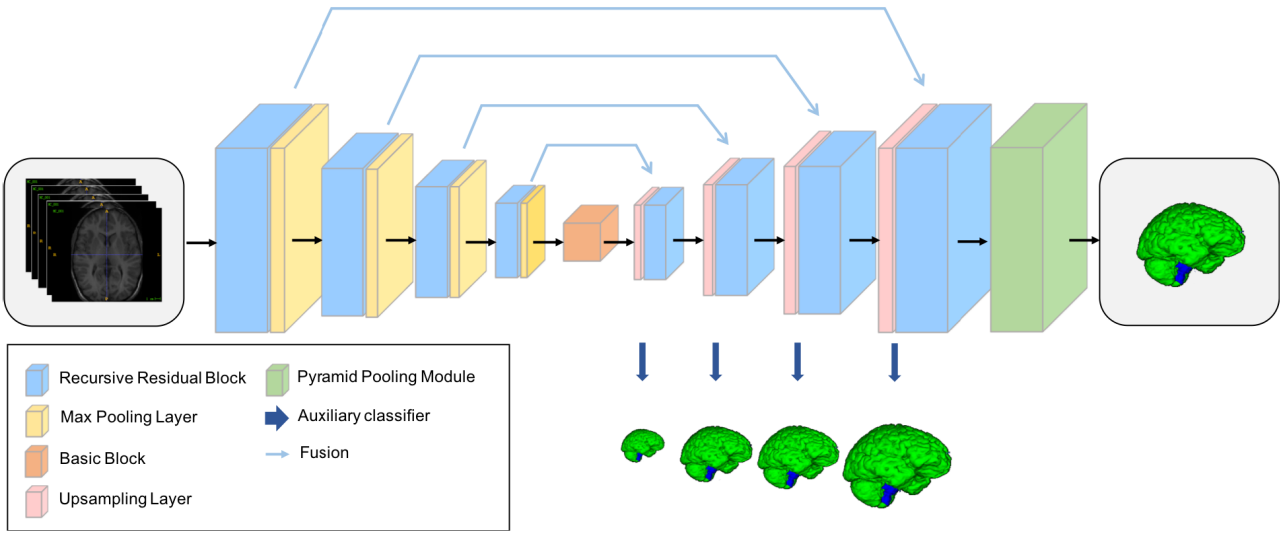


FIGURE 4. Architecture of the proposed RP-Net. The boxes and arrows denote the different operations. The orange box represents a basic block which consists of two convolution layers and each stacked by weight-BN-ReLU layers.

consists of a recursive residual block with three residual units as illustrated in Fig. 2(c) and a $2 \times 2 \times 2$ max pooling layer with strides of 2. In the upsampling path, each stage includes an upsampling layer with convolution layer and followed by a recursive residual block. At the end of the path, we use the pyramid pooling module shown in Fig. 3 to collect different levels of volumetric contextual information. All convolutional layers in RP-Net contain $3 \times 3 \times 3$ filter kernels. To reduce information loss by pooling operator, we implement skip connection to our network by connecting the corresponding recursive residual blocks with the same resolutions in the upsampling path and downsampling path. The element-wise addition is applied to feature map before feeding to the recursive residual block in the upsampling path. Note that all layers in our network including convolution, pooling, and upsampling layers are implemented in a 3D manner, so our network can take full advantages of volumetric information.

In addition, we introduce a deep supervision mechanism [38] in the RP-Net to accelerate the convergence speed and make the features of hidden layers more meaningful. We add auxiliary supervision to some hidden layers through auxiliary classifiers. Specifically, we use upsampling layer followed by the element-wise addition to get the feature map with the same size of the input, and then employ a $1 \times 1 \times 1$ convolution layer to obtain auxiliary coarse predictions. A $1 \times 1 \times 1$ convolution layer at the end of the network is to generate the main prediction. The purpose of the auxiliary loss function is to optimize the training process, but during testing, we abandon auxiliary branches and only use the well-optimized main branch for the final prediction.

G. IMPLEMENTATION DETAILS

Our implementation is using Python based on the public platform Keras [39] which supporting 3D operations. All the training and experiments are conducted on a workstation equipped with one NVIDIA GTX 1080 Ti GPU. The RP-Net is randomly initialized under the default setting of Keras with no pretraining on any external dataset. To reduce overfitting, we adopt data augmentation as mentioned in Subsection II-B. The network is trained with Adam method with a batch size of 1 due to the limited physical memory on GPU cards. The learning rate is set at $1e^{-4}$ initially. We use categorical cross-entropy loss to train the master classifier and auxiliary classifiers. Of the four auxiliary loss, we set the balancing weight to 0.2. Due to the limited GPU memory, we randomly crop $128 \times 128 \times 48$ sub-volumes from every sample as input when training the network.

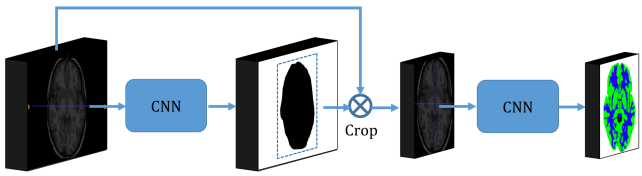


FIGURE 5. Overview of the testing phase.

In the testing phase, we only use the master branch for prediction. The testing phase goes through two stages shown in Fig. 5. We use overlapped sliding windows strategy in both two stages. In the first stage, the full image put into the network. The sub-volume size is $128 \times 128 \times 48$ and the stride is $120 \times 120 \times 40$.The output result is used to crop the coarse region of interest. In the second stage, CLAHE and normalization are performed again on the region of interest, so that the interference of pixels near the brain region can be avoided. The network and the strategy are as the same as the first stage, and the stride is smaller as $32 \times 32 \times 12$. The whole final volume prediction is provided by the average of the probability maps of the sub-volumes in the second stage.

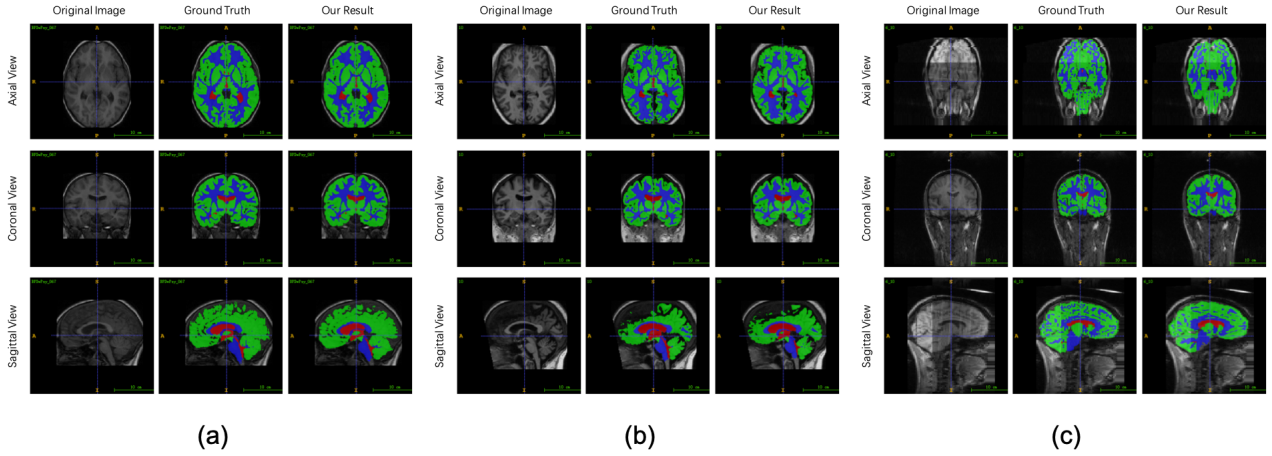


FIGURE 6. Brain segmentation results (green, blue, and red colors represent the GM, WM, and CSF, respectively) of (a) BPDwPsy_067 image on a axial slice (number 128), a coronal slice (number 64), and a sagittal slice (number 128) from CANDI, (b) 10 image on a axial slice (number 128), a coronal slice (number 64), and a sagittal slice (number 128) from IBSR18, (c) 6_10 image on a axial slice (number 128), a coronal slice (number 32), and a sagittal slice (number 128) from IBSR20.

III. EXPERIMENTS AND RESULTS

A. EVALUATION METRIC

The evaluation metric of the segmentation is dice similarity coefficient (DSC), which is calculated for each tissue type (i.e., GM, WM, and CSF). The DSC [40] measures the overlap between automatic segmentation and manual segmentation with range 0 (no overlap) to 1 (perfect agreement). It is defined as

$$DSC(A, M) = \frac{2\,|A \cap M|}{|A| + |M|} \times 100\%, \tag{4}$$

where A is the set of automatic segmentation results, M is the set of manual segmentation results.

In this paper, the results of automatic segmentation and manual segmentation all are a binary image, so the DSC between two binary images is defined as

$$DSC(A, M) = \frac{2\sum_i^N p_i q_i}{\sum_i^N p_i^2 + \sum_i^N q_i^2} \times 100\%, \tag{5}$$

where N is the total number of pixels in the image, p_i and q_i is the pixel of manual segmentation and automatic segmentation respectively.

B. EXPERIMENT DETAILS

The performance of RP-Net is evaluated on 141 images using CANDI and IBSR dataset. From CANDI dataset, 61 images are used for creating the training set, and 21 images are used for making the validation set. To make the test set more adequate, the remaining 21 images of CANDI dataset and all of IBSR20 and IBSR18 dataset are utilized to the test set. The train/validation/test set is summarized in Table. 2.

In our experiments, we perform brain-tissue segmentation by voxelwise classification. Every voxel is classified as either WM, GM, CSF, or background. We choose the model that achieves the best score of mean DSC on the validation set and evaluate it on the test set.

TABLE 2. The train/validation/test data summary.

set	dataset	# images	modality
train	CANDI	61	T1-w
validation	CANDI	21	T1-w
test	CANDI	21	T1-w
	IBSR18	18	T1-w
	IBSR20	20	T1-w

C. EVALUATION ON CANDI AND IBSR DATASET

Fig. 7 shows the average performance in terms of DSC. On CANDI dataset, the mean DSC for CSF, GM, and WM are 87.86%, 93.43%, and 90.81%, respectively. While RP-Net scores 88.18%, 92.08%, and 91.21% on IBSR18 dataset and 81.06%, 87.91% and 85.89% on IBSR20 dataset for the same classes. Quantitative analyses indicate high accuracy for all three tissues. As it can be observed, the mean DSC of CSF, GM and WM are close between IBSR18 and CANDI dataset but the performance on IBSR20 dataset is the worst among three datasets. Images of IBSR20 are more difficult to segment than images of IBSR18 because some of them have acquisition artifacts and irregularities. Fig. 6 shows examples of brain segmentation results and original images on three datasets. These results indicate that our method performs well in segmenting brain tissues in a visual view.

D. ABLATION ANALYSIS OF RP-NET ON IBSR DATASET

The RP-Net contains recursive residual blocks with three residual units, a pyramid pooling module and deep supervision mechanism. We perform a series experiments to evaluate the performance of these modules. The RP-Net without pyramid pooling module is regarded as $\text{CNN}_{\text{recur+ds}}$, and $\text{CNN}_{\text{recur}}$ defines as removing deep supervision on the basis of $\text{CNN}_{\text{recur+ds}}$. Besides, we replace the recursive residual block to the residual block as shown in Fig. 2(b) named CNN_{res}. Fig. 8 shows the result of the four networks.

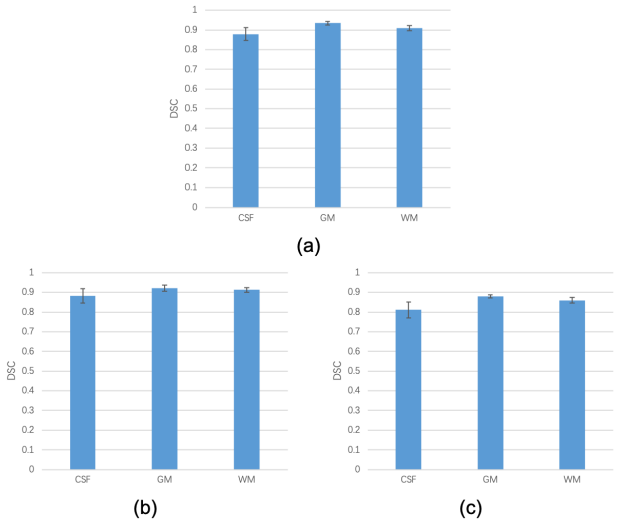


FIGURE 7. Bar plots of DSC (with standard deviation as error bar) of CSF, GM and WM using (a) CANDI dataset, (b) IBSR18 dataset, and (c) IBSR20 dataset.

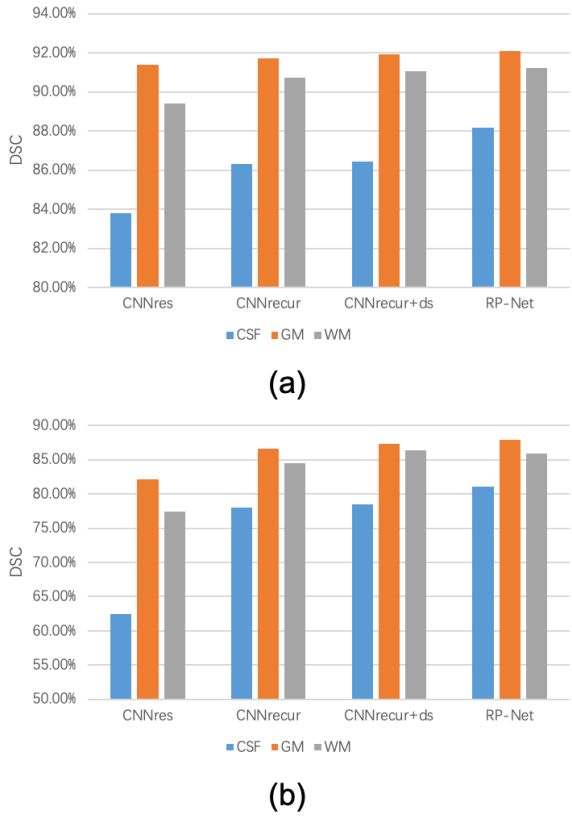


FIGURE 8. Bar plots of the DSC for CSF, GM and WM with four networks using (a) IBSR18 dataset and (b) IBSR20 dataset.

1) EFFECTIVENESS OF RECURSIVE RESIDUAL BLOCK

To analyze the effectiveness of recursive residual block, we compare the proposed method with and without the recursive residual block. Both CNN_{recur} and CNN_{res} are trained with the same training strategies. They both have U-Net structures. The difference is that CNN_{res} has stages with the residual block that the input of each unit is the output of previous unit and CNN_{recur} has stages with the recursive residual block that the input of each unit is the result of the first convolution layer on the block. As results shown in Table. 3, CNN_{recur} achieve better performance than CNN_{res}, with 1.38% and 9.03% improvement on the average DSC on IBSR18 and IBSR20 dataset respectively.

2) EFFECTIVENESS OF PYRAMID POOLING MODULE

To validate the effectiveness of the pyramid pooling module, we compare our network with and without the pyramid pooling module. As a result shown in Table. 3, our network with pyramid pooling module achieved 0.68% and 0.92% improvement on the average DSC on IBSR18 and IBSR20 respectively. PR-Net performs better on CSF segmentation with 1.73% and 2.54% improvement of mean DSC on IBSR18 and IBSR20 than $CNN_{recur+ds}$ while only drops by 0.44% on the mean DSC of WM on IBSR20.

3) EFFECTIVENESS OF DEEP SUPERVISION

We compare the performance of CNN with deep supervision or not. Specifically, $CNN_{recur+ds}$ has four auxiliary classifiers but CNN_{recur} has not. The average DSC of $CNN_{recur+ds}$ improve 0.23% on IBSR18 and 1.03% on IBSR20 than CNN_{recur}. And $CNN_{recur+ds}$ has 1.87% improvement on the mean DSC of WM. It shows different deep supervised layers' results in Fig. 9. We can see the layers from early to later can provide location information from coarse to fine.

E. COMPARISON WITH OTHER METHODS ON IBSR DATASET

IBSR dataset has been extensively used in numerous studies on brain structures segmentation. Table. 5 lists some of the studies using the IBSR dataset as segmentation benchmark and their results on the IBSR dataset. They use part of the data or all for evaluating the method they proposed. It can be seen that our results are quite competitive.

To validate the effectiveness and robustness of our method, we also compare our RP-Net with three segmentation methods 3D U-Net, 3D-like FCN [24] and VoxResNet [25] by

TABLE 3. DSC of CSF, GM, WM and the average of them for IBSR18 and IBSR20.

	IBSR18				IBSR20			
	CSF	GM	WM	Average	CSF	GM	WM	Average
CNN_{res}	83.79%	91.40%	89.40%	88.20%	62.41%	82.09%	77.45%	73.98%
CNN_{recur}	86.30%	91.72%	90.73%	89.58%	77.96%	86.62%	84.46%	83.01%
$CNN_{recur+ds}$	86.45%	91.93%	91.05%	89.81%	78.52%	87.26%	**86.33%**	84.04%
RP-Net	**88.18%**	**92.08%**	**91.21%**	**90.49%**	**81.06%**	**87.91%**	85.89%	**84.96%**

TABLE 4. DSC of CSF, GM, WM and the average of them for IBSR18 and IBSR20.

	IBSR18				IBSR20			
	CSF	GM	WM	Average	CSF	GM	WM	Average
ours	**88.18%**	**92.08%**	**91.21%**	**90.49%**	**81.06%**	**87.91%**	**85.89%**	**84.96%**
3D U-Net	84.99%	90.95%	89.97%	88.64%	68.74%	83.26%	80.94%	77.64%
3D-like FCN	85.61%	90.75%	90.04%	88.80%	77.61%	84.41%	83.49%	81.84%
VoxResNet	83.33%	88.17%	86.35%	85.95%	74.43%	81.98%	79.90%	78.77%

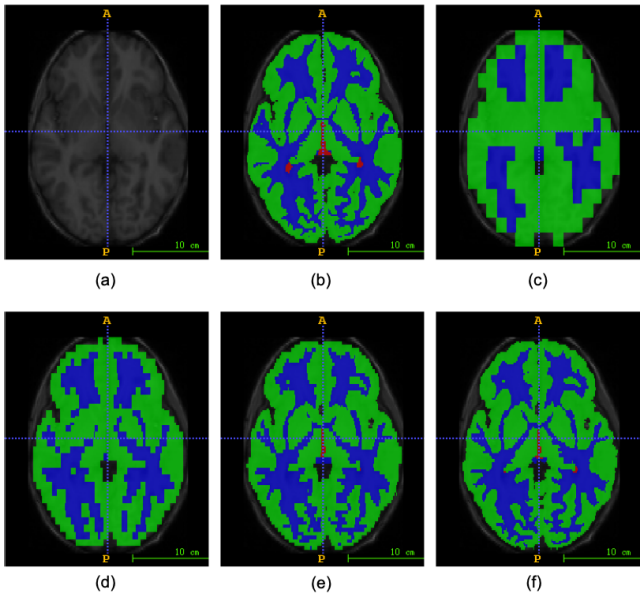


FIGURE 9. Different deep supervised layers' segmentation results (axial view). (a) is the original image, (b) is ground truth, (c) to (f) are the segmentation results of deep supervised layers respectively.

using the same training set and validation set. U-Net has widely used in medical image segmentation. The structure of 3D U-Net we compared is mentioned above. 3D-like FCN is a 2.5D CNN which use only a single modality and get the best performance in the MRBrains13 competition among the methods using single modality. VoxResNet is one of the principal methods in MRBrains13 competition. It is a 3D CNN integrating auto-context with multi-modality information. Since we only use single modality, we evaluate the architecture of VoxResNet by using T1-weighted scans only.

Table. 4 shows the comparison of the CSF, GM and WM segmentation performance on the IBSR18 and IBSR20 dataset. We can see that our method outperforms 3D U-Net, 3D-like FCN and VoxResNet on IBSR18, with 1.85%, 1.69% and 4.54% improvement on the average DSC respectively. Moreover, when evaluating on IBSR20, we achieve better performance with 7.32%, 3.12% and 6.19% improvement than the three methods.

IV. DISCUSSION

Brain tissue segmentation plays an important role in clinical diagnosis. While manual segmentation is tedious and time-consuming, automated segmentation is quite challenging because of the complex anatomical environment of the brain and the large variations of brain tissues. In this paper,

TABLE 5. Summary of brain tissues segmentation methods.

Article	Method	Dataset	Structures	DSC
Akselrod-Ballin et al. (2006) [41]	Atlas	IBSR20	CSF	51%
			GM	81%
			WM	80%
Tamajka and Benesova (2016) [42]	Supervoxels	IBSR18	CSF	67%
			GM	86%
			WM	85%
Meena Prakash and Shantha Selva Kumari (2016) [12]	Fuzzy C Means	IBSR18	CSF	32.42%
			GM	**92.6%**
			WM	87.22%
		IBSR20	GM	87%
			WM	83%
Kumar et al. (2018) [43]	U-SegNet	IBSR18	CSF	66.58%
			GM	90.33%
			WM	89.23%
Deng et al. (2018) [44]	modified U-Net	IBSR18	CSF	86.78%
			GM	91.85%
			WM	90.7%
ours	RP-Net	IBSR18	CSF	**88.18%**
			GM	92.08%
			WM	**91.21%**
		IBSR20	CSF	**81.06%**
			GM	**87.91%**
			WM	**85.89%**

we present an automatic brain segmentation method based on 3D CNN, named RP-Net. Our method segments the brain into WM, GM, and CSF. RP-Net can incorporate multi-scale features and fully explore volumetric context information. A portion of CANDI dataset is used for training. In the testing phase, MR images fed into the network twice. The first stage's result is used to crop the region of interest as the input of the second stage. It can avoid the interference of border pixels around the brain region. To show the generalization capability of our method, we evaluate our method on CANDI and IBSR dataset. The IBSR dataset contains IBSR18 and IBSR20, and IBSR20 is more challenge to be segmented. As shown in Fig. 7, the average DSC of CSF, GM, and WM is 90.7% on CANDI, 90.49% on IBSR18 and 84.96% on IBSR20. Examples of brain segmentation results of test data are shown in Fig. 6. Experimental results demonstrate that the RP-Net achieves high-quality segmentation of brain tissue.

RP-Net consists of recursive residual blocks, a pyramid pooling module, and deep supervision mechanism. In our experiments, we assess the effectiveness of these modules. To prove the influence of the recursive residual block, we give a comparison of CNN with recursive residual blocks and CNN with residual blocks. From Fig. 8, it can be observed that recursive residual blocks improve the network performance compared to residual blocks. The pyramid pooling module is extended by replacing 2D pooling operator with

3D pooling operator. Then it can obtain both local and global volumetric context information. After adding the pyramid pooling module, the performance of CSF segmentation achieves a great improvement (1.73% on IBSR18, 2.54% on IBSR20 of DSC). To demonstrate the effectiveness of deep supervision mechanism, we compare CNN with deep supervision or not. As shown in Table. 4, using the auxiliary loss can improve the score of DSC by 0.23% on IBSR18, 1.03% on IBSR20. Fig. 9 shows that it helps to generate more distinctive features of different level.

The segmentation accuracy of our method was compared to recently proposed methods for the task of brain segmentation such as 3D U-Net, 3D-like FCN, and VoxResNet. Statistics in Table. 4 show that RP-Net outperforms other methods using only T1-weighted modality. Our method yield a mean DSC value of 88.18%, 92.08% and 91.21% for CSF, GM and WM on IBSR18 dataset while 81.06%, 87.91% and 85.89% of CSF, GM and WM on IBSR20. This highlights the potential of the proposed automatic brain tissue segmentation architecture in clinical applications.

V. CONCLUSION

In this paper, we propose the RP-Net for brain tissue segmentation from 3D MR images. Our method adopts 3D fully convolutional architecture and it is effective for handling 3D MR images. The incorporation of the recursive residual block, pyramid pooling module, and deep supervision mechanism improve the network performance. We evaluate our method on both CANDI and IBSR. The result of IBSR20 indicates the robustness of our method. With a single-modality basis, our method excelled others and achieved the very competitive results on brain segmentation on the IBSR dataset.

REFERENCES

[1] G. B. Frisoni, N. C. Fox, C. R. Jack, Jr., P. Scheltens, and P. M. Thompson, "The clinical use of structural MRI in Alzheimer disease," *Nature Rev. Neurology*, vol. 6, no. 2, pp. 67–77, 2010.

[2] K. Trojachanec, I. Kitanovski, I. Dimitrovski, and S. Loshkovska, "Longitudinal brain MRI retrieval for Alzheimer's disease using different temporal information," *IEEE Access*, vol. 6, pp. 9703–9712, 2018.

[3] A. Giorgio and N. De Stefano, "Clinical use of brain volumetry," *J. Magn. Reson. Imag.*, vol. 37, no. 1, pp. 1–14, 2013.

[4] A. G. van der Kolk, J. Hendrikse, J. J. Zwanenburg, F. Visser, and P. R. Luijten, "Clinical applications of 7 T MRI in the brain," *Eur. J. Radiol.*, vol. 82, no. 5, pp. 708–718, 2013.

[5] C. N. Devi, A. Chandrasekharan, V. K. Sundararaman, and Z. C. Alex, "Neonatal brain MRI segmentation: A review," *Comput. Biol. Med.*, vol. 64, pp. 163–178, Sep. 2015.

[6] M. Cabezas, A. Oliver, X. Lladó, J. Freixenet, and M. B. Cuadra, "A review of atlas-based segmentation for magnetic resonance brain images," *Comput. Methods Programs Biomed.*, vol. 104, no. 3, pp. e158–e177, 2011.

[7] D. L. Collins, C. J. Holmes, T. M. Peters, and A. C. Evans, "Automatic 3-D model-based neuroanatomical segmentation," *Hum. Brain Mapping*, vol. 3, no. 3, pp. 190–208, 1995.

[8] 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.

[9] I. Despotović, B. Goossens, and W. Philips, "MRI segmentation of the human brain: Challenges, methods, and applications," *Comput. Math. Methods Med.*, vol. 2015, Oct. 2015, Art. no. 450341.

[10] H. A. Vrooman *et al.*, "Multi-spectral brain tissue segmentation using automatically trained k-Nearest-Neighbor classification," *NeuroImage*, vol. 37, no. 1, pp. 71–81, 2007.

[11] M. Rouaïnia, M. S. Medjram, and N. Doghmane, "Brain MRI segmentation and lesions detection by em algorithm," *Int. J. Med., Health, Biomed., Bioeng. Pharmaceutical Eng.*, vol. 2, no. 12, pp. 379–382, 2008.

[12] R. M. Prakash and R. S. S. Kumari, "Fuzzy C means integrated with spatial information and contrast enhancement for segmentation of MR brain images," *Int. J. Imag. Syst. Technol.*, vol. 26, no. 2, pp. 116–123, 2016.

[13] H. Huang, F. Meng, S. Zhou, F. Jiang, and G. Manogaran, "Brain image segmentation based on FCM clustering algorithm and rough set," *IEEE Access*, vol. 7, pp. 12386–12396, 2019.

[14] Q. Li *et al.*, "Glioma segmentation with a unified algorithm in multimodal mri images," *IEEE Access*, vol. 6, pp. 9543–9553, 2018.

[15] M. del Fresno, M. Vénere, and A. Clausse, "A combined region growing and deformable model method for extraction of closed surfaces in 3D CT and MRI scans," *Computerized Med. Imag. Graph.*, vol. 33, no. 5, pp. 369–376, 2009.

[16] S.-H. Wang, Y.-D. Lv, Y. Sui, S. Liu, S.-J. Wang, and Y.-D. Zhang, "Alcoholism detection by data augmentation and convolutional neural network with stochastic pooling," *J. Med. Syst.*, vol. 42, no. 1, p. 2, 2018.

[17] S.-H. Wang, K. Muhammad, J. Hong, A. K. Sangaiah, and Y.-D. Zhang, "Alcoholism identification via convolutional neural network based on parametric ReLU, dropout, and batch normalization," *Neural Comput. Appl.*, pp. 1–16, Dec. 2018.

[18] Y.-D. Zhang, C. Pan, X. Chen, and F. Wang, "Abnormal breast identification by nine-layer convolutional neural network with parametric rectified linear unit and rank-based stochastic pooling," *J. Comput. Sci.*, vol. 27, pp. 57–68, Jul. 2018.

[19] Y.-D. Zhang, C. Pan, J. Sun, and C. Tang, "Multiple sclerosis identification by convolutional neural network with dropout and parametric ReLU," *J. Comput. Sci.*, vol. 28, pp. 1–10, Sep. 2018.

[20] S.-H. Wang *et al.*, "Multiple sclerosis identification by 14-layer convolutional neural network with batch normalization, dropout, and stochastic pooling," *Frontiers Neurosci.*, vol. 12, p. 818, Nov. 2018.

[21] Y.-D. Zhang, K. Muhammad, and C. Tang, "Twelve-layer deep convolutional neural network with stochastic pooling for tea category classification on GPU platform," *Multimedia Tools Appl.*, vol. 77, no. 17, pp. 22821–22839, 2018.

[22] S.-H. Wang, J. Sun, P. Phillips, G. Zhao, and Y.-D. Zhang, "Polarimetric synthetic aperture radar image segmentation by convolutional neural network using graphical processing units," *J. Real-Time Image Process.*, vol. 15, no. 3, pp. 631–642, 2018.

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

[24] Y. Xu, T. Géraud, and I. Bloch, "From neonatal to adult brain MR image segmentation in a few seconds using 3D-like fully convolutional network and transfer learning," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, 2017, pp. 4417–4421.

[25] H. Chen, Q. Dou, L. Yu, J. Qin, and P.-A. Heng, "VoxResNet: Deep voxelwise residual networks for brain segmentation from 3D MR images," *NeuroImage*, vol. 170, pp. 446–455, Apr. 2017.

[26] J. Dolz, K. Gopinath, J. Yuan, H. Lombaert, C. Desrosiers, and I. B. Ayed. (2018). "HyperDense-Net: A hyper-densely connected CNN for multi-modal image segmentation." [Online]. Available: https://arxiv org/abs/1804.02967

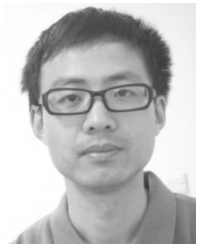
[27] D. N. Kennedy, C. Haselgrove, S. M. Hodge, P. S. Rane, N. Makris, and J. A. Frazier, "CANDIShare: A resource for pediatric neuroimaging data," *Neuroinformatics*, vol. 10, no. 3, pp. 319–322, 2012.

[28] J. A. Frazier *et al.*, "Diagnostic and sex effects on limbic volumes in early-onset bipolar disorder and schizophrenia," *Schizophrenia Bull.*, vol. 34, no. 1, pp. 37–46, 2007.

[29] T. Rohlfing, R. Brandt, R. Menzel, and C. R. Maurer, Jr., "Evaluation of atlas selection strategies for atlas-based image segmentation with application to confocal microscopy images of bee brains," *NeuroImage*, vol. 21, no. 4, pp. 1428–1442, Apr. 2004.

[30] 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.* Munich, Germany: Springer, 2015, pp. 234–241.

[31] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Athens, Greece: Springer, 2016, pp. 424–432.
[32] S. Ioffe and C. Szegedy. (2015). "Batch normalization: Accelerating deep network training by reducing internal covariate shift." [Online]. Available: https://arxiv.org/abs/1502.03167
[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 770–778.
[34] K. He, X. Zhang, S. Ren, and J. Sun, "Identity mappings in deep residual networks," in *Proc. Eur. Conf. Comput. Vis.* Amsterdam, The Netherlands: Springer, 2016, pp. 630–645.
[35] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 3147–3155.
[36] K. He, X. Zhang, S. Ren, and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," in *Proc. Eur. Conf. Comput. Vis.* Zürich, Switzerland: Springer, 2014, pp. 346–361.
[37] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. EEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.
[38] C.-Y. Lee, S. Xie, P. Gallagher, Z. Zhang, and Z. Tu. (2014). "Deeply-supervised nets." [Online]. Available: https://arxiv.org/abs/1409.5185
[39] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://keras.io
[40] C. Duan *et al.*, "A coupled level set framework for bladder wall segmentation with application to MR cystography," *IEEE Trans. Med. Imag.*, vol. 29, no. 3, pp. 903–915, Mar. 2010.
[41] A. Akselrod-Ballin, M. Galun, M. J. Gomori, R. Basri, and A. Brandt, "Atlas guided identification of brain structures by combining 3D segmentation and SVM classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assisted Intervent*. Copenhagen, Denmark: Springer, 2006, pp. 209–216.
[42] M. Tamajka and W. Benesova, "Automatic brain segmentation method based on supervoxels," in *Proc. IEEE Int. Conf. Syst., Signals Image Process. (IWSSIP)*, 2016, pp. 1–4.
[43] P. Kumar, P. Nagar, C. Arora, and A. Gupta. (2018). "U-SegNet: Fully convolutional neural network based automated brain tissue segmentation tool." [Online]. Available: https://arxiv.org/abs/1806.04429
[44] Y. Deng, Y. Sun, Y. Zhu, M. Zhu, and K. Yuan. (2018). "A strategy of MR brain tissue images' suggestive annotation based on modified U-Net." [Online]. Available: https://arxiv.org/abs/1807.07510



LIANSHENG WANG received the Ph.D. degree in computer science from The Chinese University of Hong Kong, in 2012. He is currently an Associate Professor with the Department of Computer Science, Xiamen University, Xiamen, China. His research interests include medical image processing and analysis.

CONG XIE received the B.S. degree from Xiamen University, Xiamen, China, in 2018, where she is currently pursuing the master's degree with the Department of Computer Science. Her research interests include medical image processing and machine learning.

NIANYIN ZENG received the B.Eng. degree in electrical engineering and automation, and the Ph.D. degree in electrical engineering from Fuzhou University, in 2008 and 2013, respectively. From 2012 to 2013, he was a Research Associate with the Department of Electrical and Electronic Engineering, The University of Hong Kong. From 2017 to 2018, he was an ISEF Fellow founded by the Korea Foundation for Advance Studies, and also a Visiting Professor with the Korea Advanced Institute of Science and Technology. He is currently an Associate Professor with the Department of Instrumental and Electrical Engineering, Xiamen University. His current research interests include intelligent data analysis, and computational intelligent and time-series modeling and applications.