Towards Clinical Diagnosis: Automated Stroke Lesion Segmentation on Multi-Spectral MR Image Using Convolutional Neural Network

ZHIYANG LIU [ID][1], (Member, IEEE), **CHEN CAO** [ID][2], **SHUXUE DING** [ID][1,3], (Member, IEEE), **ZHIANG LIU**[4], **TONG HAN**[2], AND **SHENG LIU**[5]

[1]Tianjin Key Laboratory of Optoelectronic Sensor and Sensing Network Technology, College of Electronic Information and Optical Engineering, Nankai University, Tianjin 300071, China
[2]Key Laboratory for Cerebral Artery and Neural Degeneration of Tianjin, Department of Medical Imaging, Tianjin Huanhu Hospital, Tianjin 300350, China
[3]School of Computer Science and Engineering, The University of Aizu, Fukushima 965-8580, Japan
[4]School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin 150006, China
[5]School of Basic Medical Sciences, Tianjin Medical University, Tianjin 300071, China

Corresponding author: Chen Cao (ccaomed@hotmail.com)

ABSTRACT The patient with ischemic stroke can benefit most from the earliest possible definitive diagnosis. While a quantitative evaluation of the stroke lesions on the magnetic resonance images (MRIs) is effective in clinical diagnosis, manually segmenting the stroke lesions is commonly used, which is, however, a tedious and time-consuming task. Therefore, how to segment the stroke lesions in a fully automated manner has recently extracted extensive attentions. Considering that the clinically acquired MRIs usually have thick slices, we propose a 2D-slice-based segmentation method. In particular, we use multi-spectral MRIs, i.e., diffusion weighted image, apparent diffusion coefficient, and T2-weighted image, as input, and propose a residual-structured fully convolutional network (Res-FCN). The proposed Res-FCN is trained and evaluated on a large data set with 212 clinically acquired MRIs, which achieves a mean dice coefficient of 0.645 with a mean number of false negative lesions of 1.515 per subject. The proposed Res-FCN is further evaluated on a public data set, i.e., ISLES2015-SISS, which presents a very competitive result among all 2D-slice-based segmentation methods.

INDEX TERMS Deep learning, stroke lesion segmentation, residual network (ResNet), convolutional neural network (CNN), fully convolutional network (FCN).

I. INTRODUCTION

Ischemic stroke is the most common cerebrovascular disease and one of the most frequent causes of death and disability worldwide [1]. A patient with ischemic stroke can benefit most from the earliest possible definitive diagnosis, and imaging plays an essential role in the assessment of patients [2]. Due to its excellent soft tissue contrast, the magnetic resonance imaging (MRI) has become the modality of choice for clinical evaluation of ischemic stroke lesions. For a quantitative analysis of stroke lesion in MRI images, the expert manual segmentation still serves as a common approach to compute the sizes, shapes and volumes of stroke lesions. However, it is a tedious and time consuming task and is non-reproducible. Therefore, the development of fully automated and accurate stroke lesion segmentation methods has become an active research field [3].

Conventionally, the lesion segmentation was treated as an abnormality detection problem, where a healthy atlas is established, and the lesions are detected according to the differences in tissue appearances [4]–[7]. The brain appearance, however, differs from patient to patient, and the lesions may also cause deformation in brain structure. Moreover, the MRIs acquired from different machines suffer from different levels of noise and deformations, which leads to incorrect detection and segmentation. Therefore, many machine-learning methods have been proposed, where

the features are learnt from massive training data, and high segmentation accuracy can be achieved. For instance, random-forest-based methods have been used in the literature [8]–[11], which present good performance in brain tumor segmentation by using hand-crafted features.

Note that the performance of the random-forests-based methods heavily relies on the manually annotated features. To achieve better performance, it is preferable to make the machine find the features from the data by itself. Deep learning is a machine-learning approach that uses artificial neural network with many hidden layers to extract features from data at progressively higher levels of abstraction [12]. In recent years, the deep-learning-based methods have been widely used in object classification and semantic segmentation thanks to its recent breakthrough in convolutional neural network (CNN).

The CNN was originally proposed for image classification tasks and has achieved 2.3% top-5 error in the classification of 1000 objects in Imagenet Large Scale Visual Recognition Challenge (ILSVRC) 2017. Thanks to its great potential in image analysis, great efforts have been made to apply the deep learning methods in medical image analysis [13]–[16]. Different from the images included in the ImageNet dataset which can be easily obtained from the Internet and annotated by the ordinary people, the medical images are much more difficult to be acquired, and many well-trained doctors are further required to annotate labels. Therefore, the dataset used in medical image analysis is usually much smaller than that used in the ordinary image classification tasks. For instance, the ImageNet dataset contains 1,431,167 images, while the brain tumor segmentation (BraTS) challenge 2015 has only 274 patients in the training dataset and 110 patients in the testing dataset [17]. In the sub-acute stroke lesion segmentation task of the ischemic stroke lesion segmentation (ISLES) 2015 challenge, the dataset is much smaller, with 28 patients in the training dataset and 36 patients in the testing dataset [18]. The insufficient data limits the ability to learn features from the training data, and the CNN may also become over-fitted to the training data. Despite of this, many deep-learning-based methods have been proposed for brain tumor and ischemic stroke segmentation tasks in the challenges [19]–[23], and presented good performance by using data augmentation techniques on the original data to enlarge the training dataset.

As the CNN is naturally a classifier, one of the most popular methods is to convert the image segmentation to a pixel-by-pixel classification task, and dedicated loss functions have been designed to overcome the huge class imbalance between the normal and abnormal tissue pixels [19], [22], [23]. To correctly classify a pixel, one need to include its surrounding area to provide sufficient contextual information for the classification task, which significantly increases the memory and computational costs. To perform semantic segmentations efficiently, [24] proposed a fully convolutional network (FCN) by replacing the fully-connected layers as convolutional layers. The FCN structure enables to segment

the whole image at once, and the memory- and computational costs can be significantly reduced. Inspired by [24], a 3D-Convolution-based FCN, known as DeepMedic, was proposed and won the ISLES 2015 and BraTS 2015 challenges [20], [21]. In ISLES 2015 sub-acute ischemic stroke lesion segmentation (SISS) dataset, it is able to detect sub-acute ischemic stroke lesions from 34 out of 36 patients in the testing dataset, and achieves a Dice coefficient of 0.59 on the test dataset [21]. In BraTS 2015 challenge, it achieves a Dice coefficient of 0.85 in segmenting the tumor tissues [20]. The promising results in ISLES and BraTS show the great potential of deep learning in the brain tissue segmentation tasks. Based on the high resolution data provided in the challenges, many deep-learning-based methods have been further proposed [25]–[27]. Note that the MRIs provided in the challenges, such as BraTS 2015 and ISLES 2015, are acquired for scientific usage with a high resolution of $1 \times 1 \times 1$mm per voxel, and the non-brain tissues, such as the skull, have been removed from the images in advance. In clinical practice, as it is very time-consuming to acquire the images with slices as thin as 1 mm, the clinically acquired images are usually with much thicker slices. For instance, the MRIs used in [28] has a slice thickness of 5mm, and a slice spacing of 1.5mm. Due to the thick slices, the 3D segmentation methods developed for the high resolution images with thin slices can hardly be applied directly on the clinically acquired images.

In clinical diagnosis, the diffusion weighted image (DWI) is one of the most commonly used MRI sequence to identify the ischemic stroke lesions, which measures the molecular motion of water in the brain tissues. Due to the decreased diffusion of the tissues with ischemia, the ischemic stroke lesions display as hyperintensive regions on DWIs. Recently, deep-learning-based methods have been applied to clinically acquired DWI data for acute ischemic stroke lesion segmentation. In [28], a network which consists of two Deconvolution Network (DeconvNet) [29] is developed and trained on a clinical dataset of 741 acute ischemic stroke patients. A multi-scale CNN is further developed to remove potential false positives. The mean dice coefficient, mean number of false positives and the mean number of false negatives achieved in [28] are 0.67, 3.27 and 4.07, respectively. While the former two are reasonable, the latter is relatively too large. When evaluating the performance of a segmentation method in clinical diagnosis, we should keep in mind that the false negative (FN) is much severer than the false positive (FP). It is very possible that the clinicians would pay much attention to the regions which are annotated as lesions by the segmentation algorithm, and the FPs can be possibly filtered by clinicians. The FNs, which are sometimes the lesions that are too subtle to be identified on the images, can be possibly ignored by clinicians, which may lead to severe misdiagnosis accident, especially if the misclassified lesion is the only lesion in the brain. This motivates us to study how to further reduce the FNs to develop a practical deep-learning based stroke lesion segmentation method for clinical diagnosis usage.
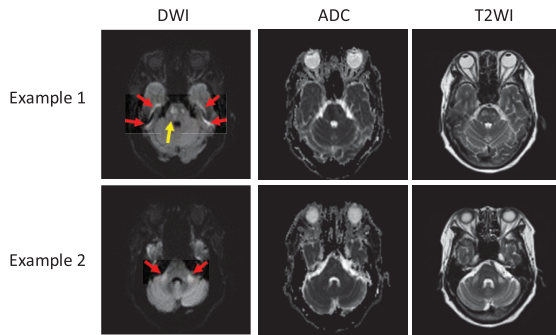
FIGURE 1. Challenge examples in ischemic stroke segmentation. The columns from left to right shows the DWI, ADC and T2WI. In example 1, the yellow arrow identifies the hyperintense that is a true acute ischemic stroke lesion, and the red arrows identify hyperintense due to magnetic susceptibility artifacts. In example 2, the red arrows identify hyperintense due to T2 shine-through effect. Best view in color.

Note that only DWI is not sufficient for the diagnosis of ischemic stroke. The hyperintensity on the DWI resulting from the magnetic susceptibility artifact and the T2 shine-through effect will be misclassified as stroke lesions if only DWI is used. For stroke diagnosis by radiologist, images with more acquisition parameters, such as apparent diffusion coefficient (ADC) map and T2-weighted image (T2WI) [30], are jointly considered. Fig. 1 presents two challenge examples that would be misclassified as a stroke lesion if only DWI is considered. In the first example, the hyperintensive region at the pons on DWI, which is identified by a yellow arrow, presents hypointensity on ADC map, and should be classified as an acute ischemia. On the other hand, the hyperintensive regions at the base of the brain, which are identified by the red arrows, present no abnormalities on ADC map and T2WI. Such hyperintensities are actually artifacts that caused by the magnetic field inhomogeneity due to the magnetic susceptibility differences between the brain tissue and air-containing areas of the skull, and should not be classified as ischemic stroke lesions [31]. The second example in Fig. 1 presents the artifacts due to T2 shine-through effect [32]. The lesions in the bilateral brachium pontis, which are identified by red arrows, appear hyperintense on DWI, ADC and T2WI. The hyperintensive regions on DWI are actually due to an increase in the T2 signal, rather than a decrease in diffusion, and should not be classified as stroke lesions. The examples shown in Fig. 1 indicate that the artifacts that appear hyperintense on DWI would probably be misclassified as stroke lesions. To correctly identify and segment the stroke lesions, multi-spectral MRIs, including DWI, ADC, and T2WI, should be jointly considered.

To further improve the performance of a CNN, a natural thought is to build the network deeper and wider, such that more features can be extracted by the convolution layers. However, the performance will become worse if we simply stack many convolution layers, as the weights of the deep convolution layers cannot be updated due to the gradient vanishing problem. Recently, a so-called residual network (ResNet) was proposed to make the network much deeper [33].
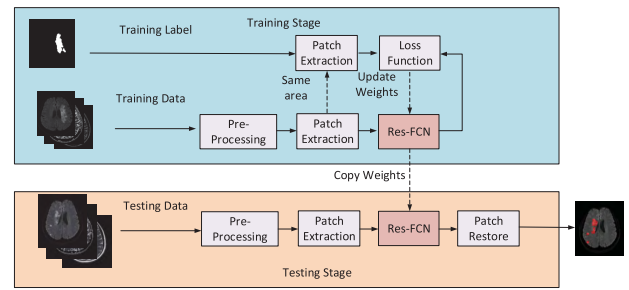

FIGURE 2. The whole pipeline of ischemic stroke lesion segmentation.

Instead of simply stacking more convolution layers, many skip connections are added between the layers. Such structure allows the gradient to pass backward through the skip connections, and all the convolution layers are able to be updated to extract features since the first training epoch. In this paper, we propose a residual-structured fully convolutional network (Res-FCN) for brain ischemic stroke lesion segmentation. Specifically, we collected the clinical data of 212 ischemic stroke patients from Nankai University affiliated Tianjin Huanhu Hospital, where 115 of them are used for training, and 97 of them for testing. By using mult-spectral MRI images, i.e., DWI, ADC and T2WI, the proposed Res-FCN is able to achieve a mean number of false negatives of 1.515 per subject on the testing dataset. We further evaluated the proposed method on a public dataset, i.e., ISLES2015-SISS, where it achieves very competitive results over all 2D segmentation methods.

II. METHOD

Fig. 2 shows the whole pipeline of our proposed residual-structured fully convolutional network (Res-FCN). The DWI, ADC and T2WI are concatenated and jointly used as input, and the output is a binary segmentation of stroke lesions. During the training process, the images and labels of patients in the training dataset are used to tune the network parameters of Res-FCN by using a gradient-based method, such that the difference between the predictions and the ground truth labels, which is measured by a loss function, is minimized. The details of our proposed method will be introduced in detail in the following subsections.

In the testing process, the images in the testing dataset, which have never been seen before by the neural network, are used to generate predictions. As the output for the i-th pixel $\hat{x}_i \in [0, 1]$ which can be interpreted as the probability that the i-th pixel is classified as lesion tissue, a threshold δ is used to generate the final binary segmentation, where the final binary output $\tilde{x}_i = 1$ if $\hat{x}_i \geq \delta$, and $\tilde{x}_i = 0$ otherwise. The predicted binary segmentation patches are finally repositioned to its corresponding position to generate the final output. The performance is then evaluated by comparing the prediction segmentations and the manually annotated ground truth.

A. PREPROCESSING AND PATCH EXTRACTION

In our method, each image slice is first normalized to zero mean and unit variance, and the DWI, ADC and T2WI images
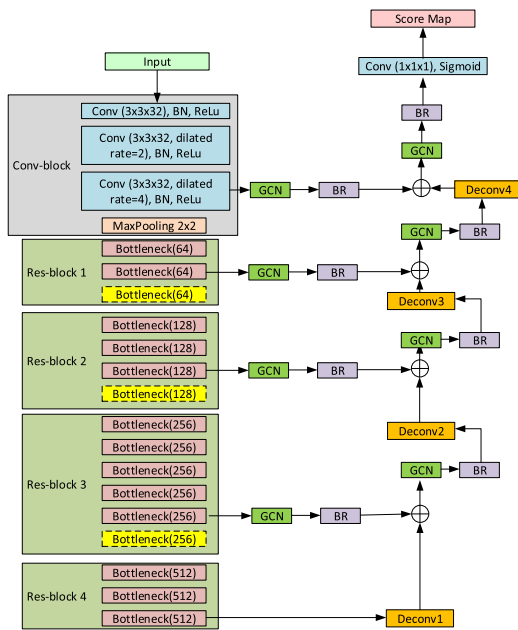
FIGURE 3. The architecture of the proposed network. The bottleneck block with downsampling, i.e., with convolution layers with stride 2, is highlighted in yellow. GCN: Global Convolutional Network. BR: Boundary Refinement. Conv: Convolutional layer. Deconv: Deconvolutional layer.

are concatenated to a 3-channel image. Note that the volumes of ischemic stroke lesions are typically much smaller than that of the normal brain tissues. If we use the whole slice as the input, the number of lesion tissue pixels would be much smaller than the number of normal tissue pixels, leading to a significant class imbalance. Therefore, we propose to use image patches, instead of the whole slice, as the input to train the network.

During the training process, we extract the patches by using a sliding window scheme with the window size $N \times N$, and a sliding step of $N/8$. To balance the number of the normal and lesion pixels, we only include the patches with lesions in the training dataset. With a relatively large N, a significant part of the normal tissue will also be included on each patch, and the features of normal tissues can still be learnt by the network. As each patient has a limited number of lesions, there is only a small number of patches available after patch extraction. To tune the massive number of network parameters, we perform data augmentation by horizontally flipping and randomly rotating the extracted patches.

In the testing process, as we have no prior information on the stroke lesions, we simply normalize the intensities of each image slice, and extract the patches using a sliding window of size $N \times N$ with a sliding step of $N/2$. After processed by the Res-FCN, for avoiding the zero padding influence, the predicted segmentation is cropped to $N/2 \times N/2$, and repositioned to its original position to generate the final segmentation result of the whole image.

B. ARCHITECTURE OF Res-FCN

The proposed network architecture is presented in Fig. 3. In the feature extraction stage, we propose to use the ResNet-50 structure as the base network. The input patches are initially processed by a stack of three atrous convolutional layers with dilation rates 1, 2 and 4, respectively. Then four Res-blocks are used to extract more features, where the Res-blocks are composed of 3,4,6 and 3 bottleneck blocks with the filters $n = 64, 128, 256$ and 512, respectively, as suggested in [33].

At the score map reconstruction stage, we use deconvolutional layers with kernel size 3×3 and stride 2 to upsample the feature maps. Inspired by U-NET [34], we propose to use the high resolution feature maps to assist reconstructing the score map. Global convolutional network (GCN) blocks and boundary refinement (BR) blocks [35] are used to refine the feature maps. The detailed structures and the effects of the essential parts in the proposed Res-FCN will be introduced in the following subsections.

1) CONVOLUTIONAL LAYERS

The basic building block to construct a CNN is the convolutional layer. Several layers can be stacked on top of each other. Each convolutional layer can be understood as extracting features from the preceding layer, and produces the feature maps as output.

Each feature map $\mathbf{O}_s$ is associated with one kernel. The feature map $\mathbf{O}_s$ is computed as

$$\mathbf{O}_s = \sum_r \mathbf{W}_s * \mathbf{X}_r + b_s, \qquad (1)$$

where $\mathbf{W}_s$ is the kernel and b_s is the bias term. $*$ is the convolution operator. $\mathbf{X}_r$ is the r-th channel of the input. For instance, for the first convolutional layer, the input is the stack of DWI, ADC and T2WI patches with a matrix size $N \times N \times 3$, and $\mathbf{X}_r$ represents the r-th channel of the original multi-spectral MRI, for $r = 1, 2, 3$. For the subsequent convolutional layers whose input is a $M \times M \times R$, $\mathbf{X}_r$ is the output of the r-th feature map of the preceding layer whose size is $M \times M$, for $r = 1, 2, \cdots, R$, where R is the number of feature maps of the preceding convolution layer.

During the training process, each convolutional layer is able to learn the features at different levels via the gradient-based method, such as the stochastic gradient descent, on a dedicated loss function related to the misclassification error, and the gradient for each layer can be computed by using the back-propagation (BP) algorithm [12].

Note that the field-of-view (FOV) of filters at a convolution layer is limited by the spatially small convolution kernels (typically 3×3), making it difficult to extract features from sufficiently large scale of spatial contextual information. If we use larger convolution kernels, the number of parameters to be tuned will grow exponentially, leading to prohibitively high computational complexity and memory cost. More importantly, a much larger number of training data will be required if we have more parameters, which is
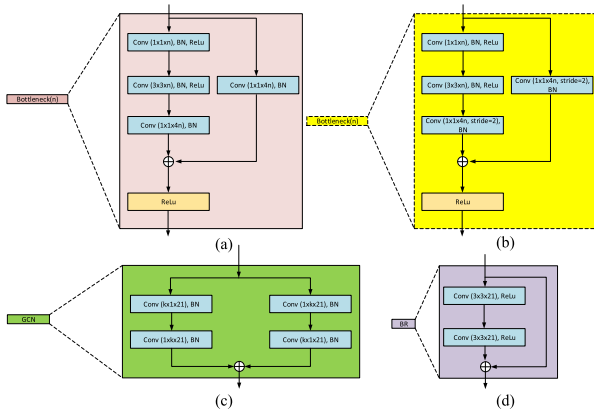
FIGURE 4. Illustration the details of (A) bottleneck block without downsampling, (B) bottleneck block with downsampling, (C) global convolution network (GCN) block, and (D) boundary refinement (BR) block.
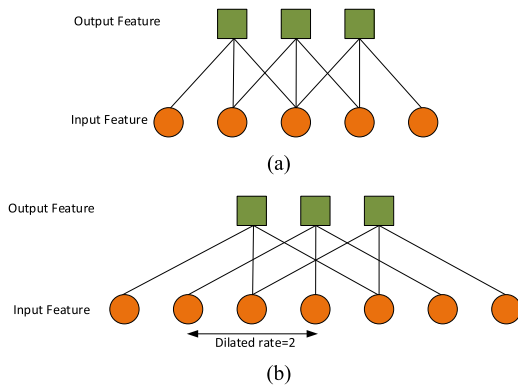

FIGURE 5. Graphic illustration of convolution operation. (A) Conventional convolution. (B) Atrous convolution with dilation rate $d = 2$.

impossible in biomedical image segmentation tasks, as the dataset is usually small. Therefore, we propose to use two methods to compensate such shortcoming.

First, inspired by DeepLab method [36], we propose to use a stack of atrous convolution (also known as dilated convolution) to extract multi-scale features from the inputs. The atrous convolution is operated as

$$O[i,j] = \sum_{s=-\lceil S/2 \rceil}^{\lfloor S/2 \rfloor} \sum_{t=-\lceil T/2 \rceil}^{\lfloor T/2 \rfloor} X_r[i+d \cdot s, j+d \cdot t] W[s,t], \quad (2)$$

where $O[i,j]$, $X[i,j]$ and $W[i,j]$ denote the (i, j)-th entry in the output feature map, input image, and the r-th convolutional kernel, respectively. d is the dilation rate. With $d = 1$, the atrous convolution converges to the conventional convolution operation in (1). Fig. 5 plots an visualized example of dilated convolution. As we can see, the atrous convolution can be interpreted as conventional convolution with a "hole" of size $d-1$ on each kernel, which enlarges the receptive field of a kernel without increasing the kernel size and the number of parameters. In our work, we propose to adopt a stack of atrous convolution layers with rate 1, 2 and 4 to extract multi-scale features from the original images.

Second, we adopt the GCN and BR blocks proposed in [35] to extract more contextual information. Instead of directly using larger kernel, the GCN block employs a combination of $1 \times k + k \times 1$ and $k \times 1 + 1 \times k$ subsequently to extract the features with receptive field of $k \times k$. Compared to the $k \times k$ kernel, the GCN structure reduces the number of parameters by a factor of $\frac{2}{k}$, which enables the probability to extract more features with considerably small number of parameters. Reference [35] justifies that the GCN structure shown in Fig. 4C is able to achieve better performance over that with a $k \times k$ kernel and that with a stack of small size kernels.

2) BOTTLENECK BLOCK

The bottleneck block is the basic block in a ResNet with more than 34 convolutional layers,[1] which are depicted in Fig. 4A and Fig. 4B. Specifically, we first use a 1×1 convolutional layer with n kernels as a bottleneck layer to reduce dimensionality of the features. Then we use a 3×3 convolutional layer with n kernels, and finally another 1×1 convolutional layer with $4n$ kernels to restore the depth. A skip connection with a 1×1 convolution with $4n$ kernels is used so that the input and output have the same size. Finally, the input of the block and the output of the final convolutional layer are added, and an activation is used after the summation. The number of filters are doubled every residual block, and at the same time their height and width are halved by using a convolutional layer with stride 2. A batch normalization (BN) layer is used after each convolution layer.

As shown in Fig. 4A, with the residual structure, the convolutional layer of each bottleneck block learns the residual part $h(\mathbf{X}) = f(\mathbf{X}) - \mathbf{X}$, where $\mathbf{X}$ and $f(\mathbf{X})$ denote the input and output feature maps, respectively. Thanks to the skip connections in the bottleneck blocks, the ResNet is able to efficiently update the deepest convolutional layers, i.e., the first convolution block shown in Fig. 3, from the beginning, and solves gradient vanishing problem in a dedicated way.

3) LOSS FUNCTION

The loss function is used to measure the error. It is also the function to be minimized during training. In this paper, we propose to use negative Dice coefficient as the loss function, which is defined as

$$f(\mathbf{p}) = -\frac{2\sum_{i \in patch} p_i g_i + \epsilon}{\sum_{i \in patch} p_i + \sum_{i \in patch} g_i + \epsilon}, \quad (3)$$

where p_i and g_i denote the segmentation results of the i-th pixel in the predicted probability map and the ground truth, respectively. $g_i = 1$ if the i-th pixel is labeled as stroke lesion, and $g_i = 0$ otherwise. $p_i \in [0, 1]$ can be interpreted as the probability of the i-th pixel being labeled as lesion. $\epsilon > 0$ is a small positive constant to avoid singularity of (3).

---

[1]In ResNet-34 or ResNet-18, a simple block is used [33]. In our method, we propose to build the network based on ResNet-50, and therefore a bottleneck block is used in this paper.

TABLE 1. Parameters used in MRI acquisition.

	Skyra		Trio		Avanto	
	DWI	T2WI	DWI	T2WI	DWI	T2WI
Repitition time (ms)	5200	3950	3100	4000	3800	4410
echo time (ms)	80	99	99	93	102	90
Flip angle ($^\circ$)	150	150	120	120	150	150
Number of excitations	1	1	3	1	3	1
Field of view (mm^2)	240×240	240×240	200×200	230×230	240×240	240×240
Matrix size	130×130	320×320	132×132	320×320	192×192	320×320
Slice thickness (mm)	5	5	6	6	5	5
Slice spacing (mm)	1.5	1.5	1.8	1.8	1.5	1.5
Number of slices	21	21	17	17	21	21

The reason we use (3) instead of categorical cross-entropy as loss function is that the numbers of normal and lesion pixels are significantly imbalanced even if we only extract the patches with lesions. With imbalanced classes, the cross-entropy loss function tends to category all pixels into the major class, i.e., the normal class in this paper, to minimize the loss function. The loss function in (3), however, tends to maximize the overlapping area and minimize the non-overlapping area, such that the class imbalance problem can be mitigated.

C. EVALUATIONS

We aim to study an automated ischemic lesion segmentation method towards clinical diagnosis. The dice coefficient (DC) is a commonly used measurement in image segmentation accuracy, which is defined as

$$DC = \frac{2|G \bigcap P|}{|G| + |P|}, \qquad (4)$$

where G and P denote the ground truth and the predicted segmentation, respectively. $|\cdot|$ denotes the area of lesion segmentations.

In clinical diagnosis, the segmentation of large and small stroke lesions are equally important. Therefore, we introduce the lesion-wise metrics to evaluate the performance. Specifically, we count the numbers of the false negative lesions and the false positive lesions at each image slice, and calculate the number of false negative and false positive lesions of each subject. The mean number of false negative lesions (m#FN) and the mean number of false positive lesions (m#FP) per subject are then calculated. These metrics are obtained using connected component analysis. A false positive lesion is defined as a 2D connected region in the predicted segmentation that has no overlap with the ground truth, while a false negative lesion is defined as a 2D connected region in the ground truth that has no overlap with the predicted segmentation.

III. EXPERIMENT RESULTS
A. DATA
In this study, 212 patients with ischemic stroke lesions were collected from Nankai University affiliated Tianjin Huanhu Hospital, where 62% of them are male, and the mean age is

56.21. All clinical images were collected from a retrospective database and anonymized prior to use. Ethical approval was granted by Tianjin Huanhu Hospital Medical Ethics Committee. MRI measurements were acquired from three MR scanners, with two 3T MR scanners (Skyra, Siemens and Trio, Siemens) and one 1.5T MR scanner (Avanto, Siemens). T2WI images were acquired using a fast spin-echo sequence. DWI images were acquired using a spin-echo type echo-planar (SE-EPI) sequence with b values of 0 and 1000 s/mm^2. The parameters are summarized in Tab. 1. Following acquisition, ADC maps were calculated from the diffusion scan raw data in a pixel-by-pixel manner as

$$ADC = \frac{\ln S_1 - \ln S_0}{b_1 - b_0}, \qquad (5)$$

where b characterizes the diffusion-sensitizing gradient pulses, with $b_1 = 1000$ s/mm^2 and $b_0 = 0$ s/mm^2 in our data. S_1 is the diffusion-weighted signal intensity with $b = 1000$s/mm^2. S_0 is the signal with no diffusion gradient applied, i.e., with $b = 0$ s/mm^2.

The T2WI, DWI and ADC images were copy referenced to ensure the same slice position so as to allow optimal image evaluation and measurement. The ischemic lesions were manually annotated by two experienced experts from Nankai University Affiliated Tianjin Huanhu Hospital. We randomly spit the whole dataset into a training set with 115 subjects, and a testing set with 97 subjects. The training set is used for training the network weights and tuning the hyperparameters, and the testing set is used for evaluating the performance only.

B. PREPROCESSING
As the images were acquired on three different machines, the matrix sizes varies. Therefore, we first resample all images to a pixel size of 1.77×1.77mm using linear interpolation, and crop the matrices to a uniform size of 128×128. Then we register the T2WI and ADC images according to the corresponding DWI images. The pixel intensity of each image slice is normalized into that of zero mean and unit variance. The patches are extracted as described in Sec. II, and the value N is set to be 64, which means that we extracted patches with size of 64×64, and the sliding window step to sample the patches is 8 pixels per step. The extracted patches are splitted into training set and validation set before data

TABLE 2. Network setting.

Layer	Filter size	Stride	Number of filters	Output shape
Input	-	-	-	$64 \times 64 \times 3$
Conv-block	3×3	1	32	$64 \times 64 \times 32$
Max Pooling	2×2	1	-	$32 \times 32 \times 32$
Res-block 1	-	-	$n = 64$	$16 \times 16 \times 256$
Res-block 2	-	-	$n = 128$	$8 \times 8 \times 512$
Res-block 3	-	-	$n = 256$	$4 \times 4 \times 1024$
Res-block 4	-	-	$n = 512$	$4 \times 4 \times 2048$
Deconv1	3×3	2	21	$8 \times 8 \times 21$
Deconv2	3×3	2	21	$16 \times 16 \times 21$
Deconv3	3×3	2	21	$32 \times 32 \times 21$
Deconv4	3×3	2	21	$64 \times 64 \times 21$
Conv	1×1	-	1	$64 \times 64 \times 1$

TABLE 3. Performance of our proposed Res-FCN with GCN kernel size $k = 9$, U-NET [34] and EDD Net [28]. The threshold δ is set to be 0.5. The bold number indicates the most significant performance.

		U-NET	EDD Net	Res-FCN($k = 9$)
DC	train	0.684	**0.812**	0.803
	test	0.541	0.626	**0.645**
m#FP	train	7.017	4.609	**4.504**
	test	7.866	5.031	**4.237**
m#FN	train	0.279	0.139	**0.130**
	test	1.918	1.753	**1.515**

augmentation, and a proportion of 0.1 patches with lesions are used for validation. The patches in the training set are augmented using the method described in Sec. II, and no data augmentation is performed for validation set.

C. SETUP

The hyperparameters of the proposed Res-FCN are shown in Tab. 2. All the parameters are initialized in the way as suggested in [33]. An ℓ_2 norm of the network parameters is added to the loss function with a kernel regularization factor 10^{-4}. We use the Adam method [37] with initial learning rate of 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$ as optimizer, and the learning rate is scaled down in a factor of $\sqrt{0.1}$ if no progress is made for 5 epochs in validation data. Without specifications, in this section, the threshold δ to generate binary segmentation is set to be 0.5.

D. RESULTS

The trained network is evaluated on the test dataset with 97 subjects. Fig. 6 plots some examples of segmentations where the lesions are located at cerebellum, cerebral hemisphere and basal ganglia. The fourth row shows an example with a small lesion at the cerebellar vermis. As Fig. 6 shows, our proposed method is sensitive to both large and small lesions. Tab. 3 summarizes the results on the training and testing dataset. For comparison, we also evaluate the results of U-Net [34] and EDD-Net [28] using the same data. Note that compared to the results reported in [28], the EDD-Net achieves worse results in DC and m#FP as we simply use a threshold $\delta = 0.5$, instead of a well-tuned value. As we will show in the following section, a higher DC can be achieved by adjusting the threshold δ.

Despite that our network is much deeper than U-Net and EDD-Net, the Res-FCN achieves the best results on the testing dataset thanks to the residual structure of the bottleneck block. As we can see from Tab. 3, the Res-FCN is able to achieve a mean number of FNs of 1.515.

E. IMPLEMENTATION AND RUNNING TIME

The experiments are performed on an Alienware Aurora R6 workstation with an Intel Core i7-7700K CPU, 48GB
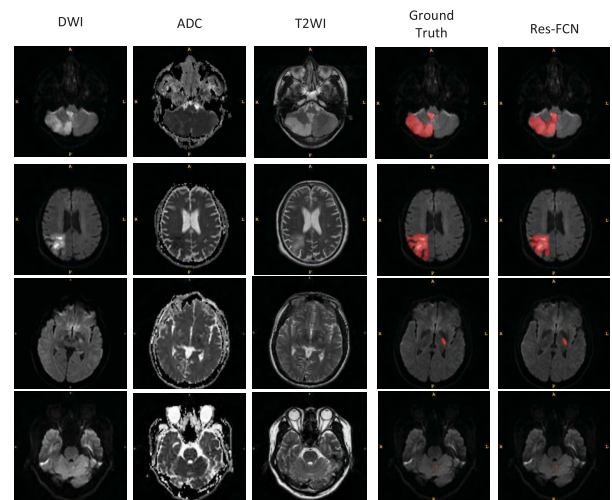

FIGURE 6. Examples of lesion segmentation. The first three columns show the original DWI, ADC, T2WI images, respectively. The fourth and the fifth columns show the manually annotated lesions and the segmentation results of Res-FCN, respectively. The segmentations are depicted on the DWI, and highlighted in red. Best view in color.

RAM and Nvidia GeForce 1080Ti GPU with 11GB memory. The workstation operates on Linux (Ubuntu 14.4) with CUDA 8.0. The network is implemented on Keras[2] with Tensorflow[3] backend. The MR image files are stored as Neuroimaging Informatics Technology Initiative (NIfTI) format, and processed using Simple Insight Toolkit (SimpleITK) [38]. The visualized results are presented by using ITK-SNAP [39].

During prediction, our method contains three stages: patch extraction, patch segmentation, and patch restoration. The number of parameters and the average running time of each stage are reported in Tab. 4. For the sake of comparison, we also present the numbers of parameters and the average running time of U-Net and EDD-Net. Despite that our proposed Res-FCN cost the most running time during prediction, it is able to finish segmentation in less than 1 second, which is sufficiently fast for diagnostic use.

---

[2]https://keras.io/

[3]https://www.tensorflow.org/

TABLE 4. Number of parameters and average running time of our proposed pipeline. The running time of U-Net and EDD-Net are also presented for comparison.

	U-NET	EDD Net	Res-FCN($k = 9$)
Number of Parameters	26.6M	3.22M	14.9M
Patch Extraction	0.12s	0.12s	0.12s
Patch Segmentation	0.46s	0.33s	0.57s
Patch Restoration	0.01s	0.01s	0.01s
Total Running Time	0.59s	0.46s	0.70s

TABLE 5. Performance of Res-FCN with a stack of three atrous convolution layers with kernel sizes 3 × 3 and dilation rate 1, 2 and 4, denoted as "atrous", and that with a stack of three conventional convolution layers with kernel size 3 × 3, denoted as "conv". The size of GCN kernels are 5, 7 and 9.

		$k = 5$		$k = 7$		$k = 9$	
		atrous	conv	atrous	conv	atrous	conv
DC	train	0.786	0.782	0.721	0.739	0.803	0.784
	test	0.600	0.589	0.623	0.568	0.645	0.613
m#FP	train	3.678	4.521	5.217	6.374	4.504	4.165
	test	5.928	8.340	4.515	7.505	4.237	4.536
m#FN	train	0.113	0.087	0.243	0.113	0.130	0.087
	test	1.546	1.629	1.619	1.690	1.515	1.660

IV. DISCUSSIONS

A. EFFECTS OF ATROUS CONVOLUTIONS AND GCN

In our proposed network, we use the atrous convolution layer stack instead of conventional convolution as the first convolutional layers. The proposed pyramid atrous convolution shows of paramount importance in extracting multi-scale spatial contextual information from the original images. To illustrate this, we compare the performance with atrous convolutions and conventional convolutions on our dataset, which are summarized in Tab. 5. The use of atrous convolution layer stack significantly increases the DC, and the m#FP and m#FN are greatly reduced, due to the fact that the atrous convolution has a larger field-of-view (FOV) of filters. In particular, with atrous convolutions, the FOVs are 3×3, 5×5 and 7×7, while with conventional convolutions, the FOVs of the three convolutional layers are all 3×3. With a larger FOV, the network is able to learn with more contextual information without increasing the kernel size.

Tab. 5 also summarizes the results with different GCN kernel sizes, i.e., $k = 5$, 7 and 9. The DC can be significantly improved by using larger GCN kernel, which is in accordance with the observation in [35]. Moreover, the mean number of FPs and FNs both decrease as the kernel size k increases, indicating that more contextual information is extremely helpful in improving the performance of lesion segmentation.

B. TRADEOFF BETWEEN FNs AND FPs

Note that the output of a CNN indicates the probability that a pixel should be labeled as lesion tissue, and a threshold δ is required to convert the probability score map to a binary

TABLE 6. Performance of Res-FCN, EDD-Net and U-Net on the testing dataset with the threshold $\delta = 1$ for U-Net and Res-FCN, and $\delta = 0.9$ for EDD-Net.

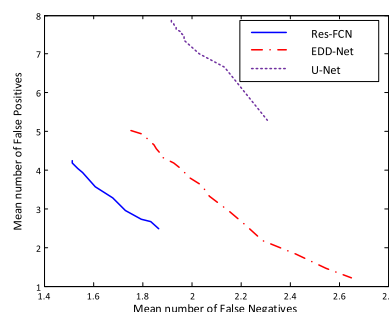|	U-Net	EDD-Net	Res-FCN($k = 9$)
DC	0.557	0.644	0.658
m#FP	5.289	2.588	2.485
m#FN	2.309	2.217	1.866

FIGURE 7. Mean number of false negatives versus mean number of false positives on the test dataset.

segmentation. The segmentation results presented previously are obtained by setting the threshold $\delta = 0.5$. In fact, the choice of the threshold can be interpreted as a tradeoff between FNs and FPs. Intuitively, as the threshold δ increases, fewer pixels will be classified as stroke lesion tissues, leading to an increasing m#FN and decreasing m#FP.

Fig. 7 plots the dependence between m#FN and m#FP of Res-FCN with GCN kernel size $k = 9$ on the test dataset. Each dot of the curve is plotted by using different values of the threshold δ, ranging from 0.5 to 1. The tradeoffs of EDD-Net and U-Net are also plotted for comparison. As the threshold δ increases, the m#FP reduces at the expense of a higher m#FN. As we can see from Fig. 7, the proposed Res-FCN presents the best FN-FP-tradeoff. In particular, for given value of m#FN, the Res-FCN has about 2 FPs less than EDD-Net, which highlights the outstanding performance of Res-FCN in clinical diagnosis.

Fig. 8 further plots the dependence between m#FN and the DC with different values of the threshold δ. As we can see from Fig. 8, the Res-FCN presents the best DC performance over EDD-Net and U-Net. The DCs of both U-Net and Res-FCN increases monotonically with the threshold δ. With $\delta = 1$, the Res-FCN achieves the highest DC of 0.658, at the expense that m#FN increases to 1.866. With EDD-Net, the DC is maximized at $\delta = 0.9$ according to our experiment, where the DC is 0.644 and the m#FN increases to 2.217. Compared to EDD-Net, our proposed Res-FCN is able to achieve higher segmentation accuracy with much fewer misclassifications.

Tab. 6 summarizes the performance on the testing dataset where the threshold δ is set to be the value that maximizes the DC. As we can see from Tab. 6, our proposed Res-FCN

TABLE 7. Performances on 2-fold cross validation of U-Net, EDD-Net and Res-FCN. 212 subjects are divided into two subset with 106 subjects each. In experiment 1, the networks are trained on Set 1 and tested on Set 2. In experiment 2, we exchange the roles of Set 1 and Set 2.

	U-Net			EDD-Net			Res-FCN($k=9$)		
	Exp 1	Exp 2	Average	Exp 1	Exp 2	Average	Exp 1	Exp 2	Average
DC	0.580	0.583	0.582	0.590	0.519	0.555	0.651	0.594	0.623
m#FP	6.387	5.547	5.967	2.424	4.311	2.368	3.698	3.113	2.906
m#FN	1.105	1.849	1.477	1.613	2.075	1.844	1.057	1.585	1.321
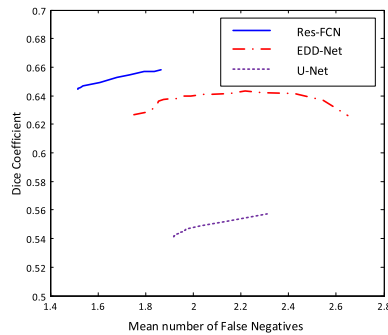
FIGURE 8. Mean number of false negatives versus mean dice coefficient on the test dataset.

TABLE 8. Evaluation results on ISLES2015-SSIS Testing Dataset on July 15, 2018.

Team	ASSD (mm)	DC	HD(mm)	Precision	Recall
kamnk1	7.87	0.59	39.61	0.68	0.60
zhanr6	7.52	0.58	38.89	0.60	0.68
fengc1	8.13	0.55	25.02	0.64	0.57
liulz2(ours)	9.21	0.55	51.04	0.61	0.60
monkf1	9.48	0.55	42.80	0.59	0.59

presents the highest DC and the lowest misdiagnosis. Compared to the results presented in [28], much lower numbers of false negatives and false positives are presented by using EDD-Net on our clinical dataset, due to the fact that we use mult-spectral images, i.e., DWI, ADC and T2WI, instead of single-spectral image, i.e., DWI, as input. Through the mutual corroboration of the different sequence, the detection of the stroke is more accurate, and the artifact can be avoided effectively. It highlights the importance to include mult-spectral MRI images in designing stroke lesion segmentation algorithms.

C. 2-FOLD CROSS VALIDATION

As our dataset is smaller compared to that commonly used for image segmentation, such as PASCAL VOC 2012, we apply a 2-fold cross-validation to evaluate its performance. In particular, we randomly split the whole dataset with 212 subjects into two complementary subsets, with 106 subjects in each set. We train the network on one subset and evaluate the performance on the other. In particular, in experiment 1, the network is trained on the subjects in Set 1, and evaluated on Set 2. In experiment 2, the roles of the two sets are exchanged. The performances on the test dataset of U-Net, EDD-Net, and the proposed Res-FCN are summarized in Tab. 7. As we can see, the proposed Res-FCN achieves the best performance, which highlights the importance in utilizing residual structure in lesion segmentation.

D. EVALUATION ON PUBLIC DATASET

To further demonstrate the performance of the proposed segmentation method, we evaluated Res-FCN on a public

dataset, i.e., ISLES2015-SISS. The ISLES2015-SISS dataset contains 28 subjects for training and 36 for testing. Each subject has four co-registered MRI sequences, i.e., T1 weighted image (T1WI), fluid attenuation inversion recovery (Flair), T2WI and DWI, which are skull-stripped and resampled to a resolution of $1 \times 1 \times 1$mm per voxel. The ground truth segmentations of the testing dataset are not released, and the evaluation results are automatically calculated after uploading the segmentations to the website.[4] The evaluation metrics of ISLES2015-SISS are average symmetric surface distance (ASSD), hausdorf distance (HD), dice coefficient (DC), precision and recall.

Since the data specification of SISS is much different from our dataset, we performed some adjustments to our method. First, the patches are extracted from three different views, i.e., axial, coronal, and sagittal views, so that the advantages of the thin slices can be utilized. Second, the patch size is set to be 128×128, and the patches are extracted by a sliding window with a sliding step of 16 in the training process. Third, we set the GCN kernel size as 11, and set the dilation rate of the first three convolution layers as 1, 4 and 8 to extract features from a larger receptive field. During training, only the patches with lesions are used, and the patches are randomly flipped along two axes. The model was trained for 500 epochs using Adam method with the same parameter as described in Sec. III-B. During the prediction process, the patches are extracted from axial, coronal, and sagittal views, and the probability score map is generated by taking average over the prediction results from three views. The binary segmentation results are finally generated by using a threshold $\delta = 0.5$.

Tab. 8 summarizes the top 5 ranking results on ISLES2015-SISS testing leaderboard on July 15, 2018. Our result is among the best result in terms of DC. We would like to

---

[4]https://www.virtualskeleton.ch/ISLES/Start2015

mention that the top ranking methods kamnk1 [20] and zhanr6 [27] use 3D patches to segment the lesions, which are more memory- and time-consuming. The method of fengc1, used conventional image analysis technique with hand-crafted features, instead of a deep-learning method [18]. From the evaluation results, we can then conclude that our proposed method achieves very competitive results over all 2D-patch-based deep-learning methods in ISLES2015-SISS dataset.

V. CONCLUSION

In this paper, we present a fully automated ischemia lesion segmentation method based on fully convolutional neural network. The DWI, ADC and T2WI are used as input, and a very deep CNN is built and efficiently trained thanks to the residual structure. The proposed Res-FCN network presents a high segmentation accuracy on the clinical MRI images with a dice coefficient of 0.645. More importantly, it presents very low false negatives, with a mean number of 1.515 per subject, which is of paramount importance in avoiding misdiagnosis in clinical scenario.

Note that there are several limitations in the present study. First, mult-spectral images have been demonstrated to provide helpful information for determining stroke ages in ischemic stroke [40]. Although our data includes both acute and subacute ischemic stroke lesions, we have not made a distinction between them. Second, our research did not include detection of hemorrhagic transformation, which is a complication of ischemic stroke. MRI-derived information about localization, timing and pathophysiology could improve decisions regarding acute management and secondary prevention. In the future, more MR images should be collected, including all types of stroke. The classification of stroke should be perfected, in order to guide the clinical treatment.

REFERENCES

[1] *Cause-Specific Mortality-Estimates for 2000–2012*, World Health Org., Geneva, Switzerland, 2012.

[2] S. L. Crichton, B. D. Bray, C. McKevitt, A. G. Rudd, and C. D. Wolfe, "Patient outcomes up to 15 years after stroke: Survival, disability, quality of life, cognition and mental health," *J. Neurol., Neurosurgery Psychiatry*, vol. 87, no. 10, pp. 1091–1098, 2016.

[3] E. A. Ashton, C. Takahashi, M. J. Berg, A. Goodman, S. Totterman, and S. Ekholm, "Accuracy and reproducibility of manual and semiautomated quantification of MS lesions by MRI," *J. Magn. Reson. Imag.*, vol. 17, no. 3, pp. 300–308, 2003.

[4] S. Doyle, F. Vasseur, M. Dojat, and F. Forbes, "Fully automatic brain tumor segmentation from multiple mr sequences using hidden Markov fields and variational EM," in *Proc. NCI-MICCAI BRATS*, vol. 1, 2013, pp. 18–22.

[5] A. Gooya, K. M. Pohl, M. Bilello, G. Biros, and C. Davatzikos, "Joint segmentation and deformable registration of brain scans guided by a tumor growth model," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, G. Fichtinger, A. Martel, and T. Peters, Eds. Berlin, Germany: Springer, 2011, pp. 532–540.

[6] P. Schmidt *et al.*, "An automated tool for detection of FLAIR-hyperintense white-matter lesions in multiple sclerosis," *NeuroImage*, vol. 59, no. 4, pp. 3774–3783, 2012.

[7] X. Liu, M. Niethammer, R. Kwitt, M. McCormick, and S. Aylward, "Low-rank to the rescue—Atlas-based analyses in the presence of pathologies," in *Medical Image Computing and Computer-Assisted Intervention— MICCAI*, P. Golland, N. Hata, C. Barillot, J. Hornegger, and R. Howe, Eds. Cham, Switzerland: Springer, 2014, pp. 97–104.

[8] A. Ellwaa *et al.*, "Brain tumor segmantation using random forest trained on iteratively selected patients," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, and H. Handels, Eds. Cham, Switzerland: Springer, 2016, pp. 129–137.

[9] L. Le Folgoc, A. V. Nori, S. Ancha, and A. Criminisi, "Lifted auto-context forests for brain tumour segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, and H. Handels, Eds. Cham, Switzerland: Springer, 2016, pp. 171–183.

[10] L. Lefkovits, S. Lefkovits, and L. Szilágyi, "Brain tumor segmentation with optimized random forest," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, and H. Handels, Eds. Cham, Switzerland: Springer, 2016, pp. 88–99.

[11] B. Song, C.-R. Chou, X. Chen, A. Huang, and M.-C. Liu, "Anatomy-guided brain tumor segmentation and classification," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, and H. Handels, Eds. Cham, Switzerland: Springer, 2016, pp. 162–170.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, May 2015.

[13] H.-C. Shin *et al.*, "Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1285–1298, May 2016.

[14] D. B. Larson, M. C. Chen, M. P. Lungren, S. S. Halabi, N. V. Stence, and C. P. Langlotz, "Performance of a deep-learning neural network model in assessing skeletal maturity on pediatric hand radiographs," *Radiology*, vol. 287, no. 1, pp. 313–322, 2018.

[15] D. S. Kermany *et al.*, "Identifying medical diagnoses and treatable diseases by image-based deep learning," *Cell*, vol. 172, no. 5, pp. 1122–1131, Feb. 2018.

[16] E.-J. Lee, Y.-H. Kim, N. Kim, and D.-W. Kang, "Deep into the brain: Artificial intelligence in stroke imaging," *J. Stroke*, vol. 19, no. 3, pp. 277–285, 2017.

[17] B. H. Menze *et al.*, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Oct. 2015.

[18] O. Maier *et al.*, "ISLES 2015—A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI," *Med. Image Anal.*, vol. 35, pp. 250–269, Jan. 2017.

[19] M. Havaei *et al.*, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.

[20] K. Kamnitsas *et al.*, "DeepMedic for brain tumor segmentation," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, and H. Handels, Eds. Cham, Switzerland: Springer, 2016, pp. 138–149.

[21] K. Kamnitsas *et al.*, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.

[22] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1240–1251, May 2016.

[23] R. S. Randhawa, A. Modi, P. Jain, and P. Warier, "Improving boundary classification for brain tumor segmentation and longitudinal disease progression," in *Brainlesion: Glioma, Multiple Sclerosis, Stroke and Traumatic Brain Injuries*, A. Crimi, B. Menze, O. Maier, M. Reyes, S. Winzeck, and H. Handels, Eds. Cham, Switzerland: Springer, 2016, pp. 65–74.

[24] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[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] S. Valverde *et al.*, "Improving automated multiple sclerosis lesion segmentation with a cascaded 3D convolutional neural network approach," *NeuroImage*, vol. 155, pp. 159–168, Jul. 2017.

[27] R. Zhang *et al.*, "Automatic segmentation of acute ischemic stroke from DWI using 3-D fully convolutional densenets," *IEEE Trans. Med. Imag.*, vol. 37, no. 9, pp. 2149–2160, Sep. 2018.

[28] L. Chen, P. Bentley, and D. Rueckert, "Fully automatic acute ischemic lesion segmentation in DWI using convolutional neural networks," *NeuroImage, Clin.*, vol. 15, pp. 633–643, Jun. 2017.

[29] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1520–1528.

[30] D. Graham, P. Cloke, and M. Vosper, *Principles and Applications of Radiological Physics*, 6th ed. Amsterdam, The Netherlands: Elsevier, 2011.

[31] J. Lu *et al.*, "Detectability and reproducibility of the olfactory fMRI signal under the influence of magnetic susceptibility artifacts in the primary olfactory cortex," *NeuroImage*, vol. 178, pp. 613–621, Sep. 2018.

[32] P. W. Schaefer, P. E. Grant, and R. G. Gonzalez, "Diffusion-weighted MR imaging of the brain," *Radiology*, vol. 217, no. 2, pp. 331–345, 2000.

[33] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.

[34] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*, N. Navab, J. Hornegger, W. Wells, and A. F. Frangi, Eds. Cham, Switzerland: Springer, 2015, pp. 234–241.

[35] C. Peng, X. Zhang, G. Yu, G. Luo, and J. Sun, "Large kernel matters— Improve semantic segmentation by global convolutional network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1743–1751.

[36] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille. (Jun. 2016). "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs." [Online]. Available: https://arxiv.org/abs/1606.00915

[37] D. P. Kingma and J. Ba. (Dec. 2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[38] B. C. Lowekamp, D. T. Chen, L. Ibáñez, and D. Blezek, "The design of SimpleITK," *Frontiers Neuroinform.*, vol. 7, p. 45, Dec. 2013.

[39] P. A. Yushkevich *et al.*, "User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability," *NeuroImage*, vol. 31, no. 3, pp. 1116–1128, Jul. 2006.

[40] R. G. Gonzalez, J. A. Hirsch, M. H. Lev, P. W. Schaefer, and L. H. Schwamm, Eds., *Acute Ischemic Stroke: Imaging and Intervention*. Berlin, Germany: Springer-Verlag, 2011.

SHUXUE DING (M'04) received the M.Sc. degree in physics from the Dalian University of Technology, Dalian, China, in 1988, and the Ph.D. degree in physics from the Tokyo Institute of Technology, Tokyo, Japan, in 1996. From 1989 to 1991 and from 1991 to 1992, respectively, he was an Assistant Professor and also an Associate Professor with the Dalian University of Technology. From 1996 to 1998, he was with Fujisoft-ABC Inc., Japan, where he was involved in algorithm design for telecommunication systems. From 1998 to 2003, he was with Clarion Co., Ltd, Japan, where he was engaged in researches in communication and signal processing, especially in speech recognition. From 2003 to 2005, he was a visiting faculty with the University of Aizu, Japan. He is currently a Professor with the School of Computer Science and Engineering, The University of Aizu. Since 2015, he has been with Nankai University, where he is a currently a Lecture Professor. He has been engaged in research in a wide range of areas of mathematical and physical engineering, such as statistical signal processing, optimization, neural computation, bio electromagnetism, and information sciences.

In particular, he has devoted himself to blind source separation and independent component analysis, and their applications in acoustic signals and vital signs. Recently, he is also conducting research in brain-style information processing, pattern recognition, compressive sensing, and sparse representation. He is also interested in speech and image processing, quantum computation, quantum and information, and the physics of information. He is a member of ASA, ACM, and IEICE.

ZHIANG LIU is currently pursuing the bachelor's degree with the School of Electrical Engineering and Automation, Harbin Institute of Technology, Harbin, China. His research interests include machine learning, signal processing, and electronic system design.

ZHIYANG LIU (S'13–M'15) received the bachelor's degree in communication engineering from Tianjin University, Tianjin, China, in 2010, and the Ph.D. degree in electronic engineering from City University of Hong Kong, Hong Kong, in 2014. Since 2014, he has been with Nankai University, where he is currently a Lecturer. His research interest includes wireless communication and deep learning.

TONG HAN received the master's and M.D. degrees in medical imaging from Tianjin Medical University, Tianjin, China, in 2003 and 2011, respectively. Since 1995, he has been with the Tianjin Huanhu Hospital, Tianjin. He is currently a Chief Physician with the Department of Medical Imaging, Tianjin Huanhu Hospital. Since 2014, he has been a Master Tutor with the School of Medical Imaging, Tianjin Medical University. His research interest includes the application of MRI in stroke diagnosis and treatment.

CHEN CAO received the bachelor's and master's degrees in clinical medicine (medical imaging) from Tianjin Medical University, Tianjin, China, in 2011 and 2013, respectively. Since 2013, he has been with Tianjin Huanhu Hospital, Tianjin, where he is currently a Physician. His research interest includes medical image analysis.

SHENG LIU received the master's degree in biomedical engineering and the Ph.D. degree in physiology from Milan University, Milan, Italy, in 2009 and 2013, respectively. Since 2013, he has been with the School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China, where he is currently a Lecturer. His research interest includes medical image analysis and cardiovascular disease.

• • •