

MI-UNet: Multi-inputs UNet Incorporating Brain Parcellation for Stroke Lesion Segmentation From T1-weighted Magnetic Resonance Images

Yue Zhang, Jiong Wu, Yilong Liu, Yifan Chen, *Member, IEEE*, Ed X. Wu, *Fellow, IEEE*, and Xiaoying Tang, *Member, IEEE*

Abstract—Stroke is a serious manifestation of various cerebrovascular diseases and one of the most dangerous diseases in the world today. Volume quantification and location detection of chronic stroke lesions provide vital biomarkers for stroke rehabilitation. Recently, deep learning has seen a rapid growth, with a great potential in segmenting medical images. In this work, unlike most deep learning-based segmentation methods utilizing only magnetic resonance (MR) images as the input, we propose and validate a novel stroke lesion segmentation approach named multi-inputs UNet (MI-UNet) that incorporates brain parcellation information, including gray matter (GM), white matter (WM) and lateral ventricle (LV). The brain parcellation is obtained from 3D diffeomorphic registration and is concatenated with the original MR image to form two-channel inputs to the subsequent MI-UNet. Effectiveness of the proposed pipeline is validated using a dataset consisting of 229 T1-weighted MR images. Experiments are conducted via a five-fold cross-validation. The proposed MI-UNet performed significantly better than UNet in both 2D and 3D settings. Our best results obtained by 3D MI-UNet has superior segmentation performance, as measured by the Dice score, Hausdorff distance, average symmetric surface distance, as well as precision, over other state-of-the-art methods.

Index Terms—Segmentation, stroke lesion, deep learning, diffeomorphic registration, brain parcellation.

I. INTRODUCTION

Cerebro-cardiovascular diseases are universal and dangerous, which may cause a high disability rate, morbidity rate, and huge economic burden. As the most quintessential manifestation of cerebrovascular diseases, stroke occurs if blood supply to the brain is interrupted or reduced and may damage the nervous system severely. Without prompt treatment, brain cells quickly die, leading to serious disability or death. Stroke is the main leading cause of deaths worldwide and approximately 795, 000 people experience a new or recurrent stroke each year in America alone [1]. Chronic stroke leads to long-term disabilities (such as paralyzation, mental disorders, language disorders, etc.), which severely affect the patients' life quality. These long-term disabilities can be effectively reduced through specialized stroke rehabilitation [2]. Neuroimaging can provide measures of brain structures, which benefit doctors in understanding

This study was supported by the Shenzhen Basic Research Program (JCYJ20190809120205578), the National Key R&D Program of China (2017YFC0112404) and the National Natural Science Foundation of China (81501546). The corresponding author is Xiaoying Tang.

Yue Zhang is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China and the Laboratory of Biomedical Imaging and Signal Processing and the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China.

Jiong Wu is with the School of Electronics and Information Technology, Sun Yat-sen University, Guangzhou, China.

Yilong Liu and Ed X. Wu are with the Laboratory of Biomedical Imaging and Signal Processing and the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, China.

Yifan Chen is with the School of Life Science and Technology, University of Electronic Science and Technology of China, Chengdu, China.

Xiaoying Tang is with the Department of Electrical and Electronic Engineering, Southern University of Science and Technology, Shenzhen, China. (e-mail: tangxy@sustech.edu.cn.)

post-stroke brain changes and thus predicting functional outcomes. High-resolution T1-weighted (T1w) anatomical magnetic resonance (MR) imaging is the mostly widely-employed noninvasive modality for examining structural brain changes after stroke. Overlaps between stroke lesions and brain regions can effectively map brain changes to behavioral performance [3]–[5], a prerequisite of which is to segment out the stroke lesions.

As such, precise segmentations are useful for quantifying stroke lesions and evaluating a patient's condition. Labels manually traced by professional radiologists are regarded as the gold standard [6]. However, fully-automated stroke lesion segmentation approaches are urged given that manual tracing is excessively labor-intensive. Automatic stroke lesion segmentation from T1w MR images is challenging. First, the location, size, and shape vary from lesion to lesion (see Fig. 1), which hinders fully-automatic segmentation. With different stroke severities, the lesion volumes vary from dozens to thousands of cubic millimeters, and the lesions can occur anywhere of the brain. Second, the boundaries of stroke lesions are generally ambiguous. Even manual tracing may be inaccurate and unreliable, having been reported to be of only 73% inter-rater Dice score on average [7].

In the past decade, various stroke lesion segmentation approaches have been proposed. These approaches can be approximately classified into two categories: model-based approaches and deep learning-based approaches. Model-based approaches mainly include the Clusterize algorithm, unsupervised fuzzy c-means, Gaussian naive Bayes, random forest, and super-voxel methods. The Clusterize algorithm consists of an automated preprocessing step, a manual cluster-delineation step, and a final modification step [8]. The manual cluster-delineation step is still labor-intensive and subjective. Seghier *et al.* then introduced an automated lesion identification toolbox which performed outlier detection using unsupervised fuzzy c-means and voxel comparisons between healthy and non-healthy tissues [9]. Griffiths *et al.* performed supervised Gaussian naive Bayes based classification using feature maps created by image algebraic operation and prior probability maps of tissue segmentation [10]. A lesion identification with neighborhood data analysis (LINDA) method has been proposed by Pustina *et al.*, taking advantages of feature detection and random forest classification [11]. Martins *et al.* employed a fully unsupervised super-voxel approach [12]. However, these methods may still have insufficient feature representation capacity because they depend on hand-crafted features.

Recently, deep learning approaches, especially convolutional neural networks (CNNs), have shown remarkable potential in medical image segmentation tasks [13]–[21]. CNN based segmentation approaches for stroke lesions can be roughly categorized into two classes, i.e., 2D based and 3D based.

In 2D CNN methods, a 3D volume is divided into multiple 2D slices and the stroke lesion is independently predicted for each slice [15], [16]. To capture long-range dependencies, Qi *et al.* proposed X-Net using depth-wise separable convolution and nonlocal blocks [15] and reported an average Dice of 48.67% on the anatomical tracing of lesions after stroke (ATLAS) dataset [7]. Yang *et al.* proposed a cross-level fusion and context inference network (CLCI-Net), making

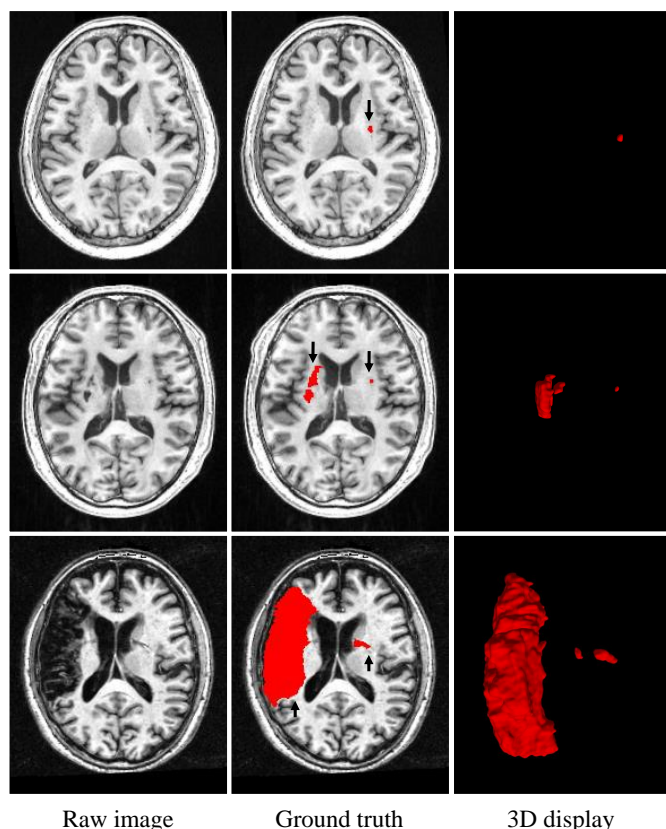


Fig. 1. Examples of T1w MR images showing large variations of stroke lesions in terms of shape, size, and location. Each row represents a different example. The red regions in the second and third columns denote the stroke lesions.

use of features at different scales to discriminate fine structures with similar intensity profiles and achieved an average Dice of 48.1% on a subset of the ATLAS dataset. In these slice-based methods, 3D MR images are converted to 2D slices at the beginning and the final 3D segmentation results are obtained by concatenating all 2D segmentation results. In such segmentation procedures, the 3D context information in MR image is ignored, and hence the predicted segmentation will have no inter-slice smoothness.

To capture characteristics of the volume data in the extra dimension, 3D CNN based methods have been proposed [18]–[22]. Kamnitsas *et al.* [22] employed an efficient dense training strategy by processing contiguous patches in one pass. They achieved an average Dice of 59% and ranked top on the ischemic stroke lesion segmentation (ISLES) 2015 challenge [23]. Zhang *et al.* proposed a deep 3D CNN equipped with dense connectivity to segment stroke lesions from diffusion-weighted MR image and achieved an average Dice of 58% on the ISLES 2015 dataset [24]. Zhou *et al.* proposed a dimension-fusion-UNet (D-UNet) by combining 2D and 3D convolutions at the encoding stage. They achieved an average Dice of 53.49% on a subset of the ATLAS dataset. Although 3D CNNs make the most use of spatial information in the volume data, the network sizes are nonetheless much larger than those of 2D CNNs [25].

It is worth being pointed out that some methods were developed using the multi-modal ISLES dataset [26], aiming at using multi-modal MR images to predict stroke lesions. However, multi-modal MR images are not typically used in routine clinics as it requires extra efforts to get the multi-modal data. On the contrary, T1w MR image is more typically used in clinical practice, as well as in stroke

rehabilitation research [7]. As such, there is a demand for developing fully-automatic stroke lesion segmentation approaches from T1w MR image.

Most stroke lesions are located at gray matter (GM) and white matter (WM) [27]–[30]. For example, Chen *et al.* showed that stroke occurs at GM given that GM has a high cerebral blood flow level due to its high metabolic demand [30]. Wang *et al.* nevertheless showed that the majority of damages caused by stroke are located in WM because WM has a low blood supply and is more susceptible to ischemia than other brain regions [29]. To the best of our knowledge, no methods have ever made use of such location prior information in stroke lesion segmentation tasks. To make use of this prior information, we propose a multi-inputs UNet (MI-UNet) stroke lesion segmentation framework and evaluate it on an open-source dataset consisting of 229 T1w MR images with manually delineated stroke lesions [7]. We first parcellate each T1w MR image into multiple whole-brain regions using the fast large deformation diffeomorphic metric mapping for image (LDDMM-image) algorithm [31]. Other than GM and WM, we also include lateral ventricle (LV) in the location prior because there are certain stroke lesions that are adjacent to the brain ventricles [32]. That brain parcellation is then sliced/sampled and concatenated with the sliced/sampled MR image to form two-channel inputs to the subsequent 2D/3D MI-UNet. In summary, we develop an automated stroke lesion segmentation method, the contributions of which are three-fold: 1) We demonstrate that the prior knowledge of the location of GM, WM, and LV is beneficial for the lesion segmentation accuracy regardless of model architecture (2D UNet or 2D Deeplab), input dimensions (2D UNet or 3D UNet) and stroke lesion sizes. 2) We demonstrate that better brain parcellations bring better stroke lesion segmentation. 3) Our method attains superior segmentation performance over other existing methods on the same ATLAS dataset [7].

The rest of this paper is arranged as follows. The detailed procedures are shown in Section II, including the registration principle using LDDMM-image and the CNN settings. The dataset and the evaluation criteria are shown in Section III. Evaluations and experimental results of the proposed method are presented in Section IV. Finally, Section V draws discussion and conclusion of the paper.

II. METHOD

The major difference between UNet and the proposed MI-UNet is shown in Fig. 2. In a typical UNet, the MR image acts as the only input. But for MI-UNet, a template MR image together with an associated brain parcellation is firstly used to obtain the brain parcellation for the input MR image, which is then used as an additional input. In Section II-A, we introduce the LDDMM-image algorithm, a registration method used to perform atlas-based segmentation to obtain brain parcellations for all MR images. The architectures of 2D MI-UNet and 3D MI-UNet are introduced in Section II-B. In Section II-C, we introduce the training and testing strategy, which further improves the segmentation performance of MI-UNet. In Section II-D, we present some implementation details.

A. LDDMM-image

As one of the cutting-edge registration methods, LDDMM-image works efficiently in producing diffeomorphic flow with constant topology [33]. In this work, we use the fast LDDMM-image algorithm to parcellate a brain into GM, WM, and LV [31], the flowchart of which is shown in Fig. 3.

As shown in Fig. 3, let I_0 represents a 3D grayscale template image and I_1 represents a target image, both of which are real-valued functions defined on the background space $\Omega \in \mathbb{R}^3$. LDDMM-image

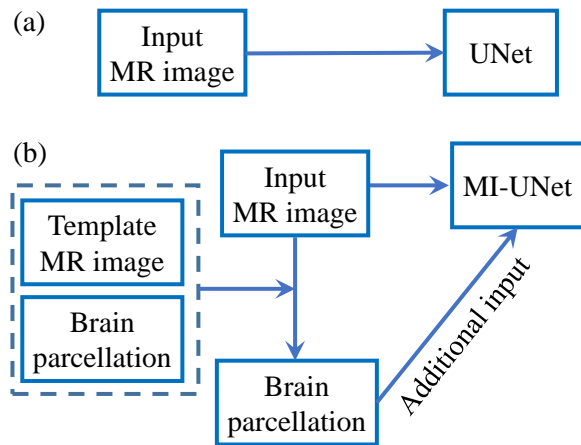


Fig. 2. Procedure of the proposed MI-UNet and comparisons with a typical UNet.

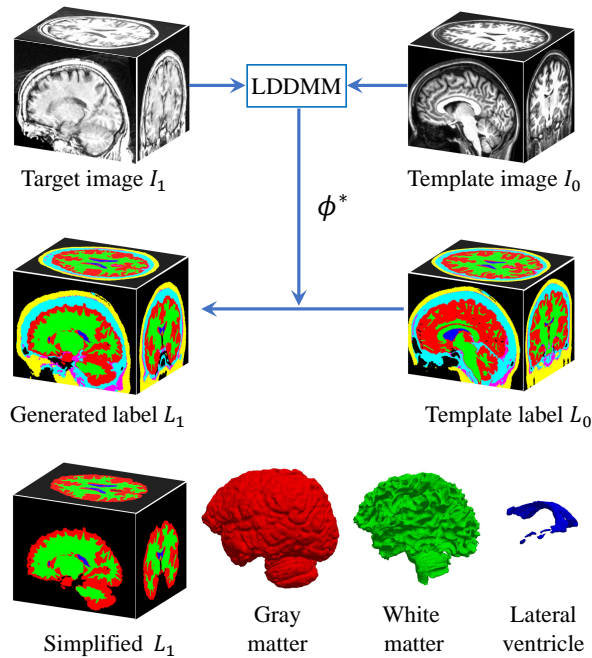


Fig. 3. Flowchart of LDDMM-image based brain parcellation into gray matter, white matter and lateral ventricle.

seeks a diffeomorphic mapping $\varphi : \Omega \rightarrow \Omega$ such that $I_0 \circ \varphi^{-1}$ is well aligned to I_1 . The time-varying diffeomorphisms are computed as

$$\phi_t = \phi_0 + \int_0^t v_\tau(\phi_\tau) d\tau, \quad (1)$$

where ϕ_0 is the identity mapping and $v_t : \Omega \times t \rightarrow R^3$ is a smooth velocity vector field, $t \in [0, 1]$. At time t , the deformed template image is $J_t = I_0 \circ \phi_{0t}$, and $\phi_{0t} = \phi_t \circ \phi_0^{-1}$ denotes the corresponding diffeomorphic coordinate transformation. To find the optimal velocity vector fields v_t^* , the following objective function is minimized [31],

$$E(v_t) = \underbrace{\frac{1}{2} \int_0^t \|Sv_t\|_{L^2}^2 dt}_{\text{regularization}} + \underbrace{\frac{1}{2\sigma^2} M(J_1, I_1)}_{\text{matching cost}}. \quad (2)$$

The objective function contains two parts: regularization and matching cost function. In the regularization term, $\|\cdot\|_{L^2}$ represents the L^2 norm of square-integrable functions, $S = -\alpha \nabla^2 + Id$ is a differential

operator, with ∇^2 and Id respectively representing the Laplacian and identity operator. The smoothness of v_t is determined by the constant $\alpha > 0$ and a smoother transformation comes from a smaller α . The matching cost function term $M(J_1, I_1)$ evaluates the degree of similarity between I_1 and J_1 . And σ determines the weight of the matching cost function relative to the regularization term. In this work, cross-correlation is used as the matching cost function [34] and σ is set to be 3, being consistent with that in Wu. *et al.* [31].

The template image used in this work has been manually divided into 7 different regions, i.e., the template label L_0 consists of 7 categories, including cerebrospinal fluid (CSF), bone marrow, LV, skull, image background, WM and GM [35]. At the very beginning, the template image is affinely aligned to the target image of interest (each image in the training set and the testing set). To find the optimal velocity vector fields v_t between the target image and the affinely aligned template image, LDDMM-image is conducted by minimizing Eq. (2). Afterwards, the optimal transformation ϕ^* is obtained through Eq. (1), which is then combined with the previously-obtained affine transformation to form the final transformation ϕ_a^* . The composed transformation ϕ_a^* is then applied to L_0 to obtain the brain parcellation (L_1) of the target image via

$$L_1 = L_0 \circ \phi_a^*. \quad (3)$$

After that, L_1 is simplified to have only the three labels of interest (GM, WM, LV) by setting all other categories to be the background. In summary, we get the parcellations of GM, WM, and LV through atlas-based segmentation and this step is independent of subsequent deep learning-based stroke lesion segmentation. Therefore, we generate brain parcellations of GM, WM, and LV for each of the training and testing images and use them as prior knowledge.

B. Network architecture

CNN is most commonly applied to analyzing visual imagery because it can automatically extract image features. UNet has an encoder-decoder style that solves segmentation tasks in an end-to-end manner [36]. As illustrated in Fig. 4, the proposed MI-UNet can be conducted in both 2D and 3D settings.

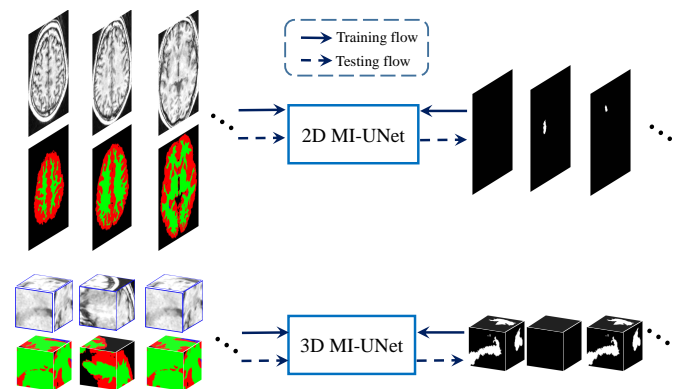


Fig. 4. Illustration of MI-UNet under 2D and 3D settings.

For each subject, its brain parcellation L_1 and T1w MR image I_1 are sliced/sampled and jointly fed into 2D/3D MI-UNet as a two-channel input. The hidden layers of UNet/MI-UNet can be reasonably categorized into two parts: (1) an encoder part that can capture multi-level, multi-resolution features by computing feature maps at different image resolutions, and (2) a decoder part that can localize stroke lesions precisely by computing the feature representations at different scales. The structure details of 2D/3D MI-UNet are shown in Table

TABLE I
ARCHITECTURES OF THE PROPOSED 2D MI-UNET AND 3D MI-UNET.

	2D UNet/MI-UNet		3D UNet/MI-UNet	
	Feature size	Parameters	Feature size	Parameters
input of UNet	$233 \times 197 \times 1$	-	$49 \times 49 \times 49 \times 1$	-
input of MI-UNet	$233 \times 197 \times 2$	-	$49 \times 49 \times 49 \times 2$	-
convolution 1	$233 \times 197 \times 64$	$[3 \times 3, 64 \text{ conv}] \times 2^a$	$49 \times 49 \times 49 \times 32$	$[3 \times 3 \times 3, 32 \text{ conv}] \times 2$
pooling	$116 \times 98 \times 64$	$[2 \times 2 \text{ max pool}]$	$24 \times 24 \times 24 \times 32$	$[2 \times 2 \times 2, \text{max pool}]$
convolution 2	$116 \times 98 \times 128$	$[3 \times 3, 128 \text{ conv}] \times 2$	$24 \times 24 \times 24 \times 64$	$[3 \times 3 \times 3, 64 \text{ conv}] \times 2$
pooling	$58 \times 49 \times 128$	$[2 \times 2 \text{ max pool}]$	$12 \times 12 \times 12 \times 64$	$[2 \times 2 \times 2, \text{max pool}]$
convolution 3	$58 \times 49 \times 256$	$[3 \times 3, 256 \text{ conv}] \times 2$	$12 \times 12 \times 12 \times 128$	$[3 \times 3 \times 3, 128 \text{ conv}] \times 2$
pooling	$29 \times 24 \times 256$	$[2 \times 2 \text{ max pool}]$	$6 \times 6 \times 6 \times 128$	$[2 \times 2 \times 2, \text{max pool}]$
convolution 4	$29 \times 24 \times 512$	$[3 \times 3, 512 \text{ conv}] \times 2$	-	-
pooling	$14 \times 12 \times 512$	$[2 \times 2 \text{ max pool}]$	-	-
convolution 5	$14 \times 12 \times 512$	$[3 \times 3, 1024 \text{ conv}], [3 \times 3, 512 \text{ conv}]$	$6 \times 6 \times 6 \times 128$	$[3 \times 3 \times 3, 256 \text{ conv}], [3 \times 3 \times 3, 128 \text{ conv}]$
upsampling	$28 \times 24 \times 1024$	$[2 \times 2 \text{ upsampling}]-[\mathcal{C}(\text{convolution 4})]^b$	$12 \times 12 \times 12 \times 256$	$[2 \times 2 \times 2 \text{ upsampling}]-[\text{convolution 3}]$
convolution 6	$28 \times 24 \times 256$	$[3 \times 3, 512 \text{ conv}], [3 \times 3, 256 \text{ conv}]$	$12 \times 12 \times 12 \times 64$	$[3 \times 3 \times 3, 128 \text{ conv}], [3 \times 3 \times 3, 64 \text{ conv}]$
upsampling	$56 \times 48 \times 512$	$[2 \times 2 \text{ upsampling}]-[\mathcal{C}(\text{convolution 3})]$	$24 \times 24 \times 24 \times 128$	$[2 \times 2 \times 2 \text{ upsampling}]-[\text{convolution 2}]$
convolution 7	$56 \times 48 \times 128$	$[3 \times 3, 256 \text{ conv}], [3 \times 3, 128 \text{ conv}]$	$24 \times 24 \times 24 \times 32$	$[3 \times 3 \times 3, 64 \text{ conv}], [3 \times 3 \times 3, 32 \text{ conv}]$
upsampling	$112 \times 96 \times 256$	$[2 \times 2 \text{ upsampling}]-[\mathcal{C}(\text{convolution 2})]$	$48 \times 48 \times 48 \times 64$	$[2 \times 2 \times 2 \text{ upsampling}]-[\mathcal{C}(\text{convolution 1})]$
convolution 8	$112 \times 96 \times 64$	$[3 \times 3, 128 \text{ conv}], [3 \times 3, 64 \text{ conv}]$	$48 \times 48 \times 48 \times 16$	$[3 \times 3 \times 3, 32 \text{ conv}], [3 \times 3 \times 3, 16 \text{ conv}]$
upsampling	$224 \times 192 \times 128$	$[2 \times 2 \text{ upsampling}]-[\mathcal{C}(\text{convolution 1})]$	-	-
convolution 9	$224 \times 192 \times 32$	$[3 \times 3, 64 \text{ conv}], [3 \times 3, 32 \text{ conv}]$	-	-
Zero Padding	$233 \times 197 \times 32$	-	$49 \times 49 \times 49 \times 16$	-
output	$233 \times 197 \times 1$	$[1 \times 1, 1] + \text{Sigmoid}$	$49 \times 49 \times 49 \times 1$	$[1 \times 1 \times 1, 1] + \text{Sigmoid}$

^a “ $3 \times 3, 64 \text{ conv}$ ” corresponds to a Conv-BN-ReLU layer with a kernel size of 3×3 and 64 features. “[\cdot] $\times d$ ” denotes that the convolution block is repeated for d times.

^b “[\cdot]-[\cdot]” denotes concatenation of two feature maps. $\mathcal{C}(\text{convolution 4})$ denotes the central-cropping operation to guarantee the same size in the concatenation operation.

I. The hyperparameters of UNet and MI-UNet are completely the same except the input size.

For 2D UNet/MI-UNet [14], each convolution block consists of two 3×3 convolutional layers, each followed by batch normalization (BN) and rectified linear unit (ReLU). The encoder part is composed of four convolution blocks, each followed by a 2×2 max-pooling layer to halve resolution of the corresponding feature maps. The decoder part consists of four 2×2 up-sampling layers, each followed by a convolution block to recover the resolution and generalize feature maps. Skip connections are employed between the encoder and decoder parts of the same resolution. To obtain the final binary segmentation, a 1×1 convolution is employed at the last layer to map the 32-component feature vector to a two-class label.

With a similar encoder-decoder structure, 3D UNet/MI-UNet is composed of 7 convolution blocks, 3 max-pooling layers, and 3 up-sampling layers [20]. The max-pooling layer and up-sampling layer of 3D UNet is one less than that of 2D UNet because the input size of 3D UNet is smaller. That is also the reason why fewer feature channels are used in 3D UNet.

C. Loss function, training and inference schemes

1) **Loss function:** The objective of training a neural network is to minimize a loss function, which generally defines a mapping from multiple variables to a real number. We train our networks using a loss combining Dice and cross-entropy [37]

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{Dice}} + \mathcal{L}_{\text{CE}}. \quad (4)$$

The volumes of stroke lesions and normal brain tissue are strongly unbalanced, which makes the segmentation challenging. To avoid the influence from a large number of background pixels, Milletari *et al.* proposed a Dice coefficient based loss and achieved satisfying

performance on a prostate segmentation task [38]. Let N denote the total number of pixels, $P = \{p_1, \dots, p_N\}$ and $G = \{g_1, \dots, g_N\}$ respectively represent the predicted map and the corresponding binary ground truth. To avoid dividing by 0, i.e. P and G are both empty, a smooth term $s = 10^{-8}$ is used in the Dice loss, namely

$$\mathcal{L}_{\text{Dice}} = - \frac{2 \sum_{n=1}^N p_n \cdot g_n + s}{\sum_{n=1}^N p_n + \sum_{n=1}^N g_n + s}, \quad (5)$$

where \cdot denotes scalar product.

The widely-used cross-entropy loss is defined as

$$\mathcal{L}_{\text{CE}} = - \frac{1}{N} \sum_{n=1}^N (g_n \cdot \log(p_n) + (1 - g_n) \cdot \log(1 - p_n)). \quad (6)$$

2) **Training scheme:** To feed 3D data into a 2D MI-UNet, we use their 2D slices along the axial plane. As shown in Fig. 4, the 2D brain parcellation and T1w MR image slices are fed into MI-UNet jointly as a two-channel input with a size of $233 \times 197 \times 2$. The 2D MI-UNet is trained for 50 epochs.

We train our proposed 3D MI-UNet for 30 epochs, with each one again composed of 20 sub-epochs. At each sub-epoch, 100 samples with a size of $49 \times 49 \times 49 \times 2$ are randomly selected from the training images and the corresponding ground truth segmentation images [20]. As shown in Fig. 4, we use 3D MR image patches and the corresponding 3D brain parcellation patches as the two-channel inputs to 3D MI-UNet.

3) **Inference scheme:** At the testing stage of 2D MI-UNet, we use the axial slices of brain parcellation and T1w MR image as the input to the trained 2D model. For 3D MI-UNet, brain parcellation and T1w MR image go through the trained 3D U-Net via sliding-window sampling. Finally, we perform thresholding processing on all prediction outcomes of 2D/3D MI-UNet. Concretely speaking, if the

probability that a pixel/voxel is predicted to be foreground is larger than 0.5, its label will be set to be one, and otherwise zero.

D. Implementation details

We have made the LDDMM-image algorithm publicly available at GitHub¹. We also made the brain parcellation used in this work publicly available at Github². The implementations of 2D UNet and 3D UNet were based on the work of Li *et al.*³ [17], Wu *et al.* [20] and Dolz *et al.* [39]⁴. The batch size was set to be 8 for 2D MI-UNet and 32 for 3D MI-UNet. Adam optimizer was used and the learning rate was set to be 0.001.

III. DATASET AND EVALUATION METRICS

A. Dataset

We evaluate our approach on the ATLAS dataset [7], consisting of 229 subjects. All images have been resampled and normalized to be 1 mm³ voxel resolution and all volumes are of the same size 233×197×189. In terms of stroke type found in the ATLAS dataset, most cases are embolic. All experiments are conducted in a five-fold cross-validation manner. For data partitioning, we first index all 229 cases according to their names and then separate the training data and testing data according to their indices. More specifically, the data having indices {1, 6, 11, ..., 226}, {2, 7, 12, ..., 227}, {3, 8, 13, ..., 228}, {4, 9, 14, ..., 229}, {5, 10, 15, ..., 225} are respectively used for testing for fold 1 to fold 5 and the others are used for training. As shown in Fig. 5, the testing data in every partition cover different lesion sizes. Through a five-fold cross-validation, every data is used as the testing data once. We also perform Wilcoxon signed-rank tests to quantify the group differences between every two folds in terms of lesion sizes. There is no statistically significant group difference (p-value>0.05).

B. Evaluation metrics

The Dice score is an evaluation metric used to quantify the overlapping ratio between two binary sets, which is defined as

$$\text{Dice} = \frac{2\text{TP}}{2\text{TP} + \text{FP} + \text{FN}}, \quad (7)$$

where TP, FP, and FN respectively represent the pixel counts of true positives (model correctly predicts), false positives (model misclassifies to be positive), and false negatives (model misclassifies to be negative). In medical image segmentation tasks, there is usually a strong size imbalance between foreground and background, and the background typically occupies over 95% of the whole volume [38]. But the foreground is what we care more, namely our object of interest. As such, here “correctly predicts” only means that a model correctly predicts a foreground voxel to be foreground. The Dice score ranges from 0 to 1, where 0 denotes no spatial overlap, and 1 denotes complete overlap.

Precision, also referred to as the positive predictive value, is the ratio of TP to the entire prediction outcome. Precision is defined as

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}, \quad (8)$$

where precision reaches its worst value at 0 and best at 1.

Recall is a measure of completeness, which is defined as the ratio of TP to the entire ground truth,

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}. \quad (9)$$

Recall also reaches its best at 1, meaning that every pixel belonging to stroke lesion is correctly predicted to be stroke lesion.

Let $S(A)$ represent the set of surface vertices of a 3D volume A , the minimum distance of any vertex v to $S(A)$ is defines as

$$d(v, S(A)) = \min_{s_A \in S(A)} \|v - s_A\|, \quad (10)$$

where $\|\cdot\|$ denotes the Euclidean distance, with a greater value indicating a larger distance.

The Hausdorff distance (HD) measures how far the surface vertices of two binary masks lie from each other. It is defined as

$$\text{HD}(G, R) = \max \left\{ \sup_{s_G} d(s_G, S(R)), \sup_{s_R} d(s_R, S(G)) \right\}, \quad (11)$$

where s_R and s_G respectively denote the surface vertices in an automated segmentation result R and the corresponding ground truth segmentation G , and sup denotes the supremum.

The average symmetric surface distance (ASSD) measures the average of all Euclidean distances between the surface vertices of two binary masks (G and R), which is defined as

$$\text{ASSD}(G, R) = \frac{\sum_{s_G} d(s_G, S(R)) + \sum_{s_R} d(s_R, S(G))}{|S(G)| + |S(R)|}. \quad (12)$$

The p-values reported below are all obtained from Wilcoxon signed-rank tests.

IV. EXPERIMENTAL RESULTS

We perform an ablation study based on the 2D model in section IV-A and show the superiority of using UNet as the baseline model and using LDDMM-image to obtain the brain parcellation. We then evaluate the segmentation performance of UNet and MI-UNet in 2D (section IV-B) and 3D settings (section IV-C).

A. Ablation analysis on 2D MI-UNet

Deeplab is one of the most widely-used semantic segmentation frameworks [40]. As such, we also investigate the performance of our proposed framework with the baseline model being Deeplab⁵. As shown in Table II, 2D UNet outperforms Deeplab in the stroke lesion segmentation task. Evidently, the proposed framework can improve the segmentation accuracy for both UNet and Deeplab. Recently, many other novel architectures have been proposed, such

TABLE II

SEGMENTATION RESULTS VIA AN ABLATION STUDY ON 2D MI-UNET IN TERMS OF THE DICE SCORE.

Baseline model	Brain parcellation		Dice [%]
Deeplab	UNet	Affine LDDMM-image	
✓			40.26
✓		✓	42.33
✓			44.68
	✓		45.54
	✓	✓	47.06
	✓		49.45

as CLCI-Net [16], ZNet [13], and WNet [41]. These networks are mainly variants of UNet [36] by adding specific efficient blocks, such as skip connection [42] and atrous spatial pyramid pooling [43]. These novel CNN architectures may further improve our lesion segmentation accuracy but at a cost of more GPU memory. In this

⁵<https://github.com/bonlime/keras-deeplab-v3-plus>

¹<https://github.com/KwuJohn/LddmmMASGD>

²<https://github.com/sustecher/StrokeSegmentation>

³https://github.com/hongweilibran/wmh_ibbmTum

⁴<https://github.com/josedolz/LiviaNET>

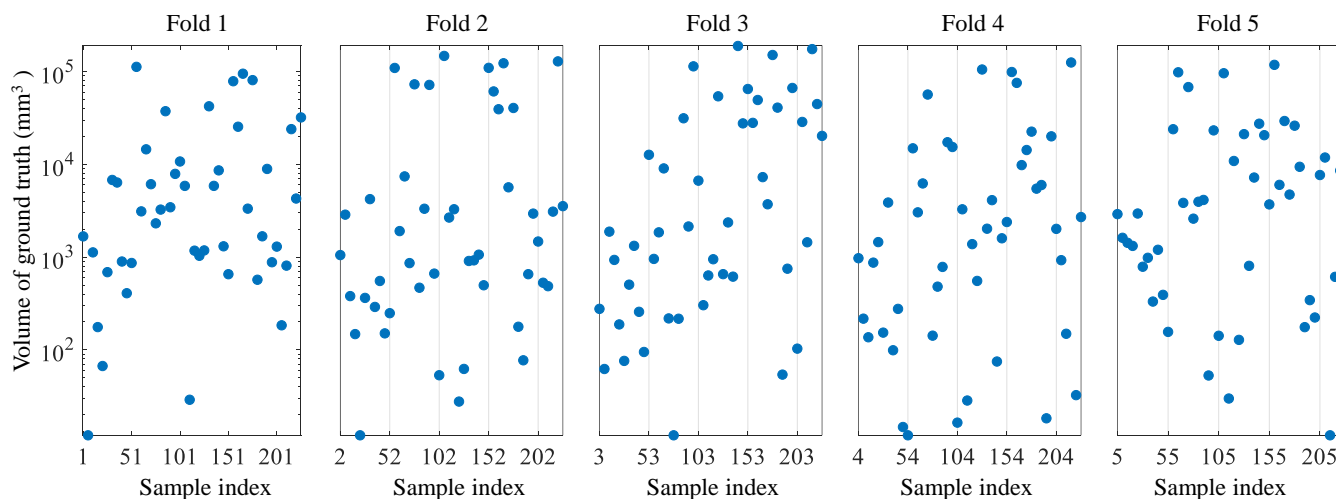


Fig. 5. Lesion size analysis of each testing data partition in the five-fold experiment.

work, our goal is to validate the importance of brain prior anatomical knowledge (registration and multi-channel). UNet is strong enough if implemented correctly and it is still the most widely-used network in many medical image segmentation challenges of MICCAI 2019. As such, we used UNet as our baseline model. We would like to kindly point out that we have released all brain parcellation results used in this work in Github, so that readers themselves can easily evaluate the performance of the proposed framework using baseline models with novel architectures.

To evaluate the effect of registration accuracy upon corresponding lesion segmentation, we also conduct 2D MI-UNet using the affine-derived brain parcellation as the additional input, which has low registration accuracy than LDDMM-image [31]. As there is no ground truth of GM/WMLV for ATLAS dataset, we evaluate the segmentation accuracy of GM/WMLV of affine and LDDMM-image using the dataset used in [35], which consists of 18 T1w MR images with human-annotated brain parcellation. With respect to average Dice score, the segmentation accuracy of GM/WMLV are 62%/69%/49% using affine and 79%/85%/83% using LDDMM-image. As shown in Table II and Fig. 6, the more accurate the registration, the more accurate the final lesion segmentation.

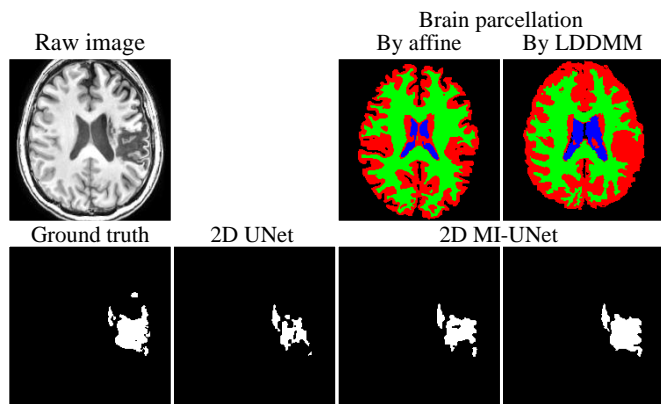


Fig. 6. Representative segmentation results of 2D UNet and 2D MI-UNet with different brain parcellation inputs.

B. 2D MI-UNet vs. 2D UNet

The prediction results of 2D UNet and 2D MI-UNet for cases of different lesion sizes (sorted in ascending order in terms of lesion size) are illustrated in Fig. 7. Clearly, results from the proposed MI-UNet (the fifth column) are closer to the ground truth (the third column) than that from UNet using only MR images as the input (the fourth column). For images with clear boundaries between foreground and background, such as the one in the last row of Fig. 7, UNet may still fail, which again demonstrates the robustness of the proposed MI-UNet.

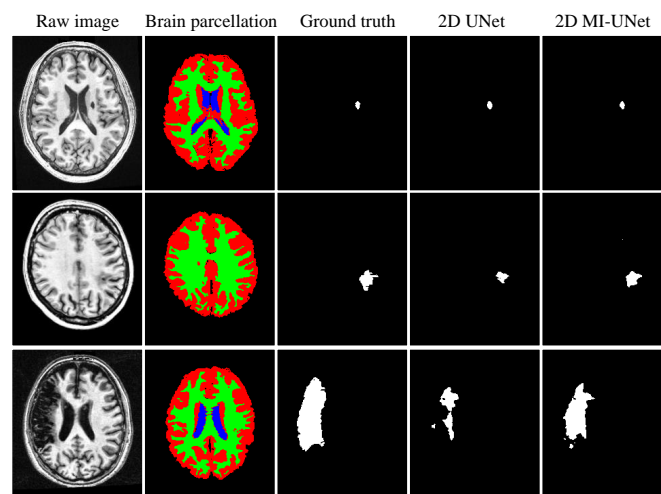


Fig. 7. Comparisons of results obtained from different methods for cases of different lesion sizes (different rows). The second column shows the brain parcellation used as the second input channel to MI-UNet.

In Fig. 8, we show some representative feature maps in the last layer of both UNet (from the first column to the third column) and MI-UNet (from the fourth column to the last column) when segmenting the image slice demonstrated in the last row of Fig. 7. From Fig. 8, we observe that the red regions in those feature maps roughly represent the low-intensity regions in the T1w MR image (including stroke lesion, CSF, LV, and background). Clearly, the proposed MI-UNet can much better distinguish lesion regions from other brain regions with similar intensity profiles when compared to

TABLE III

COMPARISONS OF DIFFERENT STROKE LESION SEGMENTATION RESULTS OBTAINED ON THE ATLAS DATASET. ↓ INDICATES THAT THE SMALLER THE BETTER.

Method	Dice[%]	HD[mm] ↓	ASSD[mm] ↓	Precision[%]	Recall[%]	ATLAS split ratio (train/test)
Clusterize [8]	23	75.00	13.59	16	79	–
ALI [9]	36	61.55	14.38	31	55	–
lesion_gnb [10]	39	58.00	10.49	30	69	–
LINDA [11]	45	42.07	12.68	50	52	–
ResUNet [44]	47.02	–	–	59.41	45.37	five-fold cross-validation
2D Dense-UNet [45]	47.41	–	–	56.13	48.75	five-fold cross-validation
X-Net [15]	48.67	–	–	60.00	47.52	five-fold cross-validation
MSDF-Net [46]	55.78	–	–	–	–	160/69
D-UNet [25]	53.49	–	–	63.31	52.43	183/46
2D UNet	45.54	65.30	20.63	50.65	51.31	five-fold cross-validation
2D MI-UNet	49.45	51.93	12.99	56.69	52.37	five-fold cross-validation
3D UNet	52.96	38.80	10.33	60.90	54.97	five-fold cross-validation
3D MI-UNet	56.72	23.94	7.00	65.45	59.38	five-fold cross-validation

the single-channel UNet.

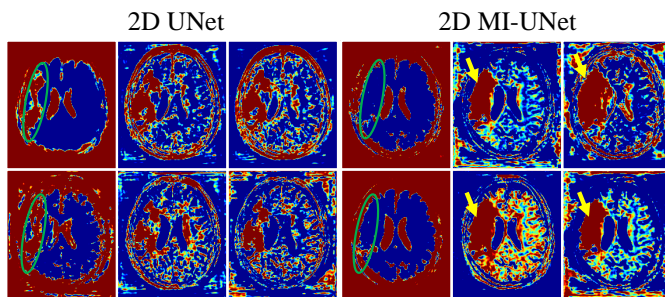


Fig. 8. Visualizations of 6 representative feature maps, out of all 32 ones, in the last layer of 2D UNet and 2D MI-UNet obtained from segmenting the image slice demonstrated in the last row of Fig. 7.

C. 3D MI-UNet vs. 3D UNet

In Fig. 9, to clearly show the difference between segmentation results of 3D UNet and 3D MI-UNet, we slice a 3D MR image and its corresponding segmentation at a fixed point along each of the three planes. Compared with 3D UNet with only MR image patches being the input, the proposed 3D MI-UNet performs better.

D. Comparisons with other methods

Table III compares MI-UNet and some existing approaches evaluated on the same ATLAS dataset, including both classical segmentation methods and deep learning-based segmentation methods. We would like to kindly point out that all the results used for comparison are directly obtained from their published papers. In other words, we do not implement their methods at our end, as we believe a direct comparison with the best results reported by the authors themselves would be the fairest way. In that table, the results of the four classical methods (Cluster, ALI, lesion_gnb, and LINDA) were obtained from Ito *et al.* [6], the results of the three 2D deep learning-based methods (ResUNet, 2D Dense-UNet, X-Net) were reported by Qi *et al.* [15], and the results of MSDF-Net and D-UNet were reported by Zhou *et al.* [25] and Liu *et al.* [46].

Our goal is to show that the additional brain parcellation can improve the segmentation accuracy over baseline models (for both 2D UNet and 3D UNet). As such, we conduct the 2D/3D U-Net and MI-UNet and obtain results by ourselves (the last four rows in Table III).

Most of the improvements (2D MI-UNet vs. 2D UNet, 3D MI-UNet vs. 3D UNet) in terms of the five evaluation metrics, are statistically significant (p -value < 0.05). From Table III, we find that generally deep learning-based methods outperform classical segmentation techniques because they can automatically capture discriminative features that outperform hand-crafted features [47]. With respect to Dice, the proposed method has the highest mean score of 49.45%/56.72% among 2D/3D deep learning-based methods. Meanwhile, our best model (3D MI-UNet) has a smaller average HD than LINDA (from 42.07 mm to 23.94 mm) [11] and a smaller ASSD (from 10.49 mm to 7.00 mm) than ALI [9]. The performance of the Cluster method is the best in terms of recall, but its Dice and precision are the lowest, indicating that Clusterize misclassifies a lot of non-lesion areas to be lesions, and hence yields the best recall. All deep learning-based methods have relatively lower recall but higher precision and Dice than traditional methods. In summary, the proposed strategy can significantly improve the segmentation performance of both 2D UNet and 3D UNet and the best result (MI-3D UNet) outperforms existing state-of-the-art methods.

E. Analyses of lesion sizes

To evaluate the performance of UNet and MI-UNet for segmenting lesions of different sizes, we divide all subjects into three categories according to their lesion sizes computed from the ground truth segmentations. Specifically, we use the 33rd and 67th percentiles of lesion sizes as two thresholds to divide all subjects into a small lesion group, a medium lesion group, and a large lesion group. To quantitatively compare the segmentation performance between UNet and MI-UNet for lesions of different sizes, we tabulate the average Dice score and p -values between 2D/3D UNet and MI-UNet in Table IV.

From Table IV, it can be observed that the additional input (brain parcellation) plays a significant role in improving the performance on segmenting lesions of all three-category sizes. These improvements are all statistically significant ($p < 0.01$). Unsurprisingly, both UNet and the proposed MI-UNet perform best for large lesions. For both UNet and MI-UNet, significant performance differences are found between any groups ($p < 0.01$). This clearly indicates that the lesion size largely affects the segmentation performance of an automated approach.

V. DISCUSSION AND CONCLUSION

Stroke lesion segmentation plays a vital role in stroke rehabilitation. In this paper, we presented a fully automatic stroke lesion seg-

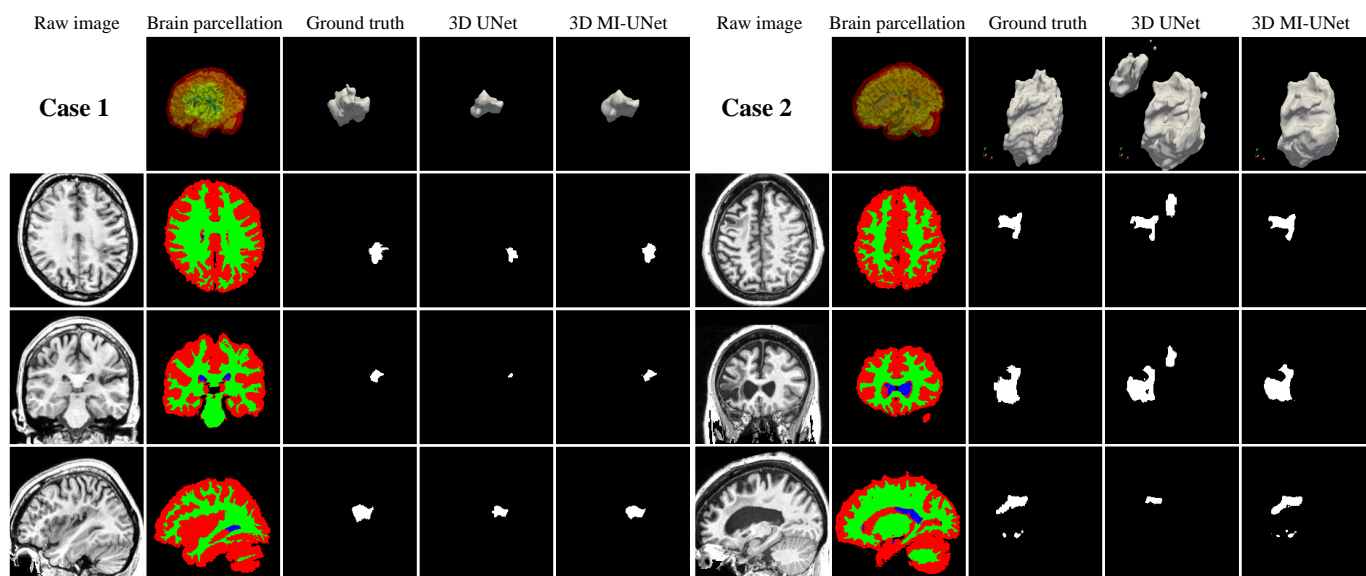


Fig. 9. Representative segmentation results obtained from 3D UNet and 3D MI-UNet. The first row shows the 3D displays of the ground truth and all automated segmentations. From the second to the fourth row, MR image and the corresponding segmentation are displayed at three planes (axial, coronal, and sagittal).

TABLE IV

DICE COMPARISONS BETWEEN UNET AND MI-UNET IN BOTH 2D AND 3D SETTINGS FOR SEGMENTING STROKE LESIONS WITHIN THE THREE CATEGORIES (SMALL LESION SIZE, MEDIUM LESION SIZE, AND LARGE LESION SIZE), THE NUMBER OF WHICH IS 77, 76, 76.

	Small lesions	Medium lesions	Large lesions	Overall
2D UNet	27.18	46.15	63.52	45.54
2D MI-UNet	30.79	51.12	66.70	49.45
p-value	0.006	$4.52 \cdot 10^{-6}$	0.002	$4.23 \cdot 10^{-8}$
3D UNet	35.14	53.45	70.53	52.96
3D MI-UNet	37.74	58.70	73.97	56.72
p-value	0.008	$7.22 \cdot 10^{-6}$	$1.44 \cdot 10^{-5}$	$2.76 \cdot 10^{-6}$

mentation approach utilizing brain parcellation as prior knowledge.

In contrast to typical UNet with only MR image being the input, the proposed MI-UNet benefits from brain parcellation by using it as an additional input channel. MI-UNet can obtain better segmentation results for baseline models of different architectures (2D UNet or 2D Deeplab), for different input dimensions (2D UNet or 3D UNet) and for lesions of varying sizes (small, medium, large). Both quantitative and qualitative analyses suggest that our automated stroke lesion segmentations are reliable. The proposed strategy can significantly improve the segmentation performance of both 2D UNet and 3D UNet and the best result (MI-3D UNet) outperforms existing state-of-the-art methods. The average intra-rater Dice/HD of the ATLAS dataset is 83%/21.02 mm and the average inter-rater Dice/HD of ATLAS dataset is 73%/22.57 mm [7]. The proposed method achieves comparable HD scores with inter-rater. However, in terms of Dice, there is still room for improvement for the proposed method.

In terms of computation time, it will take roughly 90 seconds to get the brain parcellation using LDDMM-image for an MR image of size $233 \times 197 \times 189$. The additional input will not cause extra computation burden to the baseline neural network. Including the LDDMM-image steps, it will take 95 seconds for 2D MI-UNet and 140 seconds for 3D MI-UNet to segment out stroke lesions for an MR image of size $233 \times 197 \times 189$.

This method can be easily extended to other brain lesion segmentation tasks or transferred to MR images of other modalities. Our released codes and brain parcellations for the ATLAS dataset may be helpful for others to extend the proposed framework.

A potential limitation of this work is that the registration step occupies a large part of computation time. Deep learning has been showing promising performance in registration and is very fast in the testing stage [48]. In our future work, implementing LDDMM-image in a deep learning manner may largely improve the computational efficiency of the proposed pipeline. In addition, we aim to further increase the segmentation accuracy of the proposed pipeline for small lesions, which still remains a challenging task. A potential solution is to iterate the segmentation process such that the lesion areas identified in the current iteration can be used as an additional input channel in the next iteration.

REFERENCES

- [1] S. S. Virani, A. Alonso, E. J. Benjamin *et al.*, "Heart disease and stroke statistics-2020 update: A report from the american heart association," *Circulation*, vol. 141, no. 9, pp. e139–e596, 2020.
- [2] P. Langhorne, J. Bernhardt, and G. Kwakkel, "Stroke rehabilitation," *Lancet*, vol. 377, no. 9778, pp. 1693–1702, 2011.
- [3] E. Burke Quinlan, L. Dodakian, J. See, A. McKenzie, V. Le, M. Wojnowicz, B. Shahbaba, and S. C. Cramer, "Neural function, injury, and stroke subtype predict treatment gains after stroke," *Ann. Neurol.*, vol. 77, no. 1, pp. 132–145, 2015.
- [4] J. D. Riley, V. Le, L. Der-Yeghian, J. See, J. M. Newton, N. S. Ward, and S. C. Cramer, "Anatomy of stroke injury predicts gains from therapy," *Stroke*, vol. 42, no. 2, pp. 421–426, 2011.
- [5] R. Lindenberg, V. Renga, L. Zhu, F. Betzler, D. Alsop, and G. Schlaug, "Structural integrity of corticospinal motor fibers predicts motor impairment in chronic stroke," *Neurology*, vol. 74, no. 4, pp. 280–287, 2010.
- [6] K. L. Ito, H. Kim, and S.-L. Liew, "A comparison of automated lesion segmentation approaches for chronic stroke T1-weighted MRI data," *Hum. Brain Mapp.*, vol. 40, no. 16, pp. 4669–4685, 2019.
- [7] S.-L. Liew, J. M. Anglin, N. W. Banks, M. Sondag, K. L. Ito, H. Kim, J. Chan, J. Ito, C. Jung, N. Khoshab *et al.*, "A large, open source dataset of stroke anatomical brain images and manual lesion segmentations," *Sci. Data*, vol. 5, p. 180011, 2018.
- [8] B. de Haan, P. Clas, H. Juenger, M. Wilke, and H.-O. Karnath, "Fast semi-automated lesion demarcation in stroke," *NeuroImage Clin.*, vol. 9, pp. 69–74, 2015.

- [9] M. L. Seghier, A. Ramlackhansingh, J. Crinion, A. P. Leff, and C. J. Price, "Lesion identification using unified segmentation-normalisation models and fuzzy clustering," *Neuroimage*, vol. 41, no. 4, pp. 1253–1266, 2008.
- [10] J. C. Griffiths, J. B. Allendorfer, and J. P. Szaflarski, "Voxel-based gaussian naïve bayes classification of ischemic stroke lesions in individual T1-weighted MRI scans," *J. Neurosci. Methods*, vol. 257, pp. 97–108, 2016.
- [11] D. Pustina, H. B. Coslett, P. E. Turkeltaub, N. Tustison, M. F. Schwartz, and B. Avants, "Automated segmentation of chronic stroke lesions using LINDA: Lesion identification with neighborhood data analysis," *Hum. Brain Mapp.*, vol. 37, no. 4, pp. 1405–1421, 2016.
- [12] S. B. Martins, G. Ruppert, F. Reis, C. L. Yasuda, and A. X. Falcão, "A supervoxel-based approach for unsupervised abnormal asymmetry detection in MR images of the brain," in *Int. Symp. Biomed. Imaging*. IEEE, 2019, pp. 882–885.
- [13] Y. Zhang, J. Wu, C. Wanli, C. Yifan, and X. Tang, "Prostate segmentation using Z-Net," in *Int. Symp. Biomed. Imaging*. IEEE, 2019, pp. 11–14.
- [14] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Med. Image Comput. Comput. Assist. Interv.* Springer, 2015, pp. 234–241.
- [15] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, Q. Liu, and S. Wang, "X-net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Med. Image Comput. Comput. Assist. Interv.* Springer, 2019, pp. 247–255.
- [16] H. Yang, W. Huang, K. Qi, C. Li, X. Liu, M. Wang, H. Zheng, and S. Wang, "Clci-net: Cross-level fusion and context inference networks for lesion segmentation of chronic stroke," in *Med. Image Comput. Comput. Assist. Interv.* Springer, 2019, pp. 266–274.
- [17] H. Li, G. Jiang, J. Zhang, R. Wang, Z. Wang, W.-S. Zheng, and B. Menze, "Fully convolutional network ensembles for white matter hyperintensities segmentation in MR images," *NeuroImage*, vol. 183, pp. 650–665, 2018.
- [18] G. Zeng and G. Zheng, "3D tiled convolution for effective segmentation of volumetric medical images," in *Med. Image Comput. Comput. Assist. Interv.* Springer, 2019, pp. 146–154.
- [19] K. C. Wong and M. Moradi, "SegNAS3D: Network architecture search with derivative-free global optimization for 3D image segmentation," in *Med. Image Comput. Comput. Assist. Interv.* Springer, 2019, pp. 393–401.
- [20] J. Wu, Y. Zhang, and X. Tang, "A multi-atlas guided 3D fully convolutional network for MRI-based subcortical segmentation," in *Int. Symp. Biomed. Imaging*. IEEE, 2019, pp. 705–708.
- [21] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: learning dense volumetric segmentation from sparse annotation," in *Med. Image Comput. Comput. Assist. Interv.* Springer, 2016, pp. 424–432.
- [22] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D cnn with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, 2017.
- [23] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, L. Chen *et al.*, "ISLES 2015-a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI," *Med. Image Anal.*, vol. 35, pp. 250–269, 2017.
- [24] R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. Mok, W. C. Chu, D. Wang, and L. Shi, "Automatic segmentation of acute ischemic stroke from DWI using 3D fully convolutional DenseNets," *IEEE Trans. Med. Imaging*, vol. 37, no. 9, pp. 2149–2160, 2018.
- [25] Y. Zhou, W. Huang, P. Dong, Y. Xia, and S. Wang, "D-UNet: a dimension-fusion U shape network for chronic stroke lesion segmentation," *IEEE/ACM Trans. Comput. Biol. Bioinform.*, 2019.
- [26] S. Winzeck, A. Hakim, R. McKinley, J. A. Pinto, V. Alves, C. Silva, M. Pisov, E. Krivov, M. Belyaev, M. Monteiro *et al.*, "ISLES 2016 and 2017-benchmarking ischemic stroke lesion outcome prediction based on multispectral MRI," *Front. Neurol.*, vol. 9, 2018.
- [27] L.-P. Berner, T.-H. Cho, J. Haesebaert, J. Bouvier, M. Wiart, N. Hjort, I. K. Mikkelsen, L. Derex, G. Thomalla, S. Pedraza *et al.*, "MRI assessment of ischemic lesion evolution within white and gray matter," *Cerebrovasc. Dis.*, vol. 41, no. 5–6, pp. 291–297, 2016.
- [28] C. Matute, M. Domercq, A. Pérez-Samartín, and B. R. Ransom, "Protecting white matter from stroke injury," *Stroke*, vol. 44, no. 4, pp. 1204–1211, 2013.
- [29] Y. Wang, G. Liu, D. Hong, F. Chen, X. Ji, and G. Cao, "White matter injury in ischemic stroke," *Prog. Neurobiol.*, vol. 141, pp. 45–60, 2016.
- [30] C. Chen, A. Bivard, L. Lin, C. R. Levi, N. J. Spratt, and M. W. Parsons, "Thresholds for infarction vary between gray matter and white matter in acute ischemic stroke: a ct perfusion study," *J. Cereb. Blood Flow Metab.*, vol. 39, no. 3, pp. 536–546, 2019.
- [31] J. Wu and X. Tang, "A large deformation diffeomorphic framework for fast brain image registration via parallel computing and optimization," *Neuroinformatics*, vol. 18, pp. 251–266, 2020.
- [32] X. Qian, Y. Lin, Y. Zhao, X. Yue, B. Lu, and J. Wang, "Objective ventricle segmentation in brain ct with ischemic stroke based on anatomical knowledge," *BioMed Res. Int.*, vol. 2017, 2017.
- [33] M. I. Miller, A. Trounev, and L. Younes, "On the metrics and euler-lagrange equations of computational anatomy," *Annu. Rev. Biomed. Eng.*, vol. 4, no. 1, pp. 375–405, 2002.
- [34] B. B. Avants, C. L. Epstein, M. Grossman, and J. C. Gee, "Symmetric diffeomorphic image registration with cross-correlation: evaluating automated labeling of elderly and neurodegenerative brain," *Med. Image Anal.*, vol. 12, no. 1, pp. 26–41, 2008.
- [35] J. Wu, Y. Zhang, and X. Tang, "Simultaneous tissue classification and lateral ventricle segmentation via a 2D U-net driven by a 3D fully convolutional neural network," in *Eng. Med. and Biol. Soc.* IEEE, 2019, pp. 5928–5931.
- [36] T. Falk, D. Mai, R. Bensch, Ö. Çiçek, A. Abdulkadir, Y. Marrakchi, A. Böhm, J. Deubner, Z. Jäckel, K. Seiwald *et al.*, "U-Net: deep learning for cell counting, detection, and morphometry," *Nat. Methods*, vol. 16, no. 1, p. 67, 2019.
- [37] F. Isensee, J. Petersen, A. Klein, D. Zimmerer, P. F. Jaeger, S. Kohl, J. Wasserthal, G. Koehler, T. Norajitra, S. Wirkert *et al.*, "nnu-net: Self-adapting framework for u-net-based medical image segmentation," *arXiv preprint*, 2018.
- [38] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Int. Conf. 3D Vis.* IEEE, 2016, pp. 565–571.
- [39] J. Dolz, C. Desrosiers, and I. B. Ayed, "3D fully convolutional networks for subcortical segmentation in mri: A large-scale study," *NeuroImage*, vol. 170, pp. 456–470, 2018.
- [40] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. Eur. Conf. Computer Vis.*, 2018, pp. 801–818.
- [41] W. Chen, Y. Zhang, J. He, Y. Qiao, Y. Chen, H. Shi, E. X. Wu, and X. Tang, "Prostate segmentation using 2D bridged U-net," in *Int. Jt. Conf. Neural Netw.* IEEE, 2019, pp. 1–7.
- [42] M. Drozdal, E. Vorontsov, G. Chartrand, S. Kadoury, and C. Pal, "The importance of skip connections in biomedical image segmentation," in *Deep Learning and Data Labeling for Medical Applications*. Springer, 2016, pp. 179–187.
- [43] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, 2017.
- [44] Z. Zhang, Q. Liu, and Y. Wang, "Road extraction by deep residual u-net," *IEEE Geosci. Remote. Sens. Lett.*, vol. 15, no. 5, pp. 749–753, 2018.
- [45] X. Li, H. Chen, X. Qi, Q. Dou, C.-W. Fu, and P.-A. Heng, "H-DenseUNet: hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imaging*, vol. 37, no. 12, pp. 2663–2674, 2018.
- [46] X. Liu, H. Yang, K. Qi, P. Dong, Q. Liu, X. Liu, R. Wang, and S. Wang, "MSDF-Net: Multi-scale deep fusion network for stroke lesion segmentation," *IEEE Access*, vol. 7, pp. 178 486–178 495, 2019.
- [47] I. Razzak, M. Imran, and G. Xu, "Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks," *IEEE J. Biomed. Health Inform.*, 2018.
- [48] X. Yang, R. Kwitt, M. Styner, and M. Niethammer, "Quicksilver: Fast predictive image registration—a deep learning approach," *NeuroImage*, vol. 158, pp. 378–396, 2017.