IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received February 12, 2020, accepted February 26, 2020, date of publication March 2, 2020, date of current version March 16, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2977415

Ischemic Stroke Lesion Segmentation Using Multi-Plane Information Fusion

LONG ZHANG[1], (Member, IEEE), RUONING SONG[1], YUANYUAN WANG[2], CHUANG ZHU[1], (Member, IEEE), JUN LIU[1], (Member, IEEE), JIE YANG[1], (Member, IEEE), AND LIAN LIU[3]

[1]School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics, Beijing 100191, China
[3]Interventional Neuroradiology Department, Beijing Tiantan Hospital, Capital Medical University, Beijing 100070, China

Corresponding authors: Chuang Zhu (czhu@bupt.edu.cn) and Lian Liu (liuliandaxiang@126.com)

This work was supported in part by the National Natural Science Foundation of China Grant 81972248, in part by the Beijing Municipal Natural Science Foundation Grant 7202056, in part by the Beijing Natural Science Foundation under Grant 4182044, in part by the Beijing Laboratory of Advanced Information Networks of BUPT, in part by the Beijing Key Laboratory of Network System Architecture and Convergence of BUPT, in part by the National Key Research and Development Program of China under Grant 2017YFB0802701, in part by the Fundamental Research Funds for the Central Universities under Grant 2018XKJC04, and in part by the 111 Project of China under Grant B17007.

ABSTRACT Diffusion-weighted magnetic resonance imaging (DWI) is sensitive to acute ischemic stroke and is a common diagnostic method for the stroke. However, the diagnostic result relies on the visual observation of neurologists which may vary from doctor to doctor under different circumstance. And manual segmentation is often a time-consuming and subjective process. The time from onset to thrombus removal has a significant impact on the prognosis of patients with acute ischemic stroke. The shorter the time, the better the prognosis. For this purpose we present a novel framework to quickly and automatically segment the ischemic stroke lesions on DWI. We mainly have three contributions: firstly, we design a detection and segmentation network (DSN) to solve the two kinds of data imbalance; secondly, we propose a triple-branch DSN architecture, used for extracting different plane feature respectively; thirdly, we propose a multi-plane fusion network (MPFN), which aims to make final prediction more accurate. Extensive experiments on ISLES2015 SSIS DWI sequence dataset demonstrate the superiority of our proposed segmentation method. The dice reached 62.2% and the sensitivity reached 71.7%.

INDEX TERMS Ischemic stroke, DWI, feature fusion, image segmentation, deep learning.

I. INTRODUCTION

Cerebrovascular accident (CVA) is a leading cause of disability and death in the world. There will be an estimated 15 million new CVA cases and 5 million deaths in 2018 [1]. Ischemic stroke is the most common type, making up about 87 percent of all CVA. It's caused by a blockage in an artery supplying the brain with blood. Hemorrhagic strokes which account for the remaining 13 percent, can be caused by high blood pressure or an aneurysm [2]. Patients must be effectively treated within 4.5 hours of onset, otherwise they may have a high probability of disability or even death [3]. It has been proven that the key point to the diagnosis and treatment of ischemic stroke is to obtain the location and volume of the stroke lesion [4]. Magnetic resonance imaging (MRI) and Computed Tomography (CT) are the most commonly used diagnostic methods for ischemic stroke [5]. The advantages of CT are high popularity, fast, and affordable. However, DWI, a kind of MRI, is more sensitive to the ischemic stroke lesions than the CT [6], which makes it a great advantage in the diagnosis of early acute ischemic stroke.

The lesion detection and quantification are important for the treatment of ischemic stroke [7]. However, the neurologists can only segment the the stroke lesions manually, this process can be extremely tedious and time-consuming, and the segmentation accuracy can be negatively affected by many factors, such as the complexity of the tissue structure, and the ambiguous characteristics between lesions and artifacts [8]. Therefore, it is of great significance to design an automatic segmentation system for acute ischemic stroke lesions, but it is a complicated task [9]. Some methods have been proposed to assist doctors in solving these problems [10], but the segmentation results often contain artifacts, which need to be corrected by the neurologists. In clinical

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

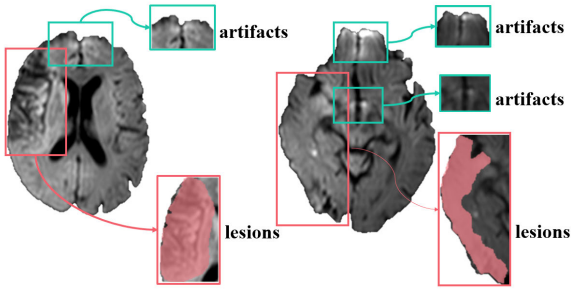


FIGURE 1. Examples of ischemic lesions in DWI. The red rectangles indicate the ischemic lesions and the blue ones show the artefacts.

practice, this kind of method is still time-consuming, so it is better to have a fully automatic segmentation method.

The deep architecture of convolutional neural network can extract a series of effective features without human intervention [11], which makes it possible for fully automatic segmentation of ischemic stroke lesions. In recent years, the deep learning (DL) methods have been widely used in medical image processing tasks to assist physicians in improving the accuracy and efficiency of medical diagnosis [12]. The segmentation of ischemic stroke lesions has been formulated as a semantic segmentation task, and the major challenges for this medical image segmentation task can be summarized as follows:

- Pixel-level and image-level data imbalance
- High cost of clean medical image dataset
- Location variation and small volume
- Lesions similar to the artifacts (see Fig. 1)

Many attempts have been done to solve these problems. The basic idea is data augmentation. Nonnegative matrix factorization (NMF) method optimizes segmentation results through 3D information and shape prior [13]. 3-D fully convolutional densenets (FCD) can learn more effective features from data due to its complex structure [14]. Automated region growing (ARG) method first reduces the range of region-growing [15] through image detection, and obtains segmentation results through seed points [16]. However, these researches have not take the two kinds of data imbalance and image artifacts into consideration.

In this paper, we focus on the ischemic stroke lesion segmentation. A fully automated method based on feature pyramidal networks (FPN) [17] is proposed in our work. Considering that there are two kinds of data imbalance in our task, we design a FPN based architecture as the detection network for extracting the patches from the entire image and utilize the U-NET [18] architecture as the segmentation network for segmenting the lesions in image patches. We call the ensemble of detection network and segmentation network DSN. In addition, it is obvious that the spatial information is of great benefit to improve learning accuracy. On the one hand, spatial information can generate smoother segmentation results, and on the other hand, it can suppress image artifacts to some extent. We introduce spatial information into our segmentation model, the DWI sequence is first segmented by the triple branches DSN, and then the segmentation results of DSN are fused by the MPFN. The most important innovation of our work is that we improve the segmentation accuracy through the information of different plane views.

We conducted experiments on the ISLES2015 SSIS DWI sequence dataset, the results demonstrate that our proposed method significantly improve the performance of ischemic stroke lesion segmentation. To summarize, the main contributions of this paper are as follows:

1) Solved the two types of data imbalance. We design DSN network, it first detect the lesions than segment on the detection bounding boxes.
2) Introduced spatial features. We design a three-branch DSN structure, which segment the stroke lesions from different plane view.
3) Suppress image artifacts. We design the network to optimize the DSNs' segmentation results and eliminate image artifacts.

II. RELATED WORK

Most of the early acute ischemic stroke imaging researches are tested on the CT data and design the image features manually [19]. First, they extract the designed image features and then use traditional machine learning algorithms to classify or detect. In fact of CT is not sensitive enough to the ischemic stroke lesions, many features only have limited contribution to the prediction accuracy. One feature that has been proven effective and widely used in researches is the symmetrical of brain, compare the differences in brain symmetric regions can identify the abnormal regions easily [20]. Further analysis can determine the lesion location and volume [21].

The performance of these methods depends on the features designed by the researchers, which may have poor generalization capability. So, it is preferable to make the model learn the feature maps from the data by itself. The deep learning algorithm has a deeper network structure that can extract more complex and abstract features from the data [22]. In recent years, it has been widely applied to tasks such as classification and semantic segmentation. Initially, deep learning algorithms were mainly used for classification tasks and achieve good results on various datasets [23]. However, there are many problems arise in applying the deep learning methods to the medical images. For example, the labeled data is scarce, which may lead to overfitting. And the label with insufficient accuracy, which severely prevent the model extracting features from the training data set. Despite these limitations, many DL-based methods have achieved good performance on the brain tumor segmentation task [24]–[27]. Because image segmentation of medical images is often regarded as a pixel-level classification problem and most of the pixels are normal tissue, which means serious imbalance in the dataset. Chen *et al.* design the ensemble of two deconvnets (EDD) first segment roughly and than the multiscale convolutional label evaluation network (MUSCLE) supress the image artifacts [28]. Partial Differential Equation (PDE) [29] model is proposed for image reconstruction,

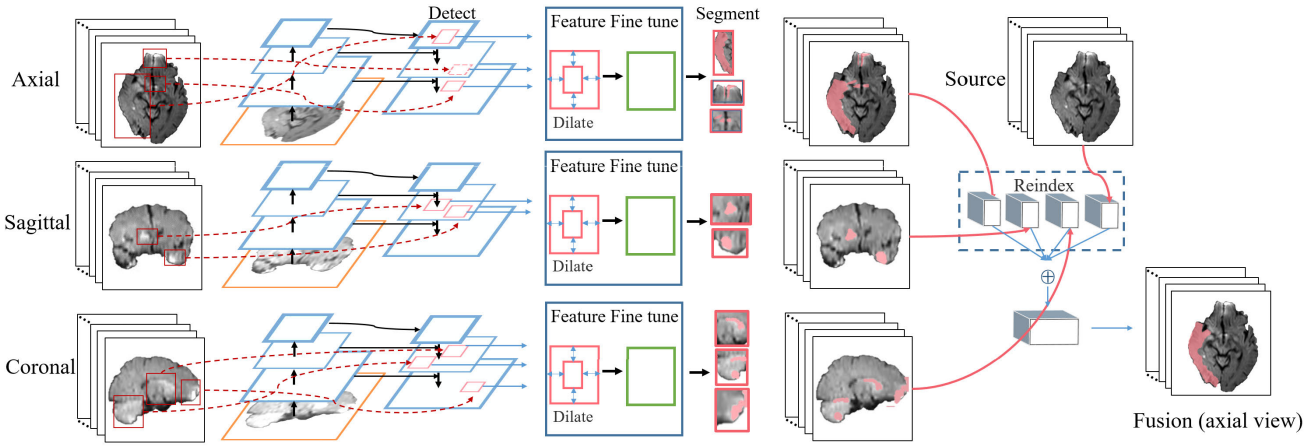


FIGURE 2. Illustration of the MULTI-PLANE Information Fusion Network. Triple branches take the different planes as input respectively. The red rectangles represent the candidate patches, red masks indicate possible lesions.

so that brain tumors will be more obvious. Other researchers use the multimodal MRI to optimize brain tumor segmentation results and achieve better performance [30]. MRI can also be used for ischemic stroke lesion segmentation, since the region of DWI abnormality can act as a gold standard for irreversible brain infarction in clinical practice. Multimodal MRI can be used to classify the onset time of ischemic stroke [31] and they may help to eliminate the artifacts in segmentation results [32]. For example, the method uses fuzzy C-means algorithm to segment the lesion on DWI, then eliminate artifacts through T1, T2 and Flair [33]. The segmentation of sub-acute ischemic stroke lesion is one of the tasks in ISLES 2015, which attracts many entries. This challenge is to segment sub-acute ischemic stroke lesions on the multi-modal MRI sequences automatically. Among the top ranked approaches, DeepMedic perform best. DeepMedic is a multi-scale 3D CNN with fully connected conditional random fields (CRFs) achieving a Dice score of 0.59 in testing. The second best performing method use a modified level-set approach embedded with the fuzzy C-means algorithm [34] while the third best method is based on random forests and contextual clustering [35]. R. Karthika's team proposed a U-NET improvement model to segment the ischemic stroke lesion on the multimodal MRI [36].

CT perfusion is also a diagnostic method for ischemic stroke. It reflects the blood flow through the fluorescent agent and can clearly show the ischemic stroke lesions. Reference [37] is an ischemic stroke lesion segmentation method on CT perfusion image. However, perfusion CT is not widely used due to its high cost and harm to the human body.

Other background methods related to the topic small lung nodules detection [38]. It detects tiny lesions through local variance calculation and fuzzy-logic and probabilistic neural network. Generalization is critical, [39] is a novel training method to preserve generalization.

The above researches are designed for a certain challenge. But none of them considered the two kinds of data imbalance and image artifacts simultaneously. In this paper, we focus on solving these issues and we also compare model performance experimentally.

III. METHOD

Our proposed approach is reffered as Multi-Plane Information Fusion Network, employs a triple branches DSN and a MPFN. The DSN first finds candidate regions of interest (ROIs) through detection, and then performs lesion segmentation on the candidate ROIs, MPFN is dedicated to smooth segmentation results and suppress image artifacts. The architecture of our work is shown in Fig. 2. This section is organzied as follows: the illustration of DSN; the illustration of MPFN; the illustration of evaluation metrics; the illustration of implementation details.

A. DSN

We find that there are two types of data imbalance in the ischemic stroke lesion segmentation task. First is pixel-level imbalance, there are more normal tissues than the abnormal. Second is slice-level imbalance, normal slices more than the abnormal. The current segmentation model does not address these data imbalance effectively. For the image level imbalance problem, we use the image augmentation method, and convert this task into a detection and segmentation task. First, DSN detects the lesion on the DWI sequence. Second, it segments the lesion on the detection results.

Our DSN is an improvement of the Mask-RCNN architecture [40]. Mask-RCNN also detects objects before segmenting. However, Mask-RCNN is an instance segmentation model, so its network structure is very complicated and have large amount of the parameters. We modify the model structure based on the characteristics of the ischemic stroke segmentation task and reduced the amount of parameters.

Fig. 3 shows the architecture details of DSN. The input is a full image, which has been resized at 1024×1024. The backbone is FPN-resnet101, which is used to extract lesion features at different scales. We choose the FPN for the following

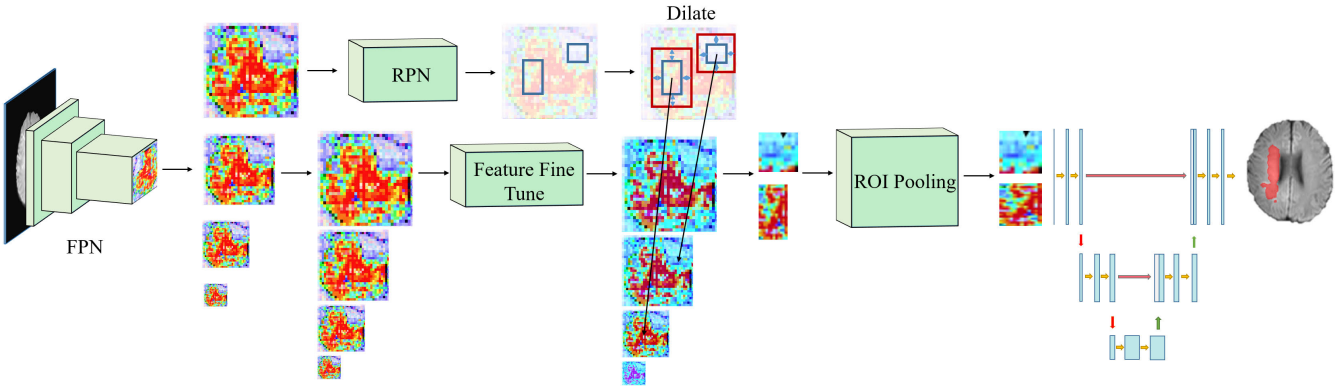


FIGURE 3. Illustration of the detection and segmentation network (DSN). The different size heatmaps represent the multi-scales feature maps.

reasons: first, objects of different scales will be assigned to different image feature levels, which addresses objects of different scales; second, FPN has a high resolution feature map for small objects. Note all scales of FPN feature maps share the same afterwards network architecture and the same model weights. In the detection task, contextual information often provides important knowledge for labeling the detection bounding boxes [41]. However, it is difficult to determine the optimal level of contextual information. On the one hand, too much contextual information may hide the true lesion. After a trade-off between performance and effectiveness, We set the depth of the FPN to 5, so we can obtain feature maps at five scales: 32×32, 64×64, 128×128, 256×256, and 512×512. These feature maps can contain Most cases.

Region proposal network (RPN) is used to generate the candidate regions, which are used for subsequent lesion segmentation. In order to reduce the information loss caused by RPN detection, the candidate regions need to be dilated, and different scale has different dilate coefficient. On the other hand, the features used for detection may not necessarily suitable for segmentation task, We design feature fine tune module modifies the multi-scale feature maps of FPN. Then we map the dilated candidate regions to the fine tuned feature maps. We use the following formula to select the best feature scale:

$$k = \lfloor k_0 + log(\frac{\sqrt{whS}}{224}) \rfloor \tag{1}$$

where k_0 is default feature level, S is the input image area, and 224 is the imagenet pretrain parameter. ROI pooling module is used for feature sampling to generate same size feature maps. Then, U-NET uses these feature maps for lesion segmentation. Generally, the deeper the U-NET, the better the performance, but the downsampling operation will cause information loss. Considering that the size of input is 28×28, we set the depth of the U-NET to 3. Finally, it maps the segmentation masks to the original image to get the segmentation results.

Our dataset is not very large, and there are pixel-level and image-level data imbalance in our data set. The sequential model DSN have two advantages to solve these problems.

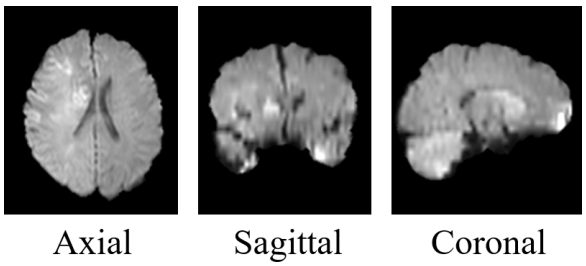


FIGURE 4. Examples of different plane view. The three pictures are from the same patient.

1) It can extract a large number of patches from image slices, which is a fundamental requirement for CNN training. 2) Control the proportion of positive and negative samples. We also optimize the training efficiency by using high-quality patches, which choosed by the IOU threshold.

We train triple branches DSN to identify ischemic stroke lesions on different planes, which are axial, coronal and sagittal (see Fig. 4).

1) LOSS FUNCTION

For training DSN's detection module, we assign a class label to each anchor. We assign a positive label to two types of anchors: 1) the anchor has highest IOU with ground truth box, or 2) the anchor has IOU with any ground truth box greater than τ_1, and we assign negative label to a non-positive anchor if IOU lower than τ_2. The values of τ_1 and τ_2 should take into account the number of positive and negative samples and the validity of the features learned by the model. So, we minimize an objective function following the multi-task loss. Our loss function is defined as:

$$\begin{aligned} \pounds &= \pounds(\{p_i\}, \{t_i\}) \\ &= \frac{1}{N_{cls}} \sum_i \pounds_{cls}(p_i, p_i^*) + \lambda \frac{1}{N_{reg}} \sum_i p_i^* \pounds_{reg}(t_i, t_t^*) \end{aligned} \tag{2}$$

where i is the anchors index, p_i is the positive softmax probability, and p_i^* is the corresponding GT prediction probability. t is predict bounding box, t^* represent the GT box corresponding to the positive anchors. And λ is loss weight

of $£_{reg}$.

$$P_i^* = \begin{cases} 1, & IOU > \tau_1 \\ 0, & IOU < \tau_2 \\ none, & otherwise \end{cases} \tag{3}$$

We set $\tau_1 = 0.7$, $\tau_2 = 0.3$. $£_{cls}$ is softmax loss used to classify the anchor boxes. $£_{reg}$ is used to train bounding box regression.

$$£_{reg}(t_i, t_i^*) = \sum_{x \in \{x,y,w,h\}} smooth_{L1}(t_i - t_i^*) \tag{4}$$

where x, y, w, and h denote the box's center coordinates and its width and height. The $smooth_{L1}(x)$ is defined as

$$smooth_{L1}(x) = \begin{cases} 0.5x^2, & if\ |x| < 1 \\ |x| - 0.5, & otherwise \end{cases} \tag{5}$$

We have controlled the proportion of positive and negative samples through the IOU. For training DSN's segmentation module, we use the Cross-Entropy loss function, because it is simple and effective. The Cross-Entropy loss for one class is given by:

$$£_{c_n} = -y_{c_n} log\widehat{y_{c_n}} - (1 - y_{c_n}) log(1 - \widehat{y_{c_n}}) \tag{6}$$

where n is the class label assumed to be in the range $1, 2, ..., N$, c_n is the *n-th* label class, $\widehat{y_{c_n}}$ is the *n-th* prediction value. And the loss function of multiclass is defined as

$$\begin{aligned} £ &= \sum_{n=1}^{N} £_{c_n} \\ &= -\sum_{n=1}^{N} (y_{c_n} log\widehat{y_{c_n}} + (1 - y_{c_n}) log(1 - \widehat{y_{c_n}})) \end{aligned} \tag{7}$$

So, our binary classification loss function is defined as

$$£ = \sum_{n=1}^{N} £_{c_n} = -\sum_{n=1}^{N} y log\widehat{y} \tag{8}$$

B. MPFN

The prediction of DSN contains many false positive clusters which have similar appearance with the small stroke lesions. So we use DSN to segment data from three different plane view, and design MPFN to combine these segmentation results and generate a more accurate ones. MPFN can greatly suppress errors in DSN prediction. In order use MPFN, the segmentation results must be adjusted to the same plane. Fig. 5 shows the workflow of our method.

The architecture of MPFN is shown in Fig. 6. The input of MPFN is 4 sequences, which are the original image sequence, axial classification probability maps, sagittal classification probability maps, and coronal classification probability maps. MPFN aims to eliminate false recognition artifacts in DSN's prediction. The volumetric slices are spatially related, the closer the distance is, the greater the correlation coefficient. Acute ischemic stroke lesions are spatially

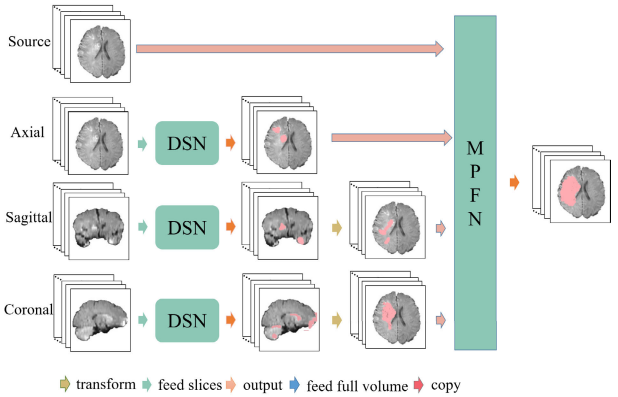


FIGURE 5. The workflow of our method. Arrows in difference colors stand for different operation.

continuous in the brain. So, the segmentation results can be optimized by the information provided by the adjacent slices. Channel Adjustment (CHA) is used to adjust the combination order of input sequences. For convenience of description, let us give an example, if the index of a picture is i, and each sequence is arranged from small to large according to the distance from the *i-th* slice. The rearrange input sequences is defined as

$$\begin{bmatrix} i_i^i & i_i^{i+1} & i_i^{i-1} & i_i^{i+2} & \cdots \\ i_a^i & i_a^{i+1} & i_a^{i-1} & i_a^{i+2} & \cdots \\ i_s^i & i_s^{i+1} & i_s^{i-1} & i_s^{i+2} & \cdots \\ i_c^i & i_c^{i+1} & i_c^{i-1} & i_c^{i+2} & \cdots \end{bmatrix} \tag{9}$$

where i_i is DWI sequence, i_a is axial plane view probability map, i_s is sagittal plane view probability map, i_c is coronal plane view probability map.

This adjustment ensure that the features learned by MPFN include the spatial correlation between the slices. Considering computational power is limited and the farther slices have smaller correlation coefficients with *i-th* slice, which means that they contribute a little to the segmentation result of *i-th* slice. So, the MPFN only take 50 slices from each kinds of sequence at a time.

The adjusted sequences are sent to the fusion module to get the final ischemic stroke lesion segmentation results. The structure of the fusion module is similar to FPN, so it can also capture multi-scale information. But unlike the FPN, it retains feature information on all scales and add a convolution layer after the last final deconvolution layer to generate the final segmentation results.

1) LOSS FUNCTION

For training MPFN, we use the Dice loss function, and the Dice loss is given by:

$$£ = 1 - DICE \tag{10}$$

where $DICE$ is dice cofficient.

$$DICE = Dice(X, Y) = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \tag{11}$$

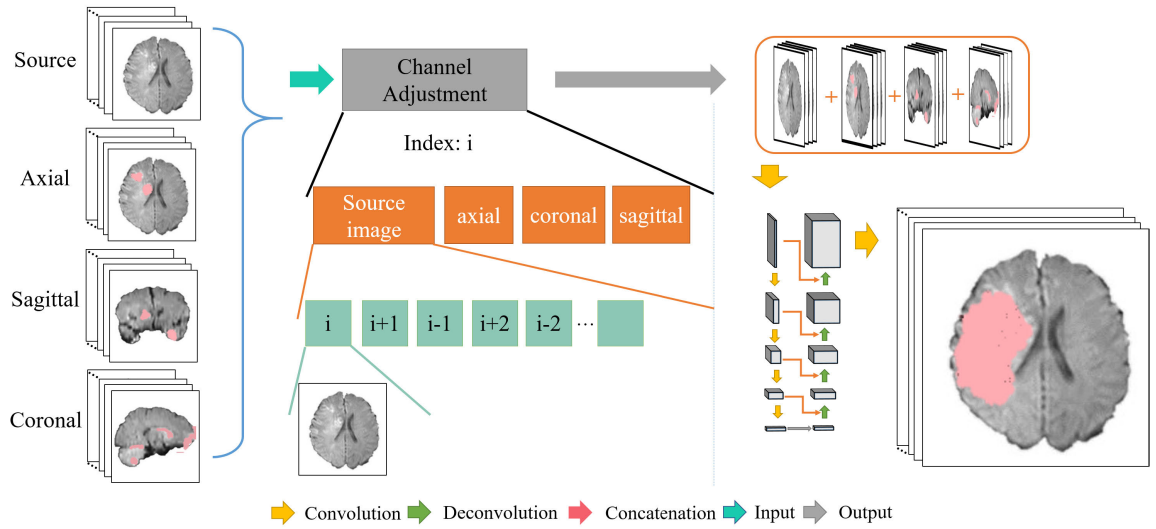


FIGURE 6. The architecture of the proposed MPFN. The orange rectangle represents the volume data and The green box represents a slice of source volume or probabilistic map. The gray rectangles in different sizes indicate data blobs in different sizes. The height shows the size of each piece of data, e.g. 32 × 32. The width shows the number of data pieces in each blob. Arrows in difference colors stand for different operation.

Therefore the loss function is defined as

$$£ = \frac{FP + FN}{2 \cdot TP + FP + FN} \tag{12}$$

C. EVALUATION METHODS

We use many criteria to evaluate our method. The Dice coefficient is used to compare the agreement with neurologists' segmentation results. It measures the overlap between the our segmentation X and the ground truth Y and is defined as

$$Dice(X, Y) = \frac{2\,|X \cap Y|}{|X| + |Y|} \tag{13}$$

IOU is similar to the Dice cofficient, which is defined as

$$IOU(X, Y) = \frac{X \bigcap Y}{X \bigcup Y} \tag{14}$$

To make a comprehensive comparison of model performance, we used five additional pixel level indicators: accuracy, sensitivity, specificity, recall and F-Measure as follows:

1) **Accuracy:**

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{15}$$

2) **Sensitivity:**

$$Sensitivity = \frac{TP}{TP + FP} \tag{16}$$

3) **Specificity:**

$$Specificity = \frac{TN}{TN + FP} \tag{17}$$

4) **Recall:**

$$Recall = \frac{TP}{TP + FN} \tag{18}$$

5) **F-measure:**

$$F - measure = \frac{(1 + \beta^2) \times Precision \times Recall}{\beta^2 \times (Precision + Recall)} \tag{19}$$

where β is a coefficient that adjusts the proportion of precision and recall.

D. IMPLEMENTATION DETAILS

The DSN and MPFN in this paper are implemented with Keras 2.0.8 and Tensorflow-1.4.1 in python 3.6 with NVIDIA GeForce GTX 1080 Ti GPU. The optimisation during training is achieved using the standard stochastic gradient descent algorithm. The learning rate is fixed as 0.001. The momentum and the weight decay is set to 0.9 and 0.0001, respectively. The weights in networks are initialized using the xavier algorithm [42]. We use resnet101 [43] as the backbone of FPN, U-NET's depth is 3. The filter size of the convolution layers are 3 × 3, the deconvolution layers are 2 × 2 and the stride is 1. The weight for detecting loss and segment loss is set to 1 and 2, respectively.

IV. DATA

In order to explain the performance gains brought by our model, we construct 2D and 3D datasets. The 2D dataset is used to verify the DSN module, and the 3D dataset is used to test the existing 3D segmentation models.

A. DATASET DESCRIPTION

In this paper, we have performed a series of comparison experiments on ISLES2015 SSIS DWI sequence dataset to confirm the validity and practicality of our method. The dataset details is shown in Table. 1. These files are all Uncompressed Neuroimaging Informatics Technology Initiative (NIfTI) format (Siemens) with the following acquisition

TABLE 1. Description of datasets.

Item	Parameters	Description
Patients Number	n_0=0, n_1=28	0=none, 1=stroke
Slices Number	n_0=3228, n_1=1056	0=none, 1=stroke
Lesion volume	μ_v=17.59 ml	min=0.001ml, max=346.1 ml
Lesion pixel ratio	μ_r=0.88%	min=0.001%, max=9.5%
Ventricular enhancement observable	n_0=18, n_1=5, n_2=1	0=none, 1=slight, 2=strong
Midline shift observable	n_0=25, n_1=2, n_2=0	0=none, 1=slight, 2=strong
Affected artery	n_1=3, n_2=20, n_3=5, n_4=3, n_5=0	1=ACA, 2=ACM, 3=ACP, 4=BA, 5=other
Lesion localization (lobes)	n_1=3, n_2=12, n_3=21, n_4=8, n_5=9, n_6=3	1=frontal, 2=temporal, 3=parietal, 4=occipital, 5=midbrain, 6=cerebellum
Lesion localization (cortical/subcortical)	n_1=15, n_2=21	1=cortical, 2=subcortical
Non-stroke white matter lesion load	μ_l=1.34	0=none, 1=small, 2=medium, 3=large
Laterality	emphn$_1$=9, n_2=19, n_3=0	1=left, 2=right, 3=both

parameters: image Dimensions: 230 × 230 × 153; voxel Dimensions: 1.0 mm × 1.0 mm × 1.0 mm; slice spacing: 1.0 mm. The acute ischemic lesions were annotated by at least three experienced experts. We use 21 (3213 images) of them to train and validate the model, the remaining 7 (1071 images) are used for testing only.

B. PRE-PROCESSING

In magnetic resonance (MR) image analysis, an essential step is the normalization of illumination. The presence of intensity inhomogeneity is common in MR images. Although this problem has minor effects on human visual image understanding due to high ability of human observation to deduce the image content from the distorted image, its impact on automated systems cannot be neglected. We try to solve this problem by using N4Bias correction [44]. On the other hand we want to keep more original information. So, we use the following three methods to process the dataset:

1) Data with N4bias correction
2) Mapping to the UINT8 data type directly
3) Three window levels and window widths concatenate

We work with neurologist to develop three sets of window levels and window widths, $[\frac{max-min}{2}, \frac{3(max-min)}{5}]$, $[\frac{max-min}{2}, max-min]$ and $[\frac{(max-min)}{4}, \frac{max-min}{2}]$. max is the maximum value of the sequence, min is the minimum value. Visualization results show that these three window levels and window widths can display lesions of different tissues. And the experiments show that the three-window pre-processing method works best.

C. DATA AUGMENTATION

If the training data is generated in the image slice level or lesion instance level, then only a small number of images (patches) are available. In fact of DSN and MPFN have a large number of parameters, it is necessary to generate more images (patches) to train the model. In order to add more data to the training set and make the data set more representative, performing data augmentation is very necessary. In addition, data augmentation applies a variety of transformations to the dataset, which can improve the generalization ability of the model and correct the distorted label data to a certain extent. During the training of the 2D model, the training data is randomly sampled and one or more of the following transformations are randomly applied:

- using origin image
- randomly crop and resize at 1024 × 1024
- randomly horizontally flip
- randomly vertically flip
- randomly affine

3D augmentation is more complex. We design a three-dimensional image augmentation method for them. This kind of image augmentation should consider more logic problems. For example:

- continuous slices use same rotate operations
- consecutive slices' use simillar parameters

TABLE 2. Performance comparison using N4bias preprocessing and without N4bias.The best performance is highlighted in bold.

method	without N4Bias		N4Bias preprocessing	
	Precision	DICE	Precision	DICE
U-NET	**35.35%**	**0.82%**	9.53%	0.33%
FCN [45]	**23.14%**	**0.53%**	10.64%	0.37%
DeepLabv3 [46]	**35.21%**	**1.23%**	26.95%	0.62%
DeepLabv3+ [47]	**16.93%**	**0.39%**	0%	0%
SegResnet [48]	**29.97%**	0.70%	20.15%	**0.71%**
UPERNET [49]	**26.25%**	**0.61%**	13.52%	0.47%

V. EXPERIMENTS AND RESULTS

In this section, we conduct a number of experiments to analyze MULTI-PLANE Information Fusion Network on ISLES2015 SSIS DWI sequence dataset. The acute ischemic stroke lesion segmentation is an essential task, so, we try many existing image segmentation models. Our experiments found that the N4Bias preprocessing method not only fail to improve the learning performance of the model, but also weak the features of the lesions. We compare the characteristics of the data before and after N4bias preprocessing, and find that N4bias algorithm consider the lesions are caused by the intensity inhomogeneities.

Table 2 shows the performance comparison. We can see that the overall results obtained with the N4bias preprocessing are worse than without the N4bias preprocessing.

TABLE 3. Performance comparison of 2D segmentation methods on our task.The best performance is highlighted in bold.

Methods	Detection?	Accuracy	Sensitivity	Specificity	Recall	DICE	IOU	F-measure
U-NET		49.89%	31.67%	99.2%	0.37%	0.73%	0.37%	0.37%
FCN		49.89%	31.6%	99.2%	0.37%	0.73%	0.37%	0.36%
DeepLabv3		49.85%	24.34%	99.12%	0.28%	0.56%	0.28%	0.28%
DeepLabv3+		49.85%	24%	99.12%	0.28%	0.56%	0.28%	0.28%
SegResnet		49.86%	26.51%	99.14%	0.32%	0.62%	0.31%	0.32%
GCN [50]		98.8%	49.7%	0%	0%	0%	0%	0%
PSPNET [51]		97.8%	49.5%	0%	0%	0%	0%	0%
UPERNET		49.9%	30.94%	99.2%	0.36%	0.72%	0.36%	0.36%
Mask RCNN	✓	98.1%	51.58%	99.42%	31.03%	38.75%	24.03%	19.37%
DSN	✓	**98.36%**	**55.14%**	**99.48%**	**35.59%**	**43.26%**	**27.61%**	**21.64%**

Therefore, we will not apply the N4bias preprocessing algorithm in the following experiments.

A. RESULT OF DSN ARCHITECTURE ANALYSIS

In order to illustrate the advantages of our DSN architecture, we test many 2D segmentation frameworks such as U-Net, DeepLab, FCN and etc. The experimental comparisons in this section are performed in a single network. The input to the model is the entire DWI slice, which size is 230×230. All the models use the same preprocessing and image enhancement methods. We try three types of loss function to train them: 1) MSE Loss, 2)Cross-Entropy Loss and 3) Focal Loss [52]. The focal loss function is defined as

$$p_t = \begin{cases} p, & y = 1 \\ 1 - p, & otherwise \end{cases} \tag{20}$$

where p is the probability of the activation layer, y denote the class label. The FocalWeight is defined as

$$FocalWeight = \begin{cases} \alpha(1-p)^\gamma, & y = 1 \\ (1-\alpha)p^\gamma, & otherwise \end{cases} \tag{21}$$

where γ is used to adjust the loss of easily classified samples, α is used to balance the proportion of positive and negative samples.We set $\gamma = 2$ and $\alpha = 0.25$. So, the loss function is defined as

$$\pounds = \sum -\alpha_t (1-p_t)^\gamma log(p_t) \tag{22}$$

Experiments show that focal loss perform better than the other two loss functions significantly. Because our task contains two types of data imbalance. First is image-level, the normal slice more than the ill ones. Second is pixel-level, the normal pixels in a slice more than ill ones. Focal loss can solve picture-level data imbalance through hyperparameter adjustment. Our DSN architecture has controlled the proportion of positive and negative samples through IOU threshold, so there is no need to use focal loss.

The performance of these models on the test dataset is shown in Table 3. It is easy to find that the 2D segmentation model performs poorly on this task, and we think that this phenomenon mainly caused by pixel-level data imbalance, most model structures do not explicitly deal with pixel-level data imbalances. Our DSN segmentation module is also U-NET, but the performance of DSN is far better than U-NET. The difference between our DSN and U-NET is the detection module, this module greatly improves the performance of the U-NET model, because the image patches sent to the U-NET contain relatively large lesions. Detection module in DSN solves two kinds of data imbalance.

Mask-RCNN also has detection and segmention modules, which make it perform better than most of 2D segmentation model. Our DSN is better than Mask-RCNN because DSN dilates the candidate regions generated by RPN and fine-tuned the multi-scale feature maps of FPN. Dilation operation can reduce the loss of information, and the features used for detection may not necessarily suitable for segmentation, so the feature fine tune module brings performance gains.

B. RESULT OF MPFN ARCHITECTURE ANALYSIS

Our data has three-dimensional information, which may improve the model performance. There are several differences between 2D and 3D image segmentation tasks:

1) Data Format: 3D data has z-axis, which can provide extra information. It is obvious that the more z-axis slices the richer information it provide.
2) Model: In general, the model based on 3D convolution has a larger amount of parameters, and the training is more difficult to converge. Therefore, the 3D model requires more data for training, otherwise may lead to cause over-fitting.
3) Task: Some pathologies have no obvious signs on 2D images, but in the three-dimensional space can reveal abnormalities. Some pathologies are sparse in three dimensions, the reception field is limited, so the 2D network may perform better.

Ischemic stroke is mainly caused by cerebral vascular occlusion, so the lesion is continuous in the three-dimensional space, and our DWI data is rich in information on the z-axis. we tried some 3D segmentation model such as UNET3D [53] and V-NET [54]. In order to save the computing expenses,

TABLE 4. Performance comparison of 3D segmentation methods on our task.The best performance is highlighted in bold.

Methods	Detection?	Accuracy	Sensitivity	Specificity	Recall	DICE	IOU	F-measure
V-Net		**99.99%**	42.64%	**99.98%**	0.27%	0.51%	0.27%	0.27%
3D-UNET		99.98%	40.1%	99.98%	45.08%	41.9%	26.54%	21.03%
DeepMedic	✓	99.6%	65.77%	99.86%	49.23%	56.31%	39.19%	28.16%
Our Method	✓	99.98%	**73.22%**	99.97%	**54.1%**	**62.2%**	**45.14%**	**31.1%**

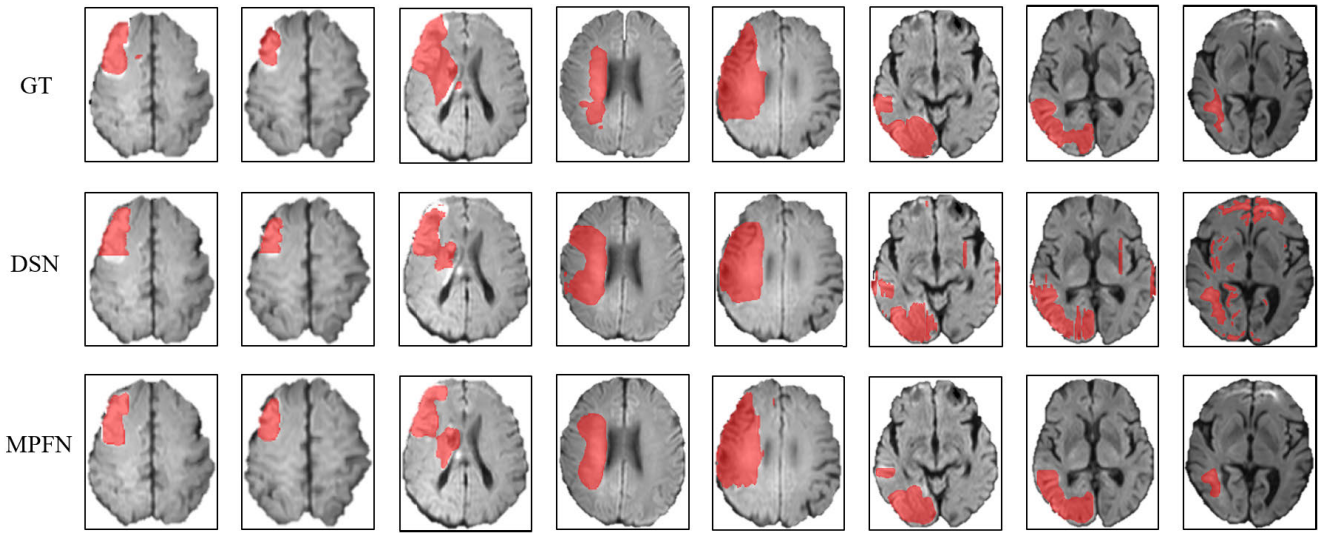


FIGURE 7. Segmentation results, for the observation convenience, we resize the image to a uniform size. The red masks represent the acute ischemic stroke lesions.

we fix the 3D kernel size to $3 \times 3 \times 3$, and dilate the reception field by the downsampling operation. The performance of these models on the test dataset is shown in Table 4. We used the Dice Loss to train.

Because it is difficult to apply focal loss to 3D segmentation models, 3D U-NET and V-NET do not address the two kinds of data imbalance. However, the experiments show that the 3D segmentation model performs better than most of 2D segmentation model. We think the information provided on the z-axis can improve the learning performance. Interestingly, the performance of the DSN and Mask-RCNN that solve the problem of data imbalance perform better than 3D U-NET and V-NET. The above experiments show that addressing data imbalances and introducing 3D information can improve the performance of acute ischemic stroke lesion segmention. So, there are two ideas to optimize the segmentation performance, one is to solve the data imbalance based on the 3D model, and the other is to introduce the 3D information to the DSN. DeepMedic is the realization of the first one, it contains a 3D detection module, MPFN is the realization of the second one, which introduces spatial information by fusing results from triple branches DSN.

Our MPFN performance is better than DeepMedic because it suppresses image artifacts more effectively. By communicating with neurologists, we learned that most image artifacts can be ruled out through multi-view observations. 3D detection module is still susceptible to interference from image artifacts. Therefore, we trained the DSN on the axial, sagittal and coronal planes, and the triple branches DSN identifies the lesions in the images from three directions, then the MPFN fuses the three kinds of results to obtain the final image segmentation results.

VI. DISCUSSION AND CONCLUSION

In this paper, we have presented a novel framework to segment the acute ischemic lesions on DWI. And several visual examples of the segmentation results are shown in Fig. 7. To solve the two kinds of data imbalance, we designed DSN architecture, which includes detection and segmentation modules. The detection one is use to extract patches with IOU greater than 0.7, which can solve the data imbalance problems effectively. Besides, we use the segmentation module to segment the lesions in the extracted patches. Experiments show that our data has rich information on the z-axis, and this information is helpful for lesion segmentation. We designed a triple branches DSN architecture, which trained on three kinds of plane. And the fusion network combines the segmention results. Different from the many other related works, our work combines the advantages of 2D and 3D segmentation models. The evaluation process is conducted using performance metrics like DICE, IOU, and average f1 score.

Extensive experiments on ISLES2015 SSIS DWI sequence dataset demonstrate the high accuracy and efficiency of our proposed segmentation method.

In the future, the experience of the neurologists will be involved to produce the attention-based model. We will try to highlight areas in a given input image that provide useful evidence for lesion segmentation.

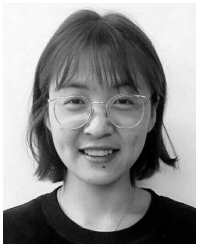
REFERENCES

[1] Wikipedia. *Stroke*. Accessed: Dec. 26, 2019. [Online]. Available: https://en.wikipedia.org/wiki/Stroke
[2] *Stroke Severity and Mortality: Types, Treatments, and Symptoms*. Accessed: Dec. 26, 2019. [Online]. Available: https://www.healthline.com/health/can-you-die-from-a-stroke
[3] X.-L. Liao, C.-X. Wang, Y.-L. Wang, C.-J. Wang, X.-Q. Zhao, L.-Q. Zhang, L.-P. Liu, Y.-S. Pan, and Y.-J. Wang, "Implementation and outcome of thrombolysis with alteplase 3 to 4.5 h after acute stroke in Chinese patients," *CNS Neurosci. Therapeutics*, vol. 19, no. 1, pp. 43–47, Dec. 2012.
[4] Z. Liu, C. Cao, S. Ding, Z. Liu, T. Han, and S. Liu, "Towards clinical diagnosis: Automated stroke lesion segmentation on multi-spectral MR image using convolutional neural network," *IEEE Access*, vol. 6, pp. 57006–57016, 2018.
[5] M. G. Lansberg, G. W. Albers, C. Beaulieu, and M. P. Marks, "Comparison of diffusion-weighted MRI and CT in acute stroke," *Neurology*, vol. 54, no. 8, pp. 1557–1561, Apr. 2000.
[6] D. Smajlović and O. Sinanović, "Sensitivity of the neuroimaging techniques in ischemic stroke," *Medicinski Arhiv*, vol. 58, no. 5, pp. 282–284, 2004.
[7] J. Crinion, A. L. Holland, D. A. Copland, C. K. Thompson, and A. E. Hillis, "Neuroimaging in aphasia treatment research: Quantifying brain lesions after stroke," *NeuroImage*, vol. 73, pp. 208–214, Jun. 2013.
[8] O. Maier, B. H. Menze, J. von der Gablentz, L. Häni, M. P. Heinrich, M. Liebrand, S. Winzeck, A. Basit, P. Bentley, and L. Chen, "ISLES 2015-A public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI," *Med. Image Anal.*, vol. 35, pp. 250–269, Jan. 2017.
[9] 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, Feb. 2003.
[10] M. G. Dwyer, N. Bergsland, E. Saluste, J. Sharma, Z. Jaisani, J. Durfee, N. Abdelrahman, A. Minagar, R. Hoque, F. E. Munschauer, and R. Zivadinov, "Application of hidden Markov random field approach for quantification of perfusion/diffusion mismatch in acute ischemic stroke," *Neurol. Res.*, vol. 30, no. 8, pp. 827–834, Jul. 2013.
[11] Y. Shen, X. He, J. Gao, L. Deng, and G. Mesnil, "Learning semantic representations using convolutional neural networks for Web search," in *Proc. 23rd Int. Conf. World Wide Web-(WWW) Companion*, 2014, pp. 373–374.
[12] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.
[13] P. McClure and F. Khalifa, "A novel NMF guided level-set for DWI prostate segmentation," *J. Comput. Sci. Syst. Biol.*, vol. 7, no. 6, pp. 209–216, 2014.
[14] R. Zhang, L. Zhao, W. Lou, J. M. Abrigo, V. C. T. Mok, W. C. W. Chu, D. Wang, and L. Shi, "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.
[15] Y.-L. Chang and X. Li, "Adaptive image region-growing," *IEEE Trans. Image Process.*, vol. 3, no. 6, pp. 868–872, Nov. 1994.
[16] N. M. Saad, S. Abu-Bakar, S. Muda, M. Mokji, and A. Abdullah, "Automated region growing for segmentation of brain lesion in diffusion-weighted MRI," in *Proc. Int. MultiConf. Eng. Comput. Sci.*, vol. 1, 2012, pp. 14–17.
[17] T.-Y. Lin, P. Dollar, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2117–2125.
[18] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, ch. 6330, pp. 234–241.
[19] C. R. Gillebert, G. W. Humphreys, and D. Mantini, "Automated delineation of stroke lesions using brain CT images," *NeuroImage: Clin.*, vol. 4, pp. 540–548, Jan. 2014.
[20] Y.-S. Tyan, M.-C. Wu, C.-L. Chin, Y.-L. Kuo, M.-S. Lee, and H.-Y. Chang, "Ischemic stroke detection system with a computer-aided diagnostic ability using an unsupervised feature perception enhancement method," *Int. J. Biomed. Imag.*, vol. 2014, Dec. 2014, Art. no. 947539.
[21] A. Wouters, P. Dupont, B. Norrving, R. Laage, G. Thomalla, G. W. Albers, V. Thijs, and R. Lemmens, "Prediction of stroke onset is improved by relative fluid-attenuated inversion recovery and perfusion imaging compared to the visual diffusion-weighted Imaging/Fluid-attenuated inversion recovery mismatch," *Stroke*, vol. 47, no. 10, pp. 2559–2564, Oct. 2016.
[22] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.
[23] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Comput. Electron. Agricult.*, vol. 147, pp. 70–90, Apr. 2018.
[24] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.-M. Jodoin, and H. Larochelle, "Brain tumor segmentation with deep neural networks," *Med. Image Anal.*, vol. 35, pp. 18–31, Jan. 2017.
[25] K. Kamnitsas, E. Ferrante, S. Parisot, C. Ledig, A. V. Nori, A. Criminisi, D. Rueckert, and B. Glocker, "DeepMedic for brain tumor segmentation," in *Proc. Int. Workshop Brainlesion: Glioma, Multiple Sclerosis, Stroke Traumatic Brain Injuries*. Cham, Switzerland: Springer, 2016, ch. 6330, pp. 138–149.
[26] K. Kamnitsas, C. Ledig, V. F. J. 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, Feb. 2017.
[27] 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.
[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, Jan. 2017.
[29] W. Wei, B. Zhou, D. Połap, and M. Woźniak, "A regional adaptive variational PDE model for computed tomography image reconstruction," *Pattern Recognit.*, vol. 92, pp. 64–81, Aug. 2019.
[30] S. Hussain, S. M. Anwar, and M. Majid, "Brain tumor segmentation using cascaded deep convolutional neural network," in *Proc. 39th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2017, pp. 1998–2001.
[31] K. C. Ho, W. Speier, S. El-Saden, and C. W. Arnold, "Classifying acute ischemic stroke onset time using deep imaging features," in *Proc. AMIA Annu. Symp. Proc.*, 2017, p. 892.
[32] Y. Kabir, M. Dojat, B. Scherrer, F. Forbes, and C. Garbay, "Multimodal MRI segmentation of ischemic stroke lesions," in *Proc. 29th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2007, pp. 1595–1598.
[33] H. Khotanlou, O. Colliot, J. Atif, and I. Bloch, "3D brain tumor segmentation in MRI using fuzzy classification, symmetry analysis and spatially constrained deformable models," *Fuzzy Sets Syst.*, vol. 160, no. 10, pp. 1457–1473, May 2009.
[34] C. Feng, D. Zhao, and M. Huang, "Segmentation of ischemic stroke lesions in multi-spectral MR images using weighting suppressed FCM and three phase level set," in *BrainLes*. Cham, Switzerland: Springer, 2015, ch. 6330, pp. 233–245.
[35] H.-L. Halme, A. Korvenoja, and E. Salli, "ISLES (SISS) challenge 2015: Segmentation of stroke lesions using spatial normalization, random forest classification and contextual clustering," in *BrainLes*. Cham, Switzerland: Springer, 2015, ch. 6330, pp. 211–221.
[36] R. Karthik, U. Gupta, A. Jha, R. Rajalakshmi, and R. Menaka, "A deep supervised approach for ischemic lesion segmentation from multimodal MRI using fully convolutional network," *Appl. Soft Comput.*, vol. 84, Nov. 2019, Art. no. 105685.
[37] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó, "Acute ischemic stroke lesion core segmentation in CT perfusion images using fully convolutional neural networks," *Comput. Biol. Med.*, vol. 115, Dec. 2019, Art. no. 103487.
[38] G. Capizzi, G. Lo Sciuto, C. Napoli, D. Polap, and M. Wozniak, "Small lung nodules detection based on fuzzy-logic and probabilistic neural network with bio-inspired reinforcement learning," *IEEE Trans. Fuzzy Syst.*, to be published.
[39] F. Beritelli, G. Capizzi, G. Lo Sciuto, C. Napoli, and M. Woźniak, "A novel training method to preserve generalization of RBPNN classifiers applied to ECG signals diagnosis," *Neural Netw.*, vol. 108, pp. 331–338, Dec. 2018.

[40] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2961–2969.
[41] J. F. Ebersole and J. F. Ebersole, Jr., "Method to aid object detection in images by incorporating contextual information," U.S. Patent 6 578 017, Jun. 10, 2003.
[42] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proc. 13th Int. Conf. Artif. Intell. Statist.*, 2010, pp. 249–256.
[43] 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.
[44] N. J. Tustison, B. B. Avants, P. A. Cook, Y. Zheng, A. Egan, P. A. Yushkevich, and J. C. Gee, "N4ITK: Improved n3 bias correction," *IEEE Trans. Med. Imag.*, vol. 29, no. 6, pp. 1310–1320, Jun. 2010.
[45] 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.
[46] 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, Apr. 2018.
[47] A. R. Choudhury, R. Vanguri, S. R. Jambawalikar, and P. Kumar, "Segmentation of brain tumors using DeepLabv3+," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, 2018, ch. 6330, pp. 154–167.
[48] X. Chen, G. Cheng, Y. Cai, D. Wen, and H. Li, "Semantic segmentation with modified deep residual networks," in *Proc. Chin. Conf. Pattern Recognit.* Singapore: Springer-Verlag, 2016, pp. 42–54.
[49] T. Xiao, Y. Liu, B. Zhou, Y. Jiang, and J. Sun, "Unified perceptual parsing for scene understanding," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 418–434.
[50] 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.*, Jul. 2017, pp. 4353–4361.
[51] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2881–2890.
[52] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal loss for dense object detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020.
[53] Ö. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2016, ch. 6330, pp. 424–432.
[54] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 565–571.

LONG ZHANG (Member, IEEE) received the B.E. degree from the Beijing University of Posts and Telecommunications (BUPT), Beijing, China, in 2017, where he is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering. His research interests include artificial intelligence, image processing, and machine learning algorithm.

RUONING SONG is currently pursuing the Ph.D. degree with the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, China. His current research interests include mobile Internet traffic analysis, cloud computing, and deep learning.



YUANYUAN WANG received the B.S. degree from the Department of Automation, North China Electric Power University, Hebei, in 2017. She is currently pursuing the M.S. degree with the School of Automation Science and Electrical Engineering, Beijing University of Aeronautics and Astronautics (BUAA), Beijing, China. Her research interests include machine learning and computer vision.

CHUANG ZHU (Member, IEEE) received the Ph.D. degree in microelectronics from Peking University, Beijing, China.

He is currently a Lecturer with the Beijing University of Posts and Telecommunications (BUPT), Beijing, where he also leads the Image Processing Group. He is also a member of the Center for Data Science, BUPT. Before that, he was a Postdoctoral Research Fellow with the School of Electronics Engineering and Computer Science, Peking University, from 2015 to 2017. His research interests include deep learning, image processing, multimedia content analysis, and machine learning algorithm optimization. He has published more than 30 publications in international magazines and conferences in these areas, including the IEEE TRANSACTIONS ON MULTIMEDIA, the journal of *Signal Processing: Image Communication*, and the IEEE International Conference on Multimedia and Expo (ICME). He has served on the Technical Program Committee for the International Conference on Computing and Pattern Recognition (ICCPR 2019).

JUN LIU (Member, IEEE) received the B.E. and Ph.D. degrees from the Department of Information Engineering, Beijing University of Posts and Telecommunications (BUPT), in 1998 and 2003, respectively. He is currently an Associate Professor and the Director of the Center for Data Science, BUPT. His research interests include deep learning, big data analysis, and stream data algorithms.

JIE YANG (Member, IEEE) received the B.E., M.E., and Ph.D. degrees from the Beijing University of Posts and Telecommunications, China, in 1993, 1999, and 2007, respectively. She is currently a Professor and the Deputy Dean of the School of Information and Communication Engineering, BUPT. She has published several articles on international magazines and conferences, including the IEEE JSAC, the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS, and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. Her current research interests include broadband network traffic monitoring, user behavior analysis, and big data analysis in the Internet and telecom.

LIAN LIU received the M.D. degree in neurosurgery from Capital Medical University, Beijing, China. He was a Medical Doctor with Capital Medical University, from 2008 to 2011. He is currently an Associate Chief Physician with the Interventional Neuroradiology Department, Beijing Tiantan Hospital. He is also a Lecturer with Capital Medical University. He has published six publications in international magazines in the past years, including *Interventional Neuroradiology*, *Clinical Neurology and Neurosurgery*, and *The Neuroradiology Journal*. His research interest includes interventional treatment of cerebral vascular disease.