IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 24 May 2023, accepted 19 June 2023, date of publication 30 June 2023, date of current version 6 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3289909

RESEARCH ARTICLE

A Novel Multi-Scale Channel Attention-Guided Neural Network for Brain Stroke Lesion Segmentation

ZHIHUA LI[1], QIWEI XING[2], YANFANG LI[1,3], WEI HE[1,3], YU MIAO[1,3], BAI JI[4], WEILI SHI[1,3], AND ZHENGANG JIANG[1,3]

[1]School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China
[2]Institute for Educational Big Data and Artificial Intelligence, Qiongtai Normal University, Haikou 571100, China
[3]Zhongshan Institute, Changchun University of Science and Technology, Zhongshan 528437, China
[4]The First Hospital of Jilin University, Changchun, Jilin 130021, China

Corresponding authors: Qiwei Xing (673659600@qq.com) and Zhengang Jiang (jiangzhengang@cust.edu.cn)

This work was supported in part by the Hainan Province Natural Science Foundation for Youths under Grant 823QN258, and in part by the Jilin Provincial Science and Technology Development Program of China under Grant 20210204101YY.

ABSTRACT Post-stroke neuroimaging is the key to the treatment of brain stroke. Typically, segmenting lesions manually is not only time-consuming, but limited by the varying morphology of lesions and the similarity of tissue intensity distribution. In recent years, with the rapidly and widely applying of deep learning technology in medical imaging, it becomes the one of the hot-spots. However, most stroke segmentation techniques could not utilize the structural symmetry information efficiently. Moreover, there are several problems, such as diverse lesions location, large changes of scale and unclear boundaries of lesions. To address these shortcomings, this paper presents a novel stroke segmentation model, called BSSNet. Firstly, the standard convolution of U-Net is replaced by the Depth-Wise Separable Convolution(DWSC) and the residual connection operation is used to reduce the loss of feature information in the encoder and decoder. Secondly, between the encoder and decoder, this paper introduces the Multi-Scale Channel Attention(MSCA) module to effectively improve the segmentation performance. Finally, the Attention-Guided Connection(AGC) module takes place of the original connection operation, which can select more context information from features with low-level guided by the features with high-level. To demonstrate the advantages of the proposed BSSNet, the comparison experiments are conducted on the open Anatomical Tracings of Lesions After Stroke(ATLAS) datasets. Experiment results shows that our model outperforms the state-of-the-art segmentation methods both in the quantitative metrics and visual effects.

INDEX TERMS Stroke lesion segmentation, ATLAS datasets, depth-wise separable convolution, multi-scale channel attention, attention-guided connection.

I. INTRODUCTION

Stroke is one of the most common cerebrovascular diseases causing death and disability worldwide [1]. Moreover, the disease has many characteristics, such as high mortality, high incidence, high recurrence, high disability rate and high prevalence [2]. Current neuroimaging analysis for stroke rehabilitation mainly uses manual segmentation slice-by-slice by the radiologists, which is time-consuming and relies heavily on subjective perception [3]. There are urgent clinical practice needs for automated computer-aided segmentation methods [4].

In recent years, early stroke segmentation methods extract the features manually from the stroke images and exploit statistical analysis approaches [5], [6], [7], [8], [9]. In addition, these methods have several shortcomings [10]. First, the shape and scale of lesions appear similar. Second, the size and location are different from each other, which limit the accuracy of automated segmentation. Third, some boundaries of stroke lesions confuse the confidential between the

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

stroke and non-stroke regions. Various approaches have been proposed for stroke lesion segmentation, such as feature extraction based on texture features, random forest, support vector machine [11], [12], [13], [14], [15], [16]. However, these methods may still have insufficient feature representation capacity because they depend on hand-crafted features.

Deep learning techniques have been applied in the field of medical image analysis due to representational capabilities of the convolutional neural networks (CNNs), especially the automated brain stroke segmentation methods, which have achieved great success [4]. Among these approaches, the recent success of U-Net [17] facilitates effective stroke segmentation tasks. Therefore, research on U-net has become one of the most popular areas [18], [19], [20], [21], [22], [23], [24], [25]. Kadry et al. [24] used VGG-UNet to study the segmentation of ischemic stroke. Yang et al. [19] proposed a multi-directional RNN that encodes spatially and blurry boundary problem and achieves a for significant refinement. Chen et al. [25] presented a novel segmentation model based on Atrous convolution to preserve context features with different sizes. However, these methods could not capture the non-local relationships. Furthermore, without the non-local relationships, it is difficult to achieve the lesions with different sizes and locations segmentation tasks.

To address these issues, this paper develops a novel segmentation model that introduces the multi-scale channel attention fusion module and the attention-guided concatenation block based on the backbone of the U-Net architecture [17]. Furthermore, to reduce the number of trainable parameters, we replace the convolutional layers with the depth wise convolution operation (DWSC). Due to the effective use of multi-scale channel attention fusion module and the attention-guided concatenation block, we are able to use local feature information to achieve the detection of subtle but critical stroke lesions. Our main contributions are as follows:

(1) We develop a novel deep neural network for stroke segmentation, called BSSNet, which uses special designs, such as AGC module, MSCA module and DWSC block to obtain detailed segmentation labels for stroke lesions.

(2) We employ an AGC module to adaptively extract more contextual features from low-level to high-level spatial attention.

(3) To better find dense contextual information and extract features with inconsistent semantics and scales, we propose a multi-scale channel attention fusion(MSCA) module as the functional block, which could further facilitate the performance of stroke segmentation.

(4) This paper replaces the original convolution layer with as the DWSC operation. With DWSC, we could reduce the size of trainable parameters.

(5) We evaluate the proposed network on the open-source datasets Anatomical Tracings of Lesions After Stroke (ATLAS) datasets. The superiority of our method is demonstrated in comparison with state-of-the-art approaches.

Table 1 describes the symbols and abbreviations used in this paper.

TABLE 1. Description of the abbreviations used in the paper.

NO	Symbol	Description
01	BSSNet	Brain stroke segmentation network
02	DWSC	Depth-wise separable convolution
03	MSCA	Multi-scale channel attention module
04	AGC	Attention-guided connection
05	ATLAS	Anatomical tracings of lesions after stroke
06	DWSC$(\cdot)$	The Depth-wise separable convolution operation
07	$H_{up}(\cdot)$	The Up-sampling block operation
08	$F_{d,j+1}$	The output of previous stage operation
09	SAB$(\cdot)$	The spatial attention block operation
10	$Cov_1(\cdot)$	The convolution layer operation
11	Concat$(\cdot)$	The concatenation operation
12	PWConv$(\cdot)$	The point-wise convolution operation
13	g$(\cdot)$	The global average pooling operation
14	$L_{DC}(\cdot)$	The dice coefficient loss function operation
15	$L_{CE}(\cdot)$	The cross-entropy loss function operation
16	f_h'	The generated by transposed convolution layer operation
17	f_l'	The refined low-level features
18	$F_{e,i}$	The DWSC block of the next stage operation
19	$F_{e,i}^m$	The feature obtained by the max-pooling layer
20	f_h	The high-level features
21	f_l	The low-level features
22	B	The batch normalization operation
23	δ	The ReLU activation function operation

II. RELATED WORK

A. DEEP LEARNING-BASED SEGMENTATION MODELS

Deep learning-based segmentation models can address the shortcomings of the traditional machine learning models, and have achieved a great success while showing a great potential in medical image analysis [26]. In particular, segmentation approaches using the encoder-decoder architecture [17], [27], [28], [29], [30], [31] is one of the research hotspots in the field of medical segmentation.

In recent years, deep learning-based methods have been widely used in the field of image analysis [32]. To use the contextual features of spatial information effectively, Olaf et.al [17] presented the skip connection into the encoder-decoder based on the original U-Net and improved the segmentation performance. Later researchers found that the residual units could facilitate the propagation of spatial information and make the training stage easier. As a result, the ResUNet became one of the most popular networks [33]. Next, in order to improve the segmentation accuracy, the extended U-Net model are improved by introducing the densely connected convolutional operation [34]. Because of up-sampling stages

in the decoder for features with low-resolution, there is no need to introduce the up-sampling operation in learning process. Besides, to effectively use global context features, the pyramid scene parsing network (PSPNet) was proposed, which could solve the problem of scene parsing [35]. Kamnitsas et al. [32] proposed a novel DeepMedic architecture to integrate local and context information and introduced a fully-connected conditional random field (CRF) in the post-processing stage.

With the rapid development of CNN techniques, many researchers utilized CNN model to achieve the task of brain stroke segmentation [4]. Clèrigues et al. [36] proposed a 3D asymmetric encoder-decoder architecture, which consists of local and global residual connections. Liu et al. [37] presented two subnetworks based on the U-Net model and replaced the convolutional layer with a dense block. Kadry et al. [24] improved the performance of VGG-Net for the task of stroke lesions segmentation. Wang et al. [38] proposed a novel network called consistent perception generative adversarial network(CPGAN) to facilitate the semi-supervised stroke lesion segmentation. Hao et al. [19] proposed a cross-level fusion and context inference network(CLCI-Net) to automatically segment the chronic stroke lesion from T1-weighted MR images. Zhou et al. [39] proposed a novel segmentation model, called D-UNet, which combines the advantages of 2D and 3D CNN modules. In addition, they designed the novel loss function, called enhanced mixing loss(EML). The X-Net was proposed to better achieve segmentation for brain stroke lesion [40]. Hui et al. [21] introduced a novel method with the partitioning-stacking prediction fusion(PSPF) techniques, who achieved precise segmentation for 3D stroke. Zhang et al. [23] combined multiple inputs from the gray white matter and lateral ventricles. Furthermore, they achieved a better segmentation performance. Unfortunately, the above brain segmentation methods generally ignore the important non-local relationship. And without the non-local relationship, it is difficult to segment lesions with different scales and locations.

B. SEGMENTATION BASED ON ATTENTION MECHANISMS

Since the attention mechanism has the strong capability of selecting multiple channel features, it has been widely used in the field of image segmentation [41], [42], [43], [44], [45]. Zhang et al. [41] proposed a coding model with global contextual information and introduced the channel attention into the segmentation mechanism. Zhang et al. [42] presented the progressive attention guided recurrent network for salient object detection. Schlempera et al. [43] proposed the additional attention gate in the network to analyses the medical image and obtain the attention maps with coarser resolution. Taghanaki et al. [46] improved the skip connection layer to extract most discriminative channel features. Zhao et al. [47] incorporated the high-level and low-level features by presenting a pyramid feature attention network. A novel channel mechanism was proposed by Fu et al. [48], which achieved better use of contextual features for segmentation. However, almost only a small number of researches have been focusing on the stroke segmentation by using the attention mechanisms.

In summary, unlike the above approaches, this paper effectively exploits and captures the long-range spatial context information from the inputs. Moreover, we exploit the high-level features by up-sampling to improve the features with low-level and easily preserve more stroke features.

III. METHODOLOGY

In this part, the general architecture of our proposed BSSNet model is described in Section III-A. Next, the implementation details of each module are described in Section III-B-III-D.

A. BSSNet ARCHITECTURE

As mentioned in previous methods [49], the long-range dependencies could be captured by the attention module and the features with multi-channels have been simulated. Therefore, we introduce the attention module in our network to achieve reliable stroke segmentation. In fact, the features with high-level and low-level include the semantic information and detailed information separately. However, most segmentation networks directly connect features with different scales. Meanwhile they do not consider the complementary relationships. Moreover, irrelevant features can easily lead to degraded segmentation performance or even incorrect segmentation results. We take advantage of an attention-guided concatenation(AGC) module and multi-scale channel attention(MSCA) module based on the original U-Net network [17]. Our improved model is called BSSNet, whose architecture is shown in Fig.1.

We take the original brain MR image as input, represented as $\boldsymbol{I}$. $\boldsymbol{M}$ denotes the corresponding ground-truth segmentation results. The encoder part introduces the depth separable convolution block to achieve the feature extraction and the reduction of trainable parameters at the same time. At the beginning of our encoder, we use five cascaded DWSC-block(see III B) with different scale to enhance the deep feature and the capability of extracting features. We denote the different stages of the operation as i ($i = 0, 1, 2, 3, 4$).

$$F_{e,i+1}, F^m_{e,i} = DWSC_i(F_{e,i}) \tag{1}$$

where DWSC$(\cdot)$ represents the DWSC block operation and $F_{e,i}$ is used as inputs to the DWSC block of the next stage operation. Taking the raw inputs as an example, $F^m_{e,i}$ indicates the feature obtained by the max-pooling layer. The output of DWSC block is $F_{e,i+1}$.

The lowest resolution feature maps $F_{m,0}$ is located at the first stage operation, which are concatenated with the highest resolution feature maps $F_{e,4}$ by AGC module (see Section III-C). Here, the AGC operation could be denoted as the function AGC and we refine the input features of decoders $F_{d,0}$ as:

$$F_{d,0} = MSCA(AGC(F_{e,4}, F_{m,0})) \tag{2}$$

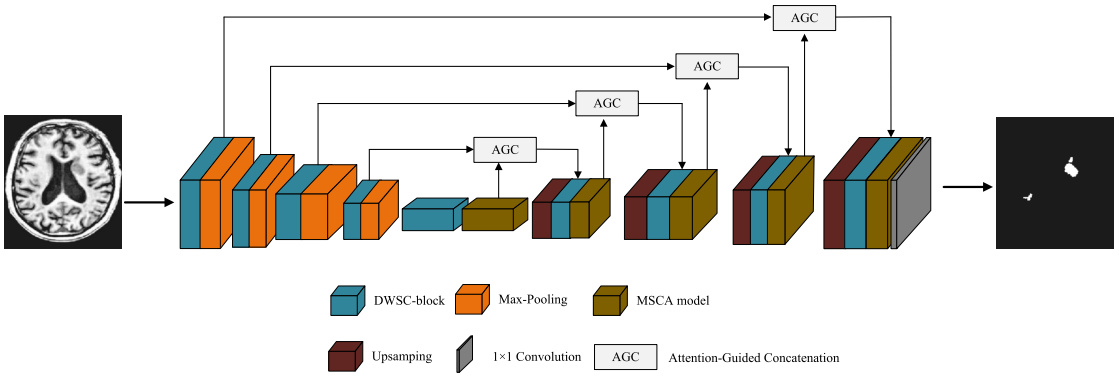

FIGURE 1. The schematic image of our BSSNet architecture to segment stroke lesion. The backbone network (based on U-Net module) consists of three parts. The encoder parts utilize the DWSC-block, including three cascaded depth convolutional layers to take the place of original convolutional layers to achieve the feature extraction and the reduction of trainable parameters at the same time. The MSCA module is employed to capture long-range dependencies and better integrate features with inconsistent semantics and scales. Furthermore, the concatenation operation allowed the extracted features to be better transferred to the decoder part. The decoder parts also employ the DWSC-block to simply the network as much as possible. The input is the original image and the corresponding output is the predicted segmentation results.

where AGC(·) denotes AGC feature concatenation, MSCA(·) is the MSCA module(see Section III-D).

To enlarge the receptive field and improve the capability of feature representation, we introduce the MSCA block to further improve the performance of stroke segmentation. Due to the symmetry of the U-Net structure, the decoder part also consists of five stage operations. For each stage, we exploit the Up-sampling block to reconstruct the original features. Next, the DWSC and AGC modules are employed to concatenate the fusion features generated by the encoder with the up-sampled features. The decoder procedure can be shown as:

$$F_{d,j+1} = ACG(DWSC(H_{up}(F_{d,j})), F_{m,j}) \tag{3}$$

Here, the value of parameter j is token from 0 to 4. $H_{up}(\cdot)$ indicates the Up-sampling block, and $F_{d,j+1}$ denotes the output of previous stage operation. Finally, we use a 1×1 convolutional layer to achieve the prediction of the brain stroke lesions. Table 2 shows the architecture details configuration of the proposed BSSNet.

TABLE 2. The proposed architecture details of BSSNet.

Instruction	Layer name	Number of channels	Kernel size	Size of feature map
Input	Input	-	-	224×192
Encoding phase	Stage-1	32	3×3	112×96
	Stage-2	64	3×3	56×48
	Stage-3	128	3×3	28×24
	Stage-4	256	3×3	14×12
Decoding phase	Stage-1	256	3×3	28×24
	Stage-2	128	3×3	56×48
	Stage-3	64	3×3	112×96
	Stage-4	32	3×3	224×192
Output	Output	1	1×1	224×192

B. DEPTH-WISE SEPARABLE CONVOLUTION BLOCK

U-Net uses standard convolution for feature extraction. However, the standard convolution requires many trainable parameters and the accuracy needs to be improved. The main function of the Depth-Wise Separable Convolution (DWSC) block is to improve the segmentation performance and reduce the number of trainable parameters. Compared to the common convolution operation, DWSC focuses on the spatial and channel directions independently. The details of DWSC block are indicated in Fig2. There are two cascade DWSC layers, with the shape of 3×3 in the DWSC block. For convenience, we set the shape of the convolution layer in the residual connection as 1×1. After each convolution layer, we add the batch normalization block by using the ReLU activation. The input of the DWSC block is the feature map, marked as $\boldsymbol{X}(\boldsymbol{X} \in R^{H\times W\times C})$ and with the DWSC, feature map could be separated into some channels. Here, H represents height; W is the width of feature map and C represents the number of channels. Next, we utilize a 3×3 convolution operation. Then we double the number of output channel with 1×1 convolution. After this, we could obtain the feature map, marked as $\boldsymbol{X}(\boldsymbol{X} \in R^{H\times W\times C_0})$. Specifically, C_0 is equal to $2C$. We use DWSC to perform the convolution in spatial direction on the created feature map. After this, we can obtain the feature map marked as $\boldsymbol{Y}(\boldsymbol{Y} \in R^{H\times W\times C})$. To add $\boldsymbol{X}$ and $\boldsymbol{Y}$, we utilize a 1×1 convolution to compact the feature dimensions. After each convolution layer in the DWSC-block, we utilize batch normalization and ReLU.

C. ATTENTION-GUIDED CONCATENATION MODULE

As we all know, the semantic information is always included in the high-level features while the detail information is preserved in the low-level features. Since the structure of stroke information is always unformed and irregular, it is critical that both using the high-level features and low-level features when achieving accurate stroke prediction. Unfortunately, most of the existing UNet-based methods don't consider the complementary relationship of the features with different scales. In addition, irrelevant features can lead to performance

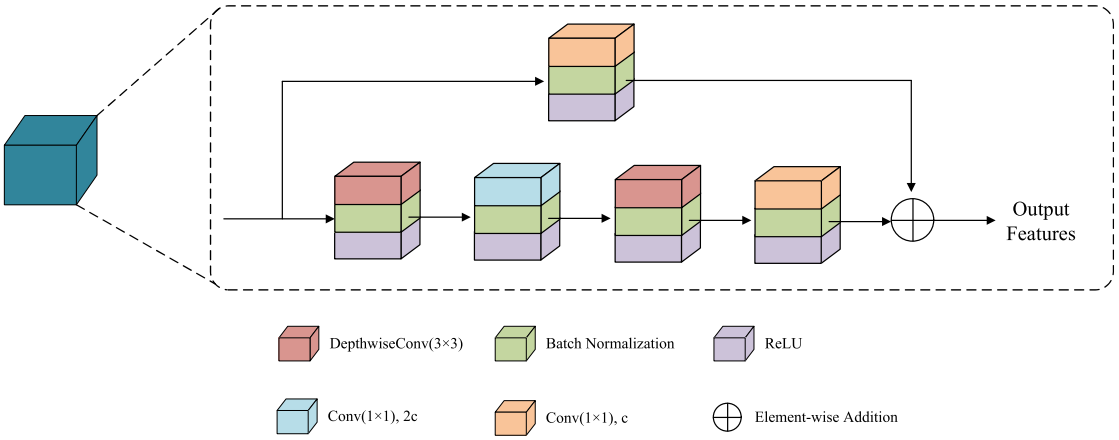

FIGURE 2. The detailed structure of the depth-wise separable convolution (DWSC) block.

degradation or even incorrect segmentation. In order to highlight valuable features, the AGC module is introduced to connect the refined low-level features and the corresponding high-level features. Specifically, due to the different characteristics of the features with different levels, we exploit the spatial attention to select effective information from the low-level features guided by the high-level features. The motivation for applying spatial attention only to low-level features is to focus on selecting more useful details (e.g., stroke edges).

We use the f_h to denote high-level features and f_l to denote low-level features. Illustrated in Fig, the input to the AGC module is f_h and f_l, and the output $\hat{f}_l$ indicated the concatenated feature. In particular, we first employ a transposed convolution layer to up-sample f_h. By up-sampling, the attention map $\boldsymbol{A}$ is generated:

$$\mathbf{A} = SAB(f_h') \tag{4}$$

Here, the spatial attention block is denoted as SAB(·). The feature f_h' is generated by transposed convolution layer with the f_h. The transposed convolution outperforms the bilinear interpolation in up-sampling ability. For ease of calculation, we set the scale of $\boldsymbol{A}$ to be the same as and normalize it and the values in $\boldsymbol{A}$ are in the range [0,1]. Then we could achieve the feature extraction from the feature f_l:

$$f_l' = \mathbf{A} \otimes f_l \tag{5}$$

where f_l' indicates the refined low-level features. $\otimes$ denotes the point-wise multiplication operation. Different from the common methods, the refined low-level features are stacked. In addition, we up-sample the high-level features at the encoder stage. Next, with a convolution layer, we could obtain the output features $\hat{f}_l$:

$$\hat{f}_l = Cov_1(Concat(f_h', f_l')) \tag{6}$$

where $Cov_1(\cdot)$ represents the convolution layer, whose kernel size is 1×1. The Concat(·) function indicates the concatenation.

The purpose of using low-level features in encoder is to enrich the boundary information of the stroke lesion. However, not all features can be used to enrich stroke effectively while our expectation is to focus on the areas where the high-level features details are easily missed. Therefore, the spatial attention block is exploited to preserve the low-level features as many as possible. That is to say, at each spatial location, the spatial attention block enlarges the capability of automatically learning adaptive features selection. The bottom of Fig.3 shows our spatial attention block. We exploit the DWSC block and a 1×1 convolution layer and followed the convolution layer, there is a batch normalized and ReLU activation. The sigmoid operation refines the feature map.

D. MULTI-SCALE CHANNEL ATTENTION MODULE

During the medical image segmentation, the dependency between pixels is an important basis for judging whether they belong to the same class or not. Target areas, e.g., lesion regions, tissue regions, have irregular shape and location. The convolution layers of the neural network are limited by the fixed size of the kernel and receptive field. It will easily lead to the convolution operation not covering the whole target areas. The current implementation of the network model is mainly a simple summation or cascade, but not the best choice. This paper improves the network based on non-local means theory [50], [51]. In particular, we propose a multiscale channel attention model to capture long-range dependencies and scaling features, as shown in Fig.4.

First, in order to fuse the different scale features in target regions, especially in small target regions, we need introduce the feature map $\boldsymbol{Y}(\boldsymbol{Y} \in R^{H \times W \times C})$ into the attention module with multi-scale as the input, shown in Fig.4. Our proposed module utilizes two branches with different scales to extract weights of attention channels. One of the branches takes advantage of global average pooling operation to extract global attention features. The other one uses the point-wise convolution operation to extract local features

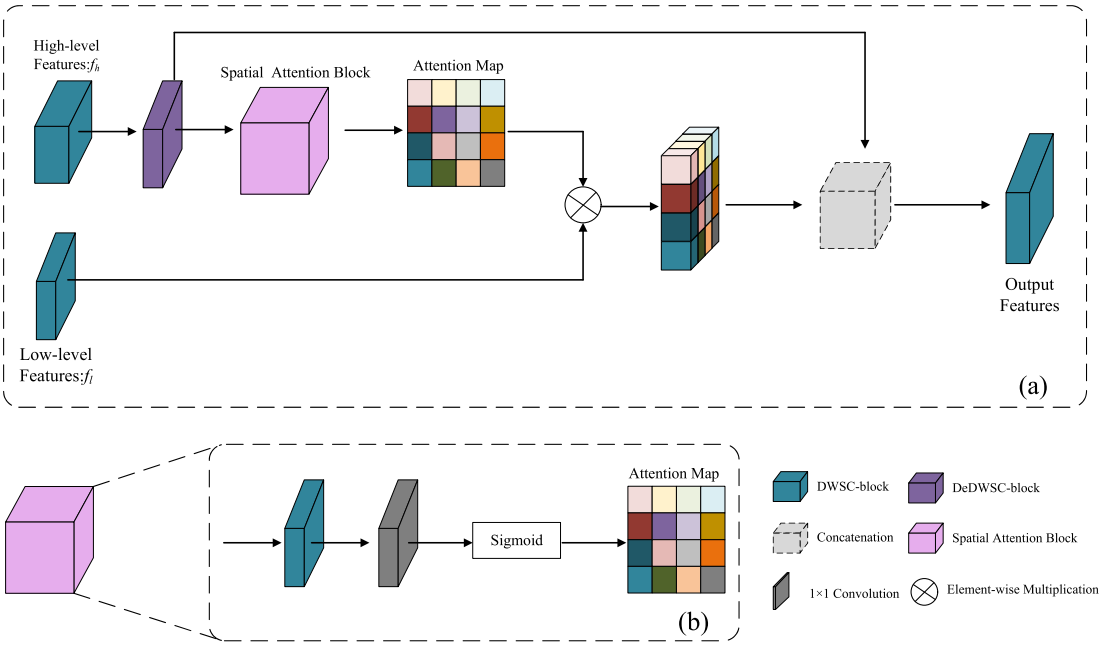

FIGURE 3. (a) The detailed structure of the AGC module, which adaptively selects useful contextual features based on spatial attention and low-level features guided by high-level features. (b) The structure of spatial attention block. The block contains one DWSC block, one convolution layer and one sigmoid operation. The attention map focused on the region that prompts the stroke prediction.

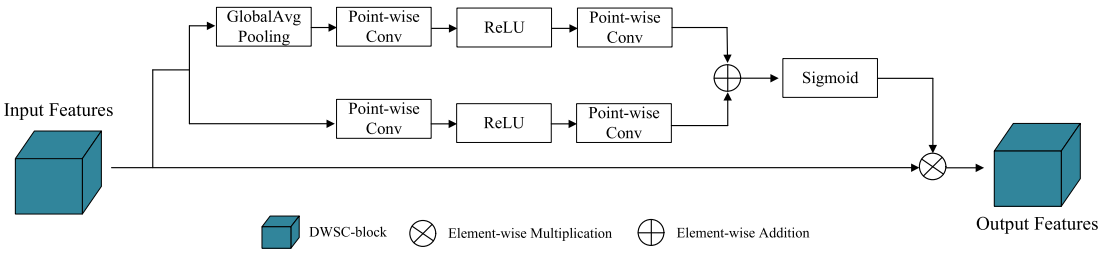

FIGURE 4. The schematic diagram of the structure of our MSCA module.

of attention channels.

$$G(Z) = B(W_2\delta(B(W_1(g(Z))))) \tag{7}$$

$$g(Z) = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X_{[:,i,j]} \tag{8}$$

where G($\boldsymbol{Z}$)(G($\boldsymbol{Z}$)$\in$ R^C) represents the global feature and global average pooling(GAP) operation is denoted as g($\boldsymbol{Z}$). B and σ represent the batch normalization operation and ReLU activity function, respectively. The size of W_1 and W_2 are $\frac{C}{r}\times C$ and $C\times\frac{C}{r}$, respectively. r represents the channel reduction ratio.

Second, the core idea of multi-scale attention mechanism is to extract features of multi-scale attention channel by changing the size of the spatial pooling. In order to keep the model lightweight and save as many parameters as possible, we just present the local contextual clues into the global features within the attention module. At each spatial location, the point-wise convolution (PWConv) is used as the local channel context aggregator. Then we could utilize a bottleneck structure to calculate the local channel context $L(\boldsymbol{X})(L(\boldsymbol{X})\in R^{C\times H\times W})$:

$$L(Z) = B(PWConv_2(\delta(B(PWConv_1(Z))))) \tag{9}$$

where the size of $PWConv_1$ and $PWConv_2$ is $\frac{C}{r}\times C\times 1\times 1$ and $C\times\frac{C}{r}\times 1\times 1$, respectively. If $G(\boldsymbol{Z})$ and $L(\boldsymbol{Z})$ are given, we could obtain the refined feature as follows:

$$Z' = Z\otimes M(Z) = Z\otimes\sigma(G(Z\oplus L(Z))) \tag{10}$$

where σ indicates the sigmoid function and $M(\boldsymbol{Z})$ represents the attention weights generated by the proposed module. $\oplus$ and $\otimes$ represent the broadcasting addition and element-wise multiplication operations, separately.

IV. IMPLEMENTATION

A. IMPLEMENTATION DETAILS

We implemented our model using the PyTorch framework. We trained our proposed BSSNet with the preprocessed

results and ground truth segmentation labels. Throughout the experiments, we set the r of Equ.(7) to 16. Based on our experimental experience, we set the momentum to 0.9, the weight decay to 0.0005, the initial learning rate is set to 10^{-4} and the batch size to 16. And use SGD(stochastic gradient descent) method to optimize the whole network. All models were trained using 150 epchs to optimize the performance of each model. The details of our experimental configurations are shown in Table 3.

TABLE 3. Experimental environment settings.

Configuration	Configuration Model
CPU	Intel(R) Xeon(R) CPU E5-2678 v3 @ 2.50GHz
GPU	NVIDIA GeForce RTX 2080s Ti×4
Operating System	Ubuntu 18.04
Memory	128 GB
Development language	Python 3.6
Development Framework	PyTorch 1.6.0

B. EVALUATION METRICS

In order to evaluate the performance of the segmentation method proposed in this paper, dice coefficient (Dice), intersection over union (IoU), precision and recall are used for segmentation accuracy evaluation.

The dice coefficient (Dice) describes the pixel similarity between the predicted segmentation labels and the corresponding ground truth labels, which could be calculated as:

$$Dice = \frac{2TP}{2TP + FN + FP} \tag{11}$$

Intersection over union (IoU) computes the ratio of intersections and unions between the predicted value and ground truth. The calculation formula of IoU is defined as:

$$IoU = \frac{TP}{TP + FP + FN} \tag{12}$$

Precision could be calculated as follows:

$$Precision = \frac{TP}{TP + FP} \tag{13}$$

The definition of Recall is as:

$$Recall = \frac{TP}{TP + FN} \tag{14}$$

In above equations, TP, TN, FP and FN denote the cases number of true positives, true negatives, false positives and false negatives, respectively. Here, it is notable that the larger Dice, Recall, Precision and IoU, the better segmentation results.

C. LOSS FUNCTION

During the stroke segmentation, due to the impact of the imbalanced data between the positive and negative samples, here this paper combines the dice coefficient (DC) loss function and the cross-entropy (CE)loss function [52]. The detail of the loss function are as follows:

$$L_{total} = L_{DC} + L_{CE} \tag{15}$$

$$L_{DC} = 1 - \frac{2\sum_i y_i \hat{y}_i + \varepsilon}{\sum_i y_i + \sum \hat{y}_i + \varepsilon} \tag{16}$$

$$L_{CE} = -\frac{1}{N}\sum_{i=1}^{N}(y_i log\hat{y}_i + (1 - y_i)log(1 - \hat{y}_i)) \tag{17}$$

where N denotes the pixel number and y_i represents the ground label. And $\hat{y}_i$ is the prediction probability of pixels as the positive samples, ε denotes the minimum constant in order to prevent the divisor from being zero. Based on experimental experience, we set ε to 1.

D. DATASETS AND PRE-PROCESSING

In this paper, we conduct our experiments on the public Anatomical Tracings of Lesion After Stroke (ATLAS, http://fcon_1000.projects.nitrc.org/indi/retro/atlas.html) dataset, which includes 229 chronic stroke cases with T1-weighted normalized MR images from 11 cohorts worldwide. The radiologists manually segmented ground truth stroke regions with MRIcron and there is at least one identified stroke lesion region in each MRI. The voxel spacing of the dataset is at $0.9 \times 0.9 \times 3.0$ mm. All the cases consist of 189 slices and the size is uniformly set to 233×197. That is to say, the ATLAS dataset includes totally 43281 2D slices.

Inspired by the previous idea [4], before training, we use two steps pre-process operation to the dataset. First, 229 subjects(contains of 149 training examples, 34 validation examples and 46 testing examples) are randomly selected. Then, the input images are crop uniformly into a resolution of 224×192. As can be seen from the cropped results, there is only a small angular deviation in part images and these can be negligible.

V. RESULTS AND DISCUSSION

A. COMPARISON WITH STATE-OF-THE-ART METHODS ON ATLAS

Here, we conduct the comparison experiments between ours and other state-of-the-art stroke segmentation methods, such as U-Net [17], 3D+2D CNN [53], MI-UNet [23], e-UNet [54] and AGMR-Net [55]. We quantitatively demonstrate the superiority of our BSSNet with the metrics Dice, Recall, Precision and IoU. The experimental results are shown in Table 4.

As Table.4 shows, BSSNet has the best overall results. Our evaluation score is 0.653(Dice), 0.621(Recall), 0.783(Precision) and 0.525(IoU), separately. Compared with other segmentation methods, our method shows the superior. Our BSSNet achieves better quantitative results than those of U-Net. MI-UNet and e-UNet both have good performance, although no prior is used. Since the more complex structure, the higher model capability, the AGMR-Net method is able to

improve the segmentation performance, which applies lots of prior knowledge of brain parcellation information. However, the more complex the network model, the more difficult the training process. The intuitive idea of proposed BSSNet is that a quick and easy, which is to look for feature information. In addition, the feature information is included in the dataset itself and no additional auxiliary information is required. To demonstrate the stability and show the distribution, we have drawn the sample distribution schematic diagram and box plot of the Dice score shown as Fig.5 and Fig.6. As previously suggested [19], Fig.5 and Fig.6 show the Dice statistical distribution plots for different models. Our final evaluation results are calculated according to the denser distribution. In other words, the higher Dice values, the more superior the segmentation performace of models. Seen from the Fig.5 and Fig.6, our method has a denser distribution at high Dice values than other state-of-the-art models. It is illustrated that our method has superior segmenation performance on the overall data, not limited to individual samples. Furthermore, seen from the Fig.6, we set the lower edge as 0 for all methods since the ATLAS dataset has many small lesion areas. From the upper edge and median line, the proposed BSSNet obtains the best results compared with other approaches.

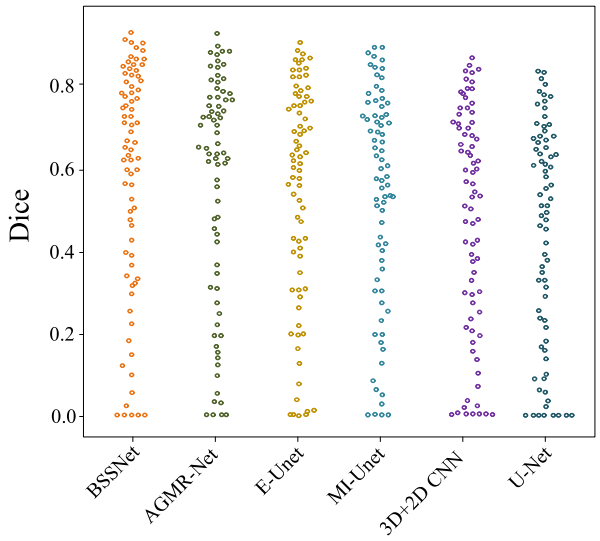

FIGURE 5. The results of the sample distribution schematic Dice score comparison between the proposed BSSNet and other state-of-the-art methods.

Fig.7 is the visual schematic of the brain segmentation on the ATLAS dataset. Shown in figure, there are still some problems, like loss of details, blurred boundaries, in the results of other state-of-the-art approaches. Note that it is important for timely treatment that the subtle lesions need to be precise detected. On the other hand, it can be inferred from the visual schematics that the proposed BSSNet could achieve better segmentation of stroke lesions in T1-weighted MRI. Experiments results further illustrate that our BSSNet outperforms other state-of-the-art methods.

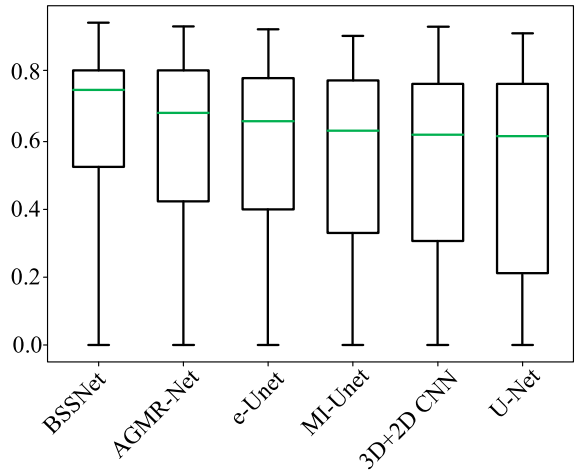

FIGURE 6. The corresponding box plot Dice score comparison results between the proposed BSSNet and other state-of-the-art methods.

TABLE 4. Quantitative comparison results between the proposed network and other state-of-the-art segmentation approaches for brain stroke.

Methods	Dice	Recall	Precision	IoU
UNet	0.497	0.456	0.625	0.362
3D+2D CNN	0.563	0.510	0.629	-
MI-UNet	0.567	0.594	0.655	-
e-UNet	0.592	0.523	0.777	0.455
AGMR-Net	0.594	0.579	0.713	0.468
BSSNet(Ours)	0.653	0.621	0.783	0.525

B. ABLATION STUDIES OF LOSS FUNCTION

In order to verify the effectiveness of loss functions, which makes the proposed BSSNet model converge faster and more efficiently. In this paper, three loss functions are applied for ablation experiments. As shown in the Table 5, it can be seen that the L_{DC} loss function obtains the highest values of 0.655 and 0.624 in Dice and Recall, respectively, while the $L_{DC} + L_{CE}$ loss function used in this paper is slightly lower than the L_{DC}, but during the experiments, this paper considers it to be within an acceptable range value. Moreover, in terms of accuracy metrics, the $L_{DC} + L_{CE}$ has the highest Precision value (0.783), while the L_{DC} has the lowest Precision value (0.629). Therefore, it indicates that the proposed BSSNet model using the $L_{DC} + L_{CE}$ loss function can effectively reduce the false positive results and competitive with the L_{DC}, L_{CE} loss function in terms of segmentation performance.

C. ABLATION STUDIES OF DIFFERENT MODULES

We perform ablation experiments to demonstrate the effectiveness of the main components of our BSSNet on the ATLAS dataset. Here, we set up the definition of Baseline, B+DWSC, B+AGC, B+MSCA, B+AGC+MSCA, B+DWSC+AGC, B+DWSC+ MSCA, B+DWSC+AGC+ MSCA(BSSNet) are as follows:

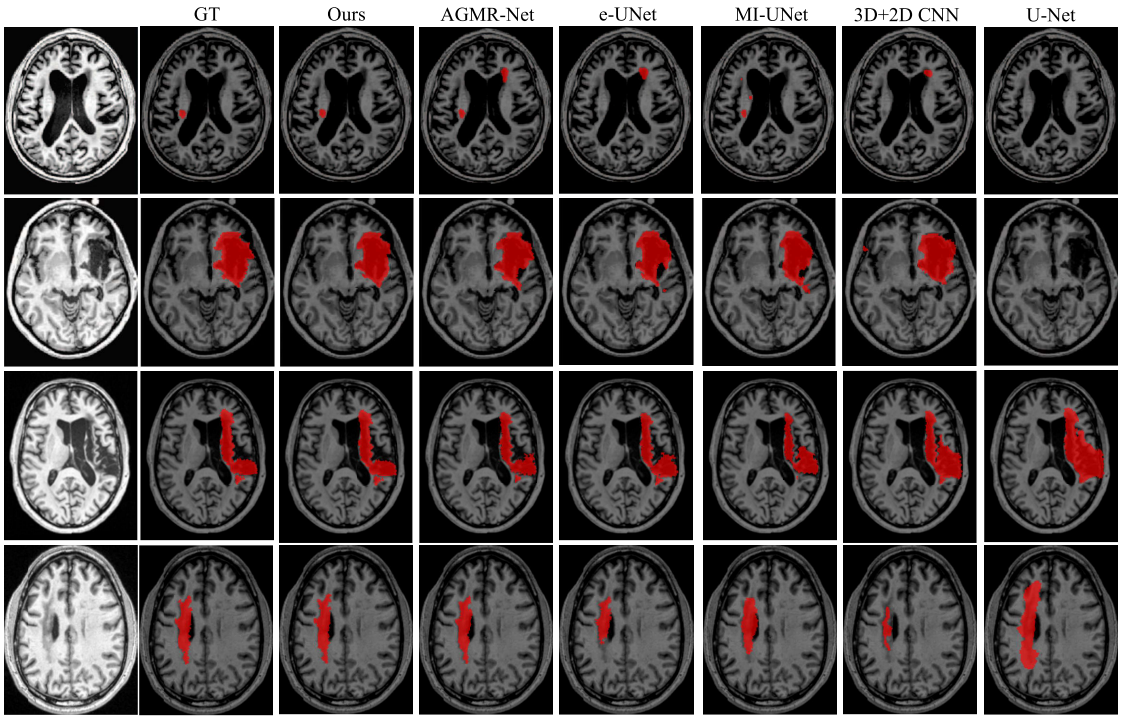

FIGURE 7. The Visual schematic of the brain stroke segmentation on the ATLAS dataset. As shown in figure, we select four stroke segmentation cases with different scale. The first columns indicate the ground truth results. The segmentation results of our proposed BSSNet, AGMR-Net, e-UNet, MI-UNet, 3D + 2D CNN and U-Net are list from the second to the sixth column, separately.

TABLE 5. Ablation analysis of three loss functions on BSSNet model.

Loss Function	Dice	Recall	Precision	IoU
L_{DC}	0.655	0.624	0.629	0.510
L_{CE}	0.584	0.554	0.633	0.486
$L_{DC} + L_{CE}$	0.653	0.621	0.783	0.525

- Baseline: B denotes the U-Net [17] baseline model.
- B+DWSC: Indicates that the standard convolution in the U-Net baseline model is replaced by the DWSC module.
- B+AGC: Indicates that the long skip connection used in the baseline model is replaced by the AGC module.
- B+MSCA: The baseline model U-Net architecture with the MSCA module added.
- B+AGC+MSCA: Indicates that the long skip connection used in the baseline model is replaced by the AGC module, and the MSCA module is added.
- B+DWSC+AGC: Indicates that the long skip connection used in the B+DWSC network is replaced by the AGC module.
- B+DWSC+MSCA: Indicates that the MSCA module has been added to the B+DWSC network.
- B+DWSC+AGC+MSCA: The BSSNet model proposed in this paper.

The comparative results of the ablation experiments are shown in Table 6. As shown in the table, our proposed network BSSNet obtained better 0.653 (Dice), 0.621 (Recall), 0.783 (Precision) and 0.525 (IoU). The specific analysis is as follows:

TABLE 6. Ablation analysis for MSCA, AGC and DWSC on the Baseline.

Serial number	Methods	Dice	Recall	Precision	IoU
Net-1	Baseline	0.497	0.456	0.625	0.362
Net-2	B+DWSC	0.606	0.537	0.588	0.436
Net-3	B+AGC	0.592	0.547	0.589	0.430
Net-4	B+MSCA	0.611	0.556	0.591	0.443
Net-5	B+DWSC+AGC	0.623	0.573	0.642	0.452
Net-6	B+DWSC+MSCA	0.648	0.591	0.744	0.498
Net-7	B+AGC+MSCA	0.641	0.564	0.741	0.481
Ours	BSSNet	0.653	0.621	0.783	0.525

1) COMPONENT ANALYSIS OF THE MSCA MODULE

For stroke segmentation, accurately predicting stroke lesions of different scales is a challenge. Based on the different information, we believe that fusing multi-scale attention features can be more effective for enhancing feature representation and stroke segmentation. As shown in Table 6, the proposed MSCA module can improve the segmentation performance of the model. For example, Net-1 vs Net-4, the latter obtained better Dice (0.611), Recall (0.556), Precision (0.591) and IoU (0.443) than the former in terms of evaluation metric

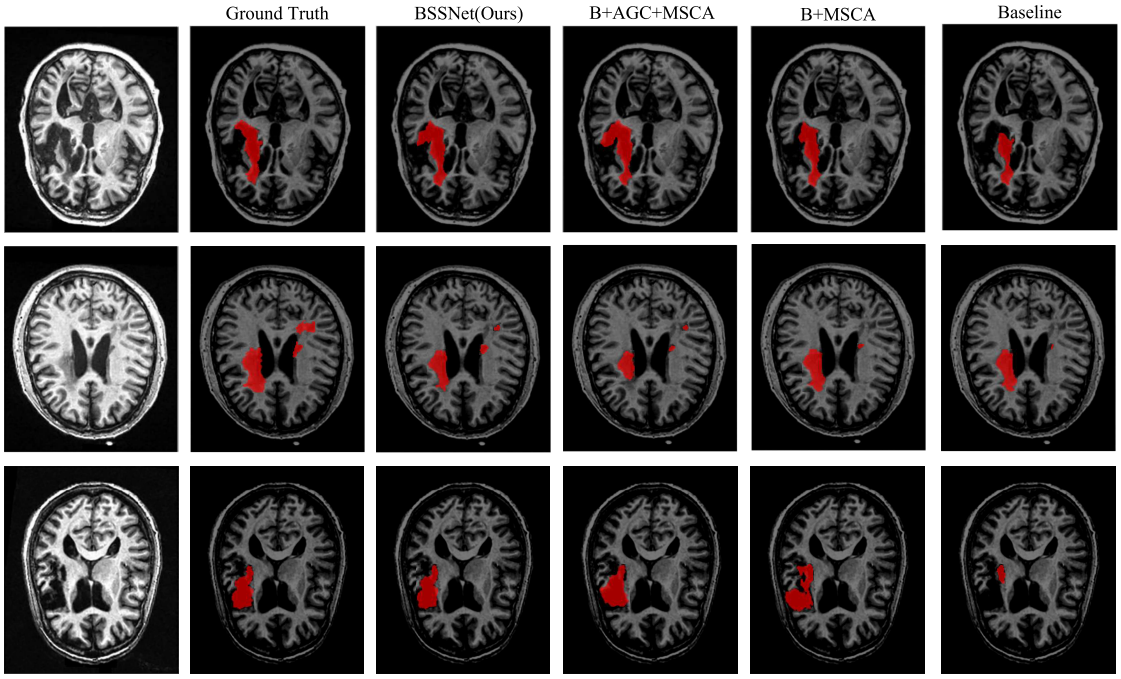

FIGURE 8. Visual comparison results: input images, ground-truths, BSSNet, B +AGC+MSCA, B+MSCA and Baseline.

values. Net-2 vs Net-6, the latter obtained better Dice (0.648), Recall (0.591), Precision (0.744) and IoU (0.498) and Net-3 vs Net-7 who obtained better Dice (0.641), Recall (0.564), Precision (0.741) and IoU (0.481). It suggests that MSCA module is better at preserving and transferring spatial features efficiently to improve segmentation performance. That is because MSCA module can extract both the low-level and high-level feature information with the different scale convolution operation.

2) COMPONENT ANALYSIS OF THE AGC MODULE

AGC block is designed to refine the low-level feature representation guided by high-level features. The low-level features are critical to the high-level features when capturing more detail information. As shown in Table 6, Net-1 vs Net-3, where the latter obtained better Dice (0.592), Recall (0.547), Precision (0.589) and IoU (0.430) than the former in terms of evaluation metric values, Net-2 vs Net-5, where the latter obtained better Dice (0.623), Recall (0.573), Precision (0.642) and IoU (0.452), and Net-4 vs Net-7, where the latter obtained better Dice (0.641), Recall (0.564), Precision (0.741) and IoU (0.481). These results show that the AGC modules has the advantages in prompting the segmentation performance.

3) COMPONENT ANALYSIS OF THE DWSC BLOCK

For brain stroke segmentation, there are two challenging problems. One is to achieve the precise labeling the stroke regions of different scales. The other is to ensure the efficiency of segmentation as much as possible. As the previous conclusion [56], DWSC operation is necessary to ensure both better segmentation results and less trainable parameters. The experimental results are provided in Table 6. Net-1 vs Net-2, where the latter obtained better Dice (0.606), Recall (0.537), Precision (0.588) and IoU (0.436) than the former in terms of evaluation metric values. Net-3 vs Net-5, where the latter obtained better Dice (0.623), Recall (0.573), Precision (0.642) and IoU (0.452) and Net-4 vs Net-6, where the latter obtained better Dice (0.648), Recall (0.591), Precision (0.744) and IoU (0.498). These results show that the DWSC blocks could efficiently fuse and transfer features mappings to improve segmentation performance.

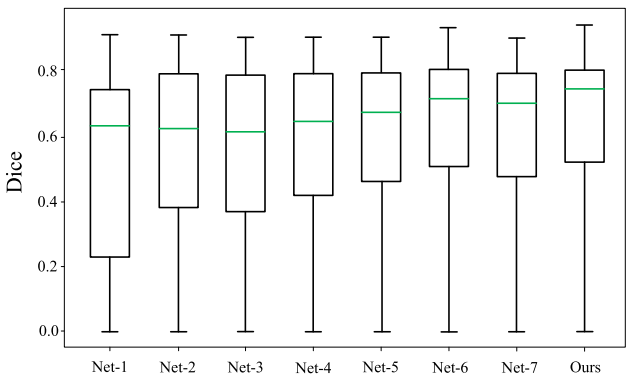

FIGURE 9. Box line diagram of different models.

Besides, we selected the Net-1(Baseline), Net-4(B + MSCA), Net-7(B + MSCA + AGC) to be compared with our proposed BSSNet model. And the visual comparsion results can be seen in Fig.8.Seen from the figure, we can see that the segmentation results of the model in this paper

have better performance than the combined model. Specifically, Baseline, B+MSCA+AGC and B+MSCA incorrectly identified normal tissues as stroke lesion regions, while our model predicted more accurate results. For small stroke lesion regions, our model showed greater ability in identifying and segmenting them. For large stroke lesion regions, our model can obtain more detailed edge information. Also, we visualize the box-line diagram of the above model. As shown in Fig.9, the excellent segmentation performance of our model is confirmed from the box-line plots.

In summary, we compared our proposed method with different networks, which consists of different modules. From the experimental results, we achieved the best results, which demonstrates the effectiveness of all components.

D. DISCUSSION

In this paper, we propose a BSSNet network to achieve automatically segmenting chronic stroke lesions from TI-weighted MR images. The effectiveness of our proposed BSSNet is further demonstrated by the analysis results of ablation experiments in Table 5, Table 6, Fig.8 and Fig.9. Seen from Table 4, Fig.5, Fig.6 and Fig.7, the advantages of our BSSNet model is demonstrated. Compared with the state-of-the-art methods (U-Net, 3D+2D CNN, MI-UNet, e-UNet and AGMR-Net), BSSNet model in this paper outperforms in terms of evaluation metrics ((Dice) 0.653, Recall (0.621), Precision (0.783) and IoU (0.525)). Meanwhile, the results are visualized in this paper shown in Fig.5, Fig.6 and Fig.7. It is clearly suggested that our method shows superiority compared to other segmentation methods.

Seen from Table 5, the L_{DC} loss function obtains the highest value in Dice and Recall, respectively, and the $LDC+L_{CE}$ loss function used in this paper is slightly lower than the L_{DC}, which is considered to be within an acceptable range of values during the experiments. Although the L_{DC} obtained the highest value in Dice and Recall, respectively, its Precision value was the lowest, while the Precision value in this paper was the highest. Therefore, it indicates that the $LDC + L_{CE}$ loss function used in the proposed BSSNet model can effectively reduce the false positive results.

From Table 6, Fig.8 and Fig.9, various different ablation studies as well as visualizations of DWSC, AGC and MSCA modules are performed in this paper to verify the effectiveness of DWSC, AGC and MSCA modules on our BSSNet model. The results of the ablation experiments of Net-1 vs Net-4, Net-2 vs Net-6 and Net-3 vs Net-7 show that MSCA module is better at preserving and transferring spatial features efficiently to improve segmentation performance. That is because MSCA module can extract both the low-level and high-level feature information with the different Second, the results of the ablation experiments of Net-1 vs Net-3, Net-2 vs Net-5 and Net-4 vs Net-7 show that the AGC module can select useful content from low-level features guided by high-level features and has an advantage in prompting segmentation performance. Finally, the results of the ablation experiments with Net-1 vs Net-2, Net-3 vs Net-5 and Net-4 vs Net-6 show that the DWSC module can reduce the loss of feature information in the encoding and decoding stages and can effectively fuse and transfer feature mappings to improve the segmentation performance.

VI. CONCLUSION

In this paper, we propose a novel stroke lesion segmentation neural network, called as BSSNet. With the systematically and thoroughly studies, we better exploit the comprehensive prior knowledge to prompt stroke detection. We utilize a series of techniques to extract the major characteristics of stroke lesions and reduce the trainable parameters. First, the original images are fed into an encoder network with DWSC block. The DWSC block is used to highlight the deep feature information and facilitate the feature learning capability of network. Next, the MSCA module is used to further preserve more necessary feature information. Then, we replace the original concatenation of U-Net with the AGC block, which could adaptively select more contextual clues and complementary features from the low-level feature maps. From the results of our extensive experiments, our proposed BSSNet has the superiority over other state-of-the-art methods. It is desirable that our main idea will be extended to other application of medical image segmentation in the future.

ACKNOWLEDGMENT

This work is supported by the Hainan Province Natural Science Foundation for Youths(823QN258) and Jilin Science & Technology Development Program of China under Grant 20210204101YY.

DATA AVAILABILITY

The datasets are available on the website:http://fcon_1000.projects.nitrc.org/indi/retro/atlas.html

CONFLICT OF INTEREST

There is no conflict of interest regarding the publication of this article.

REFERENCES

[1] N. Tomita, S. Jiang, M. E. Maeder, and S. Hassanpour, "Automatic post-stroke lesion segmentation on MR images using 3D residual convolutional neural network," *NeuroImage, Clin.*, vol. 27, Jan. 2020, Art. no. 102276.
[2] S. S. Virani, A. Alonso, E. J. Benjamin, and M. S. Bittencourt, "Heart disease and stroke statistics–2020 update: A report from the American Heart Association," *Circulation*, vol. 141, no. 9, pp. E139–E596, May 2020.
[3] Y. Wu, R. Wei, and P. G. Groot, "SARS in Hong Kong," *New England J. Med.*, vol. 349, pp. 708–709, Aug. 2003.
[4] Q. Bao, S. Mi, B. Gang, W. Yang, J. Chen, and Q. Liao, "MDAN: Mirror difference aware network for brain stroke lesion segmentation," *IEEE J. Biomed. Health Informat.*, vol. 26, no. 4, pp. 1628–1639, Apr. 2022.
[5] J. A. Fiez, H. Damasio, and T. J. Grabowski, "Lesion segmentation and manual warping to a reference brain: Intra- and interobserver reliability," *Hum. Brain Mapping*, vol. 9, no. 4, pp. 192–211, Apr. 2000.
[6] J. Montaner, A. Rovira, C. A. Molina, J. F. Arenillas, M. Ribó, P. Chacón, J. Monasterio, and J. Alvarez-Sabín, "Plasmatic level of neuroinflammatory markers predict the extent of diffusion-weighted image lesions in hyperacute stroke," *J. Cerebral Blood Flow Metabolism*, vol. 23, no. 12, pp. 1403–1407, Dec. 2003.

[7] H.-J. Wittsack, A. Ritzl, G. R. Fink, F. Wenserski, M. Siebler, R. J. Seitz, U. Mödder, and H.-J. Freund, "MR imaging in acute stroke: Diffusion-weighted and perfusion imaging parameters for predicting infarct size," *Radiology*, vol. 222, no. 2, pp. 397–403, Feb. 2002.

[8] R. G. R. Thomas, G. K. Lymer, P. A. Armitage, F. M. Chappell, T. Carpenter, B. Karaszewski, M. S. Dennis, and J. M. Wardlaw, "Apparent diffusion coefficient thresholds and diffusion lesion volume in acute stroke," *J. Stroke Cerebrovascular Diseases*, vol. 22, no. 7, pp. 906–909, Oct. 2013.

[9] 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 Mapping*, vol. 37, no. 4, pp. 1405–1421, Apr. 2016.

[10] P. Tang, P. Yang, D. Nie, X. Wu, J. Zhou, and Y. Wang, "Unified medical image segmentation by learning from uncertainty in an end-to-end manner," *Knowl.-Based Syst.*, vol. 241, Apr. 2022, Art. no. 108215.

[11] M. Bengs, F. Behrendt, M. Laves, J. Krger, R. Opfer, and A. Schlaefer, "Unsupervised anomaly detection in 3D brain MRI using deep learning with multi-task brain age prediction," *Proc. SPIE*, vol. 12033, Apr. 2022, Art. no. 1203314.

[12] Y. Li, Y. Lei, P. Wang, M. Jiang, and Y. Liu, "Embedded stacked group sparse autoencoder ensemble with L1 regularization and manifold reduction," *Appl. Soft Comput.*, vol. 101, Mar. 2021, Art. no. 107003.

[13] J. Shi, Y. Ye, D. Zhu, L. Su, Y. Huang, and J. Huang, "Comparative analysis of pulmonary nodules segmentation using multiscale residual U-Net and fuzzy C-means clustering," *Comput. Methods Programs Biomed.*, vol. 209, Sep. 2021, Art. no. 106332.

[14] K. S. Sankaran, M. Thangapandian, and N. Vasudevan, "Brain tumor grade identification using deep Elman neural network with adaptive fuzzy clustering-based segmentation approach," *Multimedia Tools Appl.*, vol. 80, pp. 25139–25169, Apr. 2021.

[15] Y. Zhang, S. Liu, C. Li, and J. Wang, "Application of deep learning method on ischemic stroke lesion segmentation," *J. Shanghai Jiaotong Univ. Sci.*, vol. 27, no. 1, pp. 99–111, Jan. 2021.

[16] Y. Yu, J. J. Heit, and G. Zaharchuk, "Improving ischemic stroke care with MRI and deep learning artificial intelligence," *Topics Magn. Reson. Imag.*, vol. 30, no. 4, pp. 187–195, Aug. 2021.

[17] 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. (MICCAI)*, Munich, Germany, 2015, pp. 234–241.

[18] G. Wang, T. Song, Q. Dong, M. Cui, N. Huang, and S. Zhang, "Automatic ischemic stroke lesion segmentation from computed tomography perfusion images by image synthesis and attention-based deep neural networks," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101787.

[19] H. Yang, W. Huang, K. Qi, C. Li, X. Liu, and M. Wang, "CLCI-Net: Cross-level fusion and context inference networks for lesion segmentation of chronic stroke,," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Shenzhen, China, 2019, pp. 266–274.

[20] L. Liu, L. Kurgan, F.-X. Wu, and J. Wang, "Attention convolutional neural network for accurate segmentation and quantification of lesions in ischemic stroke disease," *Med. Image Anal.*, vol. 65, Oct. 2020, Art. no. 101791.

[21] H. Hui, X. Zhang, F. Li, X. Mei, and Y. Guo, "A partitioning-stacking prediction fusion network based on an improved attention U-Net for stroke lesion segmentation," *IEEE Access*, vol. 8, pp. 47419–47432, 2020.

[22] V. Abramova, A. Clèrigues, A. Quiles, D. G. Figueredo, Y. Silva, S. Pedraza, A. Oliver, and X. Lladó, "Hemorrhagic stroke lesion segmentation using a 3D U-Net with squeeze-and-excitation blocks," *Computerized Med. Imag. Graph.*, vol. 90, Jun. 2021, Art. no. 101908.

[23] Y. Zhang, J. Wu, Y. Liu, Y. Chen, E. X. Wu, and X. Tang, "MI-UNet: Multi-inputs UNet incorporating brain parcellation for stroke lesion segmentation from T1-weighted magnetic resonance images," *IEEE J. Biomed. Health Informat.*, vol. 25, no. 2, pp. 526–535, Feb. 2021.

[24] S. Kadry, R. Damaševicius, D. Taniar, V. Rajinikanth, and I. A. Lawal, "Extraction of tumour in breast MRI using joint thresholding and segmentation—A study," in *Proc. 7th Int. Conf. Bio Signals, Images, Instrum. (ICBSII)*, Mar. 2021, pp. 1–5.

[25] L. 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.

[26] X. Chen, X. Wang, K. Zhang, K.-M. Fung, T. C. Thai, K. Moore, R. S. Mannel, H. Liu, B. Zheng, and Y. Qiu, "Recent advances and clinical applications of deep learning in medical image analysis," *Med. Image Anal.*, vol. 79, Jul. 2022, Art. no. 102444.

[27] K. Wang, X. Fan, and Q. Wang, "FPB-UNet++: Semantic segmentation for remote sensing images of reservoir area via improved UNet++ with FPN," in *Proc. 6th Int. Conf. Innov. Artif. Intell. (ICIAI)*, Mar. 2022, pp. 100–104.

[28] Q. Li, H. Song, F. Yang, Z. Wei, J. Fan, D. Ai, Y. Lin, X. Yu, and J. Yang, "Densely connected U-Net with criss-cross attention for automatic liver tumor segmentation in CT images," *IEEE/ACM Trans. Comput. Biol. Bioinf.*, early access, Aug. 15, 2022, doi: 10.1109/TCBB.2022.3198425.

[29] A. Kaur, G. Dong, and A. Basu, "GradXcepUNet: Explainable AI based medical image segmentation," in *Proc. Int. Conf. SMART MULTIMEDIA (ICSM)*, Marseille, France, 2022, pp. 174–188.

[30] Z. Wang, N. Zou, D. Shen, and S. Ji, "Non-local U-Nets for biomedical image segmentation,," in *Proc. AAAI Conf. Artif. Intell.*, New York, NY, USA, 2020, pp. 6315–6322.

[31] Y. Zhou, H. Chang, X. Lu, and Y. Lu, "DenseUNet: Improved image classification method using standard convolution and dense transposed convolution," *Knowl.-Based Syst.*, vol. 254, Oct. 2022, Art. no. 109658.

[32] 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.

[33] 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, May 2018.

[34] X. Li, H. Chen, X. Qi, Q. Dou, C. Fu, and P. Heng, "H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes," *IEEE Trans. Med. Imag.*, vol. 37, no. 12, pp. 2663–2674, Dec. 2018.

[35] 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. 6230–6239.

[36] A. Clèrigues, S. Valverde, J. Bernal, J. Freixenet, A. Oliver, and X. Lladó, "Acute and sub-acute stroke lesion segmentation from multimodal MRI," *Comput. Methods Programs Biomed.*, vol. 194, Oct. 2020, Art. no. 105521.

[37] L. Liu, F.-X. Wu, and J. Wang, "Efficient multi-kernel DCNN with pixel dropout for stroke MRI segmentation," *Neurocomputing*, vol. 350, pp. 117–127, Jul. 2019.

[38] S. Wang, Z. Chen, S. You, B. Wang, Y. Shen, and B. Lei, "Brain stroke lesion segmentation using consistent perception generative adversarial network," *Neural Comput. Appl.*, vol. 34, no. 11, pp. 8657–8669, Jun. 2022.

[39] 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. Bioinf.*, vol. 18, no. 3, pp. 940–950, May 2021.

[40] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, and Q. Liu, "X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies,," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Shenzhen, China, 2019, pp. 247–255.

[41] H. Zhang, K. Dana, J. Shi, Z. Zhang, X. Wang, A. Tyagi, and A. Agrawal, "Context encoding for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7151–7160.

[42] X. Zhang, T. Wang, J. Qi, H. Lu, and G. Wang, "Progressive attention guided recurrent network for salient object detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 714–722.

[43] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[44] L. Mou, Y. Zhao, L. Chen, J. Cheng, Z. Gu, and H. Hao, "CS-Net: Channel and spatial attention network for curvilinear structure segmentation,," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Shenzhen, China, 2019, pp. 721–730.

[45] Z. Fang, Y. Chen, D. Nie, W. Lin, and D. Shen, "RCA-U-Net: Residual channel attention U-net for fast tissue quantification in magnetic resonance," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Shenzhen, China, 2019, pp. 101–109.

[46] S. A. Taghanaki, A. Bentaieb, A. Sharma, S. K. Zhou, Y. Zheng, and B. Georgescu, "Select, attend, and transfer: Light, learnable skip connections," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent (MICCAI)*, Shenzhen, China, 2019, pp. 417–425.

[47] T. Zhao and X. Wu, "Pyramid feature attention network for saliency detection," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3080–3089.
[48] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3141–3149.
[49] Q. Yan, D. Gong, Q. Shi, A. van den Hengel, C. Shen, I. Reid, and Y. Zhang, "Attention-guided network for ghost-free high dynamic range imaging," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1751–1760.
[50] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2006, pp. 1735–1742.
[51] E. Hoffer and N. Ailon, "Deep metric learning using triplet network,," in *Proc. Int. Workshop Similarity-Based Pattern Recognit. (SIMBAD)*, Copenhagen, Denmark, 2015, pp. 84–92.
[52] S. Nazari-Farsani, M. Nyman, T. Karjalainen, M. Bucci, J. Isojärvi, and L. Nummenmaa, "Automated segmentation of acute stroke lesions using a data-driven anomaly detection on diffusion weighted MRI," *J. Neurosci. Methods*, vol. 333, Mar. 2020, Art. no. 108575.
[53] Y. Zhang, J. Wu, Y. Liu, Y. Chen, X. ED Wu, and X. Tang, "A 3D+2D CNN approach incorporating boundary loss for stroke lesion segmentation,," in *Proc. Int. Workshop Mach. Learn. Med. Imag. (MLMI)*, Lima, Peru, 2020, pp. 101–110.
[54] H. Shin, R. Agyeman, M. Rafiq, M. C. Chang, and G. S. Choi, "Automated segmentation of chronic stroke lesion using efficient U-Net architecture," *Biocybern. Biomed. Eng.*, vol. 42, no. 1, pp. 285–294, Jan. 2022.
[55] X. Du, K. Ma, and Y. Song, "AGMR-Net: Attention-guided multiscale recovery framework for stroke segmentation," *Computerized Med. Imag. Graph.*, vol. 101, Oct. 2022, Art. no. 102120.
[56] W. Yu, B. Lei, M. K. Ng, A. C. Cheung, Y. Shen, and S. Wang, "Tensorizing GAN with high-order pooling for Alzheimer's disease assessment," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 9, pp. 4945–4959, Sep. 2022.

ZHIHUA LI received the B.S. degree in computer science and technology from the Changchun Humanities and Sciences College, China, and the M.S. degree in computer application technology from Jilin Agricultural University, China. She is currently pursuing the Ph.D. degree in computer science and technology with the Changchun University of Science and Technology. Her current research interests include image processing, medical image processing, and deep learning.

QIWEI XING received the B.S. and M.S. degrees in computer science and information technology from Northeast Normal University and the Ph.D. degree in computer science and technology from the Changchun University of Science and Technology. He has been a Teacher with Qiongtai Normal University, since 2021. His current research interests include virtual reality, deep learning, neural networks, realistic graphics rendering, and image processing.

YANFANG LI is currently a Professor of computer science and technology with the Changchun University of Science and Technology, Changchun, China. Her current research interests include information systems and computer-aided diagnosis.

WEI HE is currently a Professor of computer science and technology with the Changchun University of Science and Technology, Changchun, China. Her current research interests include medical image processes and computer-aided diagnosis.

YU MIAO is currently an Associate Professor of computer science and technology with the Changchun University of Science and Technology. Her current research interest includes computer-aided diagnosis.

BAI JI is currently a Professor with the Department of Hepatobiliary and Pancreatic Surgery, General Surgery Center, The First Hospital of Jilin University. Her current research interests include liver cancer pathogenesis and microwave ablation.

WEILI SHI is currently an Associate Professor of computer science and technology with the Changchun University of Science and Technology. His current research interest includes computer-aided diagnosis.

ZHENGANG JIANG is currently a Professor of computer science and technology with the Changchun University of Science and Technology, Changchun, China. His current research interests include computer applications, mixed reality, and artificial intelligence.