IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 6, 2020, accepted March 21, 2020, date of publication March 24, 2020, date of current version April 7, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2983075

Attention Gate ResU-Net for Automatic MRI Brain Tumor Segmentation

JIANXIN ZHANG[1,2], (Member, IEEE), ZONGKANG JIANG[1], JING DONG[1], YAQING HOU[3], AND BIN LIU[4,5]

[1]Key Laboratory of Advanced Design and Intelligent Computing, Ministry of Education, Dalian University, Dalian 116622, China
[2]School of Computer Science and Engineering, Dalian Minzu University, Dalian 116600, China
[3]School of Computer Science and Technology, Dalian University of Technology, Dalian 116024, China
[4]International School of Information Science and Engineering (DUT-RUISE), Dalian University of Technology, Dalian 116622, China
[5]Key Laboratory of Ubiquitous Network and Service Software of Liaoning Province, Dalian University of Technology, Dalian 116622, China

Corresponding author: Bin Liu (liubin_dlut@163.com)

This work was supported in part by the National Key Research and Development Program of China under Grant 2018YFC0910506, in part by the National Natural Science Foundation of China under Grant 61972062 and Grant 61603066, in part by the Natural Science Foundation of Liaoning Province under Grant 2019-MS-011, in part by the Key Research and Development Program of Liaoning Province under Grant 2019JH2/10100030, in part by the High-Level Talent Innovation Support Program of Dalian City under Grant 2016RQ078, and in part by the Liaoning BaiQianWan Talents Program.

ABSTRACT Brain tumor segmentation technology plays a pivotal role in the process of diagnosis and treatment of MRI brain tumors. It helps doctors to locate and measure tumors, as well as develop treatment and rehabilitation strategies. Recently, MRI brain tumor segmentation methods based on U-Net architecture have become popular as they largely improve the segmentation accuracy by applying skip connection to combine high-level feature information and low-level feature information. Meanwhile, researchers have demonstrated that introducing attention mechanism into U-Net can enhance local feature expression and improve the performance of medical image segmentation. In this work, we aim to explore the effectiveness of a recent attention module called attention gate for brain tumor segmentation task, and a novel Attention Gate Residual U-Net model, i.e., AGResU-Net, is further presented. AGResU-Net integrates residual modules and attention gates with a primeval and single U-Net architecture, in which a series of attention gate units are added into the skip connection for highlighting salient feature information while disambiguating irrelevant and noisy feature responses. AGResU-Net not only extracts abundant semantic information to enhance the ability of feature learning, but also pays attention to the information of small-scale brain tumors. We extensively evaluate attention gate units on three authoritative MRI brain tumor benchmarks, i.e., BraTS 2017, BraTS 2018 and BraTS 2019. Experimental results illuminate that models with attention gate units, i.e., Attention Gate U-Net (AGU-Net) and AGResU-Net, outperform their baselines of U-Net and ResU-Net, respectively. In addition, AGResU-Net achieves competitive performance than the representative brain tumor segmentation methods.

INDEX TERMS MRI, brain tumor segmentation, U-Net, attention gate, residual module.

I. INTRODUCTION

Brain tumors are caused by abnormal cells growing in human brain. The current incidence of malignant brain tumors is relatively high, and this occurs a huge influence to humans and society [1]. The most common malignant brain tumors are gliomas, which can be further classified into high-grade gliomas (HGG) and low-grade gliomas (LGG). Magnetic resonance imaging (MRI), a typical non-invasive imaging technique that produces high-quality brain images without injury and skull artifacts while provides more comprehensive information of brain tumors, is regarded as the main technical means of diagnosing and treating brain tumors [2]. With the assistance of multimodal brain images, doctors can make quantitative analyses of the brain tumors so as to measure the maximum diameter, volume and quantity of brain lesions, thus developing the optimal diagnosis and treatment plan for patients to quantify the response of brain tumors before and

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

after treatment. In summary, brain tumor segmentation is a key step in the diagnosis and treatment of brain tumors [3]. Owing to the needs of clinical application and scientific research, brain tumor segmentation has become an important component in the field of medical images and has been widely concerned for a long time [4].

In recent years, deep neural network models represented by AlexNet [5], VGGNet [6], ResNet [7] and DenseNet [8] have been successfully utilized to a variety of computer vision tasks, which have received widespread attention from both academia and industry. In view of the strong ability of automatic extraction of high discriminant features shown by deep neural networks, they are quickly applied to the field of medical image processing and analysis [9]–[11]. In the meantime, computer-aided diagnosis of MRI brain tumor based on deep learning has also received extensive attention. Particularly, Multimodal Brain Tumor Segmentation challenge (BraTS), in conjunction with the International Association for Medical Image Computing and Computer Assisted Intervention (MICCAI) since 2012 [3], greatly promotes the development of deep learning based brain tumor segmentation methods. Generally speaking, current deep learning based brain tumor segmentation methods mainly consist of two typical types, i.e., convolutional neural network (CNN)-based methods and fully convolutional network (FCN)-based methods. CNN-based brain tumor segmentation networks utilize the idea of patch classification in small-scale images for the segmentation of brain tumor images. Pereira *et al.* [12] explore a small 3×3 convolution kernel model based on VGGNet to construct an automatic segmentation network, which wins the first place in the BraTS 2013 Challenge. Havaei *et al.* [13] construct a dual-path 2D CNN brain tumor segmentation network, which contains local and global paths by employing different size convolution kernels to extract different context feature information. However, patch-wise architectures lack spatial continuity and need large storage space, leading to low efficiency. Based on the conception of encoding and decoding, FCN-based brain tumor segmentation networks classify and predict each pixel in the whole brain image to complete the segmentation, which obviously improve the segmentation efficiency of brain tumors. FCN [14], an end-to-end semantic segmentation model that utilizes full convolution to solve the problem of pixel-by-pixel prediction, is an important milestone of image semantic segmentation task. Based on the architecture of FCN, Ronneberger *et al.* [15] propose a symmetric fully convolutional network called U-Net, which has been widely used in various medical image segmentation tasks. U-Net contains a contracting path to capture context information and an expanding path that ensures accurate location, largely improving the performance of medical image segmentation task. Researchers also introduce U-Net based architecture into the field of brain tumor segmentation. Dong *et al.* [16] develop a 2D U-Net based segmentation network and employ real-time data augmentation to refine the segmentation performance of brain tumors. Kong *et al.* [17] introduce feature pyramid module into U-Net architecture to integrate multi-scale semantic and location information, which effectively improves the segmentation accuracy. To achieve further performance improvement, dense block, dilated convolution, MultiRes block and up skip connection are also drawn into the brain tumor segmentation networks [18]–[21].

Besides, attention mechanism, which plays an important role in human perception, can effectively highlight the useful information while suppress the redundant one. Recently, attention mechanism has been receiving wide attention in a variety of computer tasks, such as natural language processing for machine translation [22], [23], natural image classification [24]–[26], salient object detection [27], [28], natural image segmentation [29]–[33] and video object segmentation [34] in computer vision fields, and medical image classification [35], [36] and medical image segmentation [37]–[41] in medical image analysis fields. There are many attempts that have embedded attention module into deep neural network architecture for improving the performance of image classification and image segmentation. Wang *et al.* [25] build a Residual Attention Network to generate attention-aware features from different modules, which change adaptively as layers going deeper and effectively improve the classification accuracy. Hu *et al.* [26] propose an attention module of Squeeze-and-Excitation (SE) block that focuses on the channel relationship and performs dynamic channel-wise feature recalibration to enhance feature expression. Li *et al.* [32] and Yu *et al.* [33] feed the features of deep layers with stronger semantics into SE-like attention block to provide high-level category information, which helps to precisely recovery details in the upsampling stage of image segmentation. Different from above attention neural networks that re-weight the important information, some researches mainly focus on capturing the long-range dependency of context. Wang *et al.* [42] firstly propose a family of pioneering non-local neural networks(NL-Nets) to capture long-range dependencies by aggregating query-specific global context to each query position, which has also been quickly applied into image segmentation tasks. Zhao *et al.* [29] design a location-sensitive module based non-local operation given in NL-Nets [42] to obtain the context long-range dependency, which achieves impressive segmentation results on several competitive scene parsing datasets. Zhang *et al.* [30] exploit a prior distribution to extend the non-local module and construct an ensemble of non-local operation with weights to pursue better segmentation performance. Fu *et al.* [31] employ dual attention modules composing of spatial and channel attention for semantic segmentation, in which the attention modules are similar to the non-local operation. In the meantime, attention mechanism has been gradually introduced into the field of medical image segmentation. Zhou *et al.* [39] explore a cross-task guided attention module to adaptively recalibrate channel-wise feature responses for brain tumor segmentation. Oktay *et al.* [37] propose attention gate module to make attention coefficients more specific to local regions, achieving promising results in pancreas

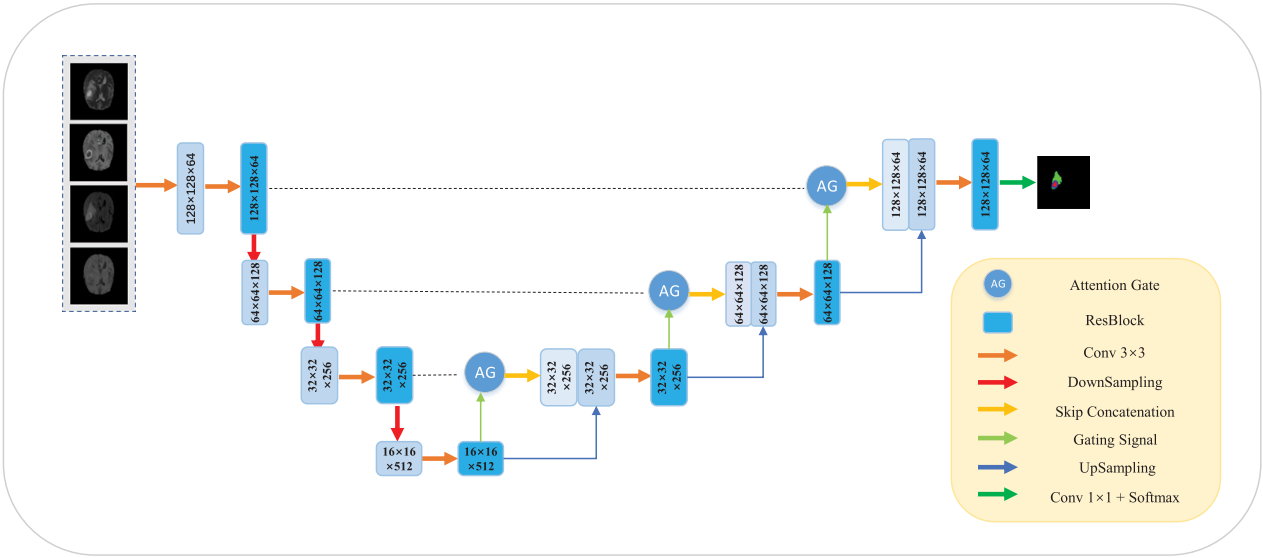

FIGURE 1. The end-to-end network architecture of AGResU-Net. AGResU-Net integrates residual modules and attention gates with a primeval and single U-Net architecture, in which a series of attention gate units are added into the skip connection for highlighting salient feature information while disambiguating irrelevant and noisy feature responses. AGResU-Net not only extracts abundant semantic information to enhance the ability of feature learning, but also pays attention to small-scale brain tumors, leading to the potential performance improvement of brain tumor segmentation task.

segmentation. Qi *et al.* [40] design a non-local operation called Feature Similarity Module to capture long-range dependencies and provide more effective context information for brain stroke lesion segmentation problem. Besides, Zhang *et al.* [43] introduce multiple attention modules that consist of region attention, spatial attention and channel attention into a modified fully convolutional network, achieving excellent performance in the ventricle segmentation field.

U-Net has achieved great success in the field of medical image segmentation, and it is also the mainstream of current MRI brain tumor segmentation methods. However, during the process of downsampling, U-Net constantly reduces the dimension of the image, which results in the poor segmentation accuracy for the small-scale tumors. It is noted that brain tumors have complex shapes and diverse sizes leading to the existence of small-scale tumors in the task of brain tumor segmentation. Considering that attention mechanism can enhance local feature expression, to solve the insufficient segmentation accuracy of small-scale tumors by U-Net, we aim to explore the effectiveness of attention gate, an efficient attention module for image segmentation task [37], for brain tumor segmentation problem, and a novel Attention Gate Residual U-Net model, i.e., AGResU-Net, is also put forward for this task. Experimental results on three brain tumor segmentation benchmarks illuminate that attention gate is benefit for the brain tumor segmentation task. The main contributions of this paper are summarized in three folds: (1) We propose an end-to-end AGResU-Net model for MRI brain tumor segmentation task, and its architecture is shown in Fig. I. AGResU-Net not only extracts more abundant semantic information, but also pays more attention to the information of small-scale brain tumors, which improves the segmentation effect of brain tumors. (2) AGResU-Net integrates residual modules and attention gates with a primeval and single U-Net architecture, in which a series of attention gate units are added into the skip connection for highlighting salient feature information while disambiguating irrelevant and noisy feature responses. On the one hand, residual modules enhance the ability of feature extraction and expression, and contribute to the classification in the process of downsampling. On the other hand, attention gates pays more attention to small-scale tumors and obtain more information about the location of small-scale tumors, so that the upsampling process is helpful to restore the location information of small-scale tumors. (3) We extensively evaluate AGResU-Net on three brain tumor segmentation benchmarks of BraTS 2017, BraTS 2018 and BraTS 2019. In addition, we also perform experiments of embedding attention gates into the basic U-Net model. Experimental results illuminate that models with attention gate units, i.e., Attention Gate U-Net (AGU-Net) and AGResU-Net, outperform their baselines of U-Net and ResU-Net, respectively. In addition, AGResU-Net achieves competitive performance than typical brain tumor segmentation methods.

II. METHODS

In this section, we first review the basic principle of attention gate. Then, we introduce the details of our AGResU-Net for brain tumor segmentation. Finally, the combined loss function adopted for supervising the learning of AGResU-Net is described.

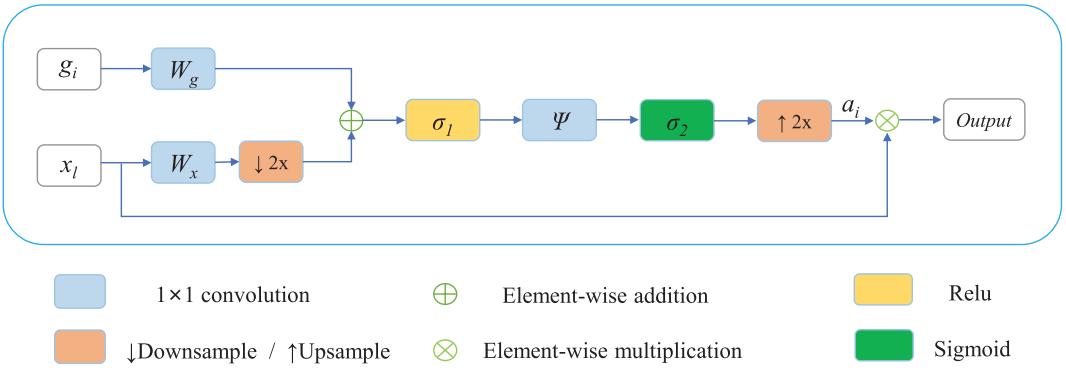

FIGURE 2. The basic schematic of attention gate. The g_i and x_l are represented as the gating signal vector and the feature map of the layer l, respectively. The σ_1 and σ_2 denote the Relu function and Sigmoid function. W_x, W_g and ψ are linear transformations. The α_i indicates the attention coefficient.

A. ATTENTION GATE

The attention gate is relative to the human visual attention mechanism, which automatically focuses on the target region and learns to suppress irrelevant feature responses in feature maps while highlighting salient feature information critical for a specific task. Recent researches have proved that deep learning models trained with the attention gate implicitly improve the performance of networks [25], [37], [38]. A typical architecture of attention gate unit can be illuminated in Fig. 2.

Given x_l be the feature map of the l layer, a gating signal vector g_i is taken advantage of each pixel i to choose the focus regions from a coarser scale. α is the attention coefficient ranging from 0 to 1, and it exports feature responses related to the target task and curbs useless feature information. The output x_{out} is the element-wise multiplication between x_l and α, whose specific formula is as follows:

$$x_{out} = x_l \cdot \alpha_i \tag{1}$$

Here, the gating coefficient α is obtained by using additive attention instead of multiplicative attention as [37]. Although additive attention requires a high computational cost, it can obtain more promising segmentation results. Due to that brain tumor segmentation is a multiple semantic classes task, we employ multi-dimensional attention coefficient [44] to focus on a subset of target regions. The multi-dimensional attention coefficient can be computed as:

$$\alpha_i = \sigma_2\left(\psi^T\left(\sigma_1\left(W_x^T x_l + W_g^T g_i + b_g\right)\right) + b_\psi\right) \tag{2}$$

where σ_1 is often chosen as Relu function $\sigma_1(r) = \max(0, r)$, and σ_2 is the Sigmoid function $\sigma_2(r) = \frac{1}{1+e^{-r}}$. W_x, W_g and ψ are linear transformations, and b_g and b_ψ are bias terms. We utilize 1×1 channel-wise convolution to perform linear transformation on the feature map x_l and the gating signal vector g_i. Besides, the xavier initialization approach [45] is employed to normalize parameters, followed by the back-propagation algorithm to update parameters in the model training.

B. ATTENTION GATE ResU-Net (AGResU-Net)

Due to the complex structures and diverse dimensional changes of brain tumors, some small-scale tumors exist in brain images. At the deeper stage of downsampling, the U-Net network has a richer feature representation capability as it learns more context information. However, spatial information and location details of the high-level output maps tend to get lost in the procedure of continuously cascaded convolutions and non-linearities transformations. This makes it difficult to improve the segmentation accuracy for small-scale brain tumors. To address this issue, we utilize attention gate units to propagate the relevant spatial details and location information of low-level feature maps in the process of upsampling. Besides, to purse more promising segmentation performance, we integrate attention gates and residual modules into U-Net architecture to produce Attention Gate ResU-Net (AGResU-Net) model, which can extract more precise dense feature information in the downsampling and well restore the spatial information and location details in the process of upsampling. Fig. I demonstrates the overall architecture of AGResU-Net, and the details of AGResU-Net are introduced as follows.

As shown in Fig. I, AGResU-Net has an encoder-decoder architecture that consists of a contracting path (an encoder on the left side) and an expanding path (a decoder on the right side). The size of the input for the network is $128 \times 128 \times 4$, in which the size of each image is 128×128 and the number of channels is 4. The contracting path consists of three residual blocks to replace the plain blocks in the original U-Net, and each residual block has two convolutional units. The size of the first layer is $128 \times 128 \times 64$ in the contracting path. For an individual convolutional unit, it includes a Batch Normalization (BN) layer, a Parametric Rectified Linear Unit (PReLU) as activation function [46]) instead of the ReLU function employed in the original U-Net architecture [15], and a convolutional layer with 3×3 filter. We utilize 3×3 convolution kernel with a stride of 2 for the downsampling. Meanwhile, the channel number of feature maps is doubled while the size of feature maps is reduced by half. The contracting path is

followed by the fourth residual unit with size of $16 \times 16 \times 512$, which serves as a bridge connecting the two paths.

Different from the existing MRI brain tumor segmentation models, the expanding path is composed of three residual blocks and three attention gate units to enhance the salient feature information while disambiguate the irrelevant and noisy feature responses. The gating signal vector is obtained by the residual block of each layer in the expanding path. The feature maps of the bottom level are taken as the initial gating signal of the expanding path. Then, the gating signal and the feature map obtained by the third residual block are combined through the attention gate unit, achieving more relevant location information of small-scale tumors to improve the performance of brain tumor segmentation. Prior to the individual residual block, there is an upsampling operation that increases the size of feature map by bilinear interpolation, subsequent a concatenation with the output feature map of attention gate and a 3×3 convolution. For each upsampling operation, the number of feature channel is reduced by half and the size of images is doubled. At the last layer of the expanding path, the multi-channel feature map is mapped to the desired number of classes utilizing a 1×1 convolution with Softmax activation function.

C. COMBINED LOSS FUNCTION

The MRI brain tumor segmentation task exhibits severe class imbalance where the healthy voxels comprise 98.46% of total voxels, 0.23% voxels belong to necrosis and non-enhancing, 1.02% to edema and 0.29% to enhancing tumor. Generalized Dice Loss (GDL) [47] is a commonly used loss function which helps narrow the gap between training samples and evaluation metric, and it is also immune to the data imbalance problem. Additionally, the weighted cross entropy (WCE) [48] has proved to be effective for multi-task training and class imbalance problem. Therefore, to provide better supervision for the model training, we utilize the combination of generalized dice loss L_{GDL} and weighted cross entropy loss L_{WCE} as a union loss function L as follows:

$$\mathrm{L} = L_{GDL} + \lambda \cdot L_{WCE} \tag{3}$$

where L_{GDL} and L_{WCE} represent the generalized dice loss and the weighted cross entropy loss, and are defined as Eq. (4) and Eq. (5), respectively.

$$L_{GDL} = 1 - 2\frac{\sum_i^N w_i \sum_k g_{ik} p_{ik}}{\sum_i^N w_i \sum_k (g_{ik} + p_{ik})} \tag{4}$$

$$L_{WCE} = -\sum_k \sum_i^N w_i g_{ik} \log (p_{ik}) \tag{5}$$

where N is the total number of labels, and w_i denotes the weight assigned to the ith label. As for the generalized dice loss, we set w_i to $\frac{1}{(\sum_k g_{ik})^2}$. p_{ik} and g_{ik} represent the value of the (ith, kth) pixel of the segmented binary image and of the binary ground truth image, respectively. Besides, λ in Eq. (3) is the hyper-parameter for controlling the balance of two functions.

III. EXPERIMENTS

In this section, we first introduce three brain tumor segmentation benchmarks utilized for model evaluation, followed by a simple description of the data preprocessing method. Then, evaluation metric and implementation details are given. Finally, compared experiment results on three benchmarks are reported and discussed.

A. DATASETS

We evaluate the effectiveness of attention gate and AGResU-Net on three benchmarks of BraTS 2017, BraTS 2018 and BraTS 2019 from Multimodal Brain Tumor Segmentation Challenge (BraTS). The BraTS 2017 database contains training dataset of 285 glioma patients, comprising of 210 HGG cases and 75 LGG cases. The validation dataset contains 46 patient subjects of unknown grade. The ground truths of the training dataset are manually labeled via expert and provided by the BraTS organizers. However, the ground truths of the validation dataset are not accessible to the public, and evaluation results on the validation dataset can only be obtained via the BraTS online website. All of the brain images are skull stripped and have the same orientation. For each patient case, there are four MRI modalities, i.e., Flair, T1, T1ce, and T2. To homogenize these data, all modalities are registered together to the T1c sequence, and then these images are resampled to 1 mm isotropic resolution on a normalized axis using a linear interpolator [3]. The labels are divided into four classes, namely healthy tissues (label 0), necrosis and non-enhancing (label 1), edema (label 2), and enhancing tumor (label 4). Fig. 3 illuminates a typical case of MRI brain image along with the ground truth.

The BraTS 2018 database is slightly different from the BraTS 2017 database. The BraTS 2018 database shares the same training data with BraTS 2017 database but has different validation dataset. It consists of 66 unlabeled patient subjects for the validation dataset, on which experimental results can only be evaluated through the BraTS website. For the BraTS 2019 database, it includes more patient cases than the BraTS 2017 and 2018 databases. Its training dataset contains 335 glioma patients, where 259 cases are HGG cases and the remaining cases are 76 LGG cases. Besides, the number of patient cases in BraTS 2019 validation dataset is expanded to 125.

B. DATA PREPROCESSING

As aforementioned, due to the contrast variations, uneven intensity and noise effects, MRI brain tumor segmentation is a challenging problem. Although deep learning based methods are more robust to the noise, data processing is still a critical and essential step. Additionally, in this work, we adopt multi-modality 3D MRI brain scan datasets, i.e., BraTS 2017, BraTS 2018 and BraTS 2019. However, multi-label brain tumor segmentation suffers from class imbalance problem, where the normal area occupying 98.46% pixels while 1.54% of pixels belong to the abnormal area. In BraTS database,

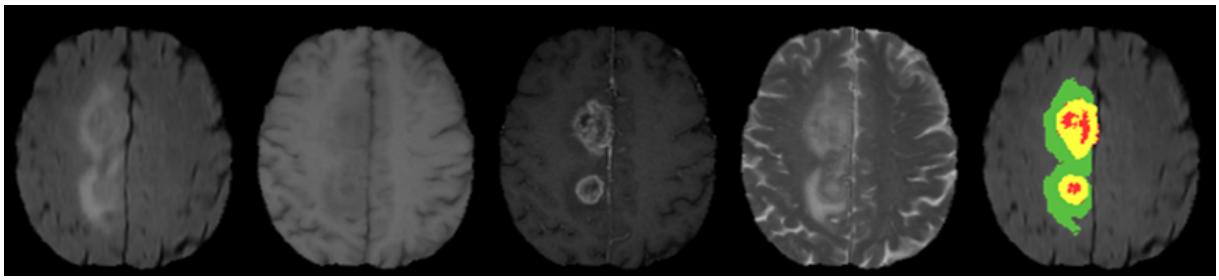

FIGURE 3. Example of the brain MRI data from the BraTS 2017 database. From left to right: Flair, T1, T1ce, T2 and the Ground Truth. Each color represents a tumor class: red—necrosis and non-enhancing, green—edema and yellow—enhancing tumor.

each 3D MRI image data has a volume dimension of 240 × 240 × 155. Considering that the axial brain image has the highest resolution and most of the volume in the dataset is obtained along the axial plane, we utilize 3D axial brain image to produce a number of 2D image slices with the size of 240 × 240. Before cutting 3D image into a series of 2D slice images, we first remove 1% highest and 1% lowest voxel intensities of the 3D images. Meanwhile, to solve class imbalance problem, we further process these 2D image slices in a patch manner by cropping each slice into several small patches whose size is 128 × 128.

Moreover, to reduce the influence of device noise, enhance the image contrast and alleviate over-fitting phenomenon, we also employ z-score normalization and Gaussian regularization [49] on 2D images. Z-score normalization operation processes each image by using mean value and standard deviation of intensities, and it can be computed as

$$z' = \frac{z - \mu}{\delta}, \tag{6}$$

where z and z' are the input image and normalized image, respectively. μ is the mean value of the input image, and δ is the standard deviation of the input image. Besides, Gaussian regularization is the addition of Gauss noise on image to improve accuracy for model training. It penalizes the noise-generated interference items for decreasing the weight square which achieves the similar effect as L2 regularization does, effectively mitigating the over-fitting phenomenon in the process of model training. After the data preprocessing, these 2D patch images are taken as the input of the brain tumor segmentation network to balance data voxels. This data preprocessing step could standardize the data and effectively mitigate the class imbalance problem, thereby improving the segmentation performance to a certain extent.

C. EVALUATION METRICS

In this work, we mainly utilize the Dice Similarity Coefficient (DSC) and Hausdorff distance (HD) for the model evaluation, which are the two most commonly used evaluation metrics for brain tumor segmentation. Between the two metrics, the DSC metric can be calculated as following.

$$DSC = \frac{2TP}{FN + FP + 2TP} \tag{7}$$

In the above equation, TP, FP and FN respectively represent true positive, false positive and false negative prediction. DSC mainly measures the area of overlap between the predicted lesion area and the actual labeled area.

Besides, the Hausdorff distance (HD) can be computed by

$$\mathrm{HD}(T, P) = \max\left\{\sup_{t\in T}\inf_{p\in P} d(t,p), \sup_{p\in P}\inf_{t\in T} d(t,p)\right\}, \tag{8}$$

where *sup* denotes the supremum, and *inf* represents the infimum. t and p represent the points on the surface T of the ground truth regions and the surface P of the predicted regions, respectively. $d(.,.)$ is the function that computes the distance between points t and p. Hausdorff distance estimates the distance between the surface of the real area and the surface of the predicted area. In addition, Hausdorff95 is a pivotal metric of the Hausdorff distance, and it is used to measure 95% quantile of the surface distance.

Due to practical clinical applications, the evaluation system generally separates the tumor structure into three regions for performance evaluation, and they are (a) the whole tumor region containing all intratumor areas with labels 1, 2, 4, (b) the core tumor region including whole tumor region but excluding edema region (labels 1, 4), and (c) the enhancing tumor region measuring label 4.

D. IMPLEMENTATION DETAILS

Our models are implemented using the Keras library with TensorFlow [50] as the backend. Each brain tumor segmentation model is trained with standard back-propagation with SGD as optimizer and Parametric Rectified Linear Unit (PReLU) function as activation function. During the training phase, we employ the patch-slice batch size for training and utilize the combined loss function given in Subsec. II-C to guide the learning of the model. Each of the MRI images is normalized by subtracting the mean and dividing by the standard deviation to alleviate the voxel intensity variance. Besides, the initial learning rate is set to 0.086, the momentum parameter is 0.97, and the hyper-parameter λ of the combined loss function is 1. All of the programs are carried out on a PC equipped with a single GeForce GTX 1080 GPU.

E. EXPERIMENT RESULTS AND DISCUSSION

Our experiments mainly consist of four parts, which are performed on BraTS 2017 training dataset, BraTS 2017 validation dataset, BraTS 2018 validation dataset, and BraTS 2019 validation dataset, respectively. Experiment results obtained on BraTS 2017 training dataset are the average value of five individual runs with cross validation. Meanwhile, evaluation results of BraTS 2017, BraTS 2018 and BraTS 2019 validation datasets are obtained via BraTS online website. In addition, the baseline results are recurrent with the same protocol utilized for the proposed model, and the other compared results are directly quoted from the literature.

1) EXPERIMENTS ON BraTS 2017 TRAINING DATASET

We first evaluate the attention gate module and AGResU-Net using the whole training dataset. We randomly split 285 labeled cases into training set, validation set and testing set at an approximate ratio of 6:2:2 [51], that is, 171 samples are taken as the training set, 57 cases are chosen for the validation set, and the remaining 57 cases are the testing set. Table 1 reports the compared segmentation results of four models, i.e., U-Net, ResU-Net, AGU-Net and AGResU-Net, where AGU-Net and AGResU-Net are the enhanced models of U-Net, ResU-Net by embedding attention gate modules. For fair comparisons, we report results of methods marked with * through publicly released codes of authors and try our best to fine-tune their parameters.

TABLE 1. Compared segmentation results with baselines on BraTS 2017 training dataset.

Methods	Whole	Core	Enhancing
U-Net *	0.870	0.762	0.703
AGU-Net (our)	0.874	0.772	0.711
ResU-Net *	0.873	0.768	0.716
AGResU-Net (our)	**0.876**	**0.772**	**0.720**

As shown in Table 1, AGResU-Net achieves the optimal segmentation performance among the four models, which obtains dice similarity coefficient(DSC) of 0.876, 0.772 and 0.720 on the whole tumor, core tumor and enhancing tumor segmentation, respectively. It is superior to the basic U-Net with 0.6%, 1.0% and 1.7% accuracy improvement on the whole tumor, core tumor and enhancing tumor, respectively. Compared with ResU-Net, AGResU-Net outperforms ResU-Net with 0.3-0.4% gains on average. When it comes to AGU-Net, AGU-Net gains 0.4%, 1.0% and 0.8% accuracy improvement over U-Net on the whole tumor, core tumor and enhancing tumor, respectively. The compared experiment results illuminate that embedding attention gates into U-Net and ResU-Net benefits the brain tumor segmentation.

Subsequently, we perform experiment using the high-grade glioma (HGG) cases of BraTS 2017 training dataset. HGG cases are the most common malignant brain tumors in adults, which tend to grow rapidly and spread faster than tumors of low-grade. As described in Subsec. III-A, there are 210 HGG cases in BraTS 2017 training dataset. In this experiment, we randomly divide the HGG cases into two parts, where 168 cases are taken as the training set and the remaining 42 cases are the testing set. Table 2 tabulates the compared results of U-Net, ResU-Net, AGU-Net and AGResU-Net. In addition, we also compare our AGResU-Net with other typical brain tumor segmentation methods to evaluate its performance, whose results are given in Table 3.

TABLE 2. Compared segmentation results with baselines on HGG cases of BraTS 2017 training dataset.

Methods	Whole	Core	Enhancing
U-Net *	0.881	0.847	0.814
AGU-Net (our)	0.884	0.854	0.825
ResU-Net *	0.886	0.857	0.823
AGResU-Net (our)	**0.891**	**0.865**	**0.830**

TABLE 3. Compared segmentation results with typical methods on HGG cases of BraTS 2017 training dataset.

Methods	Whole	Core	Enhancing
Chen et al. [52]	0.720	0.830	0.810
Kamnitsas et al. [53]	**0.900**	0.750	0.730
Dong et al. * [16]	0.831	0.801	0.750
Pereira et al. [12]	0.840	0.720	0.620
Kermi et al. * [48]	0.880	0.850	0.820
Zhao et al. [54]	0.865	0.864	0.816
AGResU-Net (our)	0.891	**0.865**	**0.830**

Table 2 illuminate that our AGResU-Net still obtains the best segmentation performance among the four models. It achieves mean DSC scores of 0.891, 0.865 and 0.830 on the whole tumor, core tumor and enhancing tumor, respectively. It outperforms U-Net by a large margin, and it is also superior to ResU-Net and AGU-Net with gains of 0.67-0.77% on average. Additionally, AGU-Net also has 0.70% average performance improvement over its baseline of U-Net, which mainly contributes to the embedded attention gate units. Compared with other typical methods, as shown in Table 3, AGResU-Net gains the optimal mean DSC scores on both of core tumor and enhancing tumor, and it is only inferior to the method proposed by Kamnitsas *et al.* [53] on whole tumor. In [53], the authors optimize the network structure by building two almost identical pathways to predict voxels of the final feature map. Compared with Chen *et al.* [52] and Kermi *et al.* [48] methods, our AGResU-Net model acquires a better segmentation performance. For models of Pereira *et al.* and Zhao *et al.*, they apply 2D CNN models with 33×33 patches as inputs to predict center voxel [12], [54]. In addition, Zhao *et al.* [54] additionally employ conditional random field as post-processing to improve the prediction performance. Without utilizing any post-processing strategy, our AGResU-Net achieves 2.40%, 3.04% and 7.05% gains over [54] on the whole tumor, core tumor and enhancing tumor, respectively. Comparison to the typical U-Net brain tumor segmentation network proposed in [16], our AGResU-Net outperforms it by a large margin with 6.0%, 6.4% and 8.0% gains on whole tumor, core tumor and enhancing tumor, respectively. The above comparison

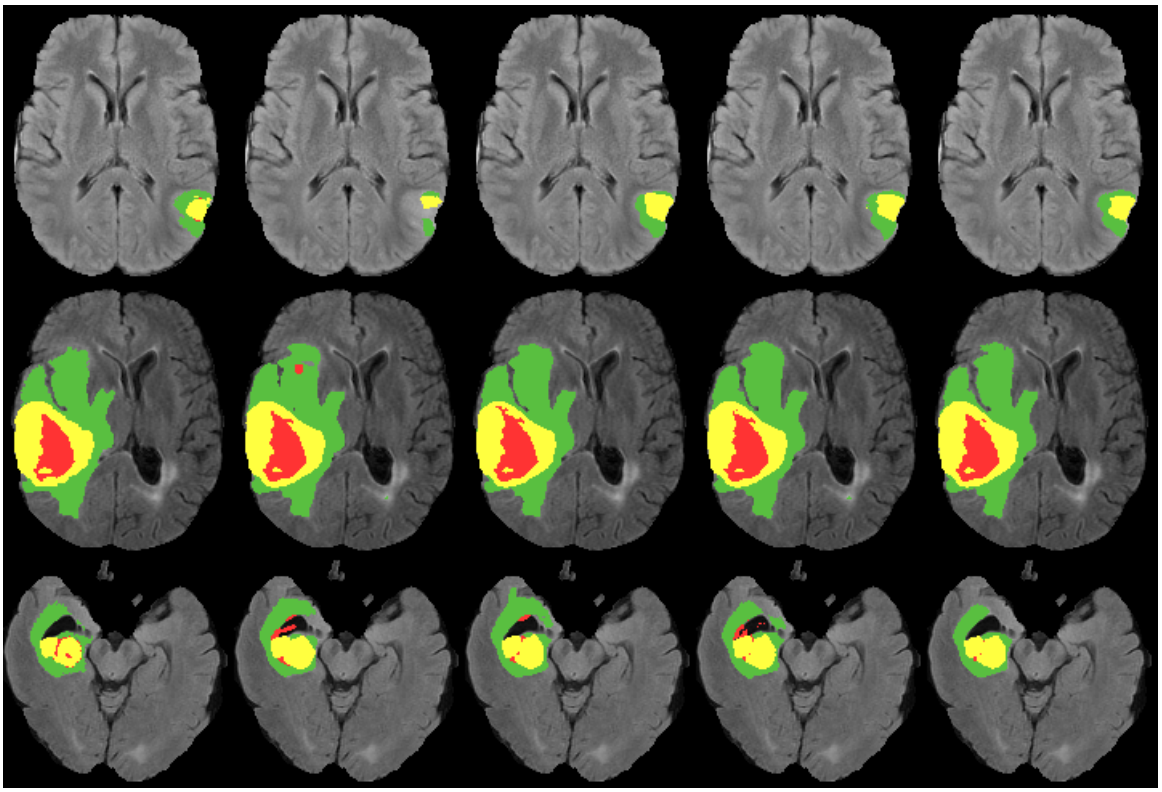

FIGURE 4. Examples of segmentation results on the HGG subset of BraTS 2017 training dataset. From left to right: Ground Truth, U-Net, AGU-Net, ResU-Net and AGResU-Net results overlaid on Flair image. Each color represents a tumor class: red—necrosis and non-enhancing, green—edema and yellow—enhancing tumor.

once again demonstrates the effect of AGResU-Net on brain tumor segmentation. Note that although these experiments are completed on the training dataset, as we split the dataset into two or three uncrossed separated datasets for independent training and testing, compared results are meaningful for evaluating the effectiveness of AGResU-Net in the segmentation of brain tumors.

Moreover, we also visualize the brain tumor segmentation results of the four models. We utilize various colors to represent different tumor classes for the segmented images, and then overlay the segmentation regions on the original brain image. Fig. 4 illuminates several typical samples with the corresponding segmentation results. In this figure, the red regions are necrosis and non-enhancing, the green regions represent edema, and the yellow regions indicate enhancing tumor. Meanwhile, images from left to right are Ground Truth, U-Net, AGU-Net, ResU-Net and AGResU-Net segmentation results overlaid on Flair image, respectively. From Fig. 4, it is seen that AGResU-Net achieves the best segmentation results of brain tumors. Then, we amplify the tumor region of the second patient case given in Fig. 4 and overlay one tumor region each time on the original Flair image, aiming to provide a more comprehensive comparison. The amplified images can be shown in Fig. 5, where the rectangle in blue is manually added to indicate the obviously different segmentation regions of the four models. It is clear to see that AGResU-Net achieves the best segmentation results of brain tumors. These segmentation results further indicate that attention gates focus on the small-scale tumor regions which help improve the performance of three tumor regions, especially for core and enhancing tumor.

TABLE 4. Compared segmentation results with baselines on BraTS 2017 validation dataset.

Methods	Whole	Core	Enhancing
U-Net *	0.852	0.759	0.698
AGU-Net (our)	0.877	0.770	0.724
ResU-Net *	0.862	0.774	0.732
AGResU-Net (our)	**0.880**	**0.781**	**0.749**

2) EXPERIMENTS ON BraTS 2017 VALIDATION DATASET

As the online evaluation of testing dataset is unavailable after the BraTS 2017 challenge, we evaluate the proposed models on the online validation dataset in this experiment. Here, We use whole training dataset of 285 MRI images to train our brain tumor segmentation models. Compared segmentation results with baselines are listed in Table 4, and comparison results with other typical methods are given in Table 5.

As shown in Table 4, AGU-Net is superior to U-Net with 2.50%, 1.10% and 2.60% gains on whole tumor, core tumor and enhancing tumor, respectively. Meanwhile, AGResU-Net

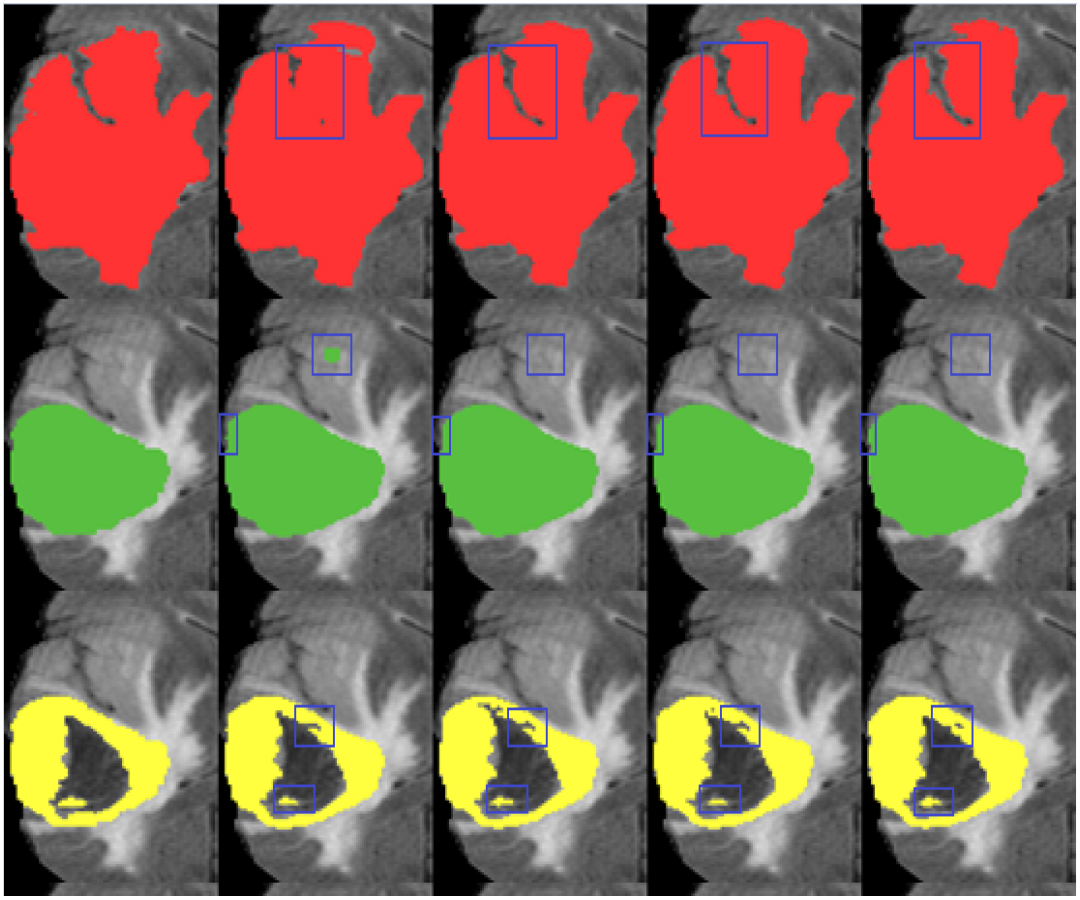

FIGURE 5. Refined segmentation results of one HGG case in BraTS 2017 training dataset. From left to right: Ground Truth, U-Net, AGU-Net, ResU-Net and AGResU-Net results overlaid on Flair image. Each color represents a tumor region: red—whole tumor, green—core tumor and yellow—enhancing tumor. Where the rectangle in blue is manually added for demonstration purpose, indicating the obviously different segmentation regions of the four models.

TABLE 5. Compared segmentation results with typical methods on BraTS 2017 validation dataset.

Methods	DSC			Hausdorff95		
	Whole	Core	Enhancing	Whole	Core	Enhancing
Islam et al. [55]	0.876	0.761	0.689	9.82	12.36	12.94
Jesson et al. [56]	0.899	0.751	0.713	**4.16**	8.65	6.98
Kamnitsas et al. [57]	**0.901**	**0.797**	0.738	4.23	**6.56**	4.50
Pereira et al. [58]	0.884	0.771	0.719	6.20	10.22	6.70
Mlynarski et al. [59]	0.886	0.784	0.711	-	-	-
Hu et al. [60]	0.850	0.700	0.650	25.24	21.45	17.98
AGResU-Net (our)	0.880	0.781	**0.749**	6.87	9.31	**3.74**

respectively obtains DSC scores of 0.880, 0.781 and 0.749 on the whole tumor, core tumor and enhancing tumor, which respectively outperforms ResU-Net by 1.80%, 0.70% and 1.70%, proving the effectiveness of the attention gate modules again. Besides, AGResU-Net achieves significant performance improvement over the baseline of U-Net with the average gain of 3.7%. Specially, it outperforms U-Net by a large margin of 5.1% on enhancing tumor segmentation. Therefore, these compared results well establish the brain tumor segmentation capability of AGResU-Net. When it comes to Table 5, AGResU-Net outperforms other top entries in the DSC value of enhancing tumor. By integrating multiple models to boost performance, Kamnitsas *et al.* [57] achieve the optimal DSC accuracy on whole tumor and core tumor segmentation. Compared with other latest models on DSC metric, AGResU-Net consistently has a large performance improvement than models of Islam and Ren [55] and Hu and Xia [60]. Meanwhile, AGResU-Net slightly lower than Jesson *et al.* [56] and Pereira *et al.* [58] on the DSC value of whole tumor. Jesson *et al.* utilize a FCN with multiple prediction layers and loss functions in different scales, which may lead to higher DSC value on the whole tumor. For Hausdorff95 metric evaluation results, our AGResU-Net achieves the Hausdorff95 distance value of 6.87, 9.31 and 3.74 on the

TABLE 6. Compared segmentation results with baselines on BraTS 2018 validation dataset.

Methods	Whole	Core	Enhancing
U-Net *	0.867	0.793	0.751
AGU-Net (our)	0.871	0.798	0.771
ResU-Net *	0.870	0.802	0.760
AGResU-Net (our)	**0.872**	**0.808**	**0.772**

whole tumor, core tumor and enhancing tumor segmentation, respectively. Specially, it gains the optimal Hausdorff95 metric on enhancing tumor segmentation. Although the best distance value on whole tumor and core tumor are respectively obtained by Jesson *et al.* [56] and Kamnitsas *et al.* [57], the compared Hausdorff95 results to a certain extent prove the effectiveness of AGResU-Net on segmenting small-scale tumors. Overall, our AGResU-Net model can achieve competitive performance and outperform several state-of-the-art methods. The comparisons also illustrate the effectiveness of our networks in conjunction with attention gate modules.

3) EXPERIMENTS ON BraTS 2018 VALIDATION DATASET

To further demonstrate the effectiveness of our method, we also perform experiment on the BraTS 2018 database. We employ the whole 285 MRI data from BraTS 2018 training dataset for models training and evaluate the models on its validation dataset. The experimental results are shown in Table 6 and Table 7.

On the one hand, Table 6 shows that comparison results among U-Net, AGU-Net, ResU-Net, and AGResU-Net are similar to those in Table 4. AGU-Net still achieves better performance than U-Net. Though it has 0.40% and 0.50% gains over U-Net on whole tumor and core tumor segmentation, it improves the enhancing tumor accuracy by 2.00%, well illuminating its effect on segmenting small-scale tumors. As the optimal model, AGResU-Net outperforms ResU-Net and AGU-Net by 0.67% and 0.40% on average, respectively. However, by embedding attention gates into ResU-Net, AGResU-Net also achieves 1.2% accuracy improvement on the enhancing tumor segmentation. In comparison, attention gate modules can improve the performance of U-Net and ResU-Net in terms of DSC scores. On the other hand, as shown in Table 7, AGResU-Net achieves very competitive performance compared with other typical brain tumor segmentation methods. AGResU-Net gains DSC values of 87.2%, 80.8% and 77.2% on whole tumor, core tumor and enhancing tumor, respectively. Comparison to some recent methods of Wang *et al.* [64], Chandra *et al.* [65] and Marcinkiewicz *et al.* [66], our AGResU-Net illuminates better segmentation performance on DSC metric. Due to Hu *et al.* [67] apply conditional random fields (CRFs) to fine segmentation results in multi-cascaded convolutional neural network, our AGResU-Net is slightly lower than it on whole tumor segmentation. When it comes to Hausdorff95 metric results, Hausdorff95 distances gained by our AGResU-Net are 5.62, 8.36 and 3.57 on the whole tumor, core tumor and enhancing tumor segmentation, respectively. The same with DSC results, it has the optimal Hausdorff95 distance value on enhancing tumor segmentation. Though AGResU-Net is inferior to Wang *et al.* [64] and Chandra *et al.* [65] on whole tumor and core tumor segmentation, it has the lowest Hausdorff95 distance value on average. Due to the whole tumor area is larger than the core tumor region and the enhancing tumor region, it is easier to segment whole tumors for U-Net model. When embedding residual structures and attention gate units individually, the performance will not be greatly improved. Meanwhile, we mainly employ attention gate units to improve the segmentation performance of small tumors, i.e., core tumor and enhancing tumor. Besides, in view of the limitation of memory size of GPU devices, we currently can only utilize 2D slices of brain image to segment brain tumors, which degrades its performance to some extent. Nevertheless, it is worth noting that though our 2D AGResU-Net is inferior to several 3D brain tumor segmentation methods on whole tumor and core tumor regions, AGResU-Net outperforms them on enhancing tumor segmentation. Therefore, AGResU-Net overall still achieves comparable performance compared with 3D models [61]–[63]. The above comparisons justify the effectiveness of attention gates. It can be concluded that AGResU-Net effectively improves the segmentation performance of small-scale tumors.

4) EXPERIMENTS ON BraTS 2019 VALIDATION DATASET

The last experiment is performed on the recent BraTS 2019 database. We adopt the similar protocol on BraTS 2017 and 2018 validation datasets to perform experiments. The differences are there are more patient cases in BraTS 2019 Database. As few results have been reported by other works, we only list the compared performance results with baselines

TABLE 7. Compared segmentation results with typical methods on BraTS 2018 validation dataset.

Methods	DSC			Hausdorff95		
	Whole	Core	Enhancing	Whole	Core	Enhancing
Çiçek et al. [61]	0.885	0.718	0.760	17.10	11.60	6.04
Chen et al. [62]	0.894	0.831	0.749	-	-	-
Nuechterlein et al. [63]	0.883	0.814	0.737	-	-	-
Wang et al. [64]	0.873	0.783	0.754	5.90	**8.03**	4.53
Chandra et al. [65]	0.872	0.795	0.741	**5.04**	9.59	5.58
Marcinkiewicz et al. [66]	0.862	0.768	0.723	-	-	-
Hu et al. [67]	0.882	0.748	0.718	12.60	9.62	5.69
AGResU-Net (our)	0.872	0.808	0.772	5.62	8.36	**3.57**

TABLE 8. Compared segmentation results with baselines on BraTS 2019 validation dataset.

Methods	Whole	Core	Enhancing
U-Net *	0.864	0.746	0.696
AGU-Net (our)	0.870	0.760	0.700
ResU-Net *	0.867	0.760	0.704
AGResU-Net (our)	**0.870**	**0.777**	**0.709**

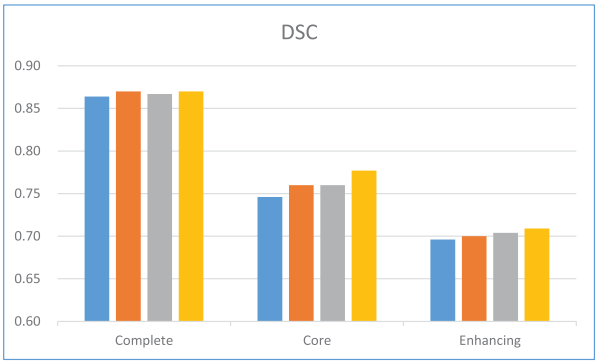

FIGURE 6. Comparison of DSC scores based on 125 patients in the BraTS 2019 validation dataset. For each tumor region, from left to right: U-Net, AGU-Net, ResU-Net and the AGResU-Net.

as in Table 8. Besides, Fig. 6 depicts bar plots of the average DSC scores for the three tumor regions on BraTS 2019 validation dataset. In this experiment, AGU-Net achieves superior segmentation performance compared to its baseline U-Net for the three tumor regions. Meanwhile, AGU-Net respectively outperforms U-Net by 0.60% (whole tumor), 1.40% (core tumor) and 0.40% (enhancing tumor). Then, compared ResU-Net with AGResU-Net, after embedding attention gates into ResU-Net, its DSC scores on three tumor segmentation regions increase by 0.30%, 1.70% and 0.50%, respectively. In particular, two models with attention gates averagely gain accuracy improvement of 1.55% over their baselines on core tumor segmentation, which is also due to the effectiveness of attention gates in improving small-scale brain tumor segmentation.

IV. CONCLUSION

In this paper, we mainly explore the effectiveness of attention gate for brain tumor segmentation task and also propose an AGResU-Net model which integrates residual modules and attention gates with a primeval and single U-Net architecture for this task. AGResU-Net adds a series of attention gate units into the skip connection for highlighting salient feature information while disambiguating irrelevant and noisy feature responses, which benefits the segmentation of small-scale brain tumors. We extensively evaluate attention gate units on three authoritative brain tumor benchmarks of BraTS 2017-2019. Experimental results illuminate that AGU-Net and AGResU-Net outperform their baselines of U-Net and ResU-Net, respectively. Additionally, AGResU-Net achieves competitive performance with typical brain tumor segmentation methods. As the 2D U-Net model has limitation on fully utilizing 3D information of MRI data, AGResU-Net loses an amount of context information and local details among different slices. In the future, we will try to explore 3D network architecture to improve segmentation performance of AGResU-Net, and extend the improved architecture to more datasets to show its generalization. Besides, as brain tumor segmentation is still a challenging task due to the complexity of MRI brain images, as well as the limitation of labeled samples for deep learning models, there is still a certain gap between our brain tumor segmentation method and current MRI scanning systems. Therefore, pursuing a clinic brain tumor segmentation model is also our future work.

REFERENCES

[1] A. Işın, C. Direkoğlu, and M. Şah, "Review of MRI-based brain tumor image segmentation using deep learning methods," *Procedia Comput. Sci.*, vol. 102, pp. 317–324, 2016.
[2] S. Bauer, R. Wiest, L.-P. Nolte, and M. Reyes, "A survey of MRI-based medical image analysis for brain tumor studies," *Phys. Med. Biol.*, vol. 58, no. 13, pp. R97–R129, Jul. 2013.
[3] B. H. Menze, A. Jakab, S. Bauer, J. Kalpathy-Cramer, K. Farahani, J. Kirby, Y. Burren, N. Porz, J. Slotboom, and R. Wiest, "The multimodal brain tumor image segmentation benchmark (BRATS)," *IEEE Trans. Med. Imag.*, vol. 34, no. 10, pp. 1993–2024, Dec. 2014.
[4] S. Cui, L. Mao, J. Jiang, C. Liu, and S. Xiong, "Automatic semantic segmentation of brain gliomas from MRI images using a deep cascaded neural network," *J. Healthcare Eng.*, vol. 2018, Mar. 2018, Art. no. 4940593, doi: 10.1155/2018/4940593.
[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[6] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[7] 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.
[8] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4700–4708.
[9] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.
[10] V. Gulshan, L. Peng, M. Coram, M. C. Stumpe, D. Wu, A. Narayanaswamy, S. Venugopalan, K. Widner, T. Madams, J. Cuadros, R. Kim, R. Raman, P. C. Nelson, J. L. Mega, and D. R. Webster, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA*, vol. 316, no. 22, p. 2402, Dec. 2016.
[11] Y. Liu, K. Gadepalli, M. Norouzi, G. E. Dahl, T. Kohlberger, A. Boyko, S. Venugopalan, A. Timofeev, P. Q. Nelson, G. S. Corrado, J. D. Hipp, L. Peng, and M. C. Stumpe, "Detecting cancer metastases on gigapixel pathology images," 2017, *arXiv:1703.02442*. [Online]. Available: http://arxiv.org/abs/1703.02442
[12] 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.
[13] 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.
[14] 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.
[15] 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.* Heidelberg, Germany: Springer, 2015, pp. 234–241.
[16] H. Dong, G. Yang, F. Liu, Y. Mo, and Y. Guo, "Automatic brain tumor detection and segmentation using u-net based fully convolutional networks," in *Proc. Annu. Conf. Med. Image Understand. Anal.* Heidelberg, Germany: Springer, 2017, pp. 506–517.

[17] X. Kong, G. Sun, Q. Wu, J. Liu, and F. Lin, "Hybrid pyramid u-net model for brain tumor segmentation," in *Proc. Int. Conf. Intell. Inf. Process.* Heidelberg, Germany: Springer, 2018, pp. 346–355.
[18] M. Shaikh, G. Anand, G. Acharya, A. Amrutkar, V. Alex, and G. Krishnamurthi, "Brain tumor segmentation using dense fully convolutional neural network," in *Proc. Int. MICCAI Brainlesion Workshop*. Cham, Switzerland: Springer, Sep. 2017, pp. 309–319.
[19] D. Liu, H. Zhang, M. Zhao, X. Yu, S. Yao, and W. Zhou, "Brain tumor segmention based on dilated convolution refine networks," in *Proc. IEEE 16th Int. Conf. Softw. Eng. Res., Manage. Appl. (SERA)*, Jun. 2018, pp. 113–120.
[20] N. Ibtehaz and M. S. Rahman, "MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation," *Neural Netw.*, vol. 121, pp. 74–87, Jan. 2020.
[21] H. Li, A. Li, and M. Wang, "A novel end-to-end brain tumor segmentation method using improved fully convolutional networks," *Comput. Biol. Med.*, vol. 108, pp. 150–160, May 2019.
[22] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," 2014, *arXiv:1409.0473*. [Online]. Available: http://arxiv.org/abs/1409.0473
[23] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 5998–6008.
[24] P. Veličković, G. Cucurull, A. Casanova, A. Romero, P. Liò, and Y. Bengio, "Graph attention networks," 2017, *arXiv:1710.10903*. [Online]. Available: http://arxiv.org/abs/1710.10903
[25] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156–3164.
[26] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.
[27] W. Wang, J. Shen, F. Guo, M.-M. Cheng, and A. Borji, "Revisiting video saliency: A large-scale benchmark and a new model," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4894–4903.
[28] W. Wang, S. Zhao, J. Shen, S. C. H. Hoi, and A. Borji, "Salient object detection with pyramid attention and salient edges," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 1448–1457.
[29] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. Change Loy, D. Lin, and J. Jia, "Psanet: Point-wise spatial attention network for scene parsing," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 267–283.
[30] H. Zhang, H. Zhang, C. Wang, and J. Xie, "Co-occurrent features in semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 548–557.
[31] 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. 3146–3154.
[32] H. Li, P. Xiong, J. An, and L. Wang, "Pyramid attention network for semantic segmentation," 2018, *arXiv:1805.10180*. [Online]. Available: http://arxiv.org/abs/1805.10180
[33] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Learning a discriminative feature network for semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1857–1866.
[34] W. Wang, H. Song, S. Zhao, J. Shen, S. Zhao, S. C. H. Hoi, and H. Ling, "Learning unsupervised video object segmentation through visual attention," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 3064–3074.
[35] J. Zhang, Y. Xie, Y. Xia, and C. Shen, "Attention residual learning for skin lesion classification," *IEEE Trans. Med. Imag.*, vol. 38, no. 9, pp. 2092–2103, Sep. 2019.
[36] M. Sun, K. Liang, W. Zhang, Q. Chang, and X. Zhou, "Non-local attention and densely-connected convolutional neural networks for malignancy suspiciousness classification of gastric ulcer," *IEEE Access*, vol. 8, pp. 15812–15822, 2020.
[37] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*. [Online]. Available: http://arxiv.org/abs/1804.03999
[38] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.
[39] C. Zhou, S. Chen, C. Ding, and D. Tao, "Learning contextual and attentive information for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2018, pp. 497–507.
[40] K. Qi, H. Yang, C. Li, Z. Liu, M. Wang, Q. Liu, and S. Wang, "X-Net: Brain stroke lesion segmentation based on depthwise separable convolution and long-range dependencies," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Heidelberg, Germany: Springer, 2019, pp. 247–255.
[41] Z. Zhuang, N. Li, A. N. Joseph Raj, V. G. V. Mahesh, and S. Qiu, "An RDAU-NET model for lesion segmentation in breast ultrasound images," *PLoS ONE*, vol. 14, no. 8, 2019, Art. no. e0221535.
[42] X. Wang, R. Girshick, A. Gupta, and K. He, "Non-local neural networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7794–7803.
[43] T. Zhang, A. Li, M. Wang, X. Wu, and B. Qiu, "Multiple attention fully convolutional network for automated ventricle segmentation in cardiac magnetic resonance imaging," *J. Med. Imag. Health Informat.*, vol. 9, no. 5, pp. 1037–1045, Jun. 2019.
[44] T. Shen, T. Zhou, G. Long, J. Jiang, S. Pan, and C. Zhang, "Disan: Directional self-attention network for RNN/CNN-free language understanding," in *Proc. 32th AAAI Conf. Artif. Intell.*, 2018, pp. 5446–5455.
[45] 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.
[46] K. He, X. Zhang, S. Ren, and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1026–1034.
[47] C. H. Sudre, W. Li, T. Vercauteren, S. Ourselin, and M. J. Cardoso, "Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*. Cham, Switzerland: Springer, 2017, pp. 240–248.
[48] A. Kermi, I. Mahmoudi, and M. T. Khadir, "Deep convolutional neural networks using U-Net for automatic brain tumor segmentation in multimodal MRI volumes, in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2018, pp. 37–48.
[49] A. V. Kolnogorov, "Gaussian two-armed bandit and optimization of batch data processing," *Problems Inf. Transmiss.*, vol. 54, no. 1, pp. 84–100, Jan. 2018.
[50] M. Abadi *et al.*, "TensorFlow: Large-scale machine learning on heterogeneous distributed systems," 2016, *arXiv:1603.04467*. [Online]. Available: http://arxiv.org/abs/1603.04467
[51] Y. Chen, Z. Lin, X. Zhao, G. Wang, and Y. Gu, "Deep learning-based classification of hyperspectral data," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 7, no. 6, pp. 2094–2107, Jun. 2014.
[52] L. Chen, Y. Wu, A. M. DSouza, A. Z. Abidin, A. Wismüller, and C. Xu, "MRI tumor segmentation with densely connected 3D CNN," *Proc. SPIE*, vol. 10574, Mar. 2018, Art. no. 105741F.
[53] K. Kamnitsas, C. Ledig, V. F. Newcombe, J. P. Simpson, A. D. Kane, D. K. Menon, D. Rueckert, and B. Glocker, "Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation," *Med. Image Anal.*, vol. 36, pp. 61–78, Feb. 2017.
[54] X. Zhao, Y. Wu, G. Song, Z. Li, Y. Zhang, and Y. Fan, "A deep learning model integrating FCNNs and CRFs for brain tumor segmentation," *Med. Image Anal.*, vol. 43, pp. 98–111, Jan. 2018.
[55] M. Islam and H. Ren, "Multi-modal pixelnet for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2017, pp. 298–308.
[56] A. Jesson and T. Arbel, "Brain tumor segmentation using a 3D FCN with multi-scale loss," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, Sep. 2017, pp. 392–402.
[57] K. Kamnitsas, W. Bai, E. Ferrante, S. McDonagh, M. Sinclair, N. Pawlowski, M. Rajchl, M. Lee, B. Kainz, and D. Rueckert, "Ensembles of multiple models and architectures for robust brain tumour segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2017, pp. 450–462.
[58] S. Pereira, V. Alves, and C. A. Silva, "Adaptive feature recombination and recalibration for semantic segmentation: Application to brain tumor segmentation in MRI," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent*. Heidelberg, Germany: Springer, 2018, pp. 706–714.
[59] P. Mlynarski, H. Delingette, A. Criminisi, and N. Ayache, "3D convolutional neural networks for tumor segmentation using long-range 2D context," *Computerized Med. Imag. Graph.*, vol. 73, pp. 60–72, Apr. 2019.
[60] Y. Hu and Y. Xia, "3D deep neural network-based brain tumor segmentation using multimodality magnetic resonance sequences," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, Sep. 2017, pp. 423–434.

[61] 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.* Heidelberg, Germany: Springer, 2016, pp. 424–432.
[62] W. Chen, B. Liu, S. Peng, J. Sun, and X. Qiao, "S3D-Unet: Separable 3D U-Net for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2018, pp. 358–368.
[63] N. Nuechterlein and S. Mehta, "3D-ESPNet with pyramidal refinement for volumetric brain tumor image segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2018, pp. 245–253.
[64] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, "Automatic brain tumor segmentation using convolutional neural networks with test-time augmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2018, pp. 61–72.
[65] S. Chandra, M. Vakalopoulou, L. Fidon, E. Battistella, T. Estienne, R. Sun, C. Robert, E. Deutsch, and N. Paragios, "Context aware 3D CNNs for brain tumor segmentation," in *Proc. Int. MICCAI Brainlesion Workshop*. Heidelberg, Germany: Springer, 2018, pp. 299–310.
[66] M. Marcinkiewicz, J. Nalepa, P. Lorenzo, W. Dudzik, and G. Mrukwa, *Automatic Brain Tumor Segmentation Using a Two-Stage Multi-Modal FCNN*. Cham, Switzerland: Springer, 2018, pp. 314–321.
[67] K. Hu, Q. Gan, Y. Zhang, S. Deng, F. Xiao, W. Huang, C. Cao, and X. Gao, "Brain tumor segmentation using multi-cascaded convolutional neural networks and conditional random field," *IEEE Access*, vol. 7, pp. 92615–92629, 2019.

JIANXIN ZHANG (Member, IEEE) received the Ph.D. degree from the Dalian University of Technology, Dalian, China, in 2009. He joined the Key Laboratory of Advanced Design and Intelligent Computing (Ministry of Education), Dalian University, in 2009, as a Lecturer, where he has been a Professor, since 2017. He joined the School of Computer Science and Engineering, Dalian Minzu University, in 2019. In 2017, he was as a Visiting Scholar with the École de technologie supérieure (ETS), Montreal, Canada. He has authored or coauthored over 40 published articles in his research area. His research area interests include computer vision, image processing, and pattern recognition.

ZONGKANG JIANG is currently pursuing the master's degree with the Key Laboratory of Advanced Design and Intelligent Computing (Ministry of Education), Dalian University, Dalian. His research interests include medical image processing and deep learning.

JING DONG received the Ph.D. degree from Jilin University, Changchun, China, in 2007. From 2007 to 2010, she joined the School of Mechanical Engineering, Dalian University of Technology, where she engaged in Postdoctoral Scientific Research. Since June 2010, she has been engaged in scientific research with the Key Laboratory of Advanced Design and Intelligent Computing, Dalian University. Her research interests include signal processing, deep learning, computer graphics, and pattern analysis.

YAQING HOU received the Ph.D. degree in artificial intelligence from the Interdisciplinary Graduate School, Nanyang Technological University, Singapore, in 2017. He was a Postdoctoral Research Fellow with the Data Science and Artificial Intelligence Research Centre, Nanyang Technological University. He is currently an Assistant Professor with the College of Computer Science and Technology, Dalian University of Technology, China. His current research interests include computational and artificial intelligence, multiagent systems, and reinforcement and transfer learning.

BIN LIU received the B.E. and M.E. degrees in mechanical engineering and the Ph.D. degree in biomedical engineering from the Dalian University of Technology, Dalian, China, in 2004, 2007, and 2009, respectively. Since October 2009, he has been a Researcher with the School of Software, Dalian University of Technology. He is currently a Professor of DUT-RU International School of Information Science and Engineering, Dalian University of Technology. His research interests include computer vision, computer graphics, medical image processing, and three-dimensional reconstruction.