Computers in Biology and Medicine

AGGN: Attention-based Glioma Grading Network with Multi-scale Feature Extraction and Multi-modal Information Fusion

--Manuscript Draft--

Manuscript Number:	CIBM-D-22-07089R1
Article Type:	Full Length Article
Keywords:	Artificial Intelligence; Glioma grading; Feature extraction; information fusion; Magnetic resonance imaging (MRI)
Corresponding Author:	Nianyin Zeng Xiamen University Xiamen, CHINA
First Author:	Peishu Wu
Order of Authors:	Peishu Wu
	Zidong Wang
	Baixun Zheng
	Han Li
	Fuad E. Alsaadi
	Nianyin Zeng
Abstract:	In this paper, a magnetic resonance imaging (MRI) oriented novel attention-based glioma grading network (AGGN) is proposed. By applying the dual-domain attention mechanism, both channel and spatial information can be considered to assign weights, which benefits highlighting the key modalities and locations in the feature maps. Multi-branch convolution and pooling operations are applied in a multi-scale feature extraction module to separately obtain shallow and deep features on each modality, and a multi-modal information fusion module is adopted to sufficiently merge low-level detailed and high-level semantic features, which promotes the synergistic interaction among different modality information. The proposed AGGN is comprehensively evaluated through extensive experiments, and the results have demonstrated the effectiveness and superiority of the proposed AGGN in comparison to other advanced models, which also presents high generalization ability and strong robustness. In addition, even without the manually labeled tumor masks, AGGN can present considerable performance as other state-of-the-art algorithms, which alleviates the excessive reliance on supervised information in the end-to-end learning paradigm.

None Declared.

Explanation of this revision

Paper number: CIBM-D-22-07089

First of all, the authors would like to express their sincere thanks to the Editor and the anonymous reviewers for their helpful comments and suggestions. The explanation of the modifications as well as corrections in this revision can be arranged as follows (comment numbers are in 1:1 correspondence with the reviewers' comments).

Reply to Co Editor in Chief:

Many thanks for your time and efforts in handling our paper. In this revision, all the comments from the three reviewers have been carefully taken into account and thoroughly implemented.

Reply to Reviewer No. 1

In this paper, a self-attention-based network (namely, the attention-based glioma grading network (AGGN)) is developed to handle the intelligent analysis of brain magnetic resonance imaging (MRI). The AGGN is composed of three meticulously designed modules (e.g. a dual-domain attention module, a multi-scale feature extraction module and a multi-modal information fusion module). The proposed AGGN is capable of reducing the reliance on supervised information of manual labels. Performance of the proposed AGGN is comprehensively evaluated on both internal and external testing sets, which yields satisfactory robustness and generalization ability.

The problem addressed is quite interesting. The paper is clearly written and well organized. Some comments are given below that might help with the presentation:

(1) **Comment**: The motivation of studying the intelligent analysis for the brain MRI should be introduced. Compared with the existing glioma grading methods, what are the essential advantages of the developed AGGN?

Reply: *Page 1, right column, line 10; Page 2, left column, line 4; Page 2, left column, line 21*

Thanks for your useful comment. In this revision, we have clarified the importance and motivation of the brain MRI intelligent analysis, and further analyzed the advantages of the proposed AGGN over other general methods.

"*Particularly, the MRI has advantages of strong specificity and sensitivity in tumor localization and pathological analysis [6], which can generate multi-modal images to reflect brain feature information at different levels by modulating imaging parameters.*"

"*In addition, most existing glioma grading algorithms have great reliance on the data with manually labeled tumor masks, while it is a time-consuming and laborious task to obtained those masks.*"

"*In addition, the ability of extracting features with strong presentation also guarantees the model robustness and generalization performance to some extents.*"

(2) **Comment**: By now, various fusion methods have been reported in the literature. What are the main advantages of the multi-modal information fusion utilized in this paper?

Reply: *Page 5, left column, line -8*

Thanks for your useful comment. Different from the general feature fusion methods, the multi-modal information fusion module used in this paper promotes the fusion of multi-scale features in different modalities. The main advantage of multi-modal information fusion module is to further integrate the enhanced semantic and detailed features. In this revision, we have further emphasized the advantages of the developed multi-modal information fusion module.

"*In the proposed AGGN, the multi-modal information fusion module is deployed to further integrate enhanced detailed and semantic features, and the structure is already illustrated in the green box of Fig. 2. In brief, fusion convolution realizes the integration of complementary advantages among the features of four modalities, where multi-scale feature maps in sizes of* 24×24, 12×12 *and* 6×6 *are fused. MB reduction block is adopted to transform the feature maps to vectors with strong presentation, which makes it feasible to further cascade the outputs of both multi-scale feature extraction and multi-modal information fusion module.*"

(3) **Comment**: In the multi-scale feature extraction module, the output of the dual-domain attention module is split into four single-modal maps. The reasons and advantages of such a setting should be proposed.

Reply: *Page 4, right column, line 2*

Thanks for your helpful comments. As you have suggested, this revision points out the reasons and advantages of dividing multi-modal MR image into four single modals for processing, which will be more conducive to understanding the structure of AGGN and the role of multi-scale feature extraction module. In this revision, we have already added the reasons and advantages of doing so.

"*As previously mentioned, the four MRI modalities have contained rich pathological information with different concerns. To realize sufficient feature extraction on each modality, output of the dual-domain attention module is further split into four single-modal maps to enter the multi-scale feature extraction module (see Fig. 2), which can*

promote in-depth analysis of the key features enhanced by attention mechanism and can also benefit the subsequent multi-modal information fusion as well."

(4) **Comment**: The computational complexity of the proposed method should be discussed.

Reply: *Page 10, left column, line 2*

Thanks for your useful comments. We have analyzed and discussed the computational cost of the proposed AGGN algorithm in detail in terms of time complexity and space complexity.

In this revision, we have followed your advice by adding model parameters and floating point operation of AGGN, which illustrates the efficient computational ability of AGGN for the glioma grading task.

"*D. Computational Complexity Analysis*

In this study, the number of model parameters (Params) and floating point operations (FLOPs) are adopted to depict the spatial and time complexity of the proposed AGGN, respectively. Excessive parameters will impede the light-weight deployment of model on edge devices, and too large FLOPs will influence the convergence during model training, which directly determines the accuracy of the model inference.

On the one hand, Params of the proposed AGGN is 16.37M, which is 9.13M fewer than that of the classical ResNet-50 model. According to Table III, the accuracy of AGGN is even 5.52% higher than that of ResNet-50, which demonstrates that the developed AGGN can effectively balance the computational costs and accuracy. It may owe to the proposed AGGN has effectively reduced the parameters by replacing large-size kernels with a series of small-sized ones. Meanwhile, the accuracy of AGGN is mainly guaranteed by the structural advantages, including employing dual-domain attention mechanism to highlight key features, realizing feature extraction on each individual modality, and integrating multi-modal information in different levels.

On the other hand, the FLOPs of AGGN are 24,790M, which mainly due to the large size of multi-channel input samples, where the data processing has consumed great deals of the FLOPs. It is also worth mentioning that during the model training, none of obvious over-fitting phenomenon has occurred, which implies that the training and inference time consumed by the proposed AGGN is acceptable.

To sum up, the proposed AGGN can effectively achieve the balance between model complexity and accuracy, which has achieved satisfactory results in the glioma grading task with considerable efficiency."

(5) **Comment**: Different metrics are adopted for performance evaluation. The practical significances of those metrics should be discussed.

Reply: *Page 6, right column, line 10*

Thanks for your thoughtful suggestions. In order to comprehensively evaluate performance of the proposed AGGN, six indicators *accuracy*, *precision*, *recall*, *specificity*, $F1$ *score* and *AUC* are utilized in the experiment section, each of which focuses on different aspects.

In this revision, by discussing the theoretical and practical significance of each metric, the logic and readability of the experimental results are both enhanced.

"*As can be seen, the accuracy describes the ratio of correct classifications of both HGG and LGG; the precision aims at all samples predicted as HGG, and calculates the proportion of correct prediction; the recall refers to the ratio of correctly identified HGG samples, which measures whether a model can screen all positive samples; similar to recall, the specificity reflects the ability of identifying negative samples of a model; the* $F1$ *score takes the harmonic average between accuracy and recall, and for all above five metrics, the larger their values are, the better the model performance is.*

In addition, the receiver operating characteristic (ROC) curve and area under this curve (AUC) are also employed for the model evaluation. Specifically, the ROC curve takes the value of $1 - specificity$ *(also known as false positive rate) as the horizontal axis and recall as the vertical one, AUC is the area enclosed by ROC curve and the two coordinate axes.*"

(6) **Comment**: Some future research topics should be discussed in the introduction or conclusion parts. For example, is it possible to consider other MRI-based tasks based on the AGGN framework?

Reply: *Page 2, left column, line 21; Page 10, right column, line 15*

Thanks for your helpful comments. Following your advice, we have expanded the future research topics in corresponding place. Especially, the applications of the proposed AGGN in MRI-based tasks and other fields are further analyzed.

In this paper, we have supplemented the future work, and have well addressed your concerns by adding more analysis on application areas.

"*In addition, the ability of extracting features with strong presentation also guarantees the model robustness and generalization performance to some extents. Therefore, it is also feasible and promising to apply the developed AGGN into other MRI-based tasks, such as the diagnosis of Parkinson's disease and Alzheimer's disease [32], [49].*"

"*In future work, we aim to 1) apply the developed AGGN framework to other MRI-based tasks such as stroke and and cancer diagnosis; 2) investigate fine-grained glioma grading methods to support quantitative analysis; 3) further optimize the*

structure of AGGN through fuzzy system and tensor decomposition techniques. [18], [23], [26], [43]"

(7) **Comment**: There are some typos throughout the paper that should be corrected.

Reply: *Page 2, right column, line 33; Page 3, left column, line 25; Page 5, right column, line 20; Page 9, right column, line 16*

Thanks for your thoughtful suggestions. In this revision, we have carefully checked our manuscript and corrected some spelling, grammar and formatting errors for further improving the quality of this manuscript.

"*Transfer learning paradigm has been introduced in [39], with two well-known CNN-based models AlexNet and GoogleNet, experimental results indicate that the pre-trained model can enhance the performance.*"

"*Feature fusion is another important operation in many DL-based methods, which promotes sufficient integration of information at different levels so as to enhance the presentation ability of features and improve the model performance.*"

"*In addition, substantial comparison experiments and ablation studies have been carried out to further validate the effectiveness and superiority of the proposed model. At first, experimental environment is briefly introduced.*"

"*Further explorations of the essential mechanism show that the volume of glioma from patients can be quite different, whereas the proposed MB conv block has adopted ACBs with different sizes to extract image features in parallel, which can obtain fine-grained texture and tissue information of multi-modal brain MRI.*"

Reply to Reviewer No. 2

This work attempts to improve the performance of glioma grading tasks without the help of mask labels. Besides, the developed framework is reasonable and the experimental results fully demonstrate the effectiveness of the methods. Although attention mechanisms have been a routine operation in deep learning, relatively few studies have been used to simultaneously select the modality and location most useful for diagnosis. I think this work is interesting. Overall, it is a well-written paper, and I have some questions and suggestions for it.

(1) **Comment**: Why do you use multimodal MR images of the brain, because they can provide more information?

Reply:

Thanks for your useful comments. Due to the four modalities in brain MRI can sufficiently present structural and functional information of tumors, the multi-modal MRI technique has become an important diagnostic tool of grading glioma in clinic.

(2) **Comment**: I don't quite understand the meaning of multimodality in this article, so I hope the authors can explain it.

Reply: *Page 1, right column, line 14*

Thanks for your questions. In this revision, we have explained the T1, T2, T1ce and FLAIR modalities of MRI in detail.

"*To be specific, four modalities are included in MRI [3], where the T1-weighted (T1) modality displays brain anatomy, and the T2-weighted (T2) one locates lesion area; fluid-attenuated inversion recovery (FLAIR) and T1-weighted contrast-enhanced (T1ce) modalities are generally used for visualization of peritumor and internal conditions so as to make further pathological analysis [8].*"

(3) **Comment**: In methodology, what is the meaning of BP in Eq.(4)?

Reply:

Thanks for your questions. BP stands for the "BN-PRelu" block in Fig. 3, which contains two sequential operations, batch normalization and parametric relu.

(4) **Comment**: The drawing of AGGN framework is beautiful, while the authors should explain what the three numbers in each bracket in Fig. 2 mean.

Reply: Thanks for your useful comments. The numbers in brackets of Fig. 2 indicate the size of feature maps (i.e., *height*, *width*, *number of channels*).

(5) **Comment**: The description of data division is a little confusing.

Reply: *Page 6, left column, line 1; Page 6, left column, line 15; Fig. 5*

Thanks for your thoughtful suggestions. In this revision, the details of how dataset is divided into training, testing and validation set, as well as the multimodal images processing, are further described.

"*The dataset is divided into training set, testing set and validation set, where the training set and testing set are independent of each other, while the validation set is obtained by further splitting the training set. To be specific, ratio of the training and testing set is* $2:1$*, where the images of training set come from 2018 BraTS. The testing set includes the internal and external subsets, which contain images from 2018 and 2019 BraTS, respectively. It should be pointed out that the main difference between internal and external testing subset is that samples of the latter belong to different data-source as those of training samples, and neither of the two subsets participates in the model training. Furthermore, one-fifth of the training samples are picked out to form the validation set for model tuning and selection.*

"*In addition, preprocessing is performed on the initial data before training the model, where tumor masks are used to screen tumor-free slices at first, and it is noticeable*

that the selected slices without tumor are not fed into the subsequent process. For slices that contain tumor tissues, the foreground region is standardized and the proportion of background is reduced so that they are center-cropped to 192×192 *in size; afterwards, four modalities are treated as four channels of the image. Finally, the dataset is divided according to the previously mentioned rules, and data augmentation operations are only performed on the training samples, including random rotation, translation and clipping. For a clear view, above preprocessing steps and dataset division are shown in Fig. 5.*"

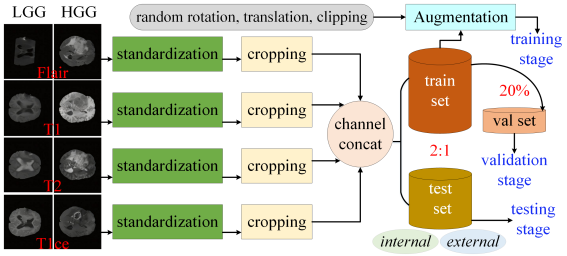


Fig.5. Flowchart of preprocessing for brain MRI datasets.

(6) **Comment**: Since attentional selection of modality and location has been done, why not conduct ablation studies on the dual-domain attention mechanism?

Reply: *Page 9, left column, line 14; Table. V; Fig. 10*

Thanks for your helpful suggestions. To further validate the performance of the proposed dual-domain attention mechanism, we have adopted an additional ablation study to evaluate our AGGN. Experimental results have shown the effectiveness of proposed attention mechanism, which is a competent and efficient module with strong ability of highlighting important features.

In this revision, we have supplemented the experiments and presented the results and discussions in Section IV-C with Table. V and Fig. 10.

"*To validate the effectiveness of core components in the proposed AGGN, substantial ablation studies are performed on the internal testing set in this subsection. The designed dual-domain attention mechanism is firstly verified and the results are reported in Table V, where AGG1, AGG2 and AGG3 refer to the model with none of attention modules, only spatial and only channel attention module, respectively. Obviously, in comparison to AGG1, on most indicators the performance has been improved to a certain extent after introducing the spatial or channel attention mechanism. It is also found that the applied dual-domain attention module in the proposed AGGN has realized significant performance enhancement, which improves accuracy, recall,* $F1$ *score and AUC by 2.92%, 3.61%, 4.23% and 1.6%, respectively.*

In addition, the ROC curves of four models listed in Table V are presented in Fig. 10, where it can be seen that the proposed AGGN has obtained the best results. In particular, when false positive rate is 0, the true positive rate of AGGN is close to 0.85 and AUC is 0.992, which implies that the proposed AGGN can accurately identify HGG with almost none of false detection. Therefore, the proposed AGGN is a reliable model that can provide a solid guarantee for the diagnosis and treatment of critical patients.

According to above results, effectiveness of the dual-domain attention mechanism is sufficiently validated. Before extracting multi-scale features, channel-domain attention is firstly introduced to low-level detail information, which determines what deserves attention in each modality of brain MRI; afterwards, spatial-domain attention is used to learn spatial dependence among high-level semantic information, so as to figure out the important locations in feature maps. As a result, the proposed AGGN can both recognize and localize the significant pathological features in brain MRI with strong robustness.

Table V
Ablation studies of dual-domain attention mechanism on internal testing set

Models	Metrics					
	accuracy	*precision*	*recall*	*specificity*	*F1 score*	*AUC*
AGG1	0.9320	1.0000	0.9160	1.0000	0.9027	0.9760
AGG2	0.9481	0.9866	0.9467	0.9532	0.9270	0.9860
AGG3	0.9476	**1.0000**	0.9349	**1.0000**	0.9238	0.9900
AGGN	**0.9612**	0.9987	**0.9521**	0.9952	**0.9450**	**0.9920**

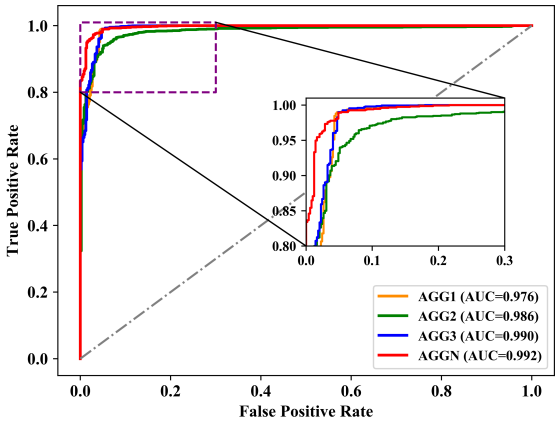


Fig. 10. ROC curves of AGG1, AGG2, AGG3, and AGGN. "

(7) **Comment**: Fig.6, Tables 2, 3, internal test means the results in validation set?

Reply:

Thanks for your useful questions. In Fig.6, Tables 2 and 3, the testing set includes the internal and external subsets, which contain images from 2018 and 2019 BraTS, respectively. It should be pointed out that the main difference between internal and external testing subset is that samples of the latter belong to different data-source as training samples, and neither of the two subsets participates in the model training. In this revision, we have further explained the difference of internal and external datasets.

(8) **Comment**: There should be a more detailed analysis and discussion of the obtained results.

Reply: *Page 6, right column, line -4; Page 7, left column, line 19; Page 8, left column, line 4; Page 9, left column, line 6*

Thanks for your thoughtful suggestions. In this revision, we have added detailed discussion in Section IV; and moreover, the analysis of the results combined with the principle of AGGN is provided as well.

"*At the same time, AGGN presents similar performance on internal and external testing sets, which demonstrates the robustness of AGGN in terms of handling various glioma MRI data; and moreover, this result indicates that AGGN can adapt to data from multi-center medical institution with strong generalization ability.*"

"*In this group of experiment, the proposed AGGN presents noticeable competitiveness in comparison to similar framework, which indicates the advantages in structural configuration.*"

"*Through this group of experiment, it is demonstrated that the proposed AGGN has overwhelming overall performance against other advanced CNN-based and domain-specific models on most metrics, which may owe to the meticulously designed and introduced dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion modules.*"

"*It is also worth mentioning that the AUC of 0.992 (without masks) is already an excellent result. Hence, it can be concluded that performance of AGGN has little reliance on the masks, which demonstrates that AGGN can overcome the high dependence of manually labeled annotations so as to achieve the end-to-end applications in practice.*"

Reply to Reviewer No. 3

In this paper, the authors have proposed an efficient convolutional neural network-based grading model for glioma magnetic resonance images, which is called attention-based

glioma grading network (AGGN). Specifically, the proposed AGGN contains three major modules of dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion. Finally, the validity and superiority of the proposed AGGN for glioma grading tasks have been demonstrated through adequate experimental validation and results analysis. In the field of medical image analysis, this work has certain theoretical research significance. Here are some comments for the authors to further improve the quality of the manuscript.

(1) **Comment**: In the introduction section, it is suggested to further analyze the difficulties in processing glioma images, as well as to elaborate on the limitations encountered by existing methods in performing the task of glioma grading.

Reply: *Page 1, right column, line 6 and line -5*

Thanks for your suggestions. As you have pointed out that there are so many difficulties in processing glioma images, and it is necessary to provide more analysis to the readers so that the limitations encountered by existing methods can be known.

In this revision, we have further elaborated the glioma grading task, analyzed the multi-modal characteristics of brain MR images and the difficulties of pre-processing and manual annotation. Moreover, the shortcomings of existing methods for glioma grading tasks are pointed out as well.

"*In clinical practice, the glioma grading task is mainly accomplished by imaging diagnosis [2], [13], [34], including computed tomography (CT) [19], magnetic resonance imaging (MRI) [22] and positron emission tomography (PET) [54], etc. Particularly, the MRI has advantages of strong specificity and sensitivity in tumor localization and pathological analysis [6], which can generate multi-modal images to reflect brain feature information at different levels by modulating imaging parameters. To be specific, four modalities are included in MRI [3], where the T1-weighted (T1) modality displays brain anatomy, and the T2-weighted (T2) one locates lesion area; fluid-attenuated inversion recovery (FLAIR) and T1-weighted contrast-enhanced (T1ce) modalities are generally used for visualization of peritumor and internal conditions so as to make further pathological analysis [8].*"

"*It is noticeable that in the context of applying DL-based models for medical image analysis [24], [27], a common and challenging issue is the robust feature extraction, which has great impact on the downstream tasks (e.g., segmentation and classification). Regarding to the glioma grading task, above problem is reflected on identifying different modalities with highly similar imaging features, and it is also tough to effectively utilize both semantic and detailed information under different MRI modalities. In addition, most existing glioma grading algorithms have great reliance on the data with manually labeled tumor masks, while it is a time-consuming and laborious task to obtained those masks.*"

(2) **Comment**:The scientific problem to be solved in this paper remains unclear, and what are the challenging issues in this work?

Reply: *Page 1, right column, line 25 and line 32*

Thanks for your useful comment. This paper aims to develop a computer-aided diagnosis system for grading glioma based on the multi-modal magnetic resonance images, therefore, the scientific problem can be described as "how to extract robust features with strong presentation by fully considering different modalities of the input". The most challenging issue is to identify different modalities with highly similar imaging features, and it is also tough to realize the effective feature fusion. In addition, it is noticeable that most existing glioma grading algorithms rely on the data with manually labeled tumor masks. Therefore, we also aim to overcome above deficiency by achieving comparable results without the masks due to it is always a time-consuming and laborious task to obtain them.

In this revision, we have clearly pointed out the major challenges in this study.

"*It is worth pointing out that inspecting the diverse information provided by multi-modal MRI is a laborious task, which inevitably increases the workload on radiologists or neurosurgeons.*"

"*Hence, it can be inferred that both the high heterogeneity of brain tumors and experiences of doctors will influence the final diagnostic results. As a result, it is necessary and beneficial to develop computer-aided diagnosis (CAD) systems to realize accurate glioma grading with less manpower [46]-[48].*"

(3) **Comment**: The authors should present more details about strengths of the developed AGGN, especially for the importance of the dual-domain attention mechanism.

Reply: *Page 2, left column, line 9*

Thanks for your useful comments. Following your suggestion, we have further introduced the advantages of the developed AGGN. To be specific, three major modules, in particular the attention mechanism, effectively capture potential correlations and key information from scattered features in different imaging modalities and improve the ability to distinguish intra-class variation and inter-class similarities of gliomas.

In this revision, we have already refined the motivation and necessity of AGGN and the individual modules presented in the introduction section.

"*In particular, by meticulously designing three modules to realize the function of dual-domain attention, multi-scale feature extraction and multi-modal information fusion, the proposed AGGN can efficiently capture potential correlations and key information from scattered features in different imaging modalities, and enhance the ability to distinguish intra-class variability and inter-class similarity existing at different grades of glioma. Based on the final fused highly discriminative features,*

the proposed AGGN can present comparable grading performance even without the manual labeled tumor masks. In addition, the ability of extracting features with strong presentation also guarantees the model robustness and generalization performance to some extents."

(4) **Comment**: The overall framework of the proposed AGGN shown in Fig. 2 seems similar to other existing models, therefore what is the major contribution of this study to related research?

Reply: *Page 2, left column, line -14; Page 3, right column, line 1*

Thanks for your helpful suggestion. As a matter of fact, the framework of deep learning-based models for a certain task can be universal, such as most object detection models consist of the backbone, neck part and detection head, which accounts for why the framework of AGGN is similar to some other models. In this study, it is worth pointing out that both the multi-scale feature extraction and multi-modal information fusion modules are meticulously designed, and the details have been already elaborated in Section III. In particular, characteristics of the brain MRI have been well considered so as to boost robust feature extraction; and with the extracted features with strong presentation, the proposed AGGN is proven powerful for grading glioma even without the assistance of manual labels, which can provide valuable experiences to alleviate the excessive reliance on supervised information in the end-to-end learning paradigm.

In this revision, we have further highlighted the major contributions of this study in corresponding places.

"*Designed multi-scale feature extraction and multi-modal information fusion modules benefit extracting discriminative features with strong presentation.*"

"*Afterwards, the pre-processed multi-modal MRI data will enter the dual-domain attention mechanism module, where the weights are assigned in both channel and spatial dimensions to highlight the key information and suppress the unimportant one in feature maps. Next, the multi-modal MRI is split, and in the followed multi-scale feature extraction module, parallel processing is performed on each single-modality map, including sequential operations of multi-branch convolution (MB conv), convolution-pooling (C-P) and multi-branch pooling (MB pool), and the final output of each pathway is in the size of* $1 \times 1 \times 256$. *It is noticeable that during above procedure, maps with three sizes on each modality are individually concatenated and fed into the multi-modal information fusion module, which contains the fusion convolution and MB reduction operations. At last, seven feature maps in the same size are cascaded and fed into the linear layers, which is responsible to accomplish the glioma grading task.*"

(5) **Comment**: More details need to be described about brain MRI datasets, including training / testing / validation data setting.

Reply: *Page 6, left column, line 1*

Thanks for your helpful comments. In this revision, we have added more descriptions of the utilized brain MR datasets, including the division of the training, testing and validation sets.

"*The dataset is divided into training set, testing set and validation set, where the training set and testing set are independent of each other, while the validation set is obtained by further splitting the training set. To be specific, ratio of the training and testing set is* $2:1$*, where the images of training set come from 2018 BraTS. The testing set includes the internal and external subsets, which contain images from 2018 and 2019 BraTS, respectively.*"

(6) **Comment**: Since the authors have highlighted that the applied modules are meticulously designed, how can they prove the advantages of the proposed architecture?

Reply: *Page 7, left column, line 4*

Thanks for your useful comment. In order to prove the architectural advantages of the proposed AGGN, another adaptive multi-modal fusion network (AMMFNet) in [35] has been adopted for comparison. It is noticeable that the AMMFNet has similar structure to the proposed AGGN, and according to the results illustrated in Fig. 7, our method has yielded improvement to different extents on all applied indicators, which indicates that those meticulously designed modules do facilitate a better feature extraction.

In this revision, we have discussed the structural advantages of the proposed AGGN in Section IV-B.

"*In this part, to validate the architectural advantages of our method, adaptive multi-modal fusion network (AMMFNet) [42] is adopted as baseline model for comparison, which is a similar glioma grading framework to the proposed AGGN. In Fig. 7, performance enhancement of AGGN on six indicators is illustrated, which shows that in comparison to AMMFNet, the proposed AGGN has improved all indicators to different extents. In particular, the most significant improvement is on specificity, which increases 11.52%. As previously mentioned, high specificity is equivalent to low false positive rate.*"

(7) **Comment**: In the experimental part, AGGN should be supplemented with a performance validation with algorithms for glioma grading, in addition to a comparison with the classical models.

Reply: *Page 8, left column, line 11*

Thanks for your useful comments. In this revision, we have supplemented an extra experiment with other state-of-the-art glioma grading algorithms to further validate the performance advantages of the proposed AGGN, where corresponding experimental results and the discussions have been displayed in Section IV-D-4).

"*4) Comparisons with state-of-the-art glioma grading algorithms*

In this part, comparison between proposed AGGN and other state-of-the-art glioma grading algorithms are presented, including multistream CNN [9], multi-scale CNN [10], CAE-GAN (convolutional autoencoder and generative adversarial network) [1], pre-trained GoogleNet [44], 3DConvNet [53] and AMMFNet [42]. It should be pointed out that in previous experiments, data from the mentioned internal and external testing sets have none of the tumor masks, while most of recently related methods for the same task require the assistance of additional tumor masks. Consequently, to make a fair comparison, in this group of experiments, tumor masks have been added to original images for training, and the results are reported in Table IV. Notice that the data of other algorithms are cited from corresponding original papers, and "–" denotes none of relevant data is provided.

Table IV
Performance of AGGN and other advanced models on tumor mask assisted data

Models	Metrics			
	accuracy	*recall*	*specificity*	*AUC*
Multistream CNN	0.9087	–	–	–
Multi-scale CNN	0.8947	–	–	–
CAE-GAN	0.9204	–	–	–
Pre-trained GoogleNet	0.9450	–	–	0.9680
3DConvNet	0.9710	0.9470	0.9680	–
AMMFNet	0.9820	**1.0000**	0.9330	0.9970
AGGN (ours)	**0.9899**	**1.0000**	**0.9678**	**0.9980**

As can be found in Table IV, with assistance of tumor masks, the proposed AGGN can present the state-of-the-art performance. In particular, the AUC value of AGGN with and without tumor masks are 0.998 and 0.992, respectively, which implies that the assistance of tumor masks does further improve the model performance. It is also worth mentioning that the AUC of 0.992 (without masks) is already an excellent result. Hence, it can be concluded that performance of AGGN has little reliance on the masks, which demonstrates that AGGN can overcome the high dependence of manually labeled annotations so as to achieve the end-to-end applications in practice.
"

(8) **Comment**: How are the so-called internal and external datasets defined, and what conclusions can be drawn from Fig. 6? More discussions should be provided.

Reply: *Page 6, left column, line 1*

Thanks for your useful suggestion. In this study, the collected data are from different sources, which are further split into the training and testing set; and furthermore, the testing set is further divided into two subsets. If the samples in a subset have shared the same data-source with the training ones, the subset is called "internal" dataset; otherwise, the subset is named the "external" dataset. According to Fig. 6, the proposed AGGN has presented satisfactory results on both datasets, which reflects that our method is reliable in glioma grading task. More importantly, this result has implied that the proposed AGGN can adapt to data from multi-center medical institution with high generalization ability.

In this revision, we have further explained how internal and external datasets are defined, and more discussions on Fig. 6 are provided in corresponding places.

"*The dataset is divided into training set, testing set and validation set, where the training set and testing set are independent of each other, while the validation set is obtained by further splitting the training set. To be specific, ratio of the training and testing set is* $2:1$*, where the images of training set come from 2018 BraTS. The testing set includes the internal and external subsets, which contain images from 2018 and 2019 BraTS, respectively. It should be pointed out that the main difference between internal and external testing subset is that samples of the latter belong to different data-source as those of training samples, and neither of the two subsets participates in the model training. Furthermore, one-fifth of the training samples are picked out to form the validation set for model tuning and selection.*

(9) **Comment**: More limitations of the study can be illustrated in conclusion.

Reply: *Page 10, right column, line 15*

Thanks for your thoughtful comments. At the end of this manuscript, we have summarized the existing limitations of the proposed AGGN, aiming at which a clear outlook of potential improvements in future work has been presented.

In this revision, we have followed your advice and clarified the future work in terms of limitations of this work, which mainly covers three perspectives of task migration adaption, quantitative lesion analysis, and model lightweighting studies.

"*Although the proposed AGGN has presented satisfactory performance on the glioma grading task, it still has some spaces for further improvement, including task migration adaption, quantitative lesion analysis, and model lightweighting studies.*"

(10) **Comment**: Future work must be clarified in the conclusion to present a clear outlook.

Reply: *Page 10, right column, line 19*

Thanks for your helpful comment. Following your advice, in this revision, we have further discussed potential future work at the end of our manuscript.

"*In future work, we aim to 1) apply the developed AGGN framework to other MRI-based tasks such as stroke and cancer diagnosis; 2) investigate fine-grained glioma grading methods to support quantitative analysis; 3) further optimize the structure of AGGN through fuzzy system and tensor decomposition techniques. [18], [23], [26], [43]*"

Highlights

- The proposed AGGN can alleviate the reliance on manually labeled tumor masks.
- Dual-domain attention is useful for selecting the modality and location of MRI.
- Multi-modal and multi-scale learning benefits analyzing brain MRI comprehensively.
- Effective fusion methods enhance the presentation ability of robust features.

AGGN: Attention-based Glioma Grading Network with Multi-scale Feature Extraction and Multi-modal Information Fusion

Peishu Wu, Zidong Wang, Baixun Zheng, Han Li, Fuad E. Alsaadi and Nianyin Zeng*

***Abstract*—In this paper, a magnetic resonance imaging (MRI) oriented novel attention-based glioma grading network (AGGN) is proposed. By applying the dual-domain attention mechanism, both channel and spatial information can be considered to assign weights, which benefits highlighting the key modalities and locations in the feature maps. Multi-branch convolution and pooling operations are applied in a multi-scale feature extraction module to separately obtain shallow and deep features on each modality, and a multi-modal information fusion module is adopted to sufficiently merge low-level detailed and high-level semantic features, which promotes the synergistic interaction among different modality information. The proposed AGGN is comprehensively evaluated through extensive experiments, and the results have demonstrated the effectiveness and superiority of the proposed AGGN in comparison to other advanced models, which also presents high generalization ability and strong robustness. In addition, even without the manually labeled tumor masks, AGGN can present considerable performance as other state-of-the-art algorithms, which alleviates the excessive reliance on supervised information in the end-to-end learning paradigm.**

***Index Terms*—Artificial intelligence; glioma grading; feature extraction; information fusion; magnetic resonance imaging (MRI)**

I. Introduction

As one of the most common primary tumors caused by the cancerization of glial cells in the brain or spinal cord, glioma accounts for nearly half of intracranial tumors and 36% of the nervous system tumors [30]. According to the criteria of the World Health Organization (WHO), glioma can be graded as four levels from I to IV [28], where low-grade glioma (LGG) includes grades I-II, and grades III-IV are the so-called high-grade glioma (HGG). It is worth mentioning that LGG may be cured by drug therapy and surgical excision, whereas radiotherapy and chemotherapy are required to cure HGG. Moreover, patients with the glioma of grade IV even suffer from low survival rate less than 10% [20]. Consequently, it is of vital significance to realize accurate preoperative grading of glioma.

In clinical practice, the glioma grading task is mainly accomplished by imaging diagnosis [2], [13], [34], including computed tomography (CT) [19], magnetic resonance imaging (MRI) [22] and positron emission tomography (PET) [54], etc. Particularly, the MRI has advantages of strong specificity and sensitivity in tumor localization and pathological analysis [6], which can generate multi-modal images to reflect brain feature information at different levels by modulating imaging parameters. To be specific, four modalities are included in MRI [3], where the T1-weighted (T1) modality displays brain anatomy, and the T2-weighted (T2) one locates lesion area; fluid-attenuated inversion recovery (FLAIR) and T1-weighted contrast-enhanced (T1ce) modalities are generally used for visualization of peritumor and internal conditions so as to make further pathological analysis [8]. Due to above four modalities can sufficiently present structural and functional information of tumors, the multi-modal MRI technique has become an important diagnostic tool of grading glioma in clinic.

It is worth pointing out that inspecting the diverse information provided by multi-modal MRI is a laborious task, which inevitably increases the workload on radiologists or neurosurgeons. For a clear view, the MRI slices of LGG and HGG are illustrated in Fig. 1(a) and Fig. 1(b), respectively. Experienced doctors generally distinguish LGG from HGG by observing the clarity of tumor contour and the presence of edema in peritumor areas. Hence, it can be inferred that both the high heterogeneity of brain tumors and experiences of doctors will influence the final diagnostic results. As a result, it is necessary and beneficial to develop computer-aided diagnosis (CAD) systems to realize accurate glioma grading with less manpower [46]–[48].

Owing to the continuous development of the deep learning (DL) techniques [29], [55], plenty of CAD methods have been proposed and applied to the intelligent analysis of brain MRI [7], [16], [25], [33], and related studies regarding to glioma grading tasks are reviewed in Section II. It is noticeable that in the context of applying DL-based models for medical image analysis [24], [50], a common and challenging issue is the robust feature extraction, which has great impact on the downstream tasks (e.g., segmentation and classification). Regarding

This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 30-135-1443). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

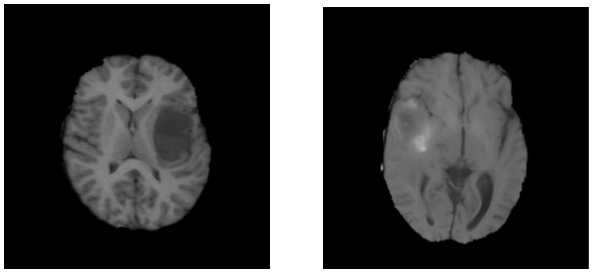
P. Wu, H. Li and N. Zeng are with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China. Email: zny@xmu.edu.cn

Z. Wang is with the Department of Computer Science, Brunel University London, Uxbridge UB8 3PH, U.K. Email: zidong.wang@brunel.ac.uk

B. Zheng is with the Polytechnic Institute, Zhejiang University, Hangzhou 310015, China and also with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China.

F. E. Alsaadi is with the Communication Systems and Networks Research Group, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia.

* Corresponding author (N. Zeng). Email: zny@xmu.edu.cn, Tel: +86-18695690380, Fax: +86-592-2182221.



(a) LGG (b) HGG

Fig. 1. MRI visualization of gliomas with different grades.

to the glioma grading task, above problem is reflected on identifying different modalities with highly similar imaging features, and it is also tough to effectively utilize both semantic and detailed information under different MRI modalities. In addition, most existing glioma grading algorithms have great reliance on the data with manually labeled tumor masks, while it is a time-consuming and laborious task to obtained those masks.

Based on above discussions, in this paper, a novel attention-based glioma grading network (AGGN) is proposed to overcome the mentioned challenges. In particular, by meticulously designing three modules to realize the function of dual-domain attention, multi-scale feature extraction and multi-modal information fusion, the proposed AGGN can efficiently capture potential correlations and key information from scattered features in different imaging modalities, and enhance the ability to distinguish intra-class variability and inter-class similarity existing at different grades of glioma. Based on the final fused highly discriminative features, the proposed AGGN can present comparable grading performance even without the manual labeled tumor masks. In addition, the ability of extracting features with strong presentation also guarantees the model robustness and generalization performance to some extents. Therefore, it is also feasible and promising to apply the developed AGGN into other MRI-based tasks, such as the diagnosis of Parkinson's disease and Alzheimer's disease [32], [49]. Major contributions of this paper are outlined as follows:

1) A novel brain MRI analysis method AGGN is proposed for grading glioma, which can reduce the reliance on supervised information of manual labels.
2) Designed multi-scale feature extraction and multi-modal information fusion modules benefit extracting discriminative features with strong presentation.
3) Evaluations on both internal and external brain MRI datasets have demonstrated superiority of the proposed AGGN, which yields satisfactory robustness and generalization ability.

The remainder of this paper is organized as follows. Related work on glioma grading is presented in Section II. The proposed AGGN and the key components are elaborated in Section III. Substantial experimental validations and comprehensive discussions are presented in Section IV. Finally, conclusions with an outlook of future work are drawn in Section V.

II. Related Work

In this section, related glioma grading methods are reviewed. As glioma grading is essentially an image classification task, representative feature extraction and fusion methods are briefly introduced as well.

A. Glioma Grading Methods

In clinic, CAD methods play an important role in grading glioma with brain MRI data, and as early as 2010, the authors in [56] have used a support vector machine (SVM) to realize the preliminary assessment of glioma grade and achieved accuracy of 82%. In [15], a two-level clustering method has been proposed for MRI preprocessing, after which an SVM is adopted to accomplish the glioma grading task. A combination of SVM and multi-layer perceptron has been adopted to glioma grading in [40], where feature selection is performed on tumor sub-regions of different modalities. In addition to SVM, other classic machine learning models have also been applied in this area, and one can refer to [17] for more information.

Owing to the rapid development of DL techniques, deep neural networks (DNNs) based glioma grading models have already become the mainstream, where convolutional neural network (CNN) is one of the most popular architectures, including 2D- and 3D-CNN according to dimension of the convolution operations. In [31], a lightweight 2D-CNN model has been developed with only basic components like convolution and pooling layers, and the proposed method has realized fast inference with low computational complexity. In [45], the authors have proposed a 3D-CNN, where the volume of interests is segmented to promote an efficient feature extraction. In particular, performance between 2D Mask R-CNN and 3D U-Net in glioma grading task has been compared in [53], and it is found that 2D model achieves higher sensitivity but lower specificity than the 3D one. Transfer learning paradigm has been introduced in [44], with two well-known CNN-based models AlexNet and GoogleNet, experimental results indicate that the pre-trained model can enhance the performance. Similarly, it is deemed in [52] that the pre-trained CNN model can extract high-dimensional information of feature maps, which benefits further grading of glioma with stronger presentation than the low-dimensional texture or shape features.

In addition, a three-stage DNN model has been developed in [36], which successively performs the rough contour segmentation, the precise contour extraction and the classification. In [1], the generative adversarial network has been utilized to solve the problems of limited samples in brain MRI. Meanwhile, an adaptive encoder has been employed to extract multi-modal features in [1], which finally achieves precision of 92% on the glioma grading task.

Although above methods have proven effective, following two important issues still deserve further improvement. Firstly, most grading models rely on tumor mask-based data, and it is difficult to achieve end-to-end training without manual annotation; secondly, it is of vital significance to efficiently capture and integrate multi-modal pathological glioma features from MRI data, which has not been well addressed in existing methods.

To overcome the mentioned problems, in the proposed AGGN, three modules are meticulously designed to realize dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion, so that highly discriminative features with strong presentation can be extracted. Details of the proposed AGGN are presented in Section III.

B. Feature Extraction and Fusion Methods

In the context of DL-based medical image processing, extracted robust features with strong presentation will have a great impact on the model performance. Representative CNN-based feature extractor include the visual geometry group (VGG) [35], GoogleNet [37], Inception v1-v4 [38], [39], residual network (ResNet) [12], re-parameterization VGG (RepVGG) [5], etc. In particular, VGG has reduced the amount of model parameters by stacking small convolution modules; GoogleNet and Inception v1 have utilized multi-branch architectures with multi-size convolutions to extract features; Inception v2-v4 models have further proposed the concept of BatchNorm, asymmetric decomposition convolution kernel and residual inception to enhance the performance; ResNet has solved the gradient explosion problem by skip connections in deep structures; in RepVGG, the dominance of both detection speed and accuracy have been achieved by decoupling the training process and inference stage.

Feature fusion is another important operation in many DL-based methods, which promotes sufficient integration of information at different levels so as to enhance the presentation ability of features and improve the model performance. One of the most representative feature fusion structures is the feature pyramid network (FPN) [21], which contains two pathways for bottom-up forward propagation and top-down sampling recovery, respectively, and the lateral connections in FPN have facilitated information fusion. It is noticeable that many FPN variants have been successfully proposed, such as path aggregation network (PANet) [27], bidirectional FPN (BiFPN) [41] and atrous spatial pyramid pooling-balanced FPN (ABFPN) [51], etc. Particularly, the ABFPN is an enhanced multi-scale feature fusion structure, which improves the model performance via sufficiently utilizing context information and generating balanced enhanced features with rich receptive fields.

III. Methodology

In this section, the proposed AGGN is elaborated with implementation details, including the designed dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion modules. To begin with, the overall framework of AGGN is illustrated in Fig. 2.

A. Overall Framework of AGGN

According to Fig. 2, in the proposed AGGN, firstly pre-processing operations including standardization, center cropping, modal splicing, data partitioning and augmentation are performed on the input brain MRI images with four modalities T1, T2, T1ce and FLAIR. Afterwards, the pre-processed multi-modal MRI data will enter the dual-domain attention mechanism module, where the weights are assigned in both channel and spatial dimensions to highlight the key information and suppress the unimportant one in feature maps. Next, the multi-modal MRI is split, and in the followed multi-scale feature extraction module, parallel processing is performed on each single-modality map, including sequential operations of multi-branch convolution (MB conv), convolution-pooling (C-P) and multi-branch pooling (MB pool), and the final output of each pathway is in the size of $1 \times 1 \times 256$. It is noticeable that during above procedure, maps with three sizes on each modality are individually concatenated and fed into the multi-modal information fusion module, which contains the fusion convolution and MB reduction operations. At last, seven feature maps in the same size are cascaded and fed into the linear layers, which is responsible to accomplish the glioma grading task.

In the following subsections, above mentioned three major modules of AGGN are presented with details.

B. Dual-domain Attention Mechanism

Attention mechanism is essentially a procedure of weighting features by pixel-wise operations in channel or spatial dimension, where the position that can reflect the detailed or semantic information of targets will be assigned large weights. In the proposed AGGN, a novel dual-domain attention mechanism is designed, and the structure is presented in Fig. 3. It is worth mentioning that the pre-processed input data are directly sent into the designed dual-domain attention module to model the target location and individual modalities, where different weights are assigned based on both channel and spatial importance of the features, so as to realize focused attention on the useful information and simultaneously suppress the useless one.

As is shown in Fig. 3, the input map will successively pass through the channel attention (CA) and spatial attention (SA) components. To be specific, in the former one, the size of feature maps in four modalities is compressed through operations of three parallel branches, where 1×1 convolution, asymmetric convolution block (ACB), average pooling and BN-PRelu (batch normalization and parametric Relu) operations are performed. It is noticeable that ACB replaces square convolution with asymmetric one equivalently [4], which can effectively avoid significant information loss and reduce the number of parameters. By concatenating the three branches, diverse information is shared and afterwards weights are assigned via activation operation and element-wise multiplication with the original input data.

In subsequent spatial attention component, average and maximum pooling layers are placed at first to compress the channel of feature maps, and the outputs are concatenated to enter series of ACB blocks to learn the parameters in spatial dimension. Similarly, after the sigmoid activation function, the spatial-domain weight assignment for different pixel regions is eventually achieved by element-wise multiplication. As a result, the applied dual-domain attention mechanism can figure

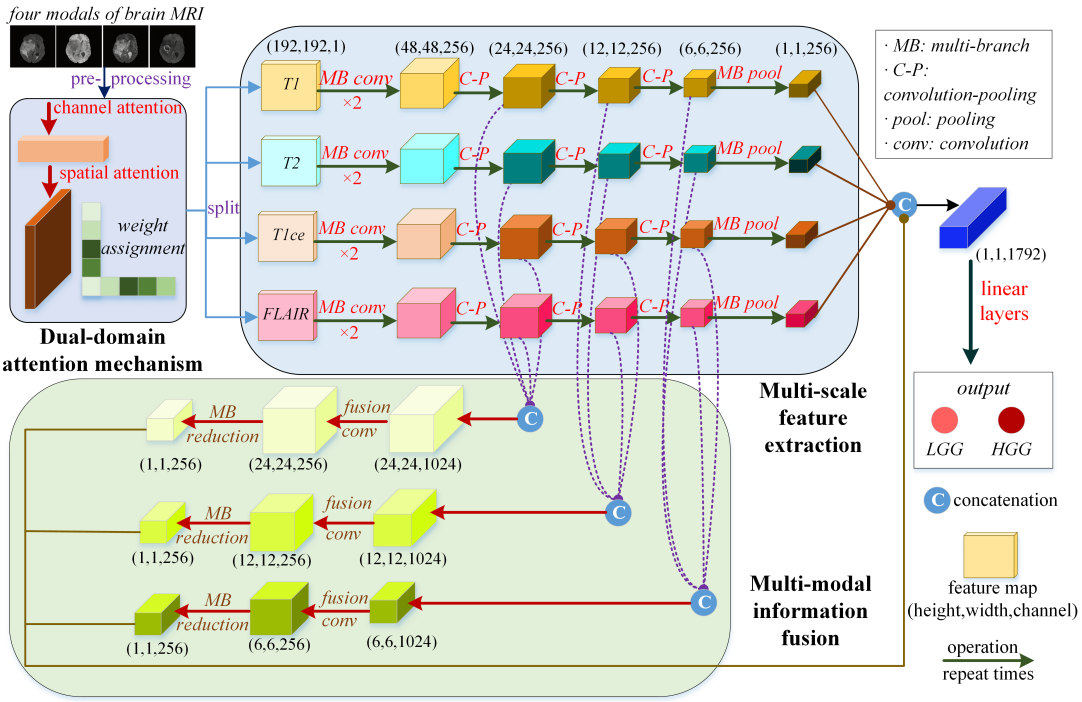


Fig. 2. Framework of the proposed attention-based glioma grading network (AGGN).

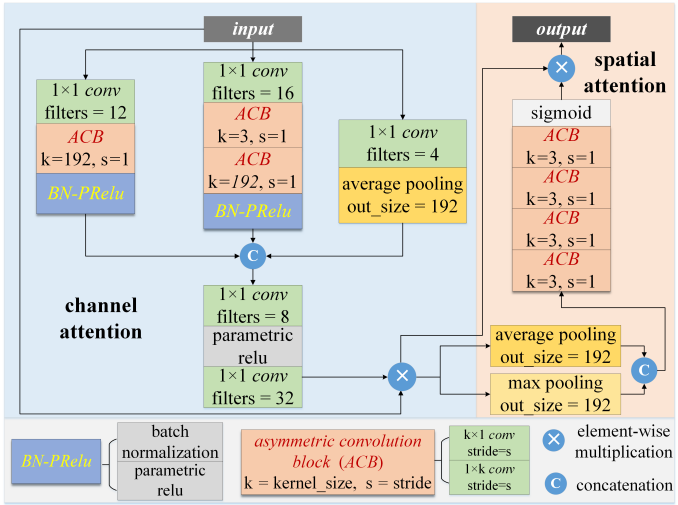


Fig. 3. Dual-domain attention mechanism module.

out both "where" and "what" the model should pay attention to. Work principle of above module is described by following equation.

$$output_{am} = SA\left\{M\left[CA\left(F_i^c\right)\right]_j^{x,y}\right\}, \quad (i = 1, ..., n;\ j = 1, 2, 3, 4) \tag{1}$$

where $F_{\cdot}^{c}$ and $M_{\cdot}^{(x,y)}$ refer to the feature map of channel c and modality of brain MRI in position (x, y), respectively; n is the number of feature maps, j denotes the modal and $output_{am}$ is the final output.

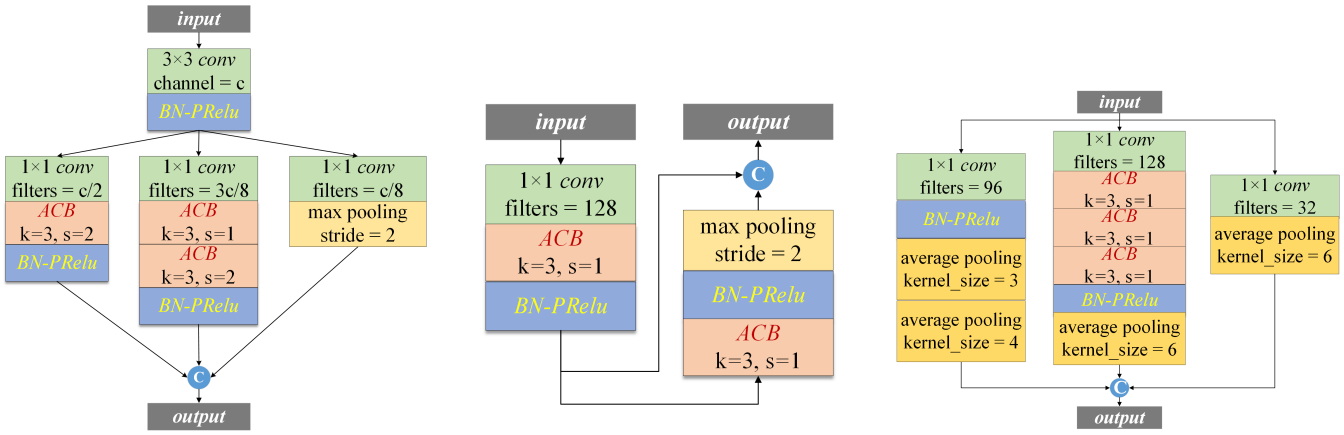
C. Multi-scale Feature Extraction

As previously mentioned, the four MRI modalities have contained rich pathological information with different concerns. To realize sufficient feature extraction on each modality, output of the dual-domain attention module is further split into four single-modal maps to enter the multi-scale feature extraction module (see Fig. 2), which can promote in-depth analysis of the key features enhanced by attention mechanism and can also benefit the subsequent multi-modal information fusion as well. On each branch, the involved MB conv, C-P and MB pool are displayed in Figs. 4(a)-4(c), respectively.

The MB conv block is used to extract the shallow features of each modal. As can be seen from Fig. 4(a), three parallel branches with different operations are included so that the extracted feature maps can contain rich information, and the last concatenation further integrates different features. Through MB conv block, the number of channels increases but the size of feature maps declines; and moreover, the applied ACB block can avoid large amount of information loss during the down-sampling procedure. In following Eq. 2, how MB conv block works is described.

$$\begin{aligned} output_{mc} = BP\left(AC_3^1\left(C_1\left(F_m\right)\right)\right) \oplus BP\left(AC_3^2\left(C_1\left(F_m\right)\right)\right) \\ \oplus MP\left(C_1\left(F_m\right)\right), \quad where\ F_m = BP\left(C_3\left(F_i\right)\right) \end{aligned} \tag{2}$$

where $output_{mc}$ is the block output, F_i and F_m refer to input and intermediate feature maps, respectively; $C_k\ (k = 1, 3)$ represents $k \times k$ standard convolution, and AC_3^m indicates the ACB operation with kernel size of three and m repetition times; BP and MP stand for BN-PRelu and maximum pooling operations, respectively.



(a) MB conv (b) C-P (c) MB pool

Fig. 4. Architectures of blocks in multi-scale feature extraction module.

Following the MB conv block, a series of C-P blocks are placed to continuously mine deep semantic features of each modality, where the ACB and maximum pooling operations with skip connection are adopted, which can be expressed as follows:

$$\begin{aligned} output_{cp} &= F_m \oplus MP\left(BP\left(AC_3^1\left(F_m\right)\right)\right), \\ where\ F_m &= BP\left(AC_3^1\left(C_1\left(F_i\right)\right)\right) \end{aligned} \tag{3}$$

It is noticeable that each branch has been equipped with three C-P blocks, and output of each C-P block will serve as input of the multi-modal information fusion module (see Fig. 2).

At the end of the multi-scale feature extraction module, MB pool blocks are deployed to generate the final output in size of $1 \times 1 \times 256$ for all modalities. It should be pointed out that the designing of MB pool is derived from improvement on the multi-receptive field pooling block in [42]. To be specific, multi-branch down-sampling is used in MB pool to convert the small-size map into a feature vector, and original convolution used in [42] is replaced by the asymmetric one with small kernels, which can expand the depth and enhance the feature extraction. In following Eq. (4), work principle of the MB pool is depicted.

$$\begin{aligned} output_{mp} = AP^2(BP(C_1(F_i))) \oplus AP^1(BP(AC^3(C_1(F_i)))) \\ \oplus AP^1(C_1(F_i)) \end{aligned} \tag{4}$$

where $output_{mp}$ is the block output, and AP^i $(i = 1, 2)$ denotes that the average pooling operation is repeated for i times.

D. Multi-modal Information Fusion Module

In the proposed AGGN, the multi-modal information fusion module is deployed to further integrate enhanced detailed and semantic features, and the structure is already illustrated in the green box of Fig. 2. In brief, fusion convolution realizes the integration of complementary advantages among the features of four modalities, where multi-scale feature maps in sizes of 24×24, 12×12 and 6×6 are fused. MB reduction block is adopted to transform the feature maps to vectors with strong presentation, which makes it feasible to further cascade the outputs of both multi-scale feature extraction and multi-modal information fusion module. MB reduction block consists of the sequential connection of the MB convolution (MC) and MB pool (MP) blocks, therefore, the block output $output_{mr}$ can be obtained by:

$$output_{mr} = MP(MC(F_i)) \tag{5}$$

where F_i denotes the input feature maps.

In addition, it is worth pointing out that the structure of fusion convolution block is similar to that of the C-P block, while the major difference is that the 2D convolution is replaced by the 3D one for fusion of feature maps with different modalities. According to Fig. 2, a finally constructed vector in size of $1 \times 1 \times 1792$ is fed into the last linear layers to obtain the glioma grading results, which is deemed to have strong presentation ability.

IV. Results and Discussions

In this section, the proposed AGGN is comprehensively evaluated on both internal and external public brain MRI dataset. In addition, substantial comparison experiments and ablation studies have been carried out to further validate the effectiveness and superiority of the proposed model. At first, experimental environment is briefly introduced.

A. Dataset Preprocessing and Experimental Settings

The experimental data used in this paper come from the 2018 and 2019 brain tumor segmentation (BraTS) challenges organized by medical image computing and computer assisted intervention society (MICCAI) [11], which are collected by 3T MRI systems of 17 institutions. The dataset includes multi-modal MRI from 326 glioma patients (250 for HGG, 76 for LGG), in which each case contains 155 slice data of four modalities, and the original size of each image is 240×240. In addition, professional radiologists have annotated and calibrated the edema, necrosis and core areas of glioma to obtain tumor masks, and grading results are determined through further pathological analysis.

The dataset is divided into training set, testing set and validation set, where the training set and testing set are independent of each other, while the validation set is obtained by further splitting the training set. To be specific, ratio of the training and testing set is 2 : 1, where the images of training set come from 2018 BraTS. The testing set includes the internal and external subsets, which contain images from 2018 and 2019 BraTS, respectively. It should be pointed out that the main difference between internal and external testing subset is that samples of the latter belong to different data-source as those of training samples, and neither of the two subsets participates in the model training. Furthermore, one-fifth of the training samples are picked out to form the validation set for model tuning and selection.

In addition, preprocessing is performed on the initial data before training the model, where tumor masks are used to screen tumor-free slices at first, and it is noticeable that the selected slices without tumor are not fed into the subsequent process. For slices that contain tumor tissues, the foreground region is standardized and the proportion of background is reduced so that they are center-cropped to 192×192 in size; afterwards, four modalities are treated as four channels of the image. Finally, the dataset is divided according to the previously mentioned rules, and data augmentation operations are only performed on the training samples, including random rotation, translation and clipping. For a clear view, above preprocessing steps and dataset division are shown in Fig. 5.

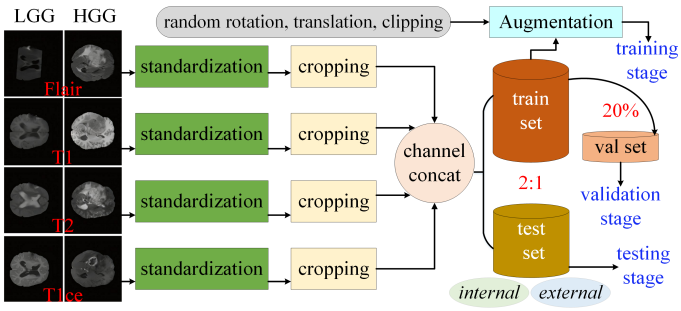


Fig. 5. Flowchart of preprocessing for brain MRI datasets.

All experiments in this study are carried out on the deep learning framework Pytorch, and the operating system is Windows 10 with NVIDIA GTX 2080Ti single GPU. Hyper-parameter settings are provided in Table I, as for model parameters, initialization of convolution and fully-connected layers adopts the Kaiming method and normal distribution, respectively.

TABLE I
HYPERPARAMETER SETTINGS

Variables	Values
Training epochs	100
Batch size	32
Optimizer	Adam
Initial learning rate	0.0001
First-order moment decay coefficient	0.9
Second-order moment decay coefficient	0.999

B. Performance Evaluation

To comprehensively evaluate performance of the proposed AGGN, four groups of experiments are carried out, which aim at verifying the generalization ability, architectural advantages, superiority against other representative CNN-based models and competitiveness in comparison to state-of-the-art glioma grading methods, respectively. Metrics $accuracy$, $precision$, $recall$, $specificity$, $F1\ score$ are adopted for performance evaluation, which can be calculated by following Eqs. (6)-(10):

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$specificity = \frac{TN}{FP + TN} \tag{9}$$

$$F1\ score = 2 \times \frac{recall \times precision}{recall + precision} \tag{10}$$

where TP/TN and FP/FN refer to the number of correct and wrong predictions of the HGG and LGG samples, respectively. As can be seen, the $accuracy$ describes the ratio of correct classifications of both HGG and LGG; the $precision$ aims at all samples predicted as HGG, and calculates the proportion of correct prediction; the $recall$ refers to the ratio of correctly identified HGG samples, which measures whether a model can screen all positive samples; similar to $recall$, the $specificity$ reflects the ability of identifying negative samples of a model; the $F1\ score$ takes the harmonic average between $accuracy$ and $recall$, and for all above five metrics, the larger their values are, the better the model performance is.

In addition, the receiver operating characteristic (ROC) curve and area under this curve (AUC) are also employed for the model evaluation. Specifically, the ROC curve takes the value of $1 - specificity$ (also known as false positive rate) as the horizontal axis and $recall$ as the vertical one, AUC is the area enclosed by ROC curve and the two coordinate axes.

1) Generalization ability of AGGN: At first, results obtained by the proposed AGGN on both internal and external testing sets are shown in Fig. 6, notice that in the former, training and testing samples share the same data-source; on the contrary, different sources are contained in the latter. As a result, this group of experiment can objectively reflect the generalization ability of the proposed AGGN. As is shown, the worst result is the $F1\ score$ on external dataset, which reaches 0.933; advantages of $precision$, $specificity$ and AUC are noticeable on both datasets, which validates that the propose AGGN is highly reliable in glioma grading task. At the same time, AGGN presents similar performance on internal and external testing sets, which demonstrates the robustness of AGGN in terms of handling various glioma MRI data; and

moreover, this result indicates that AGGN can adapt to data from multi-center medical institution with strong generalization ability.

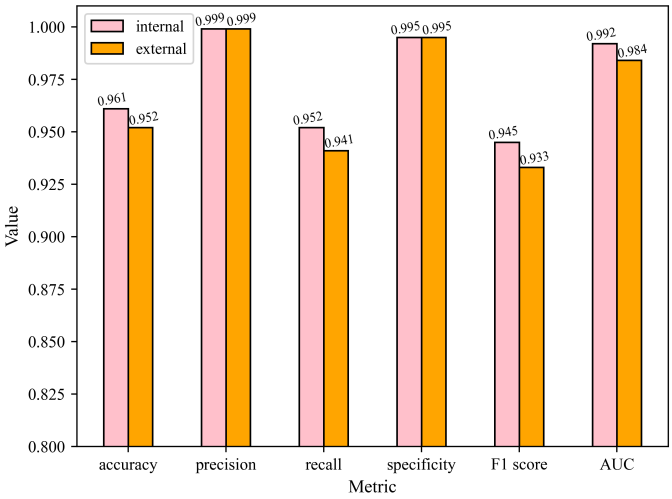


Fig. 6. Evaluation of AGGN on internal and external testing sets.

2) Architectural advantages of AGGN: In this part, to validate the architectural advantages of our method, adaptive multi-modal fusion network (AMMFNet) [42] is adopted as baseline model for comparison, which is a similar glioma grading framework to the proposed AGGN. In Fig. 7, performance enhancement of AGGN on six indicators is illustrated, which shows that in comparison to AMMFNet, the proposed AGGN has improved all indicators to different extents. In particular, the most significant improvement is on $specificity$, which increases 11.52%. As previously mentioned, high $specificity$ is equivalent to low false positive rate. Consequently, it is verified that AGGN has strong ability to correctly identify negative samples, which can effectively avoid the waste of medical resources. In addition, $precision$ is increased by 5.27%, which implies that AGGN is able to achieve accurate diagnosis of HGG. In this group of experiment, the proposed AGGN presents noticeable competitiveness in comparison to similar framework, which indicates the advantages in structural configuration.

TABLE II
PERFORMANCE COMPARISON OF PROPOSED AGGN AND FOUR CLASSIC MODELS ON INTERNAL TESTING SET

Metrics	Models				
	[12]	[14]	[35]	[39]	AGGN
$accuracy$	0.9013	0.9038	0.8785	0.9330	**0.9612**
$precision$	0.7763	0.9632	0.8848	0.9687	**0.9987**
$recall$	0.9234	0.9119	0.9476	0.9181	**0.9521**
$specificity$	0.8320	0.8747	0.7235	0.9749	**0.9952**
$F1\ score$	0.8687	0.8676	0.8506	0.9046	**0.9450**
AUC	0.9530	0.9570	0.9480	0.9780	**0.9920**

3) Comparisons with other CNN-based models: In order to further validate the competitiveness of the proposed AGGN, four other representative CNN-based models are adopted for comparison in this group of experiments, including ResNet-50 [12], DenseNet-101 [14], VGG-19 [35] and Inception-v4 [39]. For fairness, all models share the same training and testing data, and the results on internal and external datasets are reported in Table. II and Table. III, respectively. In addition, an illustration is presented in Fig. 8.

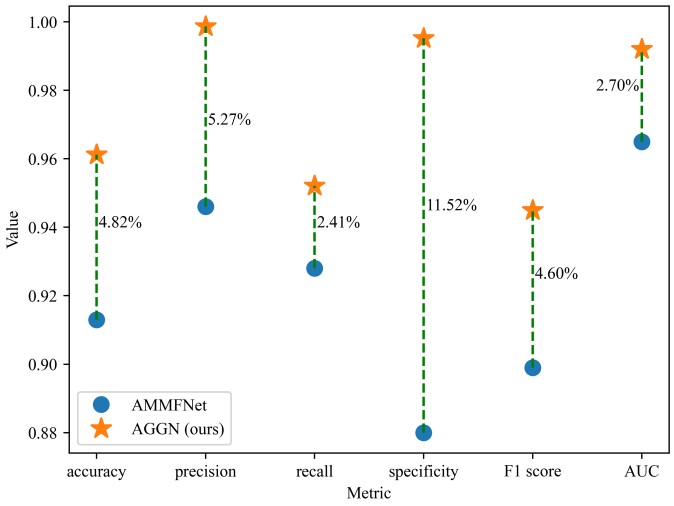


Fig. 7. Performance comparison between AGGN and AMMFNet.

TABLE III
PERFORMANCE COMPARISON OF PROPOSED AGGN AND FOUR CLASSIC MODELS ON EXTERNAL TESTING SET

Metrics	Models				
	[12]	[14]	[35]	[39]	AGGN
$accuracy$	0.8967	0.9052	0.8780	0.9446	**0.9519**
$precision$	0.9304	0.9672	0.8688	**0.9987**	**0.9987**
$recall$	0.9323	0.9122	**0.9658**	0.9325	0.9401
$specificity$	0.7890	0.8775	0.6937	0.9948	**0.9953**
$F1\ score$	0.8614	0.8639	0.8503	0.9193	**0.9334**
AUC	0.9410	0.9570	0.9580	0.9700	**0.9840**

As can be seen from Table II, all the indicators of AGGN are better than those of other representative CNN models on internal dataset, which are 2.82%, 3.0%, 0.45%, 2.03%, 4.04% and 1.4% higher than the sub-optimal model on $accuracy$, $precision$, $recall$, $specificity$, $F1\ score$ and AUC respectively. While on the external testing set, the proposed AGGN also achieves satisfactory results of 95.19%, 99.87%, 94.01%, 99.53%, 93.34% and 98.40% on above six metrics, respectively. On five out of the six indicators, AGGN has obtained the best results.

In addition, the ROC curves with magnification on the two testing sets are presented in Fig. 9, which can effectively evaluate the diagnostic ability of a model and can maintain strong stability when the distribution of positive and negative samples changes. Notice that the curve close to the upper left corner has high prediction accuracy, and accordingly, the larger AUC value, the better model performance is. As shown in Fig. 9, the ROC curve of AGGN is above all other models on both the internal and external testing sets, and the AUC

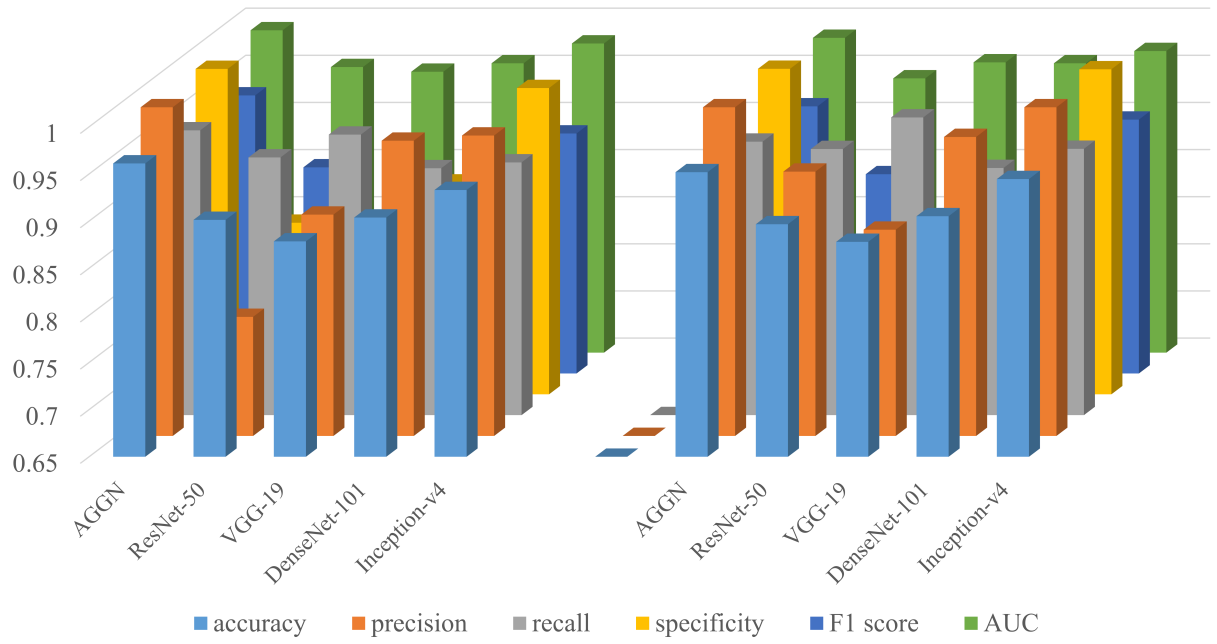
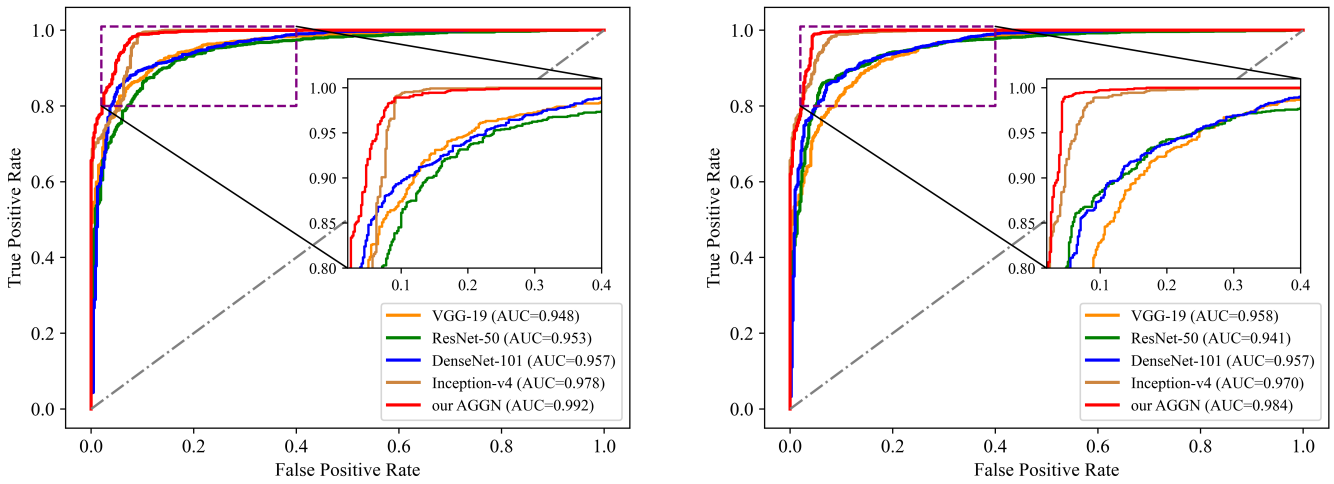


Fig. 8. Performance comparison of AGGN with advanced CNN-based models on internal (left) and external (right) testing sets.



(a) On internal testing set

(b) On external testing set

Fig. 9. ROC curves of the proposed AGGN and other CNN models.

values reach 0.992 and 0.984, respectively. Consequently, the strong generalization ability of AGGN while facing different data sources is further verified.

Through this group of experiment, it is demonstrated that the proposed AGGN has overwhelming overall performance against other advanced CNN-based and domain-specific models on most metrics, which may owe to the meticulously designed and introduced dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion modules.

4) Comparisons with state-of-the-art glioma grading algorithms: In this part, comparison between proposed AGGN and other state-of-the-art glioma grading algorithms are presented, including multistream CNN [9], multi-scale CNN [10], CAE-GAN (convolutional autoencoder and generative adversarial network) [1], pre-trained GoogleNet [44], 3DConvNet [53] and AMMFNet [42]. It should be pointed out that in previous experiments, data from the mentioned internal and external testing sets have none of the tumor masks, while most of recently related methods for the same task require the assistance of additional tumor masks. Consequently, to make a fair comparison, in this group of experiments, tumor masks have been added to original images for training, and the results are reported in Table IV. Notice that the data of other algorithms are cited from corresponding original papers, and "–" denotes none of relevant data is provided.

TABLE IV
PERFORMANCE OF AGGN AND OTHER ADVANCED MODELS ON TUMOR MASK ASSISTED DATA

Models	Metrics			
	accuracy	*recall*	*specificity*	*AUC*
Multistream CNN	0.9087	–	–	–
Multi-scale CNN	0.8947	–	–	–
CAE-GAN	0.9204	–	–	–
Pre-trained GoogleNet	0.9450	–	–	0.9680
3DConvNet	0.9710	0.9470	0.9680	–
AMMFNet	0.9820	**1.0000**	0.9330	0.9970
AGGN (ours)	**0.9899**	**1.0000**	**0.9678**	**0.9980**

As can be found in Table IV, with assistance of tumor masks, the proposed AGGN can present the state-of-the-art performance. In particular, the AUC value of AGGN with and without tumor masks are 0.998 and 0.992, respectively, which implies that the assistance of tumor masks does further improve the model performance. It is also worth mentioning that the AUC of 0.992 (without masks) is already an excellent result. Hence, it can be concluded that performance of AGGN has little reliance on the masks, which demonstrates that AGGN can overcome the high dependence of manually labeled annotations so as to achieve the end-to-end applications in practice.

C. Ablation Study

To validate the effectiveness of core components in the proposed AGGN, substantial ablation studies are performed on the internal testing set in this subsection. The designed dual-domain attention mechanism is firstly verified and the results are reported in Table V, where AGG1, AGG2 and AGG3 refer to the model with none of attention modules, only spatial and only channel attention module, respectively. Obviously, in comparison to AGG1, on most indicators the performance has been improved to a certain extent after introducing the spatial or channel attention mechanism. It is also found that the applied dual-domain attention module in the proposed AGGN has realized significant performance enhancement, which improves $accuracy$, $recall$, $F1$ $score$ and AUC by 2.92%, 3.61%, 4.23% and 1.6%, respectively.

In addition, the ROC curves of four models listed in Table V are presented in Fig. 10, where it can be seen that the proposed AGGN has obtained the best results. In particular, when false positive rate is 0, the true positive rate of AGGN is close to 0.85 and AUC is 0.992, which implies that the proposed AGGN can accurately identify HGG with almost none of false detection. Therefore, the proposed AGGN is a reliable model that can provide a solid guarantee for the diagnosis and treatment of critical patients.

According to above results, effectiveness of the dual-domain attention mechanism is sufficiently validated. Before extracting multi-scale features, channel-domain attention is firstly introduced to low-level detail information, which determines what deserves attention in each modality of brain MRI; afterwards, spatial-domain attention is used to learn spatial dependence among high-level semantic information, so as to figure out the important locations in feature maps. As a result, the proposed AGGN can both recognize and localize the significant pathological features in brain MRI with strong robustness.

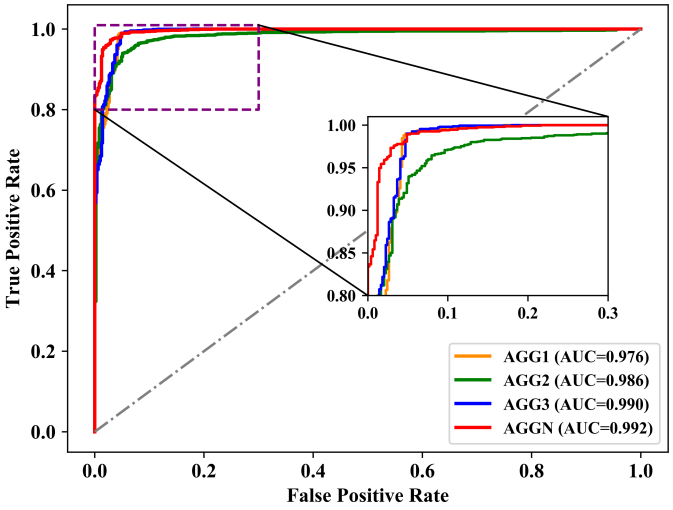


Fig. 10. ROC curves of AGG1, AGG2, AGG3, and AGGN.

In the following, ablation study results on multi-branch convolution block and the multi-modal information fusion module are reported in Table VI. It should be pointed out that except for the investigated components, other configurations of AGGN remain unchanged so as to make objective and convincing comparisons.

Firstly, as the most important component of the multi-scale feature extraction module, MB conv block is compared with multi-receptive field (MRF) conv block of AMMFNet. As can be seen from Table VI, the MB conv block has overwhelmed the MRF conv block on all metrics, which demonstrates that the multi-branch structure has superiority in dealing with glioma grading task based on brain MRI. Further explorations of the essential mechanism show that the volume of glioma from patients can be quite different, whereas the proposed MB conv block has adopted ACBs with different sizes to extract image features in parallel, which can obtain fine-grained texture and tissue information of multi-modal brain MRI. Therefore, it can be inferred that the multi-scale feature extraction module has made great contribution to the overall model performance.

Secondly, the designed multi-modal information fusion module is compared with the approach in [42]. As reported in Table VI, on four out of six indicators, proposed AGGN has achieved slight performance improvement. Specifically, AGGN improves $accuracy$, $recall$, $F1$ $score$ and AUC by 3.4%, 0.6%, 0.3% and 0.7%, respectively. It is worth mentioning that in the proposed AGGN, final input vector to the classifier is a concatenation of outputs from seven branches, and this number is fewer than that in [42]. Consequently, it can be concluded that AGGN has achieved comparable results to the model in [42] with a simplified structure. Additionally, the concatenation manner in AGGN has avoided stacking redundant features, which not only benefits highly-efficient feature fusion, but also simultaneously avoids excessively

TABLE V
ABLATION STUDIES OF DUAL-DOMAIN ATTENTION MECHANISM ON INTERNAL TESTING SET

Models	Metrics					
	accuracy	*precision*	*recall*	*specificity*	*F1 score*	*AUC*
AGG1	0.9320	1.0000	0.9160	1.0000	0.9027	0.9760
AGG2	0.9481	0.9866	0.9467	0.9532	0.9270	0.9860
AGG3	0.9476	**1.0000**	0.9349	**1.0000**	0.9238	0.9900
AGGN	**0.9612**	0.9987	**0.9521**	0.9952	**0.9450**	**0.9920**

TABLE VI
ABLATION STUDIES OF MB CONV BLOCK AND INFORMATION FUSION METHODS

Models	Metrics					
	accuracy	*precision*	*recall*	*specificity*	*F1 score*	*AUC*
MRF conv block of AMMFNet	0.9577	0.9960	0.9502	0.9856	0.9400	0.9840
MB conv block of AGGN	**0.9612**	**0.9987**	**0.9521**	**0.9952**	**0.9450**	**0.9920**
Fusion method in [42]	0.9578	**1.0000**	0.9461	**1.0000**	0.9420	0.9850
Fusion method of AGGN	**0.9612**	0.9987	**0.9521**	0.9952	**0.9450**	**0.9920**

complicating the model.

D. Computational Complexity Analysis

In this study, the number of model parameters (Params) and floating point operations (FLOPs) are adopted to depict the spatial and time complexity of the proposed AGGN, respectively. Excessive parameters will impede the light-weight deployment of model on edge devices, and too large FLOPs will influence the convergence during model training, which directly determines the accuracy of the model inference.

On the one hand, Params of the proposed AGGN is 16.37M, which is 9.13M fewer than that of the classical ResNet-50 model. According to Table III, the accuracy of AGGN is even 5.52% higher than that of ResNet-50, which demonstrates that the developed AGGN can effectively balance the computational costs and accuracy. It may owe to the proposed AGGN has effectively reduced the parameters by replacing large-size kernels with a series of small-sized ones. Meanwhile, the accuracy of AGGN is mainly guaranteed by the structural advantages, including employing dual-domain attention mechanism to highlight key features, realizing feature extraction on each individual modality, and integrating multi-modal information in different levels.

On the other hand, the FLOPs of AGGN are 24,790M, which mainly due to the large size of multi-channel input samples, where the data processing has consumed great deals of the FLOPs. It is also worth mentioning that during the model training, none of obvious over-fitting phenomenon has occurred, which implies that the training and inference time consumed by the proposed AGGN is acceptable.

To sum up, the proposed AGGN can effectively achieve the balance between model complexity and accuracy, which has achieved satisfactory results in the glioma grading task with considerable efficiency.

V. CONCLUSION

In this paper, a novel self-attention based network AGGN has been developed, which mainly consists of three meticulously designed modules, including a dual-domain attention module, a multi-scale feature extraction and a multi-modal information fusion one. Robust features with strong presentation ability are constructed by integrating outputs of the latter two modules, which are used to eventually realize the glioma grading task. Performance of the proposed AGGN has been comprehensively evaluated on both internal and external testing sets, and the results have demonstrated the superiority of AGGN against other state-of-the-art algorithms. Furthermore, substantial ablation studies have verified effectiveness of the designed three modules in AGGN, which can take full advantages of detailed and semantic information so that model performance can be greatly improved, and simultaneously computational burdens are released to some extent.

Although the proposed AGGN has presented satisfactory performance on the glioma grading task, it still has some spaces for further improvement, including task migration adaption, quantitative lesion analysis, and model lightweighting studies. In future work, we aim to 1) apply the developed AGGN framework to other MRI-based tasks such as stroke and cancer diagnosis; 2) investigate fine-grained glioma grading methods to support quantitative analysis; 3) further optimize the structure of AGGN through fuzzy system and tensor decomposition techniques. [18], [23], [26], [43]

REFERENCES

[1] M. Ali, I. Gu and A. Jakola, "Multi-stream convolutional autoencoder and 2D generative adversarial network for glioma classification", *Procedding of the 18th International Conference on Computer Analysis of Images and Patterns (CAIP)*, pp. 234-245, 2019.

[2] G. Bao, L. Ma and X. Yi, "Recent advances on cooperative control of heterogeneous multi-agent systems subject to constraints: A survey", *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 539-551, 2022.

[3] J. Cheng, J. Liu, H. Kuang and J. Wang, "A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping", *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1520-1532, 2022.

[4] X. Ding, Y. Guo, G. Ding and J. Han, "ACNet: strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks", *Procedding of the 17th IEEE International Conference on Computer Vision (ICCV)*, pp. 1911-1920, 2019.

[5] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding and J. Sun, "RepVGG: making VGG-style ConvNets great again", *Procedding of the 34th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13728-13737, 2021.

[6] E. Donaldson and A. Kirk, "Is MRI the superior test for investigating thymic pathologies in comparison to CT?", *Lung Cancer*, vol. 156, no. s1, pp. S23-S23, 2021.

[7] E. Ferrari, P. Bosco, S. Calderoni, P. Oliva, L. Palumbo, G. Spera, M. Fantacci and A. Retico, "Dealing with confounders and outliers in classification medical studies: the autism spectrum disorders case study", *Artificial Intelligence in Medicine*, vol. 108, article no. 101926, 2020.

[8] Y. Fei, B. Zhan, M. Hong, X. Wu, J. Zhou and Y. Wang, "Deep learning-based multi-modal computing with feature disentanglement for MRI image synthesis", *Medical Physics*, vol. 48, no. 7, pp. 3778-3789, 2021.

[9] C. Ge, I. Gu, A. Jakola and J. Yang, "Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks", *Procedding of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5894-5897, 2018.

[10] C. Ge, Q. Qu, I. Gu and A. Store Jakola, "3D multi-scale convolutional networks for glioma grading using MR images", *Procedding of the 25th IEEE International Conference on Image Processing (ICIP)*, pp. 141-145, 2018.

[11] M. Ghaffari, A. Sowmya and R. Oliver, "Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges", *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 156-168, article no. 2946868, 2020.

[12] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *Procedding of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[13] J. Hu, C. Jia, H. Liu, X. Yi and Y. Liu, "A survey on state estimation of complex dynamical networks", *International Journal of Systems Science*, vol. 52, no. 16, pp. 3351-3367, 2021.

[14] G. Huang, Z. Liu, L. Van Der Maaten and K.Q. Weinberger, "Densely connected convolutional networks", *Procedding of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261-2269, 2017.

[15] R. Inano, N. Oishi, T. Kunieda, Y. Arakawa, Y. Yamao, S. Shibata, T. Kikuchi, H. Fukuyama and S. Miyamoto, "Voxel-based clustered imaging by multiparameter diffusion tensor images for glioma grading", *Neuroimage-clinical*, vol. 5, pp. 396-407, 2014.

[16] S. Ismael, A. Mohammed and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks", *Artificial Intelligence in Medicine*, vol. 102, article no. 101779, 2020.

[17] J. Jeong, L. Wang, B. Ji, Y. Lei, A. Ali, T. Liu, W. Curran, H. Mao and X. Yang, "Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images", *Quantitative Imaging in Medicine and Surgery*, vol. 9, no. 7, pp. 1201-1213, 2019.

[18] J. Mao, Y. Sun, X. Yi, H. Liu and D. Ding, "Recursive filtering of networked nonlinear systems: A survey", *International Journal of Systems Science*, vol. 52, no. 6, pp. 1110-1128, 2021.

[19] L. Marginean, P. Stefan, A. Lebovici, I. Opincariu, C. Csutak, R. Lupean, P. Coroian and B. Suciu, "CT in the differentiation of gliomas from brain metastases: The radiomics analysis of the peritumoral zone", *Brain Sciences*, vol. 12, no. 1, article no. 109, 2022.

[20] M. Mittler, B. Walters and E. Stopa, "Observer reliability in histological grading of astrocytoma stereotactic biopsies", *Journal of Neurosurgery*, vol. 85, no. 6, pp. 1091-1094, 1996.

[21] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection", *Procedding of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936-944, 2017.

[22] G. Li, L. Li, Y. Li, Z. Qian, F. Wu, Y. He, H. Jiang, R. Li, D. Wang, Y. Zhai, Z. Wang, T. Jiang, J. Zhang and W. Zhang, "An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas", *Brain*, vol. 145, no. 3, pp. 1151-1161, 2022.

[23] H. Li, P. Wu, Z. Wang, J. Mao, Fuad E. Alsaadi and N. Zeng, "A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis", *Computers in Biology and Medicine*, vol. 151, article no. 106265, 2022.

[24] H. Li, P. Wu, N. Zeng, Y. Liu and Fuad E. Alsaadi, "A Survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from Systems Science Perspective", *International Journal of Systems Science*, 2022. https://doi.org/10.1080/00207721.2022.2083262.

[25] H. Li, N. Zeng, P. Wu and K. Clawson, "Cov-Net: A computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision", *Expert Systems with Applications*, vol. 207, article no. 118029, 2022.

[26] W. Li, Y. Niu and Z. Cao, "Event-triggered sliding mode control for multi-agent systems subject to channel fading", *International Journal of Systems Science*, vol. 53, no. 6, pp. 1233-1244, 2022.

[27] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path aggregation network for instance segmentation", *Procedding of the 31th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8759-8768, 2018.

[28] D. Louis, A. Perry, P. Wesseling, D. Brat, I. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling and D. Ellison, "The 2021 WHO classification of tumors of the central nervous system: a summary", *Neuro-Oncology*, vol. 23, no. 8, pp. 1231-1251, 2021.

[29] P. Lu, B. Song and L. Xu, "Human face recognition based on convolutional neural network and augmented dataset", *Systems Science & Control Engineering*, vol. 9, no. s2, pp. 29-37, 2021.

[30] Q. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko and J. Barnholtz-Sloan, "Corrigendum to: CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017", *Neuro-Oncology*, vol. 24, no. 7, pp. 1214-1310, 2022.

[31] H. Ozcan, B. Emiroglu, H. Sabuncuoglu, S. Ozdogan, A. Soyer and T. Saygi, "A comparative study for glioma classification using deep convolutional neural networks", *Mathematical Biosciences and Engineering*, vol. 18, no. 2, pp. 1550-1572, 2021.

[32] G. Pahuja and B. Prasad, "Deep learning architectures for Parkinson's disease detection by using multi-modal features", *Computers in Biology and Medicine*, vol. 146, article no. 105610, 2022.

[33] C. Sarasaen, S. Chatterjee, M. Breitkopf, G. Rose, A. Nurnberger and O. Speck, "Fine-tuning deep learning model parameters for improved super-resolution of dynamic MRI with prior-knowledge", *Artificial Intelligence in Medicine*, vol. 121, article no. 102196, 2021.

[34] B. Song, H. Miao and L. Xu, "Path planning for coal mine robot via improved ant colony optimization algorithm", *Systems Science & Control Engineering*, vol. 9, no. 1, pp. 283-289, 2021.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *Procedding of the 3rd International Conference on Learning Representations (ICLR)*, PP. 1-14, 2015.

[36] M. Soleymanifard and M. Hamghalam, "Multi-stage glioma segmentation for tumour grade classification based on multiscale fuzzy C-means", *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8451-8470, 2022.

[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions", *Procedding of the 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision", *Procedding of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818-2826, 2016.

[39] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning", *Procedding of the 31th AAAI Conference on Artificial Intelligence (AAAI)*, pp.4278-4284, 2017.

[40] P. Sun, D. Wang, V. Mok and L. Shi, "Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading", *IEEE Access*, vol. 7, pp. 102010-102020, article no. 2928975, 2019.

[41] M. Tan, R. Pang and Q. Le, "EfficientDet: scalable and efficient object detection", *Procedding of the 33th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10778-10787, 2020.

[42] L. Wang, Y. Cao, L. Tian, Q. Chen, S. Guo, J. Zhang and LH. Wang, "Adaptive multi-modality fusion network for glioma grading", *Journal of Image and Graphics*, vol. 26, no. 9, pp. 2243-2256, 2021.

[43] Y. Wang, L. Zou, L. Ma, Z. Zhao and J. Guo, "A survey on control for Takagi-Sugeno fuzzy systems subject to engineering-oriented complexities", *Systems Science & Control Engineering*, vol. 9, no. 1, pp. 334-349, 2021.

[44] Y. Yang, L. Yan, X. Zhang, Y. Han, H. Nan, Y. Hu, B. Hu, S. Yan, J. Zhang, D. Cheng, X. Ge, G. Cui, D. Zhao and W. Wang, "Glioma grading on conventional MR images: a deep learning study with transfer learning", *Frontiers in Neuroscience*, vol. 12, article no. 804, 2018.

[45] H. Yamashiro, A. Teramoto, K. Saito and H. Fujita, "Development of a fully automated glioma-grading pipeline using post-contrast T1-weighted images combined with cloud-based 3D convolutional neural network", *Applied Sciences-Basel*, vol. 11, no. 11, article no. 5118, 2021.

[46] Y. Yuan, G. Ma, C. Cheng, B. Zhou, H. Zhao, H.-T. Zhang and H. Ding, "A general end-to-end diagnosis framework for manufacturing systems", *National Science Review*, vol. 7, no. 2, pp. 418-429, 2020.

[47] Y. Yuan, X. Tang, W. Zhou, W. Pan, X. Li, H.-T. Zhang, H. Ding and J. Goncalves, "Data driven discovery of cyber physical systems", *Nature Communications*, vol. 10, no. 1, pp. 1-9, 2019.

[48] Y. Yuan, H. Zhang, Y. Wu, T. Zhu and H. Ding, "Bayesian learning-based model-predictive vibration control for thin-walled workpiece machining processes", *IEEE/ASME transactions on mechatronics*, vol. 22, no. 1, pp. 509-520, 2016.

[49] N. Zeng, H. Li and Y. Peng, "A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease", *Neural Computing and Applications*, article no. 06149-6, 2021.

[50] N. Zeng, Z. Wang, W. Liu, H. Zhang, K. Hone and X. Liu, "A dynamic neighborhood-based switching particle swarm optimization algorithm", *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9290-9301, 2022.

[51] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection", *IEEE Transactions on Instrumentation and Measurement*, vol. 71, article no. 3507014, 2022.

[52] Z. Zhang, J. Xiao, S. Wu, F. Lv, J. Gong, L. Jiang, R. Yu and T. Luo, "Deep convolutional radiomic features on diffusion tensor images for classification of glioma grades", *Journal of Digital Imaging*, vol. 33, no. 4, pp. 826-837, 2020.

[53] Y. Zhuge, H. Ning, P. Mathen, J. Cheng, A. Krauze, K. Camphausen and R. Miller, "Automated glioma grading on conventional MRI images using deep convolutional neural networks", *Medical Physics*, vol. 47, no. 7, pp. 3044-3053, 2020.

[54] B. Zinnhardt, F. Roncaroli, C. Foray, E. Agushi, B. Osrah, G. Hugon, A. Jacobs and A. Winkeler, "Imaging of the glioma microenvironment by TSPO PET", *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 1 pp. 174-185, 2021.

[55] L. Zou, Z. Wang, J. Hu, Y. Liu and X. Liu, "Communication-protocol-based analysis and synthesis of networked systems: Progress, prospects and challenges", *International Journal of Systems Science*, vol. 52, no. 14, pp. 3013-3034, 2021.

[56] F. Zollner, K. Emblem and L. Schad, "Support vector machines in DSC-based glioma imaging: suggestions for optimal characterization", *Magnetic Resonance in Medicine*, vol. 64, no. 4, pp. 1230-1236, 2010.

AGGN: Attention-based Glioma Grading Network with Multi-scale Feature Extraction and Multi-modal Information Fusion

Peishu Wu, Zidong Wang, Baixun Zheng, Han Li, Fuad E. Alsaadi and Nianyin Zeng*

***Abstract*—In this paper, a magnetic resonance imaging (MRI) oriented novel attention-based glioma grading network (AGGN) is proposed. By applying the dual-domain attention mechanism, both channel and spatial information can be considered to assign weights, which benefits highlighting the key modalities and locations in the feature maps. Multi-branch convolution and pooling operations are applied in a multi-scale feature extraction module to separately obtain shallow and deep features on each modality, and a multi-modal information fusion module is adopted to sufficiently merge low-level detailed and high-level semantic features, which promotes the synergistic interaction among different modality information. The proposed AGGN is comprehensively evaluated through extensive experiments, and the results have demonstrated the effectiveness and superiority of the proposed AGGN in comparison to other advanced models, which also presents high generalization ability and strong robustness. In addition, even without the manually labeled tumor masks, AGGN can present considerable performance as other state-of-the-art algorithms, which alleviates the excessive reliance on supervised information in the end-to-end learning paradigm.**

***Index Terms*—Artificial intelligence; glioma grading; feature extraction; information fusion; magnetic resonance imaging (MRI)**

I. Introduction

As one of the most common primary tumors caused by the cancerization of glial cells in the brain or spinal cord, glioma accounts for nearly half of intracranial tumors and 36% of the nervous system tumors [30]. According to the criteria of the World Health Organization (WHO), glioma can be graded as four levels from I to IV [28], where low-grade glioma (LGG) includes grades I-II, and grades III-IV are the so-called high-grade glioma (HGG). It is worth mentioning that LGG may be cured by drug therapy and surgical excision, whereas radiotherapy and chemotherapy are required to cure HGG. Moreover, patients with the glioma of grade IV even suffer from low survival rate less than 10% [20]. Consequently, it is of vital significance to realize accurate preoperative grading of glioma.

In clinical practice, the glioma grading task is mainly accomplished by imaging diagnosis [2], [13], [34], including computed tomography (CT) [19], magnetic resonance imaging (MRI) [22] and positron emission tomography (PET) [54], etc. Particularly, the MRI has advantages of strong specificity and sensitivity in tumor localization and pathological analysis [6], which can generate multi-modal images to reflect brain feature information at different levels by modulating imaging parameters. To be specific, four modalities are included in MRI [3], where the T1-weighted (T1) modality displays brain anatomy, and the T2-weighted (T2) one locates lesion area; fluid-attenuated inversion recovery (FLAIR) and T1-weighted contrast-enhanced (T1ce) modalities are generally used for visualization of peritumor and internal conditions so as to make further pathological analysis [8]. Due to above four modalities can sufficiently present structural and functional information of tumors, the multi-modal MRI technique has become an important diagnostic tool of grading glioma in clinic.

It is worth pointing out that inspecting the diverse information provided by multi-modal MRI is a laborious task, which inevitably increases the workload on radiologists or neurosurgeons. For a clear view, the MRI slices of LGG and HGG are illustrated in Fig. 1(a) and Fig. 1(b), respectively. Experienced doctors generally distinguish LGG from HGG by observing the clarity of tumor contour and the presence of edema in peritumor areas. Hence, it can be inferred that both the high heterogeneity of brain tumors and experiences of doctors will influence the final diagnostic results. As a result, it is necessary and beneficial to develop computer-aided diagnosis (CAD) systems to realize accurate glioma grading with less manpower [46]–[48].

Owing to the continuous development of the deep learning (DL) techniques [29], [55], plenty of CAD methods have been proposed and applied to the intelligent analysis of brain MRI [7], [16], [25], [33], and related studies regarding to glioma grading tasks are reviewed in Section II. It is noticeable that in the context of applying DL-based models for medical image analysis [24], [50], a common and challenging issue is the robust feature extraction, which has great impact on the downstream tasks (e.g., segmentation and classification). Regarding

This research work was funded by Institutional Fund Projects under grant no. (IFPIP: 30-135-1443). The authors gratefully acknowledge the technical and financial support provided by the Ministry of Education and King Abdulaziz University, DSR, Jeddah, Saudi Arabia.

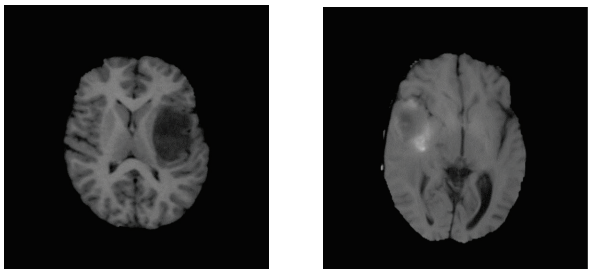
P. Wu, H. Li and N. Zeng are with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China. Email: zny@xmu.edu.cn

Z. Wang is with the Department of Computer Science, Brunel University London, Uxbridge UB8 3PH, U.K. Email: zidong.wang@brunel.ac.uk

B. Zheng is with the Polytechnic Institute, Zhejiang University, Hangzhou 310015, China and also with the Department of Instrumental and Electrical Engineering, Xiamen University, Fujian 361005, China.

F. E. Alsaadi is with the Communication Systems and Networks Research Group, Department of Electrical and Computer Engineering, Faculty of Engineering, King Abdulaziz University, Jeddah, Saudi Arabia.

* Corresponding author (N. Zeng). Email: zny@xmu.edu.cn, Tel: +86-18695690380, Fax: +86-592-2182221.



(a) LGG (b) HGG

Fig. 1. MRI visualization of gliomas with different grades.

to the glioma grading task, above problem is reflected on identifying different modalities with highly similar imaging features, and it is also tough to effectively utilize both semantic and detailed information under different MRI modalities. In addition, most existing glioma grading algorithms have great reliance on the data with manually labeled tumor masks, while it is a time-consuming and laborious task to obtained those masks.

Based on above discussions, in this paper, a novel attention-based glioma grading network (AGGN) is proposed to overcome the mentioned challenges. In particular, by meticulously designing three modules to realize the function of dual-domain attention, multi-scale feature extraction and multi-modal information fusion, the proposed AGGN can efficiently capture potential correlations and key information from scattered features in different imaging modalities, and enhance the ability to distinguish intra-class variability and inter-class similarity existing at different grades of glioma. Based on the final fused highly discriminative features, the proposed AGGN can present comparable grading performance even without the manual labeled tumor masks. In addition, the ability of extracting features with strong presentation also guarantees the model robustness and generalization performance to some extents. Therefore, it is also feasible and promising to apply the developed AGGN into other MRI-based tasks, such as the diagnosis of Parkinson's disease and Alzheimer's disease [32], [49]. Major contributions of this paper are outlined as follows:

1) A novel brain MRI analysis method AGGN is proposed for grading glioma, which can reduce the reliance on supervised information of manual labels.
2) Designed multi-scale feature extraction and multi-modal information fusion modules benefit extracting discriminative features with strong presentation.
3) Evaluations on both internal and external brain MRI datasets have demonstrated superiority of the proposed AGGN, which yields satisfactory robustness and generalization ability.

The remainder of this paper is organized as follows. Related work on glioma grading is presented in Section II. The proposed AGGN and the key components are elaborated in Section III. Substantial experimental validations and comprehensive discussions are presented in Section IV. Finally, conclusions with an outlook of future work are drawn in Section V.

II. Related Work

In this section, related glioma grading methods are reviewed. As glioma grading is essentially an image classification task, representative feature extraction and fusion methods are briefly introduced as well.

A. Glioma Grading Methods

In clinic, CAD methods play an important role in grading glioma with brain MRI data, and as early as 2010, the authors in [56] have used a support vector machine (SVM) to realize the preliminary assessment of glioma grade and achieved accuracy of 82%. In [15], a two-level clustering method has been proposed for MRI preprocessing, after which an SVM is adopted to accomplish the glioma grading task. A combination of SVM and multi-layer perceptron has been adopted to glioma grading in [40], where feature selection is performed on tumor sub-regions of different modalities. In addition to SVM, other classic machine learning models have also been applied in this area, and one can refer to [17] for more information.

Owing to the rapid development of DL techniques, deep neural networks (DNNs) based glioma grading models have already become the mainstream, where convolutional neural network (CNN) is one of the most popular architectures, including 2D- and 3D-CNN according to dimension of the convolution operations. In [31], a lightweight 2D-CNN model has been developed with only basic components like convolution and pooling layers, and the proposed method has realized fast inference with low computational complexity. In [45], the authors have proposed a 3D-CNN, where the volume of interests is segmented to promote an efficient feature extraction. In particular, performance between 2D Mask R-CNN and 3D U-Net in glioma grading task has been compared in [53], and it is found that 2D model achieves higher sensitivity but lower specificity than the 3D one. Transfer learning paradigm has been introduced in [44], with two well-known CNN-based models AlexNet and GoogleNet, experimental results indicate that the pre-trained model can enhance the performance. Similarly, it is deemed in [52] that the pre-trained CNN model can extract high-dimensional information of feature maps, which benefits further grading of glioma with stronger presentation than the low-dimensional texture or shape features.

In addition, a three-stage DNN model has been developed in [36], which successively performs the rough contour segmentation, the precise contour extraction and the classification. In [1], the generative adversarial network has been utilized to solve the problems of limited samples in brain MRI. Meanwhile, an adaptive encoder has been employed to extract multi-modal features in [1], which finally achieves precision of 92% on the glioma grading task.

Although above methods have proven effective, following two important issues still deserve further improvement. Firstly, most grading models rely on tumor mask-based data, and it is difficult to achieve end-to-end training without manual annotation; secondly, it is of vital significance to efficiently capture and integrate multi-modal pathological glioma features from MRI data, which has not been well addressed in existing methods.

To overcome the mentioned problems, in the proposed AGGN, three modules are meticulously designed to realize dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion, so that highly discriminative features with strong presentation can be extracted. Details of the proposed AGGN are presented in Section III.

B. Feature Extraction and Fusion Methods

In the context of DL-based medical image processing, extracted robust features with strong presentation will have a great impact on the model performance. Representative CNN-based feature extractor include the visual geometry group (VGG) [35], GoogleNet [37], Inception v1-v4 [38], [39], residual network (ResNet) [12], re-parameterization VGG (RepVGG) [5], etc. In particular, VGG has reduced the amount of model parameters by stacking small convolution modules; GoogleNet and Inception v1 have utilized multi-branch architectures with multi-size convolutions to extract features; Inception v2-v4 models have further proposed the concept of BatchNorm, asymmetric decomposition convolution kernel and residual inception to enhance the performance; ResNet has solved the gradient explosion problem by skip connections in deep structures; in RepVGG, the dominance of both detection speed and accuracy have been achieved by decoupling the training process and inference stage.

Feature fusion is another important operation in many DL-based methods, which promotes sufficient integration of information at different levels so as to enhance the presentation ability of features and improve the model performance. One of the most representative feature fusion structures is the feature pyramid network (FPN) [21], which contains two pathways for bottom-up forward propagation and top-down sampling recovery, respectively, and the lateral connections in FPN have facilitated information fusion. It is noticeable that many FPN variants have been successfully proposed, such as path aggregation network (PANet) [27], bidirectional FPN (BiFPN) [41] and atrous spatial pyramid pooling-balanced FPN (ABFPN) [51], etc. Particularly, the ABFPN is an enhanced multi-scale feature fusion structure, which improves the model performance via sufficiently utilizing context information and generating balanced enhanced features with rich receptive fields.

III. Methodology

In this section, the proposed AGGN is elaborated with implementation details, including the designed dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion modules. To begin with, the overall framework of AGGN is illustrated in Fig. 2.

A. Overall Framework of AGGN

According to Fig. 2, in the proposed AGGN, firstly pre-processing operations including standardization, center cropping, modal splicing, data partitioning and augmentation are performed on the input brain MRI images with four modalities T1, T2, T1ce and FLAIR. Afterwards, the pre-processed multi-modal MRI data will enter the dual-domain attention mechanism module, where the weights are assigned in both channel and spatial dimensions to highlight the key information and suppress the unimportant one in feature maps. Next, the multi-modal MRI is split, and in the followed multi-scale feature extraction module, parallel processing is performed on each single-modality map, including sequential operations of multi-branch convolution (MB conv), convolution-pooling (C-P) and multi-branch pooling (MB pool), and the final output of each pathway is in the size of $1 \times 1 \times 256$. It is noticeable that during above procedure, maps with three sizes on each modality are individually concatenated and fed into the multi-modal information fusion module, which contains the fusion convolution and MB reduction operations. At last, seven feature maps in the same size are cascaded and fed into the linear layers, which is responsible to accomplish the glioma grading task.

In the following subsections, above mentioned three major modules of AGGN are presented with details.

B. Dual-domain Attention Mechanism

Attention mechanism is essentially a procedure of weighting features by pixel-wise operations in channel or spatial dimension, where the position that can reflect the detailed or semantic information of targets will be assigned large weights. In the proposed AGGN, a novel dual-domain attention mechanism is designed, and the structure is presented in Fig. 3. It is worth mentioning that the pre-processed input data are directly sent into the designed dual-domain attention module to model the target location and individual modalities, where different weights are assigned based on both channel and spatial importance of the features, so as to realize focused attention on the useful information and simultaneously suppress the useless one.

As is shown in Fig. 3, the input map will successively pass through the channel attention (CA) and spatial attention (SA) components. To be specific, in the former one, the size of feature maps in four modalities is compressed through operations of three parallel branches, where 1×1 convolution, asymmetric convolution block (ACB), average pooling and BN-PRelu (batch normalization and parametric Relu) operations are performed. It is noticeable that ACB replaces square convolution with asymmetric one equivalently [4], which can effectively avoid significant information loss and reduce the number of parameters. By concatenating the three branches, diverse information is shared and afterwards weights are assigned via activation operation and element-wise multiplication with the original input data.

In subsequent spatial attention component, average and maximum pooling layers are placed at first to compress the channel of feature maps, and the outputs are concatenated to enter series of ACB blocks to learn the parameters in spatial dimension. Similarly, after the sigmoid activation function, the spatial-domain weight assignment for different pixel regions is eventually achieved by element-wise multiplication. As a result, the applied dual-domain attention mechanism can figure

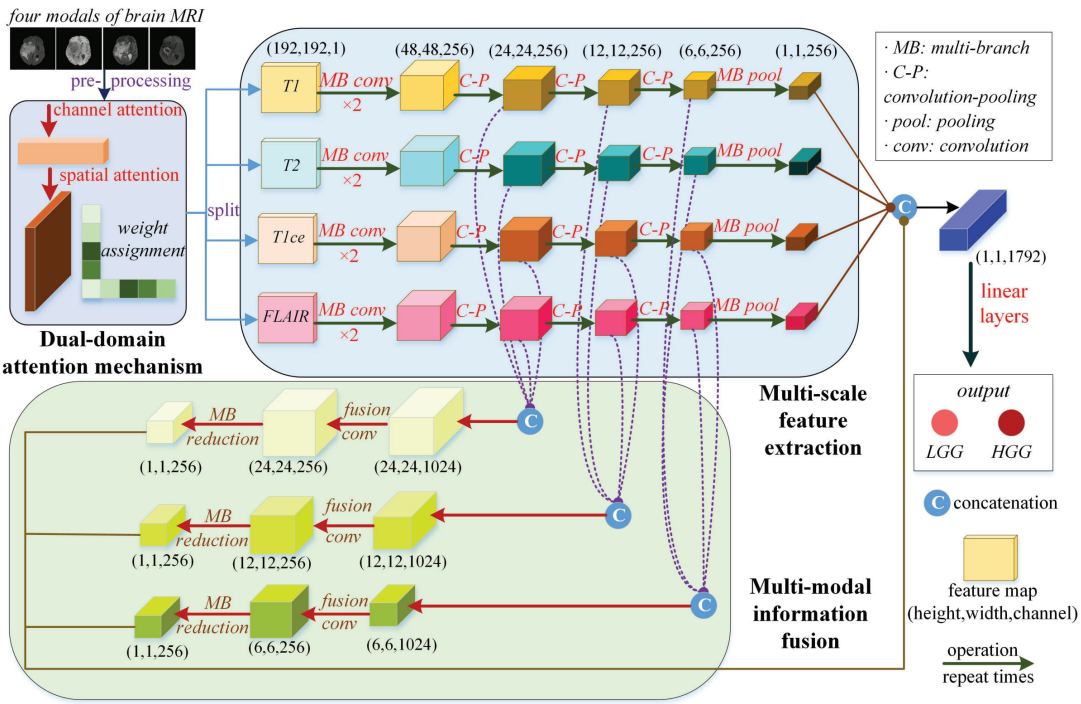


Fig. 2. Framework of the proposed attention-based glioma grading network (AGGN).

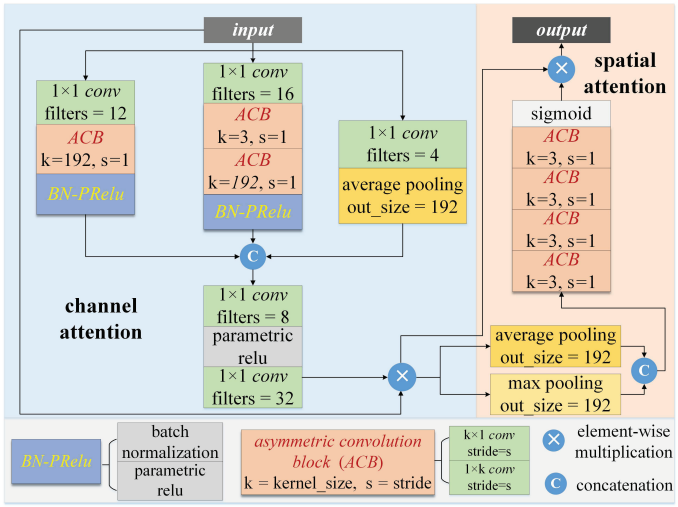


Fig. 3. Dual-domain attention mechanism module.

out both "where" and "what" the model should pay attention to. Work principle of above module is described by following equation.

$$output_{am} = SA\left\{M\left[CA\left(F_i^c\right)\right]_j^{x,y}\right\}, \quad (i = 1, ..., n;\ j = 1, 2, 3, 4) \tag{1}$$

where $F_\cdot^c$ and $M_\cdot^{(x,y)}$ refer to the feature map of channel c and modality of brain MRI in position (x, y), respectively; n is the number of feature maps, j denotes the modal and $output_{am}$ is the final output.

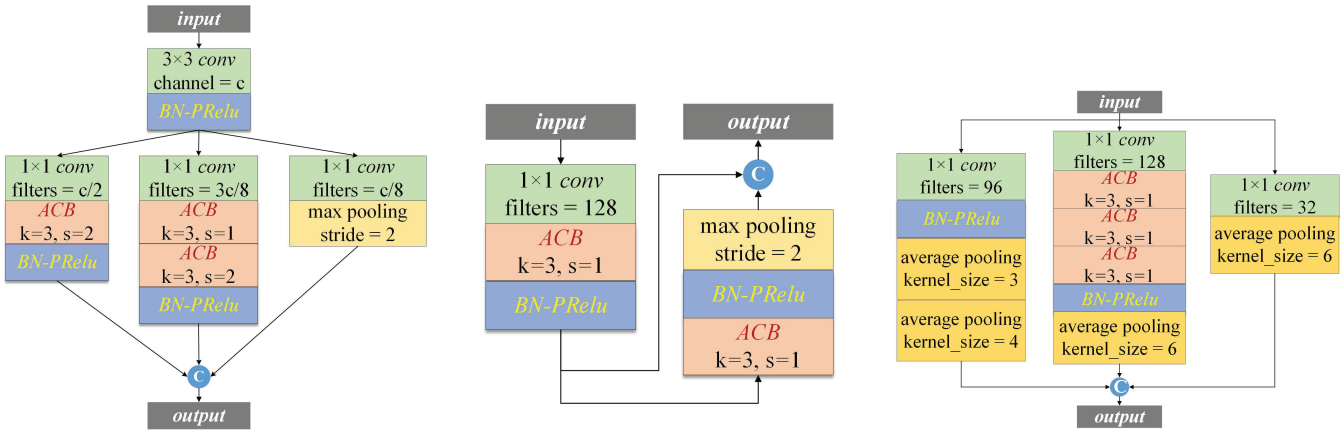
C. Multi-scale Feature Extraction

As previously mentioned, the four MRI modalities have contained rich pathological information with different concerns. To realize sufficient feature extraction on each modality, output of the dual-domain attention module is further split into four single-modal maps to enter the multi-scale feature extraction module (see Fig. 2), which can promote in-depth analysis of the key features enhanced by attention mechanism and can also benefit the subsequent multi-modal information fusion as well. On each branch, the involved MB conv, C-P and MB pool are displayed in Figs. 4(a)-4(c), respectively.

The MB conv block is used to extract the shallow features of each modal. As can be seen from Fig. 4(a), three parallel branches with different operations are included so that the extracted feature maps can contain rich information, and the last concatenation further integrates different features. Through MB conv block, the number of channels increases but the size of feature maps declines; and moreover, the applied ACB block can avoid large amount of information loss during the down-sampling procedure. In following Eq. 2, how MB conv block works is described.

$$\begin{aligned} output_{mc} = BP\left(AC_3^1\left(C_1\left(F_m\right)\right)\right) \oplus BP\left(AC_3^2\left(C_1\left(F_m\right)\right)\right) \\ \oplus MP\left(C_1\left(F_m\right)\right), \quad where\ F_m = BP\left(C_3\left(F_i\right)\right) \end{aligned} \tag{2}$$

where $output_{mc}$ is the block output, F_i and F_m refer to input and intermediate feature maps, respectively; $C_k\ (k = 1, 3)$ represents $k \times k$ standard convolution, and AC_3^m indicates the ACB operation with kernel size of three and m repetition times; BP and MP stand for BN-PRelu and maximum pooling operations, respectively.



(a) MB conv (b) C-P (c) MB pool

Fig. 4. Architectures of blocks in multi-scale feature extraction module.

Following the MB conv block, a series of C-P blocks are placed to continuously mine deep semantic features of each modality, where the ACB and maximum pooling operations with skip connection are adopted, which can be expressed as follows:

$$\begin{aligned} output_{cp} &= F_m \oplus MP\left(BP\left(AC_3^1\left(F_m\right)\right)\right), \\ where\ F_m &= BP\left(AC_3^1\left(C_1\left(F_i\right)\right)\right) \end{aligned} \tag{3}$$

It is noticeable that each branch has been equipped with three C-P blocks, and output of each C-P block will serve as input of the multi-modal information fusion module (see Fig. 2).

At the end of the multi-scale feature extraction module, MB pool blocks are deployed to generate the final output in size of $1 \times 1 \times 256$ for all modalities. It should be pointed out that the designing of MB pool is derived from improvement on the multi-receptive field pooling block in [42]. To be specific, multi-branch down-sampling is used in MB pool to convert the small-size map into a feature vector, and original convolution used in [42] is replaced by the asymmetric one with small kernels, which can expand the depth and enhance the feature extraction. In following Eq. (4), work principle of the MB pool is depicted.

$$\begin{aligned} output_{mp} = AP^2(BP(C_1(F_i))) \oplus AP^1(BP(AC^3(C_1(F_i)))) \\ \oplus AP^1(C_1(F_i)) \end{aligned} \tag{4}$$

where $output_{mp}$ is the block output, and $AP^i\ (i = 1, 2)$ denotes that the average pooling operation is repeated for i times.

D. Multi-modal Information Fusion Module

In the proposed AGGN, the multi-modal information fusion module is deployed to further integrate enhanced detailed and semantic features, and the structure is already illustrated in the green box of Fig. 2. In brief, fusion convolution realizes the integration of complementary advantages among the features of four modalities, where multi-scale feature maps in sizes of 24×24, 12×12 and 6×6 are fused. MB reduction block is adopted to transform the feature maps to vectors with strong presentation, which makes it feasible to further cascade the outputs of both multi-scale feature extraction and multi-modal information fusion module. MB reduction block consists of the sequential connection of the MB convolution (MC) and MB pool (MP) blocks, therefore, the block output $output_{mr}$ can be obtained by:

$$output_{mr} = MP(MC(F_i)) \tag{5}$$

where F_i denotes the input feature maps.

In addition, it is worth pointing out that the structure of fusion convolution block is similar to that of the C-P block, while the major difference is that the 2D convolution is replaced by the 3D one for fusion of feature maps with different modalities. According to Fig. 2, a finally constructed vector in size of $1 \times 1 \times 1792$ is fed into the last linear layers to obtain the glioma grading results, which is deemed to have strong presentation ability.

IV. Results and Discussions

In this section, the proposed AGGN is comprehensively evaluated on both internal and external public brain MRI dataset. In addition, substantial comparison experiments and ablation studies have been carried out to further validate the effectiveness and superiority of the proposed model. At first, experimental environment is briefly introduced.

A. Dataset Preprocessing and Experimental Settings

The experimental data used in this paper come from the 2018 and 2019 brain tumor segmentation (BraTS) challenges organized by medical image computing and computer assisted intervention society (MICCAI) [11], which are collected by 3T MRI systems of 17 institutions. The dataset includes multi-modal MRI from 326 glioma patients (250 for HGG, 76 for LGG), in which each case contains 155 slice data of four modalities, and the original size of each image is 240×240. In addition, professional radiologists have annotated and calibrated the edema, necrosis and core areas of glioma to obtain tumor masks, and grading results are determined through further pathological analysis.

The dataset is divided into training set, testing set and validation set, where the training set and testing set are independent of each other, while the validation set is obtained by further splitting the training set. To be specific, ratio of the training and testing set is $2:1$, where the images of training set come from 2018 BraTS. The testing set includes the internal and external subsets, which contain images from 2018 and 2019 BraTS, respectively. It should be pointed out that the main difference between internal and external testing subset is that samples of the latter belong to different data-source as those of training samples, and neither of the two subsets participates in the model training. Furthermore, one-fifth of the training samples are picked out to form the validation set for model tuning and selection.

In addition, preprocessing is performed on the initial data before training the model, where tumor masks are used to screen tumor-free slices at first, and it is noticeable that the selected slices without tumor are not fed into the subsequent process. For slices that contain tumor tissues, the foreground region is standardized and the proportion of background is reduced so that they are center-cropped to 192×192 in size; afterwards, four modalities are treated as four channels of the image. Finally, the dataset is divided according to the previously mentioned rules, and data augmentation operations are only performed on the training samples, including random rotation, translation and clipping. For a clear view, above preprocessing steps and dataset division are shown in Fig. 5.

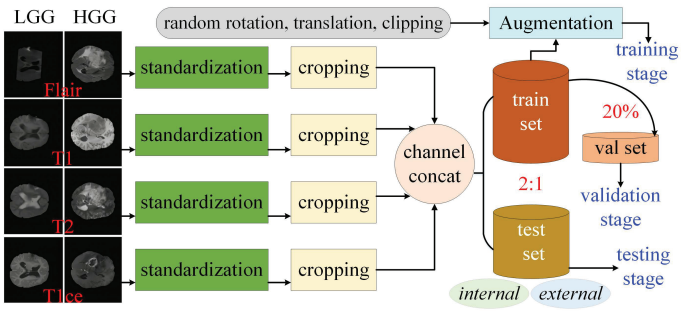


Fig. 5. Flowchart of preprocessing for brain MRI datasets.

All experiments in this study are carried out on the deep learning framework Pytorch, and the operating system is Windows 10 with NVIDIA GTX 2080Ti single GPU. Hyper-parameter settings are provided in Table I, as for model parameters, initialization of convolution and fully-connected layers adopts the Kaiming method and normal distribution, respectively.

B. Performance Evaluation

To comprehensively evaluate performance of the proposed AGGN, four groups of experiments are carried out, which aim at verifying the generalization ability, architectural advantages, superiority against other representative CNN-based models and competitiveness in comparison to state-of-the-art glioma grading methods, respectively. Metrics $accuracy$, $precision$, $recall$, $specificity$, $F1\ score$ are adopted for performance evaluation, which can be calculated by following Eqs. (6)-(10):

TABLE I
HYPERPARAMETER SETTINGS

Variables	Values
Training epochs	100
Batch size	32
Optimizer	Adam
Initial learning rate	0.0001
First-order moment decay coefficient	0.9
Second-order moment decay coefficient	0.999

$$accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{6}$$

$$precision = \frac{TP}{TP + FP} \tag{7}$$

$$recall = \frac{TP}{TP + FN} \tag{8}$$

$$specificity = \frac{TN}{FP + TN} \tag{9}$$

$$F1\ score = 2 \times \frac{recall \times precision}{recall + precision} \tag{10}$$

where TP/TN and FP/FN refer to the number of correct and wrong predictions of the HGG and LGG samples, respectively. As can be seen, the $accuracy$ describes the ratio of correct classifications of both HGG and LGG; the $precision$ aims at all samples predicted as HGG, and calculates the proportion of correct prediction; the $recall$ refers to the ratio of correctly identified HGG samples, which measures whether a model can screen all positive samples; similar to $recall$, the $specificity$ reflects the ability of identifying negative samples of a model; the $F1\ score$ takes the harmonic average between $accuracy$ and $recall$, and for all above five metrics, the larger their values are, the better the model performance is.

In addition, the receiver operating characteristic (ROC) curve and area under this curve (AUC) are also employed for the model evaluation. Specifically, the ROC curve takes the value of $1 - specificity$ (also known as false positive rate) as the horizontal axis and $recall$ as the vertical one, AUC is the area enclosed by ROC curve and the two coordinate axes.

1) Generalization ability of AGGN: At first, results obtained by the proposed AGGN on both internal and external testing sets are shown in Fig. 6, notice that in the former, training and testing samples share the same data-source; on the contrary, different sources are contained in the latter. As a result, this group of experiment can objectively reflect the generalization ability of the proposed AGGN. As is shown, the worst result is the $F1\ score$ on external dataset, which reaches 0.933; advantages of $precision$, $specificity$ and AUC are noticeable on both datasets, which validates that the propose AGGN is highly reliable in glioma grading task. At the same time, AGGN presents similar performance on internal and external testing sets, which demonstrates the robustness of AGGN in terms of handling various glioma MRI data; and

moreover, this result indicates that AGGN can adapt to data from multi-center medical institution with strong generalization ability.

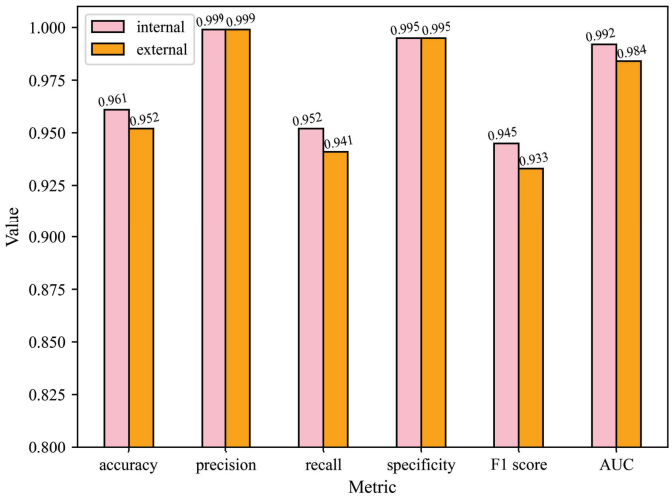


Fig. 6. Evaluation of AGGN on internal and external testing sets.

2) Architectural advantages of AGGN: In this part, to validate the architectural advantages of our method, adaptive multi-modal fusion network (AMMFNet) [42] is adopted as baseline model for comparison, which is a similar glioma grading framework to the proposed AGGN. In Fig. 7, performance enhancement of AGGN on six indicators is illustrated, which shows that in comparison to AMMFNet, the proposed AGGN has improved all indicators to different extents. In particular, the most significant improvement is on $specificity$, which increases 11.52%. As previously mentioned, high $specificity$ is equivalent to low false positive rate. Consequently, it is verified that AGGN has strong ability to correctly identify negative samples, which can effectively avoid the waste of medical resources. In addition, $precision$ is increased by 5.27%, which implies that AGGN is able to achieve accurate diagnosis of HGG. In this group of experiment, the proposed AGGN presents noticeable competitiveness in comparison to similar framework, which indicates the advantages in structural configuration.

TABLE II
PERFORMANCE COMPARISON OF PROPOSED AGGN AND FOUR CLASSIC MODELS ON INTERNAL TESTING SET

Metrics	Models				
	[12]	[14]	[35]	[39]	AGGN
$accuracy$	0.9013	0.9038	0.8785	0.9330	**0.9612**
$precision$	0.7763	0.9632	0.8848	0.9687	**0.9987**
$recall$	0.9234	0.9119	0.9476	0.9181	**0.9521**
$specificity$	0.8320	0.8747	0.7235	0.9749	**0.9952**
$F1\ score$	0.8687	0.8676	0.8506	0.9046	**0.9450**
AUC	0.9530	0.9570	0.9480	0.9780	**0.9920**

3) Comparisons with other CNN-based models: In order to further validate the competitiveness of the proposed AGGN,

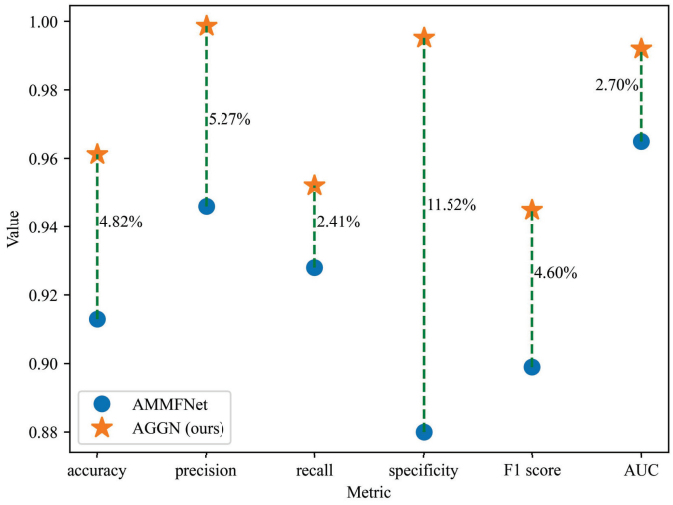


Fig. 7. Performance comparison between AGGN and AMMFNet.

TABLE III
PERFORMANCE COMPARISON OF PROPOSED AGGN AND FOUR CLASSIC MODELS ON EXTERNAL TESTING SET

Metrics	Models				
	[12]	[14]	[35]	[39]	AGGN
$accuracy$	0.8967	0.9052	0.8780	0.9446	**0.9519**
$precision$	0.9304	0.9672	0.8688	**0.9987**	**0.9987**
$recall$	0.9323	0.9122	**0.9658**	0.9325	0.9401
$specificity$	0.7890	0.8775	0.6937	0.9948	**0.9953**
$F1\ score$	0.8614	0.8639	0.8503	0.9193	**0.9334**
AUC	0.9410	0.9570	0.9580	0.9700	**0.9840**

four other representative CNN-based models are adopted for comparison in this group of experiments, including ResNet-50 [12], DenseNet-101 [14], VGG-19 [35] and Inception-v4 [39]. For fairness, all models share the same training and testing data, and the results on internal and external datasets are reported in Table. II and Table. III, respectively. In addition, an illustration is presented in Fig. 8.

As can be seen from Table II, all the indicators of AGGN are better than those of other representative CNN models on internal dataset, which are 2.82%, 3.0%, 0.45%, 2.03%, 4.04% and 1.4% higher than the sub-optimal model on $accuracy$, $precision$, $recall$, $specificity$, $F1\ score$ and AUC respectively. While on the external testing set, the proposed AGGN also achieves satisfactory results of 95.19%, 99.87%, 94.01%, 99.53%, 93.34% and 98.40% on above six metrics, respectively. On five out of the six indicators, AGGN has obtained the best results.

In addition, the ROC curves with magnification on the two testing sets are presented in Fig. 9, which can effectively evaluate the diagnostic ability of a model and can maintain strong stability when the distribution of positive and negative samples changes. Notice that the curve close to the upper left corner has high prediction accuracy, and accordingly, the larger AUC value, the better model performance is. As shown in Fig. 9, the ROC curve of AGGN is above all other models

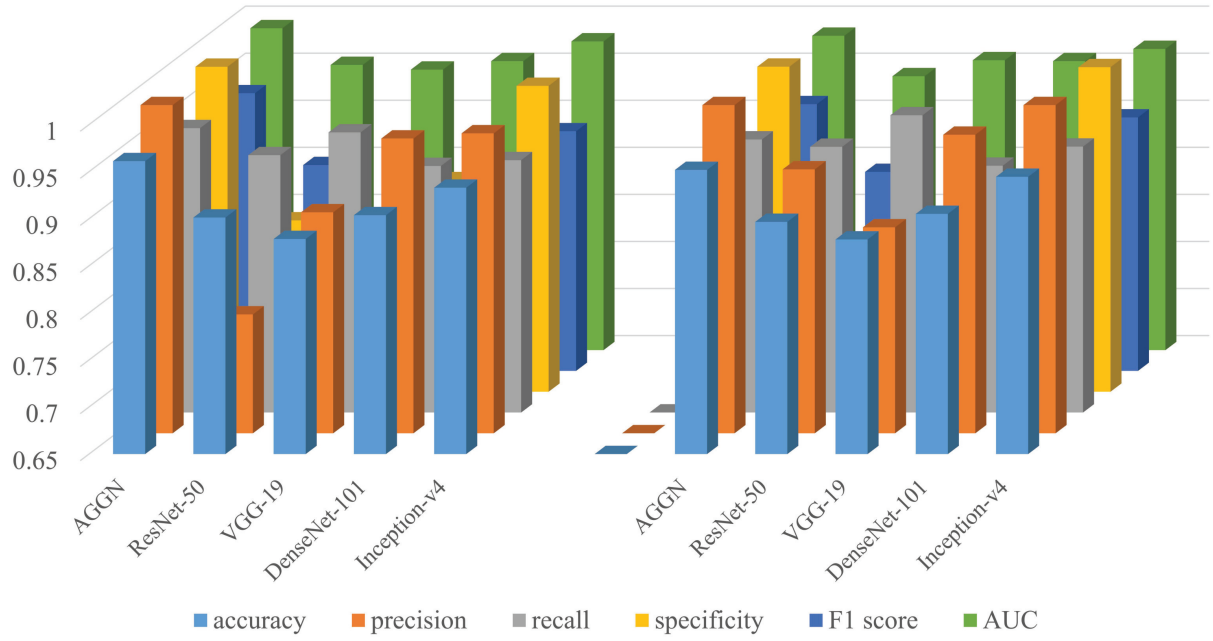
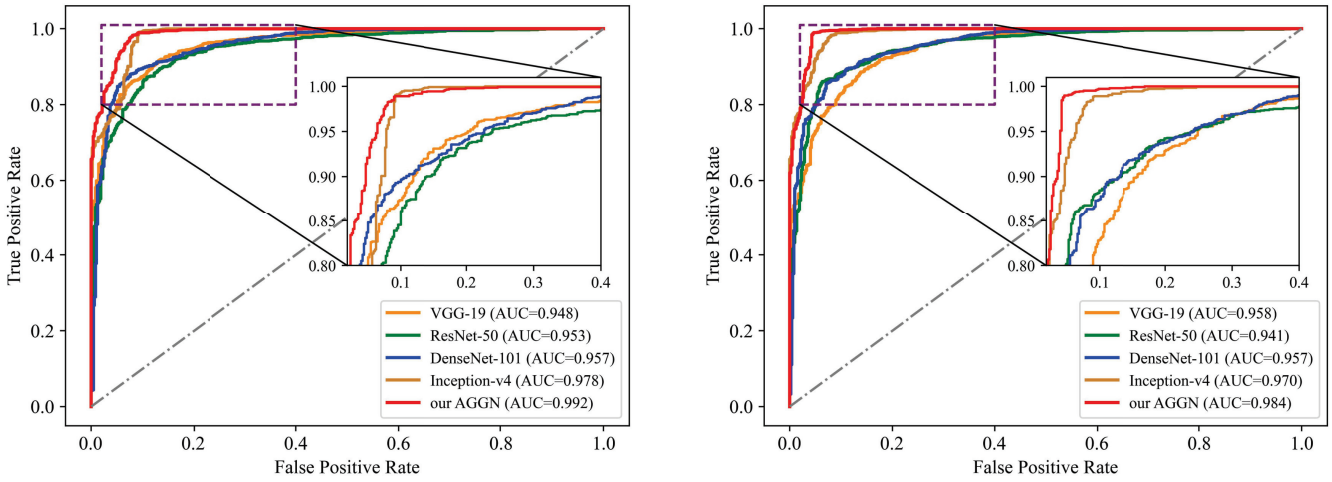


Fig. 8. Performance comparison of AGGN with advanced CNN-based models on internal (left) and external (right) testing sets.



(a) On internal testing set

(b) On external testing set

Fig. 9. ROC curves of the proposed AGGN and other CNN models.

on both the internal and external testing sets, and the AUC values reach 0.992 and 0.984, respectively. Consequently, the strong generalization ability of AGGN while facing different data sources is further verified.

Through this group of experiment, it is demonstrated that the proposed AGGN has overwhelming overall performance against other advanced CNN-based and domain-specific models on most metrics, which may owe to the meticulously designed and introduced dual-domain attention mechanism, multi-scale feature extraction and multi-modal information fusion modules.

4) Comparisons with state-of-the-art glioma grading algorithms: In this part, comparison between proposed AGGN and other state-of-the-art glioma grading algorithms are presented, including multistream CNN [9], multi-scale CNN [10], CAE-GAN (convolutional autoencoder and generative adversarial network) [1], pre-trained GoogleNet [44], 3DConvNet [53] and AMMFNet [42]. It should be pointed out that in previous experiments, data from the mentioned internal and external testing sets have none of the tumor masks, while most of recently related methods for the same task require the assistance of additional tumor masks. Consequently, to make a fair comparison, in this group of experiments, tumor masks have been added to original images for training, and the results are reported in Table IV. Notice that the data of other algorithms are cited from corresponding original papers, and "–" denotes

none of relevant data is provided.

TABLE IV
PERFORMANCE OF AGGN AND OTHER ADVANCED MODELS ON TUMOR MASK ASSISTED DATA

Models	Metrics			
	$accuracy$	$recall$	$specificity$	AUC
Multistream CNN	0.9087	–	–	–
Multi-scale CNN	0.8947	–	–	–
CAE-GAN	0.9204	–	–	–
Pre-trained GoogleNet	0.9450	–	–	0.9680
3DConvNet	0.9710	0.9470	0.9680	–
AMMFNet	0.9820	**1.0000**	0.9330	0.9970
AGGN (ours)	**0.9899**	**1.0000**	**0.9678**	**0.9980**

As can be found in Table IV, with assistance of tumor masks, the proposed AGGN can present the state-of-the-art performance. In particular, the AUC value of AGGN with and without tumor masks are 0.998 and 0.992, respectively, which implies that the assistance of tumor masks does further improve the model performance. It is also worth mentioning that the AUC of 0.992 (without masks) is already an excellent result. Hence, it can be concluded that performance of AGGN has little reliance on the masks, which demonstrates that AGGN can overcome the high dependence of manually labeled annotations so as to achieve the end-to-end applications in practice.

C. Ablation Study

To validate the effectiveness of core components in the proposed AGGN, substantial ablation studies are performed on the internal testing set in this subsection. The designed dual-domain attention mechanism is firstly verified and the results are reported in Table V, where AGG1, AGG2 and AGG3 refer to the model with none of attention modules, only spatial and only channel attention module, respectively. Obviously, in comparison to AGG1, on most indicators the performance has been improved to a certain extent after introducing the spatial or channel attention mechanism. It is also found that the applied dual-domain attention module in the proposed AGGN has realized significant performance enhancement, which improves $accuracy$, $recall$, $F1\ score$ and AUC by 2.92%, 3.61%, 4.23% and 1.6%, respectively.

In addition, the ROC curves of four models listed in Table V are presented in Fig. 10, where it can be seen that the proposed AGGN has obtained the best results. In particular, when false positive rate is 0, the true positive rate of AGGN is close to 0.85 and AUC is 0.992, which implies that the proposed AGGN can accurately identify HGG with almost none of false detection. Therefore, the proposed AGGN is a reliable model that can provide a solid guarantee for the diagnosis and treatment of critical patients.

According to above results, effectiveness of the dual-domain attention mechanism is sufficiently validated. Before extracting multi-scale features, channel-domain attention is firstly introduced to low-level detail information, which determines what deserves attention in each modality of brain MRI; afterwards,

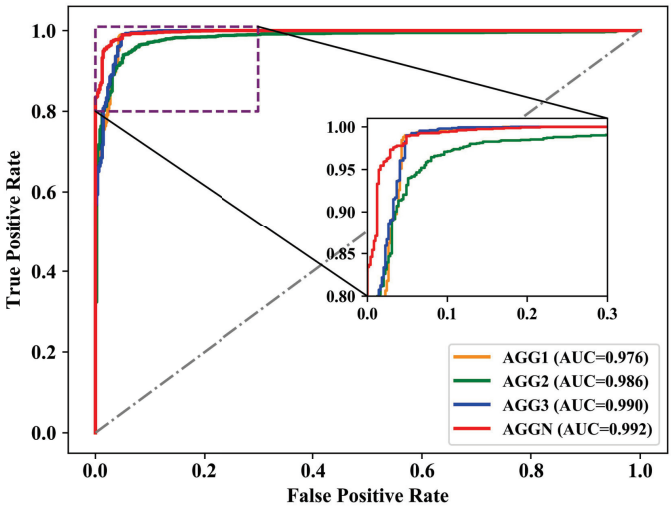


Fig. 10. ROC curves of AGG1, AGG2, AGG3, and AGGN.

spatial-domain attention is used to learn spatial dependence among high-level semantic information, so as to figure out the important locations in feature maps. As a result, the proposed AGGN can both recognize and localize the significant pathological features in brain MRI with strong robustness.

In the following, ablation study results on multi-branch convolution block and the multi-modal information fusion module are reported in Table VI. It should be pointed out that except for the investigated components, other configurations of AGGN remain unchanged so as to make objective and convincing comparisons.

Firstly, as the most important component of the multi-scale feature extraction module, MB conv block is compared with multi-receptive field (MRF) conv block of AMMFNet. As can be seen from Table VI, the MB conv block has overwhelmed the MRF conv block on all metrics, which demonstrates that the multi-branch structure has superiority in dealing with glioma grading task based on brain MRI. Further explorations of the essential mechanism show that the volume of glioma from patients can be quite different, whereas the proposed MB conv block has adopted ACBs with different sizes to extract image features in parallel, which can obtain fine-grained texture and tissue information of multi-modal brain MRI. Therefore, it can be inferred that the multi-scale feature extraction module has made great contribution to the overall model performance.

Secondly, the designed multi-modal information fusion module is compared with the approach in [42]. As reported in Table VI, on four out of six indicators, proposed AGGN has achieved slight performance improvement. Specifically, AGGN improves $accuracy$, $recall$, $F1\ score$ and AUC by 3.4%, 0.6%, 0.3% and 0.7%, respectively. It is worth mentioning that in the proposed AGGN, final input vector to the classifier is a concatenation of outputs from seven branches, and this number is fewer than that in [42]. Consequently, it can be concluded that AGGN has achieved comparable results to the model in [42] with a simplified structure. Additionally, the concatenation manner in AGGN has avoided stacking

TABLE V
ABLATION STUDIES OF DUAL-DOMAIN ATTENTION MECHANISM ON INTERNAL TESTING SET

Models	Metrics					
	$accuracy$	$precision$	$recall$	$specificity$	$F1\ score$	AUC
AGG1	0.9320	1.0000	0.9160	1.0000	0.9027	0.9760
AGG2	0.9481	0.9866	0.9467	0.9532	0.9270	0.9860
AGG3	0.9476	**1.0000**	0.9349	**1.0000**	0.9238	0.9900
AGGN	**0.9612**	0.9987	**0.9521**	0.9952	**0.9450**	**0.9920**

TABLE VI
ABLATION STUDIES OF MB CONV BLOCK AND INFORMATION FUSION METHODS

Models	Metrics					
	$accuracy$	$precision$	$recall$	$specificity$	$F1\ score$	AUC
MRF conv block of AMMFNet	0.9577	0.9960	0.9502	0.9856	0.9400	0.9840
MB conv block of AGGN	**0.9612**	**0.9987**	**0.9521**	**0.9952**	**0.9450**	**0.9920**
Fusion method in [42]	0.9578	**1.0000**	0.9461	**1.0000**	0.9420	0.9850
Fusion method of AGGN	**0.9612**	0.9987	**0.9521**	0.9952	**0.9450**	**0.9920**

redundant features, which not only benefits highly-efficient feature fusion, but also simultaneously avoids excessively complicating the model.

D. Computational Complexity Analysis

In this study, the number of model parameters (Params) and floating point operations (FLOPs) are adopted to depict the spatial and time complexity of the proposed AGGN, respectively. Excessive parameters will impede the light-weight deployment of model on edge devices, and too large FLOPs will influence the convergence during model training, which directly determines the accuracy of the model inference.

On the one hand, Params of the proposed AGGN is 16.37M, which is 9.13M fewer than that of the classical ResNet-50 model. According to Table III, the accuracy of AGGN is even 5.52% higher than that of ResNet-50, which demonstrates that the developed AGGN can effectively balance the computational costs and accuracy. It may owe to the proposed AGGN has effectively reduced the parameters by replacing large-size kernels with a series of small-sized ones. Meanwhile, the accuracy of AGGN is mainly guaranteed by the structural advantages, including employing dual-domain attention mechanism to highlight key features, realizing feature extraction on each individual modality, and integrating multi-modal information in different levels.

On the other hand, the FLOPs of AGGN are 24,790M, which mainly due to the large size of multi-channel input samples, where the data processing has consumed great deals of the FLOPs. It is also worth mentioning that during the model training, none of obvious over-fitting phenomenon has occurred, which implies that the training and inference time consumed by the proposed AGGN is acceptable.

To sum up, the proposed AGGN can effectively achieve the balance between model complexity and accuracy, which has achieved satisfactory results in the glioma grading task with considerable efficiency.

V. CONCLUSION

In this paper, a novel self-attention based network AGGN has been developed, which mainly consists of three meticulously designed modules, including a dual-domain attention module, a multi-scale feature extraction and a multi-modal information fusion one. Robust features with strong presentation ability are constructed by integrating outputs of the latter two modules, which are used to eventually realize the glioma grading task. Performance of the proposed AGGN has been comprehensively evaluated on both internal and external testing sets, and the results have demonstrated the superiority of AGGN against other state-of-the-art algorithms. Furthermore, substantial ablation studies have verified effectiveness of the designed three modules in AGGN, which can take full advantages of detailed and semantic information so that model performance can be greatly improved, and simultaneously computational burdens are released to some extent.

Although the proposed AGGN has presented satisfactory performance on the glioma grading task, it still has some spaces for further improvement, including task migration adaption, quantitative lesion analysis, and model lightweighting studies. In future work, we aim to 1) apply the developed AGGN framework to other MRI-based tasks such as stroke and cancer diagnosis; 2) investigate fine-grained glioma grading methods to support quantitative analysis; 3) further optimize the structure of AGGN through fuzzy system and tensor decomposition techniques. [18], [23], [26], [43]

REFERENCES

[1] M. Ali, I. Gu and A. Jakola, "Multi-stream convolutional autoencoder and 2D generative adversarial network for glioma classification", *Procedding of the 18th International Conference on Computer Analysis of Images and Patterns (CAIP)*, pp. 234-245, 2019.

[2] G. Bao, L. Ma and X. Yi, "Recent advances on cooperative control of heterogeneous multi-agent systems subject to constraints: A survey", *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 539-551, 2022.

[3] J. Cheng, J. Liu, H. Kuang and J. Wang, "A fully automated multimodal MRI-based multi-task learning for glioma segmentation and IDH genotyping", *IEEE Transactions on Medical Imaging*, vol. 41, no. 6, pp. 1520-1532, 2022.

[4] X. Ding, Y. Guo, G. Ding and J. Han, "ACNet: strengthening the kernel skeletons for powerful cnn via asymmetric convolution blocks", *Procedding of the 17th IEEE International Conference on Computer Vision (ICCV)*, pp. 1911-1920, 2019.

[5] X. Ding, X. Zhang, N. Ma, J. Han, G. Ding and J. Sun, "RepVGG: making VGG-style ConvNets great again", *Procedding of the 34th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 13728-13737, 2021.

[6] E. Donaldson and A. Kirk, "Is MRI the superior test for investigating thymic pathologies in comparison to CT?", *Lung Cancer*, vol. 156, no. s1, pp. S23-S23, 2021.

[7] E. Ferrari, P. Bosco, S. Calderoni, P. Oliva, L. Palumbo, G. Spera, M. Fantacci and A. Retico, "Dealing with confounders and outliers in classification medical studies: the autism spectrum disorders case study", *Artificial Intelligence in Medicine*, vol. 108, article no. 101926, 2020.

[8] Y. Fei, B. Zhan, M. Hong, X. Wu, J. Zhou and Y. Wang, "Deep learning-based multi-modal computing with feature disentanglement for MRI image synthesis", *Medical Physics*, vol. 48, no. 7, pp. 3778-3789, 2021.

[9] C. Ge, I. Gu, A. Jakola and J. Yang, "Deep learning and multi-sensor fusion for glioma classification using multistream 2D convolutional networks", *Procedding of the 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pp. 5894-5897, 2018.

[10] C. Ge, Q. Qu, I. Gu and A. Store Jakola, "3D multi-scale convolutional networks for glioma grading using MR images", *Procedding of the 25th IEEE International Conference on Image Processing (ICIP)*, pp. 141-145, 2018.

[11] M. Ghaffari, A. Sowmya and R. Oliver, "Automated brain tumor segmentation using multimodal brain scans: a survey based on models submitted to the brats 2012–2018 challenges", *IEEE Reviews in Biomedical Engineering*, vol. 13, pp. 156-168, article no. 2946868, 2020.

[12] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition", *Procedding of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770-778, 2016.

[13] J. Hu, C. Jia, H. Liu, X. Yi and Y. Liu, "A survey on state estimation of complex dynamical networks", *International Journal of Systems Science*, vol. 52, no. 16, pp. 3351-3367, 2021.

[14] G. Huang, Z. Liu, L. Van Der Maaten and K.Q. Weinberger, "Densely connected convolutional networks", *Procedding of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2261-2269, 2017.

[15] R. Inano, N. Oishi, T. Kunieda, Y. Arakawa, Y. Yamao, S. Shibata, T. Kikuchi, H. Fukuyama and S. Miyamoto, "Voxel-based clustered imaging by multiparameter diffusion tensor images for glioma grading", *Neuroimage-clinical*, vol. 5, pp. 396-407, 2014.

[16] S. Ismael, A. Mohammed and H. Hefny, "An enhanced deep learning approach for brain cancer MRI images classification using residual networks", *Artificial Intelligence in Medicine*, vol. 102, article no. 101779, 2020.

[17] J. Jeong, L. Wang, B. Ji, Y. Lei, A. Ali, T. Liu, W. Curran, H. Mao and X. Yang, "Machine-learning based classification of glioblastoma using delta-radiomic features derived from dynamic susceptibility contrast enhanced magnetic resonance images", *Quantitative Imaging in Medicine and Surgery*, vol. 9, no. 7, pp. 1201-1213, 2019.

[18] J. Mao, Y. Sun, X. Yi, H. Liu and D. Ding, "Recursive filtering of networked nonlinear systems: A survey", *International Journal of Systems Science*, vol. 52, no. 6, pp. 1110-1128, 2021.

[19] L. Marginean, P. Stefan, A. Lebovici, I. Opincariu, C. Csutak, R. Lupean, P. Coroian and B. Suciu, "CT in the differentiation of gliomas from brain metastases: The radiomics analysis of the peritumoral zone", *Brain Sciences*, vol. 12, no. 1, article no. 109, 2022.

[20] M. Mittler, B. Walters and E. Stopa, "Observer reliability in histological grading of astrocytoma stereotactic biopsies", *Journal of Neurosurgery*, vol. 85, no. 6, pp. 1091-1094, 1996.

[21] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature pyramid networks for object detection", *Procedding of the 30th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 936-944, 2017.

[22] G. Li, L. Li, Y. Li, Z. Qian, F. Wu, Y. He, H. Jiang, R. Li, D. Wang, Y. Zhai, Z. Wang, T. Jiang, J. Zhang and W. Zhang, "An MRI radiomics approach to predict survival and tumour-infiltrating macrophages in gliomas", *Brain*, vol. 145, no. 3, pp. 1151-1161, 2022.

[23] H. Li, P. Wu, Z. Wang, J. Mao, Fuad E. Alsaadi and N. Zeng, "A generalized framework of feature learning enhanced convolutional neural network for pathology-image-oriented cancer diagnosis", *Computers in Biology and Medicine*, vol. 151, article no. 106265, 2022.

[24] H. Li, P. Wu, N. Zeng, Y. Liu and Fuad E. Alsaadi, "A Survey on parameter identification, state estimation and data analytics for lateral flow immunoassay: from Systems Science Perspective", *International Journal of Systems Science*, 2022. https://doi.org/10.1080/00207721.2022.2083262.

[25] H. Li, N. Zeng, P. Wu and K. Clawson, "Cov-Net: A computer-aided diagnosis method for recognizing COVID-19 from chest X-ray images via machine vision", *Expert Systems with Applications*, vol. 207, article no. 118029, 2022.

[26] W. Li, Y. Niu and Z. Cao, "Event-triggered sliding mode control for multi-agent systems subject to channel fading", *International Journal of Systems Science*, vol. 53, no. 6, pp. 1233-1244, 2022.

[27] S. Liu, L. Qi, H. Qin, J. Shi and J. Jia, "Path aggregation network for instance segmentation", *Procedding of the 31th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8759-8768, 2018.

[28] D. Louis, A. Perry, P. Wesseling, D. Brat, I. Cree, D. Figarella-Branger, C. Hawkins, H. Ng, S. Pfister, G. Reifenberger, R. Soffietti, A. von Deimling and D. Ellison, "The 2021 WHO classification of tumors of the central nervous system: a summary", *Neuro-Oncology*, vol. 23, no. 8, pp. 1231-1251, 2021.

[29] P. Lu, B. Song and L. Xu, "Human face recognition based on convolutional neural network and augmented dataset", *Systems Science & Control Engineering*, vol. 9, no. s2, pp. 29-37, 2021.

[30] Q. Ostrom, N. Patil, G. Cioffi, K. Waite, C. Kruchko and J. Barnholtz-Sloan, "Corrigendum to: CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2013–2017", *Neuro-Oncology*, vol. 24, no. 7, pp. 1214-1310, 2022.

[31] H. Ozcan, B. Emiroglu, H. Sabuncuoglu, S. Ozdogan, A. Soyer and T. Saygi, "A comparative study for glioma classification using deep convolutional neural networks", *Mathematical Biosciences and Engineering*, vol. 18, no. 2, pp. 1550-1572, 2021.

[32] G. Pahuja and B. Prasad, "Deep learning architectures for Parkinson's disease detection by using multi-modal features", *Computers in Biology and Medicine*, vol. 146, article no. 105610, 2022.

[33] C. Sarasaen, S. Chatterjee, M. Breitkopf, G. Rose, A. Nurnberger and O. Speck, "Fine-tuning deep learning model parameters for improved super-resolution of dynamic MRI with prior-knowledge", *Artificial Intelligence in Medicine*, vol. 121, article no. 102196, 2021.

[34] B. Song, H. Miao and L. Xu, "Path planning for coal mine robot via improved ant colony optimization algorithm", *Systems Science & Control Engineering*, vol. 9, no. 1, pp. 283-289, 2021.

[35] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition", *Procedding of the 3rd International Conference on Learning Representations (ICLR)*, PP. 1-14, 2015.

[36] M. Soleymanifard and M. Hamghalam, "Multi-stage glioma segmentation for tumour grade classification based on multiscale fuzzy C-means", *Multimedia Tools and Applications*, vol. 81, no. 6, pp. 8451-8470, 2022.

[37] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke and A. Rabinovich, "Going deeper with convolutions", *Procedding of the 28th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1-9, 2015.

[38] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens and Z. Wojna, "Rethinking the inception architecture for computer vision", *Procedding of the 29th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2818-2826, 2016.

[39] C. Szegedy, S. Ioffe, V. Vanhoucke and A. Alemi, "Inception-v4, Inception-ResNet and the impact of residual connections on learning", *Procedding of the 31th AAAI Conference on Artificial Intelligence (AAAI)*, pp.4278-4284, 2017.

[40] P. Sun, D. Wang, V. Mok and L. Shi, "Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading", *IEEE Access*, vol. 7, pp. 102010-102020, article no. 2928975, 2019.

[41] M. Tan, R. Pang and Q. Le, "EfficientDet: scalable and efficient object detection", *Procedding of the 33th IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10778-10787, 2020.

[42] L. Wang, Y. Cao, L. Tian, Q. Chen, S. Guo, J. Zhang and LH. Wang, "Adaptive multi-modality fusion network for glioma grading", *Journal of Image and Graphics*, vol. 26, no. 9, pp. 2243-2256, 2021.

[43] Y. Wang, L. Zou, L. Ma, Z. Zhao and J. Guo, "A survey on control for Takagi-Sugeno fuzzy systems subject to engineering-oriented complexities", *Systems Science & Control Engineering*, vol. 9, no. 1, pp. 334-349, 2021.

[44] Y. Yang, L. Yan, X. Zhang, Y. Han, H. Nan, Y. Hu, B. Hu, S. Yan, J. Zhang, D. Cheng, X. Ge, G. Cui, D. Zhao and W. Wang, "Glioma grading on conventional MR images: a deep learning study with transfer learning", *Frontiers in Neuroscience*, vol. 12, article no. 804, 2018.

[45] H. Yamashiro, A. Teramoto, K. Saito and H. Fujita, "Development of a fully automated glioma-grading pipeline using post-contrast T1-weighted images combined with cloud-based 3D convolutional neural network", *Applied Sciences-Basel*, vol. 11, no. 11, article no. 5118, 2021.

[46] Y. Yuan, G. Ma, C. Cheng, B. Zhou, H. Zhao, H.-T. Zhang and H. Ding, "A general end-to-end diagnosis framework for manufacturing systems", *National Science Review*, vol. 7, no. 2, pp. 418-429, 2020.

[47] Y. Yuan, X. Tang, W. Zhou, W. Pan, X. Li, H.-T. Zhang, H. Ding and J. Goncalves, "Data driven discovery of cyber physical systems", *Nature Communications*, vol. 10, no. 1, pp. 1-9, 2019.

[48] Y. Yuan, H. Zhang, Y. Wu, T. Zhu and H. Ding, "Bayesian learning-based model-predictive vibration control for thin-walled workpiece machining processes", *IEEE/ASME transactions on mechatronics*, vol. 22, no. 1, pp. 509-520, 2016.

[49] N. Zeng, H. Li and Y. Peng, "A new deep belief network-based multi-task learning for diagnosis of Alzheimer's disease", *Neural Computing and Applications*, article no. 06149-6, 2021.

[50] N. Zeng, Z. Wang, W. Liu, H. Zhang, K. Hone and X. Liu, "A dynamic neighborhood-based switching particle swarm optimization algorithm", *IEEE Transactions on Cybernetics*, vol. 52, no. 9, pp. 9290-9301, 2022.

[51] N. Zeng, P. Wu, Z. Wang, H. Li, W. Liu and X. Liu, "A small-sized object detection oriented multi-scale feature fusion approach with application to defect detection", *IEEE Transactions on Instrumentation and Measurement*, vol. 71, article no. 3507014, 2022.

[52] Z. Zhang, J. Xiao, S. Wu, F. Lv, J. Gong, L. Jiang, R. Yu and T. Luo, "Deep convolutional radiomic features on diffusion tensor images for classification of glioma grades", *Journal of Digital Imaging*, vol. 33, no. 4, pp. 826-837, 2020.

[53] Y. Zhuge, H. Ning, P. Mathen, J. Cheng, A. Krauze, K. Camphausen and R. Miller, "Automated glioma grading on conventional MRI images using deep convolutional neural networks", *Medical Physics*, vol. 47, no. 7, pp. 3044-3053, 2020.

[54] B. Zinnhardt, F. Roncaroli, C. Foray, E. Agushi, B. Osrah, G. Hugon, A. Jacobs and A. Winkeler, "Imaging of the glioma microenvironment by TSPO PET", *European Journal of Nuclear Medicine and Molecular Imaging*, vol. 49, no. 1 pp. 174-185, 2021.

[55] L. Zou, Z. Wang, J. Hu, Y. Liu and X. Liu, "Communication-protocol-based analysis and synthesis of networked systems: Progress, prospects and challenges", *International Journal of Systems Science*, vol. 52, no. 14, pp. 3013-3034, 2021.

[56] F. Zollner, K. Emblem and L. Schad, "Support vector machines in DSC-based glioma imaging: suggestions for optimal characterization", *Magnetic Resonance in Medicine*, vol. 64, no. 4, pp. 1230-1236, 2010.