

Residual Channel Attention Network for Brain Glioma Segmentation

Yiming Yao^{*1}, Peisheng Qian^{*2}, Ziyuan Zhao², and Zeng Zeng^{†2}

Abstract—A glioma is a malignant brain tumor that seriously affects cognitive functions and lowers patients’ life quality. Segmentation of brain glioma is challenging because of inter-class ambiguities in tumor regions. Recently, deep learning approaches have achieved outstanding performance in the automatic segmentation of brain glioma. However, existing algorithms fail to exploit channel-wise feature interdependence to select semantic attributes for glioma segmentation. In this study, we implement a novel deep neural network that integrates residual channel attention modules to calibrate intermediate features for glioma segmentation. The proposed channel attention mechanism adaptively weights feature channel-wise to optimize the latent representation of gliomas. We evaluate our method on the established dataset BraTS2017. Experimental results indicate the superiority of our method.

Clinical relevance— While existing glioma segmentation approaches do not leverage channel-wise feature dependence for feature selection, our method can generate segmentation masks with higher accuracies and provide more insights on graphic patterns in brain MRI images for further clinical reference.

I. INTRODUCTION

Gliomas are among the most common and dreadful malignant brain tumors. It starts from non-neuronal glial cells and leads to symptoms including seizures, cranial nerve disorders, and ultimately, death [1]. Diagnosis of brain glioma at early stages is essential for subsequent treatment and recovery, which can be performed in multi-modality brain Magnetic Resonance Imaging (MRI). The 4 modalities in brain MRI includes T1, T1ce (T1 with contrast-enhanced), T2, and FLAIR (Fluid Attenuation Inversion Recovery). Segmentation of brain tumor, as demonstrated in Fig. 1, is a crucial step of the diagnosis, which localizes the glioma for quantitative investigation [2]. Manual segmentation requires expertise and labor, which is difficult to scale or satisfy the increasing demand for high-quality tumor segmentation.

Computer-aided prognosis has automated the process of brain tumor segmentation. Early works have approached the problem with classic machine learning approaches, *e.g.*, logistic regression [2], combinations of Bayesian classification and Markov Random Field [3], and hierarchical approaches [4]. In recent years, deep learning methods such as deep convolutional neural networks (DCNNs) have been validated to be more effective in feature recognition on numerous biomedical applications, including disease diagnosis [5], [6], [7], health monitoring [8], [9], [10], [11], biomedical image analysis [12], [13], [14], [15], [16], and Electroencephalography (EEG) analysis [17], [18]. Nevertheless, the problem remains challenging in brain tumor segmentation for various

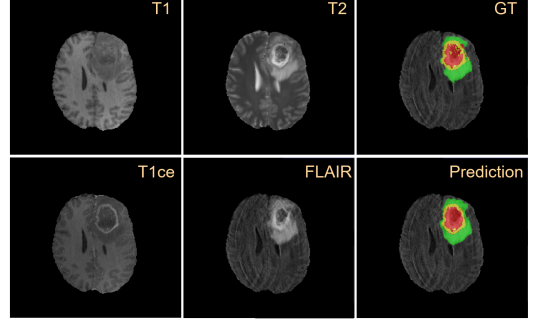


Fig. 1. The BraTS2017 MRI image slices include 4 modalities (T1, T1ce, T2, and FLAIR). In ground-truth (GT) and predicted segmentation maps, the whole tumor (WT) includes all colors (red, yellow, and green). The tumor core (TC) includes red and yellow, and the enhancing tumor (ET) is red.

reasons. First, voxel ambiguities between different tumor classes make it difficult to identify the clear boundaries of brain tumors [19]. Moreover, generalization of tumor features is a demanding task, as there are noticeable variances in the shapes, sizes, and positions of gliomas in MRI [4]. Besides, it requires more effective feature extraction and selection to overcome inhomogeneity in MRI data.

To effectively address the challenges mentioned above, we design a residual channel attention mechanism to calibrate channel-wise feature dependencies, by which the network learns to discriminate regions of brain gliomas. Our lightweight attention module includes 3D global maximum and average pooling on the hidden features’ height, width, and depth. It embeds global channel-wise feature responses to be leveraged by the intermediate layers. We connect the attention modules at each level of an encoder-decoder segmentation architecture. Attention maps generated from attention modules are added to the hidden features to recalibrate the hidden features adaptively. The addition of hidden features also acts as residual connections in the attention module for model optimization. The Main contributions in our paper are summarized in the following,

- We design a brain glioma segmentation architecture with a residual channel attention mechanism. Our proposed attention module dynamically calibrates intermediate features, which enhances the representational capacity of the network.
- We perform extensive experiments on the proposed attention module with various backbone networks and in conjunction with other feature selection and aggregation techniques. Experiments indicate the effectiveness of our approach with superior results than baselines.

^{*}Contributed equally. [†]Corresponding author, email: zengz@i2r.a-star.edu.sg. ¹School of Microelectronics, Shanghai University, Shanghai, China, 200444. ²Institute for Infocomm Research, A*STAR, 1 Fusionopolis Way, 21-01 Connexis, Singapore 138632.

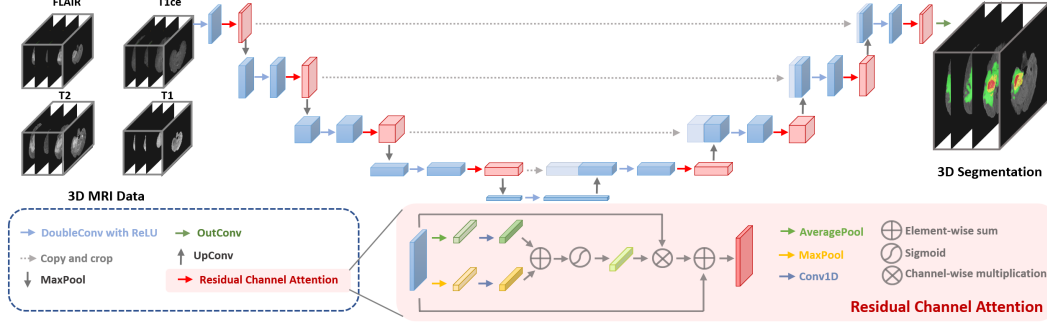


Fig. 2. The proposed segmentation architecture. Residual channel attention modules (labeled as red arrows and cubes) are appended after each DoubleConv layer. Derivation of channel descriptors and residual connections are illustrated in the enclosed rounded rectangle.

II. RELATED WORK

Brain glioma segmentation. Traditional machine learning algorithms have been considered in brain glioma segmentation. A hybrid of clustering and logistic regression algorithm was developed to segment tumors and edema represented by sparse coding [2]. Random decision forests were devised for voxel labeling in brain structures [20]. The Bayesian model was leveraged to improve affinity calculation, which was included in the segmentation by weighted aggregation (SWA) algorithm [3]. These methods are constrained by the representation capacity of hand-crafted features and require profound domain knowledge. DCNNs have achieved remarkable progress in brain glioma segmentation. Back in 2013, the possibility to apply DCNNs to brain glioma segmentation was investigated in [21]. Variations of 3D U-Net were proven effective in 3D graphical information extraction from brain MRI images [22], [23], [24]. Recently, the contextual information of tumors and surroundings were modeled in a feature interaction graph to optimize latent feature representation [19]. Despite growing evaluation results, voxel ambiguity remains a challenging problem in brain glioma segmentation, which has yet to be solved.

Channel attention. Channel attention mechanism adaptively recalibrates channel-wise features by exploiting dependencies between channels. A milestone in this field is the formulation of Squeeze-and-Excitation (SE) blocks [25], [26]. Efficient Channel Attention (ECA) was devised to alleviate dimensionality reduction in SE blocks while greatly reducing model complexity [27]. Channel attention mechanism has been used in conjunction with DCNN for medical imaging segmentation [22], [28], which inspires our work.

III. METHODOLOGY

The residual channel attention network is illustrated in Fig. 2. 3D MRI data, including all modalities, are fed as input of the network. The network includes an encoder-decoder architecture modified from U-Net [12]. The channel attention module, which is described in Sec. III-A, is appended after each double convolution module during the encoding and decoding process. Intermediate features from convolution modules are calibrated by interdependence between channels

modeled in attention modules, which is described in Sec. III-B. The output of the decoder consists of 3D segmentation maps for all tumor classes on the 3D MRI data.

A. Channel Attention Mechanism

The proposed channel attention module is the core to discover channel-wise information and relationship in brain MRI data. For input feature $X \in R^{H \times W \times D \times C}$, 2 feature descriptors, *i.e.*, the max-pooling and average-pooling results at channel c can be calculated in the following,

$$X_{max_pool} = \text{Max}(X(h, w, d)), h \in [0, H), w \in [0, W), d \in [0, D), \quad (1)$$

$$X_{avg_pool} = \frac{1}{H \times W \times D} \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} \sum_{d=0}^{D-1} X(h, w, d), \quad (2)$$

in which $X_{max_pool} \in R^{1 \times 1 \times 1 \times C}$ and $X_{avg_pool} \in R^{1 \times 1 \times 1 \times C}$. H , W , D and C represents height, width, depth, and channels of X , respectively. The 2 descriptors are then transformed in a 1D convolutional layer $Conv$ with shared weights, which efficiently captures channel interaction without dimensionality reduction. Channel-wise addition and Sigmoid activation σ are calculated on the output channel weights sequentially to acquire the combined channel attention map $M(X) \in R^{1 \times 1 \times 1 \times C}$, as shown in the following Equ. 3,

$$M(X) = \sigma(\text{Conv}(X_{max_pool}) + \text{Conv}(X_{avg_pool})). \quad (3)$$

The proposed module is inspired by Modified Efficient Channel Attention (MECA) in [28]. However, there are 2 substantial distinctions. First, there is one more dimension in the attention map in this paper, covering the depth and modalities of the data. Next, the features guided by the attention maps are passed to the subsequent Convolution layers, rather than to the skip connections in [28].

B. Feature Calibration with Residual Connection

We calibrate the intermediate feature X with the attention map $M(X)$ in 2 steps. First, we apply **channel-wise** multiplication on X and $M(X)$. Next, we perform **element-wise** summation on the multiplication result and X . The calibrated feature X' can be calculated in Equ. 4,

$$X' = (M(X) \otimes X) \oplus X, \quad (4)$$

TABLE I
EXPERIMENTAL RESULTS ON BRATS2017.

Method	Dice Score				Sensitivity			Specificity			Hausdorff95		
	Mean	ET	WT	TC	ET	WT	TC	ET	WT	TC	ET	WT	TC
Attention U-Net [24]	0.771	0.672	0.863	0.778	0.847	0.900	0.862	0.996	0.990	0.992	9.347	9.676	10.668
CANet-concat [19]	0.782	0.682	0.861	0.803	0.857	0.922	0.861	0.997	0.989	0.994	7.755	9.377	11.432
3D U-Net [22]	0.794	0.706	0.865	0.810	0.803	0.906	0.829	0.998	0.990	0.995	6.624	8.193	8.958
PRU-Net [29]	0.805	0.710	0.891	0.814	0.788	0.900	0.841	0.998	0.990	0.996	7.205	7.414	9.187
RCAN (CANet-concat)	0.798	0.724	0.883	0.788	0.765	0.906	0.830	0.998	0.998	0.997	6.439	7.093	7.941
RCAN (U-Net)	0.807	0.743	0.882	0.795	0.771	0.913	0.850	0.999	0.997	0.997	5.854	7.514	7.561

where $\otimes$ and $\oplus$ are channel-wise multiplication and element-wise summation, respectively. The summation includes a residual connection from the start of the attention module, enhancing gradient propagation in training.

IV. EXPERIMENTS

A. Dataset and Implementation

We conducted experiments on the public Multimodal Brain Tumor Segmentation Challenge 2017 (BraTS2017), consisting of 285 instances. Each instance contains 4 MR sequences, including T1, T1ce, T2, and FLAIR. The sequence's height, width, and depth are 240, 240, and 155, respectively. Ground-truth annotations contain healthy tissues, enhancing tumor (ET), whole tumor (WT), and tumor core (TC). We followed pre-processing steps in [19] and performed five-fold cross-validation with a random division.

The implementation details are described as follows. Multiple augmentations were applied, including random rotation between ± 20 degrees, random elastic deformation ($\sigma = 10$), random flipping, random scaling, and intensity shift of up to 10%. The data were randomly cropped into $128 \times 128 \times 128$ before being fed into the network with batch size $b_s = 1$. The network was randomly initialized without pre-training and was trained for 200 epochs using Adam optimizer with an initial learning rate of 0.0001. The weight decay was set to $1e-5$. The learning rate was multiplied by 0.2 after epoch 100 and 150. The experiments were carried out on an NVIDIA RTX A5000 GPU with PyTorch 1.9.1.

B. Evaluation Metrics

We follow existing works [30], [19] and use dice score, sensitivity, specificity, and Hausdorff95 distance to measure the segmentation performance. In particular, a lower Hausdorff95 distance between prediction and ground-truth indicates better model performance, while a higher dice score/sensitivity/specificity is more desired. Metrics for the 3 tumor regions, *i.e.*, ET, WT and TC are computed separately.

C. Baseline Methods

Several baselines are compared to our framework. In [22], the 3D variant of U-Net was proposed. Spatial attention was introduced to U-Net in [24]. Reversible blocks were integrated into U-Net to reduce memory cost in [29]. In [19], a feature interaction graph was combined with U-Net to raise context awareness and reduce local ambiguity. Our proposed Residual Channel Attention Networks (RCAN) are described as below,

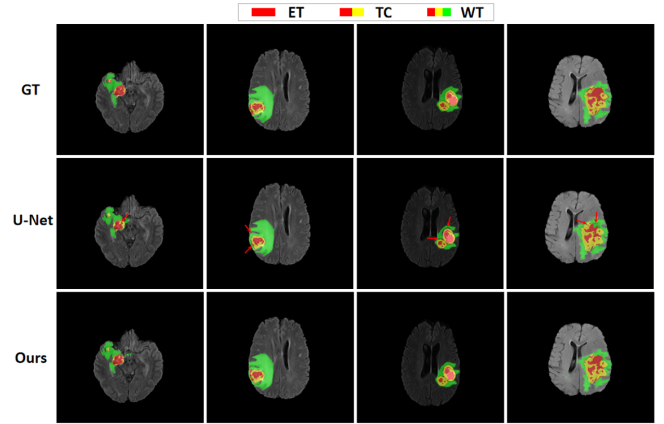


Fig. 3. Illustrations of segmentation maps from the ground-truth (GT), U-Net [22], and our RCAN (U-Net) in BraTS2017. Arrows indicate incorrect segmentation in the comparison method.

- RCAN (U-Net): 3D U-Net with residual channel attention integrated.
- RCAN (CANet-concat): Residual channel attention integrated into the backbone CANet, with the concatenated feature interaction graph [19].

D. Results and Discussion

In Table. I, experimental results of our method and baselines are presented. Comparing 3D U-Net and RCAN (U-Net), our method achieves a better result under most metrics for all tumor classes. It thus proves that our residual channel attention can calibrate the features learned by the model and enhance its performance. Moreover, the proposed network with channel attention outperforms Attention U-Net [24], which attempts to optimize the U-Net with spatial attention. Our method also achieves the lowest Hausdorff95 distance between the predicted segmentation masks and the ground-truth in ET (enhancing tumor) and TC (tumor core), indicating more accurate edge predictions in the segmentation maps. Besides, RCAN (U-Net) reaches higher specificity but lower sensitivity than other U-Net variants in ET and TC, which is a possible direction for future optimization. With CANet-concat as the backbone, we further improve the output of the original CANet-concat. The result indicates the applicability of the proposed channel attention in conjunction with other feature selection mechanisms.

In Fig. 3, we visualize selected segmentation maps. We observe that our network is capable of detecting small,

isolated enhancing tumor (ET) areas, and can also predict the correct edges for ET, where the comparison method fails. Moreover, there are visible improvements in our edge prediction for tumor core (TC) and whole tumor (WT), which are in accordance with results from Table. I.

TABLE II
EXPERIMENTAL RESULTS ON THE EXCLUSION OF COMPONENTS IN
RESIDUAL CHANNEL ATTENTION.

Experiments	Dice Score			
	Mean	ET	WT	TC
Without max pooling	0.796	0.717	0.889	0.782
Without average pooling	0.790	0.715	0.882	0.785
Without residual connection	0.793	0.688	0.892	0.795
With all components	0.807	0.743	0.882	0.795

We performed ablation studies on components in the residual channel attention network. The result is presented in Table. II. The overall performance drops when any one of the pooling operations or the residual connection is removed. The performance decreases most when the average pooling operation is excluded, even slightly worse than the baseline U-Net. In summary, the ablation studies prove the necessity of each component in the proposed architecture.

V. CONCLUSIONS

In conclusion, we have developed a novel network for brain glioma segmentation. To address class ambiguities and high variability in tumor features, we devise a residual channel attention mechanism, which calibrates latent features and enhances the representational power of the network. Experiments are conducted on the 3D brain MRI dataset, BraTS2017. The results prove the effectiveness of the proposed network quantitatively and qualitatively. In the future, domain knowledge could be incorporated to the algorithm.

REFERENCES

- [1] Peter McL Black and Jay S Loeffler, *Cancer of the nervous system*, Lippincott Williams & Wilkins, 2005.
- [2] Hoo-Chang Shin, "Hybrid clustering and logistic regression for multi-modal brain tumor segmentation," in *Proc. of Workshops and Challenges in MICCAI*, 2012.
- [3] Jason J Corso, Eitan Sharon, Shishir Dube, Suzie El-Saden, Usha Sinha, and Alan Yuille, "Efficient multilevel brain tumor segmentation with integrated bayesian model classification," *IEEE transactions on medical imaging*, 2008.
- [4] Stefan Bauer, Thomas Fejes, Johannes Slotboom, Roland Wiest, Lutz-P Nolte, and Mauricio Reyes, "Segmentation of brain tumor images based on integrated hierarchical classification and regularization," in *MICCAI BraTS Workshop*, 2012.
- [5] Ziyuan Zhao, Kerui Zhang, Xuejie Hao, Jing Tian, Matthew Chin Heng Chua, Li Chen, and Xin Xu, "Bira-net: Bilinear attention net for diabetic retinopathy grading," in *ICIP*. IEEE, 2019.
- [6] Peisheng Qian, Ziyuan Zhao, Cong Chen, Zeng Zeng, and Xiaoli Li, "Two eyes are better than one: Exploiting binocular correlation for diabetic retinopathy severity grading," in *EMBC*. IEEE, 2021.
- [7] Ranadeep Deb, Ganapati Bhat, Sizhe An, Umit Ogras, and Holly Shill, "Trends in technology usage for parkinson's disease assessment: A systematic review," *medRxiv*, 2021.
- [8] Steven Yi, Minta Lu, Adam Yee, John Harmon, Frank Meng, and Saurabh Hinduja, "Enhance wound healing monitoring through a thermal imaging based smartphone app," in *Medical Imaging 2018: Imaging Informatics for Healthcare, Research, and Applications*, 2018.

- [9] Ganapati Bhat, Yigit Tuncel, Sizhe An, and Umit Y Ogras, "Wearable IoT Devices for Health Monitoring," in *TechConnect Briefs 2019*, 2019, pp. 357–360.
- [10] Sizhe An, Yigit Tuncel, Toygun Basaklar, Gokul K Krishnakumar, Ganapati Bhat, and Umit Y Ogras, "Mgait: Model-based gait analysis using wearable bend and inertial sensors," *ACM Transactions on Internet of Things*, vol. 3, no. 1, pp. 1–24, 2021.
- [11] Sizhe An and Umit Y Ogras, "Mars: mmwave-based assistive rehabilitation system for smart healthcare," *ACM Transactions on Embedded Computing Systems (TECS)*, vol. 20, no. 5s, pp. 1–22, 2021.
- [12] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-net: Convolutional networks for biomedical image segmentation," in *MICCAI*. Springer, 2015.
- [13] Ziyuan Zhao, Xiaoyan Yang, Bharadwaj Veeravalli, and Zeng Zeng, "Deeply supervised active learning for finger bones segmentation," in *EMBC*. IEEE, 2020.
- [14] Shumeng Li, Ziyuan Zhao, Kaixin Xu, Zeng Zeng, and Cuntai Guan, "Hierarchical consistency regularized mean teacher for semi-supervised 3d left atrium segmentation," in *EMBC*, 2021.
- [15] Jiapan Gu, Ziyuan Zhao, Zeng Zeng, Yuzhe Wang, Zhengyiren Qiu, Bharadwaj Veeravalli, Brian Kim Poh Goh, Glenn Kunnath Bonney, Krishnakumar Madhavan, Chan Wan Ying, et al., "Multi-phase cross-modal learning for noninvasive gene mutation prediction in hepatocellular carcinoma," in *EMBC*. IEEE, 2020.
- [16] Ziyuan Zhao, Zeng Zeng, Kaixin Xu, Cen Chen, and Cuntai Guan, "Dsal: Deeply supervised active learning from strong and weak labelers for biomedical image segmentation," *IEEE Journal of Biomedical and Health Informatics*, 2021.
- [17] Xiaodong Qu, Qingtian Mei, Peiyan Liu, and Timothy Hickey, "Using eeg to distinguish between writing and typing for the same cognitive task," in *International Conference on Brain Function Assessment in Learning*. Springer, 2020.
- [18] Yi Ding, Nigel Wei Jun Ang, Aung Aung Phyto Wai, and Cuntai Guan, "Learning generalized representations of eeg between multiple cognitive attention tasks," in *EMBC*. IEEE, 2021.
- [19] Zhihua Liu, Lei Tong, Long Chen, Feixiang Zhou, Zheheng Jiang, Qianni Zhang, Yinhai Wang, Caifeng Shan, Ling Li, and Huiyu Zhou, "Canet: Context aware network for brain glioma segmentation," *IEEE Transactions on Medical Imaging*, 2021.
- [20] Joana Festa, Sergio Pereira, Jose Antonio Mariz, Nuno Sousa, and Carlos A Silva, "Automatic brain tumor segmentation of multi-sequence mr images using random decision forests," *Proceedings of NCI-MICCAI BRATS*, 2013.
- [21] Darko Zikic, Yani Ioannou, Matthew Brown, and Antonio Criminisi, "Segmentation of brain tumor tissues with convolutional neural networks," *MICCAI-BRATS*, 2014.
- [22] Özgün Çiçek, Ahmed Abdulkadir, Soeren S Lienkamp, Thomas Brox, and Olaf Ronneberger, "3d u-net: learning dense volumetric segmentation from sparse annotation," in *MICCAI*. Springer, 2016.
- [23] Robin Brügger, Christian F Baumgartner, and Ender Konukoglu, "A partially reversible u-net for memory-efficient volumetric image segmentation," in *MICCAI*. Springer, 2019.
- [24] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al., "Attention u-net: Learning where to look for the pancreas," *arXiv preprint arXiv:1804.03999*, 2018.
- [25] Jie Hu, Li Shen, and Gang Sun, "Squeeze-and-excitation networks," in *IEEE CVPR*, 2018.
- [26] Ziyuan Zhao, Kartik Chopra, Zeng Zeng, and Xiaoli Li, "Sea-net: Squeeze-and-excitation attention net for diabetic retinopathy grading," in *ICIP*. IEEE, 2020.
- [27] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu, "Eca-net: efficient channel attention for deep convolutional neural networks, 2020 ieee," in *CVPR*, 2020.
- [28] Changlu Guo, Márton Szemenyei, Yangtao Hu, Wenle Wang, Wei Zhou, and Yugen Yi, "Channel attention residual u-net for retinal vessel segmentation," in *ICASSP*. IEEE, 2021.
- [29] Christian F Baumgartner Robin Brügger and Ender Konukoglu, "A partially reversible u-net for memory-efficient volumetric image segmentation," in *MICCAI*. Springer, 2019.
- [30] Guotai Wang, Wenqi Li, Sébastien Ourselin, and Tom Vercauteren, "Automatic brain tumor segmentation using cascaded anisotropic convolutional neural networks," in *International MICCAI brainlesion workshop*. Springer, 2017.