Pseudo-3D CNN with inter-slice attention for glioma grading

Shanshan Du[a], Zhuoqun Cao[b], Rui-Wei Zhao[c], Xiaobo Zhang[d], Rui Feng[*bcd]
[a]School of Information Science and Technology, Fudan University, Shanghai, China; [b]School of Computer Science, Shanghai Key Laboratory of Intelligent Information Processing, Fudan University, Shanghai, China; [c]Academy for Engineering and Technology, Fudan University, Shanghai, China; [d]Children's Hospital, Fudan University, Shanghai, China

ABSTRACT

Gliomas comprise around 80 percent of all malignant brain tumors which can be further classified into the low-grade glioma and high-grade glioma categories. Compared to low-grade gliomas, high-grade gliomas show more malignant behaviour since they usually grow rapidly and frequently destroy healthy brain tissue. Therefore, it is important to determine the malignancy of gliomas for initial treatment plan. Most existing methods are based on radiomics or transfer learning which extract features from single-slice without considering that 3D features between adjacent slices can provide stronger discriminative power. In this work, we propose to incorporate the attention mechanism and pseudo-3d module into a deep convolutional network architecture, which can reduce model size and produce fine features for glioma grading. Our work focuses on the inter-slice relationship and propose an attention unit, named "Inter-Slice Attention Module", which adaptively refines intermediate feature maps by modelling dependencies between adjacent slices. We evaluate our method on an open dataset of gliomas, achieving mean accuracy of 89.47%.

Keywords: Glioma grading, neural networks, visual attention

1. INTRODUCTION

Gliomas are the brain tumors that originate from glial cells. They are the most common primary type of tumor of the brain. World Health Organization (WHO) classifies glioma into grade I to IV by both histology and molecular features. These grades strongly indicate whether the tumor is benign or malignant and tell the malignancy scales of the tumor[1,2]. In real practice, the correct diagnosis of glioma grades, such as low-grade (WHO grade II), intermediate-grade (WHO grade III) gliomas (LGG), high-grade gliomas (HGG) and glioblastoma (WHO grade IV) is very important to initial treatment decision[3-5]. For example, glioblastoma (GBM, WHO grade IV) has the highest incidence rate of 46.1%, which is more malignant and highly heterogeneous compared to lower grade gliomas[6].

At present, the radiologists evaluate the malignancy of tumor by studying the patients' head scan images. The complementary information between different imaging modalities not only imposes a burden on the radiologist, but also requires certain experience. Therefore, it is of great value to develop advanced and reliable AI methods to classify the input glioma images automatically and intelligently into LGG and HGG.

To this end, machine learning techniques has gradually been applied to the study of gliomas, especially radiomics. These methods include both shallow models[7,8] and deep learning models[9-11]. However, most of these stronger deep learning methods only extract features from single slice, ignoring the sequential information between slices, which is effect very important for human experts to make diagnosis.

In clinical diagnosis, radiologists will pay attention to individual slices that contain abnormal region and consider complementary information in images of different modalities. Specific slices can provide more discriminating features. Emphasizing these features and suppressing others will improve the representative power of intermediate features throughout network. Considering the huge computational cost and memory requirement of a very deep 3D CNN, in this work, we try to use Pseudo-3D convolutional neural network with inter-slice attention to simulate such process.

The major contributions of this paper can be summarized as follows.

* fengrui@fudan.edu.cn

Third International Conference on Computer Science and Communication Technology (ICCSCT 2022)
edited by Yingfa Lu, Changbo Cheng, Proc. of SPIE Vol. 12506, 125060I
 · 0277-786X · doi: 10.1117/12.2662626

(1) A novel inter-slice attention module with both channel-wise and spatial-wise attention submodules is proposed. The proposed attention module can work well with the Pseudo-3D features and is effective to generate better feature maps for improved glioma grading.

(2) A full glioma grading network framework with both Pseudo-3D convolutions and inter-slice attention modules are designed. It can achieve automatic glioma grading in an end-to-end manner.

(3) The effectiveness of our proposed modules and networks are verified in the experiments on the popular BRATS 2017 benchmark. Results shows that our proposed inter-slice attention module and networks can achieve improved performances compared to the state-of-the-arts.

This rest of this paper is organized as follows. In Section 2, we briefly review the recent related methods. In Section 3, our proposed network framework and inter-slice attention module are introduced. Experimental evaluations of the proposed method are conducted in Section 4. Finally, Section 5 concludes this paper.

2. RELATED WORK

Both shallow and deep models have been studied to solve the glioma grading problem. Regarding shallow models, an extreme gradient boosting classifier was used by Chen et al.[7] for the grading of gliomas based on 105 Radiomic features which are all obtained from the whole tumor region of each modality and selected by SVM-RFE, including numbers of texture features, shape features as well as first-order features. Zacharaki et al.[8] extract three kinds of features from four ROI (enhancing & non-enhancing neoplastic, necrotic, edematous) to classify tumors into different types and grades. Obviously, in radiomic-based methods, the prior information of domain expertise is very important, especially for the design of quantitative features, which leads to inefficiency and dependency on experience.

Compared to massive amounts of conventional machine learning methods that are dependent on fully hand-crafted features, deep learning methods can implicitly learn to more effectively extract representative features from the image data. The success of deep learning models demonstrates the outstanding ability of convolutional neural networks (CNN) to feature extraction. For example, Decuyper et al.[9] compare the discrimination ability of features extracted by two different methods of radiomics and pre-trained CNN. Note that the accuracy of pretrained CNN is 82% without requiring ROI annotations, 7.6% lower than radiomics-based method. Ge et al.[10] introduce a saliency-aware approach to enhance the tumor region which scale down the non-tumor regions by a factor and has shown marked improvement on the performance of glioma classification. Yang el al.[11] evaluate the performance of AlexNet and GoogleNet simply pre-trained on ImageNet to study transfer learning on glioma grading. Note that these deep learning methods only extract features from single slice and ignore to exploit the valuable sequential information between slices.

3. METHODOLOGY

In this section, we first introduce a deep learning framework based on Pseudo-3D[12] and the proposed inter-slice attention module (IAM) for glioma grading. Pseudo-3D decouples the traditional $3 \times 3 \times 3$ convolutions into 2D intra-slice convolutions and 1D inter-slice convolutions. Then, the attention mechanism based on inter-slice is introduced to refine the intermediate feature map throughout the network to enhance the discriminant feature.

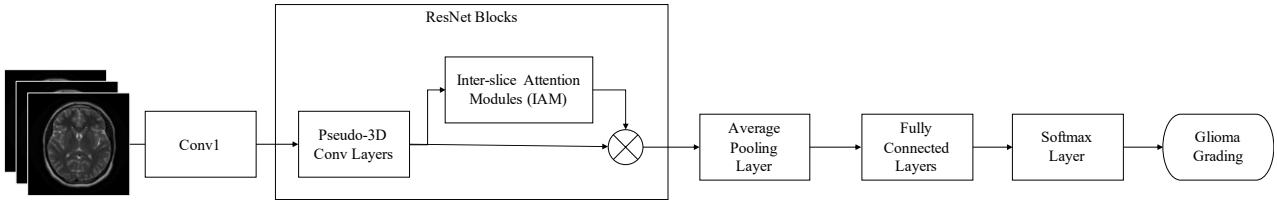

Figure 1. Framework of the proposed glioma grading networks with Pseudo-3D convolutional layers and inter-slice attention modules (IAM).

3.1 Network framework

The overall framework for the proposed end to end glioma grading networks is illustrated in Figure 1. It takes 3D MRI image volume as input. After the first convolutional layer, it will go through repeated ResNet blocks with Pseudo-3D convolutional layers and inter-slice attention modules. Then the obtained feature map will be average pooled to reduce the

size and be forwarded into the fully convolutional layers and softmax layer to produce the final glioma grading predictions. Next, the Pseudo-3D CNN backbone and the proposed inter-slice attention module will be discussed in detail.

3.2 Pseudo-3D CNN

Deep Residual Network (ResNet)[13] offers the possibility to train deep neural networks and achieves state-of-the-art performances in a wide range of computer vision tasks. In the conventional ResNet, only 2D convolutional filters are widely used to extract the image features. In order to encode the multiple dimensional information from image volumes (like MR images), the intuitive method is to replace the 2D convolutional filters of CNN with 3D filters. In this way, it not only processes the spatial information in within individual slices, but also establish the connection between adjacent slices. However, training a 3D ResNet from scratch requires large amount of computational cost and memory demand. As described in Reference[12], a standard 3 × 3 × 3 convolutional layer can be replaced by the combination of a 1 × 3 × 3 convolutional filter for intra-slice domain feature extraction plus a 3 × 1 × 1 convolutional filter for inter-slice domain feature extraction. In this way, the associations between adjacent slices can be captured by the later 3 × 1 × 1 filter, while significantly reduce the model size. Such combination of 1 × 3 × 3 convolutional layer followed by 3 × 1 × 1 convolutional filter is also abbreviated as P3D convolutional layer.

In this work, the famous ResNet-50[13] is adopted as backbone with some modifications for glioma grading. We first build the ResNet-50 and replace the 2D convolutional layers with the P3D convolutional layers according to the P3D-A architecture introduced in Reference[12]. The P3D convolutional layers were adopted to replace standard 3 × 3 × 3 convolutional layers, which makes spatial 2D filters followed by 1D filters in a cascaded manner. In brief, its calculation can be expressed as

$$\mathrm{P3D}(x) = f_{3\times1\times1,C}(f_{1\times3\times3,C}(x)), \tag{1}$$

where $f_{3\times1\times1,C}$ stands for a convolutional layer whose kernel size is 3 × 1 × 1 and output channel C; $f_{1\times3\times3,C}$ denotes a convolutional layer with kernel size of 1 × 3 × 3 and output channel C.

After the P3D convolutional layer, both ReLU an BatchNorm layers are applied. The conventional ReLU function[14] is defined as

$$f(x) = \mathrm{ReLU}(x) = \max(0, x). \tag{2}$$

While the adopted ReLU activation functions in this work are in the form of Leaky ReLU[15], whose activation function has a normal slope on the positive axis and usually much smaller slope on the negative axis. Mathematically, its calculation can be expressed as

$$f(x) = \begin{cases} x & \text{if } x > 0 \\ ax & \text{otherwise.} \end{cases} \tag{3}$$

We can preset the slope coefficient a before training. In other word, it is not learned during training. It can help alleviate the suffer from sparse gradients. The default slope coefficient is set to 10^{-2} in this work.

The BatchNorm layers perform the batch normalization operation[16]. Usually, the input feature map to deep neural networks had better be normalized for the sake of numerical computation. BatchNorm essentially performs the mean 0 and variance 1 normalization, or Whitening, to the intermediate layers of the deep neural networks. By using BatchNorm, the neurons' activation can become closer to a standard gaussian distribution, which has been proved to effective in reducing the exploding and vanishing gradients in computation.

In summary, the proposed architecture of Pseudo-3D CNN backbone is listed in Table 1.

Table 1. Detailed backbone architecture of our proposed model for glioma grading.

Layer/block	Output shape	Kernel settings
Conv1	64 × 64 × 64, 64	7 × 7 × 7, 64, s = 2
block_1	32 × 32 × 32, 256	k = 3, s = 2, Max-Pool
		1 × 1 × 1, 64
		1 × 3 × 3, 3 × 1 × 1, 64

Layer/block	Output shape	Kernel settings
		1 × 1 × 1, 256
block_2	16 × 16 × 16, 512	1 × 1 × 1, 128 1 × 3 × 3, 3 × 1 × 1, 128 1 × 1 × 1, 512
block_3	8 × 8 × 8, 1024	1 × 1 × 1, 256 1 × 3 × 3, 3 × 1 × 1, 256 1 × 1 × 1, 1024
block_4	4 × 4 × 4, 2048	1 × 1 × 1, 512 1 × 3 × 3, 3 × 1 × 1, 512 1 × 1 × 1, 2048
avgpool	4 × 4, 2048	k = 4, d = 2, Avg-Pool
dropout + fc	2	-

Note: Leaky ReLU and Batch Normalization layers are added sequentially after every convolution. Block Size: [3, 4, 6, 3]; Input Size: channel × 128 × 128 × 128. *k*, *s* and *d* stand for kernel size, stride and dimension, respectively.

3.3 Inter-slice attention

When reading MRI scans, the radiologists first obtain the regions of interest by quickly browsing the whole scan. Then they pay more attention into these regions, so as to suppress the useless information and focus on the details provided by the targets, and finally make the diagnosis.

As we mentioned before, information that can discriminate tumor grade only locate on specific slices. In order to make the model to judge like the experts, we propose a network module named Inter-slice Attention Module (IAM) to efficiently emphasize meaningful features and suppress others in this part. Recently, quite a number of studies[17,18] investigated the mechanism of attention and its role in convolutional neural networks, and the characteristics of attention mechanism are well suited to simulate such process.

Given an arbitrary feature map denoted as $x \in \mathbb{R}^{C\times D\times H\times W}$, where *C*, *D*, *H*, *W* respectively refer to the channel size, depth, height, and width of the input image volume, our IAM blocks performs feature re-calibration to get a refined feature map denoted as x'. The whole process can be formulated as

$$x' = a^s \otimes a^c \otimes x, \tag{4}$$

where $\otimes$ means the operation of element-wise multiplication; a^c and a^s are the attention maps inferred by our inter-slice channel-wise and spatial-wise attention submodules. Its framework is depicted in Figure 2.

In order to produce attention map, we need to exploit the inter-slice relationship between feature maps. Each $H \times W$ feature map along the depth axis represent contextual information captured by convolutional filters with a local receptive field. We adopt average-pooling to aggregating the spatial information, generating descriptor $x^c \in \mathbb{R}^{C\times D\times 1\times 1}$. Then, to fully capture channel-wise inter-slice dependency, a gating mechanism is employed to obtain channel-wise inter-slice attention map $a^c \in \mathbb{R}^{C\times D\times 1\times 1}$

$$a^c = \sigma(f_{3\times 1\times 1,C}(\delta(f_{3\times 1\times 1,C/r}(x^c)))), \tag{5}$$

where $f_{3\times 1\times 1,C}$ is a convolutional layer of kernel size 3 × 1 × 1 and output channel *C*; *r* is the channel reduction ratio; and σ, δ denote the sigmoid and ReLU activation function respectively.

Likewise, average-pooling is adopted to aggregating the feature along the channel axis, generating feature descriptor $x^s \in \mathbb{R}^{1\times D\times H\times W}$. Then, a single-layer network is applied to generate a spatial inter-slice attention map $M^s \in \mathbb{R}^{1\times D\times H\times W}$

$$a^s = \sigma(f_{3\times1\times3,C}(x^s)), \tag{6}$$

where $f_{3\times1\times3,C}$ stands for a three-dimensional convolutional layer of kernel size 3 × 3 × 3.

These two attention modules are placed in a sequential manner. The input feature map x is first squeezed along spatial (H and W) axis and forwarded to channel-wise submodule, generating the attention map a^s. Feature x_a^c is obtained by element-wise multiplying x and a^c. Then feature x_a^c is squeezed along the channel (C) axis and forwarded to spatial-wise submodule to generate the attention map a^s. The final output x' is obtained by further element-wise multiplying x_a^c with a^s.

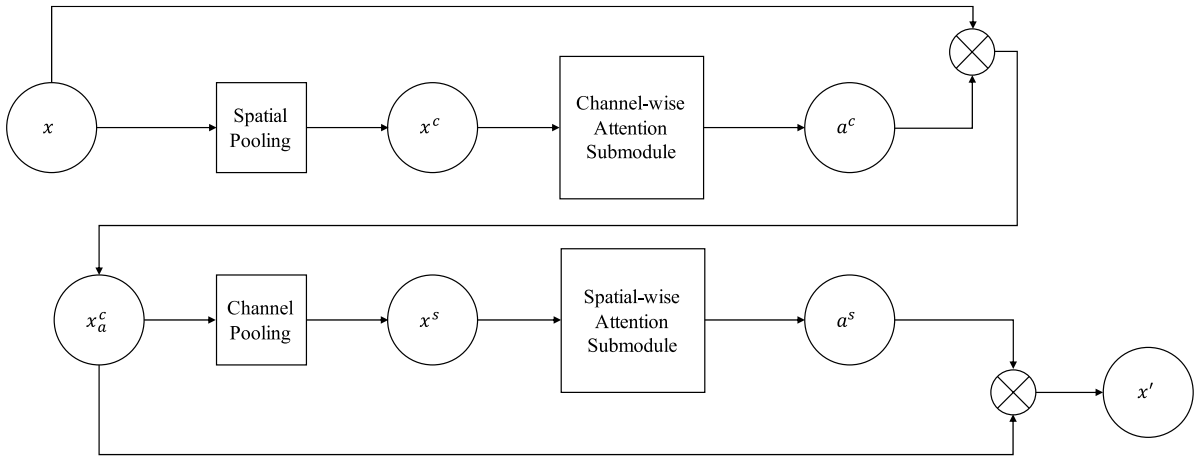

Figure 2. Framework of the proposed Inter-slice Attention Module (IAM).

Note: The ⊗ symbol refers to the element-wise multiplication operation.

4. EXPERIMENT

4.1 Dataset

The dataset evaluated in this work is BRATS 2017[19,20]. It provides clinical imaging data from 285 glioma patients, including 210 patients with Glioblastoma (HGG) and 75 patients with lower-grade glioma (LGG, WHO grade II & III). Each case contains routine pre-operative multi-parameter MRI scans from multiple institutions and has been histologically diagnosed. These scans include pre-contrast and post-contrast (1) T1-weighted (T1 and T1ce), (2) T2-weighted (T2) together with (3) T2 Fluid-Attenuated Inversion Recovery (Flair) volumes. A series of pre-processing methods were applied to each sequence, including co-registering to the same anatomic template, resampling in a standardized axial orientation by a chosen linear interpolator to 1mm voxel resolution, skull-stripping, noise reduction and histogram matching. Finally, the four categories of annotations of the tumor sub-region are obtained through the computer-aided segmentation[21] and manual correction, which respectively represent (1) the enhancing part of the tumor core, (2) the non-enhancing part of the tumor core, (3) necrotic and (4) the peritumoral edema.

4.2 Experimental setups

In the following, we briefly describe the training setting in our experiment including the pre-processing details. We will also explain our strategies to select some key parameters.

Image normalization: All volumes intensity values are normalized for computational concerns. The normalization is achieved by first subtracting the mean values and the dividing by the standard deviations. The normalization is independently performed to each example and each modality.

Cropping and padding: All volumes are cropped to fixed size bounding boxes and with the tumor region kept in center. Bounding boxes are generated from annotations for fixed image size of 128 × 128 × 128 pixels. Padding operations are applied to ensure that all image patches meet the fixed size.

Train/Validation/Test subsets: The entire BRATS 2017 dataset containing 285 subjects. They are partitioned into three subsets: 171 subjects are used as the training set, 57 subjects as the validation set, and 57 subjects as the test set. HGG and LGG maintain the same ratio in all three subsets.

Data augmentation: Only LGG is flipped left and right, in order to deal with class imbalance and avoid over-fitting.

Learning rate scheduling and others: The total number of epochs in training is 100. We use the initial learning rate of 3×10^{-4} and decayed by a factor of 0.1 at epoch 20 and 50. Batch size is set to 4.

4.3 Results

To test the effectiveness of the proposed IAM, we evaluate and compare the Pseudo-3D CNN with and without IAM on BRATS 2017 under five-fold cross-validation. The obtained performances are summarized in Table 2. It displays the average performance of proposed method, including the validation accuracy, validation AUC, test accuracy as well as the test AUC. In this experiment, the IAMs are placed after the last convolution layer of the non-identity branch of every residual module.

Table 2. Performance of Pseudo-3D CNN with and without our proposed IAM on BRATS 2017.

Model	Accuracy (validation)	AUC (validation)	Accuracy (test)	AUC (test)
Pseudo-3D CNN w/o IAM	87.72	89.52	85.96	84.04
Pseudo-3D CNN w/ IAM	91.58	91.52	89.47	86.42

Note: Average scores of validation accuracy, validation AUC, test accuracy and test AUC are reported.

In particular, the Pseudo-3D CNN model without IAM has the accuracy of 87.72% and 85.96% on the validation and test set respectively. And it has the AUC scores of 89.52% and 84.04% on the validation set and test set respectively. While the Pseudo-3D CNN model with IAM has the accuracy of 91.58% and 89.47% on the validation set and test set respectively. And it has the AUC scores of 91.52% and 86.42% on the validation set and test set respectively. It can be found that compared to Pseudo-3D CNN model without IAM, the model with our IAM achieves obvious improvements on test accuracy and AUC by more than 3% and 2% respectively. The displayed results discussed above successfully demonstrate that our proposed IAM is very effective.

In addition, Figure 3 shows the performance of the best run in all five-fold cross-validation. The proposed method exhibits relatively high accuracy and performs well in both categories.

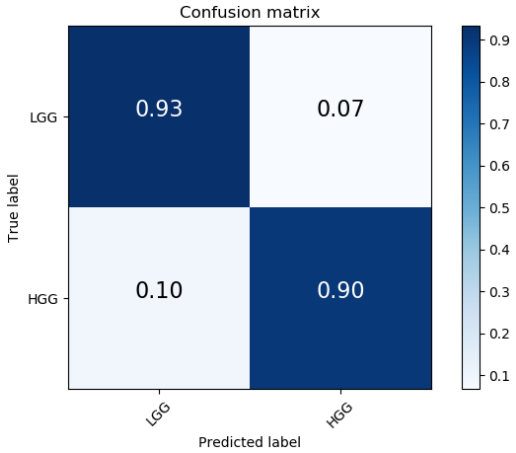

Figure 3. Confusion matrix showing the performance of proposed method on BRATS2017 dataset.

Table 3 shows comparison with 2 state-of-the-art methods. According to our evaluation, Decuyper (method 2)[9] achieves an accuracy at 83.8% on the test set. Ge[10] achieves a higher accuracy at 88.07% on the test set. Note that Ge[10] scales down

the non-tumor regions by a factor of 1/3 as a saliency-aware strategy. Finally, our model reaches the highest accuracy at 89.47%. These results are effective in demonstrating that the performance our proposed method is very competitive.

Table 3. Comparison with state-of-the-art methods on the test set of BRATS 2017.

Method	Accuracy (test)
Decuyper (method 2)[9]	83.80
Ge (with enhancement)[10]	88.07
Ours	89.47

5. CONCLUSION

In this work, we propose a Pseudo-3D CNN with inter-slice attention for glioma classification. Our results on the BRATS 2017 benchmark show that the proposed Pseudo-3D CNN and inter-slice attention module in this work are effective for learning discriminate features, which also exhibit better performance. A possible future direction is to focus on automatic segmentation of glioma with the inter-slice attention under potential class imbalance.

ACKNOWLEDGEMENTS

This work was supported (in part) by the Science and Technology Commission of Shanghai Municipality (No. 20DZ1100205, No. 21511104502, No. 20511100800, No. 20511101103).

REFERENCES

[1] Louis, D. N., Perry, A., et al., "The 2016 world health organization classification of tumors of the central nervous system: A summary Acta Neuropathologica 131(6), 803-820 (2016).
[2] Ostrom, Q. T., Bauchet, L., Davis, F. G., et al., "The epidemiology of glioma in adults: A "state of the science" review," Neuro-oncology 16(7), 896-913 (2014).
[3] DeAngelis, L. M., "Brain tumors," New England Journal of Medicine 344(2), 114-123 (2001).
[4] Caulo, M., Panara, V., et al., "Data-driven grading of brain gliomas: A multiparametric MR imaging study," Radiology 272(2), 494-503 (2014).
[5] Zhuge, Y., Ning, H., et al., "Automated glioma grading on conventional MRI images using deep convolutional neural networks," Medical Physics 47(7), 3044-53 (2020).
[6] Ostrom, Q. T., Gittleman, H., et al., "Cbtrus statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2008-2012," Neuro-oncology 17(suppl 4), iv1-iv62 (2015).
[7] Chen, W., Liu, B., et al., "Computer-aided grading of gliomas combining automatic segmentation and radiomics," International Journal of Biomedical Imaging, 2512037 (2018).
[8] Zacharaki, E. I., Wang, S., et al., "MRI-based classification of brain tumor type and grade using SVM-RFE," IEEE Inter. Symp. on Biomedical Imaging: From Nano to Macro, (2009).
[9] Decuyper, M., Bonte, S. and Holen, R. V., "Binary glioma grading: Radiomics versus pre-trained CNN features," Inter. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI), (2018).
[10] Ge, C., Qu, Q., Gu, I. Y. H. and Jakola, A. S., "3D multi-scale convolutional networks for glioma grading using MR images," IEEE Inter. Conf. on Image Processing (ICIP), (2018).
[11] Yang, Y., Yan, L. F., et al., "Glioma grading on conventional MR images: A deep learning study with transfer learning," Frontiers in Neuroscience 12, 804 (2018).
[12] Qiu, Z., Yao, T. and Mei, T., "Learning spatio-temporal representation with pseudo-3D residual networks," IEEE Inter. Conf. on Computer Vision (ICCV), (2017).
[13] He, K., Zhang, X., et al., "Deep residual learning for image recognition," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), (2016).
[14] Nair, V. and Hinton, G., "Rectified linear units improve restricted Boltzmann machines," Inter. Conf. on Machine Learning (ICML), (2010).
[15] Maas, A. L., Hannun, A. Y. and Ng, A. Y., "Rectifier nonlinearities improve neural network acoustic models," Inter. Conf. on Machine Learning (ICML), (2013).

[16] Ioffe, S. and Szegedy, C., "Batch normalization: Accelerating deep network training by reducing internal covariate shift," Inter. Conf. on Machine Learning (ICML), (2015).
[17] Hu, J., Shen, L. and Sun, G., "Squeeze-and-excitation networks," IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), (2018).
[18] Woo, S., Park, J., et al., "CBAM: Convolutional block attention module," European Conf. on Computer Vision (ECCV), (2018).
[19] Bakas, S., Akbari, H., et al., "Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features," Scientific Data 4, 170117 (2017).
[20] Menze, B., Jakab, A., et al., "The multi-modal brain tumor image segmentation benchmark (BRATS)," IEEE Transactions on Medical Imaging 34(10), 1993-2024 (2014).
[21] Wang, G., Li, W., et al., "GLISTRboost: Combining multimodal MRI segmentation, registration, and biophysical tumor growth modeling with gradient boosting machines for glioma segmentation," Inter. Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI) Brainlesion Workshop, (2017).