LKAU-Net: 3D Large-Kernel Attention-Based U-Net for Automatic MRI Brain Tumor Segmentation

Hao Li[1,2], Yang Nan[1], and Guang Yang[1,3(✉)]

[1] National Heart and Lung Institute, Faculty of Medicine, Imperial College London, London, UK
g.yang@imperial.ac.uk
[2] Department of Bioengineering, Faculty of Engineering, Imperial College London, London, UK
[3] Royal Brompton Hospital, London, UK

Abstract. Automatic segmentation of brain tumors from multi-modal magnetic resonance images (MRI) using deep learning methods has emerged in assisting the diagnosis and treatment of brain tumors. However, the diversity in location, shape, and appearance of brain tumors and the dominance of background voxels in MRI images make accurate brain tumor segmentation difficult. In this paper, a novel 3D large-kernel attention-based U-Net (LKAU-Net) is proposed to address these problems for accurate brain tumor segmentation. Advantages of convolution and self-attention are combined in the proposed 3D large-kernel (LK) attention module, including local contextual information, long-range dependence, and channel adaptability simultaneously. The network was trained and evaluated on the multi-modal Brain Tumor Segmentation Challenge (BraTS) 2020 dataset. The effectiveness of the proposed 3D LK attention module has been proved by both five-fold cross-validation and official online evaluation from the organizers. The results show that our LKAU-Net outperformed state-of-the-art performance in delineating all three tumor sub-regions, with average Dice scores of 79.01%, 91.31%, and 85.75%, as well as 95% Hausdorff distances of 26.27, 4.56, and 5.87 for the enhancing tumor, whole tumor, and tumor core, respectively.

Keywords: Brain tumor segmentation · Attention · Deep learning · MRI

1 Introduction

Segmentation on magnetic resonance imaging (MRI) is a critical step in the treatment of brain tumors, allowing clinicians to determine the location, extent, and subtype of the tumor. This aids not just in the diagnosis but also in radiotherapy planning or surgery planning. Further, the segmentation of longitudinal MRI scans can be used to monitor the growth or shrinkage of brain tumors.

G. Yang et al. (Eds.): MIUA 2022, LNCS 13413, pp. 313–327, 2022.
https://doi.org/10.1007/978-3-031-12053-4_24

Given the significance of this job, the accurate segmentation of tumor regions is typically accomplished manually by experienced radiologists in current clinical practice. This is a time-consuming and laborious process that requires considerable effort and knowledge. Besides, manual labeling results are unreproducible and may involve human bias as they are highly dependent on the radiologists' experience and subjective decision-making. Automatic or computer-assisted segmentation techniques can overcome these problems by reducing the amount of labor required while providing objective and reproducible outcomes for subsequent tumor diagnosis and monitoring.

Image analysis across multiple MRI modalities is important and beneficial for brain tumor identification. A well-known open-source multi-modal MRI dataset is collected by the Brain Tumor Segmentation Challenge (BraTS). The BraTS challenge is an annual international competition for the objective comparison of state-of-the-art brain tumor segmentation methods [1,2,16]. For each patient case, four 3D MRI modalities are provided, including the native T1-weighted (T1), the post-contrast T1-weighted (T1ce), the T2-weighted (T2), and the T2 Fluid Attenuated Inversion Recovery (FLAIR).

Machine learning methods, especially deep learning algorithms, have been intensively investigated for tumor segmentation in the BraTS challenge since 2014 [5,9,10,12,14,17,21,22,27,30]. Due to their automaticity and accuracy, they have been ranked at top of the competition in recent years. Myronenko [17] earned first place in the BraTS 2018 challenge by training an asymmetrical U-Net with a wider encoder and an additional variational decoder branch providing additional regularisation. Jiang et al. [14], the BraTS 2019 winning team, proposed a two-stage cascaded asymmetrical U-Net similar to Myronenko [17]. The first stage was used to generate a coarse prediction, while the second stage employed a larger network to refine the segmentation result. Isensee et al. [10] used a self-configuring framework named nnU-Net [11] won the BraTS 2020, which automatically adapts the conventional U-Net to a specific dataset with only minor modifications. Wang et al. [27] proposed a modality-pairing learning method using a series of layer connections on parallel branches to capture the complex relationships and rich information between different MRI modalities. A Hybrid High-resolution and Non-local Feature Network (H2NF-Net) was developed by Jia et al. [12], which exploited multi-resolution features while maintaining high-resolution feature representation by using parallel multi-scale convolutional blocks. The non-local self-attention mechanism applied therein allows aggregating contextual information across spatial positions and acquiring long-range dependence.

Although a variety of deep learning approaches have been proposed in the BraTS challenge, some common problems need to be addressed. Firstly, brain tumors can appear in any region of the brain and are extremely diverse in size, shape, and appearance [21]. Second, in most cases, the volume of the brain tumor is relatively small compared to the entire MRI scan, resulting in the data being dominated by background [4]. All these issues reduce the accuracy of brain tumor segmentation, as shown in Fig. 1. Applying long-range self-attention to allow the

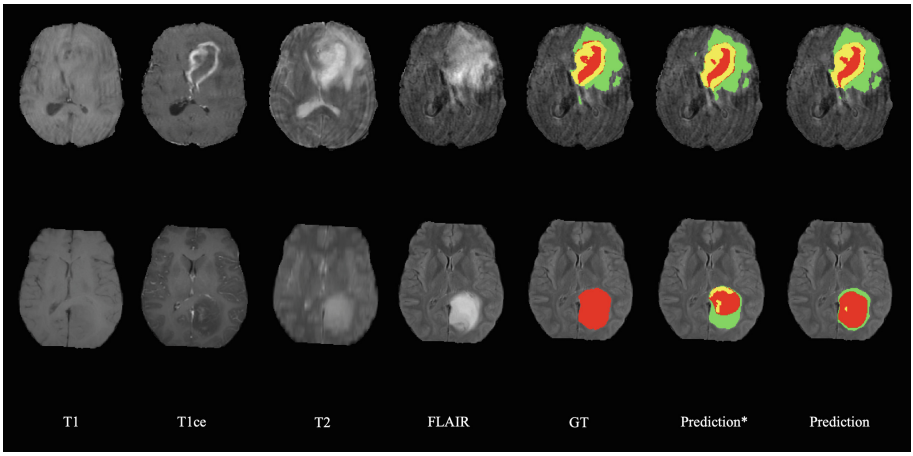

Fig. 1. Representative visual preditions of proposed LKAU-Net on the BraTS 2020 training set. Axial MRI slices in four modalities, ground truth (GT), and predictions are presented from left to right. The labels correspond to the enhancing tumor (yellow), the edema (green), and the necrotic and non-enhancing tumor (red). *: without 3D LK attention module. (Color figure online)

network to selectively learn the truly important tumor-related features is a viable solution [12]. Self-attention originated in Natural Language Processing (NLP) [23]. It is an adaptive selection mechanism based on the input characteristics. Various self-attention mechanisms have been applied in medical image segmentation [3,13,15,19,25,26,29]. Due to their effectiveness in capturing long-range dependencies, they have achieved higher performance compared to conventional Convolutional Neural Networks. However, since self-attention was developed for NLP, it has several drawbacks when it comes to medical image segmentation. First, it interprets images as one-dimensional sequences, omitting the structural information that is necessary for extracting morphological features from medical images. Second, most self-attention studies are based on 2D, since the computational cost with quadratic complexity is prohibitive for 3D scan volumes such as MRI or CT. Third, it ignores channel adaptation, which is crucial in attention mechanisms [7,8,18,24,28].

To address these issues, this paper presents a novel 3D large-kernel attention-based U-Net (LKAU-Net) for automatic segmentation of brain tumors in MRI scans. The 3D LK attention module combines the benefits of self-attention and convolution, such as long-range dependence, spatial adaptability, and local contextual information, while avoiding their drawbacks, including the disregard of channel adaptability. The main contributions of this paper are summarized as follows:

– A novel attention module, dubbed 3D LK attention, is introduced, which balances the advantages of convolution and self-attention, including long-range

dependence, spatial adaptation, and local context, while avoiding their drawbacks, such as high computational cost and neglect of channel adaptation.
– Based on the LK attention, the LKAU-Net is proposed for 3D MRI brain tumor segmentation. LKAU-Net can accurately predict tumor subregions by adaptively increasing the weight of the foreground region while suppressing the weight of the irrelevant background voxels.
– On BraTS 2020 datasets, the LKAU-Net outperformed state-of-the-art methods in delineating all three tumor sub-regions. Besides, the results validated that the new 3D LK attention module can effectively improve the accuracy of automatic MRI brain tumor segmentation.

2 Method

In this section, we presented our method, including a novel 3D LK attention module combining convolution and self-attention, and the LKAU-Net based on it for 3D MRI brain tumor segmentation.

2.1 3D LK Attention

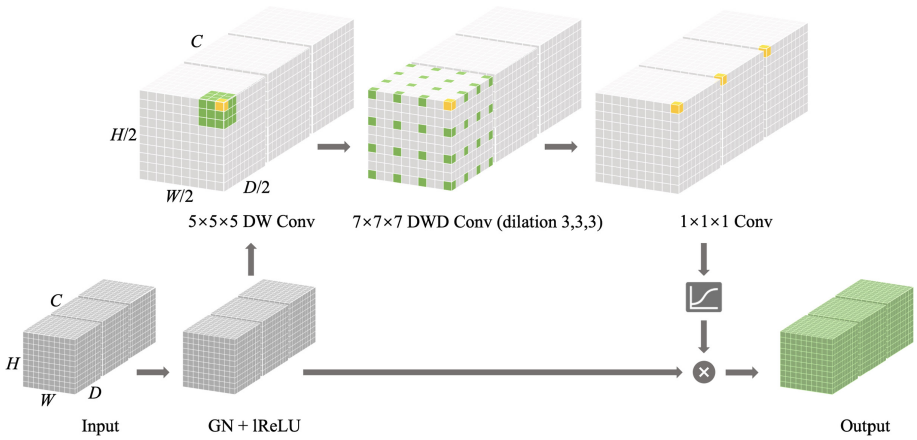

Fig. 2. 3D LK attention module: the feature map after group normalization and leaky ReLU activation is subjected to 3D LK convolution and sigmoid function activation to obtain the attention map, which is multiplied element by element with the feature map before 3D LK convolution to generate the output. The figure illustrates the decomposition of a $21 \times 21 \times 21$ convolution into a $5 \times 5 \times 5$ depth-wise convolution, a $7 \times 7 \times 7$ depth-wise dilated convolution with dilation (3, 3, 3), and a $1 \times 1 \times 1$ convolution. Colored voxels indicate the position of the kernel, and yellow voxels are the kernel centers. (This figure shows only one corner of the feature matrix of the 3D LK convolutional decomposition and ignores the zero-padding)

The attention mechanism is an adaptive selecting process that identifies the discriminative input feature and ignores the background noise. The integration of various attention mechanisms has been shown in many studies to have the potential to improve segmentation performance. The generation of an attention map, which indicates the relative importance of the space, requires the establishment of correlations between different locations. The aforementioned self-attention mechanism can be used to capture long-distance dependence but has several drawbacks outlined. Another strategy is to apply large-kernel convolution to build long-range relationships and create the attention map directly [6,8,18,24,28]. However, this strategy significantly increases the number of parameters and the computational cost.

To address these shortcomings and maximize the benefits of both self-attention and large-kernel convolution, we developed a large-kernel attention module (as shown in Fig. 2). In this module, assuming K as channel number, a $K \times K \times K$ large-kernel convolution was decomposed into a $(2d-1) \times (2d-1) \times (2d-1)$ depth-wise convolution (DW Conv), a $\frac{K}{d} \times \frac{K}{d} \times \frac{K}{d}$ depth-wise dilated convolution (DWD Conv) with dilation (d, d, d) and a $1 \times 1 \times 1$ convolution ($1 \times 1 \times 1$ Conv). Assuming that the input and output have the same dimensions $H \times W \times D \times C$, the number of parameters ($\mathrm{N_{PRM}}$) and floating-point operations (FLOPs) of the original convolution and this decomposition can be computed as follows

$$\mathrm{N_{PRM,O}} = C \times (C \times (K \times K \times K) + 1), \tag{1}$$

$$\mathrm{FLOPs_O} = C \times (C \times (K \times K \times K) + 1) \times H \times W \times D, \tag{2}$$

$$\mathrm{N_{PRM,D}} = C \times ((2d-1) \times (2d-1) \times (2d-1) + \frac{K}{d} \times \frac{K}{d} \times \frac{K}{d} + C + 3), \tag{3}$$

$$\mathrm{FLOPs_D} = C \times ((2d-1) \times (2d-1) \times (2d-1) + \frac{K}{d} \times \frac{K}{d} \times \frac{K}{d} + C + 3) \times H \times W \times D, \tag{4}$$

where the subscripts O and D denote the original convolution and the decomposed convolution, respectively. To find the optimal d for which N_{PRM} is minimal for a specific kernel size K, we let the first-order derivative of Eq. 3 equal 0 and then solved as follows

$$\frac{d}{dd^*}\left(C\left((2d^*-1)^3 + \left(\frac{K}{d^*}\right)^3 + C + 3\right)\right) = 0, \tag{5}$$

$$24d^2 - 24d - \frac{3K^3}{d^4} + 6 = 0. \tag{6}$$

In Eq. 5, the superscript $*$ is used to differentiate dilation d from derivation d. For $K = 21$, solving Eq. 6 using numerical approaches gave an optimum approximation of d of about 3.4159. With the dilation rate of 3, the number of parameters can be remarkably reduced, as detailed in Table 1. We can also

Table 1. Complexity analysis: comparison of the number of parameters N_{PRM} for a $21 \times 21 \times 21$ convolution.

C	$N_{PRM,O}$	$N_{PRM,D}$	$N_{PRM,D}/N_{PRM,O}$
32	9.48 M	16.10 k	0.17%
64	37.94 M	34.24 k	0.09%
128	151.75 M	76.67 k	0.05%
256	606.99 M	186.11 k	0.03%
512	2427.98 M	503.30 k	0.02%

The subscripts O and D denote the original convolution and the proposed decomposed convolution, respectively. C: number of channels.

observe that the decomposition becomes more efficient as the number of channels increases.

The entire 3D LK attention module can be formulated as follows

$$A = \sigma_{\text{sigmoid}} \left(\text{Conv}_{1\times1\times1} \left(\text{Conv}_{\text{DW}} \left(\text{Conv}_{\text{DWD}} \left(\sigma_{\text{lReLU}} \left(\text{GN} \left(\mathit{Input} \right) \right) \right) \right) \right) \right), \quad (7)$$

$$\mathit{Output} = A \otimes \left(\sigma_{\text{lReLU}} \left(\text{GN} \left(\mathit{Input} \right) \right) \right), \quad (8)$$

where A is the attention map, and GN denotes the group normalization. σ_{lReLU} and σ_{sigmoid} refer to leaky ReLU activation function and sigmoid activation function, respectively. The output of the 3D LK attention module is obtained by element-wise multiplication ($\otimes$) of the input features and the attention map. Using the 3D LK attention module described above, we can capture long-range relationships in a deep neural network and generate attention maps with minimal computation complexity and parameters.

2.2 3D LK Attention-Based U-Net

The U-Net [20] has been used as the backbone in many research on medical image analysis. Its ability to capture detailed object patterns using skip-architecture is highly beneficial for fine segmentation of lesions. The 3D LKAU-Net architecture, based on the U-Net, consists of an encoding path for semantic feature extraction and a decoding path for segmentation map inference with skip connections, as shown in Fig. 3.

Encoder. The encoder consists of convolution blocks for six resolution levels. Each block has two convolution layers with a filter size of $3 \times 3 \times 3$, a stride of 1, Group Normalization, and leaky ReLU activation (with slope 0.01). The input of size $4 \times 160 \times 192 \times 128$, where four channels corresponding to the four MRI modalities, is convoluted by additional 32 kernels to form initial 32-channel feature maps. Between two resolution levels, a $3 \times 3 \times 3$ convolution with a stride 2 is used for downsampling the feature maps by 2 and doubling the feature channels simultaneously to a maximum of 512. The bottleneck feature map has a size of $512 \times 5 \times 6 \times 4$, which is 1/32 of the original spatial size.

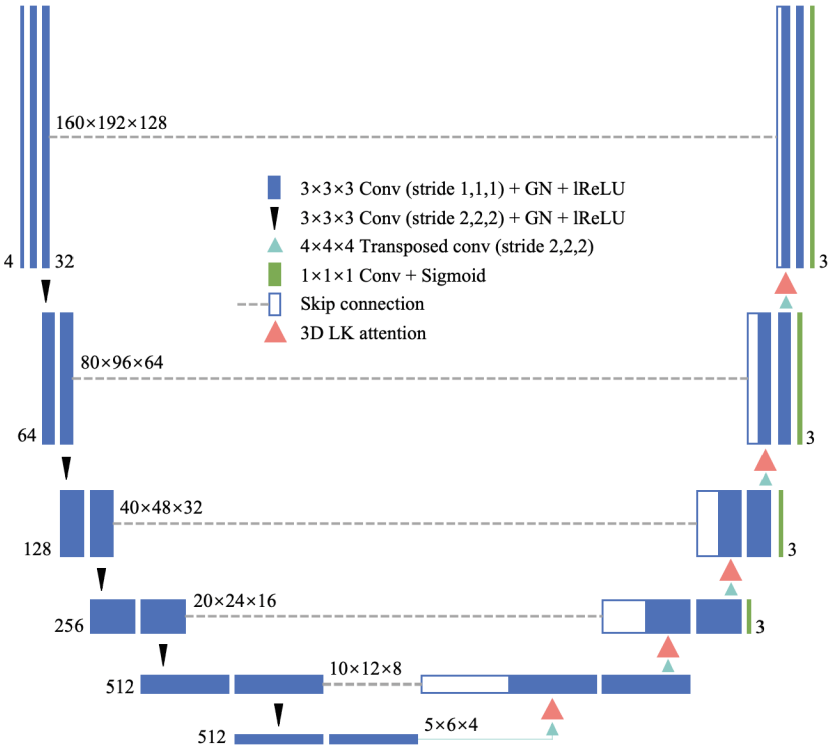

Fig. 3. The network architecture of our proposed LKAU-Net.

3D LK Attention-Based Decoder. The decoder follows the same architecture as the encoder with $4 \times 4 \times 4$ transposed convolution upsampling. The 3D LK attention module is applied to the upsampled feature maps. The output of the 3D LK attention module is then summed back to the feature maps, which are then concatenated with the skip features from of encoder at the same resolution level. The details of the 3D LK attention module at each level are summarized in Table 2. At the final layer of the decoder, a $1 \times 1 \times 1$ convolution is used to compress the feature channels to 3, followed by a sigmoid function to generate prediction probability maps of three overlapping tumor regions: ET, WT, and TC. Additional sigmoid outputs were placed at all resolutions except the two lowest levels to provide deep supervision and increase gradient propagation to previous layers.

3 Experiments

3.1 Data Acquisition

The BraTS 2020 dataset is acquired with different clinical protocols and scanners from multiple institutions. The "ground-truth" segmentation labels are anno-

Table 2. Details of LK attention modules in LKAU-Net

Resolution level	DW Conv		DWD Conv			Equal LK Conv
	Kernel	Padding	Kernel	Dilation	Padding	Kernel
10×12×8	(3, 3, 3)	(1, 1, 1)	(3, 3, 3)	(2, 2, 2)	(2, 2, 2)	(6, 6, 6)
20×24×16	(3, 3, 3)	(1, 1, 1)	(3, 3, 3)	(2, 2, 2)	(2, 2, 2)	(6, 6, 6)
40×48×32	(3, 3, 3)	(1, 1, 1)	(5, 5, 5)	(2, 2, 2)	(4, 4, 4)	(10, 10, 10)
80×96×64	(5, 5, 5)	(2, 2, 2)	(5, 5, 5)	(3, 3, 3)	(6, 6, 6)	(15, 15, 15)
160×192×128	(5, 5, 5)	(2, 2, 2)	(7, 7, 7)	(3, 3, 3)	(9, 9, 9)	(21, 21, 21)

tated by one to four raters and approved by experienced neuro-radiologists, which comprised of the GD-enhancing tumor (ET), the peritumoral edema (ED), and the necrotic and non-enhancing tumor core (NCR and NET). The evaluation of segmentation results is performed on three sub-regions of tumor: GD-enhancing tumor (ET), tumor core (TC = ET + NCR and NET), and whole tumor (WT = ET + NCR and NET + ED)(see Fig. 1). The T1, T1ce, T2, and T2-FLAIR image modalities are co-registered to the same anatomical template with an image size of 240 × 240 × 155, interpolated to the same resolution (1 mm^3), and skull-stripped. Annotations are provided only for the training data (369 cases), while the evaluation of the independent validation dataset (125 cases) should be performed on the online platform (CBICA's IPP[1]).

3.2 Preprocessing and Data Augmentation

Before entering the network, all MRI volumes were cropped to $160 \times 192 \times 128$ to reduce the computation wasted on zero voxels. The input volumes were then preprocessed by intensity normalization. The voxel intensities within each MRI modality were subtracted by the mean and divided by the standard deviation.

To minimize the risk of overfitting and optimize network performance, the following data augmentation techniques have been used: brightness, contrast, Gaussian noise, Gaussian blur, gamma, scaling, rotation, elastic transformation, and mirroring. All augmentations were applied on-the-fly throughout the training process in order to expand the training dataset indefinitely. Moreover, in order to increase the variability of the generated image data, all augmentations were applied randomly based on predetermined probabilities, and most parameters were also drawn randomly from a predetermined range U (detailed in Table 3).

3.3 Training and Optimization

The LKAU-Net was trained on the BraTS 2020 training set (369 cases) with five-fold cross-validation. The objective of the optimization is to minimize the sum of

[1] CBICA's Image Processing Portal (https://ipp.cbica.upenn.edu).

Table 3. Summary of details of 3D LK attention modules in LKAU-Net

Methods	Probability	Range
Brightness	30%	$U(0.7, 1.3)$
Contrast	15%	$U(0.6, 1.4)$
Gaussian noise	15%	Variance $\sigma \sim U(0, 1)$
Gaussian blur	20%	Kernal $\sigma \sim U(0.5, 1.5)$
Gamma augmentation	15%	$\gamma \sim U(0.7, 1.5)$
Scaling	30%	$U(0.65, 1.6)$
Rotation	30%	$U(-30, 30)$
Elastic transform	30%	$\alpha \sim U(5, 10), \sigma = 3\alpha$
Flipping	50%	Along all axes

the binary cross-entropy loss and soft Dice loss at both the final full-resolution output and the lower resolution auxiliary outputs. The adaptive moment estimator (Adam) optimizer optimized the network parameters. Each training run lasted 200 epochs with a batch size of 1 and an initial learning rate of 0.0003. All experiments were implemented with Pytorch 1.10.1 on an NVIDIA GeForce RTX 3090 GPU with 24 GB VRAM.

3.4 Postprocessing

Since the network tended to falsely predict the enhancing tumor when the prediction volume is small, the enhancing tumor region was empirically replaced with necrosis in postprocessing when the volume of predicted ET was less than 500 voxels.

3.5 Evaluation Metrics

Dice score and 95% Hausdorff distance were adopted in the BraTS 2020 challenge. Dice score measures spatial overlapping between the segmentation result and the ground truth annotation, while 95% Hausdorff distance (HD95) measures the 95th percentile of maximum distances between two boundaries. We used the Dice score in five-fold cross-validation on the training set (369 cases) and used both the Dice and HD95 scores from the official online platform to evaluate and compare the final performance of LKAU-Net on the independent validation set (125 cases).

4 Results and Discussion

We trained three network configurations of LKAU-Net, with the 3D LK attention module, without the 3D LK attention module, and with the CBAM [28] on the BraTS training set with five-fold cross-validation. This provided us with

performance estimates for 369 training cases, allowing us to compare the different network configurations internally and also externally for state-of-the-art approaches, as shown in Table 4. The increase in Dice scores from the 3D LK attention module can be seen by comparing three different network configurations of LKAU-Net. Two segmentation results are also compared visually in Fig. 1, which proves the benefit of the 3D LK attention module. Compared to the best BraTS-specific nnU-Net that was also trained with five-fold cross-validation, LKAU-Net has outperformed all but ET in Dice scores, including the average score.

Table 4. Quantitative results of proposed LKAU-Net on BraTS 2020 training set compared to state-of-the-art methods.

Methods	Dice			
	ET	WT	TC	Mean
nnU-Net baseline	80.83	91.60	87.23	86.55
nnU-Net best	**80.94**	91.60	87.51	86.68
Proposed LKAU-Net*	78.52	92.66	88.74	86.64
Proposed LKAU-Net**	78.15	92.72	88.67	86.66
Proposed LKAU-Net	78.81	**92.81**	**88.99**	**86.87**

Bold numbers: best results. *: without 3D LK attention module. **: with CBAM.

The final segmentation performance of the proposed LKAU-Net was evaluated and compared with state-of-the-art methods on the independent BraTS 2020 validation set (125 cases), with the results shown in Table 5. All segmentation results were evaluated by the Dice score and 95% Hausdorff distance (HD95) and directly obtained from the official online platform (CBICA's IPP).

Table 5. Quantitative results of proposed LKAU-Net on independent BraTS 2020 validation set compared to state-of-the-art methods.

Methods	Dice				HD95			
	ET	WT	TC	Mean	ET	WT	TC	Mean
Myronenko	64.77	84.31	72.61	73.90	41.35	13.85	18.57	24.59
Wang et al.	78.70	90.80	85.60	85.03	35.01	4.71	5.70	15.14
H2NF-Net	78.75	91.29	85.46	85.17	26.58	4.18	**4.97**	**11.91**
nnU-Net baseline	77.67	90.60	84.26	84.18	35.10	4.89	5.91	15.30
nnU-Net best	**79.85**	91.18	85.71	**85.58**	26.41	**3.71**	5.64	11.92
Proposed LKAU-Net*	78.94	91.18	84.99	85.04	29.14	4.77	6.01	13.31
Proposed LKAU-Net	79.01	**91.31**	**85.75**	85.36	**26.27**	4.56	5.87	12.23

Bold numbers: best results. *: without 3D LK attention module.

Quantitative results showed that the proposed LKAU-Net outperformed state-of-the-art methods, including the best nnU-Net model from the BraTS 2020 champion team, in segmenting all three tumor sub-regions. The proposed method achieved the highest Dice score in WT and TC segmentation and the lowest HD95 score in ET segmentation. However, due to the previously mentioned dominance of the best nnU-Net model in ET Dice, our network dropped to second place with a slight difference of 0.22 in the average Dice of the validation set. The LKAU-Net, on the other hand, performed exceptionally well on the HD95 score for ET, which may be attributed to the 3D LK attention module emphasizing the features of the correct tumor region and thus reducing incorrect and scattered ET predictions.

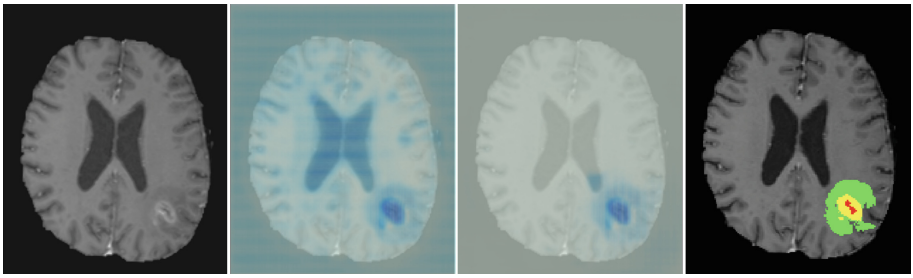

Fig. 4. The representative visual effect of the LK attention module. The first plot is the T1ce input. The second plot shows an upsampled feature map at the finest resolution in the LKAU-Net decoder, while the following plot shows the feature map after applying the 3D LK attention module. The last plot identifies the ground truth labels.

By comparing the evaluation results of LKAU-Nets with and without the 3D LK attention module, the performance improvement due to the presence of the 3D LK attention module can be seen. The performance improvement is more evident for TC and ET, as shown in Table 6. In order to validate the increase of the evaluation metrics, paired t-Tests were done, and the p-values for each category are presented in Table 6. The improvement due to the 3D LK attention module on WT Dice, TC Dice, Mean Dice, WT HD95, and TC HD95 was statistically validated. However, the paired t-Test could not verify the changes on ET due to the high penalty of Dice = 0 and HD95 = 373.13 set by BraTS 2020 for the False Positives of ET. This improvement validated the effectiveness of adaptive feature selection of the 3D LK attention module, which is visualized in Fig. 4.

Table 6. Improvement in quantitative results of LKAU-Net due to the 3D LK attention module.

	Dice				HD95			
	ET	WT	TC	Mean	ET	WT	TC	Mean
Mean results*	78.94	91.18	84.99	85.04	29.14	4.77	6.01	13.31
Mean results	79.01	91.31	85.75	85.36	26.27	4.56	5.87	12.23
Improvement	0.07	0.13	0.76	0.32	−2.87	−0.21	−0.14	−1.07
p-value	0.338	0.017	0.013	0.018	0.159	0.030	0.041	0.130

*: without 3D LK attention module.

In addition to the comparison of numbers of parameters in Table 1, we compared the running times of the original $21 \times 21 \times 21$ convolution and the proposed decomposed LK convolution, as well as the running times of LKAU-Net without and with the 3D LK attention module, respectively, in Table 7. It is worth noting that the LKAU-Net with original LK convolutions was even unable to be implemented on the current GPU of 24 GB VRAM. We recorded the time required to apply different LK convolutions to each batch including the forward and backward propagation. We can observe that the decomposition reduces the time required for the original LK convolution by about 99%, which proved its optimization on complexity. However, even so, the addition of the 3D LK attention module brought an additional training time of nearly 180 s per epoch to LKAU-Net, which is nearly double the original time. Therefore, a lighter architecture can be developed for limited computational resources and research time.

Table 7. Running time comparison: per batch for convolutions and per training epoch for LKAU-Nets.

	Running time (s)
$21 \times 21 \times 21$ Conv	9.566 ± 0.024
$21 \times 21 \times 21$ D Conv	0.088 ± 0.001
LKAU-Net*	199.873 ± 0.597
LKAU-Net	379.243 ± 1.407

The Conv and D Conv denote the original convolution and the proposed decomposed convolution, respectively. *: without 3D LK attention module.

Furthermore, we found another limitation of our approach by analyzing the segmentation results. Figure 1 shows two representative examples of predictions from five-fold cross-validation on the BraTS 2020 training set. In the first case, the network delineated all tumor sub-regions with high accuracy, despite the presence of slight artifacts. In the second case, the WT region was accurately segmented while the TC was not, which might be due to blurring in the T2

volume. This demonstrates the importance of data integrity for the accurate segmentation of medical images. To further improve the network's robustness, these exceptional cases can be covered by more diverse data acquisition or more realistic data augmentation.

5 Conclusion

In this paper, we present a 3D LK attention-based U-Net for MRI brain tumor segmentation that has outperformed state-of-the-art methods on the BraTS 2020 dataset. The 3D LK attention module combines the advantages of convolution and self-attention, exploiting local environment information, long-distance dependence, and spatial and channel adaptation. We integrated this novel attention module into the decoder of U-Net, enabling the network to focus on decisive tumor-related features. The evaluation results on the BraTS 2020 dataset showed that the 3D LK attention module could improve predictions for all three tumor sub-regions, particularly for TC and ET. As shown in Fig. 4, the 3D LK attention module also proved to be effective in adaptively selecting discriminative features and suppressing background noises. Guided by this 3D LK attention, our proposed network achieves state-of-the-art performance, with average Dice scores of 79.01%, 91.31%, 85.75%, and 95% Hausdorff distances of 26.27, 4.56, and 5.87 in segmenting the enhancing tumor, whole tumor, and tumor core, respectively.

References

1. Bakas, S., et al.: Advancing The Cancer Genome Atlas glioma MRI collections with expert segmentation labels and radiomic features. Sci. Data **4**(1), 170117 (2017). https://doi.org/10.1038/sdata.2017.117
2. Bakas, S., Reyes, M., Jakab, A., Menze, B.: Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the BRATS Challenge. arXiv:1811.02629 [cs, stat], April 2019
3. Chen, J., et al.: TransUNet: transformers make strong encoders for medical image segmentation. arXiv:2102.04306 [cs], February 2021
4. DSouza, A.M., Chen, L., Wu, Y., Abidin, A.Z., Xu, C.: MRI tumor segmentation with densely connected 3D CNN. In: Angelini, E.D., Landman, B.A. (eds.) Medical Imaging 2018: Image Processing, p. 50. SPIE, Houston, United States, March 2018. https://doi.org/10.1117/12.2293394
5. Guan, X., et al.: 3D AGSE-VNet: an automatic brain tumor MRI data segmentation framework. BMC Med. Imaging **22**(1), 6 (2022). https://doi.org/10.1186/s12880-021-00728-8
6. Guo, M.H., Lu, C.Z., Liu, Z.N., Cheng, M.M., Hu, S.M.: Visual attention network. arXiv:2202.09741 [cs], March 2022
7. Guo, M.H., et al.: Attention mechanisms in computer vision: a survey. Comput. Visual Media (2022). https://doi.org/10.1007/s41095-022-0271-y

8. Hu, J., Shen, L., Albanie, S., Sun, G., Vedaldi, A.: Gather-Excite: exploiting feature context in convolutional neural networks. In: Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 31. Curran Associates, Inc. (2018)
9. Huang, H., et al.: A deep multi-task learning framework for brain tumor segmentation. Front. Oncol. **11**, 690244 (2021) https://doi.org/10.3389/fonc.2021.690244. https://www.frontiersin.org/articles/10.3389/fonc.2021.690244/full
10. Isensee, F., Jäger, P.F., Full, P.M., Vollmuth, P., Maier-Hein, K.H.: nnU-Net for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) BrainLes 2020. LNCS, vol. 12659, pp. 118–132. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72087-2_11
11. Isensee, F., Jäger, P.F., Kohl, S.A.A., Petersen, J., Maier-Hein, K.H.: Automated design of deep learning methods for biomedical image segmentation. Nature Methods **18**(2), 203–211 (2021). https://doi.org/10.1038/s41592-020-01008-z, arXiv: 1904.08128
12. Jia, H., Cai, W., Huang, H., Xia, Y.: H^2NF-Net for brain tumor segmentation using multimodal MR imaging: 2nd place solution to BraTS challenge 2020 segmentation task. In: Crimi, A., Bakas, S. (eds.) BrainLes 2020. LNCS, vol. 12659, pp. 58–68. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72087-2_6
13. Jia, Q., Shu, H.: BiTr-Unet: a CNN-Transformer Combined Network for MRI Brain Tumor Segmentation. arXiv:2109.12271 [cs, eess], December 2021
14. Jiang, Z., Ding, C., Liu, M., Tao, D.: Two-stage cascaded U-Net: 1st place solution to BraTS challenge 2019 segmentation task. In: Crimi, A., Bakas, S. (eds.) BrainLes 2019. LNCS, vol. 11992, pp. 231–241. Springer, Cham (2020). https://doi.org/10.1007/978-3-030-46640-4_22
15. Karimi, D., Vasylechko, S.D., Gholipour, A.: Convolution-free medical image segmentation using transformers. In: de Bruijne, M., Cattin, P.C., Cotin, S., Padoy, N., Speidel, S., Zheng, Y., Essert, C. (eds.) MICCAI 2021. LNCS, vol. 12901, pp. 78–88. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-87193-2_8
16. Menze, B.H., Jakab, A., Bauer, S., Van Leemput, K.: The Multimodal Brain Tumor Image Segmentation Benchmark (BRATS). IEEE Trans. Med. Imaging **34**(10), 1993–2024 (2015). https://doi.org/10.1109/TMI.2014.2377694. http://ieeexplore.ieee.org/document/6975210/
17. Myronenko, A.: 3D MRI brain tumor segmentation using autoencoder regularization. In: Crimi, A., Bakas, S., Kuijf, H., Keyvan, F., Reyes, M., van Walsum, T. (eds.) BrainLes 2018. LNCS, vol. 11384, pp. 311–320. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-11726-9_28
18. Park, J., Woo, S., Lee, J.Y., Kweon, I.S.: BAM: Bottleneck Attention Module. arXiv:1807.06514 [cs], July 2018
19. Peiris, H., Hayat, M., Chen, Z., Egan, G., Harandi, M.: A Volumetric Transformer for Accurate 3D Tumor Segmentation. arXiv:2111.13300 [cs, eess], November 2021
20. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28

21. Soltaninejad, M., et al.: Automated brain tumour detection and segmentation using superpixel-based extremely randomized trees in FLAIR MRI. Int. J. Comput. Assisted Radiol. Surg. **12**(2), 183–203 (2017) https://doi.org/10.1007/s11548-016-1483-3
22. Soltaninejad, M., et al.: Supervised learning based multimodal MRI brain tumour segmentation using texture features from supervoxels. Comput. Methods Programs Biomed. **157**, 69–84 (2018). https://doi.org/10.1016/j.cmpb.2018.01.003, https://linkinghub.elsevier.com/retrieve/pii/S016926071731355X
23. Vaswani, A., et al.: Attention is All you Need. In: Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R. (eds.) Advances in Neural Information Processing Systems, vol. 30. Curran Associates, Inc. (2017)
24. Wang, F., et al.: Residual attention network for image classification. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 6450–6458. IEEE, Honolulu, HI, USA, July 2017. https://doi.org/10.1109/CVPR.2017.683. https://ieeexplore.ieee.org/document/8100166/
25. Wang, H., et al.: Mixed Transformer U-Net For Medical Image Segmentation. arXiv:2111.04734 [cs, eess], November 2021
26. Wang, W., Chen, C., Ding, M., Li, J., Yu, H., Zha, S.: TransBTS: multimodal brain tumor segmentation using transformer. arXiv:2103.04430 [cs], June 2021
27. Wang, Y., et al.: Modality-pairing learning for brain tumor segmentation. In: Crimi, A., Bakas, S. (eds.) BrainLes 2020. LNCS, vol. 12658, pp. 230–240. Springer, Cham (2021). https://doi.org/10.1007/978-3-030-72084-1_21
28. Woo, S., Park, J., Lee, J.-Y., Kweon, I.S.: CBAM: convolutional block attention module. In: Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y. (eds.) ECCV 2018. LNCS, vol. 11211, pp. 3–19. Springer, Cham (2018). https://doi.org/10.1007/978-3-030-01234-2_1
29. Xie, Y., Zhang, J., Shen, C., Xia, Y.: CoTr: efficiently bridging CNN and transformer for 3D medical image segmentation. arXiv:2103.03024 [cs], March 2021
30. Zhang, W., et al.: ME-Net: multi-encoder net framework for brain tumor segmentation. Int. J. Imaging Syst. Technol. **31**(4), 1834–1848 (2021). https://doi.org/10.1002/ima.22571