A REAL-TIME MRI TUMOUR SEGMENTATION METHOD BASED ON LIGHTWEIGHT NETWORK FOR IMAGING ROBOTIC SYSTEMS

Chaofan Du* and Peter Xiaoping Liu**

Abstract

Medical imaging robots typically use technologies, such as X-ray, magnetic resonance imaging (MRI), and computed tomography (CT), to generate images of the human body interior. These generated images are complex and contain a large amount of noise and interference, which requires high-precision and real-time fast image analysis algorithms to extract significant information, including tumour area, tumour location, organ and tissue, and blood vessel information. This paper proposes a novel lightweight neural network to perform tumour segmentation in brain MRI images, which could realize the high-accuracy and fast execution. To meet the real-time requirements, a lightweight module based on channel attention mechanism is presented, which constitutes an encoder–decoder architecture for the segmentation task. To enrich the feature map information, this paper designs a spatial attention mechanism to concatenate the output feature maps of the encoder and decoder correspondingly, which could realize the better fusion of high-level and low-level semantic features extracted by the network. The comparison experiments and ablation studies are conducted to improve the effectiveness of the proposed model, which could represent a higher performance. The computational cost of the proposed model shows the possibility of a real-time implementation.

Key Words

Medical imaging robots, MRI images, brain tumour segmentation, lightweight network

1. Introduction

Medical robotics is a new interdisciplinary research field that integrates multiple disciplines, such as medicine, biomechanics, material science, computer vision, and robotics. Medical robotics has important research value as a hot area of research. As usual, medical robotics could be grouped into neurosurgical robots [1], [2], rehabilitation robot [3], [4], auxiliary diagnostic robot [5], [6] and medical imaging robot [7], [8]. The medical imaging robot could be understood as robot-assisted to acquire a medical image and perform image analysis tasks. For example, robot-assisted endoscopic camera imaging is used in endoscopic surgery [9], and robot-assisted medical image registration [10] during the surgery providing valuable information to the surgeon. In addition, robot-assisted medical image analysis can be used for lesion segmentation and diagnosis tasks to assist physicians in making medical decisions. Lesion segmentation is essential in medical imaging analysis. The significance of medical image leision segmentation lies in the automatic and accurate separation of lesions from normal tissues, thereby helping physicians to better locate the site of the disease and carry out diagnosis and treatment. Robot-assisted medical image segmentation tasks require high accuracy and real-time performance simultaneously. Therefore, when developing image segmentation algorithms, it is necessary to meet the requirements for both accuracy and speed simultaneously.

Glioblastoma is a common primary brain tumour, accounting for 28% of all central nervous system tumors and 80% of all malignant brain tumors. Malignant glioblastomas are more invasive and have a poor prognosis, and can lead to patient death. Utilizing a robot-assisted medical imaging analysis system to diagnose and treat brain tumors in a timely and accurate manner is crucial in reducing mortality rates and improving patient quality of life. Segmentation of glioblastomas in MRI images provides valuable assistance in treatment planning and disease progression monitoring for tumour patients.

Deep learning has been growing up rapidly, which can be used in image analysis. Although deep learning techniques have shown great potential in natural image segmentation, applying them to medical image data segmentation still faces significant challenges. Medical

* School of Mechanical and Electronic Control Engineering, Beijing Jiaotong University, China; e-mail: 18116005@bjtu.edu.cn
** Department of Systems and Computer Engineering, Carleton University, Canada; e-mail: xpliu@sce.carleton.ca
Corresponding author: Chaofan Du; Peter Xiaoping Liu

Recommended by Jason Gu
(DOI: 10.2316/J.2024.206-1054)

datasets have class imbalance issues, and the number of samples is limited. Tumour features also have some irregularities, making automatic segmentation algorithms still face many challenges. In the brain MRI image dataset, the tumour area is the foreground information that could get more attention, while other brain tissues occupy a higher proportion of pixels in the MRI image. Therefore, the tumour segmentation problem also needs to focus on the issue of pixel imbalance. Overall, the automatic segmentation algorithm for brain MRI tumors based on robot-assisted medical imaging systems needs to face challenges including pixel imbalance, segmentation accuracy, and segmentation speed.

CNN can obtain high-dimensional features in images and share convolution kernels during computation to reduce the number of parameters in the network. It has achieved very good performance in medical image segmentation. Havaei *et al.* [11] used two training processes to construct a dual-path CNN structure that can simultaneously use large and small image patches, establishing dependencies between pixel labels. Razzak *et al.* [12] designed a network based on two sets of CNN structures that can simultaneously capture local and global features, which overcomes the instability and overfitting problems of the network by obtaining the variance of the CNN model. Mohseni Salehi *et al.* [13] proposed a network based on CNN that can automatically combine context information. It learns local and global features using 2D image blocks of different sizes and pixels with three parallel 2D convolution paths in the axial, coronal, and sagittal directions. It can also learn 3D image information without complex calculations.

Semantic segmentation models based on fully convolutional networks (FCNs) [14] have been widely used in medical image segmentation. Based on FCN, the popular U-Net network [15] builds a completely symmetrical convolutional network with U-shaped structure and skip connections, achieving high-dimensional and low-dimensional feature fusion and realizing excellent segmentation results. This method has been widely used in medical image segmentation, and a large number of improved models have been proposed based on U-Net architecture. Zahangir Alom *et al.* [16] proposed RU-Net and R2U-Net, recurrent convolutional neural networks based on the U-Net framework, combining the advantages of U-Net, ResNet, and RCNN. The Res-UNet [17] network turns the submodules into residual blocks and adds weighted attention mechanisms based on the original U-Net, enabling the model to better learn target features. To improve the capability of fast execution, some lightweight models are proposed to perform brain tumor segmentation. Tarasiewicz *et al.* [18] proposed a lightweight U-Net called Skinny which is trained over all image planes, including axial, coronal, and sagittal. Wang *et al.* [19] presented a lightweight hierarchical convolutional network called LHC-Net, in which a multi-scale strategy was adopted to replace the ordinary three-dimensional convolution with the hierarchical convolution of residual-like connections. The multi-scale feature extraction capability was improved, and parameters and computing resources were greatly reduced.

Although a considerable amount of literature has been reported to solve this problem, it still face many challenges. Complex network architectures can learn richer semantic features, but the speed of their implementation cannot meet the real-time requirements. In the existing lightweight network, some lightweight bottlenecks are usually added in the whole architecture which could effectively improve the inference speed of the network, but the accuracy decreases. In addition, the pixel imbalanced problem exists in the brain tumor segmentation task, which can also reduce the segmentation accuracy. Based on the abovementioned problems, a novel lightweight network utilizing attention mechanism is built to perform brain tumor segmentation. A lightweight component is presented to form an encoder–decoder network. The whole algorithm implementation architecture is shown in Fig. 1.

In this study, a novel lightweight component is proposed to form a novel encoder–decoder lightweight segmentation network, which could realize execution of brain tumor segmentation at a high speed and a high accuracy simultaneously. To alleviate the pixel imbalanced problem, a spatial attention module is designed to concatenate the encoder map and decoder map, which could avoid information loss. The contribution of this work is summarised as follows:

1) A lightweight component is presented, adopting a channel attention mechanism, which enables lightweight components to pay more attention to the key features of the segmentation task by enhancing the expressibility of different channels in the feature map.
2) A novel lightweight encoder–decoder architecture is designed utilizing lightweight component, which could capture image features at different scales and levels.
3) A spatial attention module is designed to connect the feature map of encoders and decoders at different levels, and to adjust the weight of the feature map through the attention mechanism, so that the network can better focus on the key information in the map and alleviate the problem of information loss caused by imbalanced pixels.

2. Methods

In this section, a lightweight component inspired by E-Net [20] is proposed. The E-Net architecture is composed of an initial block and bottleneck module. The initial block concatenates the convolution operation and maxpooling operation. Each bottleneck block has two branches. One branch includes three convolutional layers and the other branch includes maxpooling layer and padding layer. The initial block architecture and bottleneck architecture are shown in Fig. 2(a) and (b), respectively.

The bottleneck module is used to construct the encoder branch architecture, which preserves the index of maximum value in the max pooling operation to avoid information loss in downsampling. Meanwhile, the channel attention mechanism is introduced in the bottleneck module, which could focus on the importance of channels in the feature

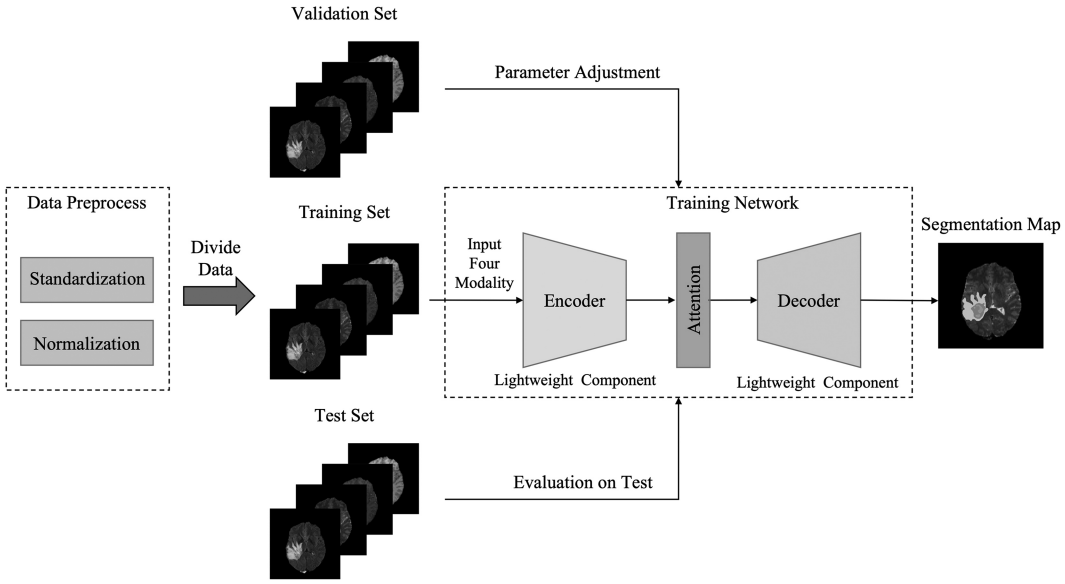

Figure 1. The implementation of brain tumour segmentation algorithm.

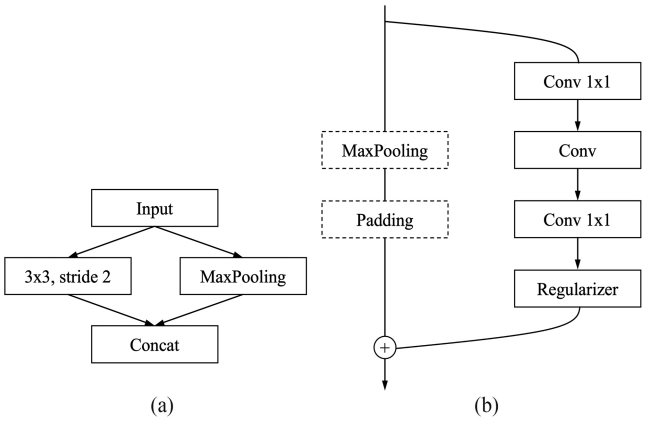

Figure 2. The initial block and bottleneck block in E-Net: (a) initial block and (b) bottleneck block.

map. The bottleneck module and upsampling operation are utilized to construct the decoder branch architecture. Inspired by E-Net [20], the presented encoder branch and the presented decoder branch are asymmetric, which could reduce the parameters of the proposed network. In addition, a novel attention module is proposed to learn to focus on target structures of varying shapes and sizes, which can be easily integrated into encoder–decoder architecture.

2.1 Encoder–Decoder Architecture

The unique structure of the encoder–decoder enables to realize effective utilization of contextual information. The encoder branch is responsible for extracting high-level semantic features from the input data, while the decoder branch progressively recovers the resolution to generate detailed information. This utilization of contextual information helps improve the accuracy of segmentation tasks. Inspired by it, the encoder–decoder architecture composed of lightweight module is proposed which is shown in Fig. 3. Each green block represents a lightweight module. The arrows in the diagram indicate the connectivity pattern of the encoder–decoder architecture. Additionally, the spatial attention mechanism is annotated, showing its position in the network structure. It connects the feature map of encoder branch and decoder branch, serving as a skip connection. The proposed structure also employs an asymmetric form of encoder and decoder branch to reduce network parameters.

2.2 Lightweight Component

The proposed lightweight block consists of a main branch and an auxiliary branch. The main branch performs max pooling operation and padding operation and then connected to the feature map generated by the auxiliary branch. The auxiliary branch consists of three main convolutional operations. The first and third convolutions adopt 1×1 kernel to compress channels, reduce data dimensionality, and decrease computational complexity. The second convolutional operation varies in different lightweight block. After the first downsampling, it performs a 3×3 convolution operation, while after the second and third downsampling, it employs dilated convolution to expand the receptive field. By using dilated convolutions with different dilation rates, the network can perform multi-scale feature extraction, capturing fine-grained and coarse-grained features simultaneously. After the convolution operation in the auxiliary branch, a batch normalization layer is added to accelerate the training process and address the internal covariate shift problem. The computation process of the proposed lightweight block is illustrated in Fig. 4.

After the concatenation of the feature maps generated by two branches, the channel attention mechanism is introduced to adaptively adjust the importance of channel features by learning channel weights. Inspired by

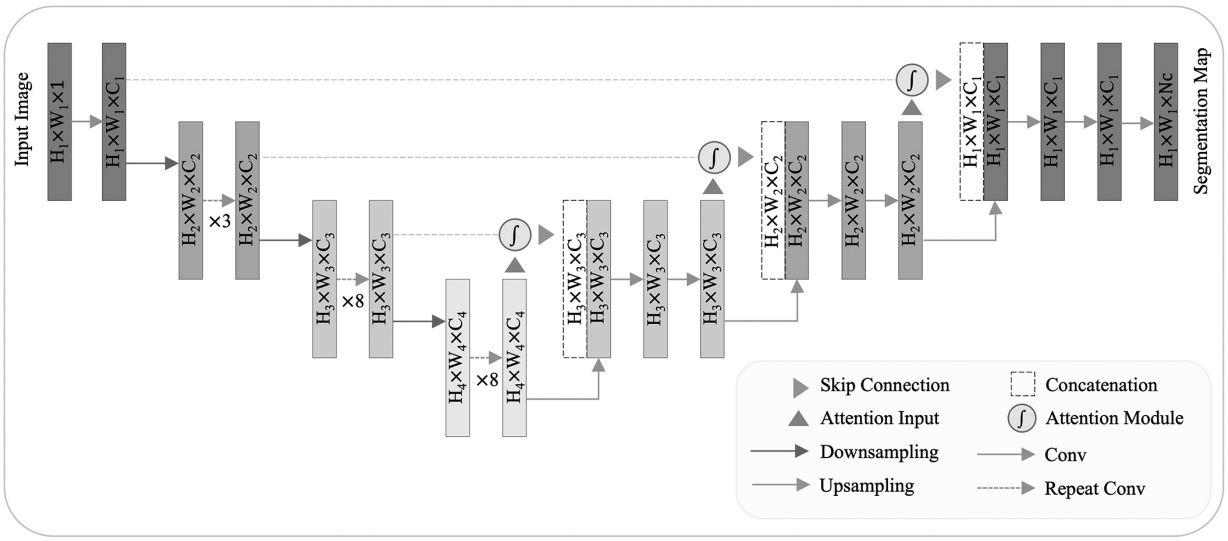

Figure 3. The encoder–decoder architecture of the lightweight network.

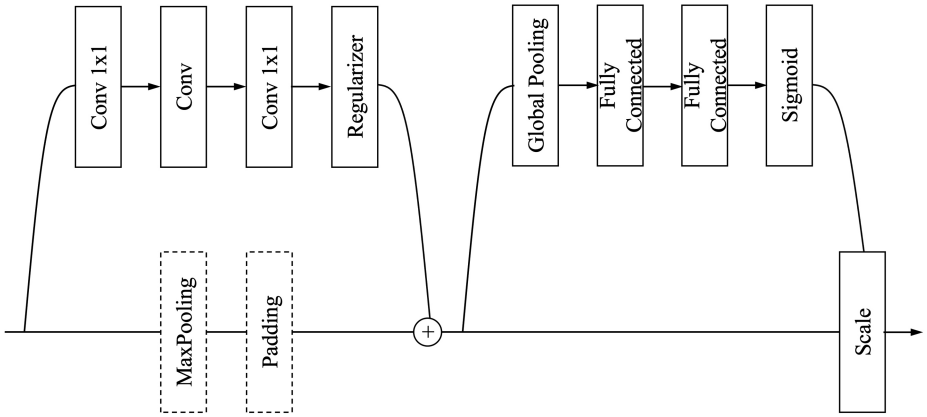

Figure 4. The computation diagram of the presented lightweight module.

SE-Net [21], the auxiliary branch continues to perform global average pooling operation and then followed by two fully connected layers and sigmoid activation, calculating the weights for each channel. The computational weights are multiplied by the original feature map in the main branch, which helps reduce the impact of redundant features and enhance the generalisation capacity of the network.

2.3 Spatial Attention Mechanism

The output of spatial attention mechanism is the element-wise multiplication of input feature-maps and attention coefficients: $\widehat{x}_i^l = x_i^l \cdot \alpha_i^l$. $\alpha_i^l \in [0,1]$ represents attention coefficients which detect prominent regions in the image and selectively remove feature responses, retaining only the activations that are pertinent to the specific task. A singular scalar attention value is calculated for each pixel vector $x_i^l \in \mathbb{R}^{F_l}$, with F_l representing the number of feature-maps in layer l. In this paper, to determine focus regions, the decoder vector x_{d_i} and the encoder vector x_{e_i} are employed for each pixel i. The formulation of attention mechanism is as follows:

$$q_{\text{att1}}^l = \sigma_1\left(W_e x_{\text{ei}}^l + W_d x_{\text{di}} + b\right) \tag{1}$$

$$q_{\text{att2}}^l = A_{\text{avg}} \cdot q_{\text{att1}}^l \circ A_{\max} \cdot q_{\text{att1}}^l \tag{2}$$

$$\alpha_i^l = \sigma_2\left(q_{\text{att2}}^l\left(x_{\text{ei}}^l,\ x_{\text{di}};\ \theta_{\text{att}}\right)\right) \tag{3}$$

Where $\sigma_1(x_i)$ corresponds to the ReLU activation function, $\sigma_2(x_i)$ corresponds to the sigmoid activation function. The graphical representation of the presented attention mechanism is depicted in Fig. 5. The main computation process of attention is formulated as two parts: q_{att1}^l and q_{att2}^l. q_{att1}^l corresponds to the left side of the vertical dotted line and the other represents the right side. A set of parameters Θ_{att} contains: linear transformations $W_e \in \mathbb{R}^{F_l \times F_{\text{int}}}$, $W_d \in \mathbb{R}^{F_d \times F_{\text{int}}}$ and bias terms $b \in \mathbb{R}^{F_{\text{int}}}$. The implementation of linear transformation is by 1×1×1 convolution computed for the input tensors. In (2), A_{avg} and $A_{\max}$ illustrate the average pooling operation and the max pooling operation. Equation (3) corresponds to the final computation result of the spatial attention mechanism.

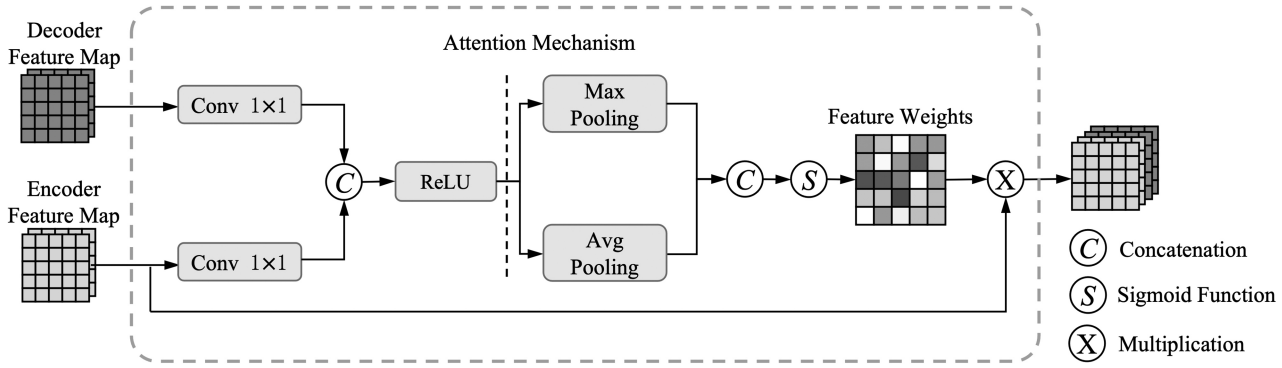

Figure 5. The architecture of the proposed spatial attention module.

The presented spatial attention module is incorporated into the lightweight encoder–decoder network to highlight salient features that are passed through the skip connection, which is shown in Fig. 3. Information derived from a coarse scale is employed in attention mechanism to resolve ambiguity and filter out irrelevant or noisy responses in skip connections. This operation is performed immediately prior to the concatenation process, aiming to merge only the relevant activations.

3. Experiments and Results

The experiments are conducted to validate the effectiveness of the proposed lightweight segmentation network and comparing with existing state-of-the-art methods. The ablation study is also performed to assess the importance of each component in the proposed network. The proposed network is trained using BraTS brain MRI dataset, which is deployed on a high-performance computing server equipped with Inter E5-2620 V4 processors with 256G memory. The computing server is also equipped with NVIDIA RTX TITAN and 24GB memory.

3.1 Datasets

The experiment of brain tumour segmentation uses a dataset of multimodal MRI volumes. Each voxel in the dataset is associated with multiple image modalities, including T1, T1c, T2, and FLAIR. The segmentation task faces the challenge of data imbalance, as there are significant variations among different types of tumors. Healthy voxels account for 98% of the total voxels, while pathological voxels, necrosis, edema, non-enhanced tumour, and enhanced tumour account for only 0.18%, 1.1%, 0.12%, and 0.38%, respectively. Besides, the medical imaging dataset contains numerous challenges due to their complex characteristics, primarily because the image acquisition process is expensive and sensitive compared to real-life data. Consequently, the collection of medical images is conducted under controlled conditions to ensure a more predictable distribution of data.

The proposed network is evaluated on the BraTS 2018 training set and leaderboard dataset. The BraTS 2018 training dataset consists of 210 cases of high-grade glioma (HGG) and 75 cases of low-grade glioma (LGG). Each case is manually annotated by one to four raters, and all segmentation results undergo expert review. The MRI images in the BraTS dataset are obtained from 19 institutions, using different protocols, magnetic field intensity, and MRI scanners. The training set provided by BraTS includes four MRI modalities (T1, T1c, T2, and FLAIR) and expert manual segmentation labels. Each MRI sequence consists of 155 slices with a size of 240×240 pixels. The label data provides pixel-level class annotations, including the background region, non-enhancing and necrotic region, edema region, and enhancing region. Additionally, the official dataset includes a test set without labels, which can be used for further evaluation of segmentation algorithms.

3.2 Evaluation Metrics

In the evaluation process, three key evaluation metrics are primarily performed on the following regions: the whole tumour (WT) region including the non-enhancing and necrotic region, edema region, and enhancing region, the tumour core (TC) region including the non-enhancing and necrotic region and enhancing region, the enhancing tumour (ET) region which evaluates the segmentation results of the enhancing region.

The segmentation evaluation metrics in this study include the dice similarity coefficient (DSC), sensitivity, specificity, precision, F1-Score. The DSC is a metric commonly used to measure two sets and is frequently employed to evaluate the performance of medical image segmentation models. Its calculation is as follows:

$$DSC = \frac{2 \times |A \cap B|}{|A| + |B|} \tag{4}$$

where A and B represent the set of pixels in the prediction results and the ground truth, respectively. $|A| + |B|$ represents the total number of non-zero pixels in set A and B, and $|A \cap B|$ represents the intersection of pixels between the prediction results and the ground truth. A higher DSC index indicates better performance of the model. The DSC is commonly used to evaluate the performance of models in binary classification tasks and image segmentation tasks.

During the experimental process, the DSC can also be calculated using the following:

$$DSC = \frac{2TP}{2TP + FN + FP} \tag{5}$$

where TP represents the number of true positive pixels, FP represents the number of false positive pixels, and FN represents the number of false negative pixels. In the evaluation of medical automatic diagnosis tasks, sensitivity refers to the ability of a system to correctly identify positive cases. It is formulated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{6}$$

Specificity refers to the ability of a system to correctly identify negative cases. It is formulated as follows:

$$\text{Specificity} = \frac{TN}{TN + FP} \tag{7}$$

where TN represents the number of true negative pixels. Precision refers to the ability of a system to identify correct positive cases in all positive cases. It is formulated as follows:

$$\text{Precision} = \frac{TN}{TN + FP} \tag{8}$$

In medical diagnostic systems, F1-Score is a common evaluation metric that combines sensitivity and precision to assess the overall performance of the model. A higher F1-Score indicates a better performance of the model in identifying true positive cases.

$$F1 - \text{Score} = 2 \times \frac{\text{Sensitivity} \times \text{Precision}}{\text{Sensitivity} + \text{Precision}} \tag{9}$$

To evaluate the real-time performance and computational efficiency, the time consumption and parameters of the proposed model are listed compared with other models. These metrics play a crucial role in assessing the recognition capability of medical diagnostic systems.

3.3 Experiment Results

The experiment is conducted based on the BraTS 2018 dataset for training and comparison, while also comparing it with existing state-of-the-art algorithms. The experimental results are shown in Table 1, which indicates that the proposed algorithm has achieved significant improvements of DSC in all three regions, with particularly noticeable increases in the TC and ET regions. The experimental comparison demonstrates the effectiveness of the proposed algorithm.

To comprehensively evaluate the overall performance of the proposed model, additional experiments are conducted to compare sensitivity and specificity metrics, and the results are shown in Table 2. The proposed method only realize a significant improvement in sensitivity for the WT region, while the improvements in TC and ET are not prominent. However, the specificity scores for all three regions have reached the highest values. The experimental results suggest that although the proposed algorithm achieves substantial gains in Dice scores, the improvements in sensitivity for TC and ET are not significant. Therefore, future experimental research should focus on additional tricks to enhance sensitivity.

Table 1
The Comparison of Dice Coefficient Score with Other Existing Models

Method	Dice Coefficient Score		
	WT%	TC%	ET%
V-Net [22]	87.61	79.53	73.64
Modified 3D U-Net [23]	90.80	84.40	78.40
Cascaded 2.5D CNN [24]	90.50	83.80	78.60
RMU-Net [24]	90.80	86.75	79.36
3D U-Net [24]	88.60	67.16	81.65
Improved 3D U-Net [24]	90.89	71.65	83.98
Ensemble OM-Net [26]	90.95	86.51	81.36
The proposed network	**93.08**	**91.03**	**87.04**

To prove the real-time performance of the proposed lightweight network, it is crucial to measure the inference speed of the network. The inference time is computed under a size of 240×240 image as shown in Table 3. The proposed model is compared with other lightweight popular segmentation model. Table 3 shows that the proposed model can achieve 21.3 ms inference time when processing one image with size of 240×240. Based on the abovementioned segmentation scores, the proposed model only has 0.88 M parameters, which could prove the performance of the proposed model.

3.4 Ablation Study

To validate the effectiveness of the proposed network model, the ablation study is conducted for comparison. According to the network framework described in this paper, the baseline model is defined as an encoder–decoder architecture composed of lightweight modules shown to the left of the block in Fig. 4. The network architecture is illustrated in Fig. 3. Subsequently, the baseline model is enhanced by incorporating spatial attention mechanism to further validate the effectiveness of the proposed spatial attention module. Finally, the lightweight module is improved by introducing the channel attention mechanism. The experiment compares the baseline model with both spatial and channel attention mechanism integrated. Table 4 presents the DSC and precision scores for three models, which can be observed that the proposed algorithm achieves the highest scores in all metrics, except for the DSC of the WT region, where it performs slightly lower. The significant improvement in precision is particularly notable.

Table 2
The Comparison of Sensitivity and Specificity with Other Existing Models

Method	Sensitivity			Specificity		
	WT%	TC%	ET%	WT%	TC%	ET%
MobileNetV2 [28]	82.34	73.58	76.13	83.19	71.26	81.45
MU-Net [24]	85.64	85.64	**95.26**	98.26	82.95	95.85
RMU-Net [24]	92.95	**90.36**	89.40	91.29	89.18	94.24
MC-Net [26]	90.65	79.70	80.95	99.51	99.81	99.80
OM-Net [26]	90.78	81.03	82.45	99.43	99.79	99.77
Improved 3D U-Net [24]	91.18	70.50	81.82	99.40	99.75	99.81
The Proposed Network	**94.44**	87.69	84.44	99.91	99.96	99.97

Table 3
The Comparison of Inference Time and Parameters with Other Existing Models

Method	Inference Time		Parameters	Parameter Size
	Time	FPS		
U-Net [15]	78.4 ms	12.8	7.76 M	29.62 MB
Attention U-Net [26]	87.3 ms	11.5	7.85 M	30.25 MB
E-Net [20]	33.9 ms	29.5	0.91 M	3.45 MB
The proposed	**21.3 ms**	**46.9**	0.88 M	3.34 MB

Table 4
The Comparison of DSC and Precision in Ablation Study

Method	DSC			Precision		
	WT%	TC%	ET%	WT%	TC%	ET%
Baseline	91.38	84.79	83.18	87.59	75.65	77.41
Baseline + Attention	**93.87**	87.12	80.99	94.07	88.34	78.15
Baseline + Attention + SE	93.08	**91.03**	**87.04**	**95.43**	**94.64**	**89.81**

Table 5 verifies the sensitivity and specificity metrics for the ablation study. It is evident from Table 5 that the introduction of the spatial attention mechanism and channel attention mechanism do not improve the sensitivity metric, although the specificity metric shows improvements. This result aligns with the experimental results in Table 2, where sensitivity does not show significant improvement. Analysis suggests that the introduction of spatial attention and channel attention mechanism significantly enhanced DSC and precision but compromised sensitivity. Therefore, to comprehensively evaluate the overall performance of the proposed model, the F1-Score is introduced.

The F1-Score is the harmonic mean of precision and sensitivity, which measures the accuracy of the model in predicting positive samples and the coverage of true positive samples. The ablation experimental results are shown in Table 6, which validates the improvements of precision through the introduction of the spatial attention module and channel attention mechanism. Consequently, this improvement also leads to an increase in the F1-Score, indicating an overall enhancement performance of the network. The experiments demonstrate that despite the decrease in sensitivity caused by the proposed attention mechanisms, the comprehensive performance metrics of the network still show improvement. This validates the effectiveness of the proposed algorithm.

To show the segmentation result of each slice in MRI images, Fig. 6 shows the comparison of prediction from the proposed model and the ground truth. The first row shows the prediction of the presented model and the second row indicates the label image, which can be observed that the segmentation results exhibit consistent overall

Table 5
The Comparison of Sensitivity and Specificity in Ablation Study

Method	Sensitivity			Specificity		
	WT%	TC%	ET%	WT%	TC%	ET%
Baseline	**95.51**	**96.45**	**89.89**	99.77	99.77	99.89
Baseline + Attention	93.66	85.92	84.05	99.87	99.87	99.94
Baseline + Attention + SE	94.44	87.69	84.44	**99.91**	**99.96**	**99.97**

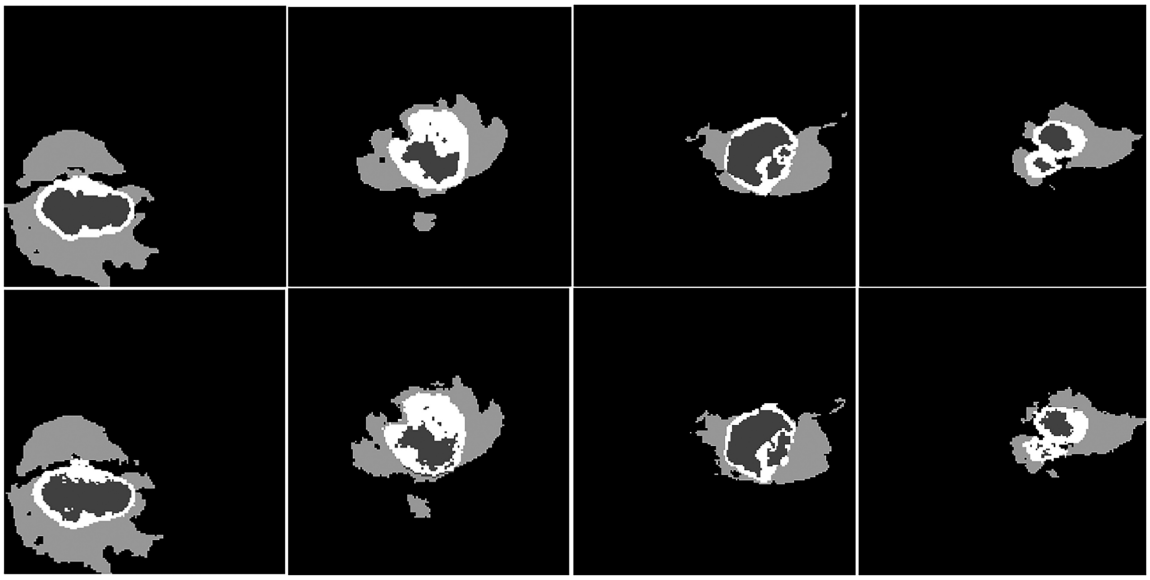

Figure 6. The visualisation segmentation results of the proposed model.

Table 6
The Comparison of Sensitivity and Specificity in Ablation Study

Method	Sensitivity	Precision	F1-Score
Baseline	**95.51**	87.59	91.38
Baseline + Attention	93.66	94.07	93.86
Baseline + Attention + SE	94.44	**95.43**	**94.93**

shapes but also need improvement in handling pixel-level details.

4. Conclusion

This paper presents a novel real-time lightweight segmentation network for brain tumour MRI segmentation with the application of medical imaging robots. A novel lightweight module with an introduction of the channel attention mechanism is designed to construct an encoder–decoder architecture. Additionally, a novel spatial attention module is devised to implement skip connection of the feature map between the encoder and the decoder. With experimental validation, the proposed network outperforms many existing networks in terms of DSC for tumor segmentation. The proposed spatial attention module and lightweight component with channel attention achieve high improvements in DSC and precision, which allows for a better performance of the proposed model. The introduction of the attention mechanism do not effectively improve the sensitivity metric, which is worth noting. The future research will focus on improving sensitivity indicators and exploring the inherent relationship between changes in sensitivity and attention mechanism.

References

[1] M.M. Mohamed, J. Gu, and J. Luo, Modular design of neurosurgical robotic system, *International Journal of Robotics and Automation, 33*(5), 2019, 542–551.

[2] J. Chen and H. Lau, Policy gradient-based inverse kinematics refinement for tendon-driven serpentine surgical manipulator, *International Journal of Robotics and Automation, 34*(3), 2019, 303–311.

[3] K. Wang, H. Yang, W. Wang, and Z. Han, Force configuration of a rigid-flexible gait rehabilitation robot, *International Journal of Robotics and Automation, 33*(6), 2018, 577–583.

[4] G. Yin, X. Zhang, and J.C. Chen, An approach for sEMG-based variable damping control of lower limb rehabilitation robot, *International Journal of Robotics and Automation, 35*(3), 2020, 171–180.

[5] V. Azimirad, Y. Salekzamani, and M. Ahmadi, Robotic diagnosis of trigger points, *International Journal of Robotics and Automation, 31*(6), 2016, 446–452.

[6] J. Seo, J. Cho, H. Woo, and Y. Lee, Development of prototype system for robot-assisted ultrasound diagnosis,

Proc. International Conf. on Control Automation and Systems (ICCAS), Busan, 2015, 1285–1288.

[7] J. An and S. Lee, Medical image analysis by robot kinematics, *Key Engineering Materials*, *326–328*, 2006, 875–878.

[8] S.E. Salcudean, H. Moradi, D.G. Black, and N. Navab, Robot-assisted medical imaging: A review, *Proceedings of the IEEE, 110*(7), 2020, 951–967.

[9] Z. Zhang, B. Rosa, O. Caravaca-Mora, P. Zanne, M.J. Gora, and F. Nageotte, Image-guided control of an endoscopic robot for OCT path scanning, *IEEE Robotics and Automation Letters, 6*(3), 2021, 5881–5888.

[10] X. Huang, J. Ren, G. Guiraudon, D. Boughner, and T.M. Peters, Rapid dynamic image registration of the beating heart for diagnosis and surgical navigation, *IEEE Transactions on Medical Imaging, 28*(11), 2009, 1802–1814.

[11] M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville, Y. Bengio, C. Pal, P.M. Jodoin, and H. Larochelle, Brain tumor segmentation with deep neural networks, *Medical Image Analysis*, *35*, 2016, 18–31.

[12] M.I. Razzak, M. Imran, and G. Xu, Efficient brain tumor segmentation with multiscale two-pathway-group conventional neural networks, *IEEE Journal of Biomedical and Health Informatics, 23*(5), 2019, 1911–1919.

[13] S.S. Mohseni Salehi, D. Erdogmus, and A. Gholipour, Auto-context convolutional neural network (Auto-Net) for brain extraction in magnetic resonance imaging, *IEEE Transactions on Medical Imaging, 36*(11), 2017, 2319–2330.

[14] E. Shelhamer, J. Long, and T. Darrell, Fully convolutional networks for semantic segmentation, *IEEE Transactions on Pattern Analysis and Machine Intelligence, 39*(4), 2017, 640–651.

[15] O. Ronneberger, P. Fischer, and T. Brox, U-Net: Convolutional networks for biomedical image segmentation, *Proc. International Conf. on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, Cham, 2015, 234–241.

[16] M. Zahangir Alom, M. Hasan, C. Yakopcic, T.M. Taha, and V.K. Asari, Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation, *arXiv:1802.06955*, 2018.

[17] X. Xiao, S. Lian, Z. Luo, and S. Li, Weighted Res-UNet for high-quality retina vessel segmentation, *Proc. International Conf. on Information Technology in Medicine and Education (ITME)*, Hangzhou, 2018, 327–331.

[18] T. Tarasiewicz, M. Kawulok, and J. Nalepa, Lightweight U-Nets for brain tumor segmentation, *in Lecture notes in computer science*, (Cham: Springer, 2021), 3–14.

[19] Y. Wang, Y. Cao, J. Li, H. Wu, S. Wang, X. Dong, and H. Yu, A lightweight hierarchical convolution network for brain tumor segmentation, *BMC Bioinformatics, 22*(5), 2022, 636.

[20] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello, ENet: A deep neural network architecture for real-time semantic segmentation, *arXiv:1804.02767*, 2016.

[21] J. Hu, L. Shen, and G. Sun, Squeeze-and-excitation networks, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 2018, 7132–7141.

[22] F. Milletari, N. Navab, and S.-A. Ahmadi, V-Net: Fully convolutional neural networks for volumetric medical image segmentation, *Proc. of 2016 fourth International Conf. on 3D Vision (3DV)*, Stanford, CA, 2016, 565–571.

[23] D. Lachinov, E. Vasiliev, and V. Turlapov, Glioma segmentation with cascaded UNet, *Proc. of the International MICCAI Brainlesion Workshop*, Cham, 2019, 189–198.

[24] G. Wang, W. Li, S. Ourselin, and T. Vercauteren, Automatic brain tumor segmentation based on cascaded convolutional neural networks with uncertainty estimation, *Frontiers in Computational Neuroscience, 13*, 2019, 56.

[25] M. U. Saeed, G. Ali, W. Bin, S. H. Almotiri, M. A. AlGhamdi, A. Ali Nagra, K. Masood, and R. Ul Amin, RMU-Net: A novel residual mobile U-Net model for brain tumor segmentation from MR images, *Electronics*, *10*(16), 2021, 1962.

[26] J. Huang, M. Zheng, and P.X. Liu, Automatic brain tumor segmentation using 3D architecture based on ROI extraction, *Proc. 2019 IEEE International Conf. on Robotics and Biomimetics (ROBIO)*, Dali, 2019, 36–40.

[27] C. Zhou, S. Chen, C. Ding, and D. Tao, Learning contextual and attentive information for brain tumor segmentation, *Proc. of the International MICCAI Brainlesion Workshop*, Cham, 2019, 497–507.

[28] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, MobileNetV2: Inverted residuals and linear bottlenecks, *Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Salt Lake City, UT, 2018, 4510–4520.

Biographies

Chaofan Du received the B.Sc. degree from Nanjing Agricultural University, China, in 2017. She is currently pursuing the master's degree and the doctor's degree with the Department of Mechanical and Electronic Control Engineering, Beijing Jiaotong University. Her current research interests include computer vision and data mining in biomedicine.

Peter Xiaoping Liu received the B.Sc. and M.Sc. degrees from Northern Jiaotong University, China, in 1992 and 1995, respectively, and the Ph.D. degree from the University of Alberta, Canada, in 2002. He has been with the Department of Systems and Computer Engineering, Carleton University, Canada, since July 2002, where he is currently a Professor. His interest includes interactive networked systems and teleoperation, haptics, surgical simulation, and control and intelligent systems. Dr. Liu has served as an Associate Editor for several journals, including *IEEE/CAA Journal of Automatica Sinica*, *IEEE Transactions on Cybernetics*, *IEEE/ASME Transactions on Mechatronics*, *IEEE Transactions on Automation Science and Engineering*, and *IEEE Transactions on Instrumentation and Measurement*. He is a licensed member of the Professional Engineers of Ontario (P.Eng), a Fellow of Canadian Academy of Engineering (FCAE), a Fellow of Engineering Institute of Canada (FEIC), and a Fellow of IEEE (FIEEE).