Machine Intelligence Research
www.mi-research.net

20(6), December 2023, 897-908
DOI: 10.1007/s11633-022-1392-6

Multitask Learning with Multiscale Residual Attention for Brain Tumor Segmentation and Classification

Gaoxiang Li   Xiao Hui   Wenjing Li   Yanlin Luo

School of Artificial Intelligence, Beijing Normal University, Beijing 100875, China

Abstract: Automatic segmentation and classification of brain tumors are of great importance to clinical treatment. However, they are challenging due to the varied and small morphology of the tumors. In this paper, we propose a multitask multiscale residual attention network (MMRAN) to simultaneously solve the problem of accurately segmenting and classifying brain tumors. The proposed MMRAN is based on U-Net, and a parallel branch is added at the end of the encoder as the classification network. First, we propose a novel multiscale residual attention module (MRAM) that can aggregate contextual features and combine channel attention and spatial attention better and add it to the shared parameter layer of MMRAN. Second, we propose a method of dynamic weight training that can improve model performance while minimizing the need for multiple experiments to determine the optimal weights for each task. Finally, prior knowledge of brain tumors is added to the postprocessing of segmented images to further improve the segmentation accuracy. We evaluated MMRAN on a brain tumor data set containing meningioma, glioma, and pituitary tumors. In terms of segmentation performance, our method achieves Dice, Hausdorff distance (HD), mean intersection over union (MIoU), and mean pixel accuracy (MPA) values of 80.03%, 6.649 mm, 84.38%, and 89.41%, respectively. In terms of classification performance, our method achieves accuracy, recall, precision, and F1-score of 89.87%, 90.44%, 88.56%, and 89.49%, respectively. Compared with other networks, MMRAN performs better in segmentation and classification, which significantly aids medical professionals in brain tumor management. The code and data set are available at https://github.com/linkenfaqiu/MMRAN.

Keywords: Brain tumor segmentation and classification, multitask learning, multiscale residual attention module (MRAM), dynamic weight training, prior knowledge.

Citation: G. Li, X. Hui, W. Li, Y. Luo. Multitask learning with multiscale residual attention for brain tumor segmentation and classification. *Machine Intelligence Research*, vol.20, no.6, pp.897–908, 2023. http://doi.org/10.1007/s11633-022-1392-6

1 Introduction

Brain tumors are characterized by a high risk and incidence, such as meningioma, glioma and pituitary tumors, which are among the major diseases threatening human health[1]. Meningiomas and pituitary tumors are usually benign, while gliomas are common malignant tumors. If the diagnosis is incorrect, it will delay patient treatment[2]. Currently, medical imaging diagnosis mainly requires doctors to determine the lesion region by observing magnetic resonance imaging (MRI) continuously and outlining the lesion manually. Manual outlining is a tedious and subjective task, and the accuracy of tumor contouring is mainly dependent on the doctor′s experience. The long time spent processing MRI also tires the doctor and leads to an increased rate of misdiagnosis. With the rapid development of deep learning in recent years, a new path has been opened for medical image processing based on deep learning. Tumor segmentation and classification is one of the applications of deep learning in the medical field. An accurate diagnostic model helps in the surgical planning and postoperative observation of patients and even helps to improve their survival rate[3]. Therefore, the automatic segmentation and classification of brain tumors are essential for the future development of clinical treatment.

Convolutional neural networks (CNNs) have shown compelling results in the processing of natural and medical images due to their powerful feature extraction[4], including residual networks (ResNet)[5], DenseNet[6], fully convolutional neural networks (FCN)[7], U-Net[8], and generative adversarial networks (GANs)[9]. Many studies perform segmentation and classification as two separate tasks. This separation often leads to ignoring the information associated with each task when acquiring features, making it difficult to obtain a more accurate model performance. Compared to single-task learning networks (STLs), multitask learning (MTL) is a general approach to improving generalization by learning tasks in parallel and is mainly divided into hard parameter sharing methods and soft parameter sharing methods[10]. Hard parameter sharing learns multiple tasks by sharing network

Research Article
Manuscript received on August 20, 2022; accepted on November 7, 2022
Recommended by Associate Editor Jun-Zhou Huang
Colored figures are available in the online version at https://link.springer.com/journal/11633

layers and task-specific layers. In soft parameter sharing, each task learns a corresponding network and shares information such as gradients. Since soft parameter sharing assigns a corresponding network to each task, it requires many additional parameters. In contrast, the multitask learning network reduces the training time by using hard parameter sharing, achieves better accuracy than single-task learning models, and significantly reduces the risk of overfitting.

Compared with natural images, brain tumor lesion regions often occupy only a tiny area in brain MRI, making it difficult for neural networks to extract effective features of the tumor. The attention mechanism, as an effective means of feature screening and enhancement, has been widely applied in many fields of deep learning. Attention-based network models enhance key information and suppress useless information by establishing dynamic weight parameters on information features[11]. Therefore, it is a good way to add an attention mechanism into the multitask learning network to make the network more focused on the tumor regions.

The loss function of a multitask learning network consists of the loss function of each task. During network training, the loss of each task may not be on the same order of magnitude. As training proceeds, the loss reduction rate of each task can also be inconsistent. This inconsistency makes the model focus on training tasks with fast loss reduction and underlearning tasks with slow loss reduction, which ultimately results in the model having better accuracy only for tasks with faster learning and underperformance for other tasks with insufficient learning[12]. Thus, reasonably setting the weights of the loss function for each task for a multitask learning network can enable each task to be adequately trained.

Previous studies usually added the doctor's prior knowledge of the tumor to the network's loss function to enhance the model's effectiveness. As a traditional image segmentation algorithm, the active contour model constrains the curve near the tumor region by establishing an energy equation. Chen et al.[13] added the active contour model as prior knowledge to the loss function of the network, making the network more focused on the shape of the tumor. However, adding overly complex prior knowledge tends to make the model difficult to converge and affects the model's accuracy. Therefore, it can improve the accuracy of the segmentation task by adding prior knowledge of brain tumors to the postprocessing of segmented images without affecting the training.

In this paper, we propose a multitask multiscale residual attention network (MMRAN) to simultaneously solve the problem of accurately segmenting and classifying brain tumors. The contributions of this paper are as follows:

1) We propose a novel multiscale residual attention module (MRAM), which can effectively extract multiscale information at a more granular level. We add it to the shared parameter layer of MMRAN to simultaneously improve the segmentation and classification performance.

2) The effect of different training methods on the model accuracy is investigated by changing the weights of the loss function of each task during training. The results show that the training method that dynamically adjusts the weights of each task can achieve better accuracy while avoiding extensive experimentation to determine the optimal weight for each task.

3) Prior knowledge of brain tumors is added to the postprocessing of segmented images to improve the final segmentation accuracy of the model through a fast and efficient traditional image processing method. By filling holes inside specific segmented images, the processed images are made complete and closer to the ground truth.

The remaining sections of this paper are as follows. Section 2 presents the work related to attention mechanisms and multitask learning algorithms. Section 3 introduces the proposed MMRAN model in detail. Section 4 analyses MMRAN's performance and compares it with other methods. Finally, we conclude our work in Section 5.

2 Related work

2.1 Attention mechanism

Attention mechanisms have been widely used in deep learning tasks such as natural language processing, speech recognition, and computer vision, with remarkable results[14].

The squeeze-and-excitation (SE) module proposed by Hu and Sun[15] enhances the expression of the neural network by explicitly modelling the correlation between feature channels and filtering out channel-specific attention. Roy et al.[16] proposed a spatial and channel SE (scSE) module consisting of channelwise attention and spatialwise attention. After computing channel attention and spatial attention in parallel, both calculations are summed up as input data for the next level. Woo et al.[17] showed experimentally that channel attention modules and spatial attention modules can be combined in parallel or sequential arrangement and that better results can be achieved by the sequential arrangement of channel attention modules first. Meanwhile, global max pooling (GMP) can collect a different target feature representation from global average pooling (GAP). The combined use of both can result in a more refined attention channel. Although the attention module can improve the model's accuracy, simply overlaying the attention module will degrade the model's performance. This is mainly because the dot product degrades the value of the deep feature map, and the output feature map corrupts the performance of the main branch[18].

Based on the above research, we improve the scSE module and add the multiscale module and residual connection to it to enhance the effectiveness of the attention module.

2.2 Medical image processing

Wu et al.[19] proposed a U-shaped network with a pyramidal self-attentive module to capture long-range dependencies and achieved better performance on the retinal vessel segmentation task. To fully utilize the correlation information among tasks, the Y-Net proposed by Mehta et al.[20] takes a two-stage structure to output segmentation and classification results. Chen et al.[21] used a multitask U-Net model, which added two classification modules in the middle and last layers of the U-Net network. Based on FCN, He et al.[22] added a fully connected layer at the end of U-Net to output segmentation and classification results.

At present, few studies apply the attention mechanism to multitask learning networks. To improve the attention of the network to the lesion region, we add the attention mechanism to our network.

3 Proposed method

3.1 Network structure of MMRAN

MMRAN adopts the hard parameter sharing method, and its structure is shown in Fig. 1. The network consists of a shared encoder, a segmentation task network, and a classification task network. The convolutional module in the network consists of a 3×3 convolutional layer, a batch normalization (BN) layer, and a rectified linear unit (ReLU) activation layer, sequentially. The maximum pooling operation is used for downsampling, while the transposed convolution operation is used for upsampling with one-half of the original number of channels. The segmentation and classification tasks share the same encoder structures, and attention modules are added to the shared encoder. Since the fully connected layer requires parameters to be determined in advance, it is generally necessary to meet this requirement by scaling the image to a fixed size. However, scaling the image can lead to the loss of key texture information, affecting the segmentation and classification results of the lesion area. Therefore, a spatial pyramid pooling (SPP) layer[23] is added before the fully connected layer for the classification task, enabling the network to handle images of arbitrary resolution size.

3.2 Multiscale residual attention module (MRAM)

As shown in Fig. 2, the MRAM consists of three main components: a multiscale module, channel attention module, and spatial attention module. The MRAM makes the following four changes to the scSE module. First, the multiscale feature map is obtained by implementing the multiscale module on the input features. Second, the GMP branch is added to the channel attention module, which is computed in parallel with the GAP branch to obtain the channel attention weight map. Third, the channel attention module and the spatial attention module are executed sequentially, and the channel attention module is executed first. Finally, the residual (RES) connection is added to the MRAM.

3.2.1 Multiscale module (MSM)

Given a feature map $F \in \mathbf{R}^{C\times H\times W}$ as input, the MRAM constructs the multiscale module to aggregate the multiscale features. As shown in Fig. 3, the features of the network go through four convolutional layers with different convolutional kernel sizes to obtain features of differ-

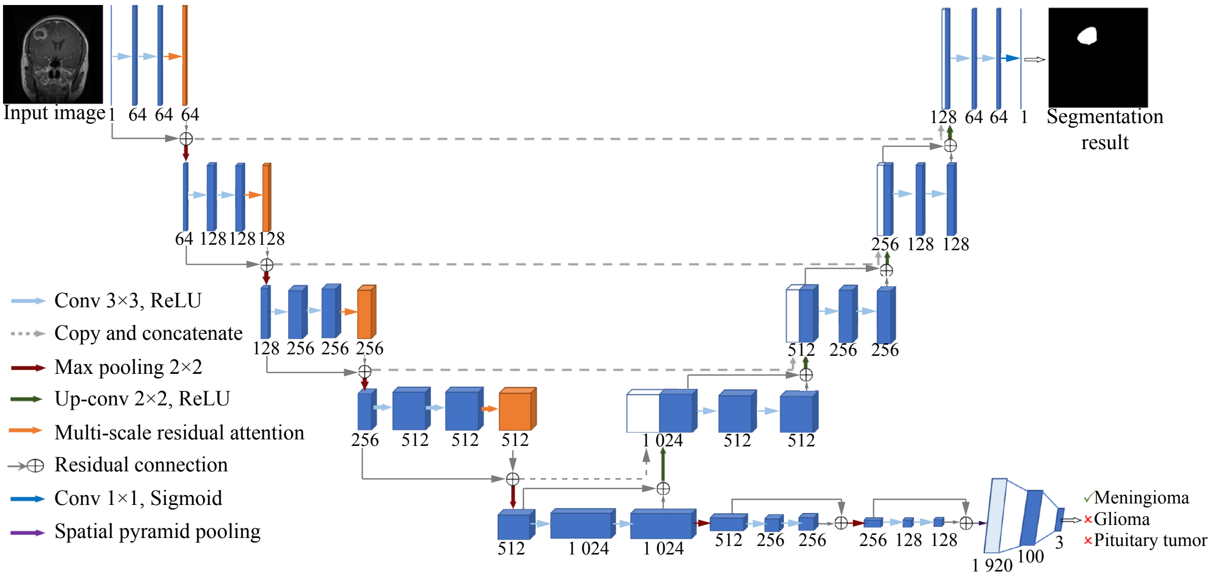


Fig. 1 Structure of the proposed network

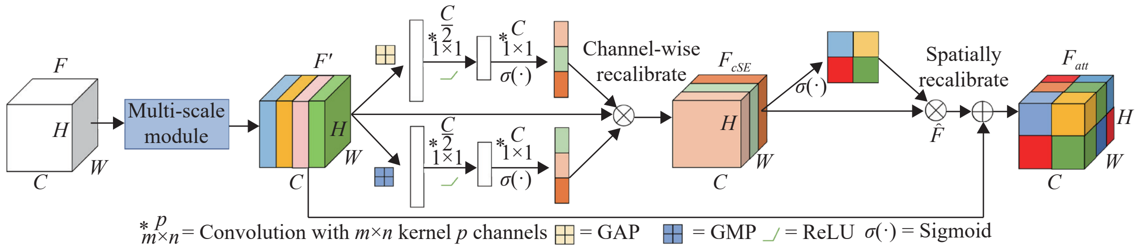


Fig. 2 Structure of multiscale residual attention module

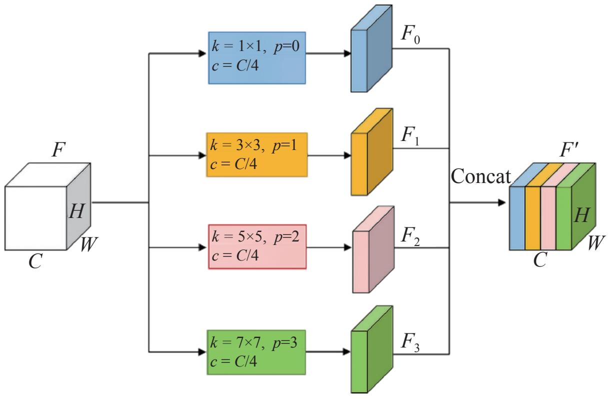


Fig. 3 Structure of the multiscale module, where k is the kernel size, s is the stride, c is the number of channels.

ent scales.

The multiscale feature map generation function is given by the following equation:

$$F_i = Conv\left(k_i \times k_i\right)(F), \quad i = 0, 1, 2, 3 \qquad (1)$$

where the i-th kernel size $k_i = 2\times(i+1)+1$ and F_i denotes the feature map with different scales. The whole preprocessed multiscale feature map can be obtained by concatenating these feature maps as

$$F' = Concat\left(\left[F_0, F_1, F_2, F_3\right]\right). \qquad (2)$$

After obtaining multiscale features, the MRAM sequentially executes the channel attention module and the spatial attention module to successively obtain a one-dimensional channel attention map $M_c \in \mathbf{R}^{C\times 1\times 1}$ and a two-dimensional spatial attention map $M_s \in \mathbf{R}^{1\times H\times W}$. The following describes the channel attention module and the spatial attention module.

3.2.2 Channel attention module

For the input feature mapping F', two different context descriptions are generated by GAP and GMP for the GAP feature $F^c_{ave} \in \mathbf{R}^{C\times 1\times 1}$ and the GMP feature $F^c_{max} \in \mathbf{R}^{C\times 1\times 1}$. The process of GAP is described as follows:

$$y_{c-avg} = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W} x_c\left(i,j\right) \qquad (3)$$

where y_{c-avg} denotes the output value of the c-th channel after GAP, $x_c\left(i,j\right)$ denotes the element located at $\left(i,j\right)$ on the c-th channel F_c of the feature mapping, and H and W denote the height and width of the channel, respectively.

The process of GMP is described as follows:

$$y_{c-max} = \max_{(i,j)\in F_c} x_c(i,j) \qquad (4)$$

where y_{c-max} denotes the output value of the c-th channel after GMP and $x_c\left(i,j\right)$ denotes the element located at $\left(i,j\right)$ on the c-th channel F_c of the feature mapping.

Next, F^c_{ave} and F^c_{max} are fed into their respective corresponding networks, each consisting of a multilayer perceptron (MLP) and a hidden layer. After the two features pass through their respective networks, the output of features from the two branches are combined using elementwise multiplication. Compared with the scSE module using elementwise summation, elementwise multiplication makes the position of the channel with high attention weight in both branches relatively more prominent. In contrast, the position of the channel with only one attention weight high or both weights low is further suppressed. In short, the channel attention is computed as follows:

$$M_c\left(F'\right) = \sigma\left(MLP\left(AP\left(F'\right)\right)\right)\odot\sigma\left(MLP\left(MP\left(F'\right)\right)\right) = \\ \sigma\left(W_1 L\left(W_0\left(F^c_{ave}\right)\right)\right)\odot\sigma\left(W_1' L\left(W_0'\left(F^c_{max}\right)\right)\right) \qquad (5)$$

where $\sigma(\cdot)$ denotes the sigmoid function, $\odot$ denotes elementwise multiplication, AP denotes the average pooling layer, MP denotes the max pooling layer, W_0 and W_1 denote the weights of the MLPs of the GAP branch, W_0' and W_1' denote the weights of the MLPs of GMP, and $L(\cdot)$ denotes the ReLU activation function.

The calculated channel attention $M_c\left(F'\right)$ is multiplied by the input feature map F to obtain the channel attention weight map.

$$F_{cSE} = M_c\left(F'\right)\odot F' \qquad (6)$$

where F_{cSE} denotes the attention weight map obtained from the channel attention calculation and $\odot$ denotes the elementwise multiplication.

3.2.3 Spatial attention module

For the result F_{cSE} calculated by the channel attention module, the spatial attention is obtained by a 1×1 convolution layer, and the spatial attention map is obtained after activation using the sigmoid function. The spatial attention is computed as follows:

$$M_s\left(F_{cSE}\right)=\sigma\left(f^{1\times1}\left(F_{cSE}\right)\right) \tag{7}$$

where $\sigma(\cdot)$ denotes the sigmoid function and $f^{1\times1}$ denotes the convolution operation with a convolution kernel size of 1×1.

Multiply the spatial attention $M_S\left(F_{cSE}\right)$ with F_{cSE} as follows:

$$\widehat{F}=M_s\left(F_{cSE}\right)\odot F_{cSE} \tag{8}$$

where $\widehat{F}$ denotes the attention weight map combining channel attention and spatial attention, and $\odot$ denotes the elementwise multiplication.

The calculated $\widehat{F}$ is summed with the feature mapping F input to the attention module, as follows:

$$F_{att}=F'+\widehat{F} \tag{9}$$

where F_{att} indicates the output of the MRAM.

To save the number of parameters, we add the MRAM to the residual module of the shared parameter layer of MMRAN, as shown in Fig. 4.

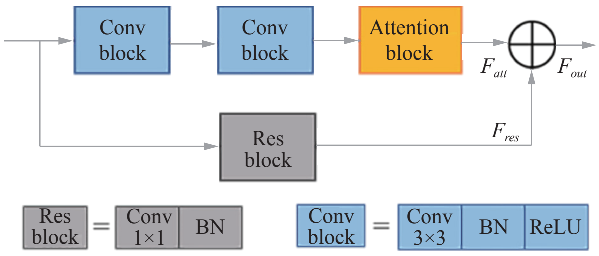


Fig. 4 Structure of the residual module

The output of the residual module is given by the following equation:

$$F_{out}=F_{att}+F_{res} \tag{10}$$

where F_{res} denotes the output of the residual connection and F_{att} denotes the output of the MRAM.

3.3 Compound loss function

Dice loss is used as the loss function for the segmentation task, which is formulated as follows:

$$Loss_{Dice}\left(G,P\right)=1-2\times\frac{\left|V_S\cap V_G\right|+\varepsilon}{\left|V_S\right|+\left|V_G\right|+\varepsilon} \tag{11}$$

where V_S denotes the set of real segmented images and V_G denotes the set of predicted segmented images. The constant ε is set to 1 to prevent the denominator from appearing equal to 0 during the calculation.

Since the incidence of each tumor is different[2], there may be differences in the size of the data for each type of tumor. The focal loss proposed by Lin et al.[24] makes the model more focused on hard-to-classify samples during training by reducing the weight of easy-to-classify samples. Therefore, focal loss is used as the loss function for the classification task, which is formulated as follows:

$$Loss_{FL}\left(p_t\right)=-\alpha_t(1-p_t)^{\gamma}\log\left(p_t\right) \tag{12}$$

where p_t denotes the probability that the sample belongs to the positive class, $\gamma\geq0$ is the moderation factor, α_t is the category weight, and $\left(1-p_t\right)^{\gamma}$ is the modulation factor. In this paper, we take $\alpha_t=0.25,\ \gamma=2$.

3.4 Network training methods

The weighted sum of Dice loss and focal loss is used as the loss function of MMRAN with the following equation:

$$Loss_{MTL}=(1-\alpha)\times Loss_{Dice}+\alpha\times Loss_{FL} \tag{13}$$

where α is the coefficient that adjusts the loss weights of the segmentation task and the classification task. The loss weights of each task are adjusted by α to optimize the training effect of both tasks.

This paper uses the following three network training methods for training.

1) Fixed weight training: Set fixed weights for each task before training starts. The parameter α is a constant and will not be changed during training.

2) Freeze training: Only one task is trained every N epochs, and the gradient backpropagation of other tasks is frozen. The parameter N is calculated by the following formula:

$$N=\frac{epoch_{all}}{c+1} \tag{14}$$

where $epoch_{all}$ is the total number of epochs and c is the total number of tasks in the network. There are two tasks in MMRAN: segmentation and classification. Thus, we take $c=2$. The loss weight of the frozen task is 0, and the loss weight of the training task is 1. This process is performed once for each task. In the last N epochs, all tasks are trained together, at which time $\alpha=0.5$.

3) Dynamic weight training: At the beginning of training, the loss of the overall network consists of the loss of the segmentation network only. As training proceeds, the weight of the loss function of the segmentation network is gradually reduced, while the weight of the loss function of the classification network is increased. At this point, α changes continuously with the number of training rounds,

and its formula is as follows:

$$\alpha = \frac{epoch_{now}}{epoch_{all}} \tag{15}$$

where $epoch_{now}$ is the current epoch and $epoch_{all}$ is the total number of epochs.

In the initial stage of training, the segmentation network has not yet produced meaningful outputs, indicating that a negative impact exists in the gradient backpropagation[25]. Therefore, we use the strategy of training the segmentation network first and then the classification network for freeze training and dynamic weight training.

3.5 Prior knowledge-based postprocessing (PKP)

Brain tumors are usually solid structures with irregular margins[26]. Large diameter tumors with high malignancy can develop internal necrosis and lead to the appearance of cavities[27]. However, the cavities caused by necrosis are no longer normal brain areas and should be excised. Since the segmentation algorithm is a pixel-level image annotation, each pixel is categorized as a tumor region or a normal region, so that not all pixels inside the segmented tumor region are necessarily labelled as the tumor. There may be holes inside the region that are classified as normal. Therefore, the output segmented images should be postprocessed based on prior knowledge.

The flood fill algorithm (FFA), a traditional image processing method, is used to fill the interior of the segmentation map. The segmentation results in a solid structure by filling the interior of the segmented connected region. Beginning from a random point in the connected domain, all the remaining points in the connected domain are found by searching for other points connected to each point in four directions: up, down, left, and right. The steps of the postprocessing method for segmented images are as follows:

1) The segmentation map output from the network is processed with FFA to obtain the connected region of the segmented tumor.

2) The pixel intensity values of the connected region are inverted.

3) The inverted image is merged with the segmentation map to obtain the complete solid segmentation map.

4 Model validation and analysis

4.1 Data set and data preprocessing

The images in our data set (T1-weighted contrast-enhanced MRIs) were acquired at Nanfang Hospital, Guangzhou, China, and General Hospital, Tianjin Medical University, China, from 2005 to 2010 using spin–echo-weighted images with a 512×512 matrix[28]. The pixel dimensions of the images were 0.49×0.49 mm^2, the slice thickness was 6 mm, the slice gap was 1 mm and the dose of Gd-DTPA was 0.1 mmol/kg at a rate of 2 ml/s. The data set contains information on 3 064 slices from 233 patients with brain tumors, wherein three types of brain tumors are meningiomas (708 slices), gliomas (1 426 slices) and pituitary tumors (930 slices). A single sample in the data set corresponds to one slice of a patient, while a patient has multiple slices. In addition, all tumors in the images were manually outlined by three experienced radiologists, who processed all images independently. Subsequently, the radiologists discussed together and reached a consensus on the segmentation of each tumor in each image. Every MRI has only one brain tumor.

The data preprocessing stage performs data enhancement on the data set, including horizontal flipping, vertical flipping, random angular rotation, random Gaussian noise, and random elastic deformation. All the above operations are performed with a 50% probability.

4.2 Experimental details

All networks are implemented using PyTorch 1.8.2 and Python 3.7.2 on the Ubuntu 18.04 system with a Tesla v100 GPU and 32 GB RAM. The adaptive moment estimation (Adam) algorithm is used as the optimizer, and the learning rate is set to 0.001. The training epochs for each experiment are set to 240, and the batch size is 10. The average training time for each network is nearly 28 hours. We divide the data set into train, validation, and test sets according to the ratio of 8:1:1. To facilitate the replication of the experimental results, we fixed the seeds as 1 029.

For the segmentation task, the Dice coefficient (Dice), 95% Hausdorff distance (HD95), mean intersection over union (MIoU), and mean pixel accuracy (MPA) are used as segmentation evaluation metrics. The formulas are as follows:

$$Dice = 2 \times \frac{|V_P \cap V_G|}{|V_P| + |V_G|} \times 100\% \tag{16}$$

$$HD95 = \max_{95\%}\left(\max_{x \in V_P}\min_{y \in V_G} ||x-y||_2, \max_{y \in V_G}\min_{x \in V_P} ||x-y||_2\right) \tag{17}$$

where V_G denotes the set of real segmented images and V_P denotes the set of predicted segmented images. $||\,x-y\,||_2$ denotes the Euclidean distance.

$$mPA = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k} p_{ij}} \tag{18}$$

$$mIoU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{p_{ii}}{\sum_{j=0}^{k}p_{ij}+\sum_{j=0}^{k}p_{ji}-p_{ii}} \tag{19}$$

where the notations are p_{ii} = true positive, p_{ij} = false positive, and p_{ji} = false negative.

For the classification task, accuracy (Acc), precision (Pre), recall (Rec), and F1-score (F1) are used as classification evaluation metrics. The formulas are as follows:

$$Acc = \frac{\sum_{i=1}^{n}TP_i}{\sum_{i=1}^{n}TP_i + FP_i} \tag{20}$$

$$Pre = \frac{TP}{TP+FP} \tag{21}$$

$$Rec = \frac{TP}{TP+FN} \tag{22}$$

where the notations are TP = true positive, FP = false positive, TN = true negative, and FN = false negative.

The network requires classifying three types of tumors. Therefore, after calculating the Pre and Rec for each category of tumors, the average value of each indicator is calculated as the corresponding indicator value of the model. F1 is calculated as follows:

$$F1 = \frac{2\times Pre_{ave}\times Rec_{ave}}{Pre_{ave}+Rec_{ave}} \tag{23}$$

where Pre_{ave} denotes the average Pre and Rec_{ave} denotes the average Rec.

During the training and ablation experimental phases, we use only two metrics: Dice as the segmentation metric and Acc as the classification metric. All the above metrics are calculated when the models are evaluated in the comparison experiments.

4.3 Network training method experiments

To explore the effect of different training methods on the performance of each task in the multitask learning network, experiments are conducted on the three training methods in this paper. The α in the fixed weight training takes values in the range of [0.2, 0.8] with a step size of 0.1. The results are shown in Table 1.

Table 1 shows that setting different task weights for each task of MMRAN during training has a certain degree of influence on the final model accuracy. In the fixed-weight training, the greater the weight of the loss function for a task, the greater its contribution of the overall loss function, and the better the model tends to perform for that task. However, the accuracy of other tasks is diminished. This may be because during the training process, the model focuses more on tasks with large weights, while the impact from tasks with small weights is ignored. The dynamic weight training method obtains the best Dice and the second highest Acc while avoiding multiple experiments to determine the best weights for each task, effectively saving the time for tuning parameters.

Table 1 Results of the network training method experiment

Training method	Dice (%)	Acc (%)
Set α = 0.2	77.51	85.29
Set α = 0.3	77.45	86.60
Set α = 0.4	77.01	86.27
Set α = 0.5	76.32	86.60
Set α = 0.6	76.14	86.93
Set α = 0.7	75.28	**87.58**
Set α = 0.8	72.57	**87.58**
Freeze training	77.42	86.27
Dynamic weight training	**77.58**	86.93

4.4 Ablation experiments

To evaluate the contribution of each module in the network to the performance of the model, we analyse the results of the ablation experiments. We use the U-Net model with the classification branch added as the baseline, which uses Dice loss and focal loss and is trained by dynamic weight training. The results of the ablation experiments are shown in Table 2.

Table 2 Results of ablation experiments

Model	Dice (%)	Acc (%)
Baseline	77.58	86.93
Baseline + SE	78.95	88.24
Baseline + scSE	79.01	88.89
Baseline + scSE + MSM	79.69	89.22
Baseline + scSE + MSM + GMP	79.80	89.54
Baseline + scSE + MSM + GMP + RES	79.83	89.54
Baseline + MRAM	79.87	**89.87**
Baseline + MRAM + PKP	**80.03**	**89.87**

Compared to not using the attention module, Dice is improved by 2.29% from 77.58% to 79.87%, and Acc is improved by 2.94% from 86.93% to 89.87% after adding the MRAM. After adding MSM to scSE, the segmentation performance and classification performance of the network are improved, which indicates that aggregating multiscale information can help the network extract features better. Adding the GMP branch to the channel attention module can collect information different from that obtained by the GAP branch, improving the performance of the network. The RES module has little impact on the performance. This may be because the number of network layers in this paper is not deep, and only four attention modules are used, making it difficult to show the role of residual connections. As a plug-and-play attention module, we still retain the residual connection in MRAM so that it can be applied to networks of different depths. The MRAM achieves better model accuracy than the SE and scSE modules in both segmentation and classification performance. Since we only add the prior knowledge of brain tumors into the image postprocessing of the segmented map, the inclusion of prior knowledge will not affect the classification accuracy.

In the segmentation maps in Fig. 5, some parts are not segmented. According to prior knowledge of the tumor, the internal void of the output map tumor should be filled. After image postprocessing, the Dice of Fig. 5(a) is improved by 1.83%, and the Dice of Fig. 5(b) is improved by 5.14%. Using prior knowledge-based postprocessing, the processed images are closer to the ground truth than the unprocessed segmentation result map.

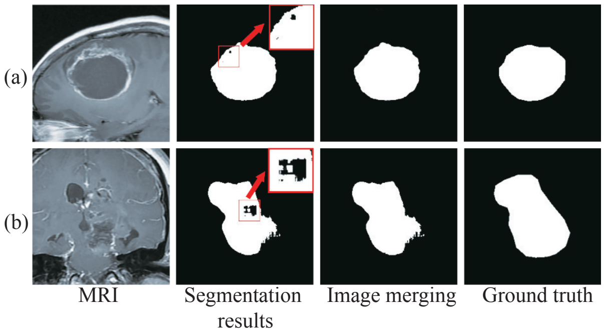


Fig. 5 Post-processing of segmented images. (a) MRI image with small holes after segmentation; (b) MRI image with large holes after segmentation.

The addition of prior knowledge slightly improves the overall segmentation performance because most images from the network segmentation are already complete solid structures, and only a few have holes inside. Nevertheless, this method results in more complete segmented images with less time (average processing time of 0.016 seconds per image) and more accuracy for single image segmentation.

4.5 Comparative experiments

To verify the effectiveness of MMRAN, a comparative analysis of performance is performed between MMRAN and other models. The parameters for networks are shown in Table 3.

4.5.1 Comparison of segmentation performance

The results of the segmentation performance evaluation are shown in Table 4.

From the results in Table 4, it can be obtained that Y-Net achieves good results with its smaller number of parameters. Our model achieves better results than other models in all four segmentation evaluation metrics, Dice, HD, MIoU, and MPA, reaching 80.03%, 6.649 mm, 84.38%, and 89.41%, respectively.

Table 3 Parameters of the networks

Net	Type	Parameters ($\times 10^6$)
Res U-Net[29]	STL	31.48
RV-GAN[30]	STL	21.93
ResNet-50[5]	STL	23.53
Datta et al.[31]	STL	23.91
Y-Net[20]	MTL	9.11
Chen et al.[21]	MTL	40.03
He et al.[22]	MTL	23.43
MB-DCNN[32]	MTL	26.72
MMRAN	MTL	42.73

Table 4 Results of the segmentation performance

Net	Dice (%)	HD95 (mm)	MIoU (%)	MPA (%)
Res U-Net[29]	75.01	9.012	80.49	84.97
RV-GAN[30]	79.46	6.832	84.26	88.39
Y-Net[20]	76.13	7.852	81.65	86.48
Chen et al.[21]	78.65	6.957	83.12	87.36
He et al.[22]	78.38	8.556	83.96	87.54
MB-DCNN[32]	78.92	6.964	84.03	88.26
MMRAN	**80.03**	**6.649**	**84.38**	**89.41**

The segmentation results of the method in this paper compared with other methods are shown in Fig. 6. For better observation, the box shows the enlarged lesion area. From the results in Fig. 6(a), the segmentation edges of the lesion region are poorly handled by Res U-Net, and there are discontinuous segmented regions. Other models also suffer from oversegmentation or undersegmentation of tumor edges. Compared with other models, the segmentation results of MMRAN match the doctor annotation better with more smooth segmentation edges. From the results in Fig. 6(b), MMRAN handles small tumor regions more accurately than others. Overall, the segmentation results of MMRAN are more accurate.

4.5.2 Comparison of classification performance

The results of the segmentation performance evaluation are shown in Table 5.

From the results in Table 5, it can be obtained that MMRAN achieved results over other models in Acc, Rec, Pre, and F1, reaching 89.87%, 90.44%, 88.56%, and 89.49%, respectively. We present the confusion matrices obtained by these models in Fig. 7. The results show that MMRAN correctly classifies most MRIs containing meningioma and is effective in classifying the other types of tumors. In the case of misclassification of glioma, each model is more likely to classify glioma as meningioma.

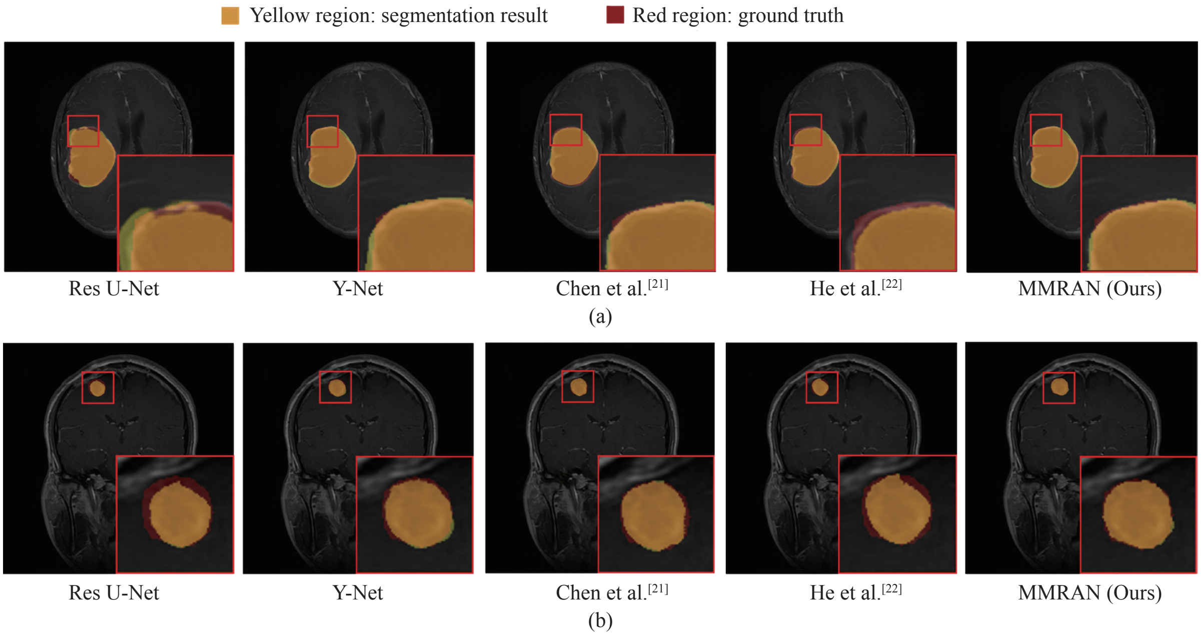


Fig. 6 Examples of the segmentation results produced by Res U-Net, Y-Net, Chen et al.'s model, He at el.'s model, and MTRAN: (a) Comparison of large tumor segmentation; (b) Comparison of small tumor segmentation.

Table 5 Results of the classification performance

Net	Acc (%)	Rec (%)	Pre (%)	F1 (%)
ResNet-50[5]	85.29	85.59	83.42	84.49
Datta et al.[31]	89.22	89.66	88.16	88.90
Y-Net[20]	87.91	88.51	86.61	87.55
Chen et al.[21]	86.93	86.79	85.32	86.05
He et al.[22]	87.58	87.64	85.89	86.75
MB-DCNN[32]	88.56	89.51	86.86	88.16
MMRAN	**89.87**	**90.44**	**88.56**	**89.49**

Meningioma and pituitary tumors are usually benign, whereas glioma is commonly malignant. Therefore, it is essential to detect glioma accurately. The area under the curve (AUC) is calculated to further evaluate the model's performance. From Table 6, the AUC of MMRAN is the highest for meningioma and glioma, reaching 99.01% and 97.04%, respectively. The AUC for pituitary tumors is second to Chen's model, reaching 95.74%. Thus, MMRAN achieves better overall performance for detecting brain tumors compared with other models.

The receiver operating characteristic (ROC) curves are plotted to evaluate the models' performance, as shown in Fig. 8.

5 Conclusions

A multitask learning network model based on a multiscale residual attention mechanism called MMRAN is proposed to improve the efficacy of segmentation and classification in brain tumors. We propose an effective plug-and-play attention module named the multiscale residual attention module (MRAM). Our proposed MRAM can effectively integrate multiscale contextual features and obtain better results than the SE module and scSE module. Experiments show that the network training methods that dynamically adjust the weights of the loss function for each task can obtain better results while avoiding multiple experiments to determine the optimal weights for each task. Adding prior knowledge of brain tumors in the postprocessing of segmented images makes the segmented image more complete and further improves the segmentation accuracy of the model. Compared with some existing methods, MMRAN achieves more accurate segmentation and classification ability on brain tumor data sets containing meningioma, glioma, and pituitary tumors.

MMRAN can also handle 3D data sets by processing each 2D slice in the Z-axis direction of the 3D image in turn. However, this approach may ignore the 3D spatial information of the lesions. Meanwhile, the multiple downsampling of the network will lose some spatial information. Avoiding the loss of tumor spatial information is the direction of our subsequent research. Extracting characteristics for small tumors is still challenging, leading to inadequate segmentation of small tumors. How to effectively extract the features of small targets is also a research direction to improve model performance. The proposed MMRAN is potentially applicable to other medical image tasks, such as segmentation and classification of lesion regions in chest CT images of pneumonia patients and choroidal vessels in the fundus. In the future, we will gradually improve the performance of the network and extend it to other image processing tasks.

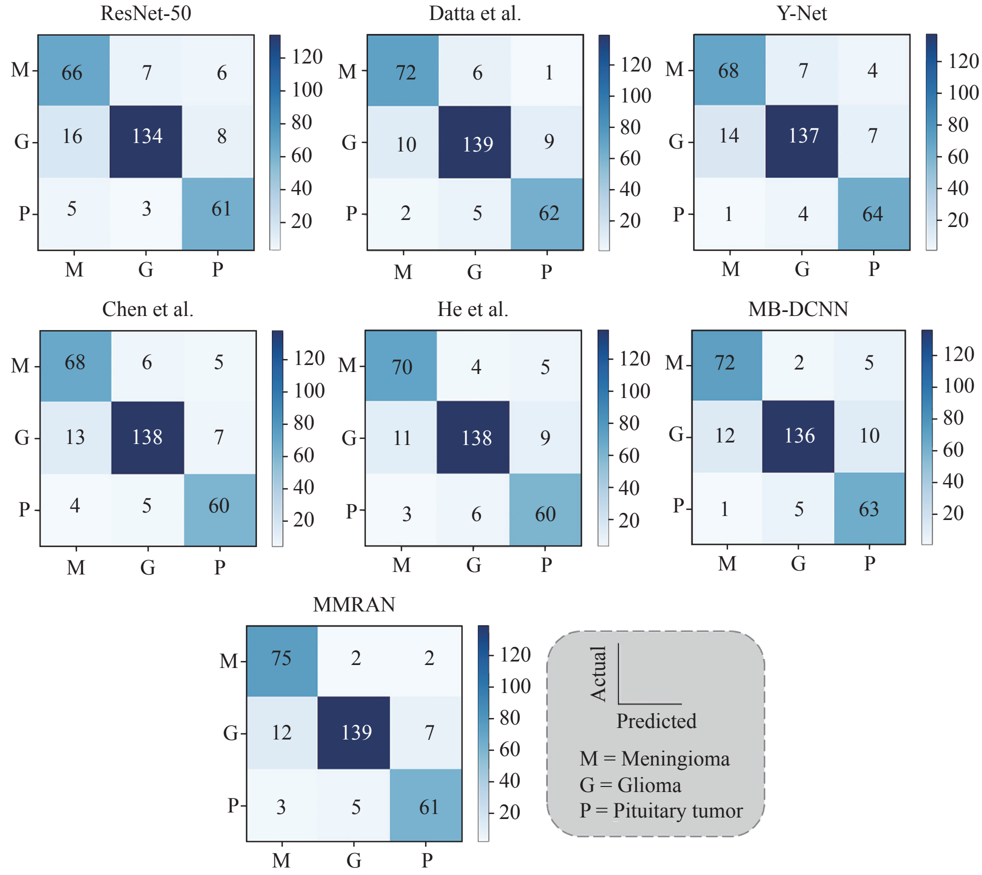
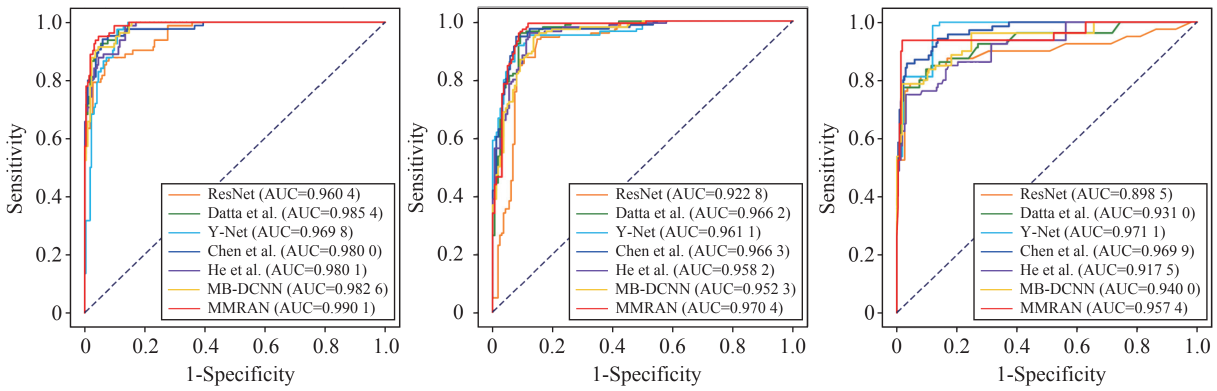


Fig. 7 Confusion matrices

Table 6 AUC results

Net	Meningioma (%)	Glioma (%)	Pituitary tumor (%)
ResNet-50	96.04	92.28	89.85
Datta et al.[31]	98.54	96.62	93.10
Y-Net[20]	96.98	96.11	**97.11**
Chen et al.[21]	98.00	96.63	96.99
He et al.[22]	98.01	95.82	91.75
MB-DCNN[32]	98.26	95.23	94.00
MMRAN	**99.01**	**97.04**	95.74



(a) ROC curves of meningioma (b) ROC curves of glioma (c) ROC curves of pituitary tumor

Fig. 8 Comparison of ROC curves

Acknowledgements

This paper was supported by National Natural Science Foundation of China (No. 61977063 and 61872020). The authors thank all the patients for providing their MRI images and School of Biomedical Engineering at Southern Medical University, China for providing the brain tumor data set. We appreciate Dr. Fenfen Li, Wenzhou Eye Hospital, Wenzhou Medical University, China, for her support with clinical consulting and language editing.

Declarations of conflict of interest

The authors declared that they have no conflicts of interest to this work.

References

[1] J. R. McFaline-Figueroa, E. Q. Lee. Brain tumors. *The American journal of medicine*, vol. 131, no. 8, pp. 874–882, 2018. DOI: 10.1016/j.amjmed.2017.12.039.

[2] C. Chen, Y. Hu, L. Lyu, S. Yin, Y. Yu, S. Jiang P. Zhou. Incidence, demographics survival of patients with primary pituitary tumors: A SEER database study in 2004–2016. *Scientific Reports*, vol. 11, no. 1, pp. 1–9, 2021. DOI: 10.1038/s41598-020-79139-8.

[3] S. Bakas, H. Akbari, A. Sotiras, M. Bilello, M. Rozycki, J. S. Kirby, J. B. Freymann, K. Farahani C. Davatzikos. Advancing the cancer genome atlas glioma MRI collections with expert segmentation labels and radiomic features. *Scientific Data*, vol. 4, no. 1, pp. 1–13, 2017.

[4] L. Hou, D. Samaras, T. M. Kurc, Y. Gao, J. E. Davis J. H. Saltz. Patch-based convolutional neural network for whole slide tissue image classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 2424–2433, 2016.

[5] K. He, X. Zhang, S. Ren J. Sun. Deep residual learning for image recognition. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, USA, pp. 770–778, 2016.

[6] G. Huang, Z. Liu, L. Van Der Maaten K. Q. Weinberger. Densely connected convolutional networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Hawaii, USA, pp. 4700–4708, 2017.

[7] J. Long, E. Shelhamer T. Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Boston, USA, pp. 3431–3440, 2015.

[8] O. Ronneberger, P. Fischer T. Brox. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Munich, Germany, pp. 234–241, 2015.

[9] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville Y. Bengio. Generative adversarial nets. *Advances in Neural Information Processing Systems*, Montreal, Canada, Article number 27, 2014.

[10] S. Ruder. An overview of multi-task learning in deep neural networks, [Online], Available: https://arxiv.org/abs/1706.05098, 2017.

[11] S. Chaudhari, V. Mithal, G. Polatkan R. Ramanath. An attentive survey of attention models. *ACM Transactions on Intelligent Systems and Technology*, vol. 12, pp. 1–32, 2021.

[12] S. Vandenhende, S. Georgoulis, W. Van Gansbeke, M. Proesmans, D. Dai L. Van Gool. Multi-task learning for dense prediction tasks: A survey, [Online], Available: https://arxiv.org/abs/2004.13379, 2021.

[13] X. Chen, B. M. Williams, S. R. Vallabhaneni, G. Czanner, R. Williams Y. Zheng. Learning active contour models for medical image segmentation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, IEEE, Los Angeles, USA, pp. 11632–11640, 2019.

[14] M. H. Guo, T. X. Xu, J. J. Liu. Attention mechanisms in computer vision: A survey. *Computational Visual Media*, vol. 8, pp. 331–368, 2022. DOI: 10.1007/s41095-022-0271-y.

[15] J. Hu, L. S. Sun. Squeeze-and-excitation networks. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Salt Lake City, USA, pp. 7132–7141, 2018.

[16] A. G. Roy, N. Navab, C. Wachinger. Concurrent spatial and channel "squeeze & excitation" in fully convolutional networks. In *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Granada, Spain, pp. 421–429, 2018.

[17] S. Woo, J. Park, J. Y. Lee I. S. Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision*, Springer, Munich, Germany, pp. 3–19, 2018.

[18] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang X. Tang. Residual attention network for image classification. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, USA, pp. 3156–3164, 2017.

[19] C. Z. Wu, J. Sun, J. Wang, L. F. Xu, S. Zhan. Encoding-decoding network with pyramid self-attention module for retinal vessel segmentation. *International Journal of Automation and Computing*, vol. 18, no. 6, pp. 973–980, 2021. DOI: 10.1007/s11633-020-1277-0.

[20] S. Mehta, E. Mercan, J. Bartlett, D. Weaver, J. G. Elmore L. Shapiro. Y-Net: Joint segmentation and classification for diagnosis of breast biopsy images. In *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Granada, Spain, pp. 893–901, 2018.

[21] E. Z. Chen, X. Dong, X. Li, H. Jiang, R. Rong J. Wu. Lesion attributes segmentation for melanoma detection with multi-task U-Net. In *Proceedings of the 16th IEEE International Symposium on Biomedical Imaging*, Venezia, Italy, pp. 485–488, 2019.

[22] T. He, J. Hu, Y. Song, J. Guo Z. Yi. Multi-task learning for the segmentation of organs at risk with label dependence. *Medical Image Analysis*, vol. 61, Article number 101666, 2020. DOI: 10.1016/j.media.2020.101666.

[23] K. He, X. Zhang, S. Ren J. Sun. Spatial pyramid pooling in deep convolutional networks for visual recognition. *IEEE*

Transactions on Pattern Analysis and Machine Intelligence, vol. 37, pp. 1904–1916, 2015. DOI: 10.1109/TPAMI.2015.2389824.

[24] T. Y. Lin, P. Goyal, R. Girshick, K. He P. Dollar. Focal loss for dense object detection. In *Proceedings of IEEE International Conference on Computer Vision*, Venezia, Italy, pp. 2980–2988, 2017.

[25] J Božič, D. Tabernik, D Skočaj. End-to-end training of a two-stage neural network for defect detection. In *Proceedings of the 25th International Conference on Pattern Recognition*, IEEE, Milan, Italy, pp. 5619–5626, 2021.

[26] Q. T. Ostrom, M. Adel Fahmideh, D. J. Cote, I. S. Muskens, J. M. Schraw, M. E. Scheurer, M. L. Bondy. Risk factors for childhood and adult primary brain tumors. *Neuro-oncology*, vol. 21, pp. 1357–1375, 2019. DOI: 10.1093/neuonc/noz123.

[27] C. H. Wu, Y. J. Liao, T. Y. Lin, Y. C. Chen, S. S. Sun, Y. W. H. Liu S. M. Hsu. A volume-equivalent spherical necrosis-tumor-normal liver model for estimating absorbed dose in yttrium-90 microsphere therapy. *Medical Physics*, vol. 43, pp. 6082–6088, 2016. DOI: 10.1118/1.4965044.

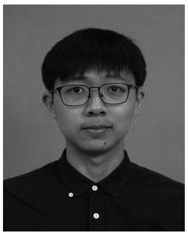
[28] Yang W, Feng Q J, Yu M. Content-based retrieval of brain tumor in contrast-enhanced MRI images using tumor margin information and learned distance metric. *Medical physics*, vol. 39, no. 11, pp. 6929–6942, 2012.

[29] X. Xiao, S. Lian, Z. Luo, S. Li. Weighted Res-UNet for high-quality retina vessel segmentation. In *Proceedings of the 9th International Conference on Information Technology in Medicine and Education*. IEEE, Hangzhou, China, pp. 327−331, 2018.

[30] S. A. Kamran, A. Sharif, A. Tavakkoli, S. L. Zuckerbrod, K. M. Sanders, S. A. Baker. RV-GAN: Segmenting retinal vascular structure in fundus photographs using a novel multi-scale generative adversarial network. In *Proceedings of International Conference on Medical Image Computing and Computer-assisted Intervention*, Springer, Strasbourg, France, pp. 34−44, 2021.

[31] S. K. Datta, M. A. Shaikh, S. N. Srihari. Soft Attention Improves Skin Cancer Classification Performance. In *Proceedings of Interpretability of Machine Intelligence in Medical Image Computing, and Topological Data Analysis and Its Applications for Medical Data*, Springer, Strasbourg, France, pp. 13−23, 2021.

[32] Y. Xie, J. Zhang, Y. Xia. A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Transactions on Medical Imaging*, vol. 39, no. 7, pp. 2482–2493, 2020. DOI: 10.1109/TMI.2020.2972964.



Gaoxiang Li received the B. Sc. degree in information and computing science from College of Science, Zhejiang University of Technology, China in 2020. He is currently a master student in computer application technology at School of Artificial Intelligence, Beijing Normal University, China.

His research interest is medical image processing.

E-mail: 202021210031@mail.bnu.edu.cn

ORCID iD: 0000-0003-4262-7699

Xiao Hui received the B. Sc. degree in computer science and technology from School of Artificial Intelligence, Beijing Normal University, China in 2020. She is currently a master student in computer application technology at School of Artificial Intelligence, Beijing Normal University, China.

Her research interests include medical image processing and virtual surgery.

E-mail: 202021210029@mail.bnu.edu.cn

Wenjing Li received the B. Sc. degree in computer science and technology from School of Artificial Intelligence, Beijing Normal University, China in 2021. She is currently a master student in computer application technology at School of Artificial Intelligence, Beijing Normal University, China.

Her research interest is medical image processing.

E-mail: 202121081305@mail.bnu.edu.cn

Yanlin Luo received the B. Sc. degree in mathematics and the M. Sc. degree in computer software and theory from School of Mathematics and Statistics, Lanzhou University, China in 1990 and 1993, respectively, and the Ph. D. degree in applied mathematics from School of Mathematical Sciences, Zhejiang University, China in 1997. She is currently an associate professor at School of Artificial Intelligence, Beijing Normal University, China.

Her research interests include medical image processing and visualization and virtual reality.

E-mail: luoyl@bnu.edu.cn (Corresponding author)

ORCID iD: 0000-0002-7881-2768

MIR | Machine Intelligence Research

Citation: G. Li, X. Hui, W. Li, Y. Luo. Multitask learning with multiscale residual attention for brain tumor segmentation and classification. *Machine Inteligence Research*, vol.20, no.6, pp.897–908, 2023. https://doi.org/10.1007/s11633–022–1392–6

Articles may interest you

Glaucoma detection with retinal fundus images using segmentation and classification. *Machine Intelligence Research*, vol.19, no.6, pp.563-580, 2022.
DOI: 10.1007/s11633-022-1354-z

Encoding-decoding network with pyramid self-attention module for retinal vessel segmentation. *Machine Intelligence Research*, vol.18, no.6, pp.973-980, 2021.
DOI: 10.1007/s11633-020-1277-0

Yolop: you only look once for panoptic driving perception. *Machine Intelligence Research*, vol.19, no.6, pp.550-562, 2022.
DOI: 10.1007/s11633-022-1339-y

Dual-domain and multiscale fusion deep neural network for ppg biometric recognition. *Machine Intelligence Research*, vol.20, no.5, pp.707-715, 2023.
DOI: 10.1007/s11633-022-1366-8

Machine learning for brain imaging genomics methods: a review. *Machine Intelligence Research*, vol.20, no.1, pp.57-78, 2023.
DOI: 10.1007/s11633-022-1361-0

Video polyp segmentation: a deep learning perspective. *Machine Intelligence Research*, vol.19, no.6, pp.531-549, 2022.
DOI: 10.1007/s11633-022-1371-y

Weakly correlated knowledge integration for few-shot image classification. *Machine Intelligence Research*, vol.19, no.1, pp.24-37, 2022.
DOI: 10.1007/s11633-022-1320-9

WeChat: MIR

Twitter: MIR_Journal