Scholarly Synergy Press
World Scientific Reports
2023, Volume1, Issue1, 2023, Type: Articles

Research article

Study on brain tumor image classification based on attention mechanism

Harish Chandra[1*]

1.Indian Institute of Information Technology, Allahabad, Department of Management Studies, India

*harish.chandra@iiita.ac.in

At present, deep learning has been successfully applied in the field of medical diagnosis, and it can effectively improve the diagnostic accuracy by using deep learning to predict brain tumor images. Based on traditional convolution neural networks tend to ignore the problem of key location in brain tumor image information, this paper proposes a door control channel attention conversion unit, door control channel attention switching unit by using the relationship between a small number of parameters to change the channel, under the condition of without any increase in computational cost more significantly improve classification accuracy. Also, the Sequeeze and Excitation block is introduced for the communication between the channels. Finally, a multipath attentional network model (FSnet) for brain tumor image classification is constructed based on ResNeXt network model. Experimental tests on brain MRI data sets show that FSnet has better classification performance than traditional convolutional neural networks.

Keywords: Attention mechanism; Convolutional neural network; Brain tumors; classification

Data Availability Statement

The data in this study can be provided without reservation by the corresponding author.
Also, the authors have no potential conflicts of interest.

1 Introduction

The brain is the main part of the human central nervous system and the main regulator of life functions. Although the incidence of brain tumors is relatively low, the mortality rate is high. Once a tumor develops in the brain, whether it is benign or malignant, it can cause great damage to the body. Brain tumors, also known as intracranial tumors, refer to primary tumors originating from meninges and neuroepithelium and other intracranial tissues, as well as secondary tumors from malignant tumors in other locations outside the skull to intracranial metastasis. The object of this study is brain glioma, which originates from brain glial cells and is the brain tumor with the highest mortality and morbidity, accounting for about 50% of brain tumors. Glioma, the most common primary brain tumor, is caused by cancerous changes of glial cells in the brain and spinal cord. Brain glioma is divided into several different kinds again, be like astrocytoma, mixed sex glioma. Different types of gliomas have different growth sites and pathological morphology, so their treatment strategies are also very different. According to the nature of the tumor, brain glioma can be divided into two types: benign glioma and malignant glioma. In general, benign gliomas are mature and grow slowly, so patients have a long life. Glioblastoma differentiates immaturely and grows rapidly, infiltrating into normal brain tissue and causing functional dysfunction. Glioblastoma is highly metastatic, and even with treatment, patients survive no more than two years after diagnosis. For glioma, the earlier detection and detection of intracranial lesions, and targeted treatment for patients, can reduce and reduce the harm of glioma to human health.

Among current imaging techniques, MRI is a particularly useful tool for clinical evaluation of gliomas, providing higher contrast images of the brain than CT imaging. Accurate segmentation of brain tumors is essential for cancer diagnosis, surgical planning, and postoperative evaluation. Computer-aided detection system has been used in medical image analysis for a long time. Traditional methods make the system performance close to the bottleneck, and deep learning is expected to break the deadlock and further improve the effect. Every day, hospital staff make critical decisions about patients' diagnosis and treatment through patient consultations, laboratory tests and imaging scans. Brain surgeons need to spend a lot of time and energy to correctly judge MRI images and correctly classify brain tumors and conduct

follow-up treatment. In the field of medical imaging, neural networks already have some applications, which relieve doctors of the burden of working on this difficult and error-prone project of artificial features. Ai is able to scan the data, compare it to thousands of other cases, and then suggest diagnosis and treatment options.

The purpose of this paper is to propose a high-performance brain tumor image classification model. The rest of this paper is organized as follows: Chapter 2 introduces the research status of deep learning and its application in brain tumor image classification. The third chapter introduces the model in detail. In chapter 4, the proposed model is tested and analyzed in data sets. The sixth chapter is the summary of the whole paper.

2 Related work

Deep learning is currently the hottest research field in machine learning. It establishes the basic idea of deep learning by simulating human brain neurons to establish synaptic connections and carry out hierarchical learning. In recent years, with the development of deep learning, medical image processing and computer vision have developed into a new interdisciplinary discipline. Medical images have their own advantages: large numbers of images, standardized formats can be quantified as a suitable soil for deep learning. At present, a large number of computer aided systems based on deep learning are used in the medical industry to assist doctors in diagnosis, bringing good news to doctors and patients. Not only reduce the workload of doctors still can provide more accurate diagnosis for the patient, and help doctors make better treatment for the patient and guidance, improve the cure rate of patients, and especially in the field of major diseases, such as brain tumor, lung cancer, breast cancer, cervical cancer and skin cancer diagnosis of these diseases is not less deep learning.

Classification task is one of the earliest tasks applied by deep learning in the field of medical image. It is generally divided into whole image classification and target classification. The meaning of whole image classification is to input the whole image into the deep learning model for end-to-end training. Target classification is a more detailed classification that classifies specific diseases, which can be classified by disease type or severity. For example, pulmonary nodules are usually classified as

good or bad, which is a common dichotomous problem in deep learning in the field of medical images. Target classification generally requires the collection of local information about the disease itself and information from other surrounding tissues to form contextual information. Brain tumors exhibit a high degree of variation in shape, size, and strength, and may exhibit a similar appearance to tumors from different pathologic types. Classification of brain tumors is a challenging research problem.

Among all brain tumors, glioma, meningioma and pituitary tumor have the highest incidence. Cheng Jun et al. studied the classification of 3 types of brain tumors using T1-MRI images, which was the first important work for classification using figShare, a challenging data set. The proposed method relied on manually delineating tumor boundaries to extract features from the region of interest [1]. Ismael and Abdel-Qadar [2] proposed a classification algorithm that firstly extracted statistical features using discrete wavelet transform (DWT) and Gabor filter, and then used these features to train multi-layer perceptron (MLP) classifier. In the work of Pashaei [3] et al., a CNN architecture was designed to extract features from brain MRI. The model has five learnable layers, all with 3x3 filters. The classification accuracy of CNN model is 81%. Performance was enhanced by using CNN functionality with a classifier model from extreme Learning Machine (ELM). For this work, the recall rate of pituitary classification was very high and that of meningioma was very low. This indicates that the classifier's discrimination is limited. Abiwinanda et al. [4] proposed a two-tier CNN architecture. Sultan et al. [5] proposed a CNN network containing 16 convolutional layers. Although these two models have a relatively simple structure, their accuracy is significantly improved compared with traditional machine learning algorithms. Later, Anaraki et al. [6] proposed a hybrid method combining CNN and genetic algorithm (GA) standards to improve the network architecture and further improve the classification effect. Francisco et al. proposed an automatic CNN based MRI image segmentation and classification method for meningioma, glioma and pituitary tumor. Different from the previous work based on CNN, its neural network contains three ways of processing information, which are at three spatial scales. Inspired by the inherent multi-scale operation of the human visual system (HVS) [7], the research team proposed a multi-scale tumor classification processing strategy, assuming that the multi-scale method can effectively extract the differential texture

features of different types of tumors. This model is very consistent with the HVS process. In visual stimulus processing, the visual stimulus system mainly works in two modes: preattention and attentional vision. Pre-attention mode is instantaneous and parallel, covering a wide area of the field of vision, while attentive processing acts on a limited area of the field of vision, establishing a serial search through focused attention [8]. In this process, simple cells in area V1 of the visual cortex filter stimuli from LGN (lateral geniculate nucleus) by frequency and bandwidth filtering. After filtering, the processed features are connected at the same time in the cognitive process [9]. Afshar et al. [10] used an improved CNN structure called Caps Net in brain tumor classification. Caps Net takes full advantage of the spatial relationship between the tumor and its surrounding tissue, but still lacks performance.

3 The proposed model

3.1 Attention conversion unit

GCT attention conversion unit [11] is a gated channel attention conversion unit, which is an attention mechanism combined with the gated mechanism. Instead of a full connection layer, it uses a normalized approach to model the relationship between channels. The gating mechanism uses a series of simplified parameters to model the feature relations between channels, so that GCT can transform the feature relations between channels into competitive or cooperative states, and its parameters can be trained together with the weight parameters of the network itself. The specific structure of GCT attention conversion unit is shown in Figure 1.

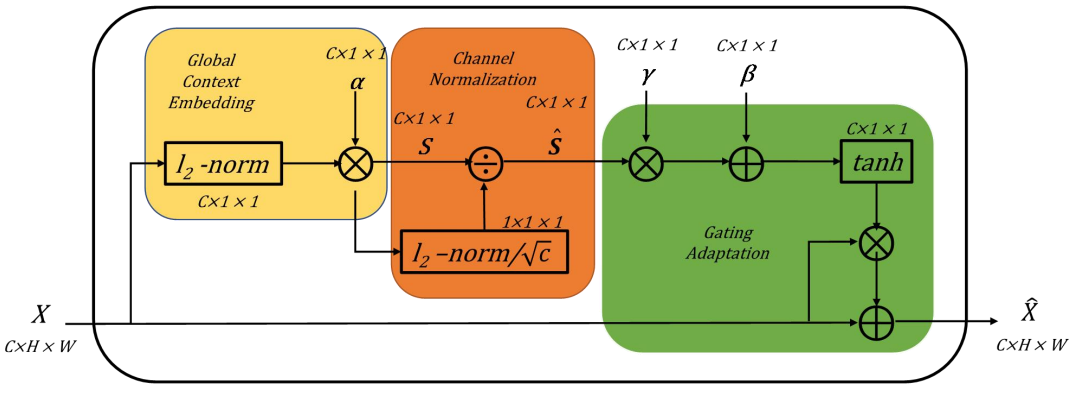

Fig.1. The structure of GCT

$X \in R^{C\times H\times W}$ is the characteristic diagram of the input network， GCT attention conversion unit transforms it into formula (1) :

$$\hat{x} = F(x|\alpha, \gamma, \beta), \alpha, \gamma, \beta \in R^C \quad (1)$$

Among them, α, γ, β is a set of trainable parameters, α can improve the adaptability of input, γ, β is the threshold control parameter. Their use leads to competitive or cooperative behavior of GCT channel attention-switching units in each channel.

Based on the basic structure of GCT attention conversion unit, this paper adds one-dimensional fast convolution to realize cross-channel access between adjacent channels, which enhances the channel modeling ability of GCT channel attention conversion unit. This gated channel attention conversion unit is called F-GCT. Its structure is shown in Figure 2.

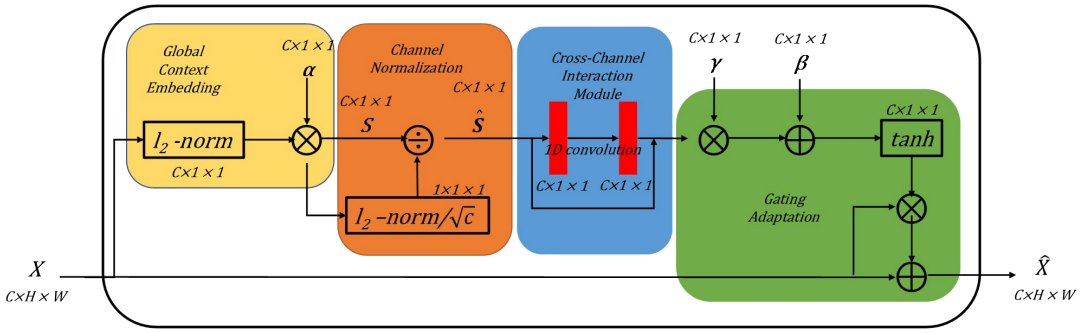

Fig.2. The structure of F-GCT

F-GCT consists of four parts: global context embedding module, channel normalization module, gated adaptive module and cross-channel interaction module. The global context embedding module uses l_2 regularization to embed the global context to increase the receptive field to avoid local confusion. Global context embedding modules do not use global average pooling because global average pooling fails in some special cases, such as nstance normalization, which is used by some networks, resulting in the mean values of various channels being fixed and resulting in constant results. The global context embedded module is defined as formula (2).

$$S_c = \alpha_c \|x_c\|_2 = \alpha_c \left\{ \left[\sum_{i=1}^{H} \sum_{j=1}^{W} \left(x_c^{i,j} \right)^2 \right] + \epsilon \right\}^{\frac{1}{2}} \quad (2)$$

Where, c represents the current c channel. ϵ is used to avoid derivative zero, and is generally set to a very small constant. F-GCT has good robustness for various regularization norms l_p. When the input values of F-GCT are all positive, l_1 norm is

equivalent to global average pooling. l_2 is the most outstanding among all regularization functions, so it is used as the regularization norm of the global context embedded module. The parameter α_c is used to adjust the importance of each channel. When α_c approaches 0, the c-th channel will not be regularized, which indicates that the most special case of the global context embedded module is that only one channel works.

Channel normalization module is a feature specification for each channel, which has the advantage that the relationship between different channels can be modeled with a few parameters. The channel normalization module is shown in Formula (3).

$$\widehat{s_c} = \frac{\sqrt{C}s_c}{\|s\|_2} = \frac{\sqrt{C}s_c}{\left[\left(\sum_{c=1}^{C} s_C^2\right)+\epsilon\right]^{\frac{1}{2}}} \tag{3}$$

Where ϵ is also set to avoid derivative 0. The role of $\sqrt{C}$ is to normalize the $\widehat{s_c}$ in the proper range. To keep s_c from appearing too small when *C* is large.

Appropriate cross-channel interaction helps to establish a channel attention mechanism. By focusing on the relationship between each channel in the feature map and k adjacent channels around the channel, the cross-channel interaction module can extract local cross-channel interaction features without reducing the number of channels. Therefore, the cross-channel interaction module effectively realizes the attentional prediction of how many neighbors are involved in a channel through a fast one-dimensional convolution of size k. It is reasonable that interaction coverage is related to channel dimension, so this article uses a function adaptive value that is related to channel dimension. This module can improve the network effect without increasing the number of parameters. Is an effective lightweight module. For each channel, only the information exchange with k neighboring channels is considered, and the number of parameters is k ×C. If the parameters of all channels are set to be shared, the final number of parameters is k. The cross-channel interaction module is shown in Formula (4).

$$\omega_c = \delta\left(\sum_{j=1}^{k} \alpha^j \widehat{s_c}\right), \widehat{s_c} \in \Omega_c^k \tag{4}$$

Ω_c^k represents k neighborhood channels of $\widehat{s_c}$, and ω_c is the weight of channel

c after cross-channel interaction. The number of parameters is k, and the number of k can be determined by itself.

The contribution of gated adaptive module is mainly the threshold mechanism, which can control the competition or cooperation between different channels. The definition of the gated adaptive module is shown in Formula (5) :

$$\widehat{x_c} = x_c[1 + \tanh(\gamma_c \omega_c + \beta_c)] \quad (5)$$

γ, β are two parameters controlling threshold function, which can be obtained by training. Gated adaptive module can promote both channel competition and collaboration. When the weight of the channel is positive, the gated adaptive module promotes the competition of the channel. When the weight of the channel is negative, the gating adaptive module promotes the cooperation of the channel.

3.2 Sequeeze and Excitation block

SEnet is the champion of Image Net 2017 classification task [12], and it is also a classic network using attention mechanism. In the past, neural networks only focus on the information fusion of local features, and do not pay attention to the relationship between global channels. SEnet is proposed to realize the communication of information between channels. The Sequeeze and Excitation block proposed by SEnet can be easily plugged into other networks for improved accuracy. SEnet consists of multiple SE blocks. Each SE block is split into two steps: Sequeeze and Excitation. Sequeeze obtains the global compression vector of the feature graph by performing global average pooling on the feature graph, while Excitation uses two full connection layers to calculate the weight occupied by each channel of the feature graph. Specifically, the workflow of SE Block is shown in Figure 3. GAP represents the average pooling layer, FC represents the full connection layer, and Scale represents the feature weight representation layer.

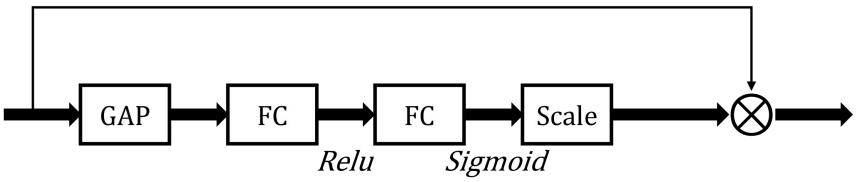

Fig.3. SE block 示意图

For the input feature graph, the global average pooling is firstly carried out to compress it, and then two consecutive fully connected layers and activation function can calculate the corresponding weight for each channel of the feature graph. The higher the weight is, the higher the importance of the channel is. Finally, the obtained weights are multiplied by the original feature image channel by channel to complete the feature image weighting.

3.3 Network structure construction

ResNeXt[16] is an improved model based on ResNet. Using the stack idea of VGG network, the network consists of a ResNeXt block. ResNeXt's most important contribution was to introduce the concept of cardinality in the model, the number of network paths, a parameter that makes the network malleable. ResNeXt can achieve high accuracy with low complexity because multipathing proved more effective for increasing the number of paths than for increasing the depth and width of the neural network. Unlike Inception's carefully constructed multipath, ResNet uses the same structure for each path, reducing the amount of network design effort. The ResNeXt structure is shown in Figure 4.

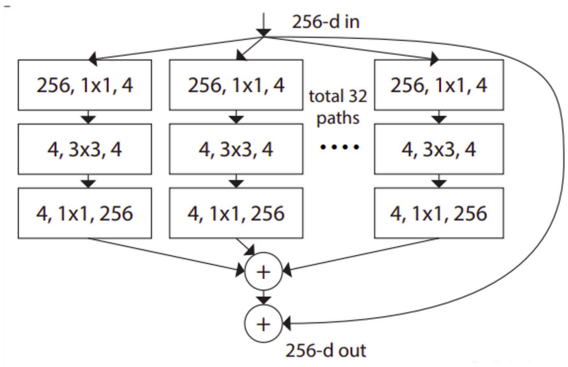

Fig.4. The structure of ResNeXt

ResNeXt combines the advantages of Inception and ResNet by concat both the incomplete structure and the feature layer, which is equivalent to merging the two models and inheriting the advantages of both. This article combines the Sequeeze and Dispatching block and the F-GCT Attention conversion unit for each ResNeXt block. A bi-attentional block is formed, as shown in Figure 5.

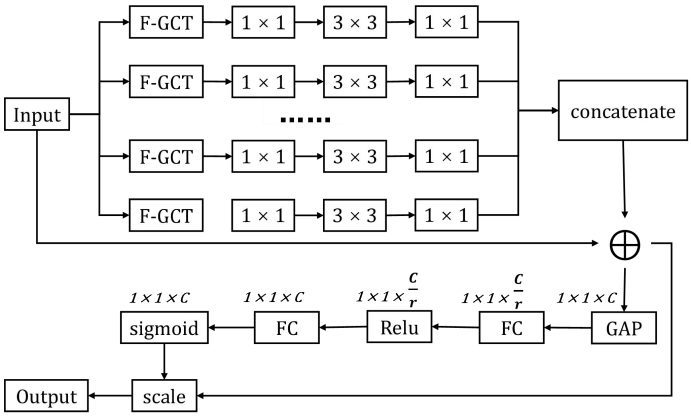

Fig.5. Biattentional block structure

In the figure above, C represents the number of channels for input or output features, $\frac{C}{r}$ represents lowering the channel dimension to the original $\frac{1}{r}$, which is usually set to 4. After the feature map is globally pooled and a full connection layer, Relu function is used to activate the feature map, and the number of channels is increased to the original C. The purpose of ascending and descending dimensions is to reduce the number of network parameters and speed up the network running time. After processing by sigmoid function, features are mapped to the range (0,1) as a mask. Finally, mask [18] was used to weight each channel and change the significance level of each channel. The process can be described as formula (6).

$$X_c = F_{\text{scale}\,(u_c, s_c)} = u_c S_c \tag{6}$$

u represents the c channel, S_c is the corresponding weight of u_c obtained after calculation, X_c is the channel multiplied by the original channel and corresponding weight, and X is the weighted feature map of all channels. This process enhances the representation of feature maps and improves the classification effect of classification models.

Based on ResNeXt, the Sequeeze and Excitation block and the F-GCT Attention conversion unit form a dual attention multipath network. We call it FSnet. Like ResNeXt, FSnet is stacked with a series of attention blocks. Each attention block is a combination of the F-GCT Attention transform unit and the Sequeeze and Excitation block. With only a few additional parameters, the optimal location for brain tumor imaging is improved.

The overall structure of FSnet is shown in Table 1. The main body of the network

can be divided into four parts, conv2, conv3, conv 4, and conv 5. The network consists of eight attention blocks. In the figure, C=32 represents 32 paths, and the next number represents the number of bolck repeats. Each block has 5 layers, so 5×8=40 layers. You start with a 7×7 convolution and end with a full connection layer for classification, so Gsnet has a total of 42 layers. The network input size is 224 × 224. The network input part is composed of a large convolution kernel with a size of 7×7 and step size of 2, and a maximum pooling kernel with a size of 3×3 and step size of 2. After convolution processing, the original input image size of 224×224 is reduced to 56×56, which greatly reduces the parameters of image processing. Then the feature map passes through each attention block in turn, the size of the feature map gradually shrinks and the number of channels gradually increases. Each attention block is split into 32 paths. For each path there are five sections: F-GCT,1×1 convolution, 3×3 convolution,1×1 convolution, and the Sequeeze and Excitation block. Finally, the network will go through a global pooling, through the full connection layer and sigmoid layer to get classification results.

Table 1 The network structure of FSnet

stage	output	FSnet
Conv1	112×112	conv, 7×7,stride2 max pool, 3×3,stride2
Conv2	56×56	$\begin{bmatrix} \mathrm{F-GCT} \\ \mathrm{conv}, 1\times 1{,}128 \\ \mathrm{conv}, 3\times 3{,}128 \quad C=32 \\ \mathrm{conv}, 1\times 1{,}256 \\ fc, [16{,}256] \end{bmatrix} \times 2$
Conv3	28×28	$\begin{bmatrix} \mathrm{F-GCT} \\ \mathrm{conv}, 1\times 1{,}256 \\ \mathrm{conv}, 3\times 3{,}256 \quad C=32 \\ \mathrm{conv}, 1\times 1{,}512 \\ fc, [32{,}512] \end{bmatrix} \times 2$
Conv4	14×14	$\begin{bmatrix} \mathrm{F-GCT} \\ \mathrm{conv}, 1\times 1{,}512 \\ \mathrm{conv}, 3\times 3{,}512 \quad C=32 \\ \mathrm{conv}, 1\times 1{,}1024 \\ fc, [64{,}1024] \end{bmatrix} \times 2$
Conv5	7×7	$\begin{bmatrix} \mathrm{F-GCT} \\ \mathrm{conv}, 1\times 1{,}1024 \\ \mathrm{conv}, 3\times 3{,}1024 \quad C=32 \\ \mathrm{conv}, 1\times 1{,}1024 \\ fc, [128{,}2048] \end{bmatrix} \times 2$
	1×1	global average pool,1000-d fc, softmax

4 Experiment and analysis

4.1 Data set

In this paper, brain MRI data set [19] was used to train the model. It is a data set collected and processed jointly by several hospitals and has reliable authority. This dataset consisted of 1426 glioma images, 708 meningioma images, and 930 pituitary tumor images. The format of the data is. Mat. Each file in this format stores a separate data structure. The data structure contains a classification label, tumor type of a specific brain image, patient ID, image data of 512×512 and type uint16, tumor boundary coordinates and mask image of tumor region. In this paper, image data and classification labels in the. Mat file are used as the input and classification labels of the model respectively.

Image data processing includes increasing image channel and adjusting image size. Since the MRI image is a gray image, this paper creates a three-channel image by copying the gray value three times. In addition, resizing the image means to adjust the size of the image in the data set to the size required by the network input.

4.2 The evaluation index

The task of image classification includes accuracy, accuracy and F1-score. These indices can be calculated from the obfuscation matrix. Confusion matrix is a technique used to summarize the performance of classification algorithms. The confusion matrix is shown in Table 2, where TP represents the number of positive samples predicted by the model and actually labeled as positive samples. FP represents the number of samples predicted as positive by the model, but actually labeled as negative. TN is the number of samples predicted by the model and actually negative; FN is the number of samples predicted by the model but actually marked as positive. The confusion matrix will summarize the correct and incorrect predictions by counting values and subdivide them according to each category, reflecting the relationship between the actual category and predicted category of each sample.

Table 2 Confusion matrix

		predicted	
		Yes	No
actual	Yes	TP	FN
	No	FP	TN

Accuracy reflects the proportion of samples in the prediction category that

predict correctly. The accuracy rate does not distinguish the positive and negative examples of the samples, but reflects the overall performance of the model algorithm. Its formula is shown in (7). The number of correctly predicted samples in the confusion matrix is divided by the total number of all predicted samples, and the accuracy rate can reflect the prediction ability of the model from an overall perspective.

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \tag{7}$$

Precision is the percentage of the sample where the prediction is correct. The formula is shown in (8). The calculated result is the proportion of positive samples in all samples with positive predictions. Accuracy is different from accuracy. It only focuses on the prediction ability of positive samples, and has nothing to do with the prediction effect of negative samples.

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

Recall rate refers to the proportion of positive samples correctly identified as positive samples among all positive samples, that is, how many positive samples can be correctly identified from these predicted samples. The formula is shown in (9), and the calculation result is the ratio of the number of correctly predicted positive samples to the total number of real positive samples.

$$\text{Recall} = \frac{TP}{TP+FN} \tag{9}$$

F1-score is one of the indicators used to measure the accuracy of the binary classification model. $F1$ is calculated by means of the harmonic mean of accuracy and recall, and its maximum value is 1 and minimum value is 0, as shown in Formula (10).

$$F1-score = \frac{2}{\frac{1}{\text{Precision}}+\frac{1}{\text{Recall}}} \tag{10}$$

$F1$-scores can be balanced between accuracy and recall, and a good $F1$ score means good accuracy and a good recall value. The combination judgment of these indexes can reasonably evaluate an algorithm model from multiple angles.

4.3 Experimental results and analysis

The laboratory server was used in the experiment. The operating system of the server was Ubuntu 16.04, CPU: Intel(R) Xeon(R) CPU @2.50GHz, GPU:GeForce GTX 1080Ti, Python version 3.6, CUDA version 10.2.

In the network initialization stage of GSnet, the offset term is 0, and the weight of initialization is evenly distributed by LeCun. When training the deep neural network model, common optimization methods such as SGD and Adam are difficult to achieve both convergence speed and convergence accuracy. In this paper, an adaptive random optimization algorithm RAdam method is used [13]. Based on the improvement of Adam, this method can control whether the adaptive learning rate is activated or not according to the divergence degree of variance, which not only maintains the fast convergence speed of Adam[14], but also improves the convergence accuracy because the gradient updating behavior in the late training is more similar to SGD[15]. Therefore, RAdam optimization method is adopted in this paper to give consideration to convergence speed and accuracy. In the training stage, the learning rate was 0.0001, the batch was 32, and the epoch was 100. Each batch of 32 images were input for training at the same time.

In this paper, the model training is carried out by using the five-fold cross-validation, which is divided into training set and test set according to the ratio of 4:1. The training set is used for model training, and the test set evaluates the classification effect of the trained model. The cross-validation method can be used to evaluate the classification effect of the model. It can not only reduce the degree of overfitting but also fully test the adaptability of the model to new data. In addition, if the data set is not large enough, cross-validation can be used to make full use of the data set to learn as much information as possible. K-fold cross-validation means that the variance generated by the network is reduced by averaging the results of K training groups, so that the performance of the model does not depend on the division of data sets. The data set partitioning process is shown in Figure 6.

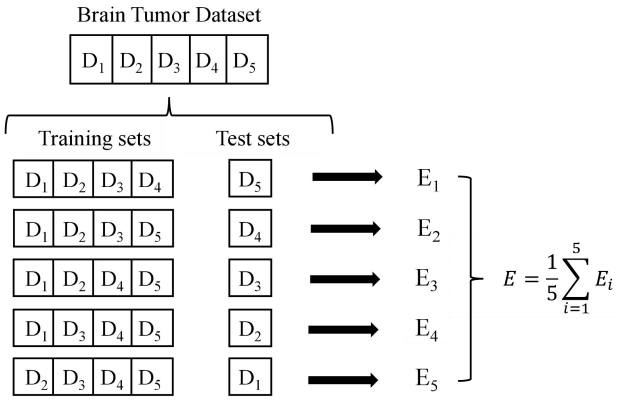

Fig.6. Cross validation

In the experiment, 2452 images were selected as the training set for the training test of the model, and the remaining 611 images were composed of the test set to test the training effect of the model. Among them, the training set included 1141 glioma images, 566 meningioma images and 744 pituitary tumor images, while the test set included 285 glioma images, 142 meningioma images and 186 pituitary tumor images.

Firstly, the FSnet network proposed in this paper was used for experiments, and the confusion matrix table as shown in Table 3 was obtained. Eleven of the predicted meningiomas were incorrectly classified as gliomas and three were incorrectly classified as pituitary tumors. Eleven of the gliomas were misclassified as meningiomas and five as pituitary tumors. Among pituitary tumors, 4 cases were misdiagnosed as meningiomas and 14 cases were misdiagnosed as pituitary tumors.

Table 3 The result Confusion matrix

		predicted		
		Glioma	meningioma	Pituitary tumor
actual	Glioma	269	11	5
	meningioma	11	128	3
	Pituitary tumor	14	4	168

The corresponding values of accuracy, precision, recall rate and F1 are shown in Table 4. Glioma is less accurate than the other two tumors, but its accuracy is higher. Meningioma had the lowest recall rate and F1-score.

Table 4 Classification performance

	Glioma	meningioma	Pituitary tumor
Accuracy	93.31%	95.26%	95.75%
Precision	94.39%	90.14%	90.32%
Recall	91.49%	89.54%	95.45%

F1-Score	92.91%	89.82%	92.81%

In order to further verify the effectiveness of FSnet network structure proposed in this paper for brain tumor classification, FSnet network and VGG16[17], VGG19[20], ResNet34[21], GoogLeNet [17], Mobilenet-V2 [18] and Inception-V3[22] are compared and analyzed. Meanwhile, f-GCT attention conversion unit in FSnet network is removed in order to verify the role of F-GCT. Defined as Snet. In order to facilitate the analysis, the glioma with the largest proportion in the data set was taken as a positive example, and the other two types were taken as negative examples to conduct further comparative analysis on the performance of the network model. The results are shown in Table 5. Among the four evaluation indexes, FSnet network model is only slightly lower than Mobilenet-V2 in precision, and all other indexes are optimal, indicating the superiority of FSnet network model. Compared with Snet, FSnet network model also has great advantages, indicating that F-GCT attention conversion unit can effectively improve the classification effect of the model.

Table 5 Model comparison results

MODEL	Accuracy	Precision	Recall	F1-Score
VGG16	87.82%	90.57%	89.78%	90.17%
VGG19	90.13%	89.88%	91.42%	90.64%
ResNet34	90.75%	89.56%	91.24%	90.39%
GoogLeNet	91.16%	90.19%	91.49%	90.83%
MobileNet-V2	91.47%	**91.81%**	90.77%	91.29%
Inception-V3	89.44%	89.34%	90.36%	89.85%
Snet	90.31%	88.96%	90.51%	89.73%
FSnet	**92.17%**	91.49%	**94.38%**	**92.91%**

5 Conclusion

In this paper, a gated channel attention conversion unit is proposed. By using a few parameters to change the relationship between channels, the gated channel attention conversion unit can significantly improve the classification accuracy without increasing the computational cost. At the same time, the Sequeeze and Congestion block is introduced for information communication between the channels to improve the accuracy of the deep learning network model for brain tumor image classification. A multipath attentional network model (FSnet) for brain tumor image classification is proposed based on ResNeXt network model. Experimental testing on brain MRI data sets found that FSnet performed better than popular high-precision models such as VGG16, VGG19, ResNet34, GoogLeNet, Mobilenet-V2 and Inception-V3.

References

[1] an M, Pang R, Le Q V. Efficientdet: Scalable and efficient object detection[C].Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 10781-10790.

[2] Ismael M R, Abdel-Qader I. Brain tumor classification via statistical features and back-propagation neural network[C].2018 IEEE international conference on electro/information technology (EIT). IEEE, 2018: 0252-0257.

[3] Pashaei A, Sajedi H, Jazayeri N. Brain tumor classification via convolutional neural network and extreme learning machines[C].2018 8th International conference on computer and knowledge engineering (ICCKE). IEEE, 2018: 314-319.

[4] Abiwinanda N, Hanif M, Hesaputra S T, et al. Brain tumor classification using convolutional neural network[C].World congress on medical physics and biomedical engineering 2018. Springer, Singapore, 2019: 183-189.

[5] Sultan H H, Salem N M, Al-Atabany W. Multi-classification of brain tumor images using deep neural network[J]. IEEE Access, 2019, 7: 69215-69225.

[6] Anaraki A K, Ayati M, Kazemi F. Magnetic resonance imaging-based brain tumor grades classification and grading via convolutional neural networks and genetic algorithms[J]. biocybernetics and biomedical engineering, 2019, 39(1): 63-74.

[7] Díaz-Pernas F J, Martínez-Zarzuela M, Antón-Rodríguez M, et al. Learning and surface boundary feedbacks for colour natural scene perception[J]. Applied Soft Computing, 2017, 61: 30-41.

[8] ulesz B, Bergen J R. Textons, the fundamental elements in preattentive vision and perception of textures, Bell Systems Tech[J]. J, 1983, 62: 1619-1645.

[9] Kruger N, Janssen P, Kalkan S, et al. Deep hierarchies in the primate visual cortex: What can we learn for computer vision?[J]. IEEE transactions on pattern analysis and machine intelligence, 2012, 35(8): 1847-1871.

[10] Afshar P, Plataniotis K N, Mohammadi A. Capsule networks for brain tumor

classification based on MRI images and coarse tumor boundaries[C].ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019: 1368-1372.

[11] Yang Z, Zhu L, Wu Y, et al. Gated channel transformation for visual recognition[C].Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 11794-11803.

[12] Hu J, Shen L, Sun G. Squeeze-and-excitation networks[C].Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7132-7141.

[13] Liu L, Jiang H, He P, et al. On the variance of the adaptive learning rate and beyond[J]. ar Xiv preprint ar Xiv, 2019:1908.03265.

[14] Kingma D P, Ba J. Adam: a method for stochastic optimization. ar Xiv[J]. ar Xiv preprint ar Xiv:1412.6980, 2014, 22.

[15] Bottou L. Stochastic gradient descent tricks[M].Neural networks: Tricks of the trade. Springer, Berlin, Heidelberg, 2012: 421-436.

[16] Mazzia, Vittorio, Francesco Salvetti, and Marcello Chiaberge. "Efficient-Caps Net: Capsule Network with Self-Attention Routing." ar Xiv preprint ar Xiv:2101.12491 (2021).

[17] Sadad T, Rehman A, Munir A, et al. Brain tumor detection and multi‐classification using advanced deep learning techniques[J]. Microscopy Research and Technique.

[18] Alaraimi S, Okedu K E, Tianfield H, et al. Transfer learning networks with skip connections for classification of brain tumors[J]. International Journal of Imaging Systems and Technology, 2021.

[19] Cheng J, Yang W, Huang M, et al. Retrieval of brain tumors by adaptive spatial pooling and fisher vector representation[J]. Plo S one, 2016, 11(6): e0157112.

[20] Simonyan K , Zisserman A . Very Deep Convolutional Networks for Large-Scale Image Recognition[J]. Computer Science, 2014.

[21] He K , Zhang X , Ren S , et al. Deep Residual Learning for Image Recognition[J]. IEEE, 2016.

[22]Szegedy C , Vanhoucke V , Ioffe S , et al. Rethinking the Inception Architecture for Computer Vision[C]// 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2016:2818-2826.