Tech Science Press

A Multi-Category Brain Tumor Classification Method Bases on Improved ResNet50

Linguo Li[1,2], Shujing Li[1,*] and Jian Su[3]

[1]Fuyang Normal University, Fuyang, 236041, China
[2]Nanjing University of Posts and Telecommunications, Nanjing, 210003, China
[3]University of Michigan, Ann Arbor, MI, 48109, USA
[*]Corresponding Author: Shujing Li. Email: lsjing1981@163.com

Abstract: Brain tumor is one of the most common tumors with high mortality. Early detection is of great significance for the treatment and rehabilitation of patients. The single channel convolution layer and pool layer of traditional convolutional neural network (CNN) structure can only accept limited local context information. And most of the current methods only focus on the classification of benign and malignant brain tumors, multi classification of brain tumors is not common. In response to these shortcomings, considering that convolution kernels of different sizes can extract more comprehensive features, we put forward the multi-size convolutional kernel module. And considering that the combination of average-pooling with max-pooling can realize the complementary of the high-dimensional information extracted by the two structures, we proposed the dual-channel pooling layer. Combining the two structures with ResNet50, we proposed an improved ResNet50 CNN for the applications in multi-category brain tumor classification. We used data enhancement before training to avoid model over fitting and used five-fold cross-validation in experiments. Finally, the experimental results show that the network proposed in this paper can effectively classify healthy brain, meningioma, diffuse astrocytoma, anaplastic oligodendroglioma and glioblastoma.

Keywords: Brain tumor; convolutional neural network; multi-size convolutional kernel module; dual-channel pooling layer; ResNet50

1 Introduction

Brain tumors are one of the most deadly cancers. Early detection plays an important role in the treatment and rehabilitation of patients. However, manual brain tumor diagnosis by physicians is a less accurate and time-consuming procedure [1]. Therefore, it is necessary to establish an intelligent and effective brain tumor classification system, which can help doctors analyze medical images and support specialists in making diagnoses at the early stage of tumor growth.

Medical image analysis has become a recent research hotspot. Machine learning and deep learning approaches in particular play core roles in computer-assisted brain image analysis, segmentation, registration, and tumor tissue classification [2–5]. At present, CNN is considered to be the most successful image processing method [6]. Instead of matrix multiplication, convolution operators are used in most layers of these networks. This contributes to the superiority of convolutional networks in solving problems with high computational costs. Another advantage of this method compared to shallow machine learning methods, is automatic feature extraction. Because of these advantages, CNN has also been widely used in the processing of medical images, such as grade classification [7], segmentation [8,9] and skull stripping of brain tumor images [10]. Establishing CAD [11] or multi-level brain tumor grading systems, supporting radiologist visualization [12] and identifying tumor types [13–15] using the deep neural network help radiologists make decisions on early diagnosis and future treatment.

2 Related Work

The detection of brain tumors has received extensive research attention, and various detection methods have been proposed in several years. Abd-Ellah [16] proposed a two-phase multi-model automatic brain tumor diagnosis system from magnetic resonance images using CNNs, in which the classification stage model consists of three parts, namely preprocessing, CNN feature extraction and error correction output code support vector machine (ECOC-SVM). The highest average accuracy of the model for benign and malignant brain tumor classification is 99.55%. Abdolmaleki et al. [17] developed a three-layer back-propagation neural network that distinguished between malignant and benign tumors using 13 different features. The magnetic resonance imaging (MRI) data they used for the experiment came from 165 patients, and the classification accuracy for benign and malignant tumors reached 91% and 94% respectively. These studies only focus on the binary classification of brain tumors, but only the benign and malignant classification is not enough for radiologists to decide on the treatment plan for the patients.

Sajjad [18] proposed a brain tumor classification method based on CNN. Firstly, the deep learning algorithm is used to segment the tumor region in MRI image, then the enhanced data set is used to fine tune the CNN model. The average accuracy of the model for four tumor classifications (Grade I-Grade IV) in radiopedia dataset is 87.38%. In 2020, Ghassemi [19] proposed a new deep learning method for MRI image classification of brain tumors. Firstly, a deep neural network is pre-trained on different data sets as a discriminator for generating confrontation network (GAN) to extract the robust features of MRI image and learn the structure of MRI image in its convolution layer, and then the softmax layer is used to replace the full connection layer of the network. The average classification accuracy of this method for meningioma, glioma and pituitary tumor is 93.01%. The above methods all use deep neural network [18,19], but with the increase of depth, the problem of gradient vanishing and over fitting will appear in the network.

3 Proposed Method

To solve the problem of gradient vanishing or gradient explosion when simply deepening the network, He proposed a deep convolutional residual network in 2015—-ResNet50 [20]. The ResNet50 network is built by residual modules. This type of modules not only makes the network free of degradation problems, but also reduces the error rate, and keeps the computational complexity at a very low level. Zhu [21] used the pre-trained ResNet50 to determine the invasion depth of gastric cancer. ResNet50 could be understood as a 5-stage one in total. After the first

stage, the image was transformed into a tensor of $112 \times 112 \times 64$. Then residual modules were introduced between Stage 2 and Stage 5 to overcome the problem of gradient disappearance and explosion. After five stages, the input was converted to a tensor of $7 \times 7 \times 2048$, and the dimension increased from 3 to 2048, indicating that information extracted was more than the original RGB pixels.

Finally, according to the extracted 2048 features, the tensor was flattened into to a 4096-dimensional vector and then sent to softmax for classification. Taking the tumor classification in this paper as an example, let the final classification number K be five, and give a training set $\{(x^1, y^1), (x^2, y^2), \ldots, (x^m, y^m)\}$, where $y^i \in \{1, 2 \ldots k\}$. Now, given an input test data x, softmax estimates the probability of each class by calculating the corresponding probability value $p(y=j|x)$ for each class, and finally outputs a k-dimensional normalized vector:

$$
h_\theta(x^i) = \begin{bmatrix} p(y=1|x^i;\theta) \\ p(y=2|x^i;\theta) \\ \cdots \\ p(y=k|x^i;\theta) \end{bmatrix} = \frac{1}{\sum_{j}^{k} e^{\theta_j^T x^i}} \begin{bmatrix} e^{\theta_1^T x^i} \\ e^{\theta_2^T x^i} \\ \cdots \\ e^{\theta_k^T x^i} \end{bmatrix} \tag{1}
$$

where $\theta_1, \ldots, \theta_k \in \Re^{n+1}$ is the model parameters, the final output vector of each element corresponding to the corresponding probability estimate of a tumor category, the sum of all elements is 1.

Due to the excellent structure of the residual network, ResNet50 was selected for the experiment and two improvements were made to it. The improved network structure is shown in Fig. 1.

In the literature [22], an inception v1 module was proposed, which improved in network width. Inspired by the inception v1 module, this paper proposed a multi-size convolutional kernel module similar to inception v1, as is shown in Fig. 1b, to replace the 7×7 convolution layer of Stage 1 in the ResNet50 network. This is because convolution kernels of different sizes can extract features of different sizes, and can also take the details of the surrounding area of each pixel and the features of the larger background (for example, whether it is close to the skull) into account [23]. This makes the features input into the residual module are more comprehensive and helps to improve the classification accuracy of the network. The difference between multi-size convolutional kernel module and inception v1 is that it used add( ) function instead of the concat( ) function to connect the various branches.

Suppose that the input characteristic dimension of the multi-size convolutional kernel module is $W \times H$, the number of channels in each branch of the module is C, and the output characteristic dimension after concat( ) is O_c, and the output characteristic dimension after using add( ) is O_a:

$$
O_c = W \times H \times 4C \tag{2}
$$

$$
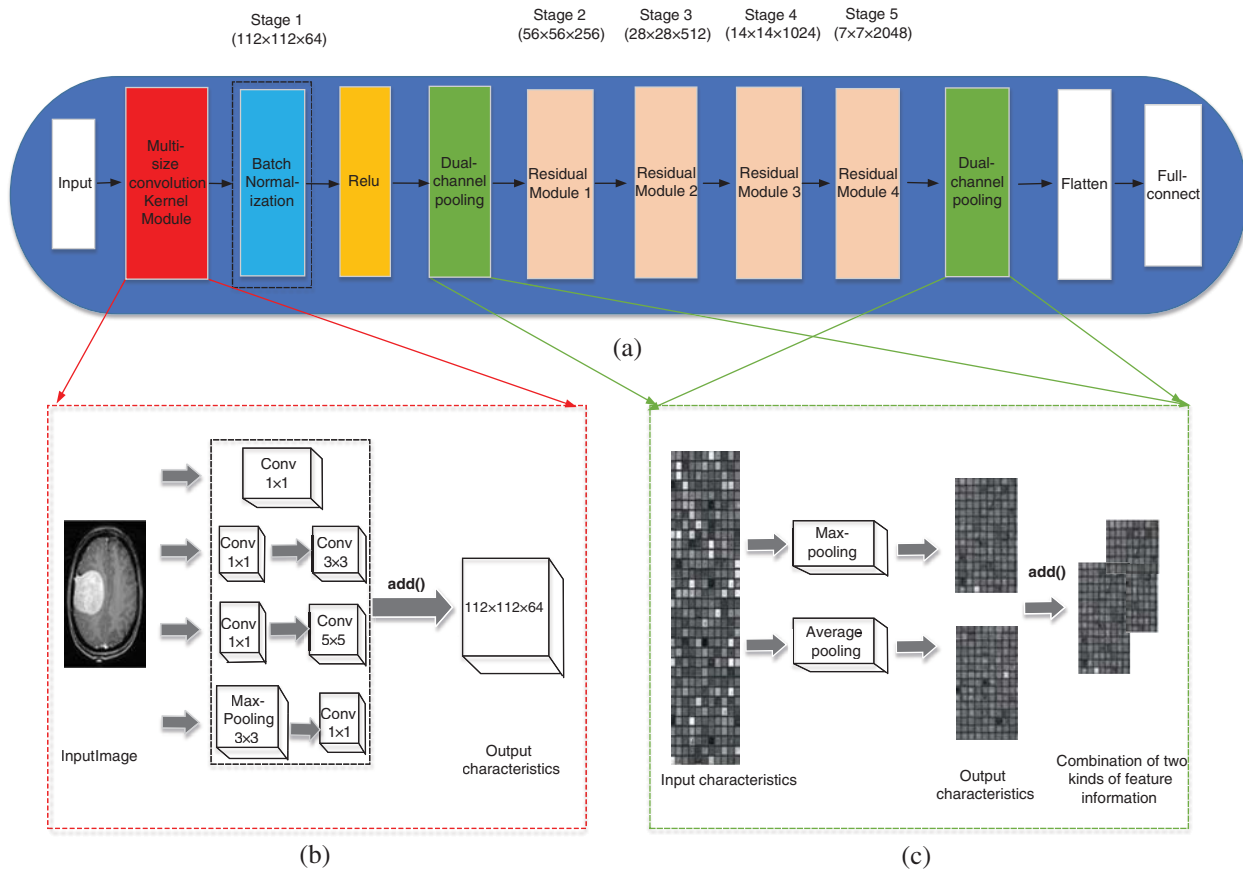O_a = W \times H \times C \tag{3}
$$

Figure 1: The improved network structure. (a) Improved ResNet50 network structure, (b) Multi-size convolutional kernel module, (c) Dual-channel pooling layer

The concat( ) function only adds the dimension describing the features of the image itself when used for the combination of channels, and the feature information under each dimension is not added. Although the add() function does not add the dimensions describing the features of the image itself, the amount of information under each dimension increases, which is obviously beneficial to the classification of the final images.

In ResNet50 network, there is only a pooling layer after Stage 1 and Stage 5, which are max-pooling and average-pooling respectively. The pooling layer can be understood as extracting significant features once again, and then obtaining more advanced features after repeated stacking. This operation can reduce the amount of data processing while retaining useful information, thus achieving feature dimension reduction, data compression and over-fitting reduction. Average-pooling can reduce the error caused by the increase of the estimated variance due to the size limit of the neighborhood and save most of the background information. Max-pooling focusing on texture information can balance the deviation of the estimated mean value caused by convolution parameter error [24]. Although CNN has been applied to the classification of brain tumors and proved to be effective, the traditional single pooling layer has limited domain for the reception of local context feature information. This traditional network mode may ignore the useful global background information, therefore, this paper tries to combine the two kinds of pooling layer

and change the single pooling layer in ResNet50 network to the dual-channel pooling layer: the features outputted at the previous step pass through the max-pooling layer and average-pooling layer simultaneously by two paths, and the output results of two paths are combined by the add() function to enter the next operation. The schematic diagram of the dual-channel pooling layer is shown in Fig. 1c.

4 Experiment and Discussion

4.1 Dataset and Preprocessing

The dataset used in this paper consisted of healthy brain images and diseased brain images, totaling 555 MRI brain images. The healthy brain images were from www.kaggle.com and 110 MRI images were selected; and the diseased brain images were collected at http://radiopaedia.org/, with a total of 445 Axrial T1 MRI images, including 210 images of meningioma (Grade I), 80 images of diffuse astrocytoma (Grade II), 75 images of anaplastic oligodendroglioma (Grade III), and 80 images of glioblastoma (Grade IV). In this paper, 555 MRI brain images are divided into training set and test set. The number of brain images of each type in training set and test set is 4:1. 444 images were used for training and 111 images were used for testing.

Rich data is the key to build deep learning model effectively. The most common method of data enhancement is to add noise or apply geometric transformation to the image, which helps to prevent over fitting of the network model. Therefore, this paper makes a series of extensions to the images in the dataset:

(1) Unify the size of the image: the size of all pictures is 224×224.
(2) Image rotation and flipping operation: CNN can learn different features when it is placed on images in different directions. Therefore, this paper uses rotation, horizontal flip and vertical flip data enhancement methods so that the same image in the training set can be enhanced to multiple images with the same essence. The operation of image rotation and flipping is described as follows: suppose that in the coordinate axis, the coordinate of a certain point in the image is (x_0, y_0), and becomes (x_1, y_1) after angular rotation a, then:

$$x_1 = x_0 \times \cos a - y_0 \times \sin a \tag{4}$$

$$y_1 = x_0 \times \sin a + y_0 \times \cos a \tag{5}$$

If the height of the image is h, the width is w, and the horizontal flipped coordinate is (x_2, y_2), then

$$\begin{bmatrix} x_1 \\ y_1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 & w \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \\ 1 \end{bmatrix} \tag{6}$$

(3) Add salt and pepper noise: salt and pepper noise is characterized by random black and white pixels propagating in the image, which is similar to the effect produced by adding Gaussian noise to the image, but has a low level of information distortion. Therefore, this paper uses salt and pepper noise as data enhancement method.

4.2 Training Details and Training Strategies

The improved network model based on ResNet50 which proposed in this paper was built using keras library of python, and the training batch size was set to 32 and the epoch was 100.

During the data enhancement experiment, the training images were traversed by the ImageData-Generator, and the random transformed data enhancement was performed on each image. The initial learning rate was set to 0.001, and the weight decay to 0.0002. And we use the stochastic gradient descent (SGD) method to train the improved network model. The expression of SGD is as follows:

Loop{
 for i = 1 to m,{
 $\theta_j := \theta_j + \alpha(y^{(i)} - h_\theta(x^{(i)}))x_j^{(i)}$ (for every j)   (7)
 }
}

where m is the number of samples, α is the learning rate, $x^{(i)}$ is the independent variable, $h_\theta(x^{(i)})$ is the predicted value, $y^{(i)}$ is the actual value, θ_j is the parameter trying to find the minimum objective function. Random gradient descent randomly selects one data to calculate the fastest descent direction instead of scanning all training data sets, which speeds up the iteration speed.

After the training was completed, the model effect was verified using K-fold cross-validation. Cross-validation is a method of model selection by estimating the generalization error of the model [25]. In order to accurately evaluate the quality of the model, a five-fold cross-validation method was used. All the experiments were performed on NVIDIA GeForce RTX 2080 Ti GPU.

4.3 Evaluation Indices

In order to comprehensively evaluate the performance of the network in this paper, the following indicators will be used as the evaluation criteria of the experimental results.

(1) Accuracy: the ratio of the number of correct predictions in the sample to the total number of samples. The calculation formula is as follows:

$$Accuracy = \frac{TP + TN}{TP + TN + FN + FP}$$ (8)

(2) Precision: the ratio of the samples with positive prediction to all samples with positive prediction. The calculation formula is as follows:

$$Precision = \frac{TP}{TP + FP}$$ (9)

(3) Recall rate: it refers to the ratio of positive samples correctly predicted to all positive samples. The calculation formula is as follows:

$$Recall = \frac{TP}{TP + FN}$$ (10)

(4) F1Score: this index takes into account both the precision and recall rate of the classification model, and can be regarded as a weighted average of the precision and recall rate of the model. The calculation formula is as follows:

$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$ (11)

Wherein TP is true positive, TN is true negative, FP is false positive and FN is false negative.

In addition, the loss function (Loss) can also be used to estimate the degree of inconsistency between the predicted value and the real value of the model, and the smaller the loss function is, the better the robustness of the model is. Therefore, this paper uses cross loss quotient to judge the performance of the model. For the sample points (x, y), y is the real label, if there areklabel values and N samples in total, the cross-entropy loss function formula of the data set is:

$$L_{\log}(Y, P) = -\frac{1}{N} \sum_{i=0}^{N-1} \sum_{k=0}^{k-1} y_{i,k} \log p_{i,k} \tag{12}$$

4.4 The Proposed Network and Performance Analysis

In order to achieve the best classification effect of the network in this paper, the size of the convolution kernel on the third branch of the multi-size convolution kernel module is set up and tested in different methods. The experiment is carried out on the first cross validation scheme. In this paper, the following experiments are carried out when the 7×7 single convolution layer in Stage 1 of ResNet50 is improved to a multi-size convolution kernel module. Method A: a commonly used 5×5 convolutional kernel was connected to the behind of 1×1 convolutional kernel. Method B: considering convolutional kernels of different sizes can extract features of different sizes, the commonly used 5×5 convolutional kernel behind the 1×1 convolutional kernel in the third branch was replaced with 7×7 convolutional kernel tentatively. Method C: considering that replacing large convolutional kernels with two small convolutional kernels could reduce the number of parameters and speed up computation, two 3×3 convolutional kernels were used to replace the 5×5 convolutional kernels behind the 1×1 convolutional kernel in the third branch tentatively. The specific parameters and experimental results of the multi-size convolutional kernel module are shown in Fig. 2.
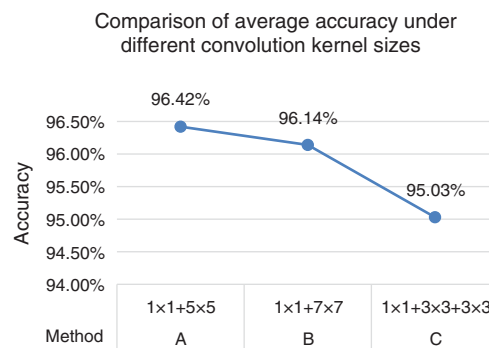

Figure 2: Comparison of average accuracy under different convolution kernel sizes

As indicated by Fig. 2, the combination methods of A, B in the third branch of the multi-size convolutional kernel module were more accurate than C combination, because a 3×3 convolutional kernel already existed in the 2nd and 4th branch, which verified that convolutional kernels of different sizes can extract features of different sizes and that richer feature information was beneficial to improve the classification accuracy. Among them, the average accuracy of A combination method was higher than that of B. This was because the tumor features on the MRI image were subtle, a small convolutional kernel was more suitable for extracting more useful

information. According to the experimental results, this paper selects A combination method with the highest accuracy.

In order to determine the optimal method of adding dual-channel pooling layer, the following experiments are carried out under the condition that the multi-size convolution kernel module is determined as the combination method in method A in Fig. 2, which is carried out on the first cross validation scheme. Experiment 1: The average-pooling layer after Stage 5 in ResNet50 was kept unchanged, and only the max-pooling layer after Stage 1 was improved into dual-channel pooling layer. Experiment 2: The max-pooling layer after Stage 1 in ResNet50 was kept unchanged, and only the average-pooling layer after Stage 5 was improved into dual-channel pooling layer. Experiment 3: The single pooling layer after Stage 1 and Stage 5 were simultaneously improved into dual-channel pooling layer and the experimental results after such improvement are shown in Fig. 3.
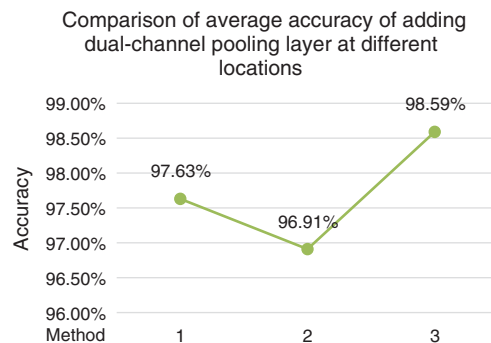
Comparison of average accuracy of adding
dual-channel pooling layer at different
locations

Figure 3: Comparison of average accuracy of adding dual-channel pooling layer at different locations

Fig. 3 shows that changing the max-pooling layer after Stage 1 into dual-channel pooling layer has a better classification effect than changing the average-pooling layer after Stage 5 into dual-channel pooling layer, which reflects that the average-pooling plays a more important role in complex brain tissue classification than max-pooling, and reveals the irreplaceability of global information in complex medical images. The experiment 3 changes the single pooling layer after Stage 1 and Stage 5 in the ResNet50 network into dual-channel pooling layer simultaneously, with a average classification accuracy reaching up to 98.59%, which indicates that the dual-channel pooling layer maximizes the advantages of max-pooling and average-pooling, and makes the high-dimensional feature information extracted by the two complement each other to make up for their respective shortcomings.

4.5 Performance Comparison of Training Set Before and After Data Enhancement Between ResNet50 and the Improved Network

In ResNet50 network and the improved network, two sets of experiments were carried out with two training sets before and after data enhancement, and the test set images were used as validation. The training set curve is drawn as shown in Fig. 4, where the dotted line represents the training curve without the operation of data enhancement. The experimental result is shown in Tab. 1.
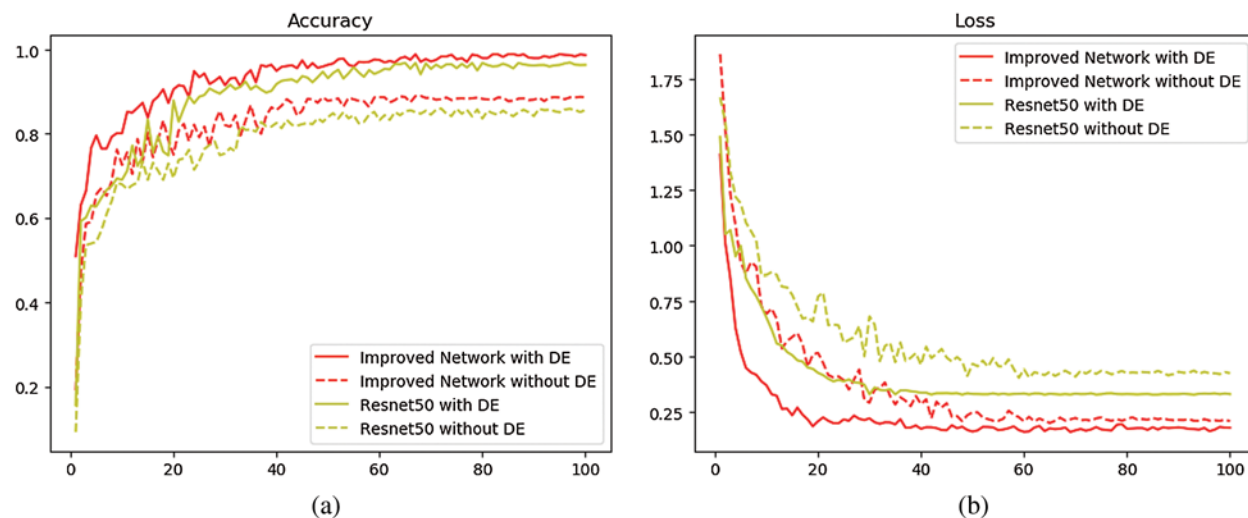
Figure 4: Comparison of the average accuracy and loss function of ResNet50 and improved network (a) Comparison of average classification accuracy between ResNet50 and improved network (b) Comparison of loss function decline between ResNet50 and improved network

Table 1: Comparison of average accuracy of two methods with or without data enhancement

Method	Without data enhancement	With data enhancement
ResNet50	85.71%	95.37%
Improved network	88.62%	98.59%

Tab. 1 and Fig. 4 show that, when the training set does not undergo data enhancement and is directly input into the network model for training, the average classification accuracy of ResNet50 network and the improved network is 85.71% and 88.62% respectively, while if the operation of data enhancement is performed on the training set, the average classification accuracy of ResNet50 is 95.37% and that of the improved network is 98.59%, both increasing by about 10%. This proves that the data enhancement technique has a positive impact on improving the classification accuracy. This is because the convolutional neural network requires a large amount of data to extract features and the stronger the ability of network learning features, the stronger the classification ability in the face of new datasets. At the same time, it is found that after 100 iterations, when the same small training sets are available, the average classification accuracy of the improved network is 2.91% higher than that of ResNet50 and the loss function is about 0.12 lower than that of ResNet50; when the same large training sets are available, the average classification accuracy of the improved network is 3.22% higher than that of the ResNet50, and the loss function is about 0.15 lower than that of the ResNet50, indicating that the features extracted by multi-size convolutional kernel module are richer than those extracted by the single convolutional layer, and that compared with the single pooling layer, the dual-channel pooling layer can combine the advantages of both max-pooling and average-pooling. All these are beneficial to improve the final classification performance. The above experimental results prove that the improvement based on ResNet50 is meaningful, and the improved network has better classification performance than ResNet50.

4.6 Comparison with Other Networks

In order to verify the significance of the improvement of ResNet50, the experimental comparison between the improved network and AlexNet, VGGNet, GooleNet and ResNet50 is carried out. The experimental results are shown in Tab. 2, and the values in the table are classified average values.

Table 2: Comparison of precision, recall, F1Score and accuracy of different networks

Network	Precision (%)	Recall (%)	F1Score (%)	Accuracy (%)
AlexNet	65.36	67.24	66.25	84.27
VGGNet	70.43	73.33	71.82	87.09
GooLeNet	76.1	77.74	76.72	89.04
ResNet50	90.3	91.85	91.04	95.37
Proposed network	96.97	96.77	96.87	98.59

The AlexNet model [26] is an eight-layer network. Its first five layers are convolutional layers, the last three are fully connected layers, and the last fully connected layer has 1000 classes of output. VGGNet is a 19-layer deep network model, including 16 convolutional layers, 2 fully connected image feature layers and 1 fully connected classification feature layer, and all convolutional kernels in the network are uniform in size, being 3×3. GooLeNet [20] is a 22-layer deep network model. Its biggest feature is the introduction of the inception structure, which solves the problem of large network parameters and high resource consumption. Five-fold cross-validation was performed on the training results of these five different network models. The following experimental result analysis is the average result of five validations.

It can be found from the above results that due to the close relationship between ResNet50 and the improved network, they can effectively solve the vanishing gradient problem by adding the residual module, which greatly reduces the number of parameters and increases the utilization of features. Both models outperform AlexNet, VGGNet and GoogLeNet in terms of precision, recall, F1Score and accuracy. Among them, AlexNet has the worst classification effect. However, the improved network has the highest results with regard to the four indices. The average classification precision of the improved network reaches 96.97%, and the average classification accuracy reaches 98.59%. The results show that the improved network can extract more features and it has effectively combined the advantages of max-pooling and average-pooling, so that the background information and the detail information extracted by the two can complement each other, thus the final recognition rate is relatively high.

5 Conclusions

For the multi classification of brain tumors, an improved convolutional neural network based on ResNet50 is proposed in this paper. The network uses multi-scale convolution core module and dual channel pooling layer to improve the single 7×7 convolution layer and single pooling layer in ResNet50. The multi size convolution core module is composed of convolution branches with different size convolution cores, which can extract rich feature information from the input image; the dual channel pooling layer combines the advantages of maximum pooling and average pooling, so that the extracted detail information and background information can complement each

other. Accuracy, precision, recall and f1score were used to evaluate the network performance, and the five-fold cross validation method was used to analyze the classification effect of the network. Finally, the average classification accuracy of healthy brain, meningioma, diffuse astrocytoma, anaplastic oligodendroglioma and glioblastoma was 98.59%.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1]   S. Yazdani, R. Yusof, A. Karimian, M. Pashna and A. Hematian, "Image segmentation methods and applications in MRI brain images," *IETE Technical Review*, vol. 32, no. 6, pp. 413–427, 2015.

[2]   J. Jiang, P. Trundle and J. Ren, "Medical image analysis with artificial neural networks," *Computerized Medical Imaging and Graphics*, vol. 34, no. 8, pp. 617–631, 2010.

[3]   D. J. Hemanth, C. K. S. Vijila, A. I. Selvakumar and J. Anitha, "Performance improved iteration-free artificial neural networks for abnormal magnetic resonance brain image classification," *Neurocomputing*, vol. 130, pp. 98–107, 2014.

[4]   X. F. Li, Y. B. Zhuang and S. X. Yang, "Cloud computing for big data processing," *Intelligent Automation & Soft Computing*, vol. 23, no. 4, pp. 545–546, 2017.

[5]   J. Cheng, R. M. Xu, X. Y. Tang, V. S. Sheng and C. T. Cai, "An abnormal network flow feature sequence prediction approach for DDoS attacks detection in big data environment," *Computers, Materials & Continua*, vol. 55, no. 1, pp. 95–119, 2018.

[6]   Y. Lecun, Y. Bengo and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[7]   G. Mohan and M. M. Subashini, "MRI based medical image analysis: Survey on brain tumor grade classification," *Biomed Signal Process Control*, vol. 39, pp. 139–161, 2018.

[8]   S. Pereira, A. Pinto and V. Alves, "Silva CA brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Transactions on Medical Imaging*, vol. 35, no. 5, pp. 1240–1251, 2016.

[9]   M. Havaei, A. Davy, D. Warde-Farley, A. Biard, A. Courville *et al.*, "Brain tumor segmentation with deep neural networks," *Medical Image Analysis*, vol. 35, pp. 18–31, 2017.

[10] J. Kleesiek, G. Urban, A. Hubert, D. Schwarz, K. Maier-Hein *et al.*, "Deep MRI brain extraction a 3D convolutional neural network for skull stripping," *Neuroimage*, vol. 129, pp. 460–469, 2016.

[11] M. Elhoseny, A. Abdelaziz, A. S. Salama, A. M. Riada, K. Muhammade *et al.*, "A hybrid model of internet of things and cloud computing to manage big data in health services applications," *Future Generation Computer System*, vol. 86, pp. 1383–1394, 2018.

[12] I. Mehmood, M. Sajjad, K. Muhammad, S. I. A. Shah, A. K. R. Sangaiah *et al.*, "An efficient computerized decision support system for the analysis and 3D visualization of brain tumor," *Multimedia Tools and Applications*, vol. 78, pp. 12723–12748, 2018.

[13] K. L. C. Hsieh, R. J. Tsai, Y. C. Yeng and C. M. Lo, "Effect of a computer-aided diagnosis system on radiologists' performance in grading gliomas with MRI," *PloS One*, vol. 12, no. 2, pp. e0171342, 2017.

[14] G. J. Litjens, J. O. Barentsz, N. Karssemeijer and H. J. Huisman, "Clinical evaluation of a computer-aided diagnosis system for determining cancer aggressiveness in prostate MRI," *European Radiology*, vol. 25, pp. 3187–3199, 2015.

[15] S. H. Wang, P. Phillips, Y. Sui, B. Liu, M. Yang *et al.*, "Classification of Alzheimer's disease based on eight-layer convolutional neural network with leaky rectified linear unit and max pooling," *Journal of Medical Systems*, vol. 42, no. 5, pp. 85, 2018.

[16] M. K. Abd-Ellah, A. I. Awad, A. Khalaf and H. F. A. Hamed, "Two-phase multi-model automatic brain tumor diagnosis system from magnetic resonance images using convolutional neural networks," *EURASIP Journal on Image and Video Processing*, vol. 1, pp. 1–10, 2018.

[17] P. Abdolmaleki, F. Mihara, K. Masuda and L. D. Buadu, "Neural networks analysis of astrocytic-gliomas from MRI appearances," *Cancer Letters*, vol. 118, pp. 69–78, 1997.

[18] M. Sajjad, S. Khan, K. Muhammad and W. Wu, "Multi-grade brain tumor classification using deep CNN with extensive data augmentation," *Journal of Computational Science*, vol. 30, pp. 174–182, 2018.

[19] N. Ghassemi, A. Shoeibi and M. Rouhani, "Deep neural network with generative adversarial networks pre-training for brain tumor classification based on MR images," *Biomedical Signal Processing and Control*, vol. 57, pp. 101678, 2020.

[20] K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *2016 IEEE Conf. on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 770–778, 2016.

[21] Y. Zhu, Q. C. Wang, M. D. Xu, Z. Zhang, J. Cheng *et al.*, "Application of convolutional neural network in the diagnosis of the invasion depth of gastric cancer based on conventional endoscopy," *Gastrointestinal Endoscopy*, vol. 89, no. 4, pp. 806–815, 2019.

[22] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed *et al.*, "Going deeper with convolutions," in *2015 IEEE Conf. on Computer Vision and Pattern Recognition*, Boston, MA, USA, pp. 1–9, 2015.

[23] H. Lee and H. Kwon, "Going deeper with contextual CNN for hyperspectral image classification," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4843–4855, 2017.

[24] J. Chang, L. Zhang, N. Gu, X. Zhang, M. Ye *et al.*, "A mix-pooling CNN architecture with FCRF for brain tumor segmentation," *Journal of Visual Communication and Image Representation*, vol. 58, pp. 316–322, 2019.

[25] H. Li, S. S. Zhuang, D. A. Li, J. M. Zhao and Y. Y. Ma, "Benign and malignant classification of mammogram images based on deep learning," *Biomedical Signal Processing and Control*, vol. 51, pp. 347–354, 2019.

[26] A. Krizhevsky, I. Sutskever and G. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS'12: Proc. of the 25th Int. Conf. on Neural Information Processing Systems*, vol. 1, pp. 1097–1105, 2012.