MACHINE LEARNING NONSEGMENTED GLIOMAS DATA EXTRACTION

Data collection

The data including clinical records and longitudinal DTI of 84 GBM patients (23 with PsP and 61 with TTP) were collected at the Wake Forest School of Medicine, the USA. DTI images were reconstructed in the matrix of 218×182 pixels.

Data Augmentation

Applied a series of affine transformations to the valid images (total 1348 slices). An affine transformation is any transformation that preserves collinearity (i.e., all points lying on a line initially still lie on a line after transformation) and ratios of distances (e.g., the midpoint of a line segment remains the midpoint after transformation). First, all images are translated 5 pixels, 10 pixels, 15 pixels, 20 pixels and rotated by 3°, 6°, 9°, 12°, 15°respectively. Next, above 9 transformations and original images are rotated by 1°, 2°, 3°, 4°, 5° separately. So, the dataset was expanded dataset to 50 times its initial scale.

Data Normalization

In order to eliminate the bias which is caused by different equipment, they applied the Z-score [29] method to normalize above dataset. After normalization, each slice in the dataset is transformed into the same magnitude, and it is suitable for further analysis and evaluation. Finally, they randomly divide one tenth of the dataset as the testing set, and the other part is regarded as the training set.

Architectures

VGG:

VGG is the first design to explore the relationship between the architecture of CNN and its depth. The depth of VGG can be added steadily by using very small convolution filters. For all convolution layers (Conv.), the size of convolution kernel is fixed to 3×3, which is considered as the smallest size to capture pixel information from left, right, up, down and center. Spatial pooling is finished by max-pooling layers (MP) with 2×2 pixel window. Similarly, the main architecture is followed by fully connected layers (FC) and a softmax layer. In this study, they applied a 17-layer VGG to diagnose PsP and TTP. The configuration of layers and corresponding channels are displayed in TABLE I.

TABLE I. Configuration of employed VGG

Layer	Type	Channels	Layer	Type	Channels	Layer	Type	Channels
1	Conv.	218×182×16	7	Conv.	54×45×64	13	Conv.	27×22×128
2	Conv.	218×182×16	8	Conv.	54×45×64	14	MP	--
3	MP	--	9	Conv.	54×45×64	15	FC	32
4	Conv.	109×91×32	10	MP	--	16	FC	2
5	Conv.	109×91×32	11	Conv.	27×22×128	17	Softmax	--
6	MP	--	12	Conv.	27×22×128			

ResNet:

Although VGG realized a deeper network, some studies 26, 25 reported that there exposed some problems in such architecture. For example, the accuracy usually degrades rapidly after saturation with the increasing of depth. To solve this degradation problem, a deep residual learning framework which called ResNet is proposed.

They used a ResNet with six residual learning blocks to classify PsP and TTP. The kernel size of all the convolution layers is 5×5. After that, following a max-pooling layer, a fully connected layer and a softmax layer. The configuration of ResNet and corresponding channels are displayed in TABLE II

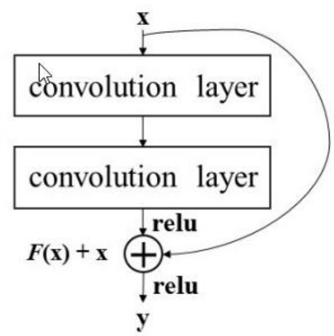

Fig.2. Residual learning block

TABLE II. Configuration of employed ResNet

Block	Layer	Type	Channels	Block	Layer	Type	Channels
1	1	Conv.	218×182×4	5	9	Conv.	14×12×64
	2	Conv.	218×182×4		10	Conv.	14×12×64
2	3	Conv.	109×91×8	6	11	Conv.	14×12×64
	4	Conv.	109×91×8		12	Conv.	14×12×64
3	5	Conv.	55×46×16	--	13	MP	--
	6	Conv.	55×46×16	--	14	FC	2
4	7	Conv.	28×23×32	--	15	Softmax	--
	8	Conv.	28×23×32				

DenseNet:

Instead of exploring extreme deep networks, DenseNet 15 exploits the potential of neural network by feature reuse. The proposed dense block (Fig. 3) connects each layer to every other layer. That is to say, the feature maps of all layers are used as inputs for subsequent layers

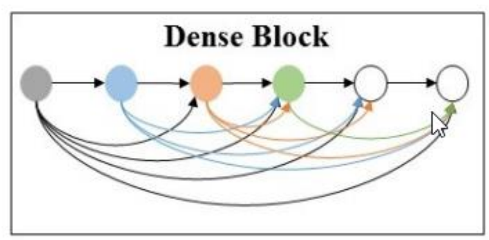

Fig.3. A dense block with six layers

We built a dense block with six convolution layers as Fig. 3. And we then set up a DenseNet with three dense blocks and two transition combinations to distinguish PsP and TTP. After that, following an average-pooling (AP) layer, a fully connected layer and a softmax layer. The configuration of DenseNet and corresponding channels are displayed in TABLE III.

Block	Layer	Type	Channels	Block	Layer	Type	Channels
Dense1	1	Conv.	218×182×8	Dense2	13	Conv.	109×91×44
Dense1	2	Conv.	218×182×12	Dense2	14	Conv.	109×91×48
Dense1	3	Conv.	218×182×16	Transition2	15	Conv.	109×91×48
Dense1	4	Conv.	218×182×20	Transition2	16	AP	55×46×48
Dense1	5	Conv.	218×182×24	Dense3	17	Conv.	55×46×48
Dense1	6	Conv.	218×182×28	Dense3	18	Conv.	55×46×52
Transition1	7	Conv.	218×182×28	Dense3	19	Conv.	55×46×56
Transition1	8	AP	109×91×28	Dense3	20	Conv.	55×46×60
Dense2	9	Conv.	109×91×28	Dense3	21	Conv.	55×46×64
Dense2	10	Conv.	109×91×32	Dense3	22	Conv.	55×46×68
Dense2	11	Conv.	109×91×36	--	23	FC	68
Dense2	12	Conv.	109×91×40	--	24	Softmax	2

Multi-Scale ensemble Network:

To take advantage of the expert knowledge and every CNN, we present an ensemble network to integrate the useful features. (Shown in Fig. 7). As displayed in TABLE I, TABLE II and TABLE III, the number of the three groups of feature maps is 64. And the size of feature map in relevant layers are 54×45 of the 7th layer in VGG, 14×12 of the 10th layer in ResNet and 55×46 of the 21st layer in DenseNet. These feature maps are taken from different convolutional levels. In order to fuse multi-level information from different down-sampling scale, we generate a group of multi-scale feature maps.

Feature visualization for PsP and TTP

Above three CNNs can realize stratification of PsP and TTP just on image-level label. That is to say, such networks have ability to distinguish subtle differences between brain lesion without region detection or segmentation. The authors have reason to believe that there are discriminative image regions to help CNNs identify the particular category. Therefore, feature visualization is a crucial step to raise the performance and reliability of deep networks.

Global average pooling (GAP) is a vectorized strategy to replace fully connected layers in deep networks. GPA can effectively decrease redundant information and avoid overfitting in fully connected. More importantly, the feature maps can be easily converted into categories confidence maps by GPA. As shown in Fig.4, GAP takes the average value of each feature map for every convolutional layer instead of directly concatenating them as fully connected layers do.

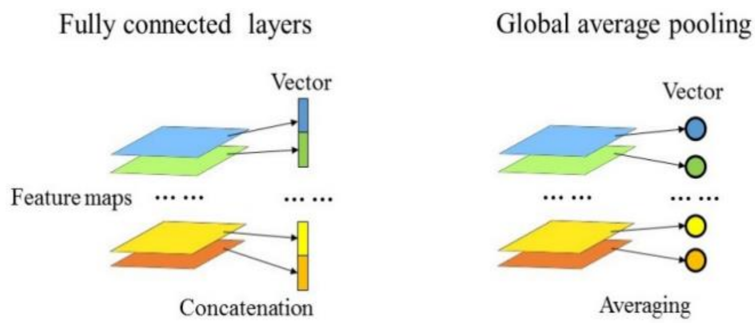

Fig.4. The comparison between fully connected and GAP

Results

Environment:

The complete experiments were performed on a workstation of Windows10 64-bit operating system with 16 GB memory and an NVIDIA GeForce GTX 1080 graphics card. The training and testing for VGG, ResNet, DenseNet and multi-scale ensemble network were developed on the DL library Keras with TensorFlow backend. The other algorithms were carried out on Python 3.5

Hyper-parameters:

In the training of VGG, optimization function is Adam with batch-size=50 and learning rate=0.001. In the training of ResNet, optimization function is Adam with batch-size=40 and learning rate=0.01. In the training of DenseNet, optimization function is Adam with batch-size=30 and learning rate=0.005. In the training of multi-scale ensemble network, stochastic gradient descent (SGD) is used as optimization function. The momentum of SGD is 0.9. The learning rate of 0.01 are considered training parameters.

Experiment results:

In this research, they realize the diagnosis of PsP and TTP with a total four deep architectures: VGG, ResNet, DenseNet and the presented multi-scale ensemble networks. The accuracy of classification, sensitivity, specificity and the area under curve (AUC) are used to evaluate the performances of these networks.

Network	Accuracy	Sensitivity	Specificity	AUC
VGG	86.24%	97.29%	64.59%	0.95
ResNet	87.68%	99.44%	64.63%	0.98
DenseNet	88.02%	99.24%	66.04%	0.98
Ensemble Network	90.20%	91.26%	88.18%	0.99

TABLE VI. The number of features for four networks

Network	VGG	ResNet	DenseNet	Ensemble Network
Number of features	2,599,104	641,296	9,413,192	162,002