Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11279-z

1210: COMPUTER VISION FOR CLINICAL IMAGES

Alzheimer's disease diagnosis based on long-range dependency mechanism using convolutional neural network

Zhao Pei[1,2] · Yuanshuai Gou[2] · Miao Ma[2] · Min Guo[2] · Chengcai Leng[3] · Yuli Chen[2] · Jun Li[4]

Received: 3 September 2020 / Revised: 19 May 2021 / Accepted: 8 July 2021 /
Published online: 31 July 2021

Abstract

Being able to collect rich morphological information of brain, structural magnetic resonance imaging (MRI) is popularly applied to computer-aided diagnosis of Alzheimer's disease (AD). Conventional methods for AD diagnosis are labor-intensive and typically depend on a substantial amount of hand-crafted features. In this paper, we propose a novel framework of convolutional neural network that aims at identifying AD or normal control, and mild cognitive impairment or normal control. The centerpiece of our method are pseudo-3D block and expanded global context block which are integrated into residual block of backbone in a cascaded manner. To be specific, we transfer pseudo-3D block in the video feature representation to extract structural MRI features. Besides, we extend the 2D global context block to the 3D model which can effectively combine the features and capture the latent associations, while simulate the global context in each dimension of structural MRI results. With the preprocessed structural MRI used as the input of the overall network, our method is capable of constructing a latent representation with multiple residual blocks to promote the classification accuracy. Intrinsically, our method reduces the complexity of conventional 3D convolutional network model applied to AD diagnosis and improves the classification accuracy over the baseline. Furthermore, our network can fully take advantage of the deep 3D convolutional neural network for automatic feature extraction and representation, and thus avoids the limitation of low processing efficiency caused by the preprocessing procedure in which a specific area needs to be annotated in advance. Experimental results on Alzheimer's Disease Neuroimaging Initiative database indicate that our proposed method reports accuracy of 89.29% on the AD/NC and 87.57% on the mild cognitive impairment/NC, whilst our approach achieves the 0.5% improvement of accuracy compared with the backbone.

Keywords Alzheimer's disease diagnosis · Long-range dependency mechanism · Convolutional neural network · Classification

✉ Zhao Pei
zpei@snnu.edu.cn

Extended author information available on the last page of the article.

1 Introduction

Alzheimer's disease (AD) is a common progressive neurodegenerative disorder characterized by the change of key parts like temporal lobe, hippocampus, parahippocampal gyrus, cingulate gyrus, thalamus, precuneus, insula, amygdala, fusiform gyrus and medial frontal cortex in our brain [11, 34]. In the process of AD formation, neurons related to AD in the brain gradually atrophy and disappear [28]. While AD has horribly sprouted in somebody for lengthy stretches of time, it is difficult for a non-expert to realize its existence, which results in the deterioration of the AD status. Although effective drugs have not been developed, fortunately, massive research efforts home in on capturing underlying representation for AD's early diagnosis and treatment [14, 18]. It is widely acknowledged that normal control (NC) populations are those who do not have any representations in structural magnetic resonance imaging (MRI), while most AD patients reveal specific symptoms based on destroyed magnitude to their health. In other words, clear clinical symptoms can be reflected by medical images generated from medical imaging technology such as MRI [29, 31]. It is expected that early AD diagnosis is beneficial for the elderly and particularly those who are prone to AD. In this paper, we aim to use the convolutional neural network(CNN) to automatically recognize structural MRI and classify them as NC or AD [1, 15, 30].

With the great progress in machine learning, more and more efforts are devoted to introducing the machine learning methods into research over the AD diagnosis, leading to significant improvement in NC or AD classification performance [4, 5, 20, 26]. It is well known that medical images benefit the diagnosis of many diseases with the help of the computer in the current era of information technology. Particularly, they are indispensable for AD's diagnosis since medical images are capable of imaging these changes intuitively. In terms of structural MRI classification, CNN based on computer-aided implementation achieves far superior performance to the human diagnosis based on empirical observation and experience. Actually, highly proficient physicians are insensitive to the areas with uniform characteristics in structural MRI when they work for a long time with low efficiency and inaccurate classification. Nevertheless, the drawback the conventional manual methods suffer can be largely overcome by CNN [15, 24] which encodes and decodes image features with sufficient robustness of structural MRI efficiently and accurately. While great success in structural MRI classification tasks is fueled by CNN-based methods, an increasing number of researchers are devoted to combining structural MRI classification tasks with CNN, since it is difficult to obtain desirable performance due to the inherent 3D property of structural MRI.

Although 3D CNN is capable of capturing the original high-dimensional features of structural MRI, the 3D CNN model with a huge number of trainable parameters is usually complex, and thus incurs expensive computational costs and tremendous memory footprint. Inspired by [22] which proposed a pseudo-3D (P3D) residual network in the field of video classification and understanding, we utilize the decoupled 3D convolution in pseudo-3D ResNet to substitute 3D convolutions in the bottleneck blocks of the backbone network. Thus, the 3D structural MRI is decomposed and the 2D feature map is obtained. In this way, we present a novel framework for classifying structural MRI collected from patients with Alzheimer's disease as NC or AD. To be specific, our method combines the advantages of 2D convolution and 3D convolution to capture potential 3D structural information of a series of related areas to AD in the brain. These information plays a crucial role in the final

AD classification [28]. The experiments demonstrate our framework is effective in accurately identifying AD based on the whole-brain structural MRI. Furthermore, our method enjoys desirable generalization capability and exhibits promising performance on the mild cognitive impairment (MCI)/NC.

Since the formation of AD is related to many areas of the brain [11], recognizing the relationship among these areas is conducive to more accurate identification of AD disease. To further enhance representation learning of our network, we introduce global context (GC) block [2] into our framework. GC block can effectively combine the features and capture the latent associations, and simulate the global context in each dimension of structural MRI. Simplifying the non-local (NL) block [27] to obtain simplified non-local (SNL) block, GC block encodes SE block [9] and SNL block into three steps, and selects the best performing one from the two parts in each step. Compared with NL block, GC block optimizes the computational process without sacrificing the accuracy. Besides, it enhances the ability of feature expression without sacrificing the original lightweight property compared with SE block. In this paper, the GC block is incorporated into our model to realize the long-range feature representation. It highlights the characteristics of regions related to AD while downplays the other irrelevant regions, and thus greatly improves the diagnosis accuracy.

In this paper, we attempt to improve the accuracy for AD diagnosis. Given a 3D structural MRI input, our framework decouples 3D convolution into 2D convolution plus 1D convolution in different dimensions. Then, expanded GC block is used to generate the feature map with sufficient descriptive power from previous module for further performance improvement. To sum up, we propose three main contributions:

First, we propose a novel framework with global context pseudo-3D module (GCP module). Particularly, GCP module can be applied to classify structural MRI as NC or AD without empirical experience involved. With the backbone combined with P3D module, our model allows fast and efficient AD diagnosis. Therefore, it effectively reduces the complexity of the conventional 3D convolution network model and alleviates the over-fitting problem in the classification of structural MRI on AD.

Second, we extract rich representation of structural MRI relevant to AD in brain which is of vital importance to the final AD classification. Overall, our method achieves significantly better performance by utilizing GC block.

Finally, we avoid selecting a specific area in advance and improves the processing efficiency by giving full play to the strong ability of automatic feature extraction and abstraction of our network.

2 Related work

With the dramatic progress of artificial intelligence(AI) in various fields, Learning based methods included in AI technology are widely applied to AD diagnosis. These methods can be divided into two categories: traditional machine learning methods [13, 17, 19, 23, 32, 35] and recent deep learning methods [3, 8, 12, 25]. Traditional methods aim to extract the features of pre-selected regions by using statistical methods, and then achieve image classification by extracting the features of these regions. They have strong capability of feature learning and have been widely used in the AD classification.

According to the scale of feature representation, traditional machine learning methods can be divided into three categories: voxel-based approaches, region-based approaches and

patch-based approaches [14]. Comparing the AD-related regions between the experimental group and the healthy control group, the medial temporal structure and hippocampal volume are related to the early markers of AD in [23]. Therefore, imaging biomarkers play an important role in AD diagnosis. Zhu et al. [35] proposed a novel feature selection method, which applied the relationship information inherent in the observation data to the diagnosis of Alzheimer's disease. Liu et al. [19] proposed a hierarchical ensemble classification method. Specifically, they solve the classification problem of local and cross-brain region related features in AD using bottom-up and local to global approach by segmenting brain images. However, traditional machine learning methods usually depend on the human experience, which often results in suboptimal performance.

Recent years have witnessed the boom of deep learning in feature extraction. It has been applied to AD diagnosis and accurate matching between the 3D depth network and the 3D images obtained by MRI. After analyzing structural MRI and classified AD based on gray matter density of hippocampus volume and region of interest, Cheng et al. [3] proposed a new method to extract features from local image blocks with multiple neural networks, revealing promising performance for AD classification. Borrowing the idea of decomposing large-scale problems into small-scale tasks, Lian et al. [14] proposed a hierarchical integrated classification method. After a brain image is divided into several local brain regions, local image features generated from local brain regions and regional correlation are combined and used as the input of advanced feature extractor, leading to better classification results. In [12], a novel method is proposed for AD diagnosis using 3D-CNN, It is capable of learning to capture the general characteristics of AD biomarkers and adapting to data sets in different fields. Suk et al. [25] combined sparse regression model with deep neural network for the first time, proposing a method for clinical AD diagnosis. It selects different feature subsets by multiple sparse regression models and take the results of sparse regression model as the target.

Since the traditional machine learning schemes usually operate on a single region and patch, it is difficult for them to characterize the global structure information of the structural MRI in our work. It is easy to over-fit the simplistic networks, while it is difficult to capture the global context information from the simplistic networks. With GC block integrated, our model couples machine learning and the complex network paradigm for AD/NC or MCI/NC. It effectively extract long-range context, such that the features of related areas in the structural MRI on AD can be fully captured. This ensures our approach works in the tasks of feature extraction and classification related to Alzheimer's disease diagnosis [14]. It is well known that the model complexity and high dimensionality of 3D structural MRI largely hinder hindering the accuracy of AD classification. Massive efforts are devoted to alleviating this problem. Gao et al. [6] extracted 2D slice and 3D feature information of CT by deep learning, and then fuse the two CNN features for subsequent processing. Zhang et al. [33] proposed an unsupervised deep learning method based on Local Deep-Feature Alignment(LDFA). It is capable of learning the local and global features of the data sample set simultaneously. Particularly, our model integrates P3D block simplifies 3D convolutions by 2D convolutional filters on spatial dimension plus 1D convolutions on depth dimension, reducing the computational cost and improving the processing efficiency.

In the rest of the paper, we elaborate how to preprocess the selected data from ADNI and the structure of our model in Section 3. Experimental setup and the analyses of experiment results are presented to evaluate our model in Section 4. The discussion of our framework is shown in the Section 5 before this paper is concluded in Section 6.

3 Materials and methods

In this section, we will give more details of our work about materials and components in the model. The preprocessed steps of materials are introduced in Section 3.1, and the components in the model are elaborated in the rest sections.

3.1 Materials

We train and evaluate our model on the data from the Alzheimer's disease Neuroimaging Initiative[1] (ADNI) [10]. ADNI dataset is launched in 2003 by the National Institute on Aging, the National Institute of Biomedical Imaging and Bioengineering, the Food and Drug Administration, private pharmaceutical companies and nonprofit organizations. It serves as one of most significant databases for AD diagnosis.

We preprocess the data containing 548 examples in which 146 examples belongs to AD, 256 examples indicate NC and the rest are MCI from the ADNI. Similar to [34], the preprocessing process is completed with the property like voxel size for normalized images changed by a series of steps using specialized tools of CAT[2] attached to SPM[3]. We then resliced the image into $224 \times 224 \times 91$.

Figure 1 intuitively shows the CAT imaging of both AD and NC subjects before and after data preprocessing. It can be clearly observed that after registration and segmentation, white matter and gray matter closely related to AD is preserved, which is conducive to improving classification performance. It is worth mentioning that there is the blurry difference between the preprocessed images from different subjects. Therefore, the learning-based AD diagnosis automatically extracting features of data for computer-aided treatment can improve the diagnosis accuracy and reduce the manual burden.

3.2 The overall architecture

The overall architecture of our framework is shown in Fig. 2. Our framework contains a GCP module in which P3D block is combined with GC block used as bottleneck in the backbone. [7] gives detailed introduction of the backbone of ResNet. The GCP module consists of GC extractor with P3D block.

As illustrated in Fig. 2, our framework mainly consists of two components, a backbone based on ResNet and an embedded module combining P3D and GC blocks. We use ResNet as our backbone to classify structural MRI, since it is observed that it can effectively handle the 3D structural MRI with sufficient trainable parameters. Besides, the models with extremely complex structure exhibit excellent capability of learning the representations of the input data, leading to the mismatch between the models and the hand-crafted ADNI database. It is necessary that the input of our framework be distributed within a 3D cube due to the inherent structural MRI 3D property. Finally, the post processing strategy based on the last softmax layer is used for the final prediction.

[1] http://adni.loni.usc.edu/
[2] http://www.neuro.uni-jena.de/cat/
[3] https://www.fil.ion.ucl.ac.uk/spm-statistical-parametric-mapping/

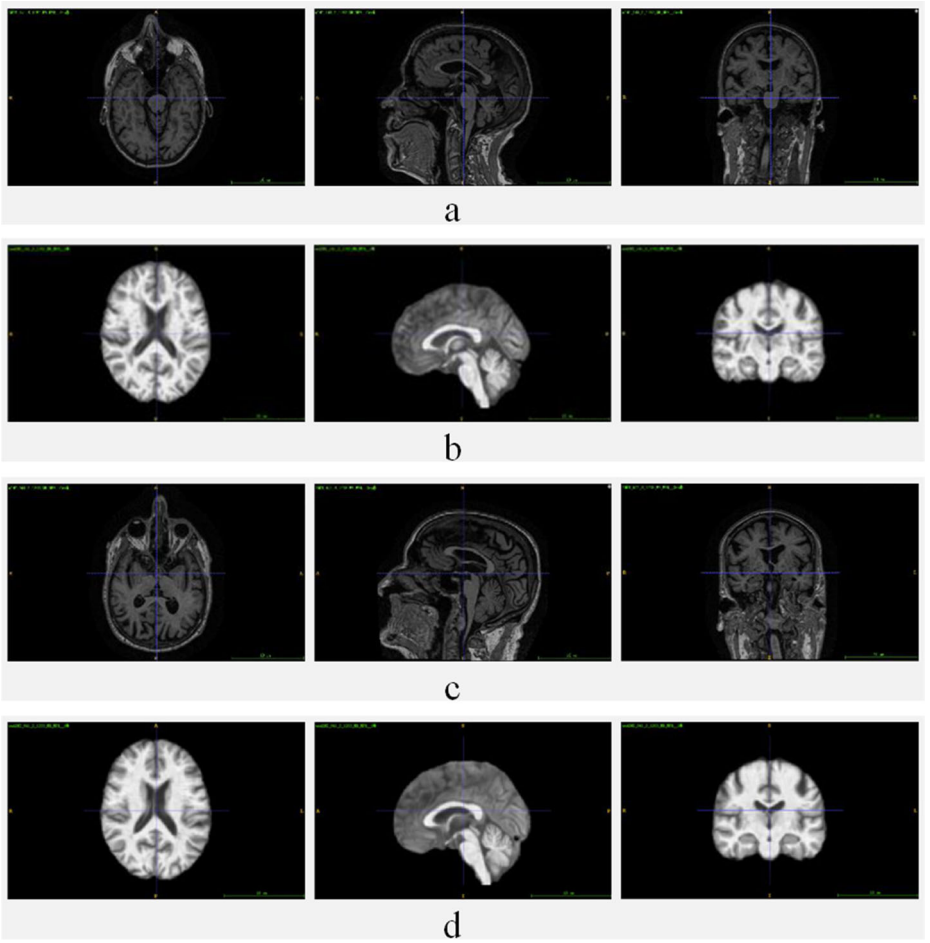

Fig. 1 Illustrative examples before and after data preprocessing. Subgraph **a** and **c** respectively show one AD and NC subject, while subgraph **b** and **d** illustrate the preprocessed one corresponding to **a** and **c** separately. In each subgraph, the three windows show transverse plane, sagittal plane and coronal plane of the current 3D sample respectively from left to right

To accelerate the convergence of the training process, the preprocessed MRI image is resized and normalized to obtain a 3D image. The normalization process can be expressed as follows

$$I_{hwd} = \frac{I_{hwd}}{I_{max}} \tag{1}$$

where h, w and d denote the position of sampling point in the MRI image I, and I_{max} indicates the maximal value in the structural MRI. The input of our framework is a 3D image which is processed by the backbone with our bottleneck and the input size of each sample is rescaled to $224 \times 224 \times 91 \times 1$. The samples are firstly processed by $7 \times 7 \times 1$ convolution and $3 \times 3 \times 3$ max pooling. Subsequently, the resulting activations pass through consecutive GCP modules containing P3D and GC block used as bottleneck in the backbone. Finally,

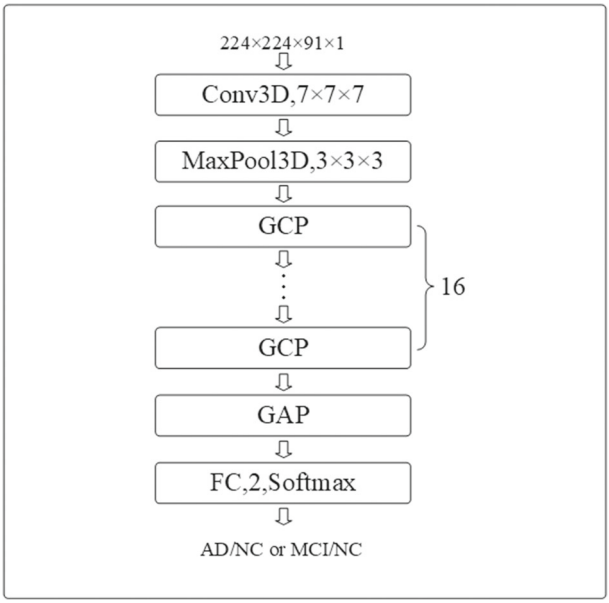

Fig. 2 The overall architecture of our framework. The input of our framework is a 3D image which is processed by the backbone with our bottleneck. The output of our framework is a vector of 2×1 indicating the identified category of the input data. The number of GCP blocks is equal to 16. GAP means global average pooling

fully-connected layer with the activation of softmax is followed after global average pooling. The output of our framework is a vector of 2×1 indicating the category of the input.

3.3 GCP module

In this part, we will introduce the GCP module in details, as shown in Fig. 3. We firstly elaborate why P3D is effective for AD diagnosis followed by applying GC block to 3D MRI after reviewing original 2D GC block [2]. Next, we demonstrate spatial-temporal representation learned by P3D allows effective description of the key features of structural MRI in AD subjects, while we analyze how GC block captures the long-range dependency. In particular, the P3D block decouples 3D convolution to two 2D convolution operations, efficiently decreasing the complexity of our network compared with the baseline. Besides, the global context block efficiently extracts latent representation of the input data and promotes the discriminating capability of our proposed network.

In P3D block [22], three pseudo-3D architectures are proposed in respectively stacked, parallel and mixed fashion. The stacked P3D architecture shown in Fig. 3 consists of 2D filters to extract latent representation in 2D direction and 1D filters to extract latent representation along the depth of input. Besides, it includes two $1 \times 1 \times 1$ filters respectively distributed in the beginning and the end of kernels 2D and 1D filters components. In other terms, the input is convoluted by filters $1 \times 1 \times 1$, $1 \times 3 \times 3$, $3 \times 1 \times 1$ and $1 \times 1 \times 1$ in a cascaded manner, and the convolution result is aggregated to the input. Considering its success in a great deal of deep learning tasks, we apply it to our model for feature representation.

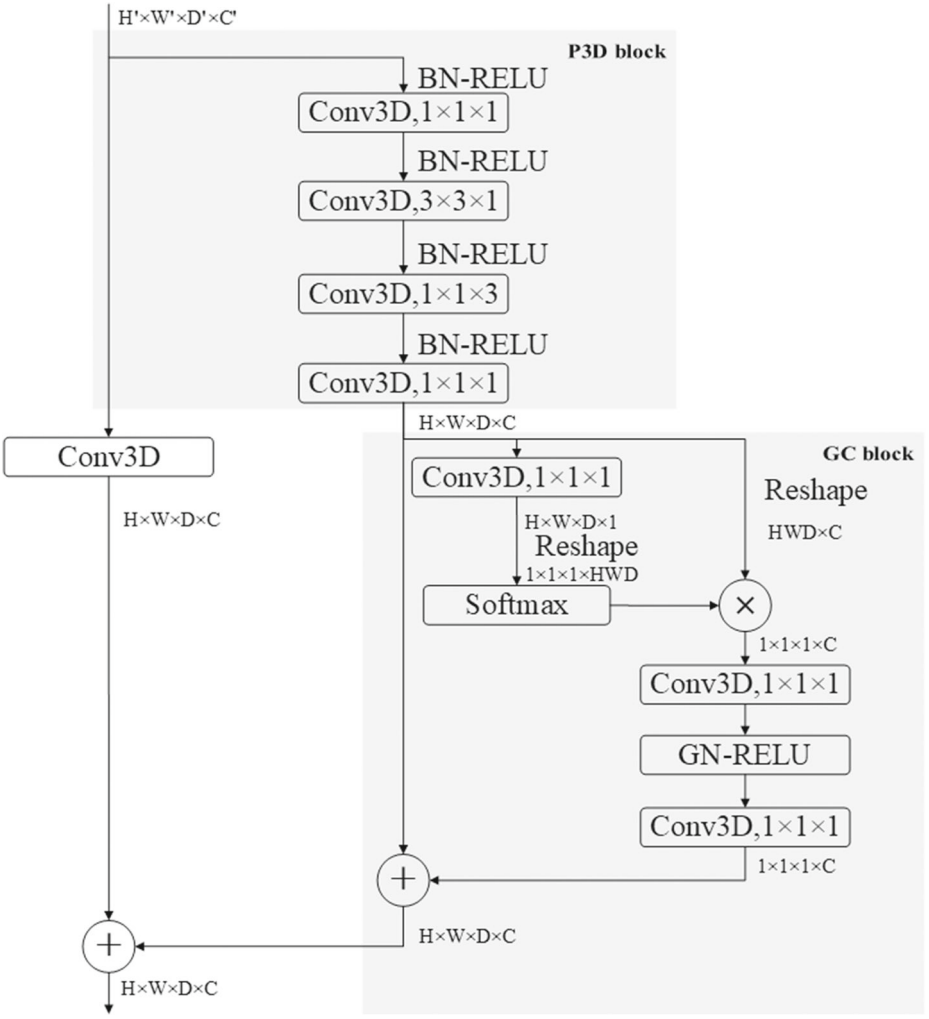

Fig. 3 The architecture of GCP module. It consists of a P3D block and a 3D GC block. Note that the convolution operation in the shortcut flow is discarded when the dimension of input feature is consistent with that of the output in the module

If not specified, P3D in the following text refers specifically to the cascaded architecture of P3D. Please refer to [22] for more details of P3D.

In order to speed up the training process and improve the generalization performance of the network, BatchNormalization(BN) is added before each activation function of ReLU. The processing pipeline of P3D block can be mathematically formulated as

$$D = x + d(x) \tag{2}$$

where x denotes input in this block, $d(\cdot)$ a family of convolutions with BN and ReLU before each of filter, while D the final result of P3D in our method.

P3D enables decomposing 3D input into 2D and 1D input without significant loss of descriptive power. The resulting representation is used to classify and reducing expensive

computational cost. Furthermore, the representation along depth in 3D MRI is relevant, making it easier to learn 3D feature evolution by embedded P3D. As discussed earlier, the deterioration of AD is closely associated with the changes on areas related to AD clearly rejected by 3D feature.

In GC block [2], In term of MRI classification in our method, the key component for performance improvement is GC block which can capture long-range representation based on the centering pixels of a patch. By adding the output features of bottleneck transformation structure to the original feature map one by one, the global context feature map is aggregated to each position of the original feature map, and thus the block can extract the features of relation distributed in multiple disease-related areas more effectively. In [2], the input size of GC block is $H \times W \times C$ (H, W and C indicate the height, width and channel of the input sample respectively). After being processed by context modeling module, the block is resized to $1 \times 1 \times C$, and then processed by transformation module. Finally, the original feature and the transformed result are aggregated to generate the output of GC block.

Due to the inherent properties of MRI data, we have extended the 2D GC block to the 3D model shown in the Fig. 3. In our method, the size of the input data in the GC module is $H \times W \times D \times C$. Firstly, the data is convoluted by a filter of size $1 \times 1 \times 1$, and then resized into $HWD \times C$. After softmax operation, the result is multiplied by the resized feature to obtain the intermediate result of size $1 \times 1 \times 1 \times C$. In the transformation step, we firstly reduce channel C by reduction_ratio times (reduction_ratio is set to be the same as that of [2]), and then perform $1 \times 1 \times 1$ convolution followed by Group Normalization(GN) and activation of ReLU. After another $1 \times 1 \times 1$ convolution operation followed by ReLU, the addition between temporary result and the original input are performed, yielding the output of GC block. Similar to [2], the corresponding processing pipeline can be formulated as

$$G = F(x_G, T(C(x_G))) \tag{3}$$

where x_G represents the input of GC block, $C(\cdot)$ context modeling part, $T(\cdot)$ transform part, $F(\cdot)$ fusion part with addition and G the result of GC block respectively.

In GCP module, convolution layers other than P3D convolution module adopt convolution kernel with size of $1 \times 1 \times 1$. Before the last convolutional layer in the transformation part of GC block, the group normalization is used to normalize the layer features. Context modeling has great advantages in extracting features of all positions, whilst transformation module is also advantageous in capturing channel information features, providing a sound basis for the fusion of the two modules to describe the features on each axis of 3D image. The later experiments also show that GC block improves the detection accuracy, demonstrating its effectiveness in processing 3D MRI images.

Mathematically, the first residual block in our framework used for processing structural MRI and 3D feature learning is constructed as

$$R_{PG} = I_{hwd} + F(d(I_{hwd}), T(C(d(I_{hwd})))) \tag{4}$$

where R_{PG} denotes the feature generated from the bottleneck in the backbone. Meanwhile, R_{PG} is also the intermediate result in our model, since it will be used as the input of the next residual block. In other words, it will replace the position of I_{hwd} when the next residual block works. In this way, it will proceed until the last residual block of our network is reached, leading to the final output R_{last}.

Furthermore, the output layer of our model uses fully-connected (FC) layer to generate an one-hot vector for each input image. Mathematically, it can be described as

$$z = FC(R_{last}) \tag{5}$$

where FC(.) denotes the fully-connected layer of our model. Finally, the softmax function is used to generate the predicted category as follows

$$P_i = \frac{e^{z_i}}{\sum_{k=1}^{N} e^{z_k}} \tag{6}$$

where N is the total number of categories in the AD, e represents the base number of the natural logarithm function while z indicates the output vector with the same dimension as the category number. Besides, i implies the category index, P_i is the probability value of the class i, k lists the indexes of all categories, whilst z_k suggests the k-th value of the output vector.

In the forward propagation of the model, the cross entropy is used as the loss function for model optimiztion. The value of the loss function indicates the training error of the whole training set. The loss function can be expressed as

$$L = -\sum_{i=1}^{N} y_i \log P_i \tag{7}$$

where y_i is the indicator variable. In addition, P_i represents the probability that the predicted samples of the model belong to category i. The back propagation algorithm calculates the gradient of the loss function for all the weights in the network according to the chain derivation rule. For parameters update, this gradient is multiplied by a learning rate α , and then reversed before adding it to the weight.

4 Results

In this section, we first introduce our experimental setup including the trainable parameters and the hyper-parameters involved in our experiment. Then, common metrics are presented to evaluate our method. Finally, comparative studies in which our proposed method is compared with the classic approaches are carried out.

4.1 Experimental setup

We train the proposed method from scratch using ADNI dataset to assist the research over the disease diagnosis by combining genetics, imaging and clinical data. In our scenario, we select 548 examples from ADNI dataset containing 146 AD, 256 NC and 146 MCI. 402 samples are grouped into the training set and the test set to feed into our proposed network in AD/NC or MCI/NC. It is noteworthy that each of the selected examples is preprocessed to remove the areas irrelevant to our study. Particularly, the preprocessing details are shown in Section 3.1. In addition, we train the proposed network in a five-fold cross validation manner such that we can make full use of the limited training data and alleviate the risk of the overfitting to some extent. In particular, we randomly divide our data into five subsets, of which four subsets are used for training while the rest for validation. Thus, final result is obtained by averaging five intermediate results. In order to dynamically balance the efficiency and accuracy, we adaptively adjust the learning rate in the training process. More specifically, set as 1×10^{-3} initially, the learning rate changes to the initial value multiplied by the coefficients of 0.5, 0.1, 0.05 and 0.01 respectively when the epoch grows up to 5, 10, 50 and 80. Therefore, more accurate results can be obtained with learning rate decay, preventing the loss function from oscillating over a period of time. In addition, in order to downplay

the unimportant features in the network, L2 regularization is adopted and the regularization factor is set to 1×10^{-4}.

For performance measure, ROC curve is used as evaluation metric in our experiments. It records the relationship between false positive rate and true positive rate. The area under the curve is termed area under curve (AUC) encoding the prediction accuracy. The larger AUC value indicate the higher prediction accuracy. The optimal AUC value equals 1, suggesting the best prediction accuracy is achieved. In Fig. 4 from the random fold among 5 folds, our method achieves promising performance by reporting high prediction accuracy with the AUC close to 1. This fully suggests the decomposed 3D module followed by a module capturing long-rang latent representation is capable of encoding significant information for AD and NC classification based on 3D structural MRI. In addition, the confidence interval (CI) of ROC curve is given in Fig. 4, which reflects the performance of our method in estimating the real value of totality.

In the training and validation process, we have used cross entropy as loss function which is effective in our task. Figure 5 shows the training and validation curve of one random fold in our five-fold cross validation. It can be observed that the training loss and validation loss decrease with the increasing iterations. When the iterations amount to roughly 100, the convergence is reached. Overall, the training loss is lower than the validation loss, and the difference between the training loss and the validation loss is within a reasonable range since all the training examples are involved in the network training.

Specifically, the accuracy score refers to the ratio of the number of correctly predicted samples to all the samples taken into account. Figure 6 shows the training and validation accuracy curves of one random fold of the five-fold cross validation. Overall, we observe the performance improves as the number of iterations increases. It is easy to find that the curve fluctuates greatly in the early stage of training, and tends to be stable in the late stage of training due to the fast update of model parameters in the early stage of training but the slow update in the late stage of training because of convergence.

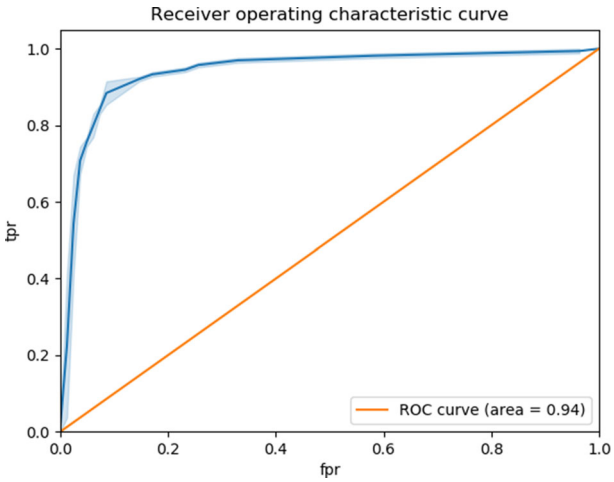

Fig. 4 The ROC curve achieved by our method. tpr indicates true positive rate, fpr false positive rate. ROC curve is mainly used for the prediction accuracy of X to y. Now it is usually used in the field of medicine to judge whether a certain factor plays an extremely important role in diagnosis of a certain disease

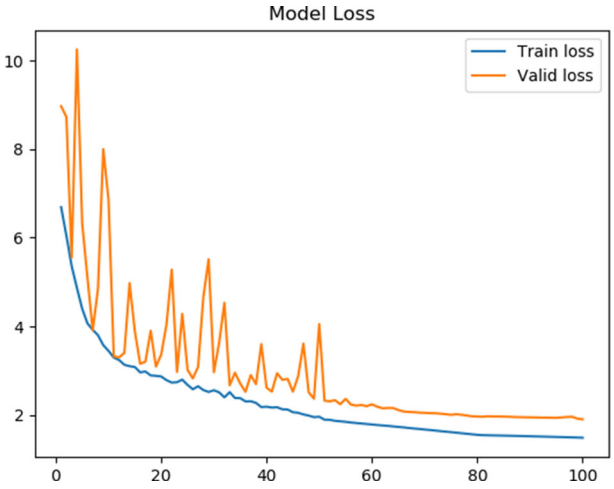

Fig. 5 The Loss curve of our model. The horizontal axis represents the number of iterations and the vertical axis is loss value. There are two loss curves corresponding to training with the navy and validation with the darkorange respectively

4.2 Comparative studies

In order to verify the effectiveness of our method, we have compared our method with the other related methods in our comparative studies. Table 1 shows the results of ours and those achieved by other methods. Note that we directly present the results of these competing methods reported in the corresponding literatures. It can be seen that our method achieves great performance, achieving a very high accuracy, which shows that our method has strong capability to distinguish AD/NC. Although other methods have slightly higher

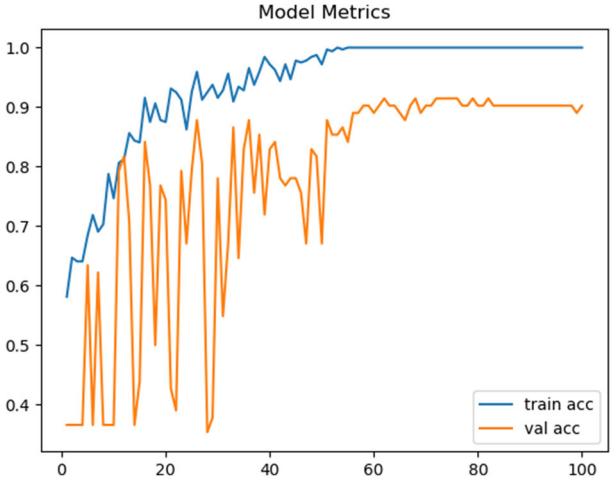

Fig. 6 Accuracy curve achieve by our method. The horizontal axis represents the number of iterations and the vertical axis accuracy value. There are two loss curves corresponding to training with the navy and validation with the darkorange respectively

Table 1 Comparison of our method and different classic approaches

Methods	Sample sizes	Accuracy	AUC	Recall
Cheng et al. [3]	199 AD, 229 NC	0.8715	0.9226	0.8636
Liu et al. [16]	65 AD, 77 NC	0.8776	-	0.8857
Lian et al. [14]	358 AD, 429 NC	0.9000	0.9500	0.8200
Andres et al. [21]	70 AD, 68 NC	0.9000	0.9500	0.8600
Ours	146 AD, 256 NC	0.8929	0.9323	0.9219
Ours(MCI/NC)	146 MCI, 256 NC	0.8757	0.9069	0.9140

accuracy, we achieve higher sensitivity, which indicates that our method can effectively diagnose AD from structural MRI samples. This can be explained by the fact we aggregate the global context feature map to each position of the original feature map. In this way, we effectively aggregate massive region-specific features such that the clinical feature is extracted more effectively from the subject dataset. In addition, in order to verify the generalization of our method and to effectively diagnose the early stage of AD, we evaluate our method on MCI/NC and obtain 87.57% accuracy, as shown in Table 1. On the whole, our method shows strong capability in the discrimination of AD, which benefits from the fact that our proposed method improves the sensitivity to pathological characteristics and the classification accuracy for structural MRI.

5 Discussion

In order to better illustrate the important role each part of our model plays in the performance, ablation studies are conducted in our experiments. Table 2 reveals the performance of our model with one or more additional components embedded. Each result is obtained by training and validation on the same data set. The results also reveal the complementarity among different components of our model. First, the accuracy at 0.8879 is achieved by the backbone network. Despite exhibiting superiority to some earlier architectures, further performance improvement can be expected, since only the backbone network is used for evaluation. Then, we integrate the P3D block into the backbone of the backbone architecture, achieving the accuracy at 0.8330. Replacing the basic block in ResNet with P3D decreases the model performance, suggesting that P3D block decomposing 3D module to 2D and 1D elements mainly contributes to reduce the complexity of the model of AD classification. Next, we explore the complementary effect between P3D and GC block. To be specific, we simultaneously combine P3D and GC on the basis of ResNet. The experimental results reported with 0.8929 demonstrate the prediction accuracy is improved by incorporating the two blocks into ResNet model. As expected, the best experimental results are obtained in this case. This is due to the strong long-range feature fusion capability of GC block. In this way, the model effectively extracts the features closely related to the direct

Table 2 The performance of our method when one or more components are embedded

Backbone	P3D	GC	Accuracy	AUC	Recall
√			0.8879	0.9166	0.9375
√	√		0.8330	0.8874	0.8788
√	√	√	0.8929	0.9323	0.9219

area like hippocampus, while ignoring the features of other irrelevant parts. This also shows the significantly beneficial effect of GC block along with the complementary effect between GC and P3D blocks in effectiveness and complexity respectively.

6 Conclusions

In this paper, we propose a novel method for AD diagnosis. Integrating P3D and GC blocks, our method allows accurate information aggregation. With GCP used as a block in the backbone, we demonstrate the promising performance achieved by our proposed convolutional neural network, and show that GC block and P3D block extract complementary information in the classification of structural MRI. Our method regards the preprocessed MRI as the input of the overall network, and constructs a latent representation by multiple basic blocks to promote high classification accuracy. With the capability of aggregating features and representing global context, our approach is beneficial for the diagnosis of more AD-relevant diseases such as Depression, Hypothyroidism and others in future work. Experiments on the public ADNI dataset show that both of the two modules have significant contributions to the final prediction results.

Acknowledgements This work is supported by the National Natural Science Foundation of China under Grant 61971273, Grant 61877038, Grant 61702251, Grant 61703096, the Key Research and Development Program in Shaanxi Province of China under Grant 2021GY-032 and the Fundamental Research Funds for the Central Universities under Grant GK202003077, Grant GK202105006. The authors would like to gratefully thank NVIDIA Corporationfor for the support of the Titan XP GPU used in our work.

Author Contributions Conceptualization, Z.P.; Methodology, M.M. and M.G.; Software, Z.P. and Y.G.; Supervision, M.M.; Resources, M.G. and C.L.; Data curation, Y.G. and Y.C.; Writing—original draft, Z.P. and Y.G.; Writing—review and editing, M.M., C.L. and J.L. All authors have read and agreed to the published version of the manuscript.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Akkus Z, Galimzianova A, Hoogi A, Rubin D, Erickson B (2017) Deep learning for brain mri segmentation: State of the art and future directions. J Digit Imaging 30(4):449–459, 06
2. Cao Y, Xu J, Lin S, Wei F, Hu H (2019) Gcnet: Non-local networks meet squeeze-excitation networks and beyond. In: 2019 IEEE/CVF International conference on computer vision workshop (ICCVW), pp 1971–1980
3. Cheng D, Liu M, Fu J, Wang Y (2017) Classification of mr brain images by combination of multi-cnns for ad diagnosis. In: Society of photo-optical instrumentation engineers society of photo-optical instrumentation engineers (SPIE) conference series, page 1042042, 07
4. Escudero J, Zajicek JP, Ifeachor E (2011) Machine learning classification of mri features of alzheimer's disease and mild cognitive impairment subjects to reduce the sample size in clinical trials. In: 2011 Annual international conference of the IEEE engineering in medicine and biology society, pp 7957–7960
5. Falahati F, Westman E, Simmons A (2014) Multivariate data analysis and machine learning in alzheimer's disease with a focus on structural magnetic resonance imaging. J Alzheimer's Disease 41(3):685–708,04
6. Gao XW, Hui R, Tian Z (2017) Classification of ct brain images based on deep learning networks. Comput Methods Prog Biomed 138:49–56

7. He K, Zhang X, Ren S, Sun J (2016) Deep residual learning for image recognition. In: 2016 IEEE Conference on computer vision and pattern recognition (CVPR), pp 770–778
8. Hosseini-Asl E, Keynton R, El-Baz A (2016) Alzheimer's disease diagnostics by adaptation of 3d convolutional network. In: 2016 IEEE International conference on image processing (ICIP), pp 126–130
9. Hu J, Shen L, Sun G (2018) Squeeze-and-excitation networks. In: 2018 IEEE/CVF Conference on computer vision and pattern recognition, pp 7132–7141
10. Jack C, Bernstein M, Fox N, Thompson P, Alexander G, Harvey D, Borowski B, Britson P, Whitwell J, Ward C, Dale A, Felmlee J, Gunter J, Hill D, Killiany R, Schuff N, Fox-Bosetti S, Lin C, Studholme C, Weiner M (2008) The alzheimer's disease neuroimaging initiative (adni): Mri methods. J Magnetic Resonance Imaging 27:685–691, 05
11. Jin D, Xu J, Zhao K, Hu F, Yang Z, Liu B, Jiang T, Liu Y (2019) Attention-based 3d convolutional network for alzheimer's disease diagnosis and biomarkers exploration. In: 2019 IEEE 16Th international symposium on biomedical imaging (ISBI 2019), pp 1047–1051
12. Khvostikov A, Aderghal K, Benois-Pineau J, Krylov A, Catheline G (2018) 3d cnn-based classification using smri and md-dti images for alzheimer disease studies. arXiv: Computer Vision and Pattern Recognition 01
13. Klöppel S, Stonnington C, Chu C, Draganski B, Scahill R, Rohrer J, Fox N, Jack C, Ashburner J, Frackowiak R (2008) Automatic classification of mr scans in alzheimer's disease. Brain 131(3):681–689, 04
14. Lian C, Liu M, Zhang J, Shen D (2020) Hierarchical fully convolutional network for joint atrophy localization and alzheimer's disease diagnosis using structural mri. IEEE Trans Pattern Anal Mach Intell 42(4):880–893
15. Liu M, Cheng D, Yan W (2018) Alzheimers Disease Neuroimaging Initiative: Classification of alzheimer's disease by combination of convolutional and recurrent neural networks using fdg-pet images. Frontiers in Neuroinformatics 12
16. Liu S, Liu S, Cai W, Pujol S, Kikinis R, Dagan D, Feng F (2014) Early diagnosis of alzheimer's disease with deep learning. 1015–1018, 04
17. Liu S, Song Y, Cai W, Pujol S, Kikinis R, Wang X, Feng D (2013) Multifold bayesian kernelization in alzheimer's diagnosis 16,303–310, 09
18. Liu M, Zhang J, Adeli E, Shen D (2019) Joint classification and regression via deep multi-task multi-channel learning for alzheimer's disease diagnosis. IEEE Trans Biomed Eng 66(5):1195–1206
19. Liu M, Zhang D, Shen D (2014) Hierarchical fusion of features and classifier decisions for alzheimer's disease diagnosis. Hum Brain Mapp 35:1305–1319, 04
20. Moradi E, Pepe A, Gaser C, Huttunen H, Tohka J (2015) Machine learning framework for early mri-based alzheimer's conversion prediction in mci subjects. NeuroImage 104:398–412
21. Ortiz A, Munilla J, Gorriz J, Ramírez J (2016) Ensembles of deep learning architectures for the early diagnosis of the alzheimer's disease. Int J Neural Syst 26:03
22. Qiu Z, Yao T, Mei T (2017) Learning spatio-temporal representation with pseudo-3d residual networks. In: 2017 IEEE International conference on computer vision (ICCV), pp 5534–5542
23. Risacher S, Saykin A, West J, Shen L, Firpi H, Mcdonald B (2009) Baseline mri predictors of conversion from mci to probable ad in the adni cohort. Current Alzheimer Res 6:347–361, 08
24. Sarraf S, Tofighi G (2016) Classification of alzheimer's disease structural MRI data by deep learning convolutional neural networks. arXiv:1607.06583
25. Suk H-I, Lee S-W, Shen D (2017) Deep ensemble learning of sparse regression models for brain disease diagnosis. Med Image Anal 37:101–113
26. Varol E, Gaonkar B, Erus G, Schultz R, Davatzikos C (2012) Feature ranking based nested support vector machine ensemble for medical image classification. In: 2012 9Th IEEE international symposium on biomedical imaging (ISBI), pp 146–149
27. Wang X, Girshick R, Gupta A, He K (2018) Non-local neural networks. In: 2018 IEEE/CVF Conference on computer vision and pattern recognition, pp 7794–7803
28. Wang S, Wang H, Shen Y, Wang X (2018) Automatic recognition of mild cognitive impairment and alzheimers disease using ensemble based 3d densely connected convolutional networks. In: 2018 17Th IEEE international conference on machine learning and applications (ICMLA), pp 517–523
29. Ye DH, Pohl KM, Davatzikos C (2011) Semi-supervised pattern classification: Application to structural mri of alzheimer's disease. In: 2011 International workshop on pattern recognition in neuroimaging, pp 1–4
30. Zhang C, Adeli E, Zhou T, Chen X, Shen D (2018) Multi-layer multi-view classification for alzheimer's disease diagnosis. AAAI Conf Artif Intell 2018:4406–4413,02
31. Zhang D, Shen D (2012) Alzheimer's Disease Neuroimaging Initiative. Predicting future clinical changes of mci patients using longitudinal and multimodal biomarkers. PLOS ONE 7(3):e33182,03

32. Zhang D, Wang Y, Zhou L, Yuan H, Shen D (2011) Multimodal classification of alzheimer's disease and mild cognitive impairment. NeuroImage 55(3):856–867
33. Zhang J, Yu J, Tao D (2018) Local deep-feature alignment for unsupervised dimension reduction. IEEE Trans Image Process 27(5):2420–2432
34. Zhou T, Liu M, Thung K, Shen D (2019) Latent representation learning for alzheimer's disease diagnosis with incomplete multi-modality neuroimaging and genetic data. IEEE Trans Med Imaging 38(10):2411–2422
35. Zhu X, Suk H-I, Wang L, Lee S-W, Shen D (2017) A novel relational regularization feature selection method for joint regression and classification in ad diagnosis. Med Image Anal 38:205–214

Affiliations

Zhao Pei[1,2] · Yuanshuai Gou[2] · Miao Ma[2] · Min Guo[2] · Chengcai Leng[3] · Yuli Chen[2] · Jun Li[4]

Yuanshuai Gou
gys@snnu.edu.cn

Miao Ma
mmthp@snnu.edu.cn

Min Guo
guomin@snnu.edu.cn

Chengcai Leng
ccleng@nwu.edu.cn

Yuli Chen
chenyuli@snnu.edu.cn

Jun Li
lijuncst@njnu.edu.cn

1 Key Laboratory of Modern Teaching Technology, Ministry of Education, Xi'an 710062, China

2 School of Computer Science, Shaanxi Normal University, Xi'an 710119, China

3 School of Mathematics, Northwest University, Xi'an 710127, China

4 School of Computer Science, Nanjing Normal University, Nanjing 210046, China