JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 37, 347-363 (2021)
DOI: 10.6688/JISE.202103_37(2).0005

An Explainable Diagnostic Method for Autism Spectrum Disorder Using Neural Network

MINGKANG ZHANG, YANBIAO MA, LINAN ZHENG, YUANYUAN WANG, ZHIHONG LIU, JIANFENG MA, QIAN XIANG, KEXIN ZHANG AND LICHENG JIAO
Department of Computer Science and Technology
Xidian University
Xi'an, 710071 P.R. China
E-mail: mk_zhang@stu.xidian.edu.cn

Autism spectrum disorder (ASD), also known as autism, is a mental illness caused by disorders of the nervous system. Autism is mainly characterized by developmental disorders, accompanied by abnormalities in social skills, communication skills, interests, and behavioral patterns. Autism cannot be completely cured by existing medical means, and its symptoms can only be relieved through acquired intervention. The best intervention period for autistic patients is before the age of six. But relying on existing methods, most patients with autism have missed the best intervention period when they are diagnosed. In order to allow the subject to be diagnosed with autism in a timely manner, we proposed a method that uses a deep neural network to analyze the subject's magnetic resonance imaging (MRI) and evaluate the performance for early screening of ASD. Our primary analysis of patients with functional magnetic resonance imaging (fMRI) also compared with structural magnetic resonance imaging (sMRI). Experiments have shown that fMRI is more sensitive to autism than sMRI. In addition, we explain the classification results of fMRI.

Keywords: autism spectrum disorder (ASD), 3D convolutional neural network (3D CNN), magnetic resonance imaging (MRI), network visualization, network interpretation

1. INTRODUCTION

Autism spectrum disorder (ASD) is a congenital neurodevelopmental disorder that cannot be effectively detected by traditional medical methods. However, with the rise of the *AI+ Medical* model, deep neural networks (DNN) have increasingly been used in many aspects of the medical field, which brings a new dawn to the detection of ASD. So far, researchers have used deep convolutional neural networks to identify Computed Tomography (CT) images, such as pneumonia by CT images of the lungs [1], classifying tuberculosis patients based on CT images [2], and a large number of researchers have achieved good results in the classification of brain images, especially brain tumor images [3, 4]. In particular, for brain image recognition, the use of deep convolutional neural networks to analyze the sMRI of the patient's brain [5, 6] is also a common approach. However, in current clinical applications, the diagnostic criteria for autism in various countries

Received November 9, 2019; revised February 24, 2020; accepted April 1, 2020.
Communicated by Jimson Mathew.

are still the scale method proposed in *Diagnostic and Statistical Manual of Mental Disorders (Fifth Edition)* (DSM-V), which is provided by American Psychiatric Association. The table method requires doctors to observe and evaluate the social behavior, language, movements and repetitive behaviors of patients, subjective, high misdiagnosis rate and long diagnosis period.

Though many scholars have done researchs on ASD, most of them are for symptom analysis of patients who have already identified autism [7, 8]. In addition, some scholars have developed a classifier based on nerve images to identify functional connectivity anomalies [9]. However, these are all researches for adult autism which is difficult to prognose. We hope to determine whether the subject suffers from autism in childhood, especially around one year old, because the earlier the age of detection and treatments, the more obvious the prognosis improvement of autism.

Our designs Facing the low accuracy and the long diagnosis period of traditional medical methods of ASD diagnosis, we explored the possibility of applying DNN to early screening for ASD. After multiple comparisons, we collaborated with two local hospitals to collect sMRI data from 1,102 subjects and fMRI data from 2,352 subjects. Based on these, we use 3D CNN [10] to replace traditional 2D CNN to extract features from sMRI. At the same time, we combine the 3D CNN with ConvLSTM [11] to extract features from fMRI.

Our design is based on two key insights. First, the convolutional neural network is very effective at extracting feature information from images. If the acquisition of ASD is related to different features of one or more brain regions, it will also be extracted by our network. Second, fMRI has a rich set of timing characteristics. To take advantage of these timing features, we first apply the 3D convolution to a single time point, and the convolved results will be sent to ConvLSTM. Since the last time point contains the most abundant information, we take the output of the last time node of ConvLSTM as a result.

Evaluations and experiment We focus on evaluating the performance of neural networks under fMRI, and find it has an excellent performance in early autism classification. Meanwhile, we introduce a new means - deconvolution, and try to use this method to mark the classification of fMRI to explain why it has such excellent performance. Last, we hope we can give back the results of our explanation to doctors as a medical diagnosis of autism.

Contributions The main contribution of this paper can be summarized as follows:

- We establish a database of sMRI and fMRI for patients with ASD in early stage.
- We have found a correlation between ASD and MRI in the brain, especially with fMRI. Moreover, we try to analyze the fMRI of the brain with a deep neural network and obtained excellent classification results.
- We explain the classification results of fMRI.

The use of neural networks to analyze magnetic resonance imaging (MRI) of subjects suspected of having ASD is a new and potentially promising screening tool. Compared to the existing screening method, especially in using neural networks for early screening of fMRI, we have achieved even better results.

2. DATA SET

Our team collaborated with two local Grade-A Tertiary Hospital to collect 1,102 sets of sMRI data and 2,352 sets of fMRI data. sMRI data includes 530 children with autism and 572 normal children. fMRI data includes 1,123 children with autism and 1,229 normal children. Since both sMRI and fMRI have a small amount of data, we have not separately divided the dev set. We randomly selected 70% of fMRI data as the training set and the remaining 30% as the test set. The same is true for sMRI.

2.1 Collecting Testers

Normal children are recruited from local kindergartens. The criteria for entry are as follows (refer to *Autism Diagnostic Interview-Revised (ADI-R)* assessment results that do not meet the children's autism score criteria):

- According to *Wechsler Intelligence Scale for Children (WISC)* test, the total IQ is greater than or equal to 70 points;
- Habitual use of the right hand.
- do not take any psychotropic drugs before collection.
- There is no history of mental illness in themselves and their family.

We guarantee that children and parents volunteer to participate in the study and can cooperate with MRI. The normal children who participated in the study were 4 to 10 (7 ± 3) years old; the total IQ score was (91.06 ± 12.40).

Children in the autism group are diagnosed by the local top three hospitals and meet the diagnosis of *American Diagnostic and Statistical Manual of Mental Disorders (Fourth Edition) (DSM-IV)* The most basic requirements for recruiting autistic children as follows:

- Comply with the criteria for determining autism.
- Habitual use of the right hand.
- There is no history of mental illness in themselves and their family.
- No other psychotropic drugs are used before collection.

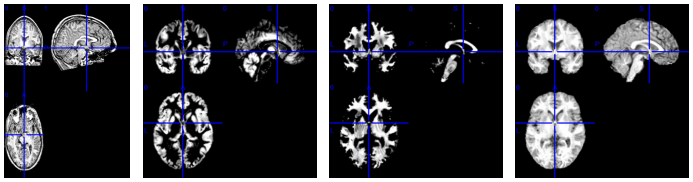
Recruited children need to exceed the autism disorder diagnosis threshold of assessment results in ADI-R; According to WISC test, the total IQ is greater than or equal to 70 points; Testers and their families are willing to participate in the research; The children recruited in the study are 4 to 10 (7 ± 3) years old; The total IQ is (83.87 ± 15.51) points.

There is no significant difference in age ($P = 0.321$) and total IQ ($P = 0.154$) between the two groups. All tested children and their parents agree to participate in the study, and the parents sign informed consent.

2.2 Pre-processing of sMRI Data

T1 structural images are analyzed with Matlab 2012a using Statistical Parametric Mapping[1] and VBM toolbox[2]. Imaging data are pre-processed by realignment, bias-correction, tissue classification and spatial normalization. Tissue classification is used

[1] SPM8, Welcome Department of Cognitive Neurology, London, UK; https://www.fil.ion.ucl.ac.uk/spm/
[2] http://dbm.neuro.uni-jena.de/vbm/download/



(a) (b) (c) (d)

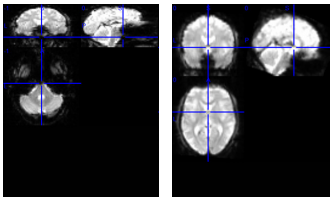
Fig. 1. The sMRI data; (a) the original sMRI data before pre-processing; (b) the gray matter data after pre-processing; (c) the white matter data after pre-processing; (d) the Cerebrospinal Fluid (CSF) data after pre-processing. By preprocessing the raw image data, we divided the structural magnetic resonance imaging of the brain into gray, white matter and CSF. The gray matter data is widely used in image analysis, which will be used in the next experiment.

to segment the brain into GM, WM, and cerebrospinal fluid images based on an adaptive Maximum a Posterior (MAP) technique [12]. To this end, images are normalized to the Montreal Neurological Institute (MNI) template using the Diffeomorphic Anatomical RegisTration using Exponentiated Lie algebra (DARTEL) technique [13] with a predefined tissue probability map. In addition, non-linear normalization is only incorporated to take into account volume changes caused by spatial normalization that could cause certain brain regions to shrink or expand. This is done by multiplying the voxel values by the non-linear components derived from the spatial normalization step. The voxel size is $1.5\times1.5\times1.5$ mm^3. Finally, all normalized, segmented and modulated images are smoothed with an 8-mm full-width at half-maximum (FWHM) isotropic Gaussian kernel. The sMRI images before and after processing are shown in Fig. 1.

2.3 Pre-processing of fMRI Data

Slow fluctuations of brain activity are fundamental features of the resting brain, and their presence is vital in determining correlated activity between brain regions and defining resting state networks. The relative magnitude of these fluctuations can differ between brain regions and between subjects, and thus may act as an index of individual difference or dysfunction. Resting-state images are preprocessed using SPM8 and Data Processing Assistant for Resting-State fMRI[3]. Specifically, the first five time points are removed to minimize non-equilibrium effects in the fMRI signal, and then slice-timing, realignment-based head movement correction, and spatial normalization (voxel size of $3\times3\times3$ mm^3) are performed. Demeaning or detrending is performed and head-motion parameters, white-matter signals, cerebrospinal-fluid signals and global signals are regressed out as nuisance covariates [14]. fMRI time-points that are severely affected by motion are removed using a "scrubbing method" (FD value $>$ 0.5mm, and ΔBOLD of DVARS $>$ 0.5%), and less than 5% time points were removed per subject. Finally, Band-pass temporal filtering (0.01-0.08Hz) is used to remove the effects of very low-frequency drift and high-frequency noise, and all images are smoothed with a 6-mm FWHM Gaussian kernel. The fMRI images before and after processing are shown in Fig. 2.

[3]DPARSF, https://sourceforge.net/projects/restingfmri/



(a) (b)

Fig. 2. The fMRI data; (a) the original fMRI data before pre-processing; (b) the fMRI data after pre-processing. The pre-processed image is more regular than the image before pre-processing, making it easier for the network to analyze.

3. NETWORK STRUCTURE

3.1 Network Structure under the Principle of sMRI

Here, We have improved the traditional 3D convolution process. The concept of 3D convolutional networks is first proposed in a paper on human behavior recognition in surveillance video [10]. Researchers in this work stack multiple two-dimensional video frames to form a three-dimensional cube, and then apply a three-dimensional convolution kernel to feature extraction. However, for sMRI data, it is already a three-dimensional structure. We directly apply a three-dimensional convolution kernel directly on it. For convenience, we call our improved 3D convolution process "3D CNN which is applied to spatial dimensions". The single-step three-dimensional convolution process is shown in Fig. 3.

sMRI data is a three-dimensional structure, and we use 3D CNN which is applied to spatial dimensions to extract features. When the data dimension drops to a low level, we roll it out and send it to the fully connected layer for disease risk prediction. The network structure is shown in Table 1.

Table 1. Neural network under the principle of sMRI.

Layer	Layer Type	Output Shape	#Param
1	Input Layer	121×145×121×1	0
2	Conv3D+Relu	121×145×121×4	6
3	Conv3D+BN+Relu	60×72×60×3	246+12
4	Conv3D+Relu	60×72×60×8	32
5	Conv3D+BN+Relu	29×35×29×8	1736+32
6	Conv3D+Relu	29×35×29×10	90
7	Conv3D+BN+Relu	14×17×14×10	2710+40
8	Fatten	33320	0
9	FC+Relu	256	8530176
10	FC+Relu	128	32896
11	FC+Softmax	2	258

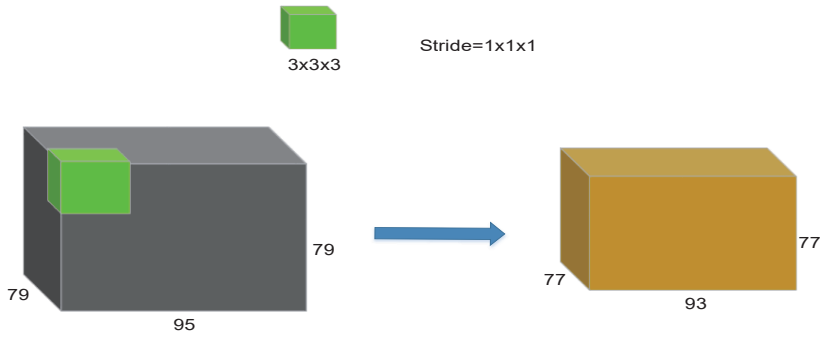


Fig. 3. Single 3D convolution kernel convolution calculation process. For example, a convolution operation of a size of 3×3×3 is used to convolve a 79×95×79 cube. The convolution step is 1 in all directions, so a convolution is obtained with a cube of size 77×93×77.

3.2 Network Structure under the Principle of fMRI

Here, we use 3D CNN which is applied to spatial dimensions and ConvLSTM. ConvLSTM is first proposed in precipitation nowcasting [11]. Researchers in this work build an end-to-end trainable model for precipitation nowcasting by stacking multiple ConvLSTM layers and forming an encoding-forecasting structure. In ConvLSTM, the three gates are updated with gate Γ_u, the forgotten gate Γ_f, and the output gate Γ_o are given by Eqs. (1), (2) and (3).

Unlike ordinary LSTM, in ConvLSTM, $x^{<t>}$ is a two-dimensional input. $a^{<t-1>}$ is the activation value of the previous layer, so the dimension is the same as $x^{<t>}$. W is the corresponding weight parameter, and its subscript can well reflect which gate value is associated with it. It can be seen that, in addition to the input of the layer $x^{<t>}$ and the upper output $a^{<t-1>}$, the gate value also "peep" the memory cell value of the upper layer $c^{<t-1>}$, this is due to the addition of a peephole. Finally, the memory cell $c^{<t>}$ and activation $a^{<t>}$ of the t-th layer are given by Eqs. (4) and (5).

$$\Gamma_u = \sigma(W_{xu} \cdot x^{<t>} + W_{au} \cdot a^{<t-1>} + W_{cu} \cdot c^{<t-1>} + b_u) \tag{1}$$

$$\Gamma_f = \sigma(W_{fu} \cdot x^{<t>} + W_{af} \cdot a^{<t-1>} + W_{cf} \cdot c^{<t-1>} + b_f) \tag{2}$$

$$\Gamma_o = \sigma(W_{ou} \cdot x^{<t>} + W_{ao} \cdot a^{<t-1>} + W_{co} \cdot c^{<t-1>} + b_o) \tag{3}$$

$$\mathbf{c}^{<t>} = (\Gamma_u \cdot tanh(W_{xc} \cdot x^{<t>} + W_{oc} \cdot a^{<t-1>}) + \Gamma_f \cdot c^{<t-1>} \tag{4}$$

$$\mathbf{a}^{<t>} = \Gamma_o \cdot tanh(c^{<t>}) \tag{5}$$

Since the data at a single time point of fMRI is also a three-dimensional structure, we first use 3D convolution all applied to spatial dimensions to the data at each time point. We select images of 20 consecutive time points in the middle as the input to the convolution layer (Note that the image at each time is also a three-dimensional structure).

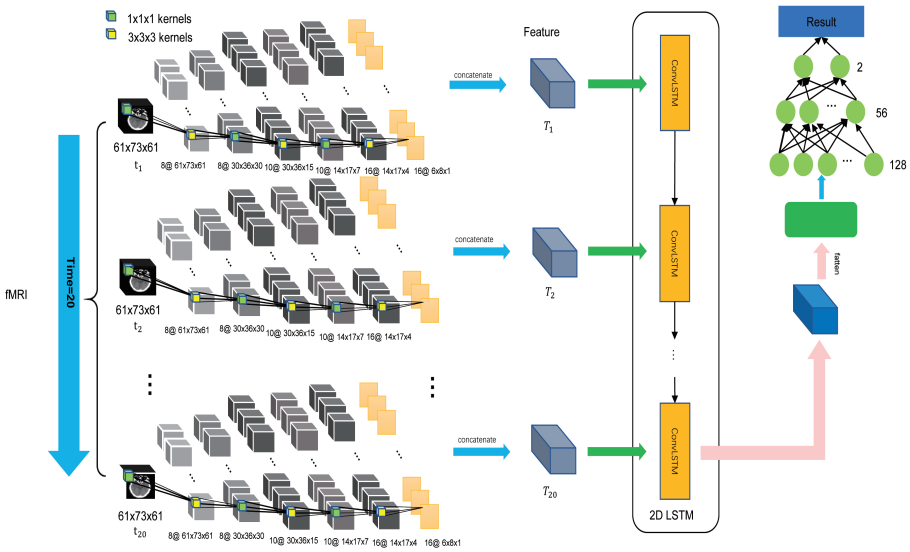


Fig. 4. Neural network structure under fMRI.

There are three reasons why we do this: First, the middle data is stable enough. Second, when the number of time point increases to a certain scale, existing computing resource is not capable to accommodate all the data, thus resulting in memory overflow. Last, the accuracy rate is not significantly improved even if the data of all time points is input to the network. Conversely, less time points can save a lot of computing resources. The convolution process for the convolution portion of our network is shown in Table 2.

Table 2. Convolution process of neural networks at a single time point of fMRI.

Layer	Layer Type	Output Shape
1	Input Layer	61×73×61×1
2	Conv3D+BN+Relu	61×73×61×8
3	Conv3D+BN+Relu	30×36×30×8
4	Conv3D+BN+Relu	30×36×15×10
5	Conv3D+BN+Relu	14×17×7×10
6	Conv3D+BN+Relu	14×17×4×16
7	Conv3D+BN+Relu	6×8×1×16

Table 2 shows a description of the convolution process at a single point in time for fMRI. We selected data for a total of 20 time points. Table 2 lists only the convolution process at a single point in time. The convolution process at the other 19 points is exactly the same so that we do not list them. After convolutional process, we add the convolution results of the 20 time points and send them to ConvLSTM. Since the last time point of ConvLSTM contains the richest information, we select the output of the last time point and send it to the fully connected layer to predict the risk of acquiring ASD. The network structure is shown in Fig. 4.

4. EXPERIMENTS

In this section, we evaluate the performance of our network on fMRI data, and compare it to the performance of the network on sMRI data. The results show that our network is very sensitive to fMRI data classification and bad on sMRI data.

4.1 Experimental Setup

We divide our data set as described in Chapter 2. In the training process for sMRI, we set the learning rate to 0.00008, the batch size to 16, and use the Stochastic Gradient Descent method (SGD) for training with a learning rate attenuation of 0.5e-6. The number of epochs is 150. In the training process for fMRI, we set the learning rate to 0.001, the batch size to 16, and the Adam optimization algorithm for training with a learning rate attenuation of 0.5e-6. The number of epochs is 150.

For each training epoch, we calculated the cross entropy loss and accuracy on the training set and test set and output. Currently, because there are not public MRI data sets for ASD, we are unable to list the results of other researchers to compare with us. In fact, we are prepared to expose the data we have collected so that other people can conduct relevant research.

4.2 The Performance of Our Network under fMRI and sMRI

In the process of training fMRI data, we recorded the loss value and accuracy value of each generation of training. After the end of the training, we draw the loss and accuracy of the image according to the loss value and accuracy value in the training process, as shown in Fig. 5 (a). We did the same on sMRI data, but since the network was not well trained on sMRI data, we applied the early stop operation when we trained to 100 generations, which is shown in Fig. 5 (b). Finally, we list the accuracy of the network with different hyperparameters on fMRI and sMRI, as shown in Table 3.

5. INTERPRETATION METHOD

We have got excellent classification results on fMRI data. We hope to interpreting the network for fMRI by deconvolution to mark the voxels that are classified. These labeled voxels reflect why a fMRI is so classified, and can be returned to doctors as a basis for diagnosis as a potential biological cause.

5.1 Some Visualization Methods

The interpretability problem of the CNN model, also known as the visualization of CNN, was discussed after the birth of the network itself. In fact, there is not a very good way to explain CNN so far. The currently available visualization methods are deconvolution and guided-backpropagation, through which we can see the features learned by the deeper convolutional layers of the CNN model to some extent.

Both deconvolution and guided-backpropagation are based on backpropagation. When the three methods interpret the same two-dimensional picture, the interpretation result obtained by backpropagation is usually so noisy that almost impossible to explain.

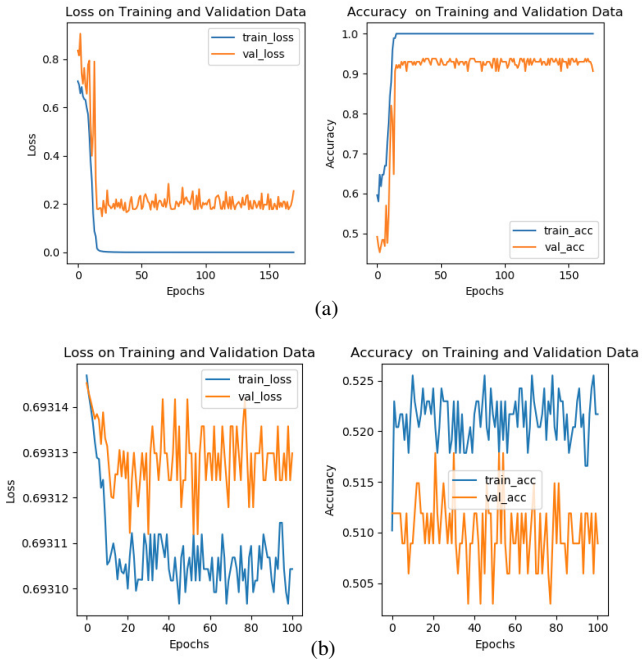


Fig. 5. Comparison of network loss and accuracy; (a) is the performance of our designed network under the principle of fMRI; and (b) is the performance of our designed network under the principle of sMRI.

When using deconvolution to interpret, it is possible to approximate the pixel profile that affects the classification result, but there is a large amount of noise outside the contour. Finally, when using guided-backpropagation to interpret, there is essentially no noise, and features can be clearly concentrated on the pixels that affect the classification results.

5.2 Selection of Visualization Methods

In essence, both deconvolution and guided-backpropagation are derived from back-propagation, which is the derivation of the input. The first appearance of the concept "deconvolution" was proposed by Zeiler in his paper [15], but the name "deconvolution" was not specified at that time. The formal use of the term deconvolution is in its subsequent work [16].

Although we can see the internals of the CNN model by means of backpropagation, deconvolution and guided-backpropagation, none of them are suitable as a way to explain our classification results. The three methods are not sensitive enough to the categories, but directly show all the features that can be extracted, even guided-backpropagation. When guided-backpropagation is applied in high-dimensional scenes (such as interpreting our 3D CNN results), the interpretation effect will be significantly worse. In other words, we cannot explain our fMRI classification results by these three methods. To solve this problem, we must consider other methods.

5.3 A New Way: the CAM Algorithm

Just as a thermogram can tell us what hot objects are in the dark, we hope to find an algorithm that can locate the neural network interpretation principle. For a deep convolutional network, after multiple convolutions and pooling, its last layer contains the most abundant spatial and semantic information. The information contained in the fully-connected layer is difficult for humans to understand, and is difficult to visualize. Therefore, in order to better explain the classification results of the network, the last convolutional layer must be utilized well.

The Class Activation Mapping (CAM) algorithm does a good job of this. This algorithm draws on a famous paper published by Lin Min et al. [17], in which the fully-connected layer is replaced with Global Average Pooling (GAP). GAP has very significant advantages. First, the input can be of any size without fully-connected layer. Secondly, GAP can make full use of spatial information, and it is more robust and not easy to produce over-fitting. The most important point is that in the mlpconv layer (the last convolutional layer), the feature map with the same number of target categories is forced to be generated, and finally sent to the sigmoid layer through the GAP layer to obtain the result, so that each feature map is given a very clear meaning, which is also called category confidence maps.

5.4 Using Improved CAM Algorithm to Interpret fMRI

Although the CAM algorithm already has a good interpretation effect, there is a very obvious disadvantage that it requires modification of the structure of the original model, which needs the retraining of the model so that the training cost is greatly increased. In particular, if the model is already up and running, it is almost impossible for us to retrain it. Therefore, we improved the CAM algorithm on the basis of the original, as shown in Fig. 6, in order to apply to the actual situation. The improved algorithm process will be described in detail below.

First, we select a patient's fMRI image, and input it into the trained network model for calculation. We extract 16 three-dimensional feature maps of the penultimate convolution layer of the first sample point in the fMRI sequence, and then calculate the neuron importance weight of each feature map to the classification task α_m^c, which is shown as Eq. (6), where "Z" is calculated by Eq. (7). In Eq. (6), $\alpha_m^c(t)$ represents the contribution of the m_{th} feature map of the t_{th} sample point of fMRI to the classification results of c-class; w, l and h represent the width, length and height of the feature map; y^c represents the score of the c-class before the softmax; $A_{ijk}^m(t)$ represents the activation value of m_{th} feature map of the t_{th} sample point of fMRI. From this we have obtained the m_{th} feature map for the c-class, which is shown as Eq. (8).

In our project, we set $T = 20$ (The sampling point of a single fMRI is much larger than 20, and we have selected the samples with the most stable 20 time points in the middle part), then derive the activation map I of the fMRI of a single autism spectrum disorder patient, which is shown as Eq. (9).

$$\alpha_m^c(t) = \frac{1}{Z} \sum_i \sum_j \sum_k \frac{\partial y^c}{\partial A_{ijk}^m(t)} \tag{6}$$

$$Z = w \cdot l \cdot h \tag{7}$$

$$L^c_{Grad-CAM}(t) = ReLU(\sum_m \alpha^c_m(t) \cdot A^m(t)) \tag{8}$$

$$I = \sum_{t=1}^{T} L^c_{Grad-CAM}(t) \tag{9}$$

5.5 Abnormal Brain Area Extraction

In order to accurately locate the brain region, it is necessary to suppress the region with a lower activation level to highlight the region with a higher activation level. The suppression method we use as Eq. (10). Then, we normalize the $I(i,j,k)$ to make the voxel value of the brain thermogram in the 0-1 range, which is shown as Eq. (11). Fig. 7 shows the effect before and after suppression.

To verify the effect of suppression, we randomly select 116 samples including 54 patients with ASD and 62 normal subjects. All the tested people are distributed between 9 and 13 years old, with an average age of 11 years. Then, we pick up 54 fMRI thermograms of the ASD group and weight them together. After comparing the different effects of suppressing 10 rounds, 20 rounds and 30 rounds, we chose to suppress the activation map after 30 rounds, as shown in Fig. 8. Meanwhile, we performed the same processing on the images of 62 normal subjects, and pick up 54 of them as brain activation maps representing normal persons. Finally, by comparing the activation maps between normal subjects and ASD group, we obtain 15 brain activation regions shared by the ASD patient group and the normal group, as shown in Fig. 9, which can be used as markers to distinguish patients with ASD from normal people.

Finally, we list the top 10 key brain regions based on the value of $I(i,j,k)$, as shown in Table 4. Among them, the left frontoparietal network (LFPN) is related to the perception of speech and semantics, mainly distributed in the Broca area, the Wernick area, the medial frontal lobes and the caudate nucleus. The Broca area is primarily responsible for language production, and may be related to grammar organization. The Wernick area is responsible for the acceptance and understanding of the language. We give the difference between the normal brain area and the abnormal brain area as shown in Fig. 10. Patients with ASD are generally unable to respond to other people's words and remain silent. Therefore, it may be that ASD patients have greater problems in language acceptance. And secondly, language understanding is problematic.

$$I(i,j,k) = \begin{cases} I(i,j,k) & I(i,j,k) \geqslant \frac{\sum_{i,j,k} I(i,j,k)}{w_I l_I h_I} \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

$$I(i,j,k) = \frac{I(i,j,k)}{I_{max}(i,j,k)} \tag{11}$$

6. DISCUSSIONS

We have now enabled visualization of deep networks, targeting areas of deep learning models that are of interest to fMRI in ASD patients. However, in the visualization, the

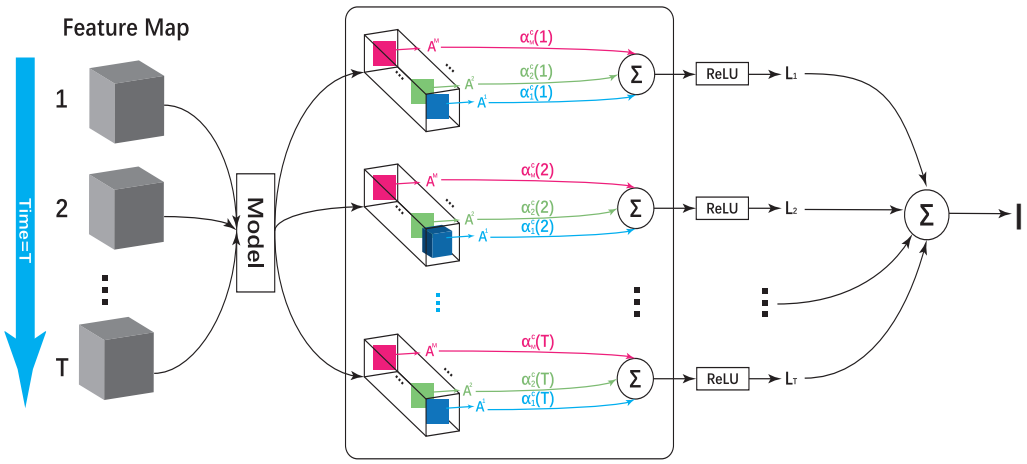


Fig. 6. Improved CAM algorithm.

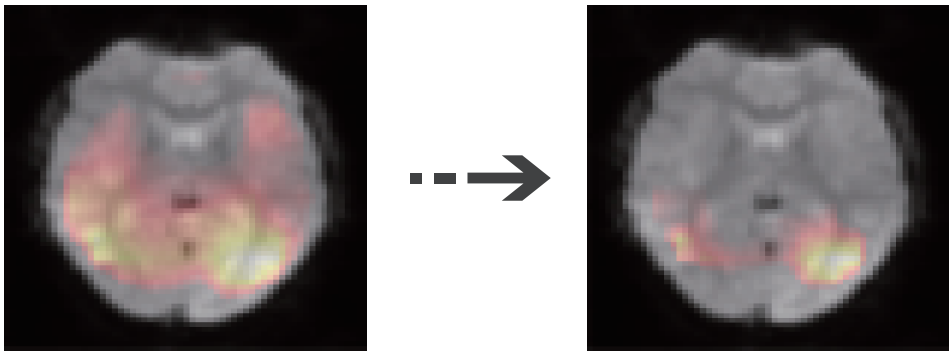


Fig. 7. Schematic diagram of suppression effect.

Table 3. The top 10 brain regions with the greatest difference.

Encephalic region	Activation
Angular_L	0.90648
Parietal_Inf_L	0.84867
Occipital_Sup_L	0.845
Parietal_Sup	0.84263
Precuneus_R	0.83135
SupraMarginal_R	0.82886
Occipital_Inf	0.8231
Parietal_Inf_R	0.82194
Occipital_Mid	0.82037
Angular_R	0.81803

resolution of the heat map is lower due to the expansion of the convolution kernel by the difference. Therefore, we expect to design a dedicated fMRI neural network in the next work to obtain more detailed brain region abnormalities, such as specific brain regions with functional connections and abnormal activity patterns. Finally, we plan to link these brain regions to sMRI of ASD patients so that we can better understand the underlying mechanisms of the disease and introduce personalized rehabilitation programs in the field of autism treatment.

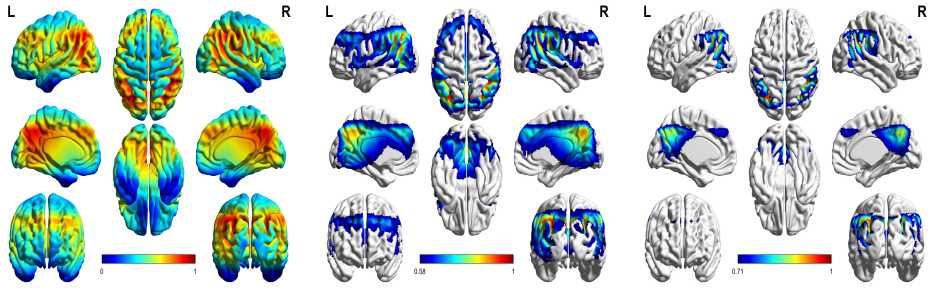


Fig. 8. Comparison of effects of suppression after different rounds; (a) is the thermogram of the brain after 10 rounds of suppression; (b) is the thermogram of the brain after 20 rounds of suppression; (c) is the thermogram of the brain after 30 rounds of suppression.

7. SUMMARY

In this paper, we explore the performance of neural network in the early screening of ASD under the principle of fMRI, and compare with the performance on sMRI. We select 3D CNN as the way of feature extraction, and add recurrent neural network to extract the features of time. Compared to sMRI, we have achieved excellent results on fMRI. In this regard, we use Grad-CAM algorithm to analyze the reasons for the excellent performance on fMRI. At the same time, by comparing the thermograms of the ASD patients with normal group, we locate the possible brain region that triggers ASD. Finally, we present some treatment recommendations for patients with ASD.

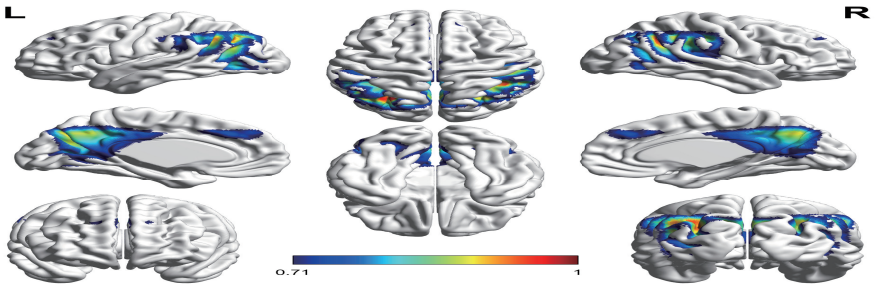


Fig. 9. The marked areas represent regions of interest that contain brain regions with different functional connections and brain regions with different patterns of activity.

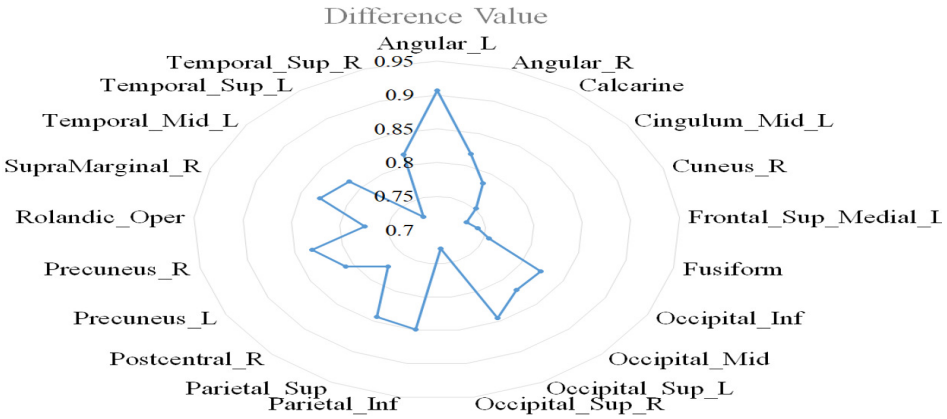


Fig. 10. The degree of difference between abnormal brain regions and normal brain regions.

ACKNOWLEDGEMENT

Thanks to the Autism Brain Imaging Data Exchange (ABIDE 1 and ABIDE 2, http://fcon_1000.projects.nitrc.org/indi/abide/) for providing data for our research. Thanks to Electronic Science and Technology of Xi'an University for equipment support for the Institute of Artificial Intelligence. Thanks to Xi'an Jiaotong University for the technical support of the Institute of Biomedical Engineering.

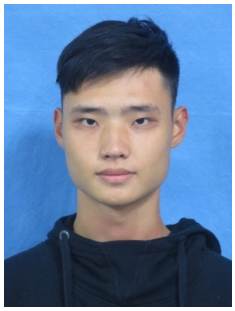
REFERENCES

1. P. Rajpurkar, J. Irvin, K. Zhu, B. Yang, H. Mehta, T. Duan, D. Ding, A. Bagul, C. P. Langlotz, K. Shpanskaya, M. P. Lungren, and A. Y. Ng, "Chexnet: Radiologist-level pneumonia detection on chest x-rays with deep learning," *CoRR*, Vol. abs/1711.05225, 2017.
2. L. Li, H. Huang, and X. Jin, "Ae-cnn classification of pulmonary tuberculosis based on ct images," in *Proceedings of the 9th International Conference on Information Technology in Medicine and Education*, 2018, pp. 39-42.
3. X. W. Gao and H. Rui, "A deep learning based approach to classification of ct brain images," in *Proceedings of Sai Computing Conference*, 2016, pp. 28-31.
4. D. Nie, H. Zhang, E. Adeli, L. Liu, and D. Shen, "3d deep learning for multi-modal imaging-guided survival time prediction of brain tumor patients," *Medical Image Computing and Computer-Assisted Intervention*, 2016, pp. 212-220.
5. C. G. L. B. Khagi and G. Kwon, "Alzheimer's disease classification from brain mri based on transfer learning from cnn," in *Proceedings of the 11th Biomedical Engineering International Conference*, 2018, pp. 1-4.
6. R. Vinoth and C. Venkatesh, "Segmentation and detection of tumor in mri images using cnn and svm classification," in *Proceedings of International Conference on Emerging Devices and Smart Systems*, 2018, pp. 21-25.

7. M. N. M. Nor, R. Jailani, and N. M. Tahir, "Analysis of emg signals of ta and gas muscles during walking of autism spectrum disorder (asd) children," in *Proceedings of IEEE Symposium on Computer Applications Industrial Electronics*, 2016, pp. 226-230.
8. M. Liu, Y. An, X. Hu, D. Langer, C. Newschaffer, and L. Shea, "An evaluation of identification of suspected autism spectrum disorder (asd) cases in early intervention (ei) records," in *Proceedings of IEEE International Conference on Bioinformatics and Biomedicine*, 2013, pp. 566-571.
9. N. Yahata, J. Morimoto, R. Hashimoto, G. Lisi, K. Shibata, Y. Kawakubo, H. Kuwabara, M. Kuroda, T. Yamada, and F. Megumi, "A small number of abnormal brain connections predicts adult autism spectrum disorder," *Nature Communications*, Vol. 7, 2016, No. 11254.
10. J. Shuiwang, Y. Ming, and Y. Kai, "3d convolutional neural networks for human action recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 35, 2013, pp. 221-231.
11. X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional lstm network: A machine learning approach for precipitation nowcasting," *Advances in Neural Information Processing Systems*, Vol. 28, 2015, pp. 802-810.
12. J. C. Rajapakse, J. N. Giedd, and J. L. Rapoport, "Statistical approach to segmentation of single-channel cerebral mr images," *IEEE Transactions on Medical Imaging*, Vol. 16, 1997, pp. 176-186.
13. A. John, "A fast diffeomorphic image registration algorithm," *Neuroimage*, Vol. 38, 2007, pp. 95-113.
14. J. D. Power, A. Mitra, T. O. Laumann, A. Z. Snyder, B. L. Schlaggar, and S. E. Petersen, "Methods to detect, characterize, and remove motion artifact in resting state fmri," *Neuroimage*, Vol. 84, 2014, pp. 320-341.
15. M. D. Zeiler, D. Krishnan, G. W. Taylor, and R. Fergus, "Deconvolutional networks," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2010, pp. 2528-2535.
16. M. D. Zeiler, G. W. Taylor, and R. Fergus, "Adaptive deconvolutional networks for mid and high level feature learning," in *Proceedings of International Conference on Computer Vision*, 2011, pp. 2018-2025.
17. M. Lin, Q. Chen, and S. Yan, "Network in network," *Computer Science*, arXiv: 1312.4400.

Mingkang Zhang received his bachelor degree in Engineering from Harbin Institute of Technology in 2017, and pursuing his mater of engineering degree in Xidian University.

Yanbiao Ma is a college student in Xidian University, researched in interpretable deep learning and evolutionary computation.



Linan Zheng is with the School of Computer Science and Technology at Xidian University as an undergraduate student since 2017.

Yuanyuan Wang received her B.Sc. degree form Xidian University in 2017, and pursuing her M.S. degree in Xidian University.

Zhihong Liu received his B.Sc. degree from National University of Defense Technology (NUDT), China, in 1989, his M.S. degree in Computer Science from Air Force Engineering University, China, in 2001, and his Ph.D. degree in Cryptography from Xidian University in 2009. Now he is with the School of Cyber Engineering at Xidian University. His research areas include mobile computing and information security. (liuzhihong@mail.xidian.edu.cn)

Jianfeng Ma received the B.S. degree in Mathematics from Shaanxi Normal University, Xi'an, China, in 1985, and the M.S. and the Ph.D. degrees in Computer Software and Telecommunication Engineering from Xidian University, Xi'an, in 1988 and 1995, respectively. From 1999 to 2001, he was a Research Fellow with the Nanyang Technological University of Singapore, Singapore. He is currently a Professor and the Ph.D. Supervisor with the Department of Computer Science and Technology, Xidian University. His current research interests include information and network security, wireless and mobile computing systems, and computer networks.

Qian Xiang received his B.Sc. degree from Xidian University in 2019, and pursuing his M.S. degree in Beihang University. His research interests include face recognition and 3D vision. (qianxiang@buaa.edu.cn)

Kexin Zhang Undergraduate in Xidian University. He wins the National Scholarship and Huawei Scholarship in 2019.

Licheng Jiao Member of Science and Technology Department of Ministry of Education Expert Group of Artificial Intelligence Science and Technology Innovation of Ministry of Education.