Thesis Defense

An Attention Network for Autism Diagnosis with Multi-modal Fusion Features Functional Connectivity

Respondent：Mr. Yang Zongxian

Instructor：Yu Li

CONTENTS

PART.01

Background and significance

Background and significance

What is autism?

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that can have a significant impact on a persons social interactions, communication and behavior. The pathological causes of autism are still undetermined and may be related to genetic, neurodevelopmental, and environmental factors, with genetic factors and synaptic connectivity abnormalities now recognized by the academic community as the main causes of ASD.

Why is automated diagnosis of autism so important?

A The earlier the treatment, the better.

Relevant studies have pointed out that the earlier the intervention, the better the outcome of the intervention, and that successful intervention can enable children with autism to achieve basic social functioning in adulthood.

B Difficulty in detecting autism

Most of the symptoms of ASD occur at school age, making early autism difficult to detect by parents or physicians, making the development of automated, low-cost methods for early diagnosis of autism essential.

PART.02

Research Status

Research status

Deep learning-assisted ASD diagnostic approach

Deep learning-assisted autism diagnostic approaches can be categorized into two main groups according to the type of data, i.e., based on extrinsic features and based on brain imaging data

External features can be divided into eye movement data and facial expression data.

Brain image data can be further divided into EEG-based data and MRI-based data.

MRI-based autism diagnosis has the advantages of non-invasiveness, high spatial resolution, and long-term monitoring, which provides an important tool and method for in-depth analysis of the brain basis of autism and aids in the diagnosis, and is also known as a research hotspot in recent years.

Deep learning-assisted ASD diagnostic modalities	
Based on external characteristics	Based on brain imaging data
Using eye movement data	Using EEG data
Using facial expressions	Using MRI data
	Using multimodal data

Research status

Development of rs-fMRI-based ASD diagnosis

priori or not | year | data volume | Selected models | accuracy

priori or not	year	data volume	Selected models	accuracy
not priori	2015	79	Grassmann Manifold	63.29%
	2016	154	SVM	63%
	2017	110	DNN	86.36%
	2019	871	CNN	65.3%
	2020	930	AE-MKFC	61%
	2022	1050	DNN	70.9%
priori	2019	1029	MT-GCN	67.3%
	2020	871	Hi-GCN	72.81%
	2022	871->5779	ViT-B/16	76.69%

Research status

Development of rs-fMRI-based ASD diagnosis

In ASD diagnosis using MRI data, resting-state MRI is favored by researchers for its validity, simplicity and large sample integration. The table on the right shows the evolution of research on autism diagnosis based on functional brain (FC) connectivity.

It can be seen that the amount of data for ASD diagnosis with FC connectivity has gradually evolved from less than a hundred samples in the beginning to nearly 1k samples now, while at the same time, as the number of samples increases, the performance of the old method decreases due to lack of robustness. The improvement of the model in turn led to a further increase in performance.

With the popularization of deep learning technology this year, a priori models represented by graph network technology have been designed and achieved some results. However, the a priori model still has some disadvantages compared to the non-priori model, such as poor objectivity and scalability, complex model algorithms, and high arithmetic requirements.

PART.03

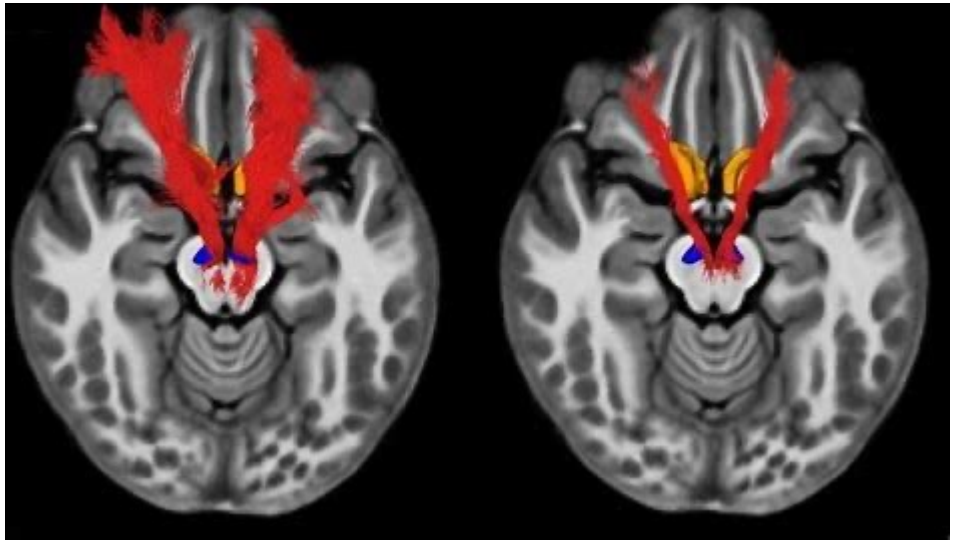
Research Content And Method

Content of the study

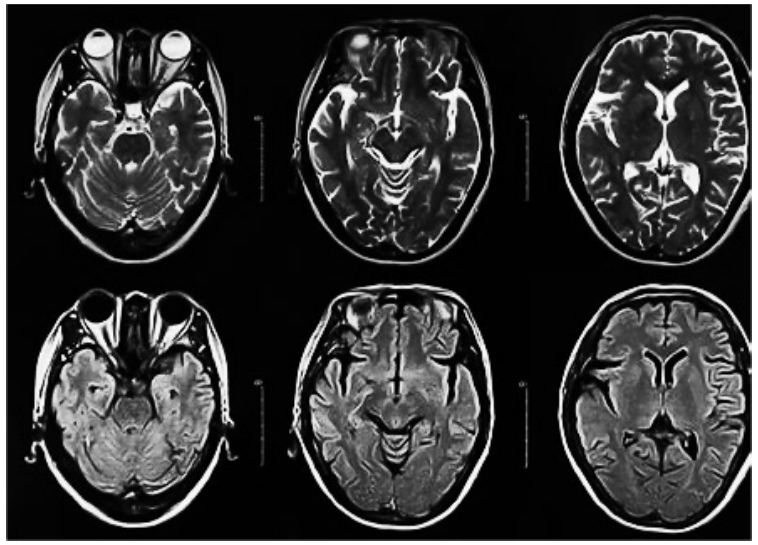
Data Source

ABIDE database

The ABIDE (Autism Brain Imaging Data Exchange) database brings together brain imaging data from autistic and typically developing individuals from all over the world, including multiple imaging modalities such as structural MRI, functional MRI, and DTI, covering data from multiple research centers and multiple scanning devices.



MRI imaging in the ABIDE database



Example of ABIDE data

Content of the study

Dataset

The table on the right provides an overview of the ABIDE data used in this study, and it is important to note that while rs- fMRI data facilitates the integration of large samples, there are still individual idiosyncrasies between the different sites. For example, one study used data from other sites to test a DNN model that performed better (acc=86.36%) at the UM site, with far less accurate results.

Therefore, in order to eliminate the idiosyncratic characteristics of each site itself and obtain a robust automated diagnostic method, the data selected and compared in this study are as shown in the figure on the right, combined with the comprehensive dataset obtained from multiple sites, and accordingly, the experimental conclusions obtained under this dataset should have a certain degree of generalizability.

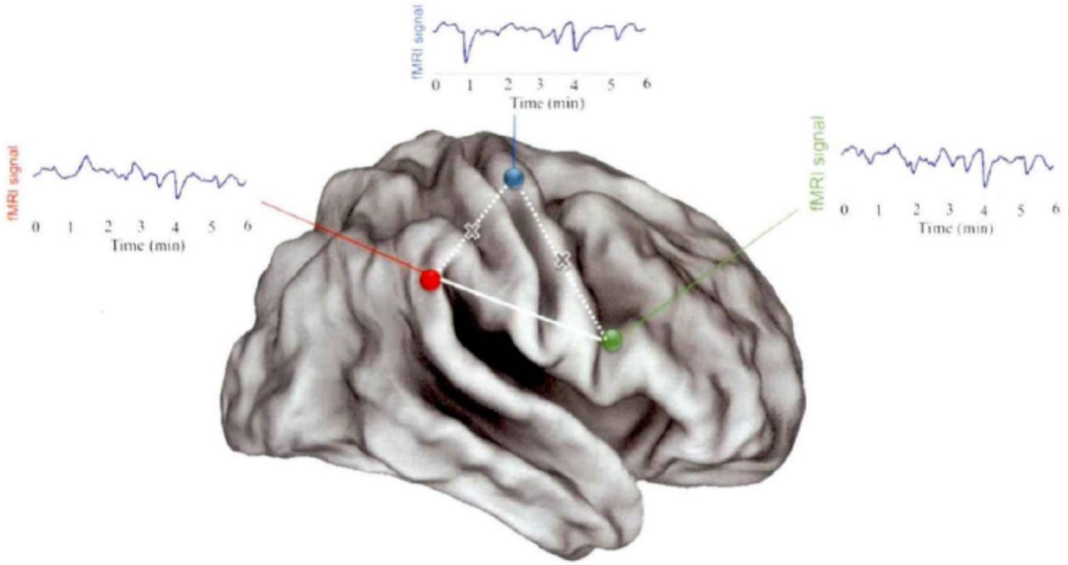
Data outlook in this study

Site	ASD or TD		Gender		numbers	Age ranges
	ASD	TD	male	female		
CALTECH1	5	10	10	5	15	17~56
CMU	6	5	7	4	11	19~33
KKI	12	21	24	9	33	8~13
LEUVEN	26	30	49	7	56	12~32
NYU	74	98	136	36	172	6~39
OHSU	12	13	25	0	25	8~15
OLIN	14	14	23	5	28	10~24
SBL	12	14	26	0	26	20~49
SDSU	8	19	21	6	27	9~17
STANFORD	12	13	18	7	25	8~13
TRINITY	19	25	44	0	44	12~26
UCLA	48	37	74	11	85	8~18
UM	47	73	93	27	120	8~29
USM	43	24	67	0	67	9~50
YALE	22	19	25	16	41	7~18
总计	403	468	727	144	871	6~58

Content of the study

Introduction to Functional Brain Connectivity (FC) Methods

The brain relies on neuronal functional interactions between different brain regions to accomplish tasks, and the essence of brain functional connectivity is to compress the complex BOLD temporal signals in rs-fMRI into correlation information between different brain regions based on different predefined templates of functional connectivity for analysis.

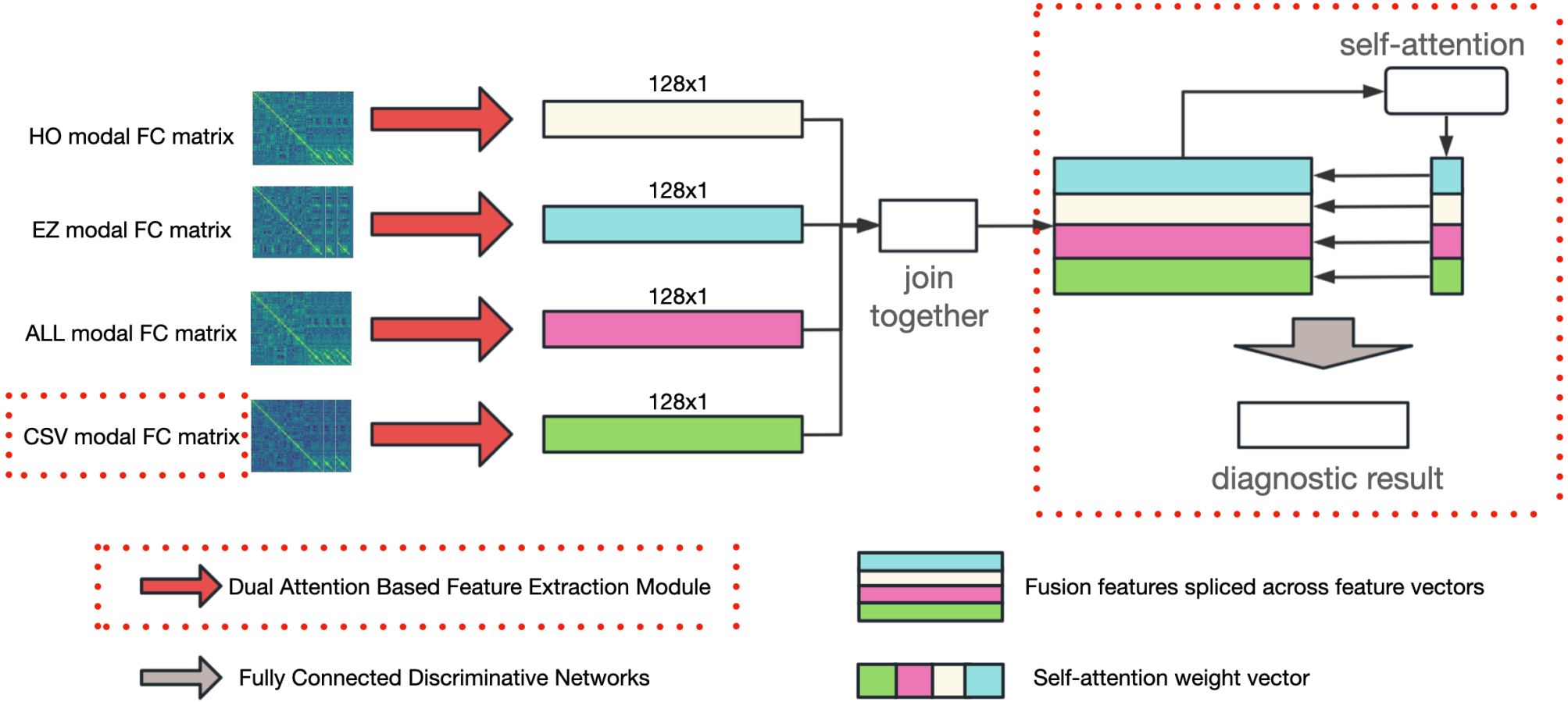


根据不同脑区fMRI信号之间的相关性确定FC关系

1. **AAL template** — Generated based on automated anatomical labeling technology
2. **HO Template** — Generated based on Harvard-Oxford Brain Mapping
3. **EZ Template** — Individual-based functional connectivity data generation

Content of the study

general overview

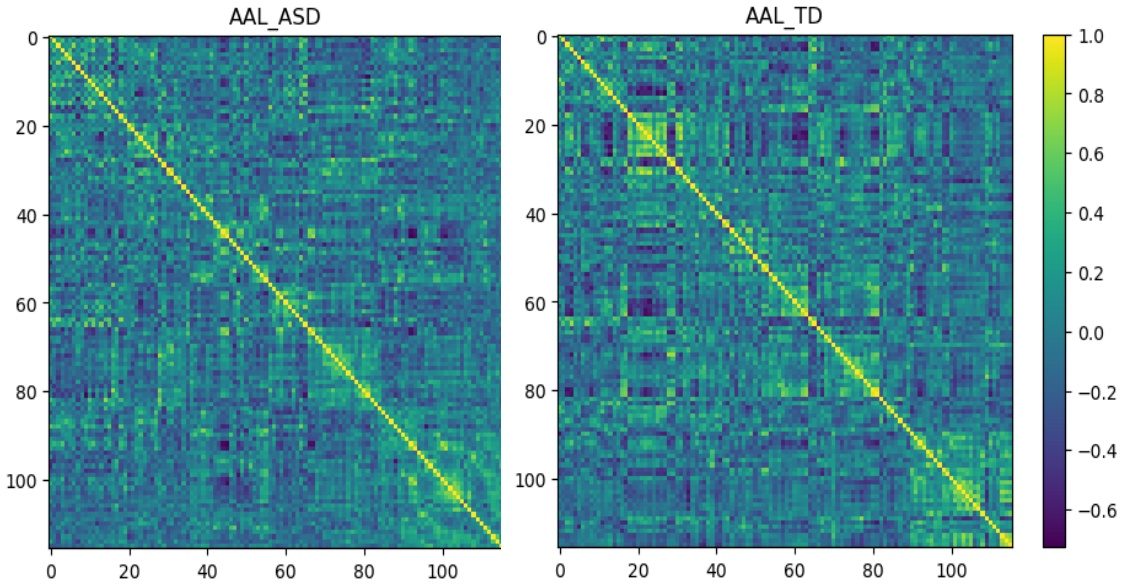


research approach

Why Convolutional Neural Networks (CNN)?

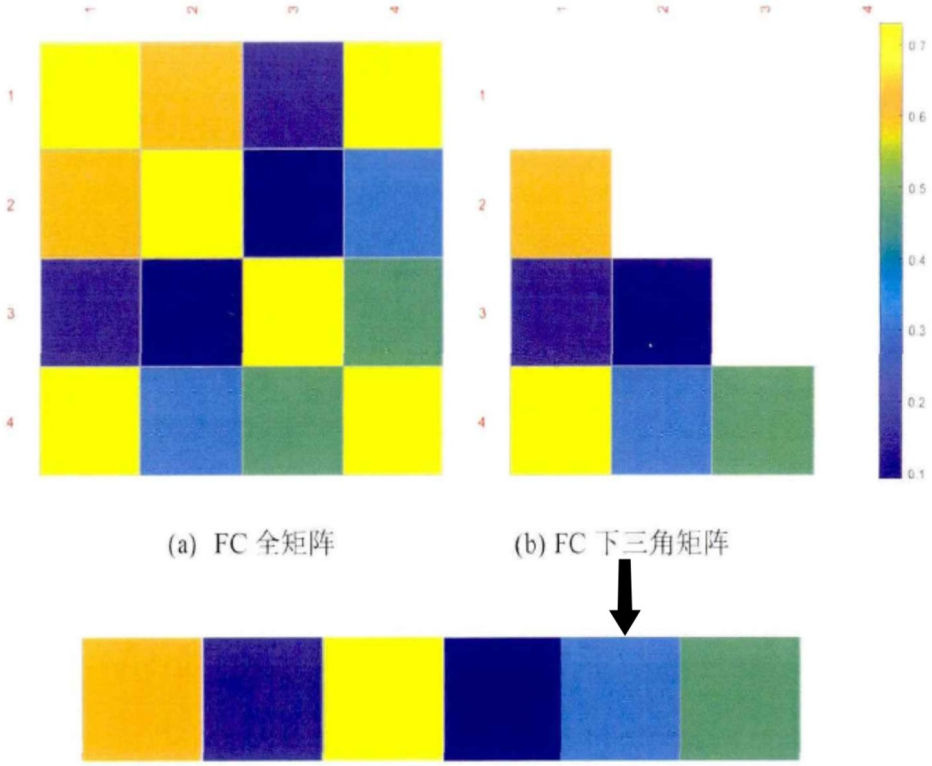
1 Appropriate data format

The correlation matrix output from rs-fMRI after FC analysis is two-dimensional and can be viewed as a grayscale image data and analyzed using proven computer vision techniques, which broadens the technological reserve compared to the poor functional connectivity analysis methods with less information.



research approach

Why CNN?



2 CNN for more information

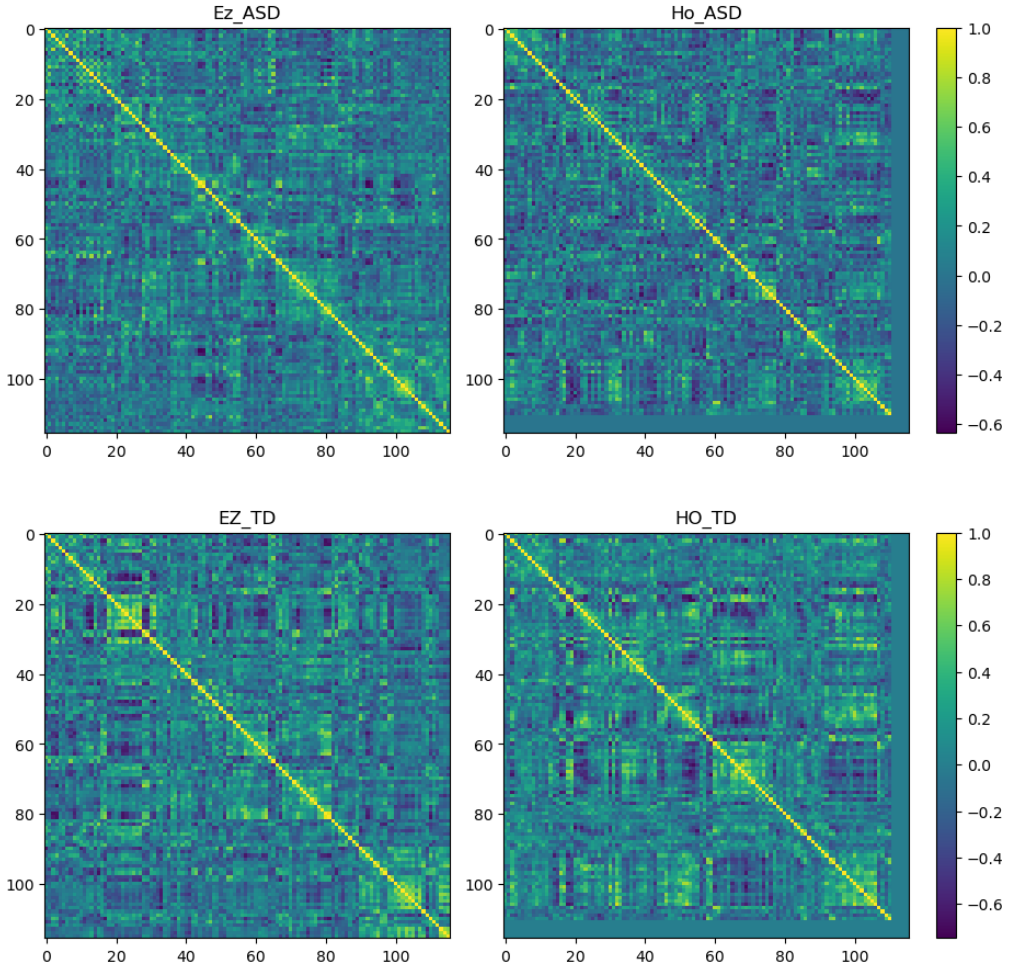
Compared with traditional methods, the nxn convolutional blocks in CNN can consider both horizontal and vertical correlation information.

Traditional methods

The traditional approach to data processing for FC is to utilize the symmetry of the FC matrix by cropping the fully connected matrix along the diagonal, setting one side of the diagonal to zero, and stretching and splicing the other side along the transversal or vertical direction into a one-dimensional vector, after which this vector is analyzed

research approach

Why CNN?



2 CNN for more information

Compared with traditional methods, the nxn convolutional blocks in CNN can consider both horizontal and vertical correlation information.

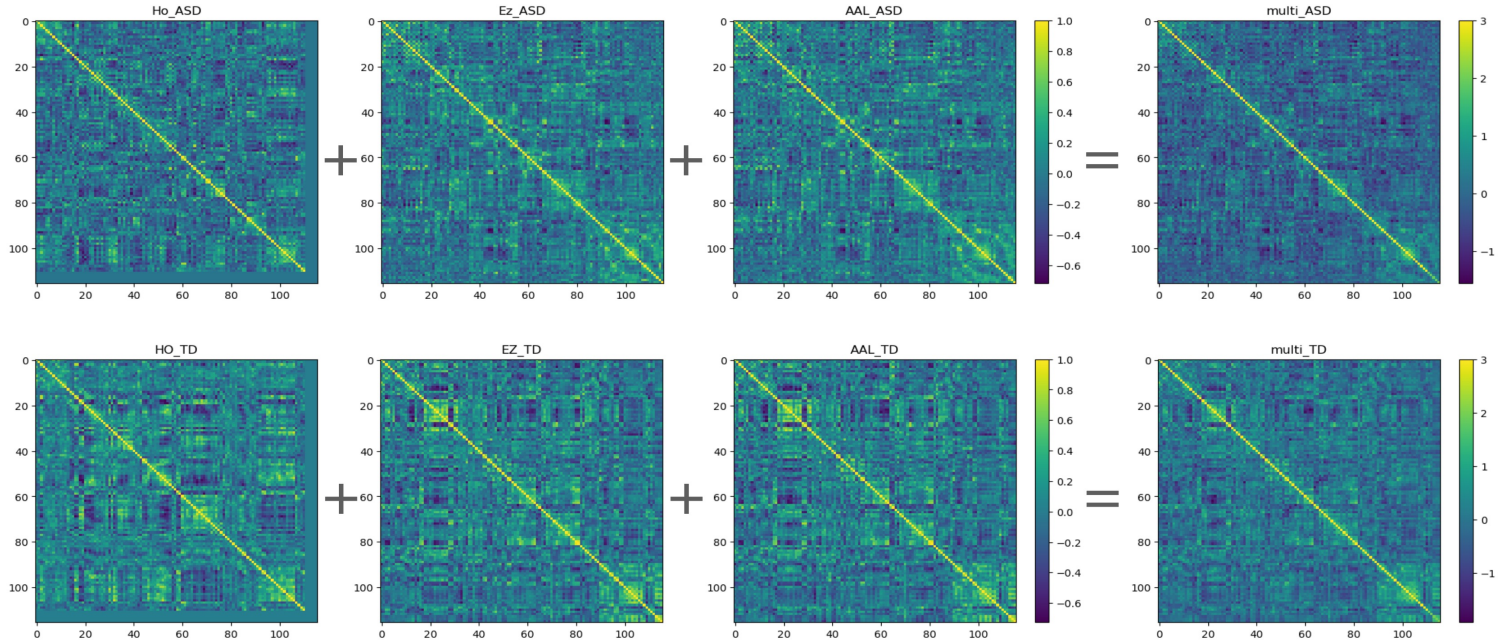
Advantages of the CNN method

1. **Capable of obtaining more information**; through the observation of the correlation matrix, it is found that the TD samples, relative to the ASD samples, the connectivity matrix shows strong continuous specificity in both the horizontal and vertical directions, so if only the vertical direction is taken into account, part of the information must have been missed.
2. S**implifies the processing flow of correlation matrix data**; compared with the complex operation of setting threshold - transforming into triangular matrix - stretching and splicing into connection vectors in traditional methods, CNN only needs to consider FC matrix as a grayscale image input model.

research approach

Perspective–level lightweight feature fusion approach

One of the reasons for using CNNs was explained earlier is that it is possible to try to utilize computer vision methods for FC matrices. Here, this study proposes a view–level lightweight feature fusion approach, which operates by simply summing the functionally connected correlation matrices obtained from aal, ho, and ez at the pixel level to obtain a correlation matrix enriched with the three modal features



Visualization of Individual ASD vs. Individual TD FC Matrix at Caltech Site

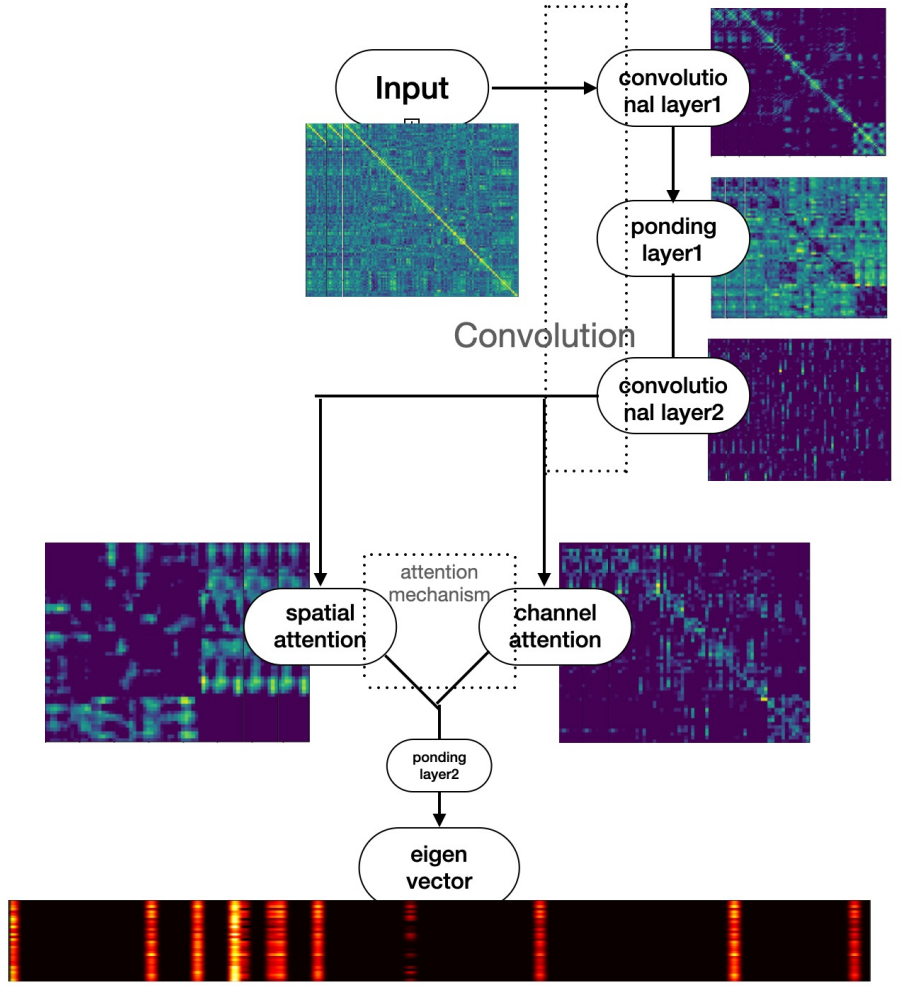
research achievement

Feature extraction module based on channel attention and spatial attention

The feature extraction module consists of two parts, the convolutional layer and the attention mechanism, and the flow is shown in the right figure.

① The convolutional layer performs preliminary feature extraction of the input data by CNN, including two convolutional layers and one fully connected layer. Each convolutional layer is followed by a maximum pooling layer, with a convolutional kernel size of 3×3, a step size of 1, and a padding of 1 to keep the feature map consistent with the original image size.

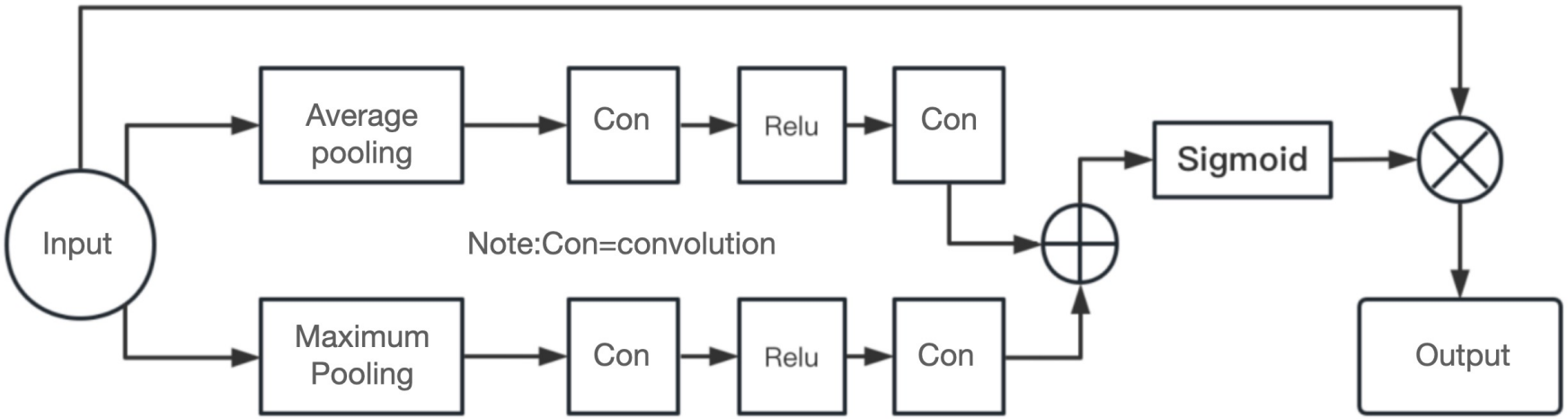
② The attention mechanism is introduced to enhance the attention to key features, including channel attention and spatial attention, which focus on the channel dimension and spatial distribution of features, respectively, and adjust the weights of input features through operations such as pooling, convolution and activation function to enhance the attention to key features.



research achievement

Channel Attention Module

Channel attention focuses on the channel dimensions of features. First, the spatial information of each channel is compressed using adaptive mean pooling and maximum pooling. Then the features are processed by two 1x1 convolution operations and ReLU activation function to process the results of average pooling and maximum pooling, respectively, and these two sets of features are summed up. Finally, the weights of each channel are obtained by sigmoid function to adjust the input features for each channel.

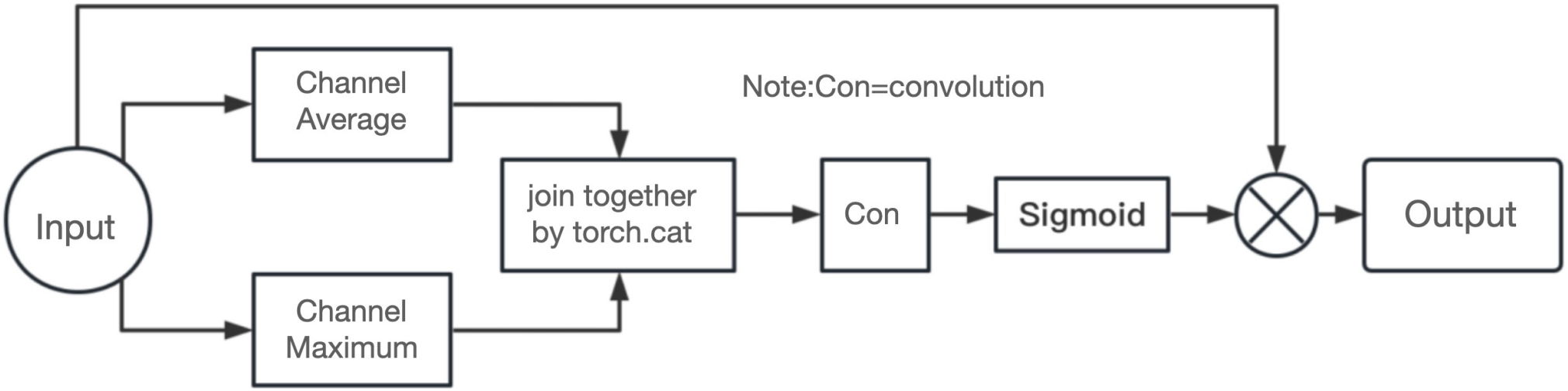


Channeled Attention Network

research achievement

Spatial Attention Module

Spatial attention focuses on the spatial distribution of the features and obtains two sets of features by computing the mean and maximum values for each channel, which are then spliced in the channel dimension. Then this new set of features is processed by a convolution operation, and finally the weights of each position are obtained by a sigmoid function, which is used to adjust each position of the input features.

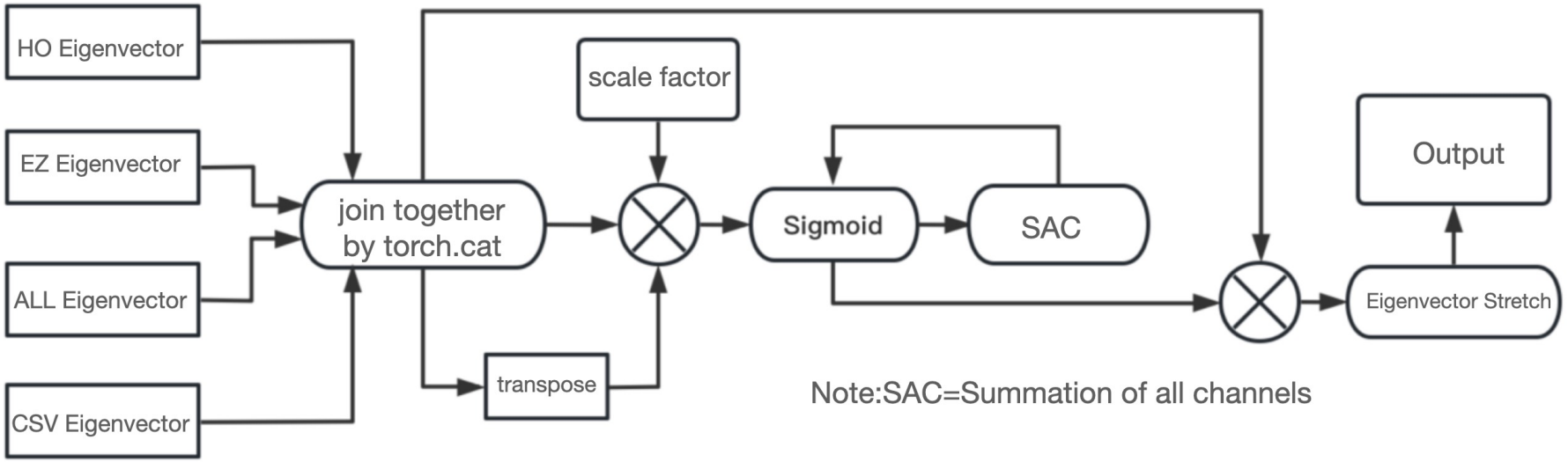


Spatial Attention Network

research achievement

Self-attention based feature fusion discrimination module

In this study, the functional connectivity matrix features of four modalities (fusion matrix, HO, EZ, and AAL) were obtained through feature extraction to describe different brain network characteristics. In order to effectively utilize the advantages of these features, a feature fusion module based on the self-attention mechanism was designed. First, the features of the four modalities are spliced into a new feature mapping. Then, the similarity between the feature mappings is calculated and the importance of different modalities is indicated by the attention weights. The features of each modality are multiplied by the weights and summed to obtain the fused functional connectivity matrix feature. Finally, the features are fed into the classifier for autism diagnosis.



Self-Attention Network

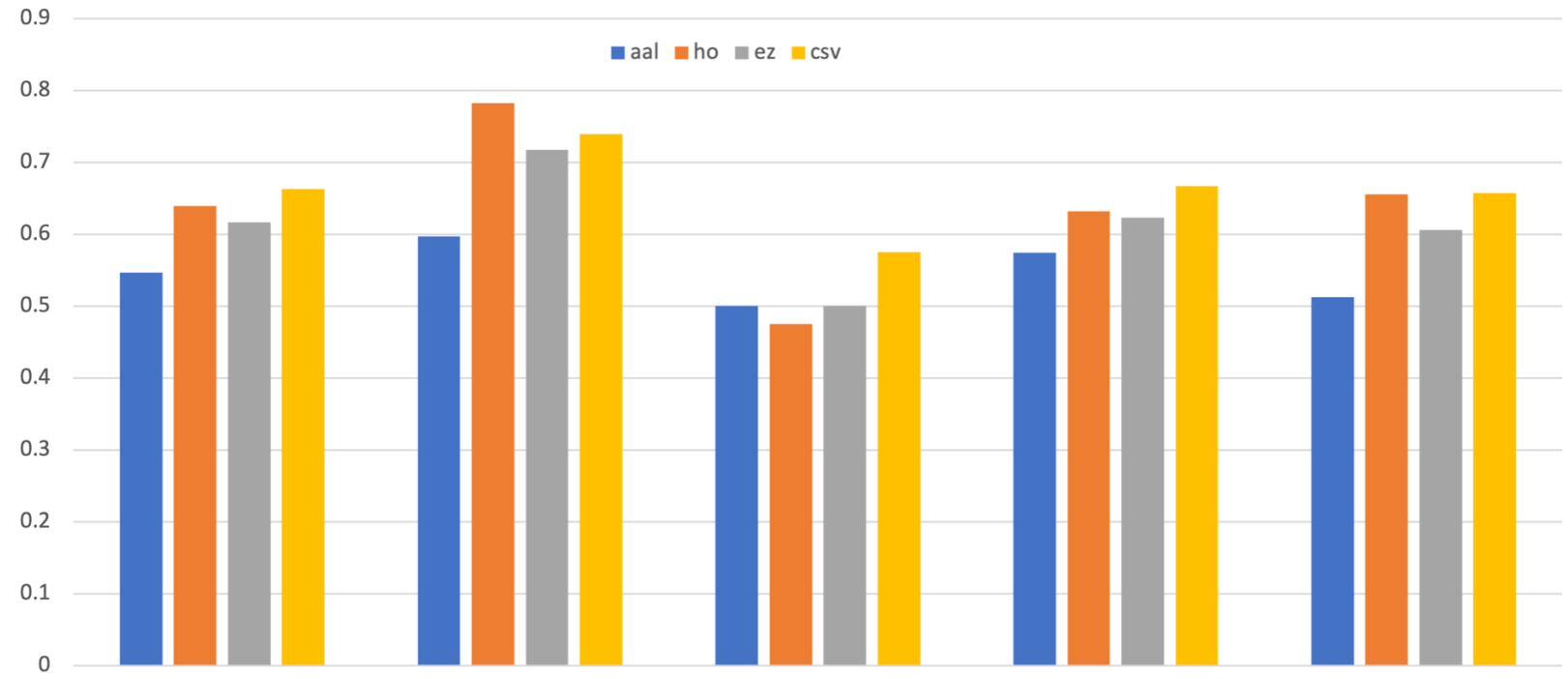
PART.04

Experimental Results And Summary Of Our Work

Ablation Experiment

Perspective-level lightweight feature fusion approach

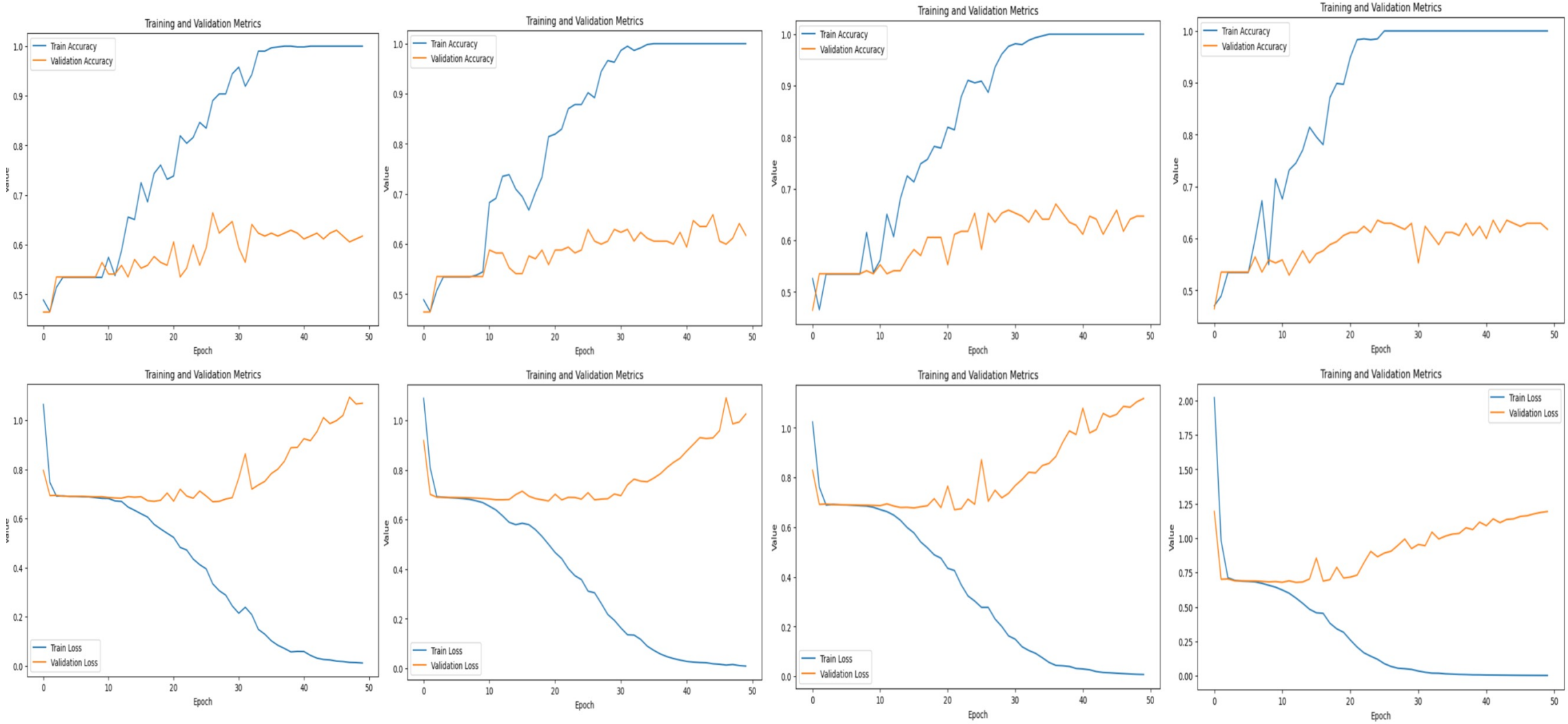
In this study, ablation experiments were done based on the underlying CNN network, and it can be seen that, except for a slight decrease in the sensitivity of the ez modality, the csv fusion method achieved a considerable degree of improvement in all other metrics, which proves the effectiveness of the method.



	Accuracy	Sensitivity	Specificity	PPV	NPV
aal	0.5465	0.597	0.5	0.5745	0.5128
ez	0.6395	**0.7826**	0.475	0.6316	0.6552
ho	0.6163	0.7174	0.5	0.6226	0.6061
csv	**0.6628**	0.7391	**0.575**	**0.6667**	**0.6571**

ablation experiment

Perspective-level lightweight feature fusion approach



AAL HO EZ CSV

Ablation Experiment

Feature extraction module based on channel attention and spatial attention

In this study, the corresponding feature extraction module is added to the basic CNN network to construct the discriminator, and ablation experiments are designed to explore the contribution of each attentional module to the model performance, and the individual performance metrics obtained are shown below:

CNN：basic network　　SCN：Single-Channel Attention Networks

SSN：Single-space attention network　　DAN：Channel and Space Dual Attention Combining Networks

ho					
	acc	sensitivity	specificity	PPV	NPV
CNN	0.6395	0.7826	0.475	0.6316	0.6552
SCN	0.6395	0.6957	0.575	0.6531	0.6216
SSN	0.6744	0.6304	0.725	0.725	0.6304
DAN	0.6279	0.8261	0.4	0.6129	0.6667

csv					
	acc	sensitivity	specificity	PPV	NPV
CNN	0.6628	0.7391	0.575	0.6667	0.6571
SCN	0.6279	0.6957	0.55	0.64	0.6111
SSN	0.6047	0.6957	0.5	0.6154	0.5882
DAN	0.6463	0.5	0.75	0.697	0.566

ez					
	acc	sensitivity	specificity	PPV	NPV
CNN	0.6163	0.7174	0.5	0.6226	0.6061
SCN	0.5581	0.413	0.725	0.6333	0.5179
SSN	0.5465	0.3261	0.8	0.6522	0.5079
DAN	0.5698	0.587	0.55	0.6	0.5366

aal					
	acc	sensitivity	specificity	PPV	NPV
CNN	0.5465	0.597	0.5	0.5745	0.5128
SCN	0.5465	0.4783	0.625	0.5946	0.5102
SSN	0.5233	0.6087	0.425	0.549	0.4857
DAN	0.5465	0.6522	0.425	0.566	0.5152

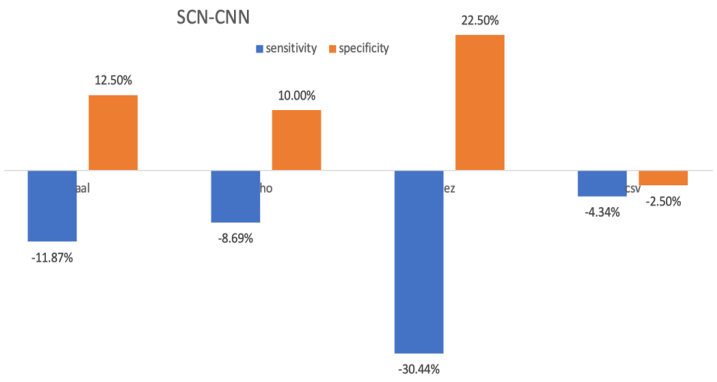
Ablation Experiment

Feature extraction module based on channel attention and spatial attention

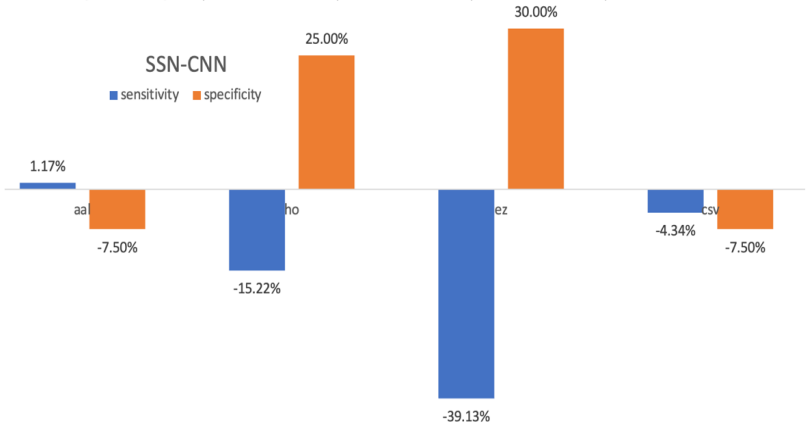
Analyzing the above data, the specificity and sensitivity metrics of the SCN, SSN, and DAN networks are differed from those of the CNN network, as shown in the following figure, which shows that, except for the single–channel CSV modality, the introduction of the various attention modules in all the other modalities improves one of the sensitivities or specificities to a certain degree and suppresses the other metrics.

This represents that after the introduction of the attention mechanism, the model can autonomously learn and strengthen the ability to extract the features of ASD or TD according to the unique characteristics of the data of each modality, thus improving the performance of the model.

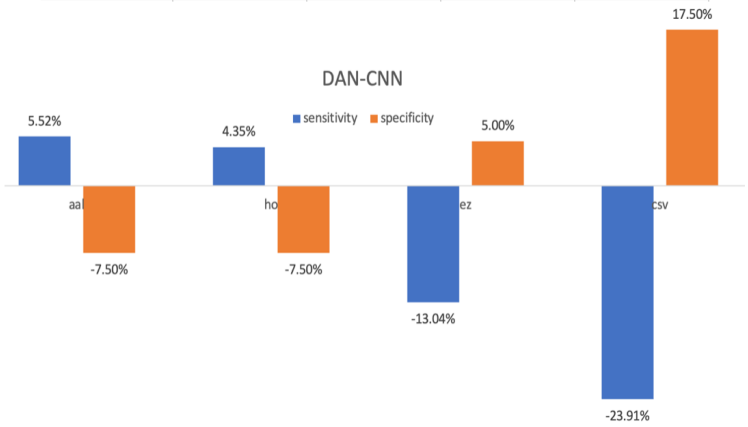
SCN-CNN	aal	ho	ez	csv
sensitivity	-11.87%	-8.69%	-30.44%	-4.34%
specificity	12.50%	10.00%	22.50%	-2.50%



SSN-CNN	aal	ho	ez	csv
sensitivity	1.17%	-15.22%	-39.13%	-4.34%
specificity	-7.50%	25.00%	30.00%	-7.50%



DAN-CNN	aal	ho	ez	csv
sensitivity	5.52%	4.35%	-13.04%	-23.91%
specificity	-7.50%	-7.50%	5.00%	17.50%



Ablation Experiment

Self-attention based feature fusion discrimination module

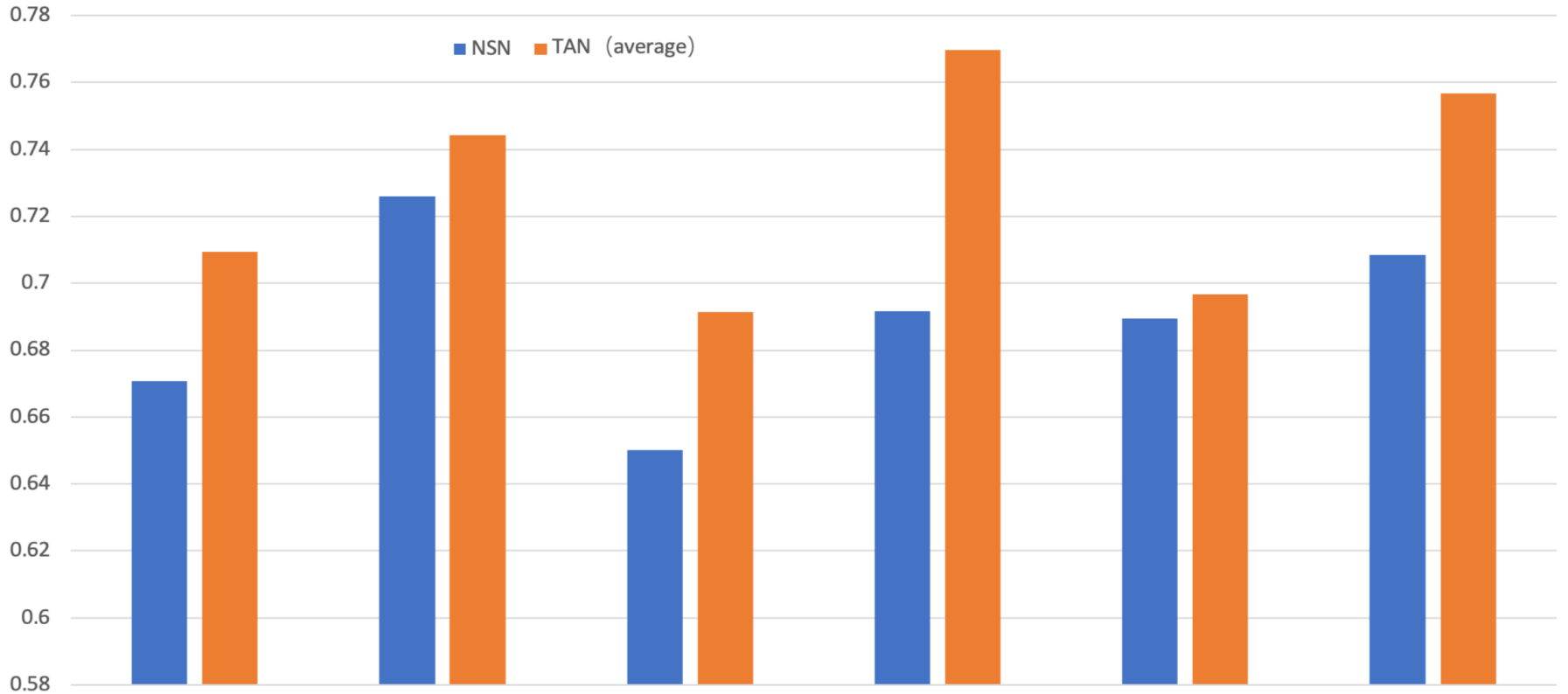
In order to effectively utilize the features of brain functional connectivity matrices of different modalities to improve the diagnostic performance of the model, this study constructed a self-attention based feature fusion discriminant module based on DAN and named it TAN(Three Attention Network). A series of ablation experiments were designed in this part aiming to analyze the contribution of the self-attention module to the performance of the model. As a comparison, the network using a simple fully connected layer as the feature fusion discriminant module is named NAN (No Self Attention). The performance metrics for each of the 5 folds of cross-validation of TAN are shown in the table below.

表 4-4 TAN 网络的 5 折交叉验证结果

	accuracy	sensitivity	specificity	ppv	npv
1 折	0.6924	0.7226	0.6728	0.7522	0.6917
2 折	0.7143	0.7487	0.6941	0.7721	0.7079
3 折	0.7158	0.7522	0.6994	0.7766	0.6912
4 折	0.7069	0.7453	0.6925	0.7697	0.6973
5 折	0.7181	0.7525	0.6987	0.7779	0.7054
average	**0.7095**	**0.74426**	**0.6915**	**0.7697**	**0.6967**

Ablation Experiment

Self-attention based feature fusion discrimination module



	Accuracy	Sensiticity	Specificity	PPV	NPV	F1
NAN	0.6707	0.7261	0.65	0.6917	0.6895	0.7084
TAN	0.7095	0.7442	0.6915	0.7697	0.6967	0.75

TAN和NAN各个性能指标的对比

Comparison Experiment

Comparative experiments on a priori networks

Now the autism discrimination based on FC connectivity can be categorized into two kinds according to whether a priori brain science knowledge is required or not, i.e., autism diagnostic network designed with a priori knowledge is required and non-priori autism diagnostic network.

The former requires researchers to design the network and process the data based on their own understanding of the sample FC connections, while the latter utilizes a large number of data samples for training and automatically extracts and selects autism-related features by learning the intrinsic structure and characteristics of the data, which is more objective, robust and scalable than the method requiring a priori knowledge.

Comparing the performance metrics of the a priori network we can find that although there is a slight lack of accuracy and specificity, it is still able to achieve a small lead in sensitivity.

However, compared with the a priori method, the model in this study has some advantages because it does not need a priori knowledge and has more objectivity, scalability and robustness.

priori or not	Network	Template used	Accuracy (%)	Sensitivity (%)	Specificity (%)
priori	Hi-GCN	AAL	72.81	72.04	73.91
	VIT-B/16	AAL	**76.69**	74.18	72.91
not priori	our-model	Multi-mode	70.95	**74.42**	69.15

Performance comparison of this study with a priori networks

Comparison Experiment

Comparative experiments on non–priori networks

Now the autism discrimination based on FC connectivity can be categorized into two kinds according to whether a priori brain science knowledge is required or not, i.e., autism diagnostic network designed with a priori knowledge is required and non–priori autism diagnostic network. The former requires researchers to design the network and process the data based on their own understanding of the sample FC connections, while the latter utilizes a large number of data samples for training and automatically extracts and selects autism–related features by learning the intrinsic structure and characteristics of the data, which is more objective, robust and scalable than the method requiring a priori knowledge.

Comparing the performance metrics of non–priori networks we can find that the fusion feature method proposed in this study has different degrees of lead in accuracy and sensitivity.

And compared with the same method using CNN, the accuracy of the model in this study is improved by 5.65%. All these results fully confirm the effectiveness and superiority of our research method.

priori or not	Network	Template used	Accuracy (%)	Sensitivity (%)	Specificity (%)
not priori	MC-NFE	BASC	68.42	70.05	63.34
	DNN	CC200	70	74	63
	SVM	MSDL	67.90	58.10	**76.60**
	CNN	RSNs	65.30	---	---
	our-model	Multi-mode	**70.95**	**74.42**	69.15

Comparison of the performance of this study with non–priori networks

Findings of the study

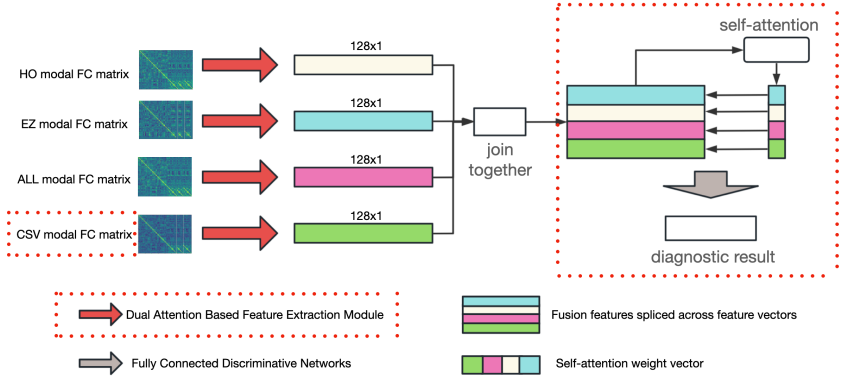
summarize

Aiming at the problem of the difficulty of early autism diagnosis detection and the subjectivity of diagnosis in the clinical field, this study proposes an automated autism diagnosis method by fusing features. And the three modules of feature extraction, feature fusion, and classifier training in the multimodal problem are optimized:

①Feature extraction: In this study, two attention mechanisms are combined to make the model take into account both local and global important information in feature extraction, and the ASD/TD feature extraction capability of the model is successfully enhanced according to the modal data characteristics.

②Feature fusion: In this study, a lightweight feature fusion method is proposed, which successfully fuses modal features and improves the models performance indexes at the same time.

③Classifier training: In this study, a self-attention-based feature fusion module is designed. The self-attention mechanism can help the model adaptively weigh the degree of dependence on different modalities, organically combine the ASD/TD-specific features extracted from each modality, and improve the performance in each dimension.



PART.05

Outlook Of Our Work

Outlook Of Our Work

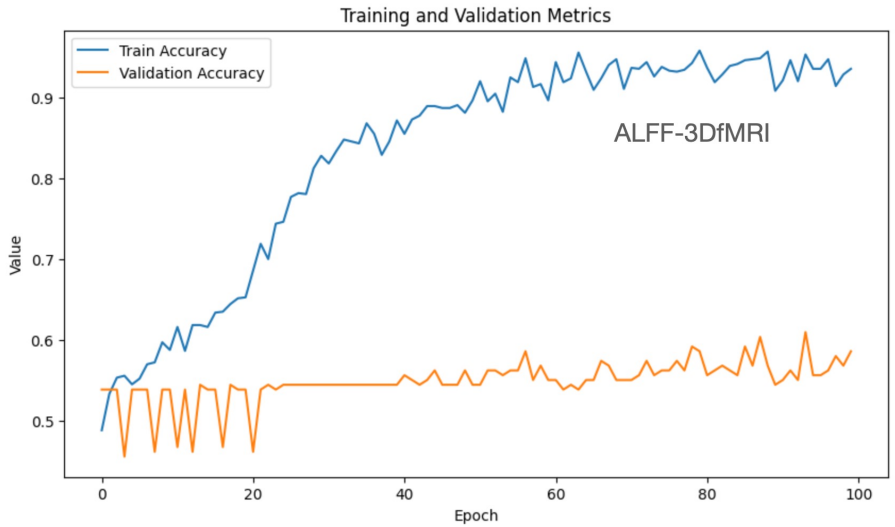
Experiencing this graduation design, I have a very deep understanding of the ABIDE database and FC connection, and also made some attempts that were not written in the thesis, so I have also gained some experiences and lessons, accordingly, I have the following outlook on the diagnosis of ASD:

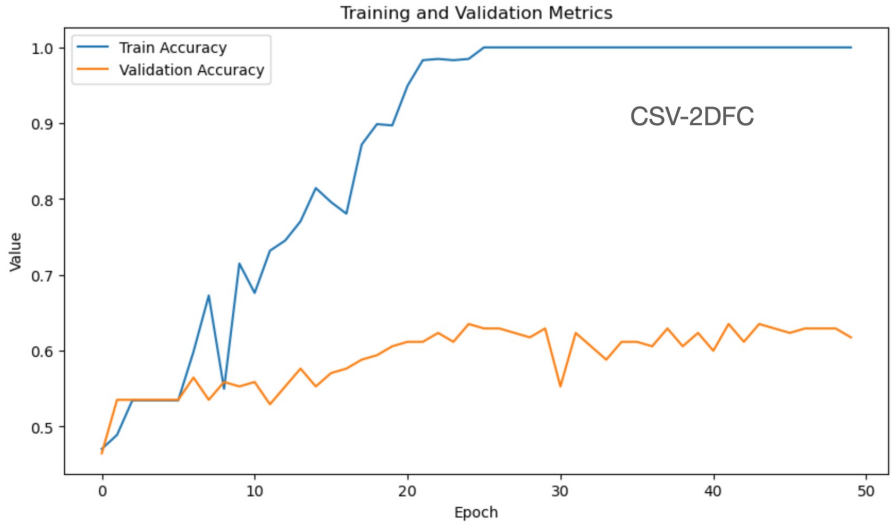
① Augmented dataset: The ABIDE I database has only more than 1000 samples, which is not enough to meet the training requirements of complex models. Based on this, Wang designed a sample augmentation model based on the ROI time series, augmented the 1000 samples into nearly 6000 data, and trained the VIT model with good results.

②Utilizing a priori information and using graph networks: a review of the literature reveals that graph networks have achieved good results in the field of functional connectivity, but graph networks are often constructed based on FC preprocessing templates, most commonly such as AAL to construct graph networks. Therefore in addition to the difficulty of requiring a priori knowledge, there may be the problem of difficulty in integrating multimodal data. However, if successfully overcome, it is believed that the accuracy of ASD diagnosis will rise to a new level.

Outlook Of Our Work

③Combining other modal data: ABIDE provides not only ROI time series, but also 3D data of various MRI images, which can be used to design a feature extraction model for 3D MRI data, extract 3D features and discriminate them together with FC features. I also conducted a similar study to design a self-attention based 3DCNN model for ALFF-3DfMRI and achieved good results. However, after integrating 3D features with FC, the accuracy decreased instead of increasing, and the adjustment was fruitless.
Analysis should be the reason that the difference between the convergence speed of the two models is too large, but due to arithmetic reasons, the solution is not fruitful.
In the future, if I have the opportunity to continue, I may follow this direction to continue to do.





Thanks