Research on EEG Recognition Based on Deep Neural Networks with Attention Mechanism

Yufeng Wang, Geng Fang, Binbin Chen, Zhengdong Chen*
School of Information Engineering, Nanjing Polytechnic Institute, Nanjing, Jiangsu, China
**Corresponding Author.*

Abstract: Electroencephalogram(EEG) is widely applied in brain-computer interface systems. Due to its high time resolution, portability, and non-invasive characteristics, it has been applied for the rehabilitation of disabled patients. At present, deep learning has also been applied in EEG feature extraction and pattern classification. However, most EEG features in deep neural networks lack discriminative ability, which limits the representation ability of these methods. To overcome this issue, two methods were proposed, named ConvNet_S_A and ConvNet_S_CA, which are based on the backbone network Shallow ConvNet. The proposed methods combine the channel attention mechanism (SE module) and the convolutional block attention module (CBAM module) to learn discriminative EEG features, respectively. Experimental results on the public EEG datasets and self-collected EEG datasets demonstrate the effectiveness.

Keywords: Electroencephalogram; Motor Imagery; Deep Neural Networks; Attention Mechanism

1. Introduction

With the Brain-computer interface(BCI)[1] system, researchers can control and operate external devices directly through their brain, thus avoiding direct limb movements. BCI can decode the dynamic changes of brain neural signals, understand the mechanism of brain activity and the presentation mode of neural signals, and provide the important clinical basis and technical means for the auxiliary diagnosis, evaluation, prediction, and rehabilitation of certain diseases(such as epilepsy).

Electroencephalogram has the advantages of noninvasive recording equipment, less environmental restrictions, and strong practicability. According to different neural patterns, EEG signals can be classified into motor imagery, P300, and other patterns. EEG signals based on motor imagery can be applied to detect the motor intentions of individuals without any external stimulation. Motor imagery means that the human brain acts by imagining its own body or part of limbs, which can be regarded as a potential cognitive processing[2].

However, EEG recognition methods in practical applications are always limited, which brings certain challenges to the construction of BCI systems with high automation and reliability. Therefore, building an accurate EEG recognition method is the key for motor imagery BCI systems to move from laboratory research to practical applications.

With the progress of machine learning technology, researchers have proposed a variety of effective methods, such as common spatial pattern, independent component analysis and other algorithms. After obtaining EEG features, a large number of classification methods, such as k-nearest neighbor and extreme learning machine are commonly applied for pattern classification of EEG features. However, traditional EEG recognition methods must rely on professionals to assist in the preprocessing and feature extraction process of EEG signals.

Nowadays, deep learning[3] has achieved great success because it can adaptively learn high-level features. Huang et al.[4] respectively applied deep belief network to represent and classify motor imagery EEG signals. The recognition results are better than those of traditional machine learning methods. Schirrmeister et al.[5] proposed Shallow ConvNet and Deep ConvNet methods respectively by simulating the FBCSP algorithm. Their effectiveness is verified on the public EEG datasets. Although the above deep learning methods abandon the traditional artificial feature extraction process and can significantly improve the recognition effect of EEG signals, they still ignore the problems: Since information collected by different electrodes have different contributions to EEG recognition, adaptively learning the weights of EEG signals from different electrodes can enhance the discriminative deep feature learning ability of the method.

2. Related Work

Some researchers have achieved fruitful results in the application and theoretical research of intelligent EEG recognition. For example, Dornhege et al.[6] first applied the common spatial pattern algorithm to distinguish EEG patterns related to motor imagery. German neuropsychologist[7] developed a device can enable patients with atretic syndrome to express their thoughts through letter spelling with the help of visual feedback training. The Language Engineering Laboratory of China has successfully created Chinese Ring Speller, which can convert EEG signals into traditional Chinese. This project has achieved a breakthrough in using EEG Chinese input zero.

With the progress of deep learning technology, An et al.[8] proposed multiple belief networks were trained and treated as multiple weak classifiers, and then the AdaBoost algorithm was applied to assemble these weak classifiers to form a strong classifier for EEG signal recognition. Tabar et al.[9] realized motor imagery EEG recognition by combining deep CNN and stacked autoencoder.

However, most of the current EEG recognition methods cannot discriminate information, which limits the representation ability of the method. Attention mechanisms can help neural networks select the most useful discriminative information. This paper proposes two deep convolutional neural networks ConvNet_S_A and ConvNet_S_CA based on different attention mechanisms to learn and recognize the discriminative features of EEG.

3. Proposed Methodology

The proposed networks use Shallow ConvNet(ConvNet_S) as the backbone network and propose two different deep EEG recognition methods based on attention mechanism by fusing channel attention mechanism(SE module) and convolutional block attention module(CBAM module) respectively.

3.1. Backbone

Shallow ConvNet(ConvNet_S)[5] is a deep convolutional neural network specially developed for EEG recognition inspired by the FBCSP algorithm. The ConvNet_S consists of 4 layers in total. Figure 1 shows the network architecture with the BCI IV IIa dataset[10] as the network input.

The input samples are 22 channels in total, and each channel has EEG signal data of 1125 sampling points, that is, the size of the input sample matrix is 22×1125. The first layer is applied for convolving the time information of EEG data. The size of 40 convolution kernels is 1×25, and the stride is 1, which is equivalent to 40 time domain filters, corresponding to the steps of extracting different frequency features in the FBCSP algorithm. The second layer uses 40 convolution kernels of size $E \times 1$, and the stride is 1. E represents the number of electrodes. The layer is equivalent to 40 spatial filters to extract spatial features, corresponding to the spatial filtering step in the CSP algorithm.

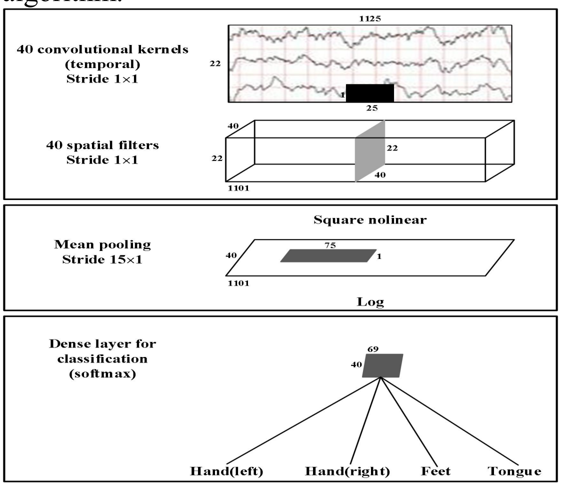


Figure 1. Structure of ConvNet_S

Then, a squared nonlinear activation function and an average pooling layer with a pooling size of 1×75 and the stride of 15 locally average the features extracted by the second layer. The above several steps correspond to the log-variance feature extraction step in the FBCSP algorithm, which is also similar to the feature extraction of differential entropy.

The last layer is a fully connected layer plus a $softmax$ activation function. The $softmax$ function maps the outputs of multiple neurons onto an interval of (0,1) that sums up to 1. This is equivalent to the classification step in the FBCSP algorithm.

Note that ConvNet_S adds a batch normalization(BN) layer before the squared nonlinear activation function and a dropout layer with the parameter set to 0.5. The network was trained with a batch size of 64 and optimized by the Adam algorithm.

Compared with the FBCSP algorithm, ConvNet_S integrates all steps into a deep convolutional neural network, adjusts the parameters of all steps adaptively during the network training process, and can be directly applied for multi-classification problems.

3.2. EEG Recognition Based on SE Module

3.2.1. SE Module

According to the importance of different channels, a discriminative feature representation of the data is extracted. Jie et al.[11] proposed the channel attention mechanism(SE module), which is effectively integrated with the deep neural network and can adaptively select important information between different channels.

The SE module adopted in this paper is shown in Figure 2.

Let $\boldsymbol{X}=\{\boldsymbol{X}_1, \boldsymbol{X}_2, \cdots, \boldsymbol{X}_c\}$ denote the input sequence, where the size of each input $\boldsymbol{X}_c(c=1,2,\cdots,C)$ is $H\times W$. To improve the global spatial information, the SE module uses the global average pooling to generate its channel statistical information, which is denoted by $\boldsymbol{z}=[\boldsymbol{z}_1, \boldsymbol{z}_2, \cdots, \boldsymbol{z}_c]$, and the cth element in the channel statistical information $\boldsymbol{z}$ can be obtained by the expression as follows:

$$\boldsymbol{z}_c = H_{GP}(\boldsymbol{X}_c) = \frac{1}{H\times W}\sum_{i=1}^{H}\sum_{j=1}^{W}\boldsymbol{X}_c(i,j) \quad (1)$$

In the expression, $\boldsymbol{X}_c(i,j)$ represents the value of $\boldsymbol{X}_c$ at position (i,j) and $H_{GP}(\cdot)$ represents the global average pooling.

To improve the versatility of the method, the SE module uses two fully connected layers. The first layer uses the ratio of r to reduce the dimension, and its weight and bias are $\boldsymbol{W}_1 \in R^{\frac{C}{r}\times C}$ and $\boldsymbol{b}_1$, respectively. The second layer is the dimension ascending layer, whose weights and bias are $\boldsymbol{W}_2 \in R^{C\times\frac{C}{r}}$ and $\boldsymbol{b}_2$, respectively. Therefore, the expression of the SE module is as follows:

$$\boldsymbol{s} = sigmoid\big(\boldsymbol{W}_2(ReLU(\boldsymbol{W}_1\boldsymbol{z}+\boldsymbol{b}_1)+\boldsymbol{b}_2)\big) \quad (2)$$

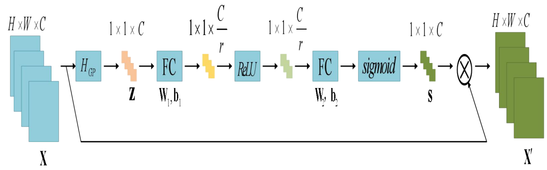


Figure 2. Structure of SE Module

In the expression, $\boldsymbol{s}=[\boldsymbol{s}_1, \boldsymbol{s}_2, \cdots, \boldsymbol{s}_C]$ represents the importance information contained in each channel, which is applied as the weight to re-weight the input data. The output of the channel c after weighting can be expressed as follows:

$$\boldsymbol{X}_c^{'} = H_{scale}(\boldsymbol{X}_c, \boldsymbol{s}_c) = \boldsymbol{s}_c\boldsymbol{X}_c \quad (3)$$

In the expression, $\boldsymbol{X}^{'}=\{\boldsymbol{X}_1^{'}, \boldsymbol{X}_2^{'}, \cdots, \boldsymbol{X}_C^{'}\}$ represents the weighted output of the network based on the SE module, and $H_{scale}(\boldsymbol{X}_c, \boldsymbol{s}_c)$ represents the multiplication of the input of the cth channel and the attention information of the cth channel.

3.2.2. ConvNet_S_A: A Novel Method for EEG Recognition Based on SE Module

Considering that the SE module can learn the important information of different channels of the EEG feature map, a novel method named ConvNet_S_A is proposed which combines the SE module into ConvNet_S to explore the influence of different channel information of the EEG feature map on EEG signal recognition.

In Figure 3, the SE module is added after Conv2. Let $\boldsymbol{C}=\{\boldsymbol{C}_1, \boldsymbol{C}_2, \cdots, \boldsymbol{C}_n\}$ denote the spatial feature extracted by the spatial convolutional layer of ConvNet_S, and $\boldsymbol{C}_i=[\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_m](i=1,2,\cdots,n)$ denote the ith spatial feature, then $\boldsymbol{c}_j(j=1,2,\cdots,m)$ denote the jth channel feature in the ith spatial feature. The SE module is applied to adaptively solve the importance weight of each channel, and the contribution to the final classification result is determined according to the channel weight.

The average pooling of the spatial features extracted from each spatial convolutional layer is applied as the channel statistical information. The specific expression is as follows:

$$\boldsymbol{u}_j = \frac{1}{H\times W}\sum_{h=1}^{H}\sum_{w=1}^{W}\boldsymbol{c}_j, \forall j=1,2,\cdots,m \quad (4)$$

In the expression, $\boldsymbol{u}_j(j=1,2,\cdots,m)$ represents the information of the jth channel, and further maps this information, the specific formula is as follows:

$$\boldsymbol{v} = sigmoid\big(\boldsymbol{W}_2(ReLU(\boldsymbol{W}_1\boldsymbol{u}+\boldsymbol{b}_1)+\boldsymbol{b}_2)\big) \quad (5)$$

In the expression, the $sigmoid$ function maps the mean information of each channel to the interval of (0,1), and $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m]$ represents the weight of each channel. The larger the value, the stronger importance of the corresponding feature channel.

In the end, the obtained weights are assigned to each feature channel, so the features extracted from the j th channel of spatial features through the SE module can be expressed as follows:

$$\boldsymbol{c}_j^{'} = \boldsymbol{v}_j \boldsymbol{c}_j \tag{6}$$

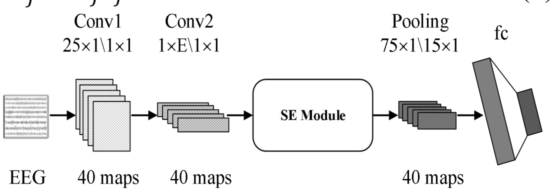


Figure 3. Neural Network Structure Based on SE Module

The spatial attention features $\boldsymbol{C}_i^{'} = [\boldsymbol{c}_1^{'}, \boldsymbol{c}_2^{'}, \cdots, \boldsymbol{c}_m^{'}]$ can be obtained by multiplying the spatial features $\boldsymbol{C}_i = [\boldsymbol{c}_1, \boldsymbol{c}_2, \cdots, \boldsymbol{c}_m]$ extracted by the spatial convolutional layer of the backbone network ConvNet_S with their corresponding weights $\boldsymbol{v} = [\boldsymbol{v}_1, \boldsymbol{v}_2, \cdots, \boldsymbol{v}_m]$. Finally, the obtained features are passed through the pooling layer and the fully connected layer.

3.3. EEG Recognition Based on CBAM Module

3.3.1. CBAM Module

The structure of the CBAM[12] module adopted in this paper is shown in Figure 4. It consists of two independent attention modules of channel shown in Figure 5 and space shown in Figure 6, that is, it takes into account both one-dimensional channel attention features and two-dimensional spatial attention features.

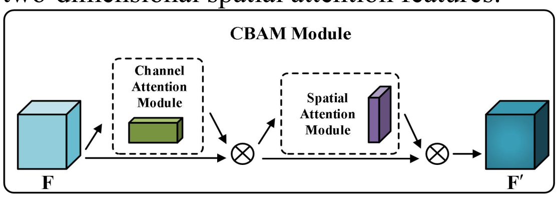


Figure 4. Structure of CBAM Module

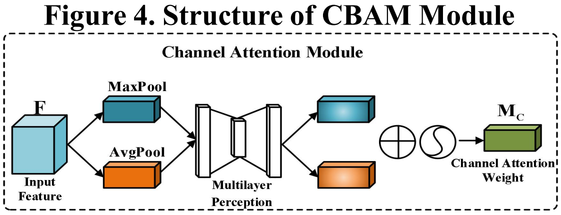


Figure 5. Structure of Channel Attention Module

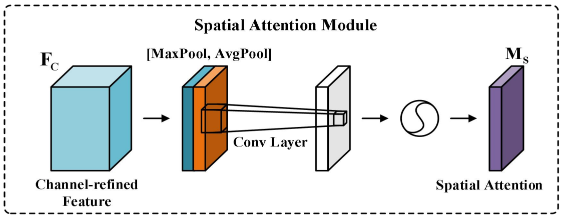


Figure 6. Structure of Spatial Attention Module

The channel attention module is closely related to the SE module. However, as you can see from Figure 5, the structure is slightly different from the SE module. The module adopts both the average pooling and the max pooling, which retains more information than the single average pooling and can obtain a better recognition effect. The operation process is as follows: First, the output feature map of the convolutional layer is applied as the intermediate input $\boldsymbol{F}$ through average pooling and max pooling respectively. Then, the output vector is obtained by sharing Multilayer Perception(MLP). Finally, the elements of each vector are added and passed through the $sigmoid$ function. The expression is as follows:

$$\boldsymbol{M_C}(\boldsymbol{F}) = sigmoid\big(MLP(AvgPool(\boldsymbol{F})) + MLP(MaxPool(\boldsymbol{F}))\big) \tag{7}$$

In the expression, $\boldsymbol{F} \in R^{C\times H\times W}$ is the module input data size, $\boldsymbol{M_C} \in R^{C\times 1\times 1}$ is the channel attention feature map, $sigmoid$ is the nonlinear activation function, $AvgPool$ and $MaxPool$ are two pooling methods commonly applied in neural networks, and the features selected by them focus on different levels and are complementary.

Similarly, the average pooling and max pooling operations of the spatial attention module are applied to the feature maps, respectively. The output feature matrices are then merged into a feature map. Finally, the two-dimensional spatial attention is generated through the convolution layer, and the feature map is generated through the $sigmoid$ function. The expression is as follows:

$$\boldsymbol{M_S}(\boldsymbol{F}) = sigmoid\big(f^{7\times 7}([AvgPool(\boldsymbol{F}); MaxPool(\boldsymbol{F})])\big) \tag{8}$$

In the expression, $\boldsymbol{M_S} \in R^{H\times W}$ is the learned spatial attention feature map and $f^{7\times 7}$ represents the convolution kernel of 7×7.

3.3.2. ConvNet_S_CA: A Novel Method for EEG Recognition Based on CBAM Module

This paper designs a novel EEG recognition method ConvNet_S_CA which combines the CBAM module into the ConvNet_S. This method includes convolution, average pooling, and full connection modules in the backbone network.

To explore the relationship between channel and space information in the EEG features extracted by convolution, ConvNet_S_CA adds a CBAM module after the spatial convolution layer Conv2 of the backbone network ConvNet_S, as shown in Figure 7.

Assuming that the features $\boldsymbol{G}$ are the features extracted by the spatial convolutional layer of ConvNet_S, the channel attention module is applied to learn the importance of EEG signals between different electrode channels. The expression is:

$$\boldsymbol{M_C}(\boldsymbol{G}) = \sigma\Big(MLP\big(AvgPool(\boldsymbol{G})\big) + MLP\big(MaxPool(\boldsymbol{G})\big)\Big) = \sigma\Big(\boldsymbol{W}_1\Big(\boldsymbol{W}_0\big(\boldsymbol{G}^C_{avg}\big)\Big) + \boldsymbol{W}_1\Big(\boldsymbol{W}_0\big(\boldsymbol{G}^C_{max}\big)\Big)\Big) \quad (9)$$

In the expression, $\boldsymbol{W}_0 \in R^{\frac{C}{r}\times C}$ and $\boldsymbol{W}_1 \in R^{C\times\frac{C}{r}}$ are the weights of MLP, respectively, and r is the compression rate. $\boldsymbol{G}^C_{avg}$ and $\boldsymbol{G}^C_{max}$ denote the generated spatial matrix and $\sigma(\cdot)$ denote the $sigmoid$ activation function.

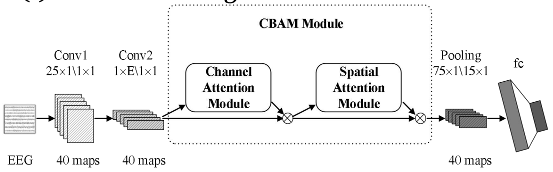


Figure 7. Neural Network Structure Based on CBAM Module

The spatial attention module is to learn the correlation between different frequency bands. The specific expression is:

$$\boldsymbol{M_S}(\boldsymbol{G}) = \sigma\big(f^{7\times7}([AvgPool(\boldsymbol{G}); MaxPool(\boldsymbol{G})])\big) = \sigma\Big(f^{7\times7}\big(\big[\boldsymbol{G}^S_{avg}; \boldsymbol{G}^S_{max}\big]\big)\Big) \quad (10)$$

In the expression, $f^{7\times7}$ represents the convolution kernel of 7×7 and $\sigma(\cdot)$ denotes the $sigmoid$ activation function.

4. Experiments

4.1. Datasets

(1) BCI IV IIa dataset[10]: This dataset contains nine subjects(A1-A9), and each subject is collected in two periods. The subjects performed four categories of tasks: left hand, right hand, two feet, and tongue with 144 samples per category of motor imagery task. In the experiment, 288 EEG data from the first stage were applied as the training set and 288 data from the second stage were applied as the test set.

(2) Low Limbs MI-BCI dataset[13]: This dataset contains EEG data from 10 subjects(S1-S10). Each subject performed the lower limb motor imagery task with or without visual guidance provided by the virtual system for a total of 200 trials. In the experiment, the first 140 EEG data were applied for training and the remaining data were applied for testing.

4.2 Evaluation Criterion

The evaluation indicators of the proposed method mainly include the following:

(1) *Accuracy*: Classification accuracy(ACC) is the most common classifier performance metric.

(2) *F1*: The comprehensive evaluation metric, which combines recall and precision, is shown in expression 11:

$$F1 = 2 * \frac{Recall*Precision}{Recall+Precision} \quad (11)$$

4.3 Experimental Results

In this paper, the 5th-order Butterworth filter is applied as the input data after preprocessing the above EEG raw data with 4-38Hz bandpass filtering. The batch size is set to 64 and 20 respectively during method training in two different datasets. The Adam optimization algorithm was applied to optimize the method, the learning rate was set to 10^{-3}, and the loss function was a cross-entropy loss function.

Table 1 shows the EEG signal classification results of the ConvNet_S, ConvNet_S_A, and ConvNet_S_CA methods on the BCI IV IIa dataset. It can be found from the table that the ACC and F1 indicators of novel methods are greatly improved in most cases. Compared with ConvNet_S, the proposed ConvNet_S_A greatly improves the recognition accuracy of 6 out of 9 subjects, and the average ACC and F1 are both increased by more than 1.7%. Similarly, the classification accuracy of ConvNet_S_CA is greatly improved by 5 subjects out of 9 compared with ConvNet_S, and the average ACC and F1 are both improved by more than 2.1%. The above results show that the novel methods proposed can effectively improve the recognition performance.

Table 1. Classification Performance of Different Methods on BCI IV IIa Dataset

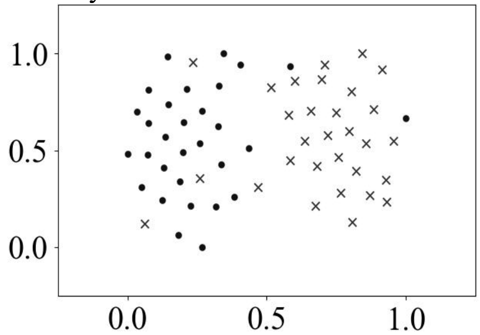
Subjects	Criterion	Contrast Algorithm		
		ConvNet_S	ConvNet_S_A	ConvNet_S_CA
A1	ACC	0.8281	0.8156	0.8219
	F1	0.8225	0.8155	0.8215
A2	ACC	0.5625	0.5938	0.5875
	F1	0.5627	0.5942	0.5861
A3	ACC	0.9219	0.9281	0.9344
	F1	0.9186	0.9224	0.9341
A4	ACC	0.6563	0.7094	0.7188
	F1	0.6560	0.7056	0.7185
A5	ACC	0.7031	0.7188	0.6875
	F1	0.7010	0.7181	0.6872
A6	ACC	0.5781	0.5938	0.5938
	F1	0.5786	0.5927	0.5944
A7	ACC	0.7656	0.8531	0.8781
	F1	0.7623	0.8538	0.8735
A8	ACC	0.8438	0.8375	0.8344
	F1	0.8424	0.8378	0.8321
A9	ACC	0.8281	0.7969	0.8250
	F1	0.8245	0.7935	0.8247
Avg	ACC	0.7431	0.7608	0.7646
	F1	0.7410	0.7595	0.7634

To further verify the effectiveness, this paper tested the EEG recognition effect on the self-collected dataset Low Limbs MI-BCI. Table 2 shows the classification results of the proposed methods are better than other methods in most subjects. Compared to ConvNet_S, ConvNet_S_A significantly improves the recognition accuracy of 8 out of 10 subjects. Similarly, ConvNet_S_CA improves substantially for all subjects. The results also show that the proposed methods can effectively improve the recognition effect of EEG.

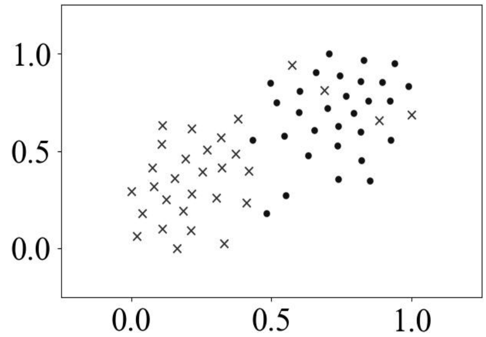
Table 2. Classification Performance of Different Methods on Low Limbs MI-BCI Dataset

Subjects	Criterion	Contrast Algorithm		
		ConvNet_S	ConvNet_S_A	ConvNet_S_CA
S1	ACC	0.7167	0.7000	0.8833
	F1	0.7136	0.6970	0.8823
S2	ACC	0.6500	0.6500	0.7500
	F1	0.6419	0.6491	0.7494
S3	ACC	0.7833	0.7833	0.8500
	F1	0.7813	0.7813	0.8496
S4	ACC	0.7833	0.7833	0.8667
	F1	0.7813	0.7813	0.8631
S5	ACC	0.8667	0.9500	0.9667
	F1	0.8631	0.9499	0.9666
S6	ACC	0.6500	0.7667	0.8500
	F1	0.6267	0.7661	0.8496
S7	ACC	0.9833	0.9500	0.9833
	F1	0.9832	0.9499	0.9832
S8	ACC	0.9000	0.9500	0.9667
	F1	0.8990	0.9499	0.9666
S9	ACC	0.7333	0.8000	0.9667
	F1	0.7277	0.7980	0.9666
S10	ACC	0.9333	0.9333	0.9333
	F1	0.9332	0.9332	0.9332
Avg	ACC	0.8000	0.8267	0.9017
	F1	0.7951	0.8256	0.9010

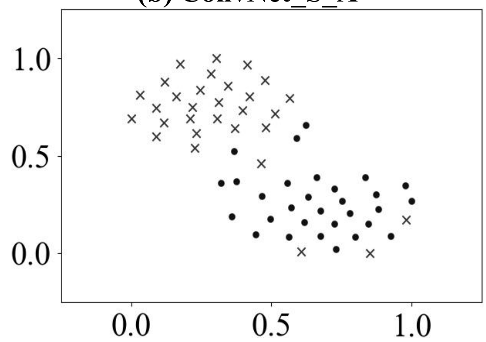
To evaluate the feature learning ability of the proposed ConvNet_S_A and ConvNet_S_CA methods, the t-SNE[14] embedded representations of the features learned by the ConvNet_S and the novel method are given on the Low Limbs MI-BCI dataset S10 subjects. "•" is the imagined movement with lower limbs, "×" is the imagined movement without lower limbs, and its visualization effect is shown in Figure 8. Compared with the backbone network ConvNet_S, the EEG features learned by the proposed ConvNet_S_A and ConvNet_S_CA methods have better inter-class separability, and the separability is more obvious.



(a) ConvNet_S



(b) ConvNet_S_A



(c) ConvNet_S_CA

Figure 8. Visualization of Features on Low Limbs MI-BCI Dataset of Subject S10

4.4. Discussion

To verify that the attention mechanism can effectively learn EEG signal features, parameter sensitivity analysis on relevant datasets is conducted. Firstly, the values of different parameters in the attention mechanism module network are experimented on

Table 3. Accuracies of Network Structures with Different Ratios r

Methods	Average Accuracy
ConvNet_S	0.7431
ConvNet_S_A(r=4)	0.7608
ConvNet_S_A(r=5)	0.7524
ConvNet_S_A(r=8)	0.7549
ConvNet_S_A(r=10)	0.7507
ConvNet_S_CA(r=4)	0.7646
ConvNet_S_CA(r=5)	0.7538
ConvNet_S_CA(r=8)	0.7392
ConvNet_S_CA(r=10)	0.7517

the dataset BCI IV IIa. When the value of r is set to 4 shown in Table 3, the EEG recognition effect is the best, and the accuracy of the two new methods can be more than 76%. This indicates that the neural network can achieve the best feature expression ability under this parameter. And the accuracy drops slightly when the parameter is 5. As the value of the parameter increases, the recognition accuracy of the neural network for EEG signals also decreases, which is not conducive to the recognition of EEG features.

5. Conclusion

Features in deep neural networks contain different types of information, which contribute to different degrees of EEG signal recognition. In the paper, the attention module is added to adaptively select important information and suppress relatively unimportant information. The experimental results verify the effectiveness of novel the method in EEG recognition.

References

[1] Lebedev M A, Nicolelis M. Brain-machine interfaces: past, present and future. Trends in Neurosciences, 2006, 29 (9): 536-546.
[2] Hatem A, Peter B, Hotz B S, et al. What disconnection tells about motor imagery: evidence from paraplegic patients. Cerebral Cortex, 2004, 15 (2): 131.
[3] Lecun Y, Bengio Y, Hinton G. Deep learning. Nature, 2015, 521 (7553): 436.
[4] Huang J, Xu X, Zhang T. Emotion classification using deep neural networks and emotional patches//2017 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). IEEE, 2017.DOI:10.1109/BIBM.2017.8217786.
[5] Schirrmeiste R T, Springenberg J T, Fiedere L D J, et al. Deep learning with convolutional neural networks for EEG decoding and visualization. Human Brain Mapping, 2017. DOI:10.1002/hbm.23730.
[6] Dornhege G. Combined optimization of spatial and temporal filters for improving brain-computer interfacing. IEEE transactions on bio-medical engineering, 2006, 53 (11): 2274.
[7] Hinterberger T, A Kübler, Kaiser J, et al. A brain–computer interface (BCI) for the locked-in: comparison of different EEG classifications for the thought translation device. Clinical Neurophysiology, 2003, 114 (3): 416-425.
[8] An X, Kuang D, Guo X, et al. A Deep Learning Method for Classification of EEG Data Based on Motor Imagery.

Springer International Publishing, 2014. DOI:10.1007/978-3-319-09330-7_25.

[9] Tabar Y R, Halici U. A novel deep learning approach for classification of EEG motor imagery signals. Journal of Neural Engineering, 2017, 14 (1): 016003.

[10]Naeem M, Brunner C, Leeb R, et al. Seperability of four-class motor imagery data using independent components analysis. Journal of Neural Engineering, 2006, 3 (3): 208-216.

[11]Hu J, Shen L, Albanie S, et al. Squeeze-and-Excitation Networks. IEEE Transactions on pattern analysis and machine intelligence, 2020, 42 (8) 2011-2023.DOI:10.1109/TPAMI. 2019. 2913372.

[12]Woo S, Park J, Lee J Y, et al. CBAM: Convolutional Block Attention Module. Springer, Cham, 2018. DOI:10.1007/978-3-030-01234-2_1.

[13]Hang W, Feng W, Liang S, et al. Deep stacked support matrix machine based representation learning for motor imagery EEG classification. Computer Methods and Programs in Biomedicine, 2020, 193: 105466.

[14]Maaten L J P V D, Hinton G E. Visualizing High-Dimensional Data using t-SNE. Journal of Machine Learning Research, 2008, 9: 2579-2605.

