EAI Endorsed Transactions
on Scalable Information Systems

Research Article EAI.EU

Channel space weighted fusion-oriented feature pyramid network for motor imagery EEG signal recognition

Wenhao Yang[1,*]

[1]School of Physical Education and Sport, Zhengzhou University of Science and Technology, Zhengzhou, 450064, China

Abstract

In order to solve the problems of weak generalization ability and low classification accuracy in motor imagery EEG signal classification, this paper proposes a channel space weighted fusion-oriented feature pyramid network for motor imagery EEG signal recognition. First, the short-time Fourier transform is used to obtain the EEG time-frequency map. Then, it builds a new feature pyramid network(FPN). The attention mechanism module is integrated into the FPN module, and the channel spatial weighted fusion-oriented feature pyramid network is proposed. This new structure can not only learn the weight of important channel features in the feature map, but also learn the representation of important feature areas in the network layers. Meanwhile, Skip-FPN module is added into the network structure, which fuses more details of EEG signals through short connections. The Dropout layer is added to prevent network training from over-fitting. In the classification model, we improve the AdaBoost algorithm to automatically update the base learner according to the classification error rate. Finally, the proposed model is used to classify the test data and the Kappa value is used as the evaluation index. Compared with the state-of-the-art motor image EEG signal recognition methods, the proposed method achieves better performance on the BCI Competition IV 2b data set. It has good generalization ability and can improve the classification effect.

Keywords: motor imagery EEG signal recognition, short-time Fourier transform, attention mechanism, FPN, channel spatial weighted fusion.

Received on 16 August 2021, accepted on 28 August 2021, published on 09 September 2021

doi: 10.4108/eai.3-9-2021.170906

*Corresponding author. Email: zxcvfdsa5024@foxmail.com

1. Introduction

Electroencephalogram (EEG) is an important biological signal formed by the synchronous summing up of postsynaptic potentials of a large number of neurons in the cerebral cortex [1]. EEG signal is a non-stationary and non-linear time series signal. Because of its high complexity, low signal-to-noise ratio, strong randomness and high dimension, it is difficult to control in brain-computer interface (BCI).

The brain-computer interface uses the EEG signals generated by the brain to directly establish information exchange and control channels with peripheral control devices such as computers without passing through peripheral nerves and muscle tissues [2]. Motor imagery (MI) EEG is one of the most commonly used EEG signals in BCT, and is currently a research hotspot in the field of medical rehabilitation [3].

In the study of BCI, the traditional EEG processing methods include three steps: preprocessing, feature extraction and classification. The common spatial pattern (CSP)[4] method is a popular feature extraction method, which maximizes the difference in energy between the spatial components of the two types of motion imagery to achieve classification by using spatial filtering. Other common spatial patterns are filter bank common spatial pattern (FBCSP) [5], power spectrum analysis [6], wavelet packet decomposition [7] and independent component analysis (ICA)[8]. Most of methods extract single features with manually feature selection and lack multi-feature fusion. Classification methods mainly adopt machine learning classification methods such as linear discriminant method, Bayesian classifier and support vector machine [9,10].

With the development of deep learning technology, convolutional neural network (CNN) has become the first deep network structure that has been truly successfully trained [11]. The CNN can learn and extract features by itself in image classification, and can fuse various features of EEG signals through images. In the study of deep learning-based BCI, Fu et al. [12] combined deep learning and data enhancement methods, using empirical mode decomposition to construct the EEG frame of artificial motor imagery. It used wavelet transform to obtain the time-frequency map of EEG, and constructed two kinds of network structures: convolutional neural network and wavelet neural network. Tabar et al. [13] proposed a new method by combining CNN and stacked auto-encoder (SAE). The short-time Fourier transform (STFT) was used to obtain the EEG time-frequency map as the input of the network. The classification accuracy of motor imagery EEG was 77.6%, which was better than that of SVM. Phang et al. [14] used working memory EEG data and azimuthal equidistant projection (AEP) method to obtain EEG power projection maps, and the recurrent neural networks (RNN) and CNN were combined to classify EEG images. Yang et al. [15] combined CSP and CNN to classify motor imagery EEG, and proposed the enhanced CSP to obtain paired projection matrix as the input image of the network, which achieved higher classification accuracy than FBCSP method. Li et al. [16] proposed a multi-wavelet basis function time-frequency analysis and conditional Granger causality method to conduct time-frequency analysis of motor imagery EEG signals to obtain the input images of CNN. Compared with the winners of BCI Competition IV 2a, the classification accuracy has improved by 12.15%. Gao et al. [17] developed a framework combining recurrence plots and convolutional neural network to achieve fatigue driving recognition. Meng et al. [18] proposed a motor imagery classification algorithm based on recurrence plot convolution neural network. Shi et al. [19] proposed a sleep quality detection and management method based on electroencephalogram (EEG). The detection of sleep quality was mainly achieved by staging sleep EEG signals. First, wavelet packet decomposition (WPD) preprocessed the collected original EEG to extract the four rhythm waves of EEG. Second, the relative energy characteristics and nonlinear characteristics of each rhythm wave were extracted. The multisample entropy (MSE) values of different scales were calculated as the main features, and the rest were auxiliary features. Finally, the long short-term memory (LSTM) model was applied to classify the extracted sleep features, and the final result was obtained. Fernandez-Blanco et al. [20] presented a study focused on the scoring of sleeping EEG signals to measure if the increase of the pressure on the features due to a reduction of the number though different techniques resulted in a benefit. But the above methods still cannot solve the problem of EEG signal classification effectively.

Inspired by the application of deep learning network in MI-EEG, this paper proposes a channel space weighted fusion-oriented feature pyramid network. Our main contributions are as follows. First, the short-time Fourier transform is used to obtain the EEG time-frequency map. Then, it builds a new feature pyramid network(FPN). The attention mechanism module is integrated into the FPN module, and the channel spatial weighted fusion-oriented feature pyramid network is proposed. This new structure can not only learn the weight of important channel features in the feature map, but also learn the representation of important feature areas in the network layers. In the classification model, we improve the AdaBoost algorithm to automatically update the base learner according to the classification error rate.

This paper is organized as follows. Section 2 introduces the related works. In section 3, we detailed give the proposed motor imagery EEG signal recognition method. Section 4 gives the experimental results and analysis. There is a conclusion in section 5.

2. Related works

2.1. Mask R-CNN model

Mask R-CNN can extend Faster R-CNN by adding a branch for predicting an object mask in parallel with the existing branch for bounding box recognition. Mask R-CNN is simple to train and adds only a small overhead to Faster R-CNN. Moreover, Mask R-CNN is easy to generalize to other tasks. Therefore, in this paper, Mask R-CNN is selected as the benchmark recognition model [21], and the basic framework is shown in figure 1. Previous deep learning-based object recognition processes include candidate region selection, feature extraction, classification and boundary regression. The Mask R-CNN model is mainly composed of CNN, Region Proposal Networks (RPN) and three branch networks (positioning, classification and semantic segmentation).

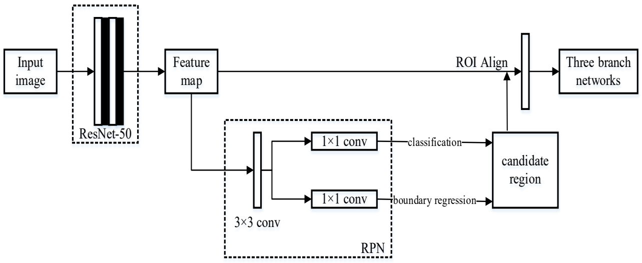

Figure 1. Mask R-CNN

2.2. ResNet-50

In this paper, ResNet-50 [22] is used as the basic network to extract EEG features. ResNet-50 is made up of 50 basic residual blocks. The structure of the basic residual block is shown in figure 2. Where, the input is x and the expected output is $H(x)$. The learning goal of residual block is the difference between output and input, that is, $F(x)=H(x)-x$, which directly transfers input information to output. It protects the integrity of information, and simplifies the learning goal and difficulty. The residual function $F(x)$ uses three convolution layers, where the convolution kernel sizes are 1×1, 3×3 and 1×1 respectively. The middle 3×3 convolution reduces the computation under the first reduced-dimension 1×1 convolution layer. In the second 1×1 convolution layer, the reduction is done, which not only preserves the precision but also reduces the calculation amount.

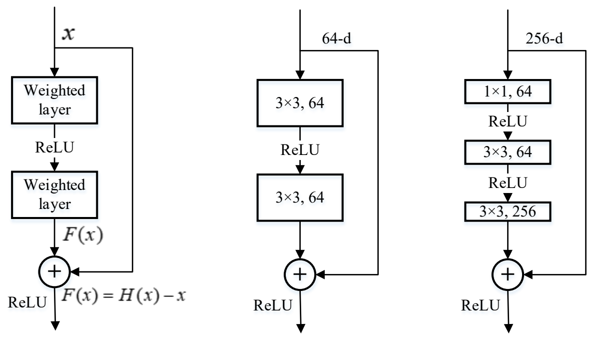

Figure 2. Base residual block

2.3. Feature pyramid network

Feature pyramid network (FPN) [23] can fuse shallow and deep features to obtain more robust semantic information. Simple targets can be distinguished by shallow features. Deep features can be used to distinguish complex targets. Resnet-50 can extract features representing EEG signals. Multi-scale features extracted by ResNet-50 can be fused by FPN module to obtain more representative and reliable features, so as to improve network performance.

FPN is divided into the bottom-up path, top-down path, and a horizontal connection, as shown in figure 3. In the bottom-up path, the EEG signals successively pass through five stages: conv1, conv2, conv3, conv4, conv5. ResNet-50 uses the feature activation of the last residual structure as the output, and represents these residual module outputs as {C1,C2,C3,C4,C5}. {C1,C2,C3,C4,C5} has a step size of 14, 8, 16, and 32 relative to the input image. After C2,C3,C4,C5, 1×1×256 convolution is connected to generate feature layers M2,M3,M4 and M5 with channel number 256. In the top-down path, M5 is up-sampled twice, and then added with M4. After that, 3×3 convolution kernel (in order to eliminate the alias effect of up-sampling) is used to process the fused feature map to generate the final required feature map P4. So P3= M3+2×M4, P2=M2+2×M3. P5 is generated by M5 through the convolution layer of 3×3 alone, and P6 is the feature layer formed after P5 passing down-sampling.

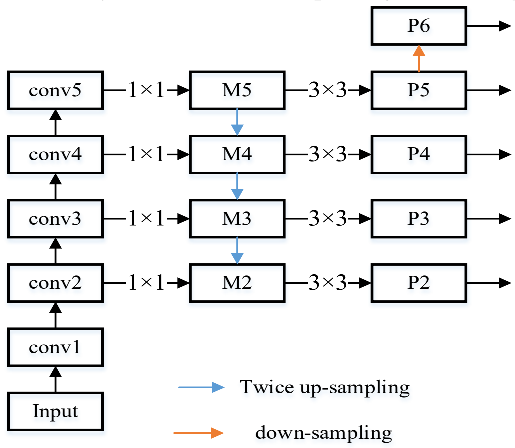

Figure 3. FPN structure

3. Proposed motor imagery EEG signal recognition method

3.1. Data set

The experimental data comes from the BCI Competition IV 2b data set (http://www.bbci.de/competition/iv/) collected by the Brain-computer Interface Laboratory of Graz Technical University [24]. The data set records nine subjects' EEG data sets: imagery left hand movements and imagery right hand movements. Each subject will collect five groups of EEG data through three electrode channels, C3, Cz and C4. The first-three groups are training data, including 400 motor imagery experiments. The latter two groups are testing data, including 320 motor imagery experiments. In the five groups of EEG data, the first two groups have 120 experiments respectively, and were neuro-feedback data without recognition results. The last three groups have 160 experiments, which are neuro-feedback data with recognition results. Experimental paradigms with and without feedback are shown in figure 4 and figure 5.

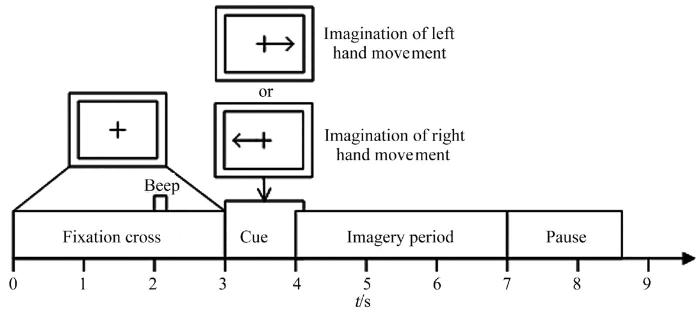

Figure 4. MI without recognition feedback

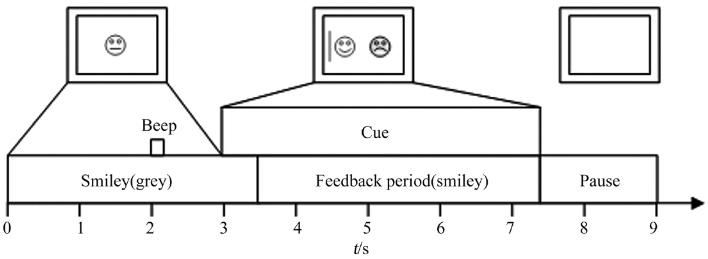

Figure 5. MI with recognition feedback

The neuro-feedback experiment with recognition results is a gray human face displayed on the screen two seconds before the start of the experiment as a cue to prepare for the experiment. At the third second, the task prompt appears, and the subjects perform the corresponding motor imagery task according to the task prompt [25]. Based on the recognition results, the system will give a green smiley face that moves in the right direction or a red smiley face that moves in the wrong direction on the screen.

The EEG data recorded in the channel are filtered with a band pass of 0.5-100 Hz, the sampling frequency is 250 Hz, and the power frequency interference is eliminated by a 50Hz notch filter.

3.2. Time-frequency image input

Compared with traditional EEG feature extraction and classification methods, this study converts the temporal signals of EEG into images and automatically extracts features and classifies them through the proposed pyramid network. When subjects perform unilateral limb motor imagery, the energy of μ rhythm (8~13 Hz) and β rhythm (13~30 Hz) in the contralateral motor sensory region of the brain decrease in a specific frequency segment. The energy of μ rhythm and β rhythm of the related motor sensory areas on the same side increase. This phenomenons are known as event related desynchronization (ERD) and event related synchronization (ERS) [26]. According to this phenomenon, the three channels (C3, Cz, C4) of EEG data that most relate to the motor sensory region of the cerebral cortex are selected in this paper.

EEG is a kind of bioelectrical signal that contains rich time frequency information. The multi-channel acquisition method provides the spatial features. In this study, the motor imagery data with 2s of subjects is selected for STFT analysis [27]. The frequency bands of μ rhythm and β rhythm are 6~13 Hz and 17~30 Hz respectively. STFT performs the Fourier transform of the signal in the window through the window function translation on the time axis, and obtains the time-frequency image with the size of 257×32. The image sizes of μ and β rhythm are 16×32 and 23×32 respectively. In order to make the influence of μ and β rhythm bands on subsequent feature learning and recognition consistent, a Bi-cubic function is constructed to reduce the image of β rhythm to 15×32 [28]. According to the acquisition method of EEG, its spatial features are also important features for EEG classification. In order to make use of their spatial features, the three channels (C3, Cz, C4) are processed with the same time-frequency analysis, and the feature information is fused to obtain a 93×32 gray image, as shown in figure 6 and figure 7.

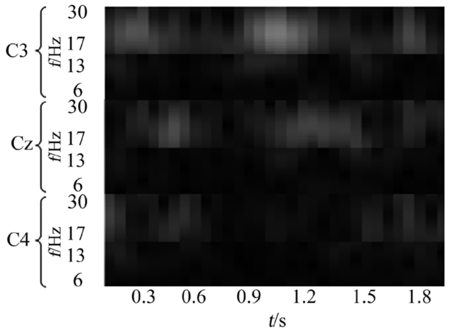

Figure 6. Sample input image of left hand MI

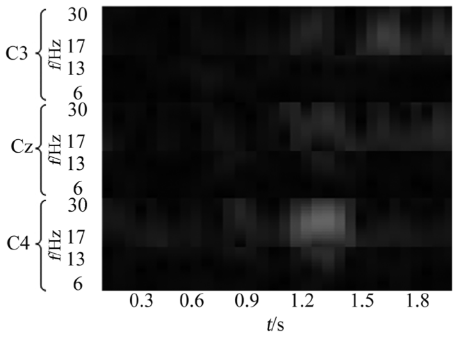

Figure 7. Sample input image of right hand MI

The obvious ERD/ERS phenomenon can be seen from figures 6 and 7. In figure 6, when subjects perform left hand motor imagery, the energy of C3 channel in the 17~30Hz band is higher than that of C4 channel, especially, in the period of 0.9s to 1.2s, the energy value is significantly higher than that of the corresponding band of C4 channel. The brightness of the time-frequency chart of the 6~13 Hz band in the C4 channel is darker than that of the corresponding band of the C3 channel. Especially, in the period of 0.6~1.2s, the energy value is significantly lower than that of the corresponding band in the C3 channel. When the subjects perform imagery right hand, they conduct the opposite. The 17~30 Hz band of C4 channel has higher energy value than the corresponding band of C3. The 6~13 Hz of C3 channel is darker and lower than that of C4 channel.

3.3. Motor imagery feature extraction based on proposed network

Due to lack shallow information, the traditional Mask R-CNN has the defects of unrefined feature extraction and incorrect classification. Therefore, this paper proposes a C-mask (cell-mask) network model. C-mask network conduct multi-scale fusion process combining with feature fusion module. C-mask network combines with channel space module to re-calibrate features and improve the accuracy of EEG feature extraction.

Skip connection

Because of the lack of shallow feature information, the classification of EEG features is wrong. In this paper, Skip-FPN [29] module is proposed for EEG signal feature fusion, which can fuse more reliable and usable features, as shown in figure 8.

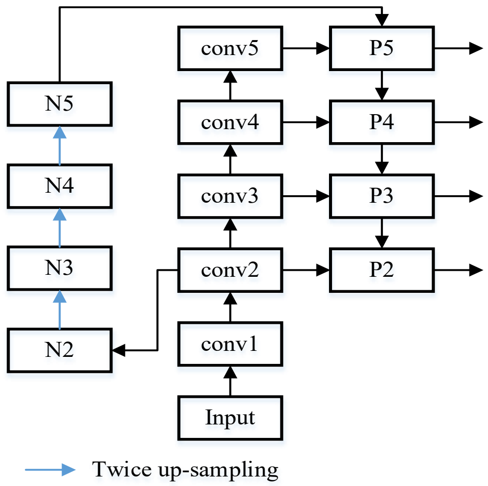

Figure 8. Skip-FPN module

In the bottom-up process of FPN algorithm, the transmission of features from the bottom layer to the top layer needs 50 network layers in conv1, conv2, conv3, conv4 and conv5, resulting in serious information loss of the bottom layer. Although P5 in FPN has indirectly acquired the underlying features, the flow line is too long [30]. Therefore, the Skip-FPN module is proposed in this paper, and the feature of the last residual structure in conv2 is named as N2. Then it performs up-sampling with step size 2 to generate N3. Similarly, the N4 and N5 are generated. After N5, it connects a 1×1×256 convolution layer, generating the same channel number as P5 and adding them up. In this way, the shallow features are transferred to the top layer through the three network layers, which can accelerate the acquisition of shallow feature information and improve the accuracy of EEG feature extraction.

Channel space weighted module

In this paper, the channel spatial weighted feature pyramid network (CSFPN) module is proposed by combining the channel spatial weighted module and feature pyramid network (FPN) module, as shown in figure 9.

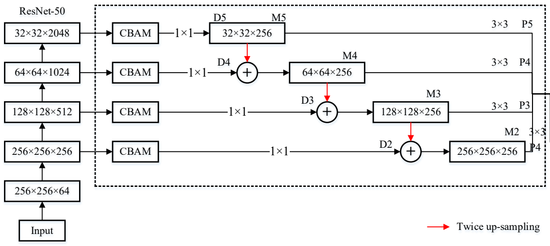

Figure 9. Structure of CSFPN

The first part in figure 9 is the ResNet-50 network, which consists of five phases. The feature activation of the last residual structure in each stage is represented as {C1,C2,C3,C4,C5}. The output sizes of C2,C3,C4,C5 are 256×256×256, 128×128×512, 64×64×1024 and 32×32×2048 respectively. The second part is the proposed CSFPN module in this paper. Because the extracted features in the shallow layer of the network contain many details, while the extracted features in the deep layer are more abstract, the CSFPN module is used to carry out channel space weighted for the generated features, and re-calibrate the features, so as to enhance the effective features and suppress the useless features.

As shown in figure 5, the feature graph F with size of 256×256×25 output by C2 is input into the convolutional block attention (CBAM) module. Firstly, the input feature graph F is compressed in the spatial dimension, and the global average pooling and maximum pooling of one space are performed to generate two 1×1×256 channel descriptions F_{avg}^{s} and $F_{\max}^{s}$ respectively. The spatial information of feature mapping is aggregated and then sent to a two-layer neural network with 1×1 convolution kernel to extract the information. The neuron number in layer 1 is 256/16, and the neuron number in layer 2 is 256. After that, the channel features are obtained through a Sigmoid function, i.e.

$$\begin{aligned} M_c(F) &= \delta(MLP(avg(F)) + MLP(avg(F))) \\ &= \delta(W_1(W_0(F_{avg}^{s})) + W_1(W_0(F_{\max}^{s}))) \end{aligned} \quad (1)$$

Where $W_0 \in R^{C/r}$, $W_1 \in R^{C}$. MLP shares the connection layer. F_{avg}^{s} and $F_{\max}^{s}$ denote the average pooling and max-pooling respectively. δ is the Sigmoid operation. r denotes the zoom ratio.

Secondly, two 256×256×1 channel descriptions F_{avg}^{s} and $F_{\max}^{s}$ are obtained by global maximum pooling and average pooling based on channel features. After that, the addition operation is performed based on the channel. After a convolution operation with kernel size of 7×7, the dimension is reduced to a channel. Finally, spatial attention features are generated through Sigmoid function, i.e.

$$\begin{aligned} M_s(F) &= \delta(f^{7\times 7}([avg(F); \max(F)])) \\ &= \delta(f^{7\times 7}([F_{avg}^{s}; F_{avg}^{s}])) \end{aligned} \quad (2)$$

Where δ is the Sigmoid operation. $f^{7\times 7}$ denotes the convolution kernel size.

In the same way, the feature maps of C3, C4 and C5 modules are input into the CBAM module to generate attention feature maps. The CBAM module is then connected with a 1×1×256 convolution layer to make the output feature map have the same channel number for multi-scale feature fusion. In the process of multi-scale feature fusion, after the feature map with 1×1 convolution for D4, the M4 is obtained. Similarly, the M3 and M2 are obtained. M layer feature maps are then convoluted by 3×3 (to reduce the aliasing effect caused by the nearest neighbor interpolation, and the surrounding ranges are the same) to obtain the final features of P2, P3, P4, and P5 layers.

CBAM module uses the global information of feature map after convolution layer to dynamically model the dependence of channel and space, so as to improve the feature learning ability of network. This module enables the network to learn important features and compress unnecessary features, so that the network can selectively optimize the parameters according to the importance of features. The structure of CBAM is shown in figure 10. The attention map of the feature map is calculated through the channel dimension, and then the attention map is multiplied with the input feature map to form F' for adaptive feature learning. Then, the attention map is calculated through the spatial dimension, and the attention map is multiplied with the input feature map to form a new feature map.

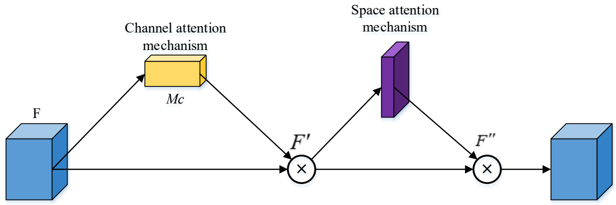

Figure 10. CBAM model

Data enhancement process

In deep learning network training, it appropriately increases the amount of data, which can improve the generalization ability of the model, reduce the over-fitting in the training process, and improve the robustness of the model. In this study, the motor imagery data with a total length of 4s is sliced. For motor imagery data ($X_\lambda \in R^{I,T}$) with 4s after band pass filtering of 0.5~30Hz, a sliding window with a duration of 2s is adopted to intercept time series segments along the time axis. The interception process is shown in formula (3):

$$x_\lambda = \{X_\lambda^{1\cdots E, t\cdots t+(f_s\times 2s)} \mid t \in 1\cdots T(f_s \times 2s)\} \quad (3)$$

Where X_λ represents the MI segment with a length of 4s, and E represents the electrode channel. T is the total number of sample points in the time series. f_s is the sampling frequency. $\lambda = 1,2,\cdots,M$ is the experimental ordinal number. M represents the experiment number of single subject. The step length of the sliding window is 0.5s. In the motor imagery data with a total length of 4s, a total of 5 frames are captured, each frame is 2s in length, and the interval between frames is 0.5s. The specific process is shown in formula (4):

$$\begin{aligned} x_\lambda = \{X_\lambda^{1\cdots E, t\cdots t+(f_s\times 2s)} \mid t \in 1, 1+(f_s \times 0.5s), \\ \cdots, 1+(f_s \times 1s), 1+(f_s \times 1.5s), 1+(f_s \times 2s)\} \end{aligned} \quad (4)$$

The category label of each frame is equal to the real category label in this experiment. After data slice processing, the data volume is expanded to five times of the original data.

Improved AdaBoost algorithm

In the classification process of this network, we adopt an improved AdaBoost algorithm for EEG. AdaBoost (Adaptive Boosting) algorithm is a typical Boosting algorithm. Its advantages lie in updating the weight of training samples, changing the sample distribution, connecting multiple base learners and assigning weights separately. Finally, it can obtain the strong learner through the additive model. The motor imagination data clipped by sliding windows overlaps each other between windows in time, that is, there is a certain correlation between the training data samples processed by slices, which results in information redundancy. In different time segments, the motor imagination features of training samples may appear in advance or lag, so the importance of different time segments for classification is also different. In order to further mine and utilize the information in the augmented samples, this paper combines the improved AdaBoost algorithm with CBAM model. This method will assign weights to the augmented training samples and update the weights iteratively, so as to reduce the influence of sample information redundancy. In this paper, we propose an AdaBoost method which can update base learner automatically.

Firstly, K CBAM learners are taken as the base learners in AdaBoost algorithm, the respective error rates of K base learners are calculated, and the base learner with the minimum error rate is selected. When the error rate of the base learner is greater than the random guessing probability of 0.5, the current base learner will be abandoned, and a CBAM base learner will be automatically pre-trained from the new random training sample and added to the base learner set. Then it is proceed with the iteration and calculates error rate. For the base learner that satisfies the condition, its weights are calculated and the sample weights are updated. After several iterations, the weights of K base learners are obtained. Finally, strong learners are obtained by the weights of K base learners and additive model. The steps of the improved AdaBoost algorithm are as follows.

1) Divide the original data into training data and test data. 90% of the data is used for training and 10% for testing, and 10 fold cross validation method is adopted.
2) Take 90% training data as input data. The input data sample set is $(x_1, y_1), (x_2, y_2), \cdots, (x_N, y_N)$, where $x_i \in X$, $y_i \in Y = \{-1, +1\}$, it is input into the CNN network structure for pre-training. The remaining 10% data is used as verification data, and K CNN-based learners of the subjects are obtained by repeating K times.
3) Initialize the weight distribution of the training data. At the beginning, each sample is assigned the same weight, and the initial weight distribution $D_1(i)$ is shown in equation (5):

$$D_1(i) = (w_1, w_2, \cdots, w_N) = (\frac{1}{N}, \cdots, \frac{1}{N}) \quad (5)$$

Where w is the sample weight and N is the number of samples.

4) According to K CNN-based learners given in Step 2, T round iterations are carried out in AdaBoost algorithm. Calculate the classification error of each base learner on sample distribution D_1 .

$$\begin{aligned} e_t &= P(H_t(x_i)) \neq y_i) \\ &= \sum_{i=1}^{N} w_i I(H_t(x_i) \neq y_i) \end{aligned} \quad (6)$$

Where $P(\cdot)$ is the probability density function of the sample, y_i is the true label of the i-th sample. w_i is the weight of the i-th sample. $I(\cdot)$ is the indicator function.

It selects the base learner h with the lowest current classification error rate as the t-th base learner H_t. If $e_t > 0.5$, then it discards the current base learner H_t, and obtains a new base learner h' from the training sample again, and then iterates and repeats the error calculation. When the above conditions are satisfied, the weight of the base learner in the final classifier is calculated:

$$\alpha_t = 0.5\ln(\frac{1-e_t}{e_t}) \tag{7}$$

Here, α_t is the weight of t-th base learner.

The weight distribution of the training sample is updated according to the classification error rate and the weight of the base learner, as shown in equations (8), (9):

$$D_{t+1} = \frac{D_t(i)\exp(-\alpha_t y_i H_t(x_i))}{Z_t} \tag{8}$$

$$Z_t = \sum_{i=1}^{N} w_{t,i} \exp(-\alpha_t y_i H_t(x_i)) \tag{9}$$

Where Z_t is the normalization factor.

5) After T round iterations, the final classifier is obtained by combining the weight α_t of each base learner, as shown in equation (10):

$$H_{final} = sign(\sum_{t=1}^{T} \alpha_t H_t(x)) \tag{10}$$

Where H_{final} is a strong learner model. sign(.) is a sign function.

The sliding window is used to intercept the data into 5 time periods on the time axis. The data sets of five time periods in the same experiment are input into the enhanced convolutional neural network model, that is, the strong learner model H_{final}. The voting method is used to obtain the label category of this experiment, that is, the category with the majority prediction in the five time segments is used as the category of this experiment.

3.4. Verification Method

For BCI Competition IV 2b data set, classifier training and testing are conducted for each subject. The classification accuracy, standard deviation and average Kappa value are used to evaluate the performance of the proposed method. In the new network model, the training model adopts the batch training. The training batch size is 50, and the total number of training rounds is 400. The structure of the network adopts the stochastic gradient descent method. The Momentum value is 0.9 and the learning rate is 0.01. 10% of the input training data is used as the verification data. Dropout layer (dropout rate=0.5) is added to prevent over-fitting during network training. In the improved AdaBoost algorithm model, the number of base learners K and iteration rounds T of each subject are set as 8 and 16 respectively.

In order to verify the feasibility of the proposed enhanced convolutional neural network model, the first three data sets of each subject are selected for 400 experiments, and a total of 2000 images are obtained after data enhancement. 90% is used for training and 10% for testing. 10 fold cross validation is added to calculate the average classification accuracy and standard bias. The commonly used classification methods, including support vector machine (SVM) method, Twin SVM method [31] and CNN method, are selected as the comparison methods.

In order to verify the validity of the model in this paper, the first three groups of data from subjects are treated as training data, and the last two groups of data are treated as testing data. In this paper, kappa coefficient is used to measure the classification accuracy. Kappa coefficient is a classification performance index that removes the influence of random classification accuracy, and its calculation method is shown in formula (11) :

$$\kappa = \frac{acc - 1/C}{1 - 1/C} \tag{11}$$

Where, acc is the classification accuracy of the sample. C is the number of classification categories.

4. Experimental results and analysis

The feasibility verification and comparison results of the classifier model in this paper are shown in table 1 and figure 11. As can be seen from table 1, the accuracy (six subjects) of this method is higher than that of the other 3 methods. For example, the classification accuracy of subject 1 is 77.3%, which is 2.6% higher than the Twin SVM method. Moreover, the standard deviation of the proposed method in this paper is 2.7%, which is lower than the other three methods, indicating that the classification accuracy of this method on this subject is more stable and will not fluctuate greatly. The final average classification accuracy of nine subjects is 76.4%, which is higher than other methods.

Table 1. Comparison of classification accuracy with different methods/%

Subjects	SVM	Twin SVM	CNN	Proposed
B01	71.8	74.7	74.5	**77.3**
B02	64.5	70.8	64.3	**65.5**
B03	69.3	66.1	71.8	**61.0**
B04	93.0	94.9	94.5	**94.6**
B05	77.5	86.1	79.5	**87.2**
B06	72.5	70.3	75.0	**76.0**
B07	68.1	73.4	70.5	**75.0**
B08	69.8	73.9	71.8	**75.0**
B09	65.0	75.2	71.1	**76.3**
Average	72.4	76.1	74.8	**76.4**

As can be seen from figure 11, the standard deviation values of the proposed method and the CNN method are close. And the standard deviation of most subjects is lower than that of SVM and Twin SVM, indicating that the accuracy of the proposed method has little fluctuation. No large deviation will occur. The new model has good robustness, and it is feasible to classify motor imagery EEG data.

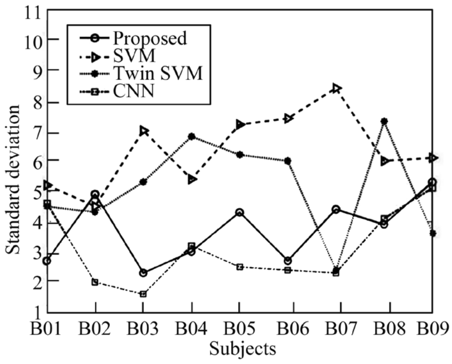

Figure 11. Standard deviation comparison with four methods

For all the data of BCI IV competition, the effectiveness verification and comparison results of the proposed classifier model are displayed in table 2. As can be seen from table 2, among the 9 subjects, the kappa coefficients of 5 subjects with the proposed method in this paper are higher than that of other methods, namely subject 2, Subject 4, subject 6, subject 7 and subject 8. For a single subject, the Kappa value of subject 4 in this study is 0.96, and the classification accuracy reaches 98%. The Kappa value of subject 7 is 0.08, which is higher than that of other optimal classification models, showing a great improvement in classification accuracy. The average Kappa value of the proposed method in 9 subjects is 0.63, which is superior to other methods, indicating that the enhanced model is effective in the classification of motor imagery EEG data.

Table 2. kappa value comparison with different methods/%

Subjects	SVM	Twin SVM	CNN	Proposed
B01	0.19	0.40	0.55	**0.48**
B02	0.12	0.21	0.21	**0.22**
B03	0.12	0.22	0.24	**0.21**
B04	0.77	0.95	0.89	**0.96**
B05	0.57	0.86	0.69	**0.81**
B06	0.49	0.61	0.53	**0.68**
B07	0.38	0.56	0.41	**0.69**
B08	0.85	0.85	0.41	**0.86**
B09	0.61	0.74	0.58	**0.76**
Average	0.46	0.60	0.51	**0.63**

According to the results in table 1 and table 2, the accuracy of the new convolutional neural network model in this paper is higher than other classification models in more than half of the 9 subjects. In Table 1, the first three groups of five data groups are adopted, and it can be seen that the accuracy of CNN method is higher than that of traditional SVM method in 8 out of 9 subjects, indicating that the convolutional neural network model has advantages in the classification of EEG time-frequency images. For the proposed method in this paper, the classification accuracy of 8 subjects is higher than that of traditional SVM method and CNN method. And the accuracy of 7 subjects is higher than that of Twin SVM method, indicating that the improved Adaboost algorithm in this paper can not only enhance the CNN model, but also effectively improve the classification accuracy. In table 2, the test data given by BCI competition are used for classification. The accuracy of 7 subjects with new method is higher than that of other methods. The average classification accuracy is the highest. It can be seen that the new convolutional neural network model can better improve the classification effect, and the model has good generalization ability. The proposed method converts motor imagery EEG signals into images and performs feature processing and classification through deep learning.

According to the results in table 1 and table 2, there are still large differences in classification accuracy between different individual subjects. For example, subject 2 and subject 3 are less accurate than the other subjects. From the time frequency diagram, the time frequency diagram of the left and right hand motor imagination of subject 3 is shown in figure 12. Compared with subject 4, who has the highest classification accuracy among the 9 subjects, the time frequency diagram of the left and right hand motor imagination is shown in figure 13. Compared with subject 4, the feature difference of ERD/ERS in time-frequency diagram of subject 3 is lower and the feature is more messy. In particular, the ERD/ERS phenomenon of subject 3 is not very obvious when the energy is abnormally high in the 17~30 Hz band of Cz channel. When it performs unilateral limb motor imagination, the mainly affected motor sensation area appears in the Cz channel, which leads to the reduction of the differentiated features that can be extracted by CNN network and the poor training effect of classifier.

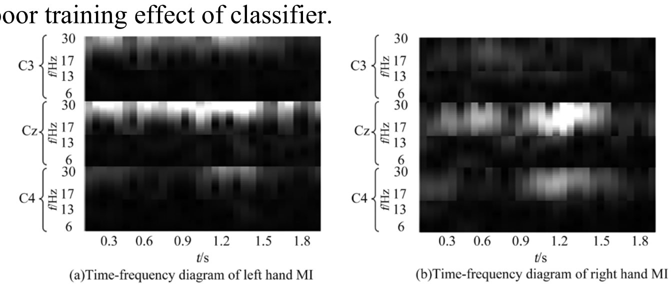

Figure 12. Time-frequency diagram of MI in subject 3

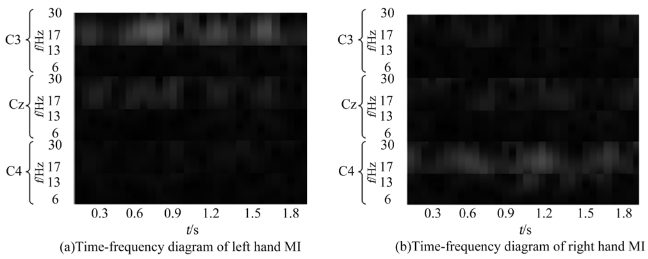

Figure 13. Time-frequency diagram of MI in subject 4

The classification accuracy of most classification models in subject 2 and 3 is 60%~70%, and some classification methods are even lower than 60%, while the values with new method in this paper are all above 60%. Moreover, it can be seen from table 2 that the Kappa value of this method on subject 2 is higher than that of other methods, so this model also has certain advantages on subjects with low feature differences.

Deep learning methods generally require a large amount of data to prevent training over-fitting and also enhance the generalization ability of the model depth [32]. In this paper, the generalization ability of CNN model is enhanced by data enhancement and adding improved Adaboost algorithm. As for the setting of the base learners number and the iteration rounds number, the next step is to select the adaptive number of base learners and the number of iteration rounds according to the condition of each subject. In addition, enhancing the correlation between training data and test data is also the focus of the future research.

We also conduct time consumption to verify the effectiveness of proposed method. The result is shown in table 3. It shows that the time with proposed method is shorter than other methods.

Table 3. Comparison of time/s

Subjects	SVM	Twin SVM	CNN	Proposed
Average	5.88	2.67	1.39	**0.12**

6. Conclusion

In this paper, our main purpose is to clearly classify the motor imagery EEG signals, so a channel space weighted fusion-oriented feature pyramid network is proposed to classify the left and right hand motor imagery EEG signals. Compared with the traditional SVM method and other state-of-the-art methods on BCI Competition IV 2b data set, the proposed method is feasible and effective in the dichotomous classification of motor imagery EEG signals, which can enhance the robustness of the classification model and improve the classification accuracy.However, for the complex EEG under complicated environments, it will be limited, resulting in low EEG signal recognition rate. In the future, we will synthetically consider various factors and adopt more advanced deep learning methods to perfect the classification of motor imagery EEG signals.

Acknowledgments.

The authors appreciate the anonymous reviewers for their meaningful comments.

References

[1] Bell, I, R., Howerter, A., Jackson, N., et al. (2012) Multiweek Resting EEG Cordance Change Patterns from Repeated Olfactory Activation with Two Constitutionally Salient Homeopathic Remedies in Healthy Young Adults. *Journal of Alternative & Complementary Medicine*, 18(5): 445.

[2] Han, C., Müller, K., and Hwang, H. (2020) Enhanced Performance of a Brain Switch by Simultaneous Use of EEG and NIRS Data for Asynchronous Brain-Computer Interface. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 28(10): 2102-2112.

[3] Mishuhina, V., Jiang, X. (2021) Complex Common Spatial Patterns on Time-Frequency Decomposed EEG for Brain-Computer Interface. *Pattern Recognition*, 115(1):107918.

[4] Lu, H., Eng, H., Guan, C., et al. (2010) Venetsanopoulos, "Regularized Common Spatial Pattern With Aggregation for EEG Classification in Small-Sample Setting. *IEEE Transactions on Biomedical Engineering*, 57(12): 2936-2946.

[5] Wang, H. et al. (2020) An Approach of One-vs-Rest Filter Bank Common Spatial Pattern and Spiking Neural Networks for Multiple Motor Imagery Decoding. *IEEE Access*, 8: 86850-86861.

[6] Antelis, J, M., Rivera, C, A., Galvis, E., et al. (2020) Detection of SSVEP based on empirical mode decomposition and power spectrum peaks analysis. *Biocybernetics and Biomedical Engineering*, 40(3).

[7] Yin, S., Zhang, Y. (2019) Singular value decomposition-based anisotropic diffusion for fusion of infrared and visible images. *International Journal of Image and Data Fusion*, 10(2): 146-163.

[8] Islam, S., Boric-Lubecke, O., Lubekce, V. (2020) Concurrent Respiration Monitoring of Multiple Subjects by Phase-Comparison Monopulse Radar Using Independent Component Analysis (ICA) With JADE Algorithm and Direction of Arrival (DOA). *IEEE Access*, 8: 73558-73569.

[9] Yin, S., Liu, J., Teng, L. (2018) A new krill herd algorithm based on SVM method for road feature extraction. *Journal of Information Hiding and Multimedia Signal Processing*, 9(4): 997-1005.

[10] Zhang, H., Wen, B., Liu, J., Zeng, Y. (2019) The Prediction and Error Correction of Physiological Sign During Exercise Using Bayesian Combined Predictor and Naive Bayesian Classifier. IEEE Systems Journal, 13(4), 4410-4420.

[11] Shi, Q., Yin, S., Wang, K., et al. (2021) Multichannel convolutional neural network-based fuzzy active contour model for medical image segmentation. Evolving Systems. https://doi.org/10.1007/s12530-021-09392-3

[12] Fu, Q., Wang, H. (2020) A Novel Deep Learning System with Data Augmentation for Machine Fault Diagnosis from Vibration Signals. *Applied Sciences*, 10(17):5765.

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

[14] Phang, C., Noman, F., Hussain, H., et al. (2020) A Multi-Domain Connectome Convolutional Neural Network for Identifying Schizophrenia From EEG Connectivity Patterns. *IEEE Journal of Biomedical and Health Informatics*, 24(5): 1333-1343.

[15] Yang, H., Sakhavi, S., Ang, K., and Guan, C. (2015) On the use of convolutional neural networks and augmented CSP features for multi-class motor imagery of EEG signals classification. *In 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Milan, Italy, 25-29 Aug. 2015, 2620-2623.

[16] Li, Y., Lei, M., Zhang, X., et al. (2018) Boosted Convolutional Neural Networks for Motor Imagery EEG Decoding with Multiwavelet-based Time-Frequency Conditional Granger Causality Analysis. 2018. arXiv:1810.10353.

[17] Gao, Z., Dang, W., Wang, X., et al. (2020) Complex networks and deep learning for EEG signal analysis. *Cognitive Neurodynamics*, 11: 369-388.

[18] Meng, X, J., Qiu, S., Wan, S., et al. (2021) A motor imagery EEG signal classification algorithm based on recurrence plot convolution neural network. *Pattern Recognition Letters*, 146(4): 134-141.

[19] Shi, M., Yang, C., Zhang, D. (2021) A Smart Detection Method of Sleep Quality Using EEG Signal and Long Short-Term Memory Model. *Mathematical Problems in Engineering*, 2021: 1-8.

[20] Fernandez-Blanco, E., Rivero, D., Pazos, A. (2020) EEG Signal Processing with Separable Convolutional Neural Network for Automatic Scoring of Sleeping Stage. *Neurocomputing*, 410:220-228.

[21] Xu, Y., Li, D., Xie, Q., et al. (2021) Automatic defect detection and segmentation of tunnel surface using modified Mask R-CNN. *Measurement*, 178(4):109316.

[22] Loey, M., Manogaran, G., Taha, M., et al. (2020) Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection. *Sustainable Cities and Society*, 65(3).

[23] Gao, X., et al. (2018) An End-to-End Neural Network for Road Extraction From Remote Sensing Imagery by

Multiple Feature Pyramid Network. *IEEE Access*, 6, 39401-39414.

[24] Zhang, H., Guan, C., Ang, K., et al. (2012) BCI Competition IV-Data Set I: Learning Discriminative Patterns for Self-Paced EEG-Based Motor Imagery Detection. *Frontiers in Neuroscience*, 6:7-.

[25] Yang, T., Ma, Y., Gao, Y., et al. (2021) Recognition of Motor Imagery EEG Signals Based on Boosted Convolution Neural Network Model. *Space Medicine &. Medical Engineering*,34(2): 128-136

[26] Jiang, Y., Hau, N., and Chung, W. Semiasynchronous BCI Using Wearable Two-Channel EEG. *IEEE Transactions on Cognitive and Developmental Systems*, 10(3):681-686.

[27] Shovon, T, H., Nazi, Z, A., Dash, S., et al. (2019) Classification of Motor Imagery EEG Signals with multi-input Convolutional Neural Network by augmenting STFT. *In 5th International Conference on Advances in Electrical Engineering (ICAEE)*. 26-28 September, 2019, Dhaka, Bangladesh.

[28] Boukhtache, S., Blaysat, B., Grédiac, M., et al. (2021) FPGA-based architecture for bi-cubic interpolation: the best trade-off between precision and hardware resource consumption. *Journal of Real-Time Image Processing*, 18: 901-911.

[29] Song, W., Yao, J., Zhang, K., et al. (2021) The Impacts of Pore Structure and Relative Humidity on Gas Transport in Shale: A Numerical Study by the Image-Based Multi-scale Pore Network Model. *Transport in Porous Media*, 2021:1-25.

[30] Cheng, Q., Wang, G., Dong, Q., et al. (2020) Arbitrary-shaped text detection with adaptive convolution and path enhancement pyramid network. *Multimedia Tools and Applications*, 79(8): 29225-29242.

[31] Soman, S., Jayadeva. (2015) High performance EEG signal classification using classifiability and the Twin SVM. *Applied Soft Computing Journal*, 30:305-318.

[32] Wang, J., Li, H., Yin, S., and Sun, Y. (2019) Research on Improved Pedestrian Detection Algorithm Based on Convolutional Neural Network. 2019 International Conference on Internet of Things (iThings) and IEEE Green Computing and Communications (GreenCom) and IEEE Cyber, Physical and Social Computing (CPSCom) and IEEE Smart Data (SmartData), Atlanta, GA, USA, 14-17 July 2019, 254-258.