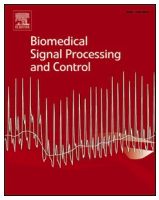
Biomedical Signal Processing and Control 72 (2022) 103338

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc



Deep CNN model based on serial-parallel structure optimization for four-class motor imagery EEG classification

Xuefei Zhao [a], Dong Liu [a], Li Ma [a,*], Quan Liu [a], Kun Chen [a], Shane Xie [b], Qingsong Ai [a]

[a] Wuhan University of Technology, Wuhan, China
[b] University of Leeds, Leeds, UK

ARTICLE INFO

Keywords:
Motor imagery electroencephalogram (MI-EEG) signal
Deep learning
Convolutional neural network
Serial-parallel (SP) structure
Multi-dimensional feature extraction

ABSTRACT

Motor imagery electroencephalogram (MI-EEG) is one of the most important brain-computer interface (BCI) signal. It is vital to analyze the MI-EEG for the manipulation of external BCI actuator. However, traditional methods usually undertake EEG feature extraction and classification separately, which may lose efficient feature information. It behaves beyond our satisfaction for multi-class MI activity evoked by space-close and cannot eliminate the influence of individual differences. To solve these problems, we propose a convolutional neural network (CNN) with an end-to-end serial-parallel (SP) structure followed by tranfer learning. In detail, we use the serial module to extract the rough features in time–frequency-space domain, and the parallel module for fine feature learning in different scales. Meanwhile, a freeze-and-retrain fune tuning transfer learning strategy is proposed to improve the cross-subject accuracy. When our model is compared with the other three typical networks, results show that the proposed model performs best with the average testing accuracy of 72.13% and the average loss of 0.47, among which one subject only takes 0.7 s to reach 89.17% as the highest one. Through transfer learning, we reduce the training parameters by 53%. The average cross-subject classification accuracy increases by approximate 15%, and the individual highest accuracy reaches 76.98%. In conclusion, the integrity and separability of SPCNN determine that we require no additional EEG signal feature analysis, which is conducive to the realization of an efficient online BCI. It can also get rid of the dependence on training time and subject data to rapidly advance BCI in the future.

1. Introduction

Brain-computer interface (BCI) is a system that establishes a bridge between human and external devices, where the signal is mainly derived from electroencephalogram (EEG) [1]. BCI can acquire and analyze signals and then interacts with the end equipment by mutual information flow [2]. Combined with biocompatible materials, sensor technology, embedded computing technology and neuroscience, BCI can be used to realize human–computer involved task scenarios, such as vehicle intelligent drive system, smart home, patient service and rehabilitation, etc[3]. In 2017, Zander et al. set up a BCI system for driver's fatigue and mental state detection using an N200-P300 event related potential (ERP). The parietal EEG alpha rhythm with two features are employed to evaluate the working condition of the system in the laboratory and real world environment, thus providing voice prompt, and activating the auxiliary driving mode of the vehicle when necessary [4]. The work can be further applied to automatic driving, and the driving safety can be guaranteed through the dual control of man and vehicle. Fan et al. used convolutional neural network (CNN) to classify different types of motor imagery EEG (MI-EEG) signals, and used ZigBee network to realize the communication between the main control center and the controlled household appliances, and realized the real-time intelligent control of home equipment [5]. Steady-state visual evoked potential (SSVEP), one of popular EEG data paradigms, is also usually analyzed to control home devices [6]. A BCI system based on deep learning only used "idea" to realize robot flexible walking and wheelchair control in 2019 [7]. Combined with artificial intelligence, SSVEP can also be used to control the robotic arm [8]. It not only improves users' quality of life, but also help stroke patients restore damaged limbs and is helpful to the remodeling of brain neural circuit.

A typical BCI paradigm consists of brain signal acquisition, data processing and interaction with the external machine. Signal acquisition and processing mainly includes EEG signal and some other signal format like fNIRS or fMRI, which are utilized in combination style[9–10]. The

* Corresponding author.
E-mail address: excellentmary@whut.edu.cn (L. Ma).

https://doi.org/10.1016/j.bspc.2021.103338
Received 7 June 2021; Received in revised form 29 October 2021; Accepted 2 November 2021
Available online 12 November 2021

EEG signal sources of BCI varies, such as SSVEP, P300, MI, and potential hybrid combination of them. When receiving a fixed frequency visual stimulus, the visual cortex of the brain will produce a continuous response related to the frequency of the stimulus, this response is called SSVEP [11]. P300 is a component of ERP [12], which is a late positive wave with a latency of about 300 ms induced by stimulus. When people imagine their body movements but no actual motor output, there will still be activation of certain brain areas called MI. In the process of MI, the cerebral cortex will generate two kinds of rhythm signals with obvious changes. The rhythmical energy of the contralateral motor sensory region of cerebral cortex is significantly decreased, while that of the ipsilateral motor sensory region is increased. This phenomenon is called event-related desynchronization (ERD)/ event-related synchronization (ERS) [13]. MI-EEG signal does not rely on additional equipment and it requires only the subject's imagination, which is widely used for its remarkable phenomenon and spontaneity.

The traditional feature extraction methods can be conducted in the time domain, frequency and spatial domain. The time domain analysis is mainly to find the statistical properties of the signal waveform, while the frequency domain analysis focuses on spectral characteristics of the signal. The method combining the two aspects is called time–frequency analysis method, for example, local feature scale decomposition (LCD) [14], discrete wavelet transformation (DWT) [15] and Winger-Ville distribution [16]. and flexible analytic wavelet transforms (FAWT) [17]. The spatial analysis mainly extracts the spatial distribution components of signals, and the common spatial pattern (CSP) is one of widely recognized methods. However, the CSP method utilizes a wide frequency band to extract features, which contains most useful information as well as redundant information. Therefore, many improved algorithms based on CSP are proposed. Taking a single EEG signal channel as an example, it is divided into multiple sub-channels according to the frequency band. The common spatio-spectral pattern (CSSP) uses the CSP algorithm to optimize the basic filter by inserting different time delay in different sub-channels [18]. The common sparse spectral spatial pattern (CSSSP) makes all sub-channels share the same spectrum mode on the basis of CSSP method [19]. The sub-band common spatial pattern (SBCSP) extracts CSP features from MI-EEG signals at multiple channels, and uses linear discriminant analysis (LDA) to reduce the dimensions of each sub-channel, and finally fuses them [20]. The filter band common spatial pattern (FBCSP) studies the mutual information between sub-channels on the basis of SBCSP, and extracts the most representative features of sub-channels [21]. The classification results of EEG signals can be obtained by constructing a classifier based on the extracted features. Feature classification methods include the spectral regression discriminant analysis (SDRA) [22], which is a combination of spectral analysis and linear regression, support vector machine (SVM) [23] and fisher discriminant analysis [23]. However, MI-EEG signals are low signal-to-noise ratio (SNR), time-varying and large individual differences. And feature extraction and classification of the methods mentioned above are discontinuous, and many of them are manual and based on prior knowledge of predecessors. Therefore, the extracted features are often not comprehensive enough to affect the classification accuracy.

In order to overcome the disadvantages of traditional methods, many classification methods based on the deep learning have been proposed recently [24]. Recurrent neural networks (RNN), CNN, or their combinations are common approaches. Some researchers use the raw multi-channel MI-EEG signals or convert them into spectral images with retained topological structure, feed them the into the complex network, extract the time and space information of the signals to realize the classification goal. Thus, many optimized networks are developed, such as shallow convolutional neural network (ShallowNet), deep convolutional neural network(DeepNet), residual convolutional neural network (ResNet), and deep separable convolutional called as EEGNet [25,26,27]. They are typical methods for the work comparison in machine learning based BCI data processing. The reason why these models can apply to the MI EEG data analysis is that the models can learn the underlying infromation within EEG and shows their own superiority. To be explicit, ShallowNet can utilize small dataset for training, like our MI EEG data of few cases and it can decrease the over-fitting risks. Deep convolutional neural network called DeepNet is based on general CNN structure and kernel. It is a typical machine learning model for data analysis because CNN is quite feasible and powerful and also DeepNet mathmatically extends and upgrades the capability to large and complex EEG signal. It can be one of the suitable methods for MI EEG data analysis [28]. The EEGNet behave as a single CNN architecture to accurately classify EEG signals from different BCI paradigms, while simultaneously being as compact as possible. It is robust enough to learn a wide variety of interpretable features over a range of BCI tasks and paradigms. However, those deep learning models still face two critical challenges: one is that the network involves too many complicated layers and huge amount of parameter to be trained, which leads to long training time [29]. On the other hand, some networks are too shallow, the extracted features are not comprehensive enough, and the classification accuracy is relatively low. At the same time, both the traditional method and the deep learning method have the problem that the training model of the same subject cannot be applied to other subjects, which results in a large number of experimental data requirements and training time. Therefore, a more reasonable and universal network architecture is needed to ensure comprehensive feature extraction and avoid network parameter redundancy. For example, the structure of the network can be diversified, not limited to the general series form of layer upon layer connection, or add some special layers to enrich the features.

To solve the above problems, this paper proposes a new network structure to optimize the classification efficiency of MI-EEG signals. The network is mainly composed of multiple convolutional layers in series, and has a parallel structure composed of multiple branches to enrich feature extraction. The main work contents are as follows: (a) a serial-parallel structure convolutional neural network (SPCNN) is proposed for multi-domain and multi-level feature extraction; (b) Compared with the existing mainstream deep learning network in terms of MI-EEG signal classification, the proposed network is proved to be advanced; (c) A freeze-and-train transfer learning strategy is proposed to improve the accuracy of cross-subject evaluating and improve the universality of the network; (d) Using a visualization method, the effectiveness of the network is verified in time and frequency domains, respectively. The structure of the article is as follows: The introduction is given in Section I. Section II introduces materials and methods used in the experiment. Section III covers the experimental results. In Section IV, the advancements of our network and the limitation and prospect of this paper are discussed. The last Section concludes the paper.

2. Materials and methods

2.1. Dataset

Data are from BCI competition 2008 2a MI-EEG dataset, obtained from nine subjects. The experimental data of each subject can be divided into two groups, namely the training set and the testing set. Each part contains 288 trials, within 72 trials for each type of MI-EEG signals [30]. The experimental data of subjects are collected in accordance with the experimental paradigm shown in Fig. 1. When t = 0 s, the monitor

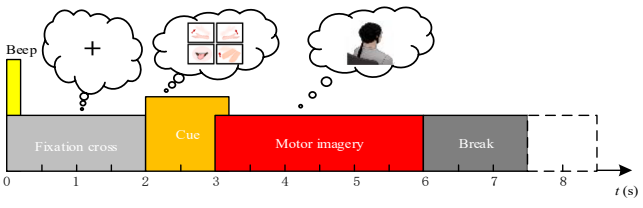


Fig. 1. Experimental paradigm for data collection.

displays a fixation cross, and the subjects prepare for MI. In addition, a short acoustic warning tone is presented; When t = 2 s, an arrow pointing of MI (both feet, left hand, right hand, tongue) appears as a prompt and stayed on the screen for 1.25 s, then the subjects begin to carry out corresponding MI, and one kind of MI action tiral lasts for 3 s from appearance of the cue to the end, in other words, the data used for our analysis are all balanced; When t = 6 s, the word "break" appears to show the end. The experimental collection device has 25 channels, of which there are 22 EEG electrodes and 3 electrooculogram (EOG) electrode, and the distribution of the main electrodes is shown in Fig. 2. It meets the international 10–20 standard with a sampling rate of 250 Hz. Then data are filtered through a 50 Hz notch filter to eliminate power frequency interference, followed by a band-pass filter of 0.5–100 Hz to retain the main components of human brain wave frequency. In this paper, the 4-second data is selected from the beginning of "Cue" prompt to the end of the MI for experiment. The one trial experimental data includes 1000 sampling points, so the input form of the network is 22 * 1000, i.e. 1,584,000 samples in each class for each subject.

2.2. Pre-processing and CNN architecture design

In order to simplify the preprocessing steps, we use wavelet de-noising tactic with soft threshold to remove the inevitable physiological and environmental noise. Wavelet threshold (WT) de-noising method is characterized by low-entropy, multi-resolution and de-correlation, which is a good non-stationary signal de-noising method [31,32]. The main steps of WT de-noising method are: (a) the input original signals are processed by wavelet decomposition; (b) The wavelet coefficients of each layer are extracted and the threshold values are calculated to conduct threshold treatment on the wavelet coefficients; (c) The signals are reconstructed by processed data.

Then, the preprocessed signals are input of the end-to-end CNN model with serial-parallel structure. Our SPCNN firstly converts the multi-channel time series into 3D from 2D shape, and extracts the features in time domain, space domain and frequency domain through the serial structure. Then the parallel structure is used to extract the futher detail features in the three domains and get them fused systematically. Finally, the classification results can be obtained after integrating the features through the double connection layer. The main structure of the model is shown in Fig. 3. As it is shown, this new model main consists of serial parallel components. Both parts are key innovation points in our model design. This framework is inherited by the traditional covolution idea in the image processing area. It means the features are extracted from general skeleton level to advanced details in further. Therefore, we come up with serial parallel stragetry for in-depth feature mining. The two hierarchy parts expand our EEG feature learning capabitly in dierction of width and depth. Beside, this new proposed model are based on the mature effective CNNs, which has been widely applied to a lot of areas. Moreover, the SPCNN undertakes the feature extraction in multi-domain, i.e time domain in the temporal and space domain by multi-channel data combination via space filters. Besides, the parallel part are based on the work of serial part and has mutli co-existing module to work , which accelerates the model training and optimization from the output of the serial part at the same time. Pseudocode I. gives the details of coding instructions.

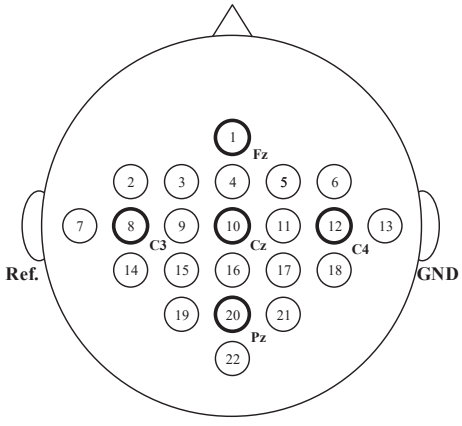


Fig. 2. Schematic diagram of EEG electrode position.

Pseudocode I. The whole work procedure.

1. Let $X(i,j)$ be the raw data;
// i = No. of the channel, j = length of data in each channel
2. Call discrete wavelet transform(DWT) package;
// DWT for de-noising
3. Call convolution neural network package;
4. Wavelet_cofficient(k) ← DWT($X(i,j)$);
5. Wavelet_component(k) ← DWT($X(i,j)$);
6. Initialize the soft threshold ε and the threshold function output $\omega(k)$;
7. For k = 1:m
//m is number of layers by the DWT decomposition
8. If (Wavelet_cofficient(k) < ε.
9. $\omega(k) = 0$;
10. else
11. $\omega(k)$ = Wavelet_cofficient(k)- ε;
12. End
13. End
14. Wavelet_cofficient_optimized (k) =$\omega(k)$;
X_denoised(i,j)15. = reconstruct _wavelet (Wavelet _cofficient_optimized, Wavelet_component(k));
16. Initialize the parameters of SPCNN serial module in the Table 1;
17. Set activation function $f_1(x) = x^2$, $f_2(x) = \log(x)$;
18. Feature_serial (x, y, z) = SPCNN_serial(*X_denoised*);
19. Initialize the parameters of SPCNN parallel module in the Table 1;
20. Feature_parrallel_1 (x, y, z) = SPCNN_ parallel_P1 (Feature_serial (x, y, z));
21. Feature_parrallel_2 (x, y, z) = SPCNN_ parallel_P2 (Feature_serial (x, y, z));
22. Feature_parrallel_3 (x, y, z) = SPCNN_ parallel_P3 (Feature_serial (x, y, z));
23. Feature (x, y, z) = concatenate (Feature_parallel_1 (x, y, z), Feature_parrallel_2 (x, y, z), Feature_parallel_3 (x, y, z));
24. Initialize the parameters of SPCNN classification module in the Table 1;
25. MI_labels = Classification (Feature (x, y, z));

2.2.1. Serial feature extraction module

Serial feature extraction module employs several temporal filters and spatial filters to extract temporal and spatial features of signals, and a series of nonlinear processing is used to extract characteristics of frequency. The specific implementation process is as follows. The two-dimensional signal of the input network is transformed into a three-dimensional form suitable for the subsequent network layer, followed by two-dimensional convolution kernels with core sizes of [t, 1] and [1, c] selected for time-domain convolution and spatial convolution respectively. The setting value of t is explained below, and c is the number of EEG electrode channels. This process extracts the temporal and spatial characteristics of the signal, which are both followed by batch normalization. Thus, the signal channel dimension is compressed to 1, and the third dimension feature originally 1 is extended to p, where p is related to the number of spatial filters. Next, there are two nonlinear activation functions, the first one is the square nonlinear function as shown in formula (1) and the second one is the logarithmic nonlinear function as shown in formula (2), both of which are used to extract the features related to the frequency band power in the signal, so as to enhance the non-linear expression ability of the network. In the end, the average pooling layer is selected to change the time dimension of the signal, and the dropout layer is used to avoid over fitting phenomenon. All the above network layers are connected by serial structure, so they together constitute serial feature extraction layer. The serial feature extraction layer is shown in Fig. 4 (a). (see Table 1)

$$f1(x) = x^2 \tag{1}$$

$$f2(x) = log(x) \tag{2}$$

where $\times$ denotes the input signal of the activation layer. In the two nonlinear layers of the network, square and log functions were used as the activation functions of the layer respectively. These two nonlinear

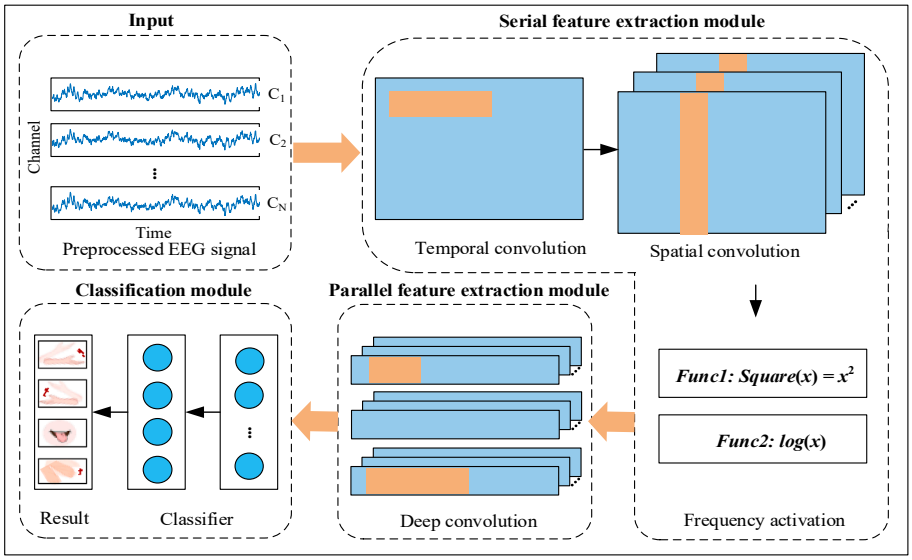


Fig. 3. Main structure of SPCNN.

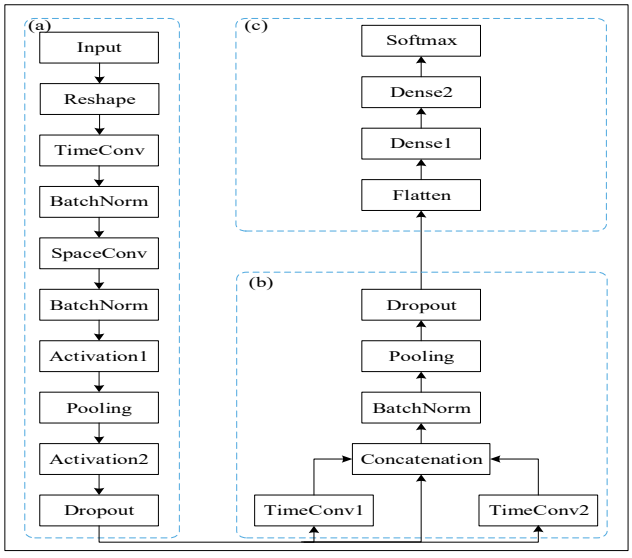


Fig. 4. Hierarchical framework of SPCNN. (a) Serial feature extraction module: input deformation, extraction of time domain, spatial domain and frequency domain characteristics; (b) Parallel feature extraction module: depth characteristics extraction at different scales; (c) Classification module: category determination of signal.

layers can reduce the features, simplify the network and adapt to the characteristics of EEG signals without reducing the network feature extraction capacity. In table I in [32], the authors explained the activation functions of each layer, in which the nonlinear layer is square(x) and log(x).

2.2.2. *Parallel feature extraction module*

The input of the parallel feature extraction layer is the output of the serial feature extraction layer, which is the MI-EEG signal after the preliminary three-dimensional feature extraction. In the serial feature extraction layer, the third-dimensional feature has been changed through some filters with specific convolution kernels. In the parallel feature extraction layer, we use two separate 20 filters to change the third-dimensional features of the signal for two stream path, and make different convolution kernels for these two kinds of filters, which are set

Table 1
Framework parameter design of SPCNN.

Block	Layer	Filter	No. of Parameters	Output	Option
Serial	Input			(C, T)	
	Reshape			(1, C, T)	
	TimeConv	40 (1, 13)	560	(40, 22, 1000)	Linear activationMode = same
	BatchNorm		160	(40, 22, 1000)	
	SpaceConv	40 (22, 1)	70,400	(40, 1, 1000)	Linear activationMode = valid
	BatchNorm		320	(40, 1, 1000)	
	Activation			(40, 1, 1000)	Square activation
	Pooling	(1, 75)		(40, 1, 62)	Average pooling, Stride=(1, 15)
	Activation			(40, 1, 62)	Square activation
	Dropout			(40, 1, 62)	P = 0.5
Parallel	TimeConv1	20 (1, 16)	25,600	(20, 1, 62)	Linear activationMode = same
	TimeConv2	20 (1, 8)	12,800	(20, 1, 62)	Linear activationMode = same
	Concatenation			(80, 1, 62)	
	BatchNorm		480	(80, 1, 62)	
	Pooling	(1, 3)		(80, 1, 20)	Max pooling
	Dropout			(80, 1, 20)	P = 0.5
Classifier	Flatten			2400	
	Dense1		24,010	10	Max norm = 0.25
	Dense2		44	4	Max norm = 0.25
	Softmax			4	

as [1,16] and [1,8]. In order to extract features in deep-level characteristics of space–time domain, a larger size convolution kernel and a smaller size convolution kernel are set up respectively. Then, the signals extracted by different scale deep-level features are fused by concatenate layer with that the fused signals are processed by batch normalization to make the distribution of the data suitable to its true distribution. Followed by the maximum pooling layer with convolutional kernel of [1,3] for dimension change, dropout layer with rate of 0.5 is used to avoid overfitting phenomenon again. The core part of the above network layer is connected by parallel structure, so it is called parallel feature extraction layer, as shown in Fig. 4 (b).

2.2.3. Classification module

We use a flatten layer to compress the processed multi-dimensional signal into one dimension. After that we use two different numbers of dense layers to converge the data faster for extracting the association between these features, and map them to the output space to get the output form we need. According to the probability value, a softmax layer is used to judge the final decision. The process is shown as Fig. 4 (c).

2.3. CNN model construction

This session can be divided into two stages: the first stage denotes the initialization of the network and model construction based on within-subject work procedure. The other one is to pre-train model in the same way as above and improve it with transfer learning technique. The final model is used for cross-subject evaluation.

2.3.1. Network initialization

The structure and corresponding parameter settings of the whole network are shown in Table 1. The network basically consist of three parts: serial, parallel and classifier. As described in 2.2 section, each part have different CNN based structure and the number of filers and some activation type are also given below.

where $C = 22$ is the number of electrode channels, $T = 1000$ is the number of sampling points.

2.3.2. Freeze and train approach

EEG signal is weak and there are huge differences among different individuals. Many studies show that the testing performance of classification algorithms is generally limited to the same subject. In other words, that is, part of a subject's data is used to train the model, and the rest is used to test performance. It takes a lot of time to collect experimental data and train the model. However, models built by a small number of subjects can be applied to other subjects, which can greatly improve the efficiency.

In this paper, we use the method of transfer learning to achieve the above purpose. The motivation of transfer learning is that people can use what they have learned before to solve new problems faster or better [33]. Transfer learning here can be further divided into fine-tuning strategy and freeze and train strategy. Through the latter method, we first build a model, and then integrate some new data to retrain the structure of the trained model. Freeze and train strategy can increase the universality of the trained model between different subjects, reduce the number of training parameters, shorten experimental time and improve the learning performance of the network. The time involves specifically data acquisition time, model training time and much expensive data-labelling efforts.

The dataset consists of nine subjects, so there are nine models that have been constructed within subjects, and each of them need to be tested across subjects by leave-one-out cross validation strategy. Freeze and train strategy is adopted to evaluate cross-subject performance of each subject. Taking one subject as an example, his initial training model is based on training data from the other top four subjects in within-subject experiment. Then we add this subject's training data to the former data to retrain the model, and we can get this model updated. Finally, his testing set is used to evaluate his cross-subject classification accuracy. The entire process is shown as Algorithm I.

Algorithm. I. *Freeze and train for cross-subject testing of subject A.*

Input: S_B: $D_1=\{(X^{(i)}, L^{(i)}), i = 1, 2, \ldots, n/4\}$
S_C: $D_2=\{(X^{(i)}, L^{(i)}), i = 1, 2, \ldots, n/4\}$
S_D: $D_3=\{(X^{(i)}, L^{(i)}), i = 1, 2, \ldots, n/4\}$
S_E: $D_4=\{(X^{(i)}, L^{(i)}), i = 1, 2, \ldots, n/4\}$

Procedure:
1: Together as an overall input $D_{\text{original}}=\{(X^{(i)}, L^{(i)}), i = 1, 2, \ldots, n\}$ of SPCNN.
2: Train an initial model M_0.
3: Freeze part of the network layer.
4: Combine part of the original input and part of the subject A's training set to be a new input $D_{\text{new}}=\{(X^{(i)}, L^{(i)}), i = 1, 2, \ldots, n\}$.
5: retrain the model M_0 and form a new model M.
Output: A new well-trained model M of SPCNN

There are 19 layers in SPCNN, including serial feature extraction module from the 2nd to the 9th layer, parallel feature extraction module from the 10th to the 15th layer, and classification layer from the 16th to the last layer. At first, we freeze the first 9 layers, that is, retain the parameters of the last 10 layers for retraining. The part of retraining is that after integrating the training data of the subject to be evaluated, the time–frequency-space features are preliminarily extracted by using the parameters of serial feature modules trained by other subjects, and the parameters of parallel feature module and classification module are updated to extract deep level features.

3. Results

3.1. Wavelet threshold de-noising performance

WT method is used to remove the noise in each channel. "sym4" is selected as the wavelet basis function, and the decomposition level is set to 3. They were haar, db4, sym4, coif5, bior4.4, and bior6.8, and sym4 based WT decomposition usually obtains a higher entropy. And sym4 oriented WT method often applies well to physiological and neural signal [34–35]. The threshold of one-dimensional wavelet transform is obtained by Birge-Massart algorithm, and the wavelet soft threshold de-noising method is used. Three electrodes 'C_3', 'C_4', 'C_z' in the first trial of the training data of subject 1 are selected as representatives. The experimental results are shown in Table 2 and Fig. 5.

It can be seen from Table 2 that the correlation of the signals after de-noising of the three channels is above 0.9, indicating that the signals by de-noising are significantly correlated without distortion. At the same time, the smoothness (s) value of the de-noised signal is about 0.1, indicating that the reconstructed signal has a high degree of smoothness. From Fig. 5, the original signal obviously contains a lot of noise in the form of burr. After de-noising, these burrs are basically eliminated and the de-noising is effective. The EEG signal is preprocessed effectively by wavelet de-noising.

3.2. Model training outcome

In this paper, we use the Adam optimizer [36], one of the commonly used optimizers, which has simple implementation, high calculation efficiency and less memory requirement. After the bias correction, each iterative learning rate of the optimizer is kept in a certain range, which makes the parameters more stable. The loss function selected is cross

Table 2
De-noising results of WT method (typical electrodes).

Channel	Correlation coefficient	smoothness
C_3	0.9267	0.10
C_4	0.9196	0.09
C_z	0.9329	0.12

Fig. 5. De-noising results of WT method (typical electrodes).

entropy loss function, as shown in formula (3). Cross entropy can measure the difference between two different probability distributions in the same random variable, which is expressed as the difference between the real probability distribution and the predicted probability distribution in machine learning.

$$Loss = -\sum pi(x)logqi(x), i = 1, 2, ..., n \tag{3}$$

where $p_i(x)$ denotes the i-th target probability distribution, and $q_i(x)$ denotes the i-th predicted probability distribution. The smaller the value of cross entropy loss, the better the prediction effect of the model.

We compare the proposed network SPCNN with DeepNet, ShallowNet, and EEGNet. We set the batch size to 64. In the first experiment, the number of iterations is 1000. It is observed that when the number of iterations is 400, the loss value decreased and tended to be stable. Therefore, the number of iterations in the follow-up experiment is set to 400. All networks are implemented by tensorflow and keras. Through the 10-fold cross method, the training data of the subjects are randomly divided into ten parts, of which nine are training parts, and the remaining parts are verification parts, which are carried out ten times until all the data serve as both training part to establish the model and verification part to verify. The training results are shown in Table 3.

Table 3 shows the average training classification accuracy in ten random trials of each subject and the average accuracy of all subjects under different networks. The results show that the average accuracy of subject 2 is the lowest, while that of subject 3 or subject 7 is the best. In the network proposed in this paper, for subject 2, the average accuracy is 71.65%, which was 9.8% higher than the best of the other three methods. For subject 7, the average accuracy is 93.66%, which is higher than 39% of the same subject in DeepNet. The average accuracy of our method is 83.65%, which is 16.4%, 6.3% and 4.3% higher than the other three methods respectively. DeepNet has a complex architecture, the extracted signal features might be nothing and disruptive to the classified results, and the other shallow networks could not extract enough features to assist classification. The minimum loss values of all models are shown in Table 4, and the number of parameters involved in each training of different models is shown in Table 5.

As can be seen from Table 3, Table 4 and Table 5, although the numbers of parameters of our proposed method are slightly more than those of EEGNet and ShallowNet, they are less than those of DeepNet. The performances of our network surpass other three networks. Compared with the other three models, the accuracy of our method is improved by 16.4%, 6.3% and 4.3%, and the minimum loss value is reduced by 35.4%, 20.6% and 7.9%, respectively.

3.3. Model performance

The testing set is used to evaluate the model established in the training part. Due to the individual differences of MI-EEG signals, the testing session is divided into within-subject test and cross-subject test according to the source of the training dataset.

3.3.1. Within-subject classification

The within-subject test is carried out, i.e. the training dataset and the testing dataset are from the same subject. The experimental results are shown in Table 6, Table 7 and Fig. 6.

Fig. 6 and Table 6 show the average testing accuracy of each subject using the ten-fold cross method under different networks and the average testing accuracy of corresponding network. The experimental results show that the average accuracy of subject 2 is the lowest, and that of subject 3 is the highest. In our method, the average accuracy of subject 3 can reach 89.17%, and the average accuracy of all subjects is 72.13%, which is higher than the other three methods of 17.6%, 5.5%, 3.7%. However, higher standard deviation in SPCNN means model may fit very well for the subjects while fair for others. This may be overcome by more subject data for model training, thus narrowing the gap. In our method, we choose the experiment with the highest evaluating accuracy of each subject in 10-fold cross method to draw the confusion matrix, as shown in Fig. 7.

Table 7 records the average testing time that input the testing dataset of subjects into the training model. As can be seen from the table, our method only needs 0.7 s to complete the experiment. The experimental results accord with the size of the model training parameters and prove the high efficiency of the proposed method.

Table 3
The training classification performance for all subjects (accuracy/%).

Network Subject	DeepNet	EEGNet	ShallowNet	SPCNN
Subject 1	73.13	80.98	79.29	80.63
Subject 2	55.98	54.96	65.14	71.52
Subject 3	79.46	88.80	92.19	92.64
Subject 4	74.42	69.02	65.94	75.40
Subject 5	74.60	79.15	77.81	80.09
Subject 6	54.73	72.59	71.52	74.64
Subject 7	67.28	84.96	93.98	93.66
Subject 8	83.21	86.16	86.61	93.04
Subject 9	83.84	91.65	89.02	91.25
Average	71.85 ± 10.68	78.70 ± 11.54	80.17 ± 10.98	83.65 ± 8.98

Table 4
The minimum loss value of all models.

Network Subject	DeepNet	EEGNet	ShallowNet	SPCNN
Subject 1	0.81	0.49	0.48	0.49
Subject 2	0.54	0.99	0.86	0.72
Subject 3	0.89	0.41	0.21	0.29
Subject 4	0.68	0.79	0.83	0.69
Subject 5	0.57	0.61	0.62	0.59
Subject 6	0.84	0.70	0.70	0.65
Subject 7	0.83	0.49	0.23	0.24
Subject 8	0.81	0.48	0.34	0.28
Subject 9	0.71	0.35	0.29	0.27
Average	0.74 ± 0.11	0.59 ± 0.19	0.51 ± 0.24	0.47 ± 0.19

Table 5
Model parameter.

Network	DeepNet	EEGNet	ShallowNet	SPCNN
Parameter	193,429	3268	46,164	134,374

Table 6
10-fold within-subject testing classification performance.

Network Subject	DeepNet	EEGNet	ShallowNet	SPCNN
Subject 1	66.28	74.72	73.37	78.85
Subject 2	46.94	53.92	52.99	51.04
Subject 3	74.17	83.51	87.95	89.17
Subject 4	67.40	59.34	64.34	67.60
Subject 5	63.13	71.42	63.37	63.54
Subject 6	43.40	56.94	56.18	57.40
Subject 7	59.65	72.78	77.08	82.08
Subject 8	66.84	67.95	77.88	80.80
Subject 9	63.96	74.51	72.85	78.72
Average	61.31 ± 9.99	68.34 ± 9.73	69.56 ± 11.23	72.13 ± 12.79

Table 7
Testing time.

Network	DeepNet	EEGNet	ShallowNet	SPCNN
Time/s	1.1	0.7	0.3	0.7

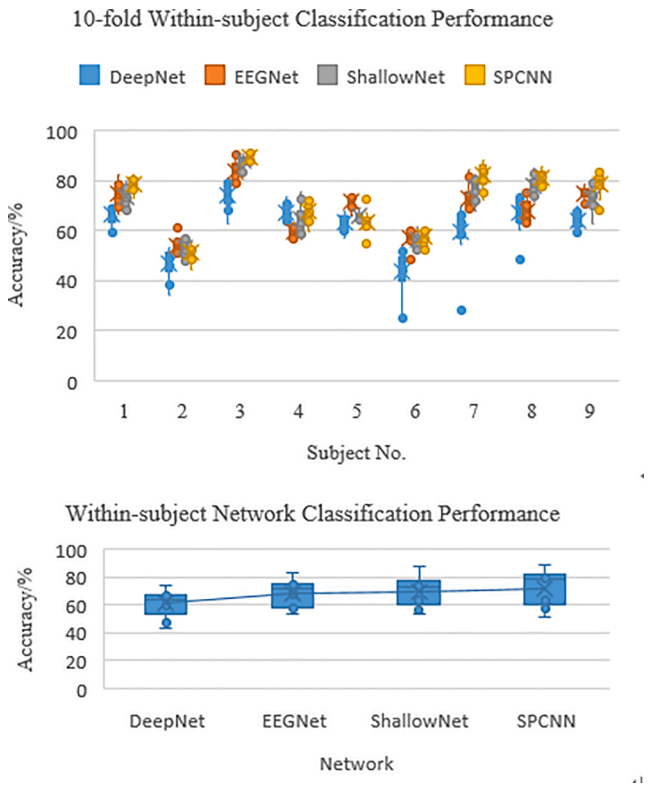


Fig. 6. Within-subject testing classification performance. (a) 10-fold cross classification performance; (b) Network classification performance.

3.3.2. *Cross-subject transfer learning*

The cross-subject evaluation is implemented by freeze and train strategy (as shown in Algorithm I), and the training dataset and the testing dataset come from different subjects.

We use 10-fold cross method to build initial model. And then the model is further used to evaluate specific subject. The average training accuracy, the loss of the model and the average cross-subject testing accuracy are shown in Table 8.

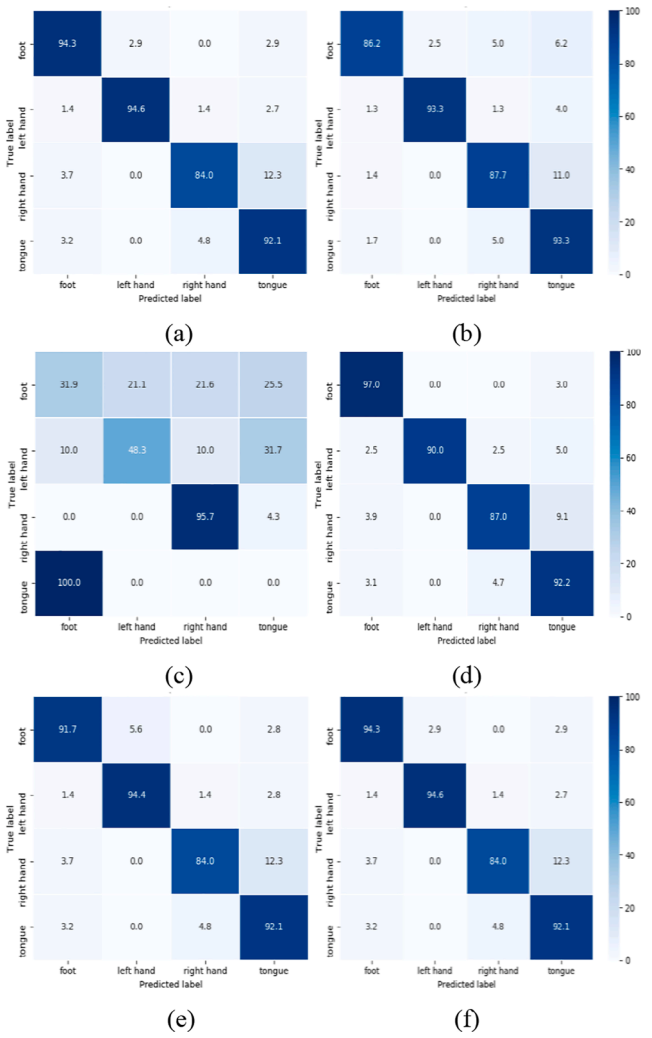


Fig. 7. Confusion matrixes of the SPSCNN evaluated with different frequency bands filtered. (a) Unfiltered (Acc = 90.97%); (b) α band filtered (Acc = 89.93%); (c) β band filtered (Acc = 40.28%); (d) γ1 band filtered (Acc = 91.32%); (e) γ2 band filtered (Acc = 90.28%); (f) γ3 band filtered (Acc = 90.97%).

The results in Table 8 show the overall performance of model under different network methods across all subjects. The average training accuracy of our proposed method reaches 77.24%, and the average loss as low as 0.53, which shows the best performance compared with other networks. It indicates that our network can extract more abundant and useful common features between multiple different subjects as much as possible. Although DeepNet has a deeper architecture, the results are less than ideal. It probably because the common features of multiple subjects have not been found in the deep excavation of the characteristics of the subjects, and such a deep structure is not suitable for EEG signals that vary widely in individuals. The layer of EEGNet is too few, so

Table 8
The performance on training model M_0 and cross-subject testing classification.

Network	DeepNet	EEGNet	ShallowNet	SPCNN
Average training accuracy/%	53.41	69.67	77.24	78.21
Average loss	1.12	0.81	0.57	0.53
Average cross-subject testing accuracy/%	34.65	42.81	44.24	40.19

we can only get the insufficient common features, resulting in a lower accuracy rate than that of SPCNN 8.54% and a higher loss value that of SPCNN about 50%. ShallowNet performs slightly worse than SPCNN. As for the average cross-subject testing accuracy, the performance of ShallowNet and EEGNet is slightly better than that of SPCNN, because these two networks have relatively fewer layers and only extract the typical characteristics of signals, which has little influence on different subjects. The performance of DeepNet is poor mainly because the loss during the model training is too high to extract common features and not enough basic features. In contrast, the performance of SPCNN is similar to that of the two shallow networks. The freeze and train strategy is combined to improve the universality of our network among different subjects. The training data of specific subject is fused with the original data, and input into the network for retraining, we can get the average testing classification performance of model M for all subjects in SPCNN shown as Table 8.

In order to get the best scenario of the freeze and train strategy, the performance of model M is shown in Table 9 in the case of different fused data volumes. The experimental results show that the average accuracy reaches 48.51% when the new data volume is 32, which is 8.32% higher than that without using the freeze and train strategy. Furthermore, the new data volume increases to 64 and 96 respectively, and the average accuracy raises to 51.21% and 54.74% accordingly. By this trend, when the data volume is 128, the accuracy slightly increases by about 2%. Compared with the previous results, the method of transfer learning not only makes it possible to use the model across subjects, but also improves the classification accuracy by 10%. We also find the parameters are reduced from the original 133,894 to 62,694 through the freeze and train strategy. The number of parameters are reduced by more than 50%, which not only saves the training time of the model, but also reduces the difficulty of experimental data collection. Specific difficulties include (a) the state of nervous exhaustion of subjects and the accuracy of data collection caused by the long periods of concentration in the experiment and (b) the labeling workload. It also improves the accuracy of the cross-subject model evaluating, increases the universality of the model, and enables the network to be verified in all aspects. From the results, these prove that the SPCNN model can reach a satisfied accuracy level by a certain freeze tactic. This is used to overcome the intra-subject difference and save modelling time consumption even a model already built within subject, which would be used to test another subject.

3.3.3. Different frequency domain and time domain scenarios

3.3.3.1. Frequency domain difference.. In the above experiment, in order to obtain sufficient data and rich MI-EEG signal characteristics, we select 0.5–100 Hz data segment. Next, we filter out the part of frequency band and retain different frequency bands to explore the instructive frequency domain in classification results. The whole band is divided into the core band (α band), the secondary band (β band) and the rest band (γ band) according to their importance. To avoid the effect of large filter frequency on the experimental results, we set the filter bands to α(8–13 Hz), β(14–30 Hz), γ1(31–50 Hz), γ2 (51–70 Hz) and γ3 (71–100 Hz) respectively. Taking the fifth model established by the 10-fold cross method of Subject 3 as an example, confusion matrices are drawn under different conditions. The experimental results are shown in Fig. 7.

It can be seen from Fig. 7 that the average classification accuracy of the four kinds of MI-EEG signals in the testing set is 90.97% when the whole band is retained. When α band is filtered out, the accuracy slightly decreased to 89.83%, with a significant 8.1% reduction in the accuracy of MI classification for the foot. When β band is filtered out, the average classification accuracy drops to 40.28%, making it impossible to classify the tongue correctly. Conversely, when three unrelated bands are filtered, the average classification accuracy varies little compared with the original accuracy. Further enlarges the band filter range, divides the signal into useful and unnecessary bands according to the band, and draws the confusion matrix as shown in Fig. 8.

Fig. 8 shows that when the unnecessary band is filtered out, the average classification accuracy reaches 91.32%, which is similar to the average classification accuracy before filtering. When the chief band is completely filtered out, the average classification accuracy is as low as 34.38%, which is lower than that of β band filtered alone. The experimental results are consistent with the law of MI-EEG signals, and the validity of the network is verified in frequency domain.

3.3.3.2. Time domain difference.. In the experiment, we set 4 s of the cue appearance and the whole MI process as the experimental data, and further divide it into the first 2 s and the last 2 s. In order to avoid the influence of the amount of data on the experimental results, each part of the data is repeated once to ensure that the total sampling point is 1000. The experimental results are shown in Table 10.

Table 10 shows the results of the first trial in 10-fold cross method conducted by subject 1. The experimental results show that the model accuracy of the complete data is 96.88%, and the model accuracy of 0–2 s segment is 87.5%, and that of 2–4 s segment is 71.88%. It illustrates that the data segment within 0–2 s contains more abundant classification information, which is consistent with the general law of MI-EEG signal. At the same time, the data within 2–4 s also plays a certain auxiliary role in classification. It is proved that the network is effective in time domain.

Table 9
The cross-subject testing classification performance in final model.

Freeze layer	Training Parameters	New data	Average accuracy/%
0	133,894	0	40.19
10	62,694	32	48.51
10	62,694	64	51.21
10	62,694	96	54.74
10	62,694	128	56.02

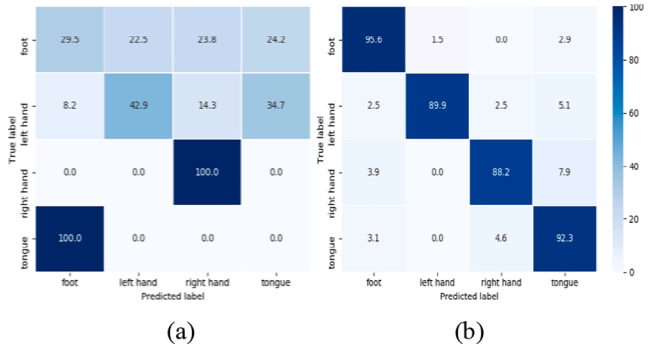
4. Discussions

4.1. Advancement of SPCNN

4.1.1. Advancement of our innovative structure

In this paper, through the parallel feature extraction layer, we use different scales to extract deep-level space-temporal features, and get more abundant signal features, which has a positive effect on the further classification of subsequent networks. As an innovative structure of the network, the experiments are carried out before and after removing parallel parts to verify its advancement. The experimental results are shown in Table 11.

Table 11 shows the results of the first trial in 10-fold cross method conducted by subject 1. The experimental results show that the accuracy



(a) (b)

Fig. 8. Confusion matrixes of the SPCNN evaluated with different frequency bands filtered (wider range). (a) The chief band filtered (Acc = 40.28%); (b) The unnecessary band filtered (Acc = 91.32%).

Table 10
Time domain validity test.

Time domain interval	0 – 2 s	2 – 4 s	0 – 4 s
Accuracy/%	87.50	71.88	96.88
Loss	0.35	0.81	0.33

Table 11
Parallel feature extraction layer testing.

parallel feature extraction layer	Yes	No
Accuracy/%	96.88	84.38
Loss	0.33	0.63

of the model is reduced by 12.9% when the parallel structure is removed, and the loss of the model is almost twice as much as the original. It illustrates that the deep-level features of parallel feature extraction layer are very important for the final classification result, and also shows the advancement of the structure.

4.1.2. Advancement of our overall structure

In the early stage of network construction, the parameters of parallel feature extraction layer have been adjusted, and the current parameter setting is the best. Therefore, this part mainly optimizes the parameters of serial feature extraction layer, including the number optimization of temporal filter and spatial filter. The parameter adjustment and its effect are shown in Table 12.

It can be concluded from Table 11 that when the number of spatial filters is twice that of temporal filters, the highest accuracy of the model reaches 83.56%, and the highest evaluating accuracy of the model reaches 72.3% when the number of both kinds of filters is set as 40. This proves that when the number of both filters is set to 40, the model is generally applicable with fewer parameters. However, in general, the overall structure we propose is superior to other networks in terms of different number of filters, which proves the advancement of our method.

4.2. Theoretical consistency of SPCNN

To demonstrate the effectiveness of the network, we verify it in the frequency domain and time domain. The experimental results show that the law of the network in different frequency domain and time domain is consistent with the theory.

4.2.1. Frequency domain difference.

Studies have shown that the different states of the human brain mainly correspond to four types of brain waves: δ(0.5–3 Hz), θ(4–8 Hz), α(8–13 Hz), and β(14–30 Hz). For different types of MI-EEG signals, their β bands have the distinctive differences, so β band is used as the core for classification, α band is also important, and the other bands with the less influence are called the secondary band here.

4.2.2. Time domain difference.

Early research on EEG signals shows that the important classification information time period of MI-EEG signals is within the first 2 s of imagination. With the progress of the experiment, the sensitivity of classification results to the latter part of the data also decreased. Therefore, we divide the whole data into 1–2 s segment and 2–4 s segment. Experimental results show that the classification accuracy of the first 2 s is higher than that of the last 2 s, which is consistent with the theory. However, the classification accuracy of the whole data segment is the highest, which indicated that the insensitive data segment still has a certain auxiliary effect on the classification. More effective MI-EEG signal data is conducive to classification.

Table 12
Parameter adjustment and effect.

TimeConv	20 (1,13)	20 (1,13)	40 (1,13)	40 (1,13)
SpaceConv	20 (22,1)	40 (22,1)	40 (22,1)	80 (22,1)
Model accuracy/%	82.09	81.80	82.74	83.56
Model loss	0.48	0.48	0.48	0.47
Parameter	31,134	53,694	71,654	134,374
Testing accuracy/%	71.78	72.08	72.30	72.13

4.3. Necessity of transfer learning applied across subject

Our paper realizes the cross-subject application of the model through the freeze and train strategy in transfer learning, and improves the accuracy of cross-subject testing to some extent. It is necessary in the case of insufficient MI-EEG data and low experimental acquisition efficiency. By establishing a basic model of MI-EEG signal, different subjects are subsequently integrated into some of their own data for model retraining, which not only increased the universality of the model, but also reduced the data acquisition time and required data.

4.4. Limitation and prospect

At present, there are still many limitations in this field. Two of the most typical problems are: (a) the amount of data is insufficient, this could be solved by importing more data, like high gamma dataset for MI in [37]; (b) the network architecture is lack of generality and explain ability. Due to the specification limitation of experimental equipment and human subject fatigue in the process of MI, the amount of MI-EEG signal needs to be expanded. In this paper, although the sampling points of the whole process from the beginning of the clue to the end of the imagination are selected as far as possible, but compared with other deep learning-related applications, such as image processing, the amount of data is still too small. Another common issue is the stationarity preservation of MI EEG data. We may take the gradient descent on an orthogonal manifold as a meaningful reference to enforce the stationarity of EEG [38–39], which could improve the accuracy. Also, there are some points we can improve even though we made the frequency band dependent model performance evaluation. The fixed hyperparameter based modern models could outperform that utilize variable hyperparameter networks for each subject. Also, the based on their similar performance results with simple model, our SPCNN may could be optimized in further to shrink efficiently.

In the future work, we will study some data augmentation methods to enrich the data set. On the one hand, Gaussian noise with different distributions can be added, but considering the non-stationary characteristics of EEG signals, it is further considered to add noise to the amplitudes of EEG-signals' spectral image to achieve data enhancement [40,41]. On the other hand, the empirical mode decomposition method can be used to decompose the original signal into multiple components, and then the components can be recombined to form a new EEG signal sample [42]. And the method of generating adversarial network can also be used to generate new EEG signal samples without supervision and realize sample expansion more intelligently [43].

At the same time, the EEG signals of different individuals are quite different, and the characteristics of different brain power sources are also inconsistent. The network which takes a lot of time to train couldn't adapt to various situations. In this paper, we adopt a freeze and train strategy in transfer learning. By pre-training a basic model, we freeze the serial module of SPCNN, and then retrain the model after merging some new data, which not only saves the training time, but also reduces the needed parameters, solving the above problems to some extent. There are several other methods of transfer learning which are expected to be applied to cross-subject testing of EEG signals in the future. The model can automatically adapt to the data characteristics of different subjects by domain adaptation method [44]. Subspace learning can map the data in the source domain and target domain to another space for geometric

or statistical operations, so as to discover the hidden structure and features of the data and realize the prediction of the target domain label [45]. However, we need to clarify that the transfer learning in our paper is pretty basic. The transfer policy should be adaptive to overcome the inter-subject difference. The policy could give clear learning target and constraints for model optimization. Currently, we just focus on the MI recognition accuracy, but the substantial conditions and goals are not well mentioned. Also, the model transfer tactic needs to be improved by introducing more advanced algorithm. The freeze and retrain skill is not parameter-wise tuneable. And subject difference should be mapped into the source and target domain adaptation representation, thus improving the learning capability efficiently.

In addition, more analysis on some other dataset should conducted. It can improve our method generalizability in different scenarios based on more and more data, therefore facilitates our future work related with human brain to machine interface application. We are undertaking our own customized experimental work now. And this can improve our method capability. Moreover, with the rapid development of BCI, people are more sensitive to the real-time performance of BCI, and the requirements for the accuracy of EEG classification are higher. In the future, we will also consider to improve the accuracy of classification based on deep learning method and apply it to online BCI, so as to realize more intelligent and rich human–computer interaction system.

5. Conclusion

In this work we propose a convolutional neural network with serial-parallel structure for MI-EEG signals decoding. Our SPCNN adopts two different structures and extracts the time–frequency and spatial features of MI-EEG signals comprehensively with different scales. Combined with transfer learning, the pre-trained universal model can be quickly retrained and the final training model obtained can be applied across subjects. Compared with other mainstream networks, the network in this paper is proven to be effective. The experimental results show that our work consistent with the theory, reduce the complexity of signal processing, and increase the model generality. Our model and strategy show great potential for the future application of BCI area.

CRediT authorship contribution statement

Xuefei Zhao: . **Dong Liu:** . **Li Ma:** Conceptualization, Methodology. **Quan Liu:** . **Kun Chen:** . **Shane Xie:** . **Qingsong Ai:** .

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

This work is supported by National Science Foundation of China under Grant: 52075398. And also funded by Wuhan Science and Technology Plan Application of Fundamental Frontier Special Project: 2020020601012220.

References

[1] Y.T. Wu, T.H. Huang, C.Y. Lin, J.T. Sheng, P.S. Wang, Classification of EEG Motor Imagery Using Support Vector Machine and Convolutional Neural Network. 2018 International Automatic Control Conference (CACS), 2018.
[2] D.J. McFarland, J.R. Wolpaw, Brain–computer interface use is a skill that user and system acquire together, Plos Biology 16 (7) (2018) e2006719, https://doi.org/10.1371/journal.pbio.2006719.
[3] G.-Z. Yang, J. Bellingham, P.E. Dupont, P. Fischer, L. Floridi, R. Full, N. Jacobstein, V. Kumar, M. McNutt, R. Merrifield, B.J. Nelson, B. Scassellati, M. Taddeo, R. Taylor, M. Veloso, Z.L. Wang, R. Wood, The grand challenges of Science Robotics. Science, Robotics 3 (14) (2018), https://doi.org/10.1126/scirobotics.aar7650.
[4] T.O. Zander, L.M. Andreessen, A. Berg, M. Bleuel, J. Pawlitzki, L. Zawallich, L. R. Krol, K. Gramann, Evaluation of a Dry EEG System for Application of Passive Brain-Computer Interfaces in Autonomous Driving, Frontiers in Human Neuroscience 11 (2017), https://doi.org/10.3389/fnhum.2017.00078.
[5] E.S. Fan, S.Y. Lu, Q. Zhu, Y.L. Geng, Study on BCI Smart House System Based on Motor Imagination, Building Electricity 37 (02) (2018) 51–54.
[6] Q. Gao, X. Zhao, X. Yu, Y.u. Song, Z. Wang, Controlling of smart home system based on brain-computer interface, Technology and health care: official journal of the European Society for Engineering and Medicine 26 (5) (2018) 769–783.
[7] M. Mahmood, D. Mzurikwao, Y.-S. Kim, Y. Lee, S. Mishra, R. Herbert, A. Duarte, C. S. Ang, W.-H. Yeo, Fully portable and wireless universal brain–machine interfaces enabled by flexible scalp electronics and deep learning algorithm, Nature Machine Intelligence 1 (9) (2019) 412–422.
[8] H.W. Li, X.G. Chen, Brain-computer interface controlled robotic arm system based on high-level control strategy, Beijing Biomedical Engineering 38 (2019) 36–41.
[9] A.M. Chiarelli, P. Croce, A. Merla, F. Zappasodi, Deep learning for hybrid EEG-fNIRS brain-computer interface: application to motor imagery classification, Journal of neural engineering 15 (3) (2018) 036028, https://doi.org/10.1088/1741-2552/aaaf82.
[10] P. Croce, A. Basti, L. Marzetti, F. Zappasodi, C.D. Gratta, EEG-fMRI bayesian framework for neural activity estimation: a simulation study, Journal of neural engineering 13 (6) (2016) 066017, https://doi.org/10.1088/1741-2560/13/6/066017.
[11] N. Gao, W.W. Zhai, J.X. Lu, L.Y. Wu, S.Y. Lu, The research on intelligent wheelchair based on brain computer interface of steady state visual evoked potential, Journal of Biomedical Engineering Research 37 (01) (2018) 6–10.
[12] G. Townsend, B.K. LaPallo, C.B. Boulay, D.J. Krusienski, G.E. Frye, C.K. Hauser, N. E. Schwartz, T.M. Vaughan, J.R. Wolpaw, E.W. Sellers, A novel P300-based brain–computer interface stimulus presentation paradigm: Moving beyond rows and columns, Clinical Neurophysiology 121 (7) (2010) 1109–1120.
[13] G. Pfurtscheller, F.H. Lopes da Silva, Event-related EEG/MEG synchronization and desynchronization: basic principles, Clinical Neurophysiology 110 (11) (1999) 1842–1857.
[14] Y. Yang, M. Zeng, J.S. Cheng, A New Time-frequency Analysis Method-the Local Characteristic-scale Decomposition, Journal of Hunan University (Natural Sciences) 39 (2012) 35–39.
[15] J. Zhou, G.Y. Yang, T. Xu, Classification of multi-class motor imagery EEG data based on spatial frequency and time-series information, Chinese Journal of Medical Physics 36 (01) (2019) 87–93.
[16] Y. Yang, S. Chevallier, J. Wiart, I. Bloch, Subject-specific time-frequency selection for multi-class motor imagery-based BCIs using few Laplacian EEG channels, Biomedical Signal Processing & Control 38 (2017) 302–311.
[17] S. Chaudhary, S. Taran, V. Bajaj, S. Siuly, A flexible analytic wavelet transform based approach for motor-imagery tasks classification in BCI applications, Computer Methods and Programs in Biomedicine 187 (2020) 105325, https://doi.org/10.1016/j.cmpb.2020.105325.
[18] S. Lemm, B. Blankertz, G. Curio, K.-R. Muller, Spatio-spectral filters for improving the classification of single trial EEG, IEEE transactions on bio-medical engineering 52 (9) (2005) 1541–1548.
[19] G. Dornhege, B. Blankertz, M. Krauledat, F. Losch, G. Curio, K.-R. Muller, Combined optimization of spatial and temporal filters for improving brain-computer interfacing, IEEE transactions on bio-medical engineering 53 (11) (2006) 2274–2281.
[20] Novi Q, Guan C, Dat TH, Xue P. Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface. Neural Engineering, 2007. CNE '07. 3rd International IEEE/EMBS Conference on. IEEE 2007; 204-207.
[21] K.K. Ang, Z.Y. Chin, H. Zhang, C. Guan, Filter Bank Common Spatial Pattern (FBCSP) in Brain-Computer Interface, IEEE International Joint Conference on Neural Networks. IEEE (2008) 2390–2397.
[22] D. Cai, X.F. He, J.W. Han, SRDA: An Efficient Algorithm for Large-Scale Discriminant Analysis, IEEE Trans on Knowledge and Data Engineering 20 (2008) 1–12.
[23] S. Li, W. Zhou, Q.i. Yuan, S. Geng, D. Cai, Feature extraction and recognition of ictal EEG using EMD and SVM, Computers in Biology & Medicine 43 (7) (2013) 807–816.
[24] A. Al-Saegh, S.A. Dawwd, J.M. Abdul-Jabbar, Deep learning for motor imagery EEG-based classification: A review, Biomedical Signal Processing and Control 63 (2021) 102172, https://doi.org/10.1016/j.bspc.2020.102172.
[25] P. Bashivan, I. Rish, M. Yeasin, N. Codella, Learning Representations from EEG with Deep Recurrent-Convolutional Neural Networks. Computer, Science (2015).
[26] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, F. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Human Brain Mapping (2017).
[27] V.J. Lawhern, A.J. Solon, N.R. Waytowich, S.M. Gordon, C.P. Hung, B.J. Lance, EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces, Journal of neural engineering 15 (5) (2018) 056013, https://doi.org/10.1088/1741-2552/aace8c.
[28] P. Croce, F. Zappasodi, L. Marzetti, et al., Deep Convolutional Neural Networks for Feature-Less Automatic Classification of Independent Components in Multi-Channel lectrophysiological Brain Recordings[J], IEEE Transactions on Biomedical Engineering (2018) 2372–2380.
[29] J. León, J.J. Escobar, A. Ortiz, J. Ortega, J. González, P. Martín-Smith, J.Q. Gan, M. Damas, R. Stoean, Deep learning for EEG-based Motor Imagery classification:

Accuracy-cost trade-off[J], PLoS ONE 15 (6) (2020) e0234178, https://doi.org/10.1371/journal.pone.0234178.
[30] Leeb R, Brunner C, Muller-Putz GR, Schlogl A. BCI Competition 2008 - Graz data set B. putz 2008.
[31] Zhang JW, Feng Y, Li W. Research on wavelet denoising method based on an improved threshold function. Electronic Design Engineering 2017; 025(009): 137-140,144.
[32] Liu R, Xu M, Zhang YZ, Li DH, Liu MM, Deng ZK, Jia RS. EOG detection and removal method for single channel electroencephalogram signal. Journal of Computer Applications 2017; 37(S1): 226-230+265.
[33] H. Wu, Y.i. Niu, F.u. Li, Y. Li, B. Fu, G. Shi, M. Dong, A parallel multiscale filter bank convolutional neural networks for motor imagery eeg classification, Frontiers in Neuroscience 13 (2019), https://doi.org/10.3389/fnins.2019.01275.
[34] F. Gerhard, L. Szegletes, Spline- and wavelet-based models of neural activity in response to natural visual stimulation, Annual International Conference of the IEEE Engineering in Medicine and Biology Society 2012 (2012) 4611–4614.
[35] D. Kharbanda, G.K. Verma, Multi-level 3D Wavelet Analysis: Application to Brain Tumor Classification, International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE) 2016 (2016) 379–384, https://doi.org/10.1109/ICMETE.2016.121.
[36] D. Kingma, B.J. Adam, A Method for Stochastic Optimization. Computer, Science (2014).
[37] S.J. Pan, Q. Yang, A Survey on Transfer Learning, IEEE Transactions on Knowledge and Data Engineering 22 (10) (2010) 1345–1359.
[38] R.T. Schirrmeister, J.T. Springenberg, L.D.J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, T. Ball, Deep learning with convolutional neural networks for EEG decoding and visualization, Hum. Brain Mapp. 38 (11) (2017) 5391–5420.
[39] S. Kumar, T.K. Reddy, V. Arora, L. Behera, Formulating Divergence Framework for Multiclass Motor Imagery EEG Brain Computer Interface, in: ICASSP 2020–2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2020, pp. 1344–1348.
[40] Y.K. Musallam, N.I. AlFassam, G. Muhammad, S.U. Amin, M. Alsulaiman, W. Abdul, H. Altaheri, M.A. Bencherif, M. Algabri, Electroencephalography-based motor imagery classification using temporal convolutional network fusion, Biomedical Signal Processing and Control 69 (2021) 102826, https://doi.org/10.1016/j.bspc.2021.102826.
[41] H. Cecotti, A.R. Marathe, A.J. Ries, Optimization of Single-Trial Detection of Event-Related Potentials Through Artificial Trials, IEEE Transactions on Biomedical Engineering 62 (9) (2015) 2170–2176.
[42] Y. Li, X.-R. Zhang, B. Zhang, M.-Y. Lei, W.-G. Cui, Y.-Z. Guo, A Channel-Projection Mixed-Scale Convolutional Neural Network for Motor Imagery EEG Decoding, IEEE Transactions on Neural Systems and Rehabilitation Engineering 27 (6) (2019) 1170–1180.
[43] Z. Zhang, F. Duan, J. Sole-Casals, J. Dinares-Ferran, A. Cichocki, Z. Yang, Z. Sun, A Novel Deep Learning Approach with Data Augmentation to Classify Motor Imagery Signals, IEEE Access 7 (2019) 15945–15954.
[44] Ge RX, Hu JZ. Research of multi-class motor imagery EEG classification based on deep learning framework. Journal of Jiangsu University of Science and Technology (Natural Science Edition) 2019; 33(4).
[45] Z. Wan, R. Yang, M. Huang, N. Zeng, X. Liu, A review on transfer learning in EEG signal analysis, Neurocomputing 421 (2021) 1–14.