Motor Imagery EEG Classification with Self-attention-based Convolutional Neural Network

2022 7th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS) | 978-1-6654-8229-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIIBMS55689.2022.9971698

Rui Zhang#
[1]Shandong Institute of Advanced Technology
Chinese Academy of Sciences
Jinan, China
[2]School of Microelectronics
Shandong University
Jinan, China
xiaozhanghaha@126.com

Guoyang Liu#
School of Microelectronics
Shandong University
Jinan, China
virter1995@outlook.com

Mingxin Su
School of Microelectronics
Shandong University
Jinan, China

Ningling Zhang, Cai Chen, Danyang Lv
Shandong Institute of Advanced Technology
Chinese Academy of Sciences
Jinan, China

Fulai Peng*
Shandong Institute of Advanced Technology
Chinese Academy of Sciences
Jinan, China
fl.peng@sdiat.ac.cn

Weidong Zhou*
School of Microelectronics
Shandong University
Jinan, China
wdzhou@sdu.edu.cn

***Abstract*—Motor Imagery-based Brain-Computer Interfaces have been widely utilized in neuro-rehabilitation. Motor Imagery electroencephalogram (MI-EEG) refers to the EEG signals that people imagine their body moving without real action. People who have motor disorders can control the external devices through electroencephalogram (EEG)-decoding. However, there are still a variety of challenges in decoding due to the complexity and non-stationarity of EEG. How to improve the accuracy and robustness of EEG-decoding remains a key question to be studied. In this work, a self-attention-based convolutional neural network (CNN) combined with Frequency-Time Band Common Spatial Pattern (FTBCSP) is first introduced for the four-class MI-EEG classification. Self-attention-based CNN is employed on raw data to obtain the channel weights and intensify the spatial information. Common Spatial Pattern (CSP), an algorithm that is widely used in MI-EEG decoding, can extract discriminative features between two classes. Features after processing by the CSP algorithm are combined with the above spatial information to accomplish classifying. We validate this method on the publicly available multiclass MI datasets and yield a mean accuracy of 78.12% which performs better than other traditional methods. It proves that the proposed approach makes full use of the temporal and spatial information of EEG and acquires outstanding classification performance on public datasets.**

Keywords—Motor Imagery, Brain-Computer Interface, Electroencephalgram (EEG), attention, Frequency-Time Band Common Spatial Pattern (FTBCSP)

* Corresponding author
# First authors contributed equally

978-1-6654-8229-5/22/$31.00 ©2022 IEEE

I. INTRODUCTION (*HEADING 1*)

Brain-Computer Interface (BCI) is an important tool that allows people to communicate with or control external electronic devices. Moreover, motor imagery-based BCI (MI-BCI) enables individuals with a motor disability to link with the outer environment and then improve their life quality. To realize these functions, researchers need to obtain the motor intention from subjects' scalp electroencephalogram (EEG) and classify these intentions accurately.

Pfurtscheller et al. [1] [2] discovered that the α and β frequency bands from different channel EEG would oscillate regularly when imaging hand moving. Then they defined this phenomenon as event-related desynchronization (ERD) and event-related synchronization (ERS). At present, this phenomenon has been broadly used in the analysis of Motor Imagery (MI) classification tasks and achieves excellent performance. Motor Imagery EEG (MI-EEG) refers to the EEG signals that people imagine their body moving without real action.

Accordingly, the MI-EEG classification methods can be divided into two main categories, namely traditional machine learning methods, and deep learning methods. As a classic traditional machine learning method, Common Spatial Pattern (CSP) [3] [4] was proved to be effective in the field of MI. It is a feature extraction algorithm that can extract the spatial components from multi-channel EEG signals and acquire the most remarkable difference between discriminative classes. Then Sub-band Common Spatial Pattern (SBCSP) [5], as well as Filter Bank Common Spatial Pattern (FBCSP) [6], was

developed from CSP and overcame the frequency limitation of CSP. SBCSP utilized a filter bank to decompose the EEG signal into multiple sub-bands and achieved the classification results through sub-band scores. And, the FBCSP added feature selection procedure to yield the most discriminative features from these filter banks after filtering. The FBCSP helps improve the situation of manually selecting the frequency band appropriately in MI-EEG classification.

Due to the advantage of realizing the end-to-end classification without manual participation, deep learning started to be widely used in MI-EEG classification. It helps to decrease some deviation of traditional methods depending on hand-crafted features and enhance the performance. For instance, Deep ConNet was used for end-to-end EEG decoding and EEG features visualization by Schirrmeister et al. [7]. Besides, EEG-Net [8] encapsulates several well-known EEG feature extraction procedures, such as the construction of optimal spatial filter and filter bank. Moreover, it achieves better classification performance in BCI than other deep learning methods and has excellent generalization performance.

The attention model was first employed in the field of translation. With the development of attention, it becomes an essential part of the convolutional neural network (CNN) and is gradually used in the area of natural language processing, EEG decoding, and computer vision. Lately, some MI-classification studies adopted CNN combined with the attention model too. For example, Liu et al. [9] presented the spatial-temporal self-attention CNN-based method to obtain accurate intention from the EEG. Ma et al. [10] proposed time-distributed attention combined with long short-term memory (LSTM) approach, and raw data was split into several time segments and then these segments were operated by class attention as well as band attention. Squeeze-and-excitation attention was also applied in MI classification [11]. These promising results achieved on attention-based approaches demonstrate the good prospects of attention mechanism in MI.

The primary objective of this work is to improve the accuracy of four-class MI classification task (left hand, right hand, feet, and tongue). We combine the frequency-time band common spatial pattern (FTBCSP) feature with self-attention-based CNN. The proposed model was tested on two public BCI datasets and compared with some classic methods. The remainder of this paper is organized as follows: Section II describes the datasets and Section III depicts the details of our presented methods. In Section IV, the results of comparison with other strategies in two BCI datasets are shown. The conclusions and future work are shown in Section V.

II. MI-BCI DATABASE

In this work, a publicly available dataset, namely BCI Competition IV IIa (BCI IV IIa), is adopted to evaluate the presented method. BCI IV IIa is sampled at 250 Hz and consisted of four MI tasks (left hand, right hand, feet, and tongue). Furthermore, the paradigm of this dataset is cue-based.

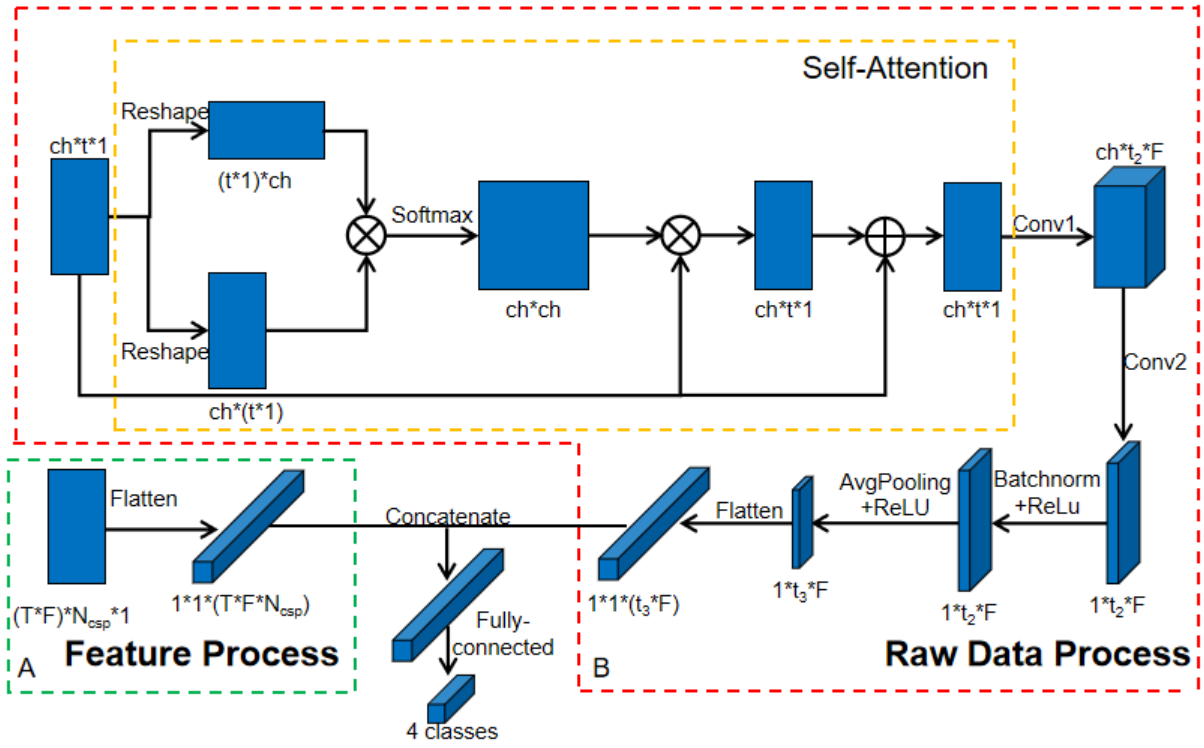

Fig. 1. The framework of the proposed model. (A) Feature process is the process of FTBCSP feature that raw data was firstly divided into frequency and temporal bands then processed by the CSP algorithm. Among the process, T=4 represents the number of temporal bands, F=11 represents the number of frequency bands, and Ncsp=24 represents the number of features from each band. (B) Raw data process includes the self-attention model and the CNN model. Ch=22 is the number of EEG channels, t=1000 denotes the time points of 4s data, F=40 is the number of maps, and t2=976 as well as t3=61 is the result of the convolution operation.

TABLE I. DETAILS OF PARAMETERS IN RAW DATA PROCESS.

details of Raw Data Process								
Layer	Self-attention	Conv1	Conv2	BatchNorm	ReLU	AvgPooling	ReLU	Flatten
Input	(22, 1000, 1)	(22, 1000, 1)	(22,976,40)	(1,976,40)	(1,976,40)	(1,976,40)	(1,61,40)	(1,61,40)
Output	(22, 1000, 1)	(22, 976, 40)	(1,976,40)	(1,976,40)	(1,976,40)	(1,61,40)	(1,61,40)	(1,1,2440)
Feature maps	1	40	40	40	40	40	40	2440
Kernel	-	(1,25)	(ch,1)	-	-	(1,75)	-	-
Stride	-	(1,1)	(1,1)	-	-	(1,15)	-	-

BCI VI IIa dataset comprises EEG data from 9 subjects with 25 channels. Each subject has two sessions and each session has 288 trials (72 trials for each of the four classes). Each channel is bandpass-filtered between 0.5 Hz and 100Hz. In this study, we only consider the EEG data from the first 22 channels because the last 3 channels contain Electrooculogram (EOG) signals. Besides, a Butterworth bandpass filter with 0.5 Hz-32 Hz and a time window of 4s length is applied to extract 3s-7s data from every trial. We separate the two sessions into a training set and a test set, both of them have the same amount of 288 trials.

III. METHODS

A. Overview

As Fig. 1 shows, the whole structure of the model can be roughly divided into two parts. The first part is the process of raw data, which includes a self-attention-based CNN model. Self-attention model is used to yield the channel weight of the raw EEG data and then the data is sent into two convolutional layers for further processing. The first one is the temporal convolutional layer and the other is the spatial convolutional layer. Accordingly, a series of pooling and nonlinear activation operations follows the convolution. The feature maps after being processed by the self-attention-based CNN are flattened. The detailed parameters of the network, such as the kernel size and the size of the stride, are shown in Table I. The second part is the CSP feature extraction process. After dividing the raw data into frequency and temporal bands, the raw data was calculated by the CSP algorithm. The frequency-time band common spatial pattern (FTBCSP) feature works as another input of the whole net model. In the end, these two feature maps are concatenated together and classified into four classes.

B. Self-Attention based CNN

The characteristics of the EEG signals varies from person to person. For example, the activated regions of the brain or the channels that respond to MI are generally various between individuals. Therefore, it is essential to find accurate channels when classifying. We employ the self-attention model before CNN to obtain the EEG channel information on various subjects.

As the raw data process part shown in Fig. 1, the input data sized $22\times1125\times1$ (take the BCI IV IIa as an example) is firstly reshaped into two matrixes sized $22\times(1125\times1)$ and $(1125\times1)\times22$ respectively. Among it, 22 represents the number of channels, 1000 refers to the time points of 4s data, and 1 denotes the number of maps. Then the two matrixes are multiplied and applied softmax function in order to acquire a channel weight matrix with a size of 22×22. After that, the map that the channel weight matrix multiplied with the raw data matrix is added to the raw data map.

Subsequently, the feature map with a size of $22\times1125\times1$ is put into the convolutional neural network. The detailed CNN framework is illustrated in Fig. 2. The first convolutional layer (Conv1) with 40 filters of size (1,25) is leveraged to gain the temporal features. The second convolutional layer (Conv2) utilize 40 filters with a size of (22,1) to capture the spatial features. Then the batch normalization is used to accelerate training process. The non-linear activation function, ReLU, is applied to improve the ability that the model learns and simulates complex data. The average pooling layer of size (1,75) can complete downsampling and reduce the redundancy of data.

C. Frequency-Time Band Common Spatial Pattern (FTBCSP)

FTBCSP means that the raw data was divided into different frequency and temporal bands before being calculated by the CSP algorithm. For the feature process, the first step is to divide temporal bands and frequency bands. As depicted in Fig. 2, the frequency band between 4-36 Hz is split into 11 subbands by the Butterworth filter. The bandwidth of the first seven subbands is 8Hz with 50% overlapping. The length of the eighth to the tenth bands is 16 Hz and they are also 50% overlapped. The bandwidth of the last one is 32 Hz. One time segment is shown in Fig. 3. Since only four seconds of the EEG data is employed and the sampling rate is 250 Hz, there are 1000 points in each trial. There are 3 temporal bands with a length of 500 points and 50% overlapping as well as a temporal band with a length of 1000 points. As a consequence, 44 frequency-time bands (11 frequency bands $\times$ 4 temporal bands) can be obtained from each trial after splitting.

After splitting, spatial filtering using the One-Versus-One (OVO) CSP algorithm is applied in each band. It can extract the spatial components of multi-channel EEG signals by finding the optimal spatial projection of two kinds of tasks.

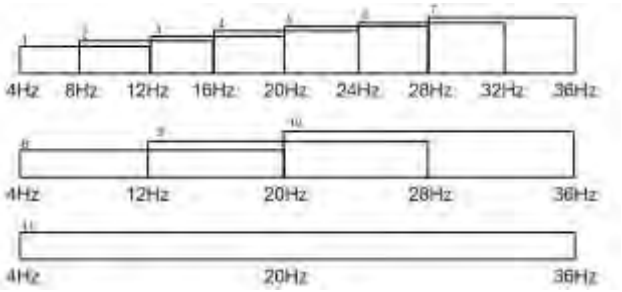

Fig. 2. Frequency band. There are 11 frequency bands in total. Seven subbands have the width of 8Hz overlapped with 50% and the length of three bands is 16Hz. The band width of the last frequency band is 32Hz (4-36Hz).

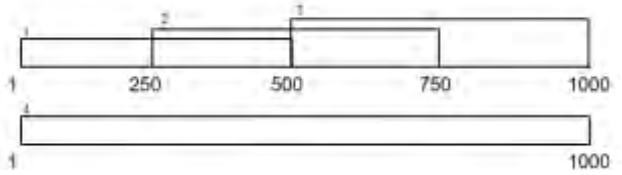

Fig. 3. Time band. Three temporal bands with 500-time points and 50% overlapping and one temporal band with all 1000-time points.

OVO CSP means that we only take two classes in one CSP task. Taking this strategy, six spatial filters will be obtained owing to combining every two different classes from four classes.

The first step of CSP is to calculate the normalized covariance matrixes R_1 and R_2 from two classes respectively.

$$R_i = \frac{X_{1,i}X_{2,i}^T}{\text{trace}(X_{1,i}, X_{2,i})} \quad (1)$$

($X_{1,i}$ represents the i trial of the first class)

Then the mixed space covariance matrix R is calculated as follows:

$$\overline{R_c} = \frac{1}{k}\sum_{i=1}^{k} R_{c,i} \quad (2)$$

(c = 1 or 2, represents different classes)

$$R = \frac{\overline{R_1}+\overline{R_2}}{2} \quad (3)$$

Next, whitening matrix $P = \frac{U_c^T}{\sqrt{\lambda_C}}$ (U_c is the eigen matrix of R and λ_C is the eigenvector of R) is then applied to get the spatial filter with the purpose of removing correlation and reducing redundancy.

$$S_c = PR_cP^T \quad (4)$$

In the end, the spatial filter is projected to the time series of the ith band for obtaining feature vector Z^i. We employ the first two m as well as the last two m rows of the eigenvectors to acquire the optimal discriminating features as follows:

$$f_p^{\,i} = \log\left(\frac{var(Z_p^{\,i})}{\sum_{i=1}^{2m} var(Z_p^{\,i})}\right)(p = 1 \dots 2m) \quad (5)$$

Here, the value of m is 1. Hence the number of features in each band is 24 (4 features multiply 6 spatial filters). As a result, the size of the feature input is $44 \times 24 \times 1$ (the number of bands $\times$ the number of features from multiple filters $\times$ 1).

IV. Results

To verify the performance of our approach, we evaluate our method by comparing it with different methods in dataset BCI IV IIa. Besides, all experiments employ the same training set and test set without any preprocessing. The classification accuracy is set as an evaluation metric as follows:

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (6)$$

(TP: true positive; TN: true negative; FP: false positive; FN: false negative)

Table II lists the accuracy of different approaches for each subject in four-class MI EEG classification task of the BCI IV IIa dataset. The proposed method can achieve a mean accuracy of 78.12% which outperforms other classical methods. For most subjects, our proposed method obtains better accuracy than other traditional methods. For subjects A06 and A08, there is only a tiny gap in accuracy between our method and the optimal method. These results prove the superiority of the proposed model.

In this study, we focus more on decoding the spatial information of MI-EEG. Self-attention model is used to acquire the channel weights and intensify spatial information in raw data. Before feature processing, FTBCSP is utilized as spatial filters to obtain significant spatial information between different classes.

TABLE II. The accuracy (%) comparison on BCI IV IIa dataset

Method	A01	A02	A03	A04	A05	A06	A07	A08	A09	Mean
EEGNet [8]	78.82	53.82	80.90	61.11	68.75	**58.68**	73.61	75.35	67.71	68.75
ST-Attention CNN [9]	72.57	54.86	78.47	55.90	69.10	50.00	69.79	67.71	68.06	65.16
Deep ConvNet [7]	48.96	44.79	52.78	43.40	46.53	35.07	54.86	44.79	42.36	45.95
Shallow ConvNet [7]	62.85	44.79	78.47	54.17	53.82	44.44	56.60	68.06	64.93	58.66
FBCSP [6]	78.13	49.65	76.74	60.42	57.29	45.14	81.60	76.74	65.28	65.66
Multiscale time-frequency method [12]	84.03	61.81	82.99	63.89	70.14	52.08	**92.01**	81.25	**80.21**	74.27
Proposed Method	**86.11**	**65.28**	**84.03**	**75.00**	**80.56**	58.33	90.63	**82.99**	**80.21**	**78.12**

In Table II, Deep ConvNet [7] shows the worst accuracy. Instead, the Shallow ConvNet [7] performs better than Deep ConvNet. The reason for the above situation may be that Deep ConvNet has higher complexity than the Shallow one. Compared with the common deep learning method, EEGNet [8], the mean accuracy in our approach is nearly 10% higher on datasets IV IIa. The comparison result indicates the effectiveness of our strategy in capturing spatial information. Furthermore, the reason why our model performs better than the traditional method FBCSP focused on spatial information is that the presented method can also take advantage of the temporal information. Due to the utilization of the self-attention CNN model, the proposed method shows better performance than the multiscale time-frequency method [12], which also employs the CSP algorithm on the frequency and temporal bands.

V. Conclusion

MI-BCI has become increasingly crucial in many fields, especially in rehabilitation. However, there are still a variety of problems in decoding MI-EEG, such as low accuracy and efficiency. We apply FTBCSP to extract the spatial feature of multi-channel EEG. Besides, self-attention is also utilized to acquire the channel weight of EEG. Hence, spatial and temporal information are fully leveraged in MI-EEG. It is also the first time that the FTBCSP features are combined with self-attention-based CNN. The proposed approach displays outstanding classification performance on the publicly available datasets.

However, some limitations such as the massive parameters in our model and the limited number of subjects still exist. In our future work, we will focus on optimizing the parameters and collecting more data to further validate the proposed method.

Acknowledgment

This research is supported in part by the project of Natural Science Foundation of Shandong Province (ZR2020QF024, ZR2020LZH009, ZR2021QH290), Jinan 20 Universities (2019GXRC040), Jinan 5150 Program for Talents Introduction, Major Basic Research Project of Shandong Natural Science Foundation (ZR2021ZD40), Key Research and Development Plan of Shandong Province (2021CXGC011304), and Shandong Institute of Advanced Technology, Chinese Academy of Sciences (YJZX003, YQCX20220106).

REFERENCIES

[1] G. Pfurtscheller, C. Neuper, D. Flotzinger, and M. Pregenzer, "EEG-based discrimination between imagination of right and left hand movement," Electroencephalography & Clinical Neurophysiology, vol. 103, no. 6, p. 642, 1997.

[2] G. Pfurtscheller, C. Brunner, A. Schl?Gl, and F. Silva, "Mu rhythm (de)synchronization and EEG single-trial classification of different motor imagery tasks," Neuroimage, vol. 31, no. 1, pp. 153-159, 2006.

[3] J. Müller-Gerking, G. Pfurtscheller, and H. Flyvbjerg, "Designing optimal spatial filters for single-trial EEG classification in a movement task," Clinical Neurophysiology, vol. 110, no. 5, pp. 787-798, 1999.

[4] H. Ramoser, J. Muller-Gerking, and G. Pfurtscheller, "Optimal spatial filtering of single trial EEG during imagined hand movement," IEEE Transactions on Rehabilitation Engineering, vol. 8, no. 4, pp. 441-446, 2000.

[5] Q. Novi, C. Guan, T. H. Dat, and P. Xue, "Sub-band Common Spatial Pattern (SBCSP) for Brain-Computer Interface," in 2007 3rd International IEEE/EMBS Conference on Neural Engineering, 2007, pp. 204-207.

[6] K. K. Ang, Z. Y. Chin, H. Zhang, and C. Guan, "Filter Bank Common Spatial Pattern (FBCSP) in brain-computer interface," in Proceedings of the International Joint Conference on Neural Networks, 2008, pp. 2390-2397.

[7] R. T. Schirrmeister et al., "Deep learning with convolutional neural networks for EEG decoding and visualization," Hum Brain Mapp, vol. 38, no. 11, pp. 5391-5420, Nov 2017.

[8] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: a compact convolutional neural network for EEG-based brain-computer interfaces," J Neural Eng, vol. 15, no. 5, p. 056013, Oct 2018.

[9] X. Liu, Y. Shen, J. Liu, J. Yang, P. Xiong, and F. Lin, "Parallel Spatial-Temporal Self-Attention CNN-Based Motor Imagery Classification for BCI," Front Neurosci, vol. 14, p. 587520, 2020.

[10] X. Ma, S. Qiu, and H. He, "Time-Distributed Attention Network for EEG-Based Motor Imagery Decoding From the Same Limb," IEEE Trans Neural Syst Rehabil Eng, vol. 30, pp. 496-508, 2022.

[11] G. A. Altuwaijri, G. Muhammad, H. Altaheri, and M. Alsulaiman, "A Multi-Branch Convolutional Neural Network with Squeeze-and-Excitation Attention Blocks for EEG-Based Motor Imagery Signals Classification," Diagnostics (Basel), vol. 12, no. 4, Apr 15 2022.

[12] G. Liu, L. Tian, and W. Zhou, "Multiscale time-frequency method for multiclass Motor Imagery Brain Computer Interface," Comput Biol Med, vol. 143, p. 105299, Feb 6 2022.