Received October 4, 2021, accepted October 26, 2021, date of publication October 29, 2021, date of current version November 8, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3124238

Filter Bank Convolutional Neural Network for SSVEP Classification

DECHUN ZHAO, TIAN WANG, YUANYUAN TIAN, AND XIAOMING JIANG, (Member, IEEE)
Chongqing Engineering Research Center of Medical Electronics and Information Technology, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

Corresponding author: Xiaoming Jiang (jiangxm@cqupt.edu.cn)

This work was supported in part by the Natural Science Foundation of Chongqing, China, under Grant cstc2018jcyjAX0163; and in part by the National Natural Science Foundation of China under Grant 61801069.

ABSTRACT Harmonics in electroencephalogram (EEG) caused by visual stimulation are the main basis of classification of steady-state visual evoked potential (SSVEP). However, the correlation of various harmonics, which could improve the classification performance especially when evoked EEG components are much weaker than spontaneous EEG components, has not been take into consideration in the design of classifier in previous studies. In this study, we proposed a filter bank convolutional neural network (FBCNN) method to optimize SSVEP classification. Three filters with passbands covering each harmonic of SSVEP signals are used to extract and separate the corresponding components, and the information from them are transformed into frequency domain. Subsequently, we introduce a novel convolutional neural network (CNN) architecture with three parallel CNN channels to extract and learn the harmonic features in passbands, and conclusions on the correlation among harmonics can finally be made by pair-add-up operations and dimension reductions to weigh the feature vectors. The proposed FBCNN is evaluated on two public datasets (Dataset1: 12-class, 10 subjects; Dataset2: 40-class, 35 subjects) to compare with other methods. The experimental results illustrate that FBCNN method improves the performance of CNN-based SSVEP classification methods and has a great potential to be applied in SSVEP-based BCI.

INDEX TERMS Brain–computer interface, convolutional neural networks, electroencephalography, filter bank, steady-state visual evoked potential.

I. INTRODUCTION

Brain-computer interface (BCI) based on electroencephalogram (EEG) measures EEG signals in a noninvasive way, extracts the specific features, subsequently, converts them into the commands of the equipment [1]. BCI provides a novel hand-free communication channel to control peripheral devices by realizing the interaction between human brain and machine, thus could facilitate lives of the disabled, help training of stroke patients with rehabilitation and even control computer games [2]. Among different EEG-based BCIs, e.g. motor imagery (MI) and event-related potentials etc., steady-state visual evoked potential (SSVEP) possesses the advantage of high information transfer rate (ITR) and few training times [3] and thus is widely used in the context of human-machine interaction.

In SSVEP, the visual stimuli with different flicker frequencies is applied and the consequent oscillations could occur in the visual cortex, which could be further detected in EEG signals in the form of the strong amplitude of corresponding frequency and harmonic. Based on the observed pattern of the detected EEG signals, the recognition algorithm could be used to find out target stimulus [3]. Generally, the performance of SSVEP-based BCI is mainly determined by three factors, namely, stimulus presentation, multiple target coding and target identification algorithm [4].

Practically, the target identification algorithm of SSVEP-based BCI can be divided into three categories: training-free methods, user-specific or user-dependent training methods and user-independent training methods [5]. Training-free methods compute the relevance between the detected signals and the potential stimuli, and hence directly determine the classification results. These methods mainly include power spectral density analysis (PSDA) [6], canonical correlation analysis (CCA) [7] and minimum energy combination (MEC) [8]. Among these methods, CCA, the most prevalent training-free method, is always treated as the baseline algorithm for SSVEP detection [9]. It aims at finding the

The associate editor coordinating the review of this manuscript and approving it for publication was Fanbiao Li.

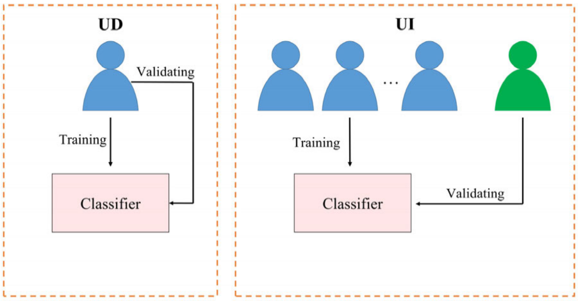


FIGURE 1. The diagram of user-dependent (UD) and user-independent (UI) training strategies.

potential correlation between the detected EEG data and a set of sinusoidal reference templates corresponding to the stimulus frequency.

In contrast with the training free method, the training methods, either user-dependent (UD) or user-independent (UI) method, incorporate the features extracted from the trail data to improve the classification accuracy [5]. The difference between UD and UI training methods lies in the training and testing strategy, which can be observed in Fig. 1. As for UD training methods, the trained model is generated from the trail data from one person, and is only suited for this individual in the classification step. However for UI training methods, a generalized model generated by the training data from multiple participants is applied to extra users, i.e., arbitrary new users have access to BCI equipment with no collection of their own training data. For example, Combination method [10], Individual Template CCA (IT-CCA) [11] and Multi-way CCA (MwayCCA) [12] are considered as the UD methods and Filter Bank CCA (FBCCA) [13] is the typical UI methods. The design strategies of filter bank will be described in the following paragraphs.

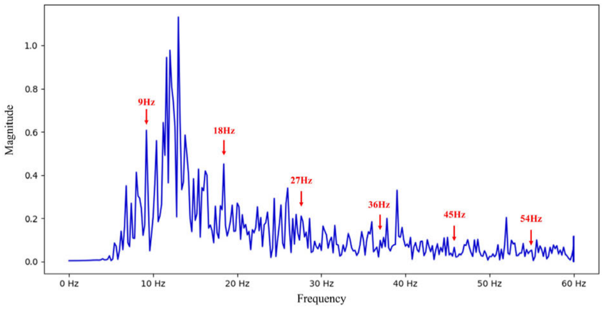


FIGURE 2. FFT spectrum of the first block and O1n EEG channel data of S1 with stimulation frequency of 9Hz in Dataset2.

Recently, deep learning has developed rapidly in view of its prominent capabilities of feature extraction and learning [14], providing new insight to the classification and the detection of EEG-based BCI. In fact, the deep learning based methods tend to outperform traditional methods in most fields including EEG signal detection [15], face recognition [16] and cross-media retrieval [17]. The advantages of convolutional neural network (CNN) over standard deep neural network (DNN) can be seen in prior researches [18], and the time delays term in neural networks has also been explored in depth [19]. In short, CNN structure plays the most popular role in different deep learning-based BCI algorithms [20].

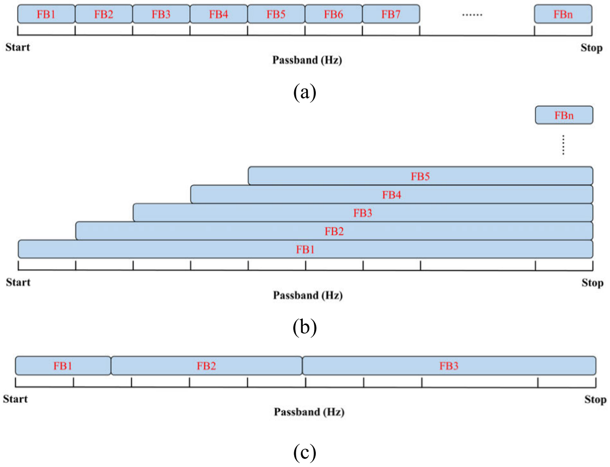


FIGURE 3. (a-b) Two typical filter bank design schema and (c) the proposed filter bank schema.

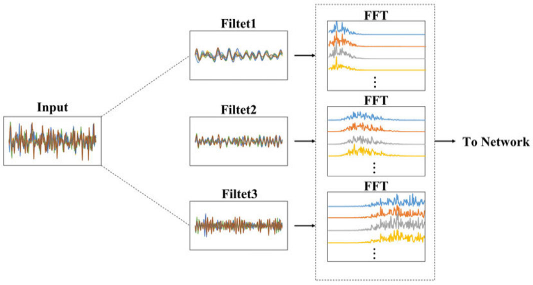


FIGURE 4. The input structure of FBCNN including three band-pass filters with different bandwidth and three FFT converters.

The CNN-based methods in SSVEP can be categorized in many perspectives. For example, the input data of the algorithms, which could impact the performance of the classifiers, can generally be classified into two broad categories: time domain data and frequency domain data. When time domain signals used as input, wider and deeper convolution operations are generally required in feature extraction step, while the neural networks with the input of frequency domain data don't need too many convolution operations [9], [15], [24]. Given that significant frequency and phase characteristics in SSVEP signals usually remain stable during stimulation, Fast Fourier Transform (FFT) [9], [21]–[23] is widely applied to transform SSVEP signals into the frequency domain before feeding into classifiers. According to the reported researches, the methods using frequency domain data as input tend to present better performance [9]. The other main difference lays in the diverse training strategies for the CNN-based methods. UD training procedure uses training and testing data from the same participant, whereas data from different

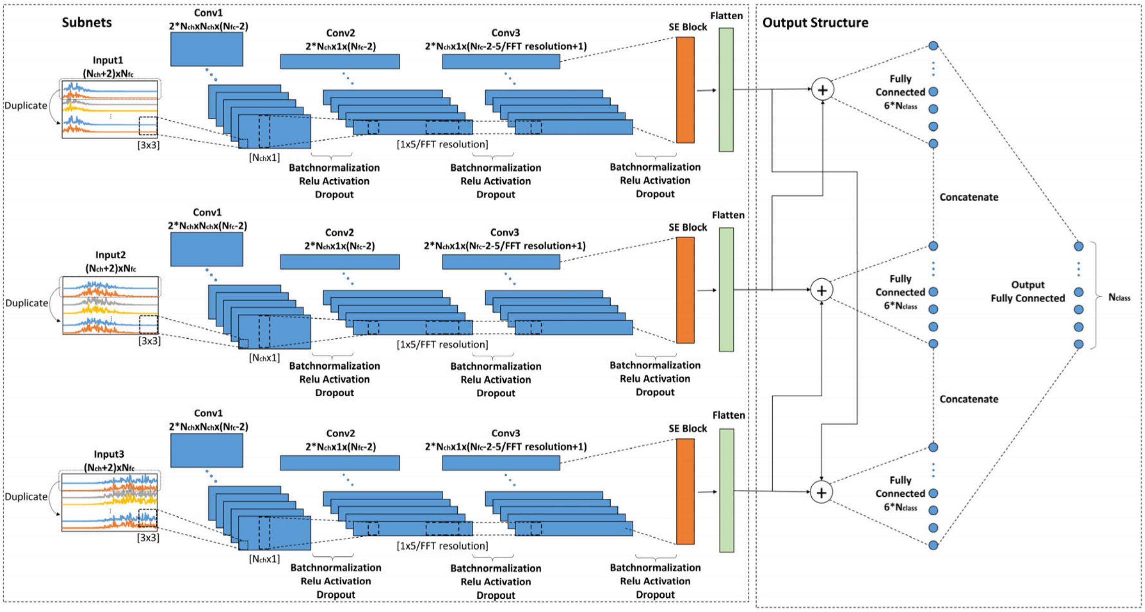


FIGURE 5. FBCNN network architecture including the subnets and the output structure.

participants are used for training and validating respectively in UI training procedure. From the related studies, classifiers based on UD training procedure perform better than that based on UI training procedure due to the generalization ability of classifiers [9]. On the other hand, results from previous studies show that the deep learning based SSVEP classification methods outperform traditional methods. For example, PSDA and CCA can classify targets directly without training, while methods based on deep learning need to be trained in advance, leading to the increase in the complexity of applications.

Although the CNN-based SSVEP classification algorithms has been widely recognized for its end-to-end characteristic, the prior knowledge for SSVEP, for example, the relevance among different SSVEP harmonics, has not received deserved attention. In some cases, as is shown in Fig. 2, it is necessary to analyze the correlation among different harmonic components of the evoked EEG signals to optimize classification. Since the spontaneous EEG components of the subjects are much stronger than the SSVEP evoked EEG components in light of inattention, individual difference or interference caused by environment etc, classifiers pay more attention on spontaneous EEG signals while neglecting the harmonics with weaker signal amplitude. Therefore, we put forward a novel input structure of the classifier and a new CNN architecture.

Our design strategy of the proposed filter bank differs from previous studies. As is shown in Fig. 3, we compare three different filter bank schema. Fig. 3(a) illustrates one typical filter bank structure using plenty of narrow band-pass filters with equally spaced bandwidths to extract independent frequency components from SSVEP signals. Fig. 3(b) shows another typical structure owing multiple filters with different bandwidths to cover all harmonics [13]. These methods aim to concentrate the feature analysis on each independent frequency band, ignoring the correlation among harmonic components. In our method, as is shown in Fig. 3(c) and Fig. 4, three filters with passbands covering the 1st, 2nd and residual harmonics of the input SSVEP signals respectively are used to extract each harmonic component before FFT operation. The 1st harmonic represents the stimulation frequency range. The detailed example of filer bandwidth can be found in the experiment section. Moreover, our neural network utilizes three parallel CNN channels to extract and learn the harmonic component information in each passband, other than mixed harmonic components in the CNN structure proposed in [9]. By means of pair-add-up operations and dimension reductions on the output features of each CNN channel, the feature vectors of each harmonic are weighted to find the correlation among them. Finally, the classification results are output through fully connected layer. By foregoing operations, we learned not only the feature of harmonics but also relevance among them, which receive less attention when single band used in CNN.

In order to compare with the method demonstrated in [9], we also implement our method in both UD and UI training procedure. Additionally, to verify the feasibility of the model and the fairness of the results, we test the proposed model using two public datasets: 1) Dataset1: a publicly available twelve class SSVEP dataset with 10 participants [10]; 2) Dataset2: a publicly available forty class SSVEP dataset with 35 participants [25]. Since Dataset2 was not used in [9], we reproduce the architecture proposed in the paper for comparison.

The paper is organized as follows. Section II outlines the proposed SSVEP classification method. Section III details the information of two datasets and training parameters. The performance of the proposed method is presented in Section IV. Section V discusses the results of the comparison. Finally, Section VI concludes the paper and prospects for the future work.

II. METHODS

In view of strong correlation between harmonics [13], different SSVEP harmonic components are used in the classification of SSVEP. To extract various harmonics for feature learning, filter bank is introduced to separate each harmonic component. Considering the computational complexity and classification accuracy [13], [26], three band-pass filters are selected to preprocess the EEG signals before FFT operations, followed by a corresponding neural network to classify the SSVEP targets, and Fig. 4 displays the input structure of the proposed method. The whole algorithm structure is called filter bank convolutional neural network (FBCNN). In the proposed method, three independent CNN subnets are used to extract features and weight contributions of different harmonics, and the attention mechanism offers an enhancement of the weighted features. The follow-up pair-add-up operations and the fully connected layers are used to fuse different harmonic features and output the classification results.

A. DATA PROCESSING

To compare the performance of various classification algorithms, different data preprocessing methods are employed in two datasets. The data preprocessing method for Dataset1 and Dataset2 is based on [9], [10], [24], and [25] respectively for comparison, and we reproduce the CNN [9] structure for Dataset2. In detail, a 4th order Butterworth band-pass filter is used to remove the artifacts which may exist in EEG signals for CCA [7] and CNN [9]. In our case, the EEG data are filtered by three 4th order Butterworth band-pass filters with different frequency bands to separate the harmonics. Then, the filtered EEG data of each SSVEP trail is divided into non-overlapping segments with 1s time window (TW).

As in the previous studies, the results of FFT for each 1s segment contain two parts: magnitude spectrum and complex spectrum [9]. The output of FFT is shown as follows:

$$FFT(x) = \mathrm{Re}\{FFT(x)\} + jIm\{FFT(x)\} \tag{1}$$

where x indicates the input time domain segment. The magnitude spectrum X_{mag} can be calculated by:

$$X_{mag} = \sqrt{\mathrm{Re}\{FFT(x)\}^2 + \mathrm{Im}\{FFT(x)\}^2} \tag{2}$$

And for complex spectrum X_{comp}, the real part and imaginary part of the FFT output components are concatenated into a single vector as:

$$X\mathrm{comp} = \mathrm{Re}\{FFT(x)\}||\mathrm{Im}\{FFT(x)\} \tag{3}$$

The magnitude spectrum only takes the magnitude information of the FFT results, without the phase information [9]. However, previous studies have shown the importance of phase related information presented in the SSVEP-based BCIs [24], [28]–[30]. Hence, it is necessary to use the complex spectrum which contains related information of both magnitude and phase simultaneously. The magnitude input I_{mag} can be defined as:

$$I_{\mathrm{mag}} = \begin{bmatrix} X_{mag}(O_1) \\ X_{mag}(O_2) \\ \vdots \\ X_{mag}(O_n) \end{bmatrix} \tag{4}$$

where O_1, O_2 and O_n represent the different EEG data channels. The following computation defines the complex input:

$$I_{\mathrm{comp}} = \begin{bmatrix} X_{comp}(O_1) \\ X_{comp}(O_2) \\ \vdots \\ X_{comp}(O_n) \end{bmatrix} \tag{5}$$

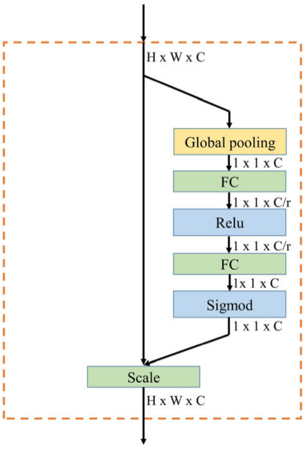


FIGURE 6. The schema of the squeeze-and-excitation block.

B. FILTER BANK CONVOLUTIONAL NEURAL NETWORK

We propose a SSVEP classification method which combines filter bank and CNN structure. As shown in Fig. 3, three filters with different bandwidth are employed in order to fulfill the fine-grained feature learning of the SSVEP data. The bandwidth of Filter1 covers the first harmonic frequency range of the SSVEP, and Filter2 is for the second harmonic, Filter3 is for the rest. Three CNN-based subnets are introduced to compose the network of FBCNN for extracting and weighting features of different SSVEP harmonic components. Fig. 5 shows the network architecture of FBCNN. Each subnet contains 6 layers: an input layer, three convolutional layers, a Squeeze-and-Excitation (SE) block and a flatten layer. The output structure consists of three add layers and fully connected layers followed by subnets. The complete network has an additional concatenate layer and an output full connected layer.

The dimension of the input layer is $(N_{ch}+2) \times N_{fc}$, where N_{ch} represents the number of the EEG channels and N_{fc} is the number of frequency components extracted from FFT result. In FBCNN, the preprocessed data of the first two EEG channels are repeated to achieve entirely convolution in the Conv1 which uses the "valid" padding mode. The input layer composition is shown as follows:

$$I = \begin{bmatrix} X(O_1) \\ X(O_2) \\ \vdots \\ X(O_n) \\ X(O_1) \\ X(O_2) \end{bmatrix} \tag{6}$$

where O_1, O_2 and O_n are different EEG data channels. The convolution kernel size of Conv1 is 3×3 which is used to extract components of different EEG channels and frequencies at the same time. Conv2 with kernel dimensions of $N_{ch} \times 1$ is employed to learn the contribution of weighted EEG channel features of Conv1. Conv3 extracts features of various weighted continuous frequency components, and the scale of the convolution kernel is set to $1 \times 5/FFT\ resolution$ empirically, which indicates the span of the frequency is 5Hz. The number of feature maps in the three convolutional layers is $2 * N_{ch}$ for the sake of extracting features of the previous layer completely and minimizing the amount of calculation. The dimension of each feature map in Conv1 layer is $N_{ch} \times (N_{fc} - 2)$, $1 \times (N_{fc} - 2)$ for Conv2, and $1 \times (N_{fc} - 2 - 5/FFT\ resolution + 1)$ for Conv3. A SE block is followed with the aim of enhancing the representational power of the network. SE block explicitly models the inner dependencies among channels of convolutional features to acquire preferable representation [31]. Fig. 6 shows the schema of the SE block, the input feature maps have the dimension of $H \times W \times C$ which represent the height, width and channels of the feature maps respectively. The parameter r is the reduction ratio of the dimensionality-reduction layer, and the ratio is 8 in this study. The input feature maps are firstly passed through a squeeze operation to produce a channel descriptor by aggregating, then followed by an excitation operation to produces a collection of per-channel modulation weights [31]. The scale layer of the SE block rescales the feature maps by:

$$X_s = u_c s_c \tag{7}$$

where u_c is the feature maps, and s_c is the scalar. The flatten layer of each subnet is used to compress the output feature vectors of SE block to one-dimension.

Subnets follow by layers to perform pair-add-up operations and fully connected layers to reduce dimension for weighting the feature vectors of each harmonic. And the number of units $6 * N_{class}$ output from fully connected layers equipped with the rectified linear unit (ReLu) activation function is in line with that of SSVEP classes. Then, the outputs of three CNNs are connected by a concatenate layer, and finally a fully connected layer possessed with the *softmax* function is used to output N_{class} units corresponding to the probability of each SSVEP class. Batch normalization, ReLu activation function and Dropout are performed on Conv1, Conv2 and Conv3 layers of each CNN channel of the FBCNN network. For neural networks, Batch normalization and Dropout are commonly applied to enhance the generalization performance and calculation speed [23], [32], [33].

C. CONVOLUTIONAL NEURAL NETWORK

The architecture of CNN is shown in Fig. 7, which was proposed in the earlier study of SSVEP classification [9]. The CNN is composed of four main layers, an input layer with the dimension of $N_{ch} \times N_{fc}$, two convolutional layers and a output fully connected layer with the unit number of N_{class}. N_{ch}, N_{fc} and N_{class} stand for the number of the EEG channels, the number of frequency components extracted from FFT result, and the number of SSVEP classes respectively. The first convolutional layer Conv1 performs 1D convolutions across the channel dimension with kernel dimension of $N_{ch} \times 1$,and the number of feature maps in the first convolutional layer is 2^*N_{ch} and each feature map has dimensions $1 \times N_{fc}$. For Conv2, the scale of convolutional kernel is $1 \times 3/FFT\ resolution$.Thus, for Dataset1 the scale is 1×10 [9], and for Dataset2 the scale is 1×15. The scale, a fixed value of 1×10 in the earlier study [9], did not function effectively compared to 1×15 on account of different sampling frequency and FFT resolution when we reproduce the CNN structure on Dataset2. Furthermore, the kernel size of Conv2 has been changed into 1×15 in order to ensure fairness.

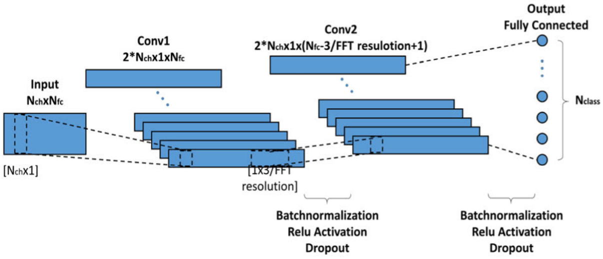


FIGURE 7. Convolutional neural network architecture for SSVEP classification.

Batch normalization and ReLu activation function are performed on the outputs of Conv1 and Conv2, and Dropout is used to avoid overfitting.

D. CANONICAL CORRELATION ANALYSIS

The training-free method CCA is introduced as a baseline to perform on the same EEG datasets. CCA, a multivariate statistical analysis method, reflects the underlying correlations between two groups of indicators. In previous SSVEP-based BCIs studies, CCA was used to calculate the correlations between EEG signals and reference signals corresponding to SSVEP classes [9], [10], [24], [25]. The canonical correlation coefficient $\rho(x, y)$ of CCA is defined as:

$$\rho(x, y) = max_{w_x, w_y} \frac{E\left[w_x^T X\, Y^T w_y\right]}{\sqrt{E\left[w_x^T X\, X^T w_x\right] E\left[w_y^T Y\, Y^T w_y\right]}} \tag{8}$$

where X and Y are input EEG signals matrices and reference signals matrices respectively, and x, y stand for the linear representation of X, Y respectively. w_x and w_y indicate the linear coefficient vectors. In this paper, sinusoidal signals are used as the reference signals Y to perform the classification in an unsupervised way. The reference signals Y is defined as:

$$Y = \begin{bmatrix} \sin(2\pi ft) \\ \cos(2\pi ft) \\ \vdots \\ \sin(2\pi Nhft) \\ \cos(2\pi Nhft) \end{bmatrix} \tag{9}$$

where f corresponds to the target frequency of the SSVEP, and N_h is the number of harmonics.

E. ALGORITHM EVALUATION

One-way repeated measures analysis of variance (ANOVA) is applied to evaluate the results of CAA, CNN and FBCNN classification methods on both datasets, and the classification accuracy of each method is input as the response variable. The statistical significance level is 0.05 for analysis and comparison. The CNN and FBCNN are compared on every subject in the two datasets using both UI and UD training approaches. Moreover, the information transfer rate (ITR) of each method is calculated as follows:

$$ITR = \frac{60}{T}(\log_2 N + P \log_2 P + (1 - P) \log_2[\frac{1 - P}{N - 1}]) \tag{10}$$

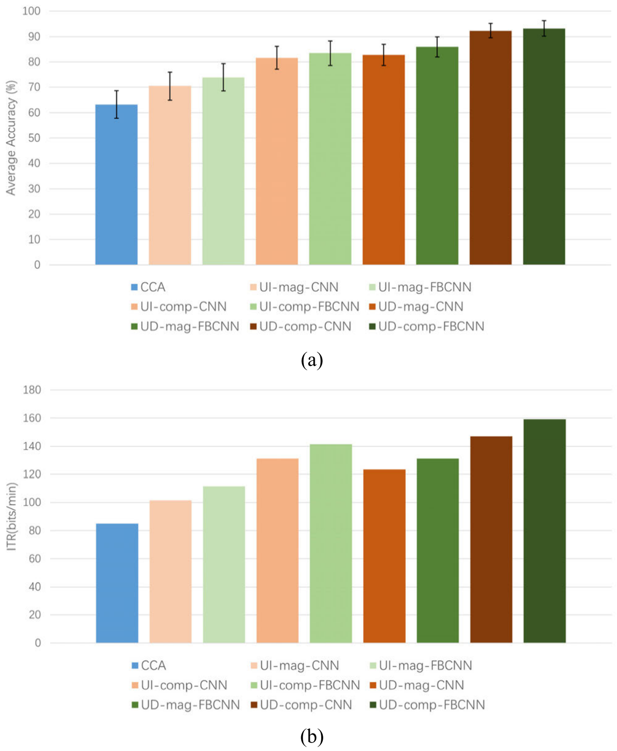
where N and T are the number of SSVEP classes and the average time for a selection, and P is the classification accuracy.

III. EXPERIMENTS

A. DATASET1 DESCRIPTION

The twelve stimuli are arranged on a 27-inch LCD monitor flashing at frequencies step of 0.5Hz ranging from 9.25Hz to 14.75Hz, with the corresponding phases ranging from 0π to 1.5π in steps of 0.5π [10].

EEG data are recorded from ten healthy subjects (9 males and 1 female, mean age: 28 years) with normal or corrected-to-normal vision using a BioSemi ActiveTwo EEG system (Biosemi, Inc.) with the sampling rate of 2048Hz. All subjects are seated in a comfortable chair in a dim room, 60 cm in front of the LCD monitor. The experiment consists of 15 blocks for each subject, and a block contains 12 trails corresponding to all 12 targets in a random order. A red square appears for 1s at the position of the target stimulus at the beginning of each trial. Participants are asked to shift their gaze to the target within the 1s duration. All stimuli start to flicker simultaneously for 4s after the red square disappears. Participants are instructed to avoid eye blinks during the 4s stimulation process.



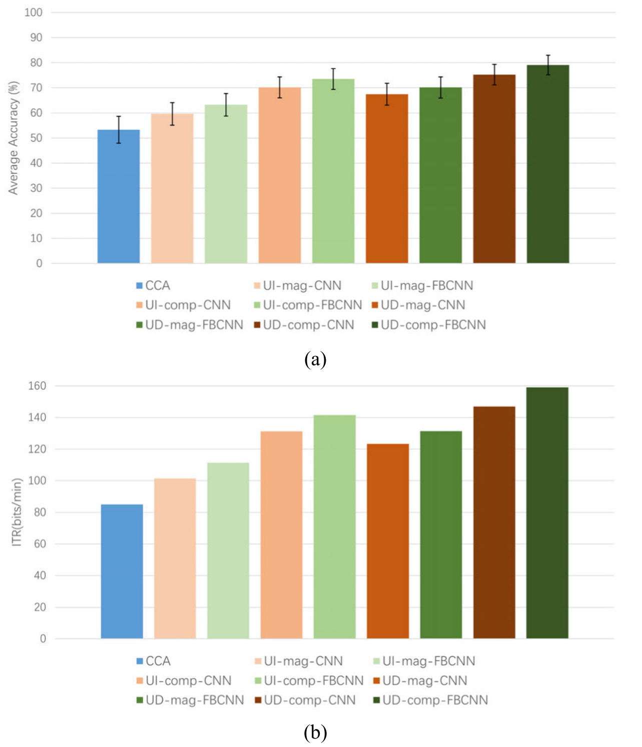
(a)

(b)

FIGURE 8. (a) Comparisons of the average classification accuracies and (b) the average ITR on Dataset1 for different methods with 1s data length. Error bar in each method indicates the standard deviation among all subjects.

B. DATASET2 DESCRIPTION

An open 40-target dataset for SSVEP-based BCIs conducts a cue-guided target selecting task. 40 characters are presented on a 23.6-in LCD monitor and the viewing distance to the LCD monitor is 0.7 meters [25]. The frequencies of 40 characters rang from 8Hz to 15.8Hz with the step of 0.2Hz, and the phase step is 0.5π. 40 targets are coded using a joint frequency and phase modulation (JFPM) approach [27]. EEG data are acquired using the Synamps2 EEG system (Neuroscan, Inc.) from Thirty-five healthy subjects (17 females, aged 17–34 years, mean age: 22 years) with normal or corrected-to-normal vision. The sampling rate is 1000 Hz. Each trial lasts 6s in total. At the beginning of each trail, subjects are asked to shift their gaze to a red square at the target location during 0.5s cue duration as soon as possible. Then, all stimuli start to flicker on the LCD monitor simultaneously for 5s after the cue. The blank in the LCD monitor lasts for 0.5s following the stimulation. The experiment includes 6 blocks for each participant, with one block containing 40 trials corresponding to all 40 targets indicated



(a)

(b)

FIGURE 9. (a) Comparisons of the average classification accuracies and (b) the average ITR on Dataset2 for different methods with 1s data length. Error bar in each method indicates the standard deviation among all subjects.

in a random order. A total of 240 trials are conducted on each participant.

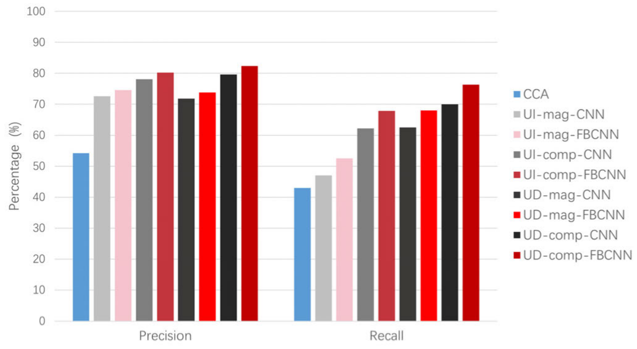


FIGURE 10. Comparisons of the average classification precision and recall on Dataset2 for different methods in this study 1s data length.

C. DATA PREPROCESSING PARAMETERS

In Dataset1, the data of all the 8 channels are filtered by a 4th order Butterworth band-pass filter between 6Hz and 80H for CCA [7], while by three 4th order Butterworth band-pass filter in our proposed method FBCNN. The bandwidth of three filters are 6Hz-16Hz, 16Hz-32Hz and 32-64Hz respectively. Then, each 4s trail is divided into 1s non-overlapping segments. Considering the visual latency, a time delay of 135ms is added in the extraction [10], and the frequency components converted by FFT with a resolution of 0.25Hz are extracted between 5Hz and 55Hz for FBCNN methods.

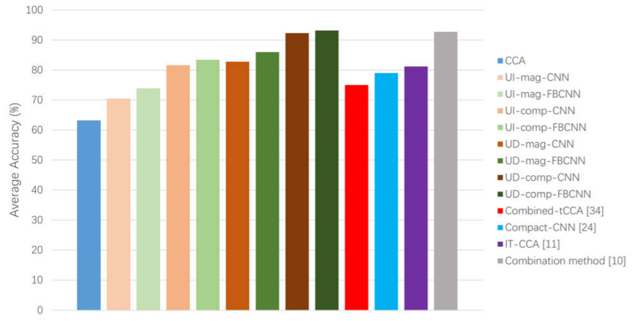


FIGURE 11. Comparison of the classification accuracies on Dataset1 for different methods in this study and other reported methods with 1s data length.

In Dataset2, O1, Oz, O2, PO3, POz, PO4, Pz, PO5 and PO6 are used as the input channels for all methods. For CCA [7] and CNN [9], the bandwidth is between 6Hz and 64Hz; for FBCNN, three band-pass filter are used to filter the data, and the bandwidth of three filters are 6Hz-18Hz, 14Hz-36Hz and 28 Hz-64Hz respectively. The first 140 milliseconds of data are removed because of the visual latency [25]. Finally, the filtered 5s EEG data of each trail is divided into 1s non-overlapping segments. The frequency domain signals transformed by FFT with a resolution of 0.2Hz are extracted between 5Hz and 55Hz for both CNN [9] and FBCNN.

TABLE 1. The details of hardware platform.

Items	Value
Operating System	Ubuntu 18.04.4 LTS
CPU	Intel(R) Xeon(R) Gold 6148
GPU	NVIDIA Tesla V100-PCIE-32GB
RAM Size	128GB
Language	Python3
Machine Learning Platform	Tensorflow2.0

D. TRAINING PROCESS

A normal distribution with a mean of 0 and a standard deviation of 0.01 is used to initialize the convolutional layers and fully connected layers of the FBCNN and CNN network. Both networks are trained by the stochastic gradient descent (SGD) optimization algorithm with the momentum of 0.9, and the categorical cross-entropy loss function with back-propagation technique is introduced to achieve the difference between predicted value and real value. The hyper parameters are obtained according to the best performance of networks on every participant during the training and testing process. Notably, the training process of CNN architecture has not been reproduced for Dataset1, since the average accuracies

TABLE 2. Classification accuracy (%) on Dataset1 for each subject with 1s data length.

Subject	UD-mag		UD-comp		UI-mag		UI-comp	
	CNN	FBCNN	CNN	FBCNN	CNN	FBCNN	CNN	FBCNN
S1	**65.13**	**75.00**	**77.91**	**91.67**	**36.52**	**49.90**	**60.50**	**66.24**
S2	**37.22**	**43.33**	57.77	57.08	20.53	21.01	36.50	31.38
S3	**82.36**	**88.61**	94.86	97.36	68.12	70.25	**77.12**	**81.89**
S4	92.08	93.47	98.33	98.11	88.96	87.19	93.88	93.43
S5	94.72	95.97	99.86	99.58	**82.98**	**86.94**	**87.89**	**94.78**
S6	96.38	99.31	99.44	99.95	**85.89**	**90.58**	93.24	96.04
S7	**88.88**	**94.58**	**94.58**	**98.75**	**77.20**	**86.28**	**86.76**	**92.72**
S8	98.05	98.47	99.16	99.58	96.11	97.13	98.33	98.03
S9	88.47	89.17	97.91	97.92	73.45	75.47	92.50	94.17
S10	80.97	81.81	90.69	91.94	74.78	74.33	86.50	85.88
Mean±STD	**82.77±16.7***	**85.97±15.9**	**92.33±11.1***	**93.19±12.4**	**70.50±22.0***	**73.91±21.7**	**81.60±18.0***	**83.46±19.5**

*Values used directly from [9]

results are directly from [9]. The FBCNN code on Dataset1 and the CNN and FBCNN codes on Dataset2 are available in https://github.com/tianwangchn/SSVEP_FBCNN_12class and https://github.com/tianwangchn/SSVEP_FBCNN_40class, respectively.

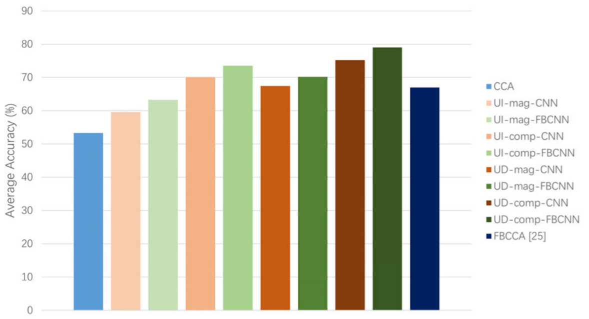


FIGURE 12. Comparison of the classification accuracies on Dataset2 for different methods in this study and the reported FBCCA with 1s data length.

For UD training method, the leave-one-session-out strategy is performed on each participant's dataset. To be specific, the network is trained and validated on the same participant's data, and 10-fold cross-validation method is used to divide the data. The sum of 1s non-overlapping segments for Dataset1 are: 648 (training) and 72 (testing), and for Dataset2 are: 1080 (training) and 120 (testing). UD-mag-FBCNN and UD-mag-CNN are used to instead of the training of the methods using magnitude spectrum features as input. And, the methods using complex spectrum features are referred to as UD-comp-FBCNN and UD-comp-CNN. For FBCNN method, the final parameters are chosen as: Learn rate (0.001), Dropout ratio (0.25), L2 Regularization (0.001), Number of Epochs (120, Dataset1), Batch size (16, Dataset1) and Epochs (150, Dataset2), Batch size (32, Dataset2). For CNN method on Dataset2, the final parameters are chosen as: Learn rate (0.001), Dropout ratio (0.25), L2 Regularization (0.0001), Number of Epochs (150), Batch size (128).

For UI training method, the leave-one-participant-out strategy is introduced to carry out the training and testing procedure. In UI method, a subject is leave out of the all subjects for testing process, and the remains are used to training the classifier. The total number of 1s non-overlapping segments for Dataset1 are: 6480 (training) and 720 (testing), and for Dataset2 are: 40800 (training) and 1200 (testing). UI-mag-FBCNN and UI-mag-CNN are used to instead of the training of the methods using magnitude spectrum features as input. And, the methods using complex spectrum features are referred to as UI-comp-FBCNN and UI-comp-CNN. For FBCNN method, the final parameters are chosen as: Learn rate (0.001), Dropout ratio (0.25), L2 Regularization (0.001), Number of Epochs (50, Dataset1), Batch size (128, Dataset1) and Epochs (50, Dataset2), Batch size (256, Dataset2). For CNN method on Dataset2, the final parameters are chosen as: Learn rate (0.001), Dropout ratio (0.25), L2 Regularization (0.0001), Number of Epochs (50), Batch size (512). Both CNN and FBCNN networks are trained and validated on the hardware platform listed in Table 1.

IV. RESULTS

A. RESULTS OF DATASET1

Fig. 8(a) shows the average classification accuracies of all classification methods across 10 participants on Dataset1. Table 2 lists the classification accuracies for all the 10 subjects on Dataset1 with 1s data length. Among all the methods, UD-comp-FBCNN achieves the highest accuracy of 93.19±12.4%. The methods of UD strategy: UD-mag-FBCNN, UD-comp-FBCNN, UD-mag-CNN and UD-comp-CNN all perform better than the methods of UI strategy: CCA, UI-mag-FBCNN, UI-comp-FBCNN, UI-mag-CNN and UI-comp-CNN. The average accuracies of all the

TABLE 3. Classification accuracy (%) on Dataset2 for each subject with 1s data length.

Subject	UD-mag		UD-comp		UI-mag		UI-comp	
	CNN	FBCNN	CNN	FBCNN	CNN	FBCNN	CNN	FBCNN
S1	**71.33**	**76.67**	**73.58**	**85.42**	**57.68**	**69.02**	**71.68**	**77.33**
S2	78.00	79.50	**88.58**	**92.92**	54.95	56.52	76.57	78.35
S3	84.67	86.17	**89.25**	**93.08**	**77.81**	**82.77**	**85.42**	**89.13**
S4	83.67	85.50	**90.50**	**94.17**	**75.28**	**79.05**	**81.79**	**85.67**
S5	**79.92**	**85.42**	**87.00**	**93.17**	**64.52**	**74.77**	**79.50**	**84.72**
S6	**70.16**	**74.25**	**80.16**	**83.75**	**71.75**	**76.48**	**83.00**	**88.27**
S7	**50.75**	**55.08**	73.88	75.25	42.83	43.70	69.06	71.03
S8	**52.67**	**57.67**	**61.66**	**67.00**	**51.94**	**55.30**	**62.00**	**68.95**
S9	61.25	62.92	**68.25**	**71.25**	62.83	62.63	69.89	73.42
S10	**69.47**	**74.67**	**78.58**	**91.25**	**66.28**	**71.82**	**80.68**	**84.48**
S11	34.83	34.41	**38.75**	**43.67**	**20.97**	**24.70**	**23.32**	**30.53**
S12	83.16	81.68	86.92	86.75	68.74	70.52	**72.00**	**78.40**
S13	**64.67**	**70.33**	**72.42**	**82.83**	**28.07**	**41.12**	**47.47**	**53.75**
S14	**78.67**	**81.92**	**81.08**	**84.92**	61.79	57.08	54.89	55.63
S15	51.92	51.92	64.42	61.32	51.91	54.27	69.72	70.77
S16	**56.00**	**63.17**	**72.00**	**77.25**	**48.42**	**54.10**	**58.19**	**63.98**
S17	54.19	54.42	70.91	72.75	**51.06**	**55.12**	**69.47**	**73.92**
S18	51.75	54.17	**62.08**	**65.50**	58.71	59.33	71.26	71.58
S19	**33.67**	**38.00**	**36.41**	**43.67**	29.09	29.30	36.62	36.43
S20	**69.33**	**73.08**	**78.00**	**87.25**	**67.91**	**75.78**	**73.69**	**79.52**
S21	78.38	78.08	85.42	86.50	66.59	66.37	76.00	76.75
S22	88.33	89.58	**91.42**	**94.67**	85.15	87.57	91.66	94.57
S23	**74.67**	**78.33**	**77.83**	**83.92**	64.15	66.82	71.13	72.67
S24	**78.29**	**82.00**	**81.92**	**85.25**	82.03	84.28	85.45	85.57
S25	76.75	77.17	80.00	81.33	**70.54**	**73.73**	**74.29**	**79.50**
S26	82.25	81.67	84.92	85.17	**72.74**	**77.08**	**76.24**	**81.10**
S27	88.16	89.42	94.16	94.25	85.14	87.92	92.13	93.60
S28	**73.16**	**77.92**	**84.08**	**90.50**	**66.30**	**71.45**	81.48	84.20
S29	**32.41**	**39.42**	46.75	49.50	**27.14**	**31.40**	**49.42**	**54.93**
S30	67.67	69.92	83.00	81.50	**46.11**	**56.33**	**65.89**	**69.12**
S31	**87.13**	**91.00**	96.50	96.58	**80.98**	**84.27**	90.24	92.45
S32	91.92	93.42	94.00	95.92	88.72	88.68	93.78	93.58
S33	**22.58**	**26.42**	**28.72**	**36.25**	21.54	19.68	27.35	27.63
S34	64.48	64.42	76.75	76.00	**53.17**	**57.77**	**75.95**	**79.17**
S35	74.58	76.25	74.08	76.17	**63.78**	**67.87**	**67.90**	**72.45**
Mean±STD	**67.45±17.4**	**70.17±16.9**	**75.26±16.2**	**79.05±15.6**	**59.62±17.9**	**63.27±18.0**	**70.15±16.7**	**73.52±16.5**

methods for data length of 1s are: CCA: 63.22±21.7%, UD-mag-CNN: 82.77±16.7%, UD-mag-FBCNN: 85.97±15.9%, UD-comp-CNN: 92.33±11.1%, UD-comp-FBCNN: 93.19± 12.4%, UI-mag-CNN: 70.5±22%, UI-mag-FBCNN: 73.91± 21.7%, UI-comp-CNN: 81.6±18%, UI-comp-FBCNN: 83.46± 19.5%. For UD approach, the average accuracies of FBCNN methods corresponding to inputting magnitude and complex feature respectively are 3.2% and 0.86%, higher than that of CNN methods. For UI approach, the average accuracies of FBCNN methods corresponding to inputting magnitude and complex feature respectively are 3.41% and 1.86%, higher than that of CNN methods. Fig. 8(b) summarizes the average ITR(bits/min) for all methods with 1s data length and 0.5s of gaze-shifting as: CCA: 59.5 bits/min, UD-mag-CNN:

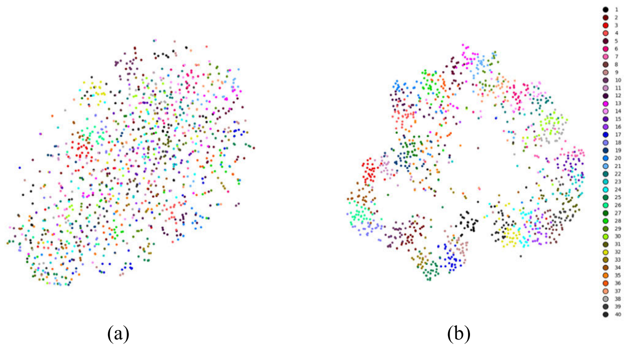


FIGURE 13. (a) Output feature clusters of the Conv2_Flatten layer (second to last layer) of UI-mag-CNN and (b) output feature clusters of the concatenate layer (second to last layer) of UI-mag-FBCNN on Dataset2.

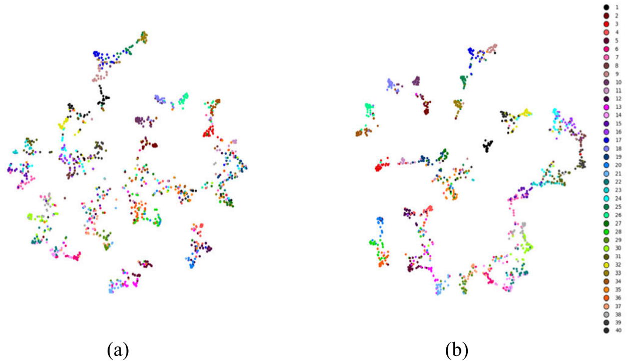


FIGURE 14. (a) Classification result clusters of UI-mag-CNN and (b) classification result clusters of UI-mag-FBCNN on Dataset2.

101.5 bits/min, UD-mag-FBCNN: 109.7 bits/min, UD-comp-CNN: 127.8 bits/min, UD-comp-FBCNN: 130.5 bits/min, UI-mag-CNN: 73.7 bits/min, UI-mag-FBCNN: 80.9 bits/min, UI-comp-CNN: 98.6 bits/min, UI-comp-FBCNN: 103.2 bits/min. These results demonstrated the strength of FBCNN methods in Dataset1. And, the UD-based training methods obtained higher performance than the UI-based training methods and CCA. The one-way repeated measures ANOVA revealed a significant difference in the classification accuracy among these methods ($p < 0.02$)

B. RESULTS OF DATASET2

Fig. 9(a) illustrates the average classification accuracies of all the methods for Dataset1 across 35 participants. Table 3 lists the classification accuracies for all the 35 subjects for Dataset1 with 1s data length. The one-way repeated measures ANOVA disclosed that the classification accuracy among these methods differ significantly ($p < 0.01$). Similar to the results on Dataset1, UD-comp-FBCNN achieves the highest accuracy of 79.05±15.6%. All of the CNN and FBCNN methods outperformed CCA. Among the UD-based methods and UI-based methods, the UD FBCNN and CNN methods outperformed the UI FBCNN and CNN methods respectively. However, the UI-comp-FBCNN acquired the accuracy of 73.52±16.5%, superior to that of UD-mag-CNN and UD-mag-FBCNN. The average accuracies of all the methods for data length of 1s are: CCA: 53.3±21.3%, UD-mag-CNN: 67.45±17.4%, UD-mag-FBCNN: 70.17±16.9%, UD-comp-CNN: 75.26±16.2%, UD-comp-FBCNN: 79.05±15.6%, UI-mag-CNN: 59.62±17.9%, UI-mag-FBCNN: 63.27±18%, UI-comp-CNN: 70.15±16.7%, UI-comp-FBCNN: 73.52± 16.5%. As for UD approach, the average accuracies of FBCNN methods corresponding to inputting magnitude and complex feature respectively are 2.72% and 3.79%, higher than that of CNN methods. As for UI approach, the average accuracies of FBCNN methods corresponding to inputting magnitude and complex feature respectively are 3.66% and 3.37%, higher than that of CNN methods. As is shown in Fig. 9(b), the average ITR(bits/min) for all methods with 1s data length and 0.55s of gaze-shifting are: CCA: 85 bits/min, UD-mag-CNN: 123.3 bits/min, UD-mag-FBCNN: 131.3 bits/min, UD-comp-CNN: 146.9 bits/min, UD-comp-FBCNN: 159.1 bits/min, UI-mag-CNN: 101.4 bits/min, UI-mag-FBCNN: 111.4 bits/min, UI-comp-CNN: 131.2 bits/min, UI-comp-FBCNN: 141.5 bits/min. Fig.10 shows the average classification precision and recall of CCA, CNN methods and FBCNN methods on Dataset2. The average classification precision and recall of all methods are: CCA: 54.2% and 43%, UD-mag-CNN: 71.8% and 62.5%, UD-mag-FBCNN: 73.8% and 68%, UD-comp-CNN: 79.6% and 70%, UD-comp-FBCNN: 82.4% and 76.3%, UI-mag-CNN: 72.6% and 47%, UI-mag-FBCNN: 74.6% and 52.5%, UI-comp-CNN: 78.1% and 62.2%, UI-comp-FBCNN: 80.2% and 67.9%. The comparative results indicate that FBCNN methods are obviously superior to CNN methods.

C. COMPUTATIONAL COMPLEXITY

The total parameters are 647.3K (Dataset1) and 297.9K (Dataset2) for UD/UI-mag-FBCNN, and 134.2K (Dataset1) and 623.2K (Dataset2) for UD/UI-comp-FBCNN. For Dataset1, the overall training time of one epoch are: UD-mag-FBCNN: 10 milliseconds, UD-mag-comp: 10 milliseconds, UD-comp-FBCNN: 1013 milliseconds and UI-comp-FBCNN: 1016 milliseconds. For Dataset2, the overall training time of one epoch are: UD-mag-FBCNN: 11 milliseconds, UD-mag-comp: 12 milliseconds, UD-comp-FBCNN: 4023 milliseconds and UI-comp-FBCNN: 6039 milliseconds. The number of floating point operations (FLOPs) are introduced for the purpose of evaluating the computational complexity of all CNN and FBCNN methods. The FLOPs of each layer are added to calculate the total FLOPs of the network. The FLOPs of CNN methods are: 0.00274G (mag) and 0.00564G (comp), and the FLOPs of FBCNN methods are: 0.0237G (mag) and 0.0493G (comp).

V. DISCUSSIONS

During recent years, the deep learning-based methods have been widely applied in SSVEP classification, fully

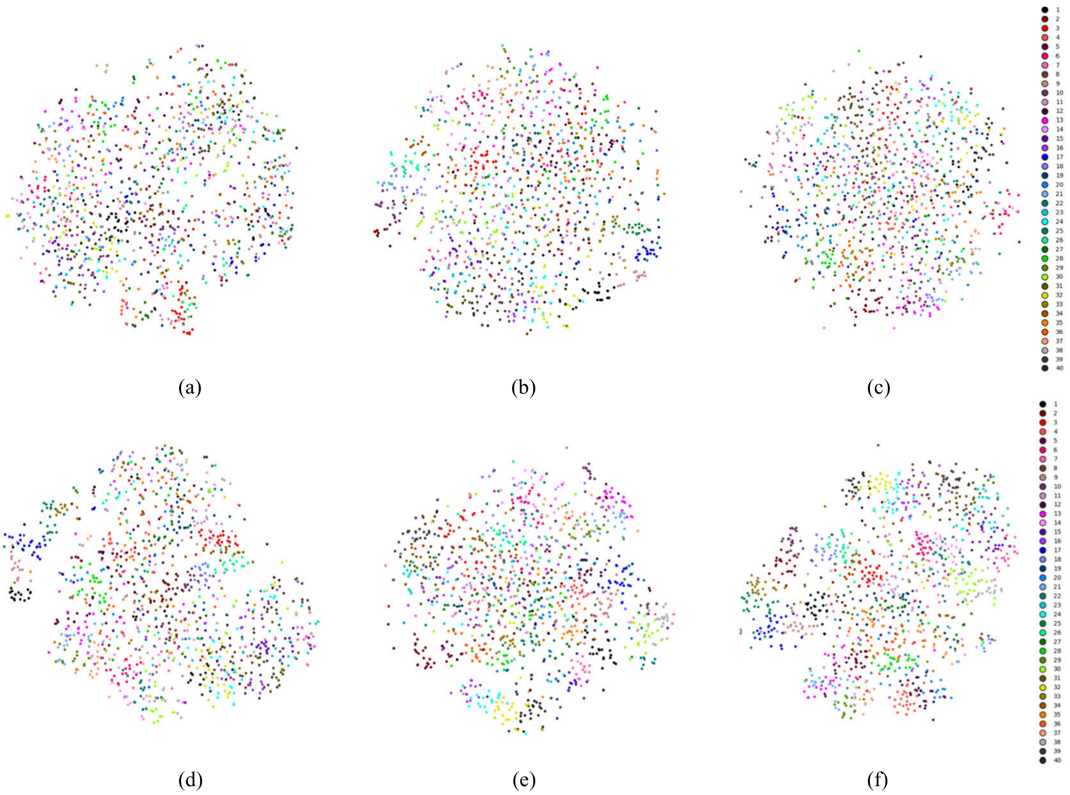

(a)
(b)
(c)
(d)
(e)
(f)

FIGURE 15. **(a-c) Output feature clusters of the three Conv3_Dropout layers of UI-mag-FBCNN and (d-f) output feature clusters of the three add layers of UI-mag-FBCNN.**

revealling the significant advantages over traditional methods in terms of classification accuracy and generalization ability [9], [21]–[24]. However, in the design of neural networks, formerly published studies failed to take into account some prior knowledge which could help to improve the performance of a classification algorithm. Also, we verify the new network on a small dataset (Dataset1 used in this paper, 146MB) and a big dataset (Dataset2 used in this paper, 3.45GB). Fig. 11 and Fig. 12 list accuracy comparisons on the same datasets between methods in this study and previous studies. The UD training methods Combination method (92.78±10.22%) [10], IT-CCA (81.17±18.84%) [10] and the UI training methods Combined-tCCA (75±24%) [34], Compact-CNN (79±15%) [24] are introduced to compare on Dataset1 with 1s data length. The FBCCA (67±18%) [25] is cited to compare on Dataset2 with 1s data length. Compared the proposed FBCNN methods with methods [9], the average accuracies of FBCNN methods are higher than CNN between 0.86% and 3.79% on both datasets. Additionally, comparing each participant in the two datasets one by one, we find that FBCNN methods owns a distinct advantage. The average classification precisions and recalls of CNN methods and FBCNN methods vary significantly on Dataset2. Both average classification precisions and recalls of FBCNN methods are higher than CNN methods. Especially, the recalls of all FBCNN methods are 5.7% higher than CNN methods in average. The above results prove it feasible to improve the classification performance by analyzing the correlation between harmonics. In some cases, the spontaneous EEG signals are much stronger than the SSVEP evoked EEG signals, thus the CNN using single band pay more attention on spontaneous EEG signals while neglecting the harmonics with weaker signal amplitude. We split whole single band into three sub-bands and feed them into parallel subnets of FBCNN to learn the feature of harmonics using each subnet. The contribution of the subnet to the final classification results, i.e. weight of each subnet, can be learned by full connected layers. In spite of an increase in the FLOPs of FBCNN methods comparing with CNN methods, FBCNN still remain a lightweight neural network with low computational complexity.

To intuitively visualize the difference between CNN and FBCNN, the t-Stochastic Neighborhood Embedding (t-SNE) approach is introduced to demonstrate the feature representations of CNN and FBCNN methods. The nonlinear dimensionality reduction algorithm t-SNE offers an access to visualize the high-dimensional features in lower dimensions [35]. UI-mag-CNN and UI-mag-FBCNN validated on S1 and trained on other subjects of Dataset2 are used to extract and demonstrate the patterns. Different colors of the clusters in Fig. 13, Fig. 14 and Fig. 15 stand for different SSVEP target labels. Fig. 13(a) presents the output feature clusters of the Conv2_Flatten layer (second to last layer) of UI-mag-CNN, and Fig. 13(b) shows the output feature clusters of the concatenate layer (second to last layer) of UI-mag-FBCNN. Fig.14 (a) and Fig. 14(b) illustrates the classification result clusters of UI-mag-CNN and UI-mag-FBCNN respectively. To compare the impact after SE and the pair-add-up operation, the output feature clusters of three Conv3_Dropout layers in UI-mag-FBCNN are presented in Fig. 15(a-c), and the output feature clusters of the three add layers of UI-mag-FBCNN are exhibited in Fig. 15(d-f). It can be observed that the output features become more clustered after SE and pair-add-up operation. The results acquired on both public datasets and the feature representations exhibit the effectiveness of FBCNN. The direct feature extraction and learning in CNN method [9], which rely on sole channel CNN structure from SSVEP frequency domain data without prior knowledge-based operation, may not be the most appropriate approach owing to incomplete feature extraction. More information in SSVEP data can be extracted with the help of the filter bank, and the attention mechanism enhances the distinguishability of features. The bandwidth and number of filters are set empirically in this paper, but work on the performance of FBCNN in different number and bandwidth of filters is still ongoing to further implement online SSVEP-based BCI applications.

VI. CONCLUSION

In summary, we propose a novel harmonic-based feature learning method for SSVEP classification, which is based on filer bank and a new CNN architecture. Two public datasets with sufficient stimulus frequencies and participants are introduced to verify the performance of all the methods for comparison. The experimental results demonstrate that UD-comp-FBCNN obtains the best performance among the compared methods and that FBCNN methods perform better than other CNN-based methods on the two datasets.

ACKNOWLEDGMENT

The authors would like to thank Prof. Dr. Jiasong Wu, Prof. Dr. Guanyu Yang, and Prof. Dr. Youyong Kong from Southeast University for their constructive suggestions and help on the network structure and they also would like to thank the improvement of article to valuable comments from all anonymous reviewers.

REFERENCES

[1] J. Wolpaw, N. Birbaumer, D. McFarland, G. Pfurtscheller, and T. Vaughan, "Brain-computer interfaces for communication and control," *Clin. Neurophys.*, vol. 113, no. 6, pp. 767–791, 2002.

[2] A. Nijholt and D. Tan, "Brain-computer interfacing for intelligent systems," *IEEE Intell. Syst.*, vol. 23, no. 3, pp. 72–79, May/Jun. 2008.

[3] Y. Wang, X. Gao, B. Hong, C. Jia, and S. Gao, "Brain-computer interfaces based on visual evoked potentials," *IEEE Eng. Med. Biol. Mag.*, vol. 27, no. 5, pp. 64–71, Sep. 2008.

[4] M. Nakanishi, Y. Wang, Y.-T. Wang, Y. Mitsukura, and T.-P. Jung, "A high-speed brain speller using steady-state visual evoked potentials," *Int. J. Neural Syst.*, vol. 24, no. 6, Sep. 2014, Art. no. 1450019.

[5] R. Zerafa, T. Camilleri, O. Falzon, and K. P. Camilleri, "To train or not to train? A survey on training of feature extraction methods for SSVEP-based BCIs," *J. Neural Eng.*, vol. 15, no. 5, Jul. 2018, Art. no. 051001.

[6] G. R. Müller-Putz, E. Eder, S. C. Wriessnegger, and G. Pfurtscheller, "Comparison of DFT and lock-in amplifier features and search for optimal electrode positions in SSVEP-based BCI," *J. Neurosci. Methods*, vol. 168, no. 1, pp. 174–181, Feb. 2008.

[7] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIS," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 6, pp. 1172–1176, Jun. 2007.

[8] O. Friman, I. Volosyak, and A. Graser, "Multiple channel detection of steady-state visual evoked potentials for brain-computer interfaces," *IEEE Trans. Biomed. Eng.*, vol. 54, no. 4, pp. 742–750, Apr. 2007.

[9] A. Ravi, N. H. Beni, J. Manuel, and N. Jiang, "Comparing user-dependent and user-independent training of CNN for SSVEP BCI," *J. Neural Eng.*, vol. 17, no. 2, Apr. 2020, Art. no. 026028.

[10] M. Nakanishi, Y. Wang, Y.-T. Wang, and T.-P. Jung, "A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials," *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0140703.

[11] Y. Wang, M. Nakanishi, Y.-T. Wang, and T.-P. Jung, "Enhancing detection of steady-state visual evoked potentials using individual training data," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. SoC*, Aug. 2014, pp. 3037–3040.

[12] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, and A. Cichocki, "Multiway canonical correlation analysis for frequency components recognition in SSVEP-based BCIS," in *Proc. 18th Int. Conf. Neural Inf. Process.*, 2011, pp. 287–295.

[13] X. Chen, Y. Wang, S. Gao, T.-P. Jung, and X. Gao, "Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface," *J. Neural Eng.*, vol. 12, no. 4, Aug. 2015, Art. no. 046008.

[14] S. Rehman, S. Tu, O. Rehman, Y. Huang, C. Magurawalage, and C.-C. Chang, "Optimization of CNN through novel training strategy for visual classification problems," *Entropy*, vol. 20, no. 4, p. 290, Apr. 2018.

[15] K. G. van Leeuwen, H. Sun, M. Tabaeizadeh, A. F. Struck, M. J. A. M. van Putten, and M. B. Westover, "Detecting abnormal electroencephalograms using deep convolutional networks," *Clin. Neurophysiol.*, vol. 130, no. 1, pp. 77–84, Jan. 2019.

[16] S. U. Rehman, S. Tu, Y. Huang, and Z. Yang, "Face recognition: A novel un-supervised convolutional neural network method," in *Proc. IEEE Int. Conf. Online Anal. Comput. Sci. (ICOACS)*, May 2016, pp. 139–144.

[17] S. U. Rehman, S. Tu, Y. Huang, and O. U. Rehman, "A benchmark dataset and learning high-level semantic embeddings of multimedia for cross-media retrieval," *IEEE Access*, vol. 6, pp. 67176–67188, 2018.

[18] S. U. Rehman, S. Tu, M. Waqas, Y. Huang, O. U. Rehman, B. Ahmad, and S. Ahmad, "Unsupervised pre-trained filter learning approach for efficient convolution neural network," *Neurocomputing*, vol. 365, pp. 171–190, Nov. 2019.

[19] X. Li, F. Li, X. Zhang, C. Yang, and W. Gui, "Exponential stability analysis for delayed semi-Markovian recurrent neural networks: A homogeneous polynomial approach," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 6374–6384, Dec. 2018.

[20] A. Craik, Y. He, and J. L. Contreras-Vidal, "Deep learning for electroencephalogram (EEG) classification tasks: A review," *J. Neural Eng.*, vol. 16, no. 3, Jun. 2019, Art. no. 031001.

[21] N.-S. Kwak, K.-R. Müller, and S.-W. Lee, "A convolutional neural network for steady state visual evoked potential classification under ambulatory environment," *PLoS ONE*, vol. 12, no. 2, pp. 1–20, 2017.

[22] T.-H. Nguyen and W.-Y. Chung, "A single-channel SSVEP-based BCI speller using deep learning," *IEEE Access*, vol. 7, pp. 1752–1763, 2018.

[23] X. Zhang, G. Xu, X. Mou, A. Ravi, M. Li, Y. Wang, and N. Jiang, "A convolutional neural network for the detection of asynchronous steady state motion visual evoked potential," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 6, pp. 1303–1311, Jun. 2019.

[24] N. Waytowich, V. J. Lawhern, J. O. Garcia, J. Cummings, J. Faller, P. Sajda, and J. M. Vettel, "Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials," *J. Neural Eng.*, vol. 15, no. 6, Dec. 2018, Art. no. 066031.

[25] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain–computer interfaces," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 10, pp. 1746–1752, Oct. 2017.

[26] Y. S. Zhang, E. Yin, F. Li, Y. Zhang, T. Tanaka, Q. Zhao, Y. Cui, P. Xu, D. Yao, and D. Guo, "Two-stage frequency recognition method based on correlated component analysis for SSVEP-based BCI," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 26, no. 7, pp. 1314–1323, Jul. 2018.

[27] Y. Zhang, P. Xu, K. Cheng, and D. Yao, "Multivariate synchronization index for frequency recognition of SSVEP-based brain-computer interface," *J. Neurosci. Meth.*, vol. 221, pp. 32–40, Jan. 2014.

[28] M. Nakanishi, Y. Wang, X. Chen, Y. Wang, X. Gao, and T.-P. Jung, "Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis," *IEEE Trans. Biomed. Eng.*, vol. 65, no. 1, pp. 104–112, Jan. 2018.

[29] X. Chen, Y. Wang, M. Nakanishi, T.-P. Jung, and X. Gao, "Hybrid frequency and phase coding for a high-speed SSVEP-based BCI speller," in *Proc. 36th Annu. Int. Conf. IEEE Eng. Med. Biol. SoC.*, Aug. 2014, pp. 3993–3996.

[30] J. Pan, X. Gao, F. Duan, Z. Yan, and S. Gao, "Enhancing the classification accuracy of steady-state visual evoked potential-based brain–computer interfaces using phase constrained canonical correlation analysis," *J. Neural Eng.*, vol. 8, no. 3, Jun. 2011, Art. no. 036027.

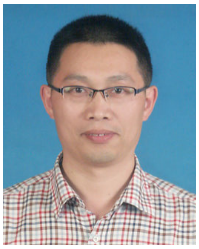
[31] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[32] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn.*, vol. 37, 2015, pp. 448–456.

[33] S. U. Rehman, S. Tu, Y. Huang, and G. Liu, "CSFL: A novel unsupervised convolution neural network approach for visual pattern classification," *AI Commun.*, vol. 30, no. 5, pp. 311–324, Aug. 2017.

[34] N. R. Waytowich, J. Faller, J. O. Garcia, J. M. Vettel, and P. Sajda, "Unsupervised adaptive transfer learning for steady-state visual evoked potential brain-computer interfaces," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2016, pp. 4135–4140.

[35] L. Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.



DECHUN ZHAO received the Ph.D. degree from the Bioengineering College, Chongqing University, in 2008. After that, he was a Teacher with the Chongqing University of Posts and Telecommunications. He became a Professor, in 2016. His research interests include brain–computer interface, micro diagnosis and treatment systems, and electromagnetic safety.

TIAN WANG was born in Chongqing, China, in 1994. He received the B.S. degree from the Chongqing University of Posts and Telecommunications, China, in 2016, where he is currently pursuing the M.S. degree. His research interests include deep learning and brain–computer interface.

YUANYUAN TIAN was born in Gansu, China, in 1997. She received the B.S. degree from the Chongqing University of Posts and Telecommunications, China, in 2019, where she is currently pursuing the M.S. degree. Her current research interests include medical image processing and deep learning.

XIAOMING JIANG (Member, IEEE) received the Ph.D. degree in human science in medical physics from Heidelberg University, Germany, in 2015. He worked as a Postdoctoral Fellow with the Institute of Experimental Physics, Ulm University, Germany, from 2014 to 2015. Since 2016, he has been with the School of Bioinformatics, Chongqing University of Posts and Telecommunications, China. His research interests include biomedical signal/imaging and processing.