IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 24, 2020, accepted January 10, 2021, date of publication January 18, 2021, date of current version January 27, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3052656

EEGNet With Ensemble Learning to Improve the Cross-Session Classification of SSVEP Based BCI From Ear-EEG

YUANLU ZHU[1,2], YING LI[1], JINLING LU[1], AND PENGCHENG LI[1,2,3]

[1]MoE Key Laboratory for Biomedical Photonics, Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology, Wuhan 430074, China
[2]HUST-Suzhou Institute for Brainsmatics, Suzhou 215000, China
[3]School of Biomedical Engineering, Hainan University, Haikou 570228, China

Corresponding author: Jinling Lu (lujinling@mail.hust.edu.cn)

This work was supported in part by the National Key Research and Development Program of China under Grant 2017YFB1002503, in part by the National Natural Science Foundation of China (NSFC) under Grant 61721092, in part by the Fundamental Research Funds for the Central Universities, HUST under Grant 2018KFYXKJC035, and in part by the Director Fund of Wuhan National Laboratory for Optoelectronics.

ABSTRACT Ear-electroencephalography (ear-EEG) using electrodes placed above hairless areas around ears is a convenient and comfortable method for signal recording in practical applications of steady-state visual evoked potential (SSVEP) based brain-computer interface (BCI). However, due to the constraint of electrode distribution behind the ear, the amplitude of SSVEP in ear-EEG signals is relatively low, which hinders the application of ear-EEG in SSVEP-based BCI. This study was aimed to improve the performance of ear-EEG in SSVEP-based BCI through re-implementing a compact convolutional neural network (EEGNet) with ensemble learning. We first evaluated the feasibility of applying widely used EEGNet models with different kernel numbers to decode SSVEP in ear-EEG signals. Then we applied an ensemble learning strategy to combine EEGNet models with different kernel numbers to improve the classification of ear-EEG signals. The ear-EEG data was from an open dataset, which acquired three sessions of SSVEP data induced by three flicker stimuli from eleven subjects. The average accuracy of EEGNet with ensemble learning for ear-EEG signals in cross-session validations at 1 s window length was 81.12% (from session 1 to session 2) and 81.74% (from session 1 to session 3), which significantly outperformed canonical correlation analysis (CCA). In addition, the network visualization indicated that EEGNet extracted features related to stimulation frequencies. The results showed promise for accurate classification of SSVEP in ear-EEG signals using deep learning models with strategies, helping to promote the SSVEP based BCI from laboratory to practical application.

INDEX TERMS Brain-computer interface (BCI), convolutional neural network (CNN), ear-electroencephalography (ear-EEG), steady-state visual evoked potential (SSVEP).

I. INTRODUCTION

Brain-computer interfaces (BCIs) are designed as a bridge to construct direct communication between the brain and external devices without relying on normal peripheral nerves and muscle tissue [1]. Electroencephalography (EEG) is one of the most widely used non-invasive methods in BCI for its low cost, portability and high temporal resolution. Several types of physiological paradigms are usually chosen to generate the output commands of the EEG-based BCI, such as motor imagery (MI) [2], P300 [3], and steady-state visual evoked potential (SSVEP) [4]. Among them, SSVEP has gained a lot of attention in BCI for its characteristics of less training, high classification accuracy and high information transfer rates (ITR) [5]. Previous studies have shown that SSVEP induced by periodic visual stimulus contains brain responses in occipital cortex at the stimulation frequency and its harmonic frequencies [6]. Therefore, SSVEP-based BCI mostly acquired EEG signals from electrodes placed above the occipital brain region (scalp-EEG), which often required a long time of preparation leading to the inconvenience in some application scenarios, such as lying patients. These

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

limitations may hinder the promotion of SSVEP-based BCI to practical applications.

Recently, several studies attempted to develop devices to detect SSVEP based on the ear-EEG which acquired signals around ears or in the external ear canal. These devices were easy to wear, and suitable for daily use with short time required for preparation. Wang *et al.* were the first to conduct offline and online experiments to evaluate the feasibility of decoding SSVEP from hairless areas. They recorded SSVEP signals both from the occipital brain region and the non-hair-bearing areas including the face, behind-ears, and neck areas [7], [8]. The results showed that SNRs of SSVEP from high to low were from the occipital, behind-the-ear, neck, and face areas, which illustrated the potential of using ear-EEG for SSVEP based BCI. Norton *et al.* introduced a soft, curved electrode systems for long-term BCI systems, which were able to adhere to the surface of the auricle and the mastoid to measure ear-EEG signals [9]. They used a speller task based on SSVEP or P300 signals to demonstrate the feasibility of the proposed electrode systems. Kappel *et al.* developed dry-contact soft-earpiece electrodes to realize long-term brain signal monitor in BCI. [10]. Bleichner *et al.* proposed a wireless device cEEGrids to record signals around ears [11]. They compared the classification of ear-EEG and scalp-EEG signals acquired simultaneously in spatial auditory attention task. The comparable results indicated the feasibility of measuring event-related potentials by ear-EEG. To improve the decoding accuracy, Kwak *et al.* proposed an error correction regression method [12] which utilized the scalp-EEG signals to enhance SSVEP in ear-EEG signals. Their proposed method significantly outperformed canonical correlation analysis (CCA) and other regression methods on decoding SSVEP in ear-EEG signals.

As shown in previous studies, the amplitude of SSVEP in ear-EEG signals is far less than that in scalp-EEG signals. Moreover, owing to the non-stationarity and high individual variation of ear-EEG signals, traditional methods using machine learning or statistical methods with fixed parameters did not work well on ear-EEG signals which limited the promotion of ear-EEG in BCI. Therefore, it is necessary to improve the classification accuracy of ear-EEG signals.

In recent years, convolutional neural network (CNN) have achieved satisfactory performance on computer vision and speech recognition [13], [14]. Different from traditional methods, CNN combined automatic feature extraction and classification to form an end-to-end decoding method. Several studies have applied CNN to enhance the performance on classification of scalp-EEG signals. Cecotti was the first to combined CNN and fast fourier transform (FFT) to decode SSVEP signals without special pre-processing [15]. The average classification accuracy at 1 s window length was 95.61%. To decode motion related information from EEG signals, Schirrmeister *et al.* designed deep and shallow CNN and implemented the visualization of brain mapping to verify the potential of CNN [16]. To promote practical applications of BCI, there were several studies used CNN to accurately classify SSVEP acquired from dry-EEG headset or single-channel electrode [17], [18]. For simulating real-life application scenarios, Kwak *et al.* used CNN to classify SSVEP acquired under static and ambulatory environment to evaluate the accommodation of CNN for EEG movement artefacts [19]. Podmore *et al.* proposed a deep convolutional neural networks (DCNNs) architecture to classify an open source SSVEP dataset which included 40 stimuli for speller task [20]. The DCNNs achieved 87% offline accuracy at 6 s window length. Waytowich *et al.* proposed EEGNet with compact structure to accurately classify EEG signals from different BCI paradigms [21], [22]. The average accuracy of a 12-class SSVEP dataset at 1 s time segment was approximately 80%, which outperformed CCA. Additionally, network visualization illustrated the feasibility of implementing asynchronous BCI with EEGNet. In total, previous studies have demonstrated that CNN designed by appropriate network architecture performed better than traditional methods on classifying SSVEP in scalp-EEG signals.

In view of good performance of CNN on the classification of scalp-EEG signals, it is worth trying to decode low SNR SSVEP in ear-EEG signals with CNN. Considering that EEGNet with compact structure required fewer training weights which was suitable for small labeled datasets, we attempted to re-implement the proposed EEGNet [21] to classify SSVEP-based BCI from ear-EEG. In [21], the EEGNet was designed with fixed kernel numbers for all subjects. Due to the individual variations between subjects, the fixed kernel number might not be optimal for each subject. Therefore, we first evaluated the performance of EEGNet with different kernel numbers for each subject. Besides, considering that the weak SSVEP in ear-EEG signals made it difficult to separate useful information from background noise, we applied ensemble learning strategy [23] which combined EEGNet with different kernel numbers to improve the classification accuracy of ear-EEG signals. The data we used were from an open dataset [12]. To evaluate the feasibility of EEGNet to classify ear-EEG signals, we applied intra-subject validation at 1 s window length and compared the performance of EEGNet with ensemble learning and CCA. Additionally, we implemented network visualization to interpret EEGNet in the process of ear-EEG signals.

II. MATERIALS AND METHODS

A. DATA FOR CLASSIFICATION

An open dataset from [12] was used to evaluate the performance of EEGNet on classifying SSVEP based BCI from ear-EEG. In their experiments, eleven subjects were arranged to sit in a comfortable chair in front of the screen at a distance of approximately 60 cm, each of whom conducted two offline sessions and one online session experiments at different days. There were three stimuli flickered at 10, 8.75, and 7.5 Hz on LCD screen. For the two offline sessions, scalp-EEG signals and ear-EEG signals were recorded simultaneously from two separate devices. Subjects were asked to attend

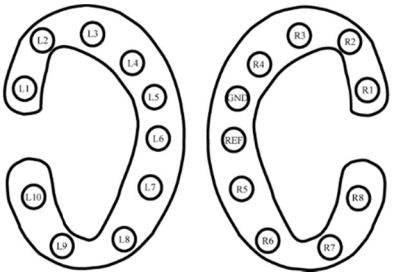


FIGURE 1. The topography distribution of the ear-EEG electrodes.

to each stimulus 50 times at a random sequence for a total of 150 trials. Each trial contained 2 s of cue, 6 s of flicker, and 2 s of rest. For the online session, only ear-EEG signals were recorded for classification and feedback. Subjects were asked to attend to each stimulus 20 times at a random sequence for a total of 60 trials. Each trial contained 2 s of cue, 3 s of flicker, and 2 s of rest. During the experiments, scalp-EEG signals were recorded from BrainAmp with eight channels (PO7, PO3, POz, PO4, PO8, O1, Oz, and O2) according to the International 10-20 system. Ear-EEG signals were recorded from a wireless device attached cEEGrid electrodes around each ear. The electrodes included twenty channels, ten channels around each ear as shown in Fig. 1. Two channels were used as ground and reference, and the remaining eighteen channels were used for data acquisition. The sampling frequency of both devices was 500Hz. A band-pass filter at 0.3-50 Hz and a 60 Hz notch filter were applied in the data acquisition process.

B. PREPROCESSING

The data of ear-EEG signals were first downsampled to 250Hz to reduce the input dimension. And then we applied a 6-50 Hz bandpass filter to remain components related to the stimulation frequencies and its harmonic frequencies and to remove physiological and environmental noise. The band-pass filtered data without other artifact correction preprocessing were directly used to evaluate the performance of deep learning model on classifying SSVEP in ear-EEG signals.

C. DATA AUGMENTATION

The data of flicker stage were segmented using a 1 s moving window with a slide step of 100 ms for data augmentation. The segmented data were then normalized between -1 and $+1$ before using as the inputs of EEGNet models.

D. DEEP LEARNING MODEL

Due to the high amplitude of SSVEP from occipital cortex, the CNN models consisted of two convolutional blocks were enough to achieve good performance on classification of scalp-EEG signals in previous studies [15], [24]. Given that the relatively low amplitude of SSVEP in ear-EEG signals, we re-implemented EEGNet [22] with three convolutional blocks for classification and the kernel sizes were modified according to the size of ear-EEG signals.

As shown in Fig. 2, EEGNet in the dotted box contained three sequential convolutional blocks. The size of network

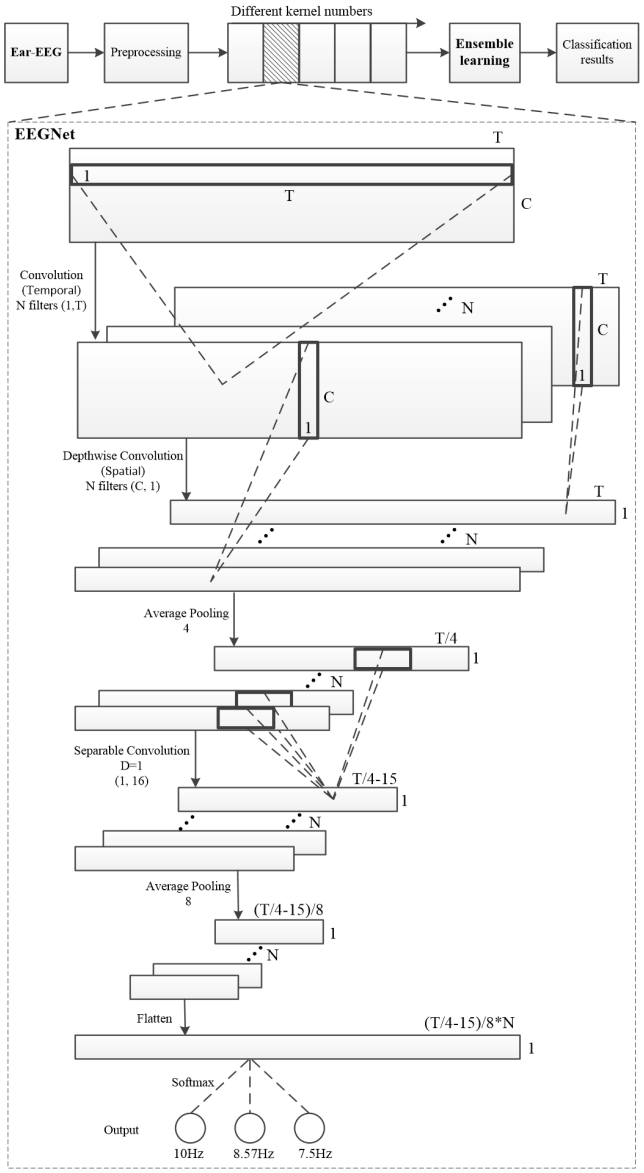


FIGURE 2. The architecture of the EEGNet with ensemble learning strategy. EEGNet in the dotted box contained three convolutional layers, the kernel size of which is set as (X, Y), representing the kernel size in dimension of channel (X), and sample points (Y), respectively. The black bold box represents a convolutional kernel. The specific kernel size of each convolutional layer is shown in the box.

input was C $\times$ T, where C represented the number of channels of the input signals and T indicated the number of sampling points in the 1 s moving window. As for 1 s window length with 250 Hz sample frequency, T was set as 250.The black bold box represented a convolutional kernel. The first block used traditional convolutional layer acted as temporal filters. The second block applied depthwise convolutional layer acted as spatial filters to integrate the information of electrode locations, in which kernel size was set as (C, 1). The depthwise convolutional layer, which independently extracted features from the outputs corresponding to each of the previous convolutional kernels, required fewer training weights, so it was able to further extract information without

increasing the complexity of the network [13], [25]. The first two convolutional blocks separately extracted temporal and spatial features to avoid information mixing. The average pooling layer was used to introduce invariance and reduce redundancy [26]. Each pooling layer in EEGNet was followed by a dropout to avoid overfitting. After that, a separable convolutional layer in the third block was designed to further extract features from the temporal and spatial filtered data. In the structure of EEGNet, batch normalization was applied after each convolutional layer, which is conducive to network convergence [27]. At last, the extracted features passed through flatten, dense, and softmax layers to output the final classification results.

E. ENSEMBLE LEARNING

Ensemble learning is based on building and combining base models to improve the classification performance [23]. Through utilizing the diversity of base models, ensemble learning is able to turn weak classifiers into strong classifiers. Previous studies [21], [22] demonstrated the feasibility of EEGNet to classify SSVEP based BCI from scalp-EEG. In [21], they used a fixed kernel number in EEGNet for all subjects. In the [20], they compared two EEGNet models with two small kernel numbers and found no significant difference in the performance of the two EEGNet models. The results were based on limited kernel numbers, and the kernel number setting of EEGNet was still dependent on experience. Therefore, we first evaluated a series of kernel numbers in the training of EEGNet to form base models, and then combined the trained EEGNet models through ensemble learning [23] to improve the classification accuracy of ear-EEG signals. The kernel number of base models in this study were set to 12, 24, 48, 96, and 144. A small number of kernels may not extract enough information but the network requires less training weights. Large number of kernels can obtain more information but may cause overfitting. We selected averaging as the combining strategy to determine the final results of ensemble learning.

F. INTRA-SUBJECT VALIDATION

Considering that each subject collected a total of 3 sessions of SSVEP data on different days, we performed cross-session validation for each subject, which trained the network based on data of first session and used data of the second and the third session to predict. The validation was able to evaluate the robustness of the network, taking into account that the acquired ear-EEG signals of the same subject on different days might fluctuate due to the environmental noise, differences in device wearing and the concentration of the subject.

The network training process used 5-fold-cross-validation. The data of the first session were divided into 5 parts in time sequence. Each part was used as the validation set in turn while the rest four parts were used to train the network. The weights of network were trained specifically for each subject considering the variance between subjects. The trained network was used to predict the data of session 2 and session 3. The final accuracy was obtained by averaging the results of all trained networks in 5-fold-cross-validation.

III. EXPERIMENTS

In this paper, the inputs of EEGNet models ware divided into training, validation, and test set. Among them, the test set was only used to finally evaluate the performance of network without participating in the training process, and the validation set was applied to determine whether to perform early stopping in the training process.

A. TRAINING PARAMETERS

We implemented the network design with keras framework [28]. The same hyperparameters were used for the training of all EEGNet models in this study. Specifically, the learning rate was set at 0.001, the batch size was 64, and the dropout rate was 0.5. The optimization algorithm was Adam, the activation function was exponential linear unit, and the cross-entropy function was used as the loss function [29]. The network training used early stopping strategy. The maximum training epoch was set to 100. When the validation loss of 10 consecutive epochs did not fall more than 0.001, the training process would be stopped, and the trained weights with the lowest validation loss in training process would be saved to predict the test set.

B. TRADITIONAL METHOD

This study adopted CCA [30] as the traditional method for comparation, which has been widely applied in the classification of SSVEP based BCI. Based on the characteristics of SSVEP, CCA method calculates the correlation between test signals and predefined sinusoidal reference signals for classification.

C. VISUALIZATION

We compared the mean amplitude spectrums and SNRs of ear-EEG signals and scalp-EEG signals at each stimulation frequency. Considering the time-locked and phase-locked characteristics of SSVEP, we first averaged the 50 trials of each stimulation frequency at each channel for each subject in the first session to improve the SNR of SSVEP. And then we conducted FFT to calculate the amplitude spectrums of the averaged trials at each channel for each subject. The mean amplitude spectrums $y(f)$ of ear-EEG and scalp-EEG signals were obtained by averaging across the amplitude spectrums of 18 and 8 channels for 11 subjects. The SNR of SSVEP was obtained by dividing the amplitude at the stimulation frequency by mean amplitude of 8 neighboring frequencies using the mean amplitude spectrum as follows:

$$SNR = 20log_{10}\frac{4 \times y(f)}{\sum_{k=1}^{4}[y(f - 0.25 \times k) + y(f + 0.25 \times k)]}$$

To verify whether the features extracted by the EEGNet contained spectrum information related to stimulation frequencies, we performed FFT and time-frequency analysis on spatially and temporally filtered signals corresponding to all

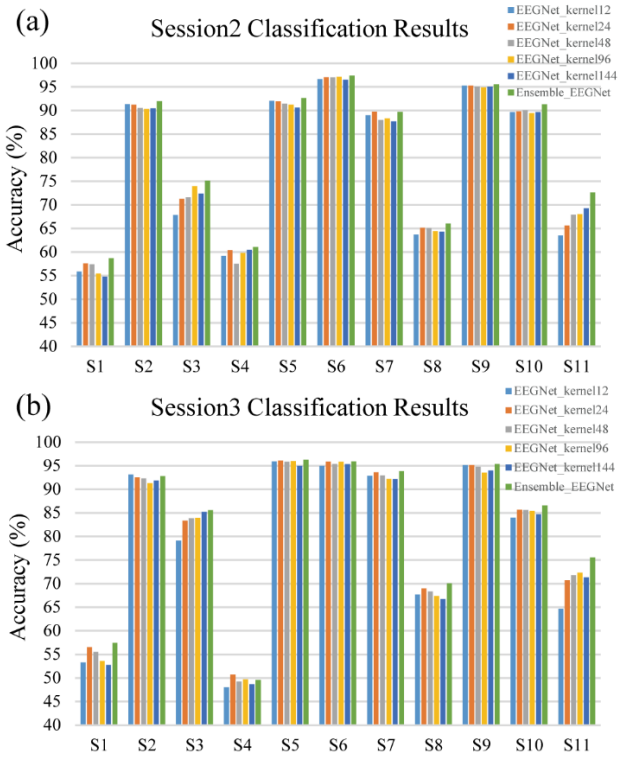


FIGURE 3. The classification results of EEGNet models with different kernel numbers and ensemble learning for each subject at 1 s window length in cross-session validations from session 1 to session 2 (a) and session 3 (b).

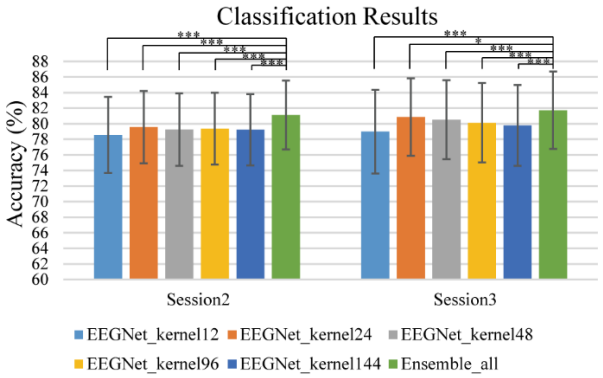


FIGURE 4. The average classification results of EEGNet with different kernel numbers and ensemble learning in cross-session validations from session 1 to session 2 and session 3. The error bar indicated the standard error. We used the Wilcoxon signed-rank test. *** indicated the $p < 0.01$. * indicated the $p < 0.05$.

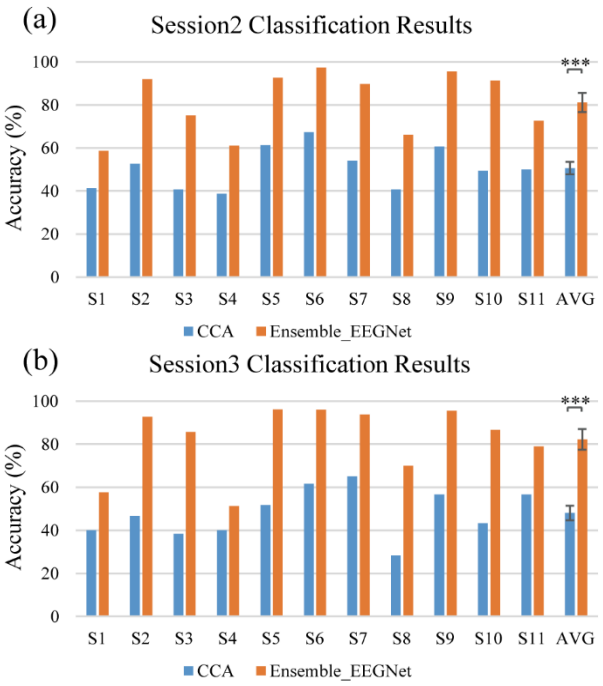


FIGURE 5. The classification results of CCA and EEGNet with ensemble learning for each subject at 1 s window length in cross-session validation of session 2 (a) and session 3 (b). The error bar indicated the standard error. We used the Wilcoxon signed-rank test. *** indicated the $p < 0.01$.

kernels of the network. In addition, T-distributed Stochastic Neighbor Embedding (t-SNE) [31] was used to reduce feature dimensions of the inputs and outputs of each convolutional layer to observe the separability of the data after the process of EEGNet.

IV. RESULT

A. INTRA-SUBJECT VALIDATION

For intra-subject validation, since the data of each subject were divided into training and validation sets for network training and test set for classification, the network performance was not affected by the variance between subjects. We first trained EEGNet models with different kernel numbers, and then applied the ensemble learning to combine the trained models.

Fig. 3 showed the accuracy of EEGNet models with different kernel numbers and the ensemble learning for each subject at 1 s window length in the cross-session validation. Fig. 3(a) showed the results of validation from session 1 to session 2 and Fig. 3(b) showed the results of validation from session 1 to session 3. As shown in Fig. 3, the optimal kernel number of EEGNet was different for each subject. The average classification accuracy of EEGNet models with different kernel numbers and ensemble learning were shown in Fig. 4. The results indicated that the classification accuracy was not always improved with the increase of kernel number of EEGNet. For both sessioin 2 and session 3, when the kernel number of EEGNet reached 24 and continued to increase, the classification accuracy decreased instead. The average results indicated that the performance of EEGNet was sufficient for a small kernel number which is consistence with [21]. The averaged accuracy of EEGNet with ensemble learning in validation from session 1 to session 2 and session 3 at 1 s window length was 81.12% and 81.74%, respectively. The results demonstrated that EEGNet with ensemble learning performed better than EEGNet with single kernel number in Wilcoxon signed-rank test ($p < 0.05$).

We further compared the performance between EEGNet with ensemble learning and CCA method in Fig. 5.

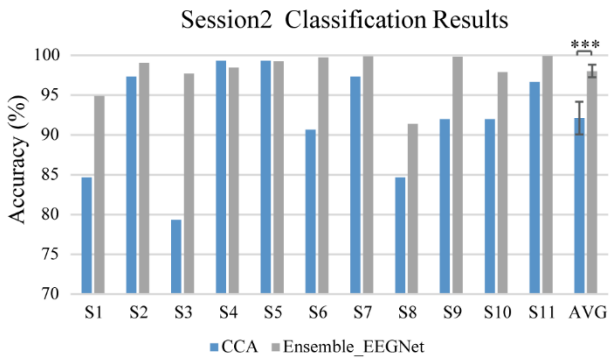


FIGURE 6. The classification accuracy of scalp-EEG signals using CCA and EEGNet with ensemble learning for each subject at 1 s window length in cross-session validation from session 1 to session 2. The error bar indicated the standard error. We used the Wilcoxon signed-rank test. *** indicated the $p < 0.01$.

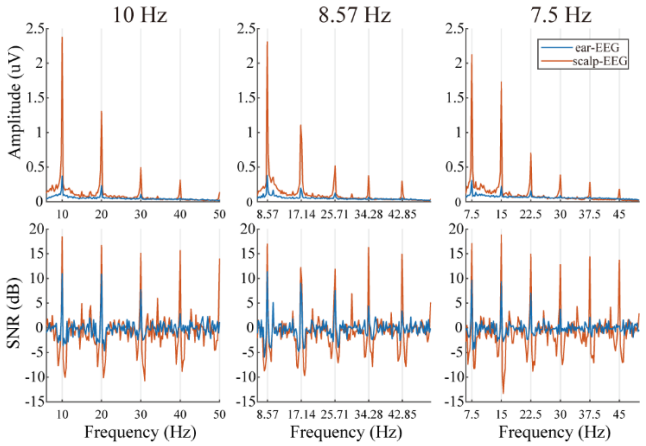


FIGURE 7. The mean amplitude spectrums and SNRs of ear-EEG signals and scalp-EEG signals for each stimulation frequency.

Specifically, Fig. 5(a) and Fig. 5(b) separately showed the classification results of each subject and the average accuracy in session 2 and session 3. We used the Wilcoxon signed-rank test for statistical analysis. EEGNet with ensemble learning significantly outperformed CCA method ($p < 0.01$). The average accuracy at 1 s window length increased from 50.61% to 81.12% in session 2 and from 48.03% to 81.74% in session 3. It was worth mentioning that the EEGNet with ensemble learning achieved satisfactory classification results at 1 s window length, which validated the potential of the EEGNet to decode SSVEP in ear-EEG signals even at a short time window.

We also evaluated the performance of EEGNet with ensemble learning using scalp-EEG signals of each subject at 1 s window length in cross-session validation. The classification results of scalp-EEG signals in session 2 were compared with CCA method. As shown in Fig. 6, EEGNet with ensemble learning also significantly performed better than CCA in the classification of scalp-EEG signals.

B. VISUALIZATION

1) AMPLITUDE SPECTRUM ANALYSIS

As for the amplitude spectrum, the best comparison should be between occipital and temporal EEG signals, but the temporal EEG signals were not provided from the open dataset. Thus, we attempted to compare the amplitude spectrums of ear-EEG signals and scalp-EEG signals. Fig. 7 showed the mean amplitude spectrums and SNRs of ear-EEG signals and scalp-EEG signals for each stimulation frequency. Spectrum and SNR of ear-EEG signals had peaks at stimulation frequencies and its harmonics, but the amplitudes were smaller than that of scalp-EEG signals.

To better interpret the EEGNet model, we attempted to visualize the trained EEGNet using the data with best classification results in intra-subject validation (Subject 6). To evaluate the feasibility of EEGNet to decode SSVEP at short time window, we used the data from the first second of trials in session 1 as input of the trained network. The kernels outputs of the second convolutional layer of the trained network were extracted for FFT and time-frequency analysis. The final results were obtained by averaging the trial results of the same stimulation frequency. The results showed that the kernel outputs contained information of the stimulation frequencies. We selected the output of a kernel for each stimulation frequency as shown in Fig. 8. The responses to stimuli flickered at 10 Hz and 8.75 Hz had higher amplitude than the response to stimulus flickered at 7.5 Hz. The results demonstrated that EEGNet extracted the features related to the stimulation frequencies in ear-EEG signals.

2) NETWORK VISUALIZATION

In order to evaluate the separability of ear-EEG signals responded to different stimuli in the process of EEGNet, we adopted t-SNE algorithm which reduced the features of each trial to two dimensions. Then, scatter plot was used to show the t-SNE results of all trails. Each point represented a single trial, and color of the point corresponded to the label of the trial, representing the stimulus type. The network used for visualization is the trained EEGNet with kernel number set to 24 in intra-subject validation with best classification accuracy (Subject 6). We used the 1 s window length with a slide step of 1 s of the training set in session 1 as the input of network visualization. The sliding window segmented a trial of 6 s into six trials.

Fig. 9(a) showed the t-SNE results of input signals. All types of trials were almost mixed together, which was difficult to be separated by a simple linear discriminator. Fig. 9(b) showed the results after the first convolutional layer. Compared to the input signals, the points of the same type started to gather, but there were still a lot of overlaps. Fig. 9(c) was the results after the second convolutional layer. The same type continued to aggregate, and the overlapping portion was gradually reduced. Fig. 9(d) was the results after the third convolutional layer. Different types of trials were able to be distinguished from each other. Fig. 9(e) represented the

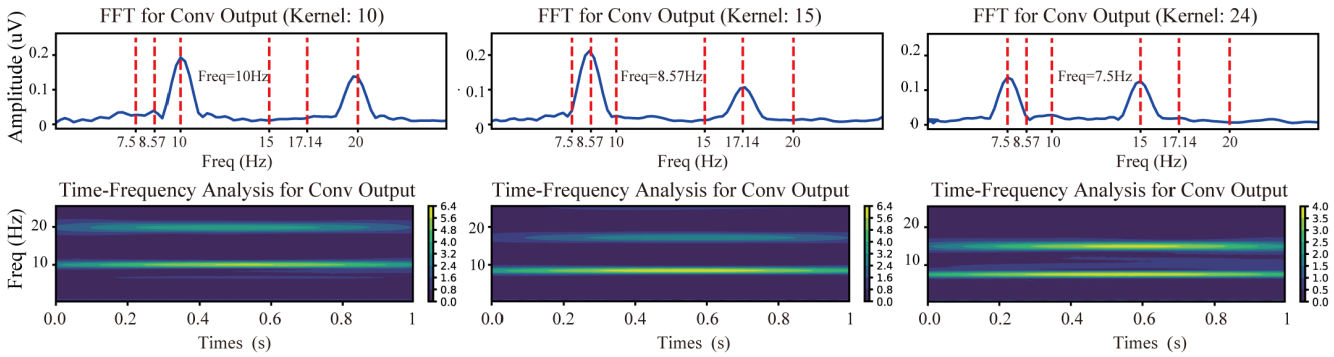


FIGURE 8. The averaged FFT and time-frequency analysis of trials corresponding to each stimulation frequency from a kernel output of the second convolutional layer of the trained EEGNet with kernel number set to 24. The input data were from the first second of trials in session 1 from Subject 6 who received best classification accuracy result.

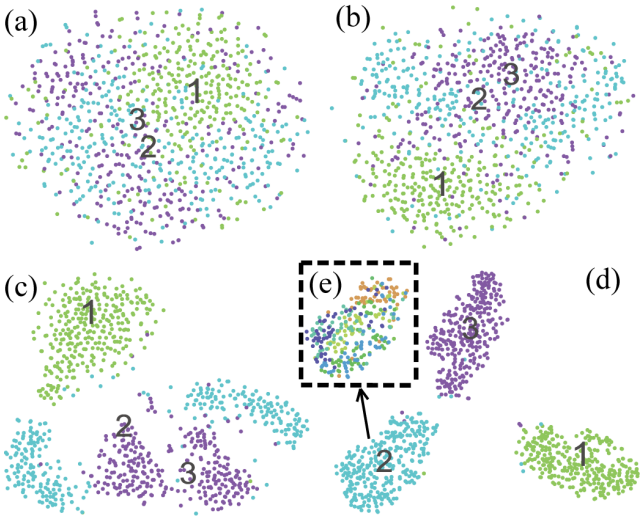


FIGURE 9. The t-SNE results of inputs and outputs of the three convolutional blocks of the trained EEGNet with kernel number set to 24 in intra-subject validation (Subject 6). We applied 1 s window length with 1 s slide step for trials with 6 s duration in session 1 to form the input signals. The subfigures represent the results of raw input (a), the output of first convolutional block (b), the output of second convolutional block (c), the output of third convolutional block (d). And (e) represented the stimulus type 2 with trials labeled in six different colors according to the sliding windows in time sequence.

stimulus type 2 which labeled the trials from each sliding window in time sequence as six different colors. Trials from the same sliding window tended to congregate which was in consistency with [22], illustrating that EEGNet had the ability to learn the phase information. As shown in Fig. 9, the data after the process of EEGNet were clearly separated into different stimulus types, indicating the effectiveness of EEGNet for the classification of SSVEP in ear-EEG signals.

V. DISCUSSION

This study was the first attempt to implement deep learning models to improve the classification of SSVEP in ear-EEG signals. Although ear-EEG signals are not measured from the area with the strongest SSVEP response, ear-EEG placed electrodes above the hairless area is selected due to its convenience for long-term and real-time application. Considering that the SSVEP measured from hairless area was relatively weak, we re-implemented the EEGNet for the ability of end-to-end feature extraction and classification with compact structure. We first evaluated the performance of EEGNet with different kernel numbers, and then applied the ensemble learning strategy which combined EEGNet models with different kernel numbers to further improve the classification of ear-EEG signals.

The dataset used in this study came from [12]. The authors in [12] proposed an error correction regression framework which took advantage of occipital EEG signals to artificially enhance the response of SSVEP in ear-EEG signals, and the processed signals were finally classified by CCA method. The regression framework increased the classification accuracy of ear-EEG signals from 80.85% to 90.3% at 6 s time window in validation from session 1 to session 2, whereas the accuracy of ear-EEG signals at 1 s window length was still below 60%. This study was aimed to improve the classification accuracy of ear-EEG signals at 1 s window length to promote the practical application of SSVEP based BCI. Given that previous studies implemented CNN models in BCI and achieved high classification accuracy of scalp-EEG signals, we attempted to re-implement the compact EEGNet to classify SSVEP in ear-EEG signals in cross-session validation.

EEGNet is a compact CNN model which is suitable for the auto feature extraction and classification of small labelled dataset [21], [22]. In this study, we re-implemented EEGNet with small modification according to the size of ear-EEG dataset. Given that there were few studies focused on the influence of kernel numbers on the deep learning models and most studies set a fixed kernel number empirically for all subjects, we first evaluated the performance of EEGNet with different kernel numbers. As shown in Fig. 3, the optimal kernel number of EEGNet was different for each subject due to individual variation. Therefore, a fixed kernel number for all subjects might ignore the variations between subjects. Besides, the performance of a network with a single kernel number for the classification of ear-EEG signals was not good

TABLE 1. The classification accuracy of ear-EEG for all subjects in validation from session 1 to session 2 at 1 s window length using EEGNet with kernel number set to 24 and different kernel sizes.

Accuracy (%) / Kernel Size / Subjects	32	64	125	250
S1	49.14	52.42	56.20	**57.59**
S2	87.27	89.41	89.88	**91.23**
S3	68.2	69.56	**71.53**	71.27
S4	53.01	57.06	59.05	**60.41**
S5	87.64	89.50	90.34	**91.94**
S6	96.95	96.84	97.05	**97.10**
S7	87.41	88.33	88.66	**89.77**
S8	58.06	62.21	62.85	**65.16**
S9	92.23	93.89	95.19	**95.28**
S10	86.58	88.35	88.54	**89.81**
S11	50.67	57.17	61.47	**65.64**
Mean±Std	74.29±18.57	76.79±17.07	78.25±15.98	**79.56±15.40**

enough as shown in Fig. 4. Considering that ensemble learning had the ability to improve the classification through the combination of weak classifiers, we used the trained EEGNet models with different kernel numbers as base models for ensemble learning to utilize the diversity of different models. Fig. 4 showed that the average classification accuracy of EEGNet with ensemble learning at 1 s window length was above 80%, which was significantly higher than that of EEGNet with single kernel number. We also compared the performance of EEGNet with ensemble learning and CCA method. Fig. 5 illustrated that EEGNet with ensemble learning significantly outperformed traditional CCA method which had been wildly used in the classification of SSVEP in scalp-EEG signals.CCA used the correlation between the test signals and the predefined reference signals for classification, but due to ear-EEG signals had a low amplitude of SSVEP as shown in Fig. 7, it may be difficult to accurately decode SSEVP in ear-EEG signals just by calculating the correlation. As for network visualization, the results shown in Fig. 8 demonstrated that EEGNet extracted task-related features. The separability of the data from different stimulus types was progressively enhanced in the process of EEGNet as shown in Fig. 9.

We also evaluated the influence of kernel size on the performance of EEGNet. Table 1 showed the classification results of ear-EEG signals for all subjects in validation from session 1 to session 2 at 1 s window length using EEGNet with kernel number set to 24 and different kernel sizes of the first convolutional layer which acted as temporal filters.

The accuracy increased with the kernel size of EEGNet. Given the compact structure of EEGNet and the weak response in data of 1 s window length, the larger the kernel size, the more information can be learned and extracted from the first convolutional layer. Considering the low amplitude of SSVEP in ear-EEG signals, the temporal and spatial filters were important for the feature extraction. Therefore, we chose 250 as the kernel size of all EEGNet models in this study.

Although the SSVEP in ear-EEG signals is relatively weak as shown in Fig. 7, the results of cross-session validation and network visualization validated the robustness and effectiveness of EEGNet with ensemble learning to accurately decode SSVEP in ear-EEG signals at a short window length. Compared with traditional SSVEP based BCI which acquired signals from electrodes placed above occipital brain region, the measurement of ear-EEG signals from hairless areas can simplify the preparation of BCI to improve the user experience and to promote the practical application of SSVEP based BCI. Future research will focus on the detection of SSVEP from other hairless areas, and explore the optimal distribution of electrodes placed above hairless area. Besides, other deep learning models and learning strategies will be applied in online experiments to control external devices for practical promotion.

VI. CONCLUSION

This study was the first attempt to implement the deep learning model to decode SSVEP in ear-EEG signals. We evaluated the influence of kernel numbers on performance of EEGNet, and implemented ensemble learning to improve classification accuracy. We compared the EEGNet with ensemble learning and CCA in cross-session validation. In addition, we attempted to interpret the network by visualization. The results showed that the average accuracy of EEGNet with ensemble learning at 1 s time window was above 80%, which demonstrated the ability of deep learning model to accurately classify SSVEP from hairless areas. The good performance of EEGNet with ensemble learning indicated the potential of promoting SSVEP based BCI to practical application by using ear-EEG.

ACKNOWLEDGMENT

The authors would like to thank Seong-Whan Lee and Vernon J Lawhern for their open data and codes.

REFERENCES

[1] J. R. Wolpaw, N. Birbaumer, W. J. Heetderks, D. J. McFarland, P. H. Peckham, G. Schalk, E. Donchin, L. A. Quatrano, C. J. Robinson, and T. M. Vaughan, "Brain-computer interface technology: A review of the first international meeting," *IEEE Trans. Rehabil. Eng.*, vol. 8, no. 2, pp. 164–173, Jun. 2000.

[2] J. R. Wolpaw, D. J. McFarland, G. W. Neat, and C. A. Forneris, "An EEG-based brain-computer interface for cursor control," *Electroencephalogr. Clin. Neurophysiol.*, vol. 78, no. 3, pp. 252–259, Mar. 1991.

[3] L. A. Farwell and E. Donchin, "Talking off the top of your head: Toward a mental prosthesis utilizing event-related brain potentials," *Electroencephalogr. Clin. Neurophysiol.*, vol. 70, no. 6, pp. 510–523, Dec. 1988.

[4] M. Cheng, X. Gao, S. Gao, and D. Xu, "Design and implementation of a brain-computer interface with high transfer rates," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 10, pp. 1181–1186, Oct. 2002.

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

[6] C. S. Herrmann, "Human EEG responses to 1–100 Hz flicker: Resonance phenomena in visual cortex and their potential correlation to cognitive phenomena," *Exp. Brain Res.*, vol. 137, nos. 3–4, pp. 346–353, 2001.

[7] Y.-T. Wang, Y. Wang, C.-K. Cheng, and T.-P. Jung, "Measuring steady-state visual evoked potentials from non-hair-bearing areas," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc.*, Aug. 2012, pp. 1806–1809.

[8] Y. Te Wang, M. Nakanishi, Y. Wang, C. S. Wei, C. K. Cheng, and T. P. Jung, "An online brain-computer interface based on SSVEPs measured from non-hair-bearing areas," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 25, no. 1, pp. 14–21, Jan. 2017.

[9] J. J. S. Norton, D. S. Lee, J. W. Lee, and W. Lee, "Soft, curved electrode systems capable of integration on the auricle as a persistent brain–computer interface," *Proc. Nat. Academy Sci. USA*, vol. 112, no. 13, pp. 3920–3925, 2015.

[10] S. L. Kappel, M. L. Rank, H. O. Toft, M. Andersen, and P. Kidmose, "Dry-contact electrode ear-EEG," *IEEE Trans. Biomed. Eng.*, vol. 66, no. 1, pp. 150–158, Jan. 2019.

[11] M. G. Bleichner, B. Mirkovic, and S. Debener, "Identifying auditory attention with ear-EEG: cEEGrid versus high-density cap-EEG comparison," *J. Neural Eng.*, vol. 13, no. 6, 2016, Art. no. 066004.

[12] N.-S. Kwak and S.-W. Lee, "Error correction regression framework for enhancing the decoding accuracies of ear-EEG brain–computer interfaces," *IEEE Trans. Cybern.*, vol. 50, no. 8, pp. 3654–3667, Aug. 2020.

[13] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, *arXiv:1704.04861*. [Online]. Available: https://arxiv.org/abs/1704.04861

[14] 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.

[15] H. Cecotti, "A time–frequency convolutional neural network for the offline classification of steady-state visual evoked potential responses," *Pattern Recognit. Lett.*, vol. 32, no. 8, pp. 1145–1153, 2011.

[16] R. T. Schirrmeister, J. T. Springenberg, L. D. J. Fiederer, M. Glasstetter, K. Eggensperger, M. Tangermann, F. Hutter, W. Burgard, and T. Ball, "Deep learning with convolutional neural networks for EEG decoding and visualization," *Hum. Brain Mapping*, vol. 38, no. 11, pp. 5391–5420, Nov. 2017.

[17] T.-H. Nguyen and W.-Y. Chung, "A single-channel SSVEP-based BCI speller using deep learning," *IEEE Access*, vol. 7, pp. 1752–1763, 2019.

[18] N. K. Nik Aznan, S. Bonner, J. Connolly, N. Al Moubayed, and T. Breckon, "On the classification of SSVEP-based dry-EEG signals via convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2018, pp. 3726–3731.

[19] 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, 2017, Art. no. e0172578.

[20] J. J. Podmore, T. P. Breckon, N. K. N. Aznan, and J. D. Connolly, "On the relative contribution of deep convolutional neural networks for SSVEP-based bio-signal decoding in BCI speller applications," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 27, no. 4, pp. 611–618, Apr. 2019.

[21] 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, 2018, Art. no. 056013.

[22] N. R. Waytowich, V. J. Lawhern, J. O. Garcia, and J. Cummings, "Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials," *J. Neural Eng.*, vol. 15, no. 6, Mar. 2018, Art. no. 066031.

[23] N. C. Oza, *Online Ensemble Learning*. Berkeley, CA, USA: Univ. of California, Berkeley, 2001.

[24] 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.

[25] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1800–1807.

[26] O. Faust, Y. Hagiwara, T. J. Hong, O. S. Lih, and U. R. Acharya, "Deep learning for healthcare applications based on physiological signals: A review," *Comput. Methods Programs Biomed.*, vol. 161, pp. 1–13, Jul. 2018.

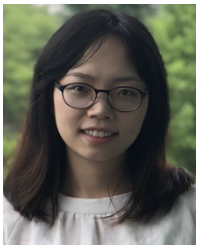
[27] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. 32nd Int. Conf. Mach. Learn. (ICML)*, vol. 1, 2015, pp. 448–456.

[28] F. Chollet. (2015). *Keras*. [Online]. Available: http//keras.io

[29] Y. Roy, H. Banville, I. Albuquerque, A. Gramfort, T. H. Falk, and J. Faubert, "Deep learning-based electroencephalography analysis: A systematic review," *J. Neural Eng.*, vol. 16, no. 5, Aug. 2019, Art. no. 051001.

[30] G. Bin, X. Gao, Z. Yan, B. Hong, and S. Gao, "An online multi-channel SSVEP-based brain–computer interface using a canonical correlation analysis method," *J. Neural Eng.*, vol. 6, no. 4, Aug. 2009, Art. no. 046002.

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



YUANLU ZHU received the B.S. degree in biomedical engineering from South-Central University for Nationalities, Wuhan, China, in 2015. She is currently pursuing the Ph.D. degree in biomedical engineering with the Huazhong University of Science and Technology. Her research interests include applications of hybrid brain-computer interface (BCI) and biomedical signal processing.

YING LI received the B.S. and M.S. degrees in biomedical engineering from the Huazhong University of Science and Technology, Wuhan, China, in 2006 and 2009, respectively. Her research interests include hybrid brain-computer interface (BCI) and biomedical signal processing.

JINLING LU received the Ph.D. degree in biomedical engineering from the Huazhong University of Science and Technology, in 2007. From 2008 to 2010, she worked as a Postdoctoral Researcher with the Institute of Bioimaging, Singapore Science and Technology Bureau. Since 2007, she has been working with the Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology. Her research interests include optical molecular imaging and neural engineering.

PENGCHENG LI received the Ph.D. degree in biomedical engineering from the Huazhong University of Science and Technology, in 2003. Since 2003, he has been working with the School of Life Science and Technology, School of Optoelectronics, and the Wuhan National Laboratory for Optoelectronics, Huazhong University of Science and Technology. He is currently the Secretary General of Biomedical Photonics Branch of the Chinese Society of Biomedical Engineering. His research interests include biomedical photonics and neural engineering.