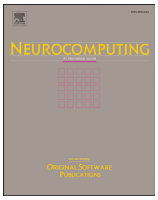
Neurocomputing 656 (2025) 131561

Contents lists available at ScienceDirect

Neurocomputing

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



# LGFCNN: A synergistic framework integrating graph-based spatial filter and lightweight CNN for SSVEP recognition

Rui Ma [a], Yu Cao [a], Sheng Quan Xie [a], Mingming Zhang [b], Jun Li [c,*], Zhi-Qiang Zhang [a,*]

[a] University of Leeds, Woodhouse Lane, Leeds, LS2 9JT, United Kingdom
[b] Southern University of Science and Technology, 1088 Xueyuan Avenue, Shenzhen, 518055, Guangdong, China
[c] College of Intelligent Systems Science and Engineering, Hubei Minzu University, 39 Xueyuan Road, Enshi, 445000, Hubei, China

ARTICLE INFO

Communicated by K. Liu

*Keywords:*
Steady-state visual evoked potentials (SSVEP)
Neural signal decoding
Graph-based spatial filter
Lightweight convolutional neural network (CNN)
Data augmentation

ABSTRACT

Optimizing the feature representation and decoding efficiency of the steady-state visual evoked potentials (SSVEPs) is critical to enhancing the performance of neural signal decoding systems. Current deep learning models often overlook the physical topological information of EEG channels, resulting in suboptimal feature extraction and limited recognition performance. To address these challenges, this study proposes a synergistically designed SSVEP recognition framework to alleviate data insufficiency, improve the feature representation, and enhance decoding efficiency. Specifically, a slicing-and-scaling technique is adopted to improve the model generalization under limited-sample scenarios. A graph-based spatial filter leverages the topological relationships among EEG channels to suppress redundant information and enhance spatial feature quality. A lightweight convolutional neural network (CNN) with fewer parameters is developed to efficiently extract discriminative temporal–spatial features for accurate SSVEP classification. Experimental results on two public benchmark datasets and one self-collected dataset demonstrate that the proposed framework outperforms baseline deep learning models, yielding improvements of at least 6.8 %, 8.5 %, and 0.5 % in peak average classification accuracy, respectively. The maximum average information transfer rates (ITRs) achieved on the three datasets were 221.4 bits/min,106.7 bits/min, and 133.9 bits/min, respectively. By simultaneously reducing model complexity and improving decoding performance, the proposed framework offers an effective and promising approach for efficient neural signal decoding in SSVEP recognition.

## 1. Introduction

The non-invasive brain-computer interface (BCI) system establishes a direct link between the human brain and external electronic devices, enabling individuals to interact with these devices without relying on the peripheral nervous system [1–3]. Given its advantages of higher temporal resolution, lower cost, and greater portability [4,5], electroencephalography (EEG) is widely used in most non-invasive BCI systems to measure human neural activity, providing real-time monitoring of neural responses. In particular, steady-state visual evoked potentials (SSVEP) based EEG-BCIs, which reflect periodic neural responses elicited by visual stimuli, have achieved significant advancements in fields such as medical assistive technologies [6], daily life assistance [7], industrial device control [8,9], and entertainment and virtual reality [10]. Achieving superior recognition accuracy is a crucial requirement for developing robust SSVEP-based EEG-BCI systems.

Traditional prior knowledge-based methods are widely used in SSVEP classification due to their computational efficiency and high classification accuracy [11]. By constructing spatial filters based on common templates or optimized signal components, these methods aim to enhance the signal quality of the measured SSVEPs by improving the signal-to-noise ratio (SNR) [12]. The former methods include canonical correlation analysis (CCA) [13], filter bank CCA (FBCCA) [14], deep multiset CCA (MsetCCA) [15,16], and multivariate variational mode decomposition CCA (MVMD-CCA) [17]. These methods extract time-domain features by computing the correlation coefficients between EEG signals and predefined sinusoidal reference signals. The latter methods, such as task-related component analysis (TRCA) [18] and task-discriminant component analysis (TDCA) [19], aim to maximize the reproducibility of individual SSVEP trials thereby improving the effectiveness of spatial filters. Although these traditional methods

* Corresponding authors.
*Email addresses:* 1995007@hbmzu.edu.cn (J. Li), z.zhang3@leeds.ac.uk (Z.-Q. Zhang).

https://doi.org/10.1016/j.neucom.2025.131561
Received 23 May 2025; Received in revised form 9 September 2025; Accepted 12 September 2025
Available online 15 September 2025

have demonstrated notable performance in SSVEP classification, their reliance on static templates and handcrafted features may limit their ability to capture the non-stationary and complex neural dynamics, thereby reducing robustness in scenarios with varying signal conditions [20].

Benefiting from the automatic feature extraction and end-to-end optimization capabilities of deep learning (DL) algorithms, fully data-driven DL models have been increasingly developed to process SSVEP signals. By learning effective representations directly from raw EEG data, these models contribute to enhancing both the quality of the extracted features and the reliability of subsequent recognition. Several DL models have demonstrated the state-of-the-art (SOTA) performance in improving the classification accuracy of SSVEP-BCIs. For instance, Waytowich et al. [21] proposed EEGNet, one of the most widely used convolutional neural network (CNN) models, which effectively exploits temporal-spatial information across all channels by a depth-wise separable convolution block and achieves higher classification accuracy than the standard CCA method. Ding et al. [22] introduced the Filter Bank-based time-domain CNN (FB-tCNN) to fully utilize both fundamental and harmonic information. Zhang et al. [23] adopted complex spectrum features as the input to a DL network, referred to as FB-CCNN. Pan et al. [24] developed a model integrating the CNN and long-short term memory (LSTM) network to capture temporal dependencies within temporal-spatial features of SSVEP signals. Additionally, Wang et al. [25] and Chen et al. [26] introduced attention modules in CNN-based architectures to enhance classification performance, as features from different domains or channels may contribute unequally. Despite the significant progress achieved by these DL algorithms in SSVEP-BCIs, the majority of them capture spatial dependencies primarily based on channel index order, rather than explicitly considering the physical topological characteristics of the electrode layout.

The topological characteristics of EEG signals play a crucial role in shaping the spatial dependence between electrodes [27]. In fact, brain activity generates similar waveforms across different electrode locations, particularly among closely spaced electrodes. This spatial proximity often causes adjacent electrodes to record highly similar signals, leading to information redundancy [28]. In addition, the spatial relationships among EEG channels are represented by their indices in traditional CNN models [29]. As a result, the local spatial filtering considers only the left and right neighboring channels of a given channel while neglecting anterior and posterior neighboring channels. This limitation is particularly pronounced for edge channels, where the next indexed channel is not necessarily physically adjacent. Although several studies have employed graph neural networks (GNNs) to model topological relationships among EEG channels for SSVEP classification [30–33], these approaches typically construct trainable, multi-layer graph structures to model inter-channel dependencies in EEG data, aiming to enhance spatial feature representation for SSVEP classification. However, iterative neighborhood aggregation across multiple layers might result in feature over-smoothing and substantially increase the number of trainable parameters, leading to higher model complexity and computational demands. Notably, GNN-based models focus on spatial feature modeling rather than explicitly addressing information redundancy.

To leverage the topological characteristics of EEG while mitigating the large number of trainable parameters in the DL model, this study proposes an SSVEP processing framework composed of three sequential yet complementary modules, enhancing both neural signal representation and classification performance. Specifically, a random walk Laplacian matrix based on graph theory is employed as a topology-aware spatial filter, reducing inter-channel redundancy in SSVEP signals. A lightweight CNN with fewer training parameters is then applied to extract discriminative temporal–spatial features for multi-class classification. To address the overfitting issue caused by limited training data in DL models, a slicing-and-scaling method is employed as a data augmentation strategy to increase sample diversity. Although these modules are implemented independently and sequentially, their integration creates a synergistic effect, as each module enhances the effectiveness of the next, collectively resulting in improved neural signal representation, classification accuracy, and low computational complexity.

The key contributions of this work can be summarized as follows:

1. A new SSVEP decoding framework is proposed, synergistically integrating slicing-and-scaling data augmentation, topology-aware graph spatial filtering, and lightweight CNN architecture. This integrated design allows each module to complement the others in enhancing overall model performance while maintaining low computational complexity.
2. The slicing-and-scaling augmentation improves model robustness under limited data conditions; the graph-based spatial filter reduces inter-channel information redundancy by leveraging EEG topology; and the compact CNN efficiently extracts discriminative temporal–spatial features with minimal parameters. The integrated approach supports enhanced SSVEP feature decoding and stable recognition performance within individual subjects.
3. Experimental results on two public and one self-collected dataset demonstrated that the proposed framework substantially improves recognition accuracy and information transfer rate (ITR) compared to representative baseline methods, verifying the effectiveness and practical utility of the design.

## 2. Methods

This study proposes a synergistic framework for SSVEP recognition. The entire framework of the proposed method is illustrated in Fig. 1. In the following subsections, the details of the method will be introduced.

### 2.1. SSVEP datasets and preprocessing

In this study, two public datasets and one self-collected dataset were used to evaluate the effectiveness of the proposed method.

(1) *Dataset I: 40-Target Benchmark Dataset* [34]: The first dataset contains data from 35 healthy subjects. EEG data were recorded from 64 electrodes at a downsampling rate of 250 Hz. Subjects were presented with a screen including 5 ×8 stimulus matrix that represented 40 characters. The frequencies of the stimuli ranged from 8 Hz to 15.8 Hz with 0.2 Hz interval, and the phases ranged from 0 to $1.5\pi$ with $0.5\pi$ interval. The whole experiment was composed of 6 blocks, with each block including 40 trials corresponding to the 40 stimuli. Each trial started with a 0.5-s target cue, followed by 5 s of flickering at the target frequency. Subsequently, there was a 0.5-s blank on the screen before the next trial.

(2) *Dataset II: 40-Target Beta Dataset* [35]: The second dataset was collected from 70 healthy subjects. The 40 SSVEP stimulus frequencies were same as those in Dataset I; however, the experiment was conducted without laboratory electromagnetic shielding and was presented using QWERTY virtual keyboard. Unlike Dataset I, the entire experiment consisted of four blocks, and the stimulation duration for each trial varied among subjects. Specifically, for subjects 1–15, the stimulation duration was 2 s, while for subjects 16–70, it was 3 s.

(3) *Dataset III: 12-Target self-collected Dataset*: The third dataset was a self-collected dataset, which was recorded using 16 electrodes from 11 healthy subjects. The sampling frequency was 256 Hz without downsampling. There were 12 SSVEP stimuli flickering with a frequency range of 9.25–14.75 Hz (interval was 0.5 Hz), with the phase range of $0\pi$ to $1.5\pi$ (interval was $0.5\pi$). The experiment consisted of ten blocks for each subject. Each block included 12 trials, each corresponding to a different stimulus pattern. Each trial began with a 0.5-s target cue, followed by 5 s of stimulus flickering, and concluded with a 0.5-s rest period.

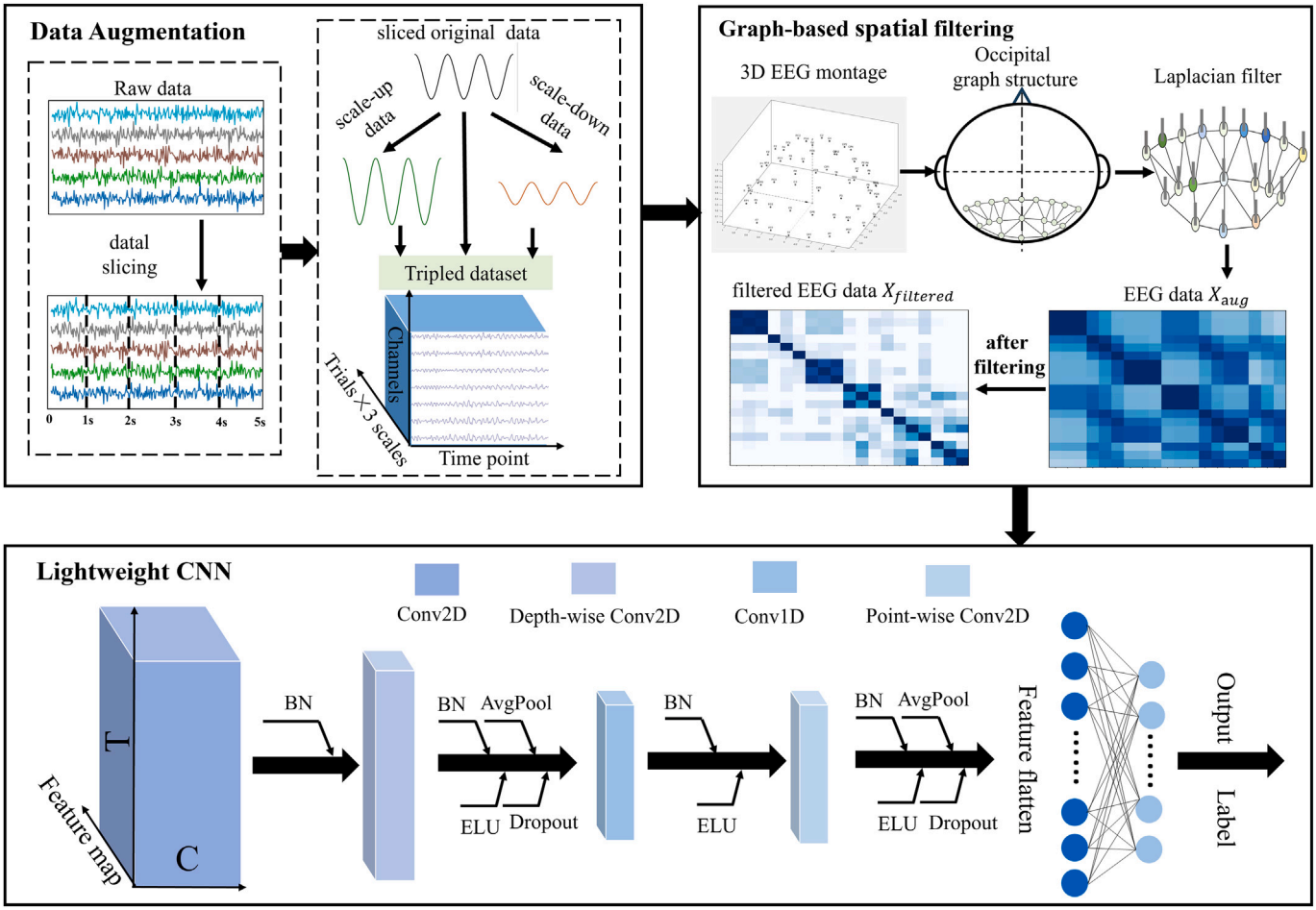



**Fig. 1.** The whole framework of the proposed lightweight convolutional neural network with graph-based filter for SSVEP Recognition.

To ensure the quality of EEG data, several preprocessing steps were applied to both datasets, which were conducted using the EEGLAB [36] toolbox in MATLAB 2024 a (The MathWorks, Natick MA). Baseline correction and 6–33 Hz band-pass filtering were performed on both datasets. Nineteen EEG electrodes were selected for the two public datasets (Fig. 2(a)), and nine for the self-collected dataset (Fig. 2(b).

### 2.2. Data augmentation

Given the limited number of trials available in both datasets to train the DL model, this study employed a slicing-and-scaling techniques to augment the dataset. Specifically, after preprocessing, EEG data from each trial was divided into five/three/two non-overlapping one-second time segments according to different datasets, and amplitude scaling technique was applied to generate additional samples. Although amplitude scaling has been explored in previous EEG research, its integration and systematic evaluation within the SSVEP decoding framework—especially in combination with topology-aware spatial filtering and DL models—remain limited. This data augmentation approach not only increases data diversity but also helps mitigate overfitting and improve classification performance.

The sliced one-second segment was denoted as $X \in \mathbb{R}^{C\times T}$. C represented the number of EEG channels and T denoted the time series length. For the subsequent model performance analyses requiring shorter data length, each 1-s segment was further truncated from the beginning to generate segments of 0.2, 0.4, 0.6, and 0.8 s. For each shorter duration, only the initial portion (e.g., 0–0.2 s) of each 1-s segment was used, thereby ensuring the number of samples remained consistent across all segment lengths. Next, data augmentation was performed using a scaling method, following the rules below:

$$X_{scaleup} = (1+\alpha)\times X \tag{1}$$

$$X_{scaledown} = (1-\alpha)\times X \tag{2}$$

where $X_{scaleup}$ and $X_{scaledown}$ were the data obtained by amplifying and reducing the amplitude of the original signal, respectively; $\alpha$ was the scaling control factor that was set to 0.1 in this study. $X$, $X_{scaleup}$ and $X_{scaledown}$ together formed the final training dataset, where each sample

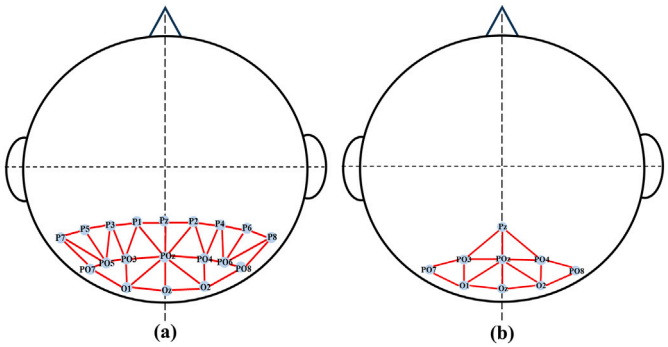



**Fig. 2.** The electrode configuration and predefined connectivity for Dataset I (a), Dataset II (a) and Dataset III (b).

was denoted as $X_{aug}$. Through this data augmentation method, the size of SSVEP data for each subject was increased to 1800 samples (5 trials × 3 scales × 6 blocks × 40 classes) for Dataset I, 960 or 1440 samples (2 or 3 trials × 3 scales × 4 blocks × 40 classes) for Dataset II, and 1200 samples (5 trials × 3 scales × 10 blocks × 12 classes) for Dataset III.

### 2.3. Graph-based spatial filter

It is widely accepted that EEG signals represent a mixture of multiple brain sources [37,38]. A single EEG channel can simultaneously receive signals from multiple brain regions, with activity from each source contributing to responses across multiple channels. Consequently, different channels may contain overlapping information leading to redundancy.

Let $r \in \mathbb{R}^3$ denote a spatial location in three-dimensional(3D) Euclidean space. The scalp potential $V(r)$ is a scalar field that reflects the cumulative effect of multiple cortical current sources at position $r$ and the electric field $E(r)$ is defined as the negative spatial gradient of the potential [39,40]:

$$E(r) = -\nabla V(r) \tag{3}$$

where $\nabla$ is the vector differential operator, $E$ represents the direction and magnitude of electric field propagation at each point in space. According to Ohm's law in conductive media, the current density $J(r)$ is proportional to the electric field at that point:

$$J(r) = \sigma(r) \cdot E(r) = -\sigma(r) \cdot \nabla V(r) \tag{4}$$

where $\sigma(r)$ is the electrical conductivity tensor at location $r$. This equation forms the foundation for relating measured scalp potentials to underlying cortical current sources [39]. EEG signals are typically analyzed under the quasi-static approximation, which neglects electromagnetic wave propagation and assumes that the current density field $J(r)$ obeys the law of current conservation:

$$\nabla \cdot \boldsymbol{J}(r) = 0 \tag{5}$$

Therefore, let substitute Eq. (4) it into the Eq. (5) yields the Poisson equation [40]:

$$\nabla \cdot (\sigma(\mathbf{r}) \nabla V(\mathbf{r})) = -\rho_s(\mathbf{r}) \tag{6}$$

If there are sources (such as neural discharges), the right-hand side is no longer zero and the Eq. (6) is modified as follows:

$$\nabla \cdot (\sigma(\mathbf{r}) \nabla V(\mathbf{r})) = -\rho_s(\mathbf{r}) \tag{7}$$

where $\rho_s$ denotes the source current density, indicating the presence and intensity of local neural sources. Assuming the conductivity is homogeneous and isotropic [39], i.e., $\sigma(r) = \sigma$, a constant scalar, the Eq. (7) simplifies to

$$\nabla^2 V(\mathbf{r}) = -\frac{\rho_s(\mathbf{r})}{\sigma} \tag{8}$$

Here, $\nabla^2 V$ is denoted as the Laplacian operator of the potential field [40]. Mathematically, the Laplacian at a given location quantifies how much the electric potential at that point differs from the average potential of its neighboring points. When the potential field is spatially smooth, indicating minimal differences between adjacent points, the resulting Laplacian values approach zero. Conversely, sharp local changes in potential, such as those produced by focal cortical sources, yield large Laplacian values. This inherent property allows the Laplacian operator to function as a spatial high-pass filter, attenuating spatially widespread, low-frequency background activity while amplifying localized, high-frequency spatial components. Thus, the Laplacian operator effectively emphasizes source-related neural activity and suppresses broadly distributed or volume-conducted signals.

To implement this principle in a discrete electrode space, the spatial configuration of EEG channels can be modeled as a graph and a graph Laplacian operator [41,42], which serves as a discrete analogue of the continuous Laplacian operator $\nabla^2 V$. A weighted graph is defined as $G = \{N, L\}$, where $N$ denotes the set of EEG channels (nodes) and $L$ represents the links between nodes. The connectivity is encoded in a weighted adjacency matrix $A = \in \mathbb{R}^{C \times C}$ where $C$ is the number of channels. The weight $A_{ij}$ between channels $i$ and $j$ is defined as:

$$A_{ij} = \begin{cases} \frac{1}{d_{ij}}, & i \text{ is adjacent to } j \\ 1, & i = j \\ 0, & \text{otherwise} \end{cases} \tag{9}$$

where $d_{ij}$ is the Euclidean distance between the 3D coordinates of channels $i$ and $j$. In this study, the graph structure is constructed based on the physical distance between the EEG channels. The standard physical 3D coordinates for each channel are provided by EEGLAB [36]. As shown in Fig. 2, usually the channels located to the left, right, front, and back of a given channel in the visual cortex were selected as its neighbor nodes in this work.

Following the graph construction process described above, a graph-based random walk Laplacian matrix is introduced to enhance the feature differentiation among EEG channels. Specifically, let $\mathrm{D} \in \mathbb{R}^{C \times C}$ be the diagonal matrix whose elements are $D_{ii} = \sum_{j=1}^{C} A_{ij}$. A random walk transition matrix $P$ is defined as $P = D^{-1}A$, where each element $p_{ij}$ indicates the probability of transitioning from channel $c_i$ to channel $c_j$ in a random walk. Based on $P$, the random walk Laplacian matrix $L^{rw} \in \mathbb{R}^{C \times C}$ is formulated as:

$$L^{\mathrm{rw}} = I - P \tag{10}$$

where I is the identity matrix. $L^{rw}$ quantifies each channel's deviation from the weighted average of its neighbors on the graph, mimicking the divergence of a gradient field. It can be seen as a spatial filter to enhance the information difference of the EEG data $X_{aug} \in \mathbb{R}^{C \times T}$ as follows:

$$X_{filtered} = L^{\mathrm{rw}} X_{aug} \tag{11}$$

At each time point $t$, the filtered signal of channel $c_i$ is defined as:

$$X_{filered}[c_i, t] = X[c_i, t] - \sum_{c_j \in N(c_i)} P_{c_i c_j} X[c_j, t] \tag{12}$$

where $N(x_i)$ denotes the set of spatial neighbors of the channel $c_i$. This operation effectively subtracts the weighted average of the spatial neighbors from the current channel, retaining only the component that differs from its local surroundings. If the signals of neighboring channels $X[c_j, t]$ are highly similar due to volume conduction or spatially widespread activity, then $X[c_i, t] \approx X[c_j, t]$, and the weighted average of the neighbors approaches $X[c_i, t]$. Thus, the subtraction yields a value close to zero, indicating that the local spatial information is redundant. In contrast, when $X[c_i, t] \gg X[c_j, t]$ or $X[c_i, t] \ll X[c_j, t]$, $X_{filered}[c_i, t]$ becomes large in magnitude, highlighting local spatial discrepancies. Therefore, $L^{rw}$ approximates the Laplacian operator $\nabla^2 V$ of the scalp potential in the discrete electrode space, enabling spatial filtering that preserves source-related, high-frequency spatial patterns while attenuating low-frequency, volume-conducted background activity. Fig. 3 illustrates a simplified example demonstrating the graph-based spatial filtering procedure, where inter-node signal differences are enhanced, as reflected by the color changes.

### 2.4. The lightweight CNN architecture and the design of loss function

Given the significantly smaller number of EEG samples compared to the large number of parameters in deep neural networks, this study proposes a lightweight convolutional neural network with a graph-based filter, termed LGFCNN. The network consists of four convolutional

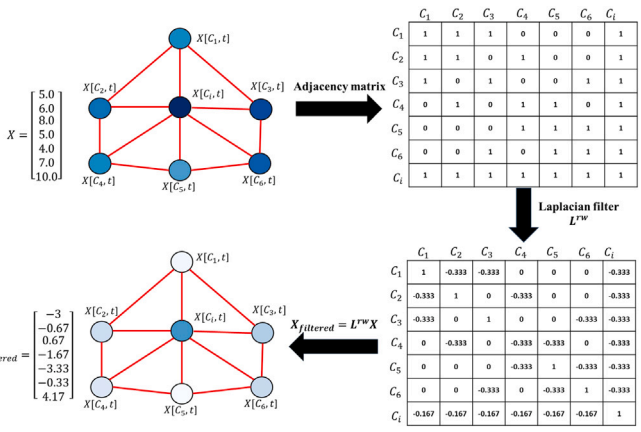



**Fig. 3.** A simplified example demonstrating the graph-based spatial filtering procedure, with node colors representing signal activations pre- and post-filtering.

**Table 1**
The architecture of LGFCNN.

| Layer type | Output dimension | Kernel size | Options |
|---|---|---|---|
| Input | (1,C,T) | | |
| Conv2D | (48,C,T) | $1 \times T/2$ | stride = (1,1)<br>padding = 'same' |
| BN | | | |
| Depth-wise Conv2D | (48,1,T) | $C \times 1$ | stride = (1,1)<br>depth = 1 |
| BN | | | |
| Activation | | | ELU |
| Average Pooling | (48,1,T/4) | $1 \times 4$ | stride = (1,4) |
| Dropout | | | ratio = 0.5 |
| Conv1D | (48,1,T/4) | 3 | stride = 1<br>padding = 1 |
| Point-wise Conv2D | (48,1,T/4) | $1 \times 1$ | stride = 1<br>padding = 'same' |
| BN | | | |
| Activation | | | ELU |
| Average Pooling | 48,1,T/32) | $1 \times 8$ | stride = (1,8) |
| Dropout | | | ratio = 0.5 |
| Flatten | $(48 \times T/32)$ | | |
| Dense | $K$ | | |
| Activation | | | softmax |

blocks and a fully connected classifier. The Conv2D slides along the timeline to extract local temporal features; the Depthwise Conv2D serves as a spatial filter; the Conv1D further extracts the contextual features in the time dimension to enhance the ability for time modeling; the pointwise Conv2D performs feature fusion and mapping. After applying the graph-based filter, the data $X_{filtered}$ is reshaped as $(1, C, T)$. The detailed network parameters are presented in Table 1, where $K$ denotes the number of SSVEP classes.

To avoid the overconfidence and improve the generalization capability of the model, this study introduces a joint loss function, $Loss_{cl}$. The standard cross-entropy loss function is depicted as follows:

$$Loss_{ce} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{i,k}\log(p_{i,k}) \tag{13}$$

$$y_{i,k} = \begin{cases} 1, & k = \text{true label} \\ 0, & k \neq \text{true label} \end{cases} \tag{14}$$

where $N$ denotes the batch size, $y_{i,k}$ denotes the label of the $i^{th}$ sample to be of the class $k$, and $p_{i,k}$ is the probability that the model predicts the $i^{th}$ sample to be of the class k. While this approach ensures that the model primarily focuses on actual categories, it may also lead to overconfidence in predictions and insufficient generalization. The label smoothing method converts true labels from "hard labels" to "soft labels", making a light reduction in the probability of the correct category, and evenly distributing the reduction to the other categories. The converted soft target label $y'_{i,k}$ of each sample i and class k is defined as:

$$y'_{i,k} = \begin{cases} (1-\epsilon), & k = \text{true label} \\ \frac{\epsilon}{K-1}, & k \neq \text{true label} \end{cases} \tag{15}$$

where $\epsilon$ is the smoothing parameter, set to 0.1 in practice. Therefore, the cross entropy after label smoothing is:

$$Loss_{lsr} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y'_{i,k}\log(p_{i,k}) \tag{16}$$

The joint loss function is defined as:

$$Loss_{cl} = Loss_{ce} + \alpha_{lsr} \times Loss_{lsr} \tag{17}$$

where, $\alpha_{lsr}$ is a weighting parameter that controls the proportion of the regularization loss $Loss_{lsr}$ in the total loss. In this study, $\alpha_{lsr}$ was set to 0.1.

### 2.5. Training strategy

A subject-dependent experiment was conducted to evaluate the performance of the proposed model. All original SSVEP trials were first partitioned into training and testing sets with a ratio of 85 % to 15 %. In the testing set, only the slicing method was applied to the original trials. The training set was further divided into training and validation subsets using five-fold cross-validation, with each fold allocating 80 % of the data for training and 20 % for validation. During this process, the validation subset included only sliced segments of the original trials, while the training subset employed both slicing and scaling for data augmentation. The same data partitioning and augmentation procedures were applied consistently across all deep-learning models. All models were trained using the early stopping strategy, in which training was stopped if the validation accuracy did not improve by at least 0.01 % within 50 epochs. The model with the best performance on the validation set was selected and subsequently evaluated on the testing set. The proposed model was implemented using the PyTorch framework and executed on an NVIDIA Quadro GV100 with a memory size of 32GB. The Adam optimizer was employed to update model parameters and minimize the loss function. The learning rate and weight decay were both set to 0.001. For each fold, the batch size and total iteration epochs were set to 128 and 500 for Dataset I, and 64 and 500 for Dataset II and Dataset III, respectively.

### 2.6. Performance evaluation

#### 2.6.1. Evaluation metrics

The classification accuracy and information transfer rate (ITR) were applied as performance evaluation metrics. The accuracy is computed as:

$$Accuracy = \frac{TP}{AP} \tag{18}$$

where TP denotes the number of correctly predicted samples and AP is the number of all samples predicted. The ITR is defined as:

$$ITR = \frac{60}{T}\left[\log_2 K + P\log_2 P + (1-P)\log_2\left(\frac{1-P}{K-1}\right)\right] \tag{19}$$

where T, K, and P represent the time length, the number of targets, and the classification accuracy, respectively. Specifically, an additional 0.5 s was added to the parameter T to account for the gaze movement time, as suggested in [43]

#### 2.6.2. Statistical analysis

All experimental results were reported as the average classification accuracies and ITRs across all subjects unless otherwise specified. Error bars in the figures represent the standard error of the mean (SEM). Furthermore, the Wilcoxon Signed-Rank test with Bonferroni correction was conducted to compare the baseline models (EEGNet [21], CCNN [44], AttentCNN [45], SSVEPNet [24], and SSVEPformer [20]) with the proposed LGFCNN model to assess statistical significance. The asterisks shown in figures indicate significant differences between the proposed model and the other baseline models (*p<0.05, **p<0.01, ***p<0.001).

## 3. Results

### 3.1. The overall SSVEP classification performance

#### 3.1.1. Comparison with DL model

The distribution of classification accuracies across all subjects for the proposed models and other baseline models on data segments longer than 0.6 s is illustrated in Fig. 4. The width of the violin plots represents the density of data points at different accuracy levels. Among all data lengths and datasets, the median accuracy of LGFCNN is generally above that of all models. Furthermore, the distribution of accuracies achieved by LGFCNN was most concentrated, indicating higher consistency in performance. Specifically, for Dataset I, the bottom of the violin in LGFCNN has almost no extremely low-score samples below 60 %, while other models have a distinct long tail in the low-accuracy region.

To further compare the overall performance across all subjects, the average classification accuracy and ITR of all models on both datasets are presented in Fig. 5. For Dataset I, as shown in Fig. 5(a), LGFCNN achieved the highest accuracy among all baseline models when the data length was greater than or equal to 0.4 s. Specifically, the accuracies for data lengths of 0.4 s, 0.6 s, 0.8 s and 1 s were $62.4 \pm 9.5\%$, $86.3 \pm 8.0\%$, $88.7 \pm 6.3\%$ and $93.1 \pm 4.3\%$, respectively. The corresponding ITRs (Fig. 5(d)) of LGFCNN also outperformed those of the baseline models across all data lengths that were greater than or equal to 0.4 s. The highest average ITR achieved by LGFCNN was $221.4 \pm 34.4$bits/min at 0.6 s. For Dataset II, which also comprises 40 SSVEP targets, LGFCNN consistently outperformed the other models across data lengths ranging from 0.6 to 1 s. The highest ITR of LGFCNN was achieved at the 1-s data length. However, its overall classification accuracy was lower than that achieved on Dataset I. In terms of the Dataset III, the accuracies and ITRs of LGFCNN were generally higher than those of all baseline models across all data lengths. However, at 0.8 s, LGFCNN's accuracy was 0.8 % and 0.7 % lower than that of CCNN and SSVEPformer, respectively. Furthermore, the significance level between the LGFCNN and the baseline models was also shown in Fig. 5. The results indicated that LGFCNN generally achieved a significant performance improvement across most data lengths on all datasets, with p-values consistently below 0.05.

The t-distributed stochastic neighbor embedding (t-SNE) technique was applied to visualize the output features of different models, and the result was shown in Fig. 6. To quantitatively evaluate the quality of the feature clusters, three standard metrics were employed: the Calinski-Harabasz (CH) Index, the Silhouette Coefficient, and the Davies-Bouldin (DB) Index. Higher CH Index and Silhouette Coefficient values, alongside a lower DB Index, indicate better cluster quality, reflecting greater between-cluster separation, improved sample alignment within clusters, and enhanced intra-cluster compactness, respectively. Fig. 6(a) indicates that features of 40 classes were well differentiated by the proposed LGFCNN model, as features belonging to the same class clustered closely together. Although the output features of EEGNet exhibited compact clusters, several classes of EEGNet's features overlapped, indicating a lower discriminative ability. Besides, for the SSVEPformer, it presented the highest Silhouette Coefficient and the lowest DB Index, indicating tighter within-class compactness and reduced inter-class similarity at the sample level. However, its CH Index was lower than that of LGFCNN, reflecting weaker dispersion between class centroids. In terms of other models, the output features of all classes were largely intermixed, showing no clear boundaries among these features.

#### 3.1.2. Comparison with conventional methods

To more comprehensively evaluate the performance of the proposed LGFCNN, additional experiments were conducted on 1-s data segments, comparing LGFCNN with the classical template-based methods eTRCA and TDCA under two training protocols. All methods were evaluated using the leave-one-out block strategy on the same dataset, which comprised both original and augmented data. The protocols were defined as follows: (1) CBFTrain – only the initial 1-s segment from each block was used for training and testing; and (2) CBATrain – all 1-s segments from each block were used for training and testing. As shown in Fig. 7, LGFCNN achieved the highest accuracies under the CBATrain protocol, but its performance dropped sharply under the CBFTrain protocol. In contrast, both eTRCA and TDCA performed better under the CBFTrain protocol, yet showed substantial declines in accuracy when evaluated with CBATrain. Overall, without distinguishing between protocols, LGFCNN achieved the best accuracies with the lowest standard deviations on Dataset I and Dataset III, whereas TDCA yielded the highest accuracy on Dataset II.

### 3.2. Effectiveness analysis of data augmentation and graph-based spatial filter

Fig. 8 presents the t-SNE visualization of learned feature distributions before and after data augmentation. It is evident that after data augmentation, the features of different classes become more compact within clusters and are better separated from each other. Feature clustering metrics for augmented data showed significant improvement compared to those for data before augmentation. Besides, to examine the sensitivity of data augmentation scale factor $\alpha$, the average accuracy across all subjects under different data lengths and scale factors is shown in

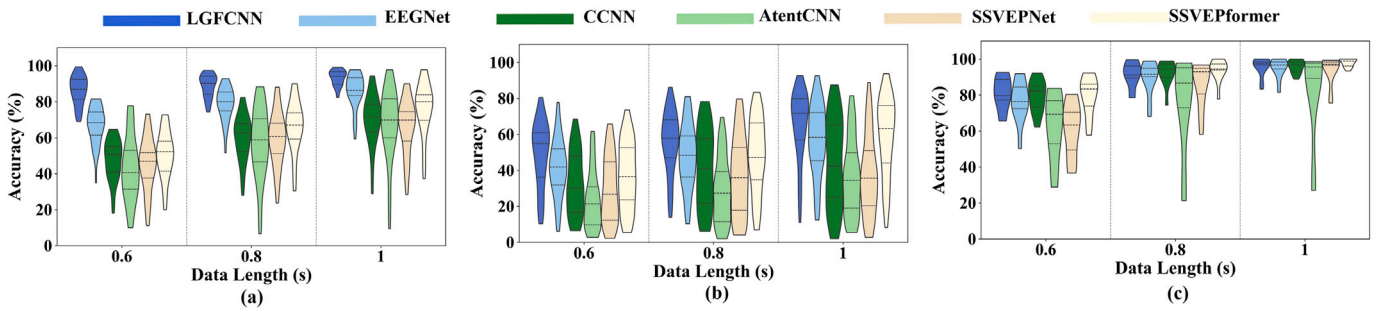



**Fig. 4.** Distributions of classification accuracy across all subjects using five DL models at different data lengths for Dataset I (a), Dataset II (b), Dataset III (c). The proposed LGFCNN genearally showed the most concentrated accuracy distribution.

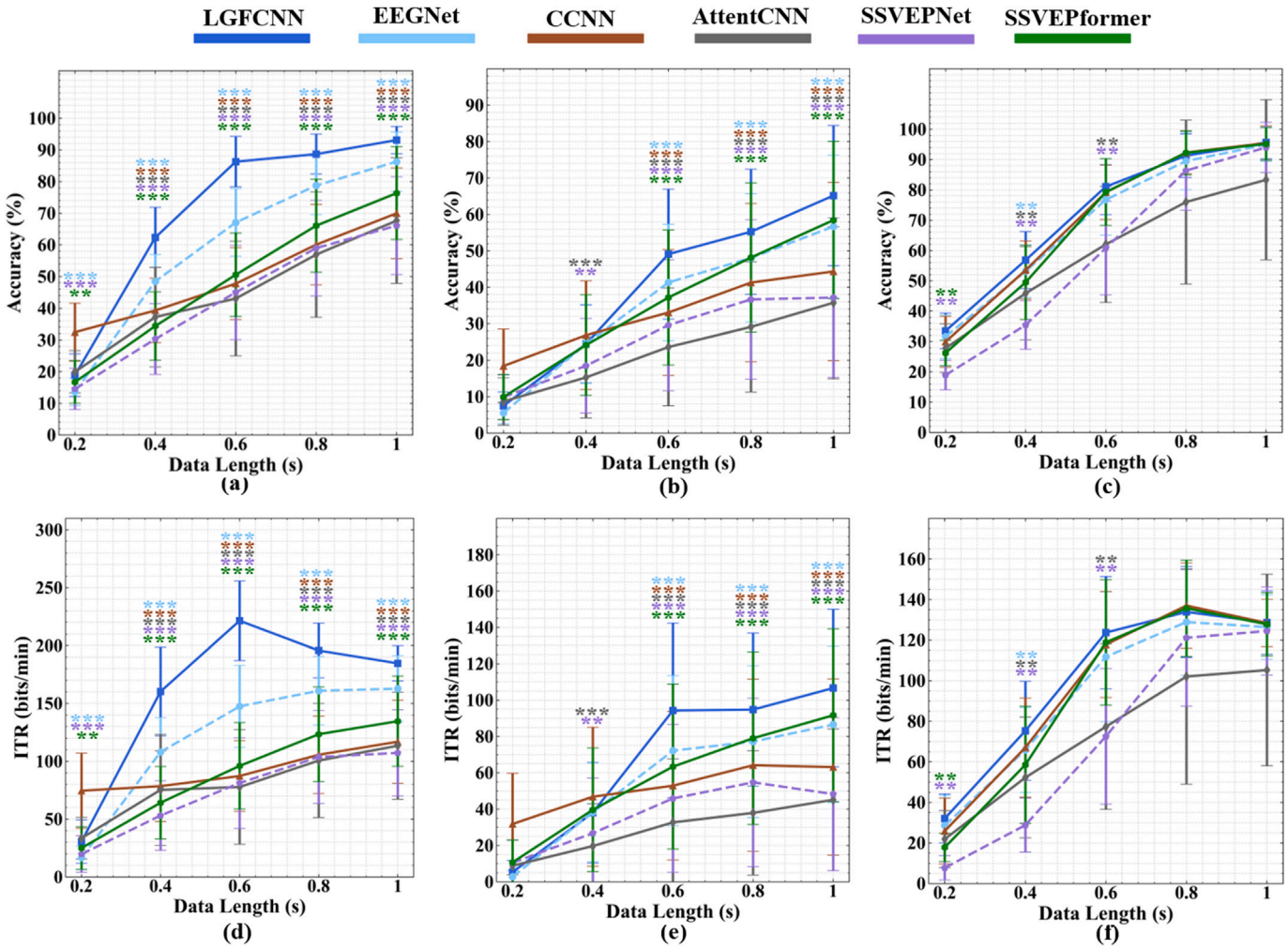



**Fig. 5.** Average classification accuracy and ITR of five models for Dataset I (a,d), Dataset II (b,e) and Dataset III (c,f) at different data lengths. Blue, brown, gray, purple and green asterisks indicate significant differences between LGFCNN and EEGNet, CCNN, AttentCNN, SSVEPNeta and SSVEPformer respectively (*p<0.05, **p<0.01, ***p<0.001).

Fig. 9. Across all values of $\alpha$, the accuracy of a given data length exhibited minimal fluctuation. Although the optimal $\alpha$ may depend on the characteristics of the dataset, $\alpha = 0.1$ appears to be the optimal choice, as it maintains stable performance across different time lengths and offers additional benefits for shorter data lengths, such as 0.6 s.

To analyze the effect of the graph-based spatial filter, the correlation coefficients among channels were calculated and visualized in the form of heat maps, as shown in Fig. 10. It could be observed that the correlations between a given channel and the other channels were generally high, indicating minimal information differences among them. Furthermore, combined with the topological information shown in Fig. 2, the correlation coefficient increased as the distance between channels decreased. However, after applying the random walk Laplacian filter, the information differences between channels were greatly enhanced, emphasizing the importance of neighboring channels for a given channel.

To further evaluate the effectiveness of the data augmentation method, the random walk Laplacian filter, and their combined usage with the proposed LGFCNN, four ablation experiments were conducted on Dataset I. The data with or without augmentation was named as "W/ Augmentation" and "W/O Augmentation", respectively; the model with and without graph-based filter was named as "W/ GSF" and "W/O GSF", respectively. Therefore, the four experiments were: (1)W/ Augmentation & W/ GSF; (2)W/ Augmentation & W/O GSF; (3)W/O Augmentation & W/ GSF; (4)W/O Augmentation & W/O GSF. Fig. 10(a) shows average accuracies of each experiment at different data lengths and the significance level comparison. In all cases of data length, except for 0.2 s, the "W/ Augmentation & W/ GSF" condition demonstrated the best performance, showing a significant improvement compared to the conditions "with/without GSF or augmentation" methods. Specifically, under the same condition of W/ GSF or W/O GSF ((1) vs. (3) or (2) vs. (4)), the use of augmentation improved accuracy ranging from 6.2 % to 12.5 %. Regarding the effect of GSF, it contributed to an accuracy improvement ranging from 2.8 % to 8.2 %. Fig. 10(b) presents the ITR performance comparison of all ablation experiments. It is noted that the trend change of all conditions was consistent with that of accuracy. However, the highest ITR of all models was achieved at a data length of 0.6 s. Furthermore, as the data length decreased to 0.2 s, both accuracy and ITR dropped significantly, and the use of GSF did not provide additional benefits when data augmentation was applied.

### 3.3. Analysis of the number of EEG channels

To evaluate the performance of the proposed model with different numbers of EEG channels, the average accuracies for two channel configurations (9 channels vs. 19 channels) across all data lengths in Dataset I are presented in Fig. 11. The overall results indicated that larger number of channels generally improves classification accuracy,

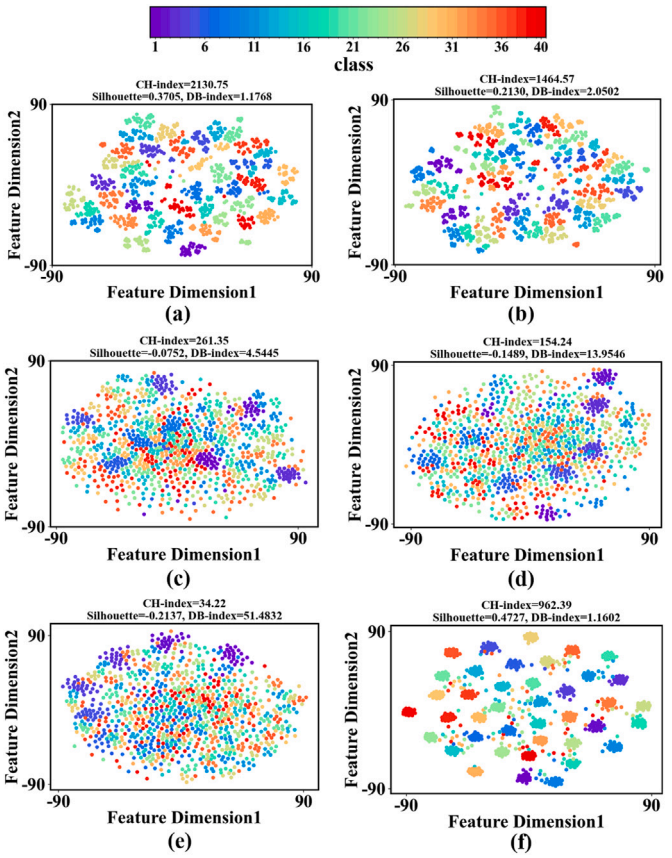



**Fig. 6.** t-SNE visualization of output features for a representative subject using six models on Dataset I: (a) LGFCNN, (b) EEGNet, (c) CCNN, (d) AttentCNN, (e) SSVEPNet and (f) SSVEPformer. LGFCNN demonstrates clearer intra-class clustering and greater inter-class separability.

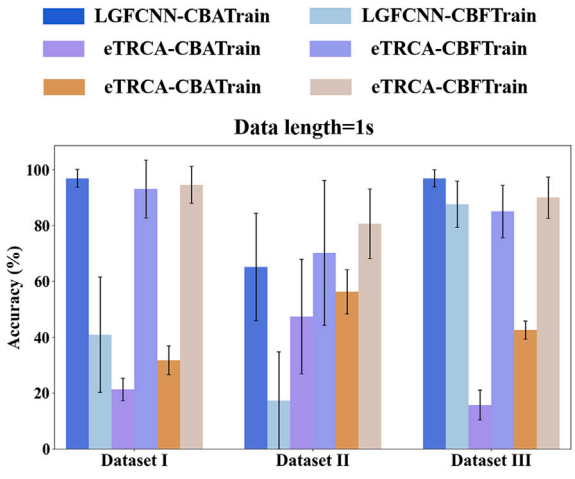



**Fig. 7.** Accuracy comparison of LGFCNN, eTRCA, and TDCA under CBATrain and CBFTrain protocols (data length = 1 s). LGFCNN performed best with CBATrain, while eTRCA and TDCA were more competitive with CBFTrain.

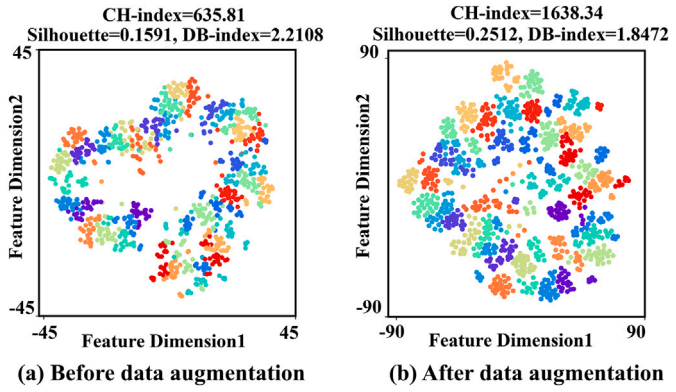



**Fig. 8.** t-SNE visualization of output features for data before and after augmentation in Dataset I.

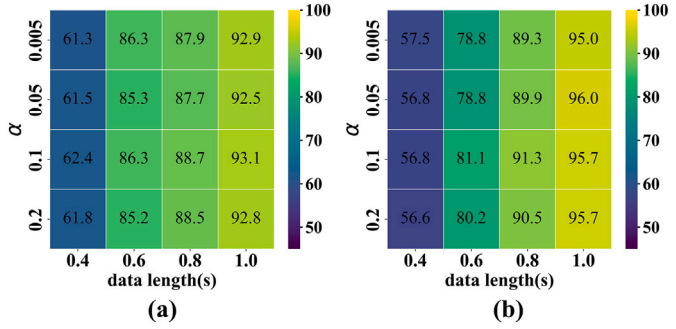



**Fig. 9.** Parameter sensitivity analysis for the augmentation scale factor $\alpha$ on Dataset I (a) and Dataset III (b). The model exhibited the most stable and robust classification performance when $\alpha$ was set to 0.1.

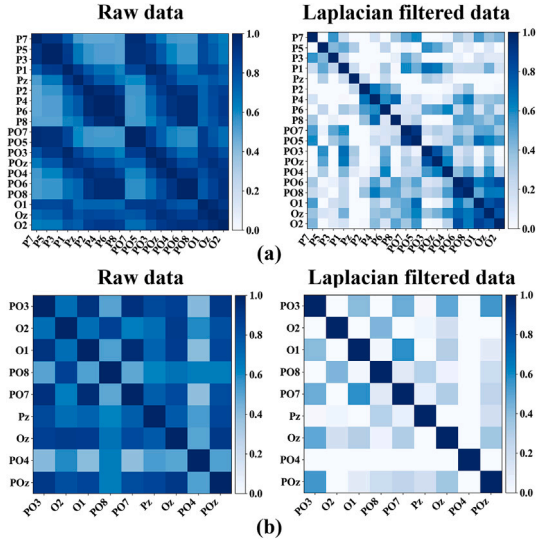



**Fig. 10.** Heat maps of the correlation coefficients matrix of EEG channels for (a) Dataset I and (b) Dataset II. Information difference was enhanced after applying graph-based spatial filter.

especially for longer data lengths. Specifically, for 9-channel configurations, accuracy was relatively low, primarily concentrated around 20 % for 0.2 s, increasing to around 40 % at 0.4 s. For longer data lengths, accuracy improved in both configurations; however, the 19-channel setup exhibited a more pronounced enhancement, with a significant portion of results exceeding 80 %. The 9-channel configuration exhibited a wider spread in accuracy, suggesting greater variability in performance. While both configurations benefited from longer data segments, the 19-channel setup achieved superior accuracy with reduced variance, suggesting that an electrode arrangement incorporating more

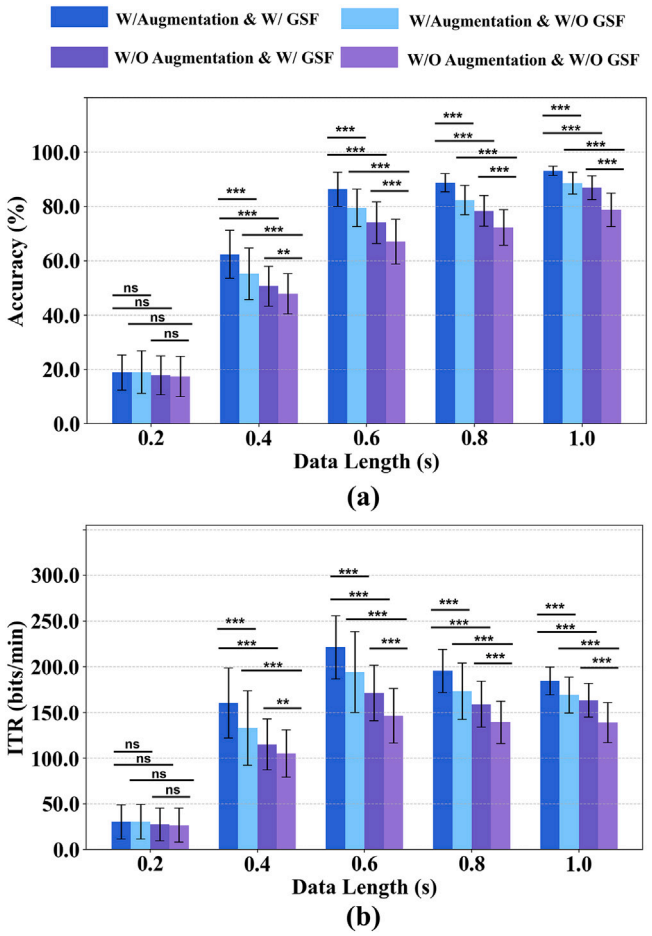



**Fig. 11.** Ablation experiment results for Dataset I in terms of accuracy (a) and ITR (b). The full framework incorporating both the graph-based spatial filter and data augmentation achieves the best overall performance.

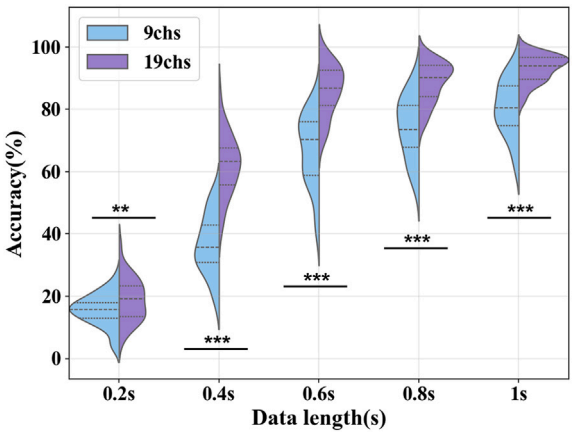



**Fig. 12.** Average classification accuracy for different number of channels on Dataset I.The 19-channel configuration led to a significant improvement in classification accuracy. (**p<0.01, ***p<0.001).

**Table 3**
Ablation experiments on the last two convolution modules of LGFCNN.

| Data length (s) | Modules | | ACC (%) | |
|---|---|---|---|---|
| | Conv1D | Point-wise Conv2D | Dataset I | Dataset II |
| 0.6 | ✓ | ✓ | 86.3±7.9 | 81.1±9.2 |
| | ✓ | | 82.2±7.8 | 80.50±8.4 |
| | | ✓ | 77.5±9.3 | 78.7± 9.7 |
| 0.8 | ✓ | ✓ | 88.6±6.3 | 86.4±13.1 |
| | ✓ | | 87.1±5.5 | 90.7±7.8 |
| | | ✓ | 85.7±5.8 | 90.1±7.1 |

connectivity information among channels enables more robust SSVEP feature extraction for classification tasks.

Additionally, experiments were conducted on a 9-channel configuration for all DL models using Dataset I to further evaluate the performance of the proposed LGFCNN model. Table 2 shows the average accuracy of each model at the data lengths of 0.6 s and 1 s. It demonstrates that even with reduced channels, the proposed LGFCNN model consistently had superior performance compared to other DL models.

### *3.4. Parameter analysis of loss function and DL neural network structure*

To investigate the effect of the regularization factor $\alpha_{lsr}$, Fig. 12 presents box plots illustrating the classification accuracy of the proposed model across different values of $\alpha_{lsr}$ on Dataset I and Dataset III, with the data length set to 0.8 s. For Dataset I, accuracy peaked at $\alpha_{lsr} = 0.1$, but declined as $\alpha_{lsr}$ increased to 0.25 and 0.5, suggesting that moderate regularization may enhance performance, whereas excessive regularization could introduce instability. In Dataset III, classification accuracy remained relatively stable, exhibiting a slight decline as $\alpha_{lsr}$ increased, suggesting lower sensitivity to regularization. Overall, introducing regularization factor $\alpha_{lsr} = 0.1$ improves accuracy compared to the non-regularized case $\alpha_{lsr} = 0$, suggesting that a moderate level of regularization can enhance performance.

To evaluate the impact of key modules of LGFCNN, an ablation study was conducted on the Conv1D and Point-wise Conv2D modules using data lengths of 0.6 s and 0.8 s. As shown in Table 3, it was obvious that models incorporating Conv1D consistently achieved the highest or near-highest accuracy, indicating its important contribution to further extract features after the preceding modules. While Conv1D alone performed reasonably well, the addition of Point-wise Conv2D further enhanced accuracy. Moreover, the optimal neural network architecture comprises both Conv1D and Point-wise Conv2D, as it maximizes classification performance.

To demonstrate the lightweight characteristic of the proposed model, Table 4 summarizes the computational complexity of LGFCNN and comparative baseline models, evaluated on Dataset I with 1 s EEG segments. It shows the number of parameters, the number of floating-point operations, inference time, training time, test time and classification accuracy. It should be noted that the reported training time for each model already includes the time required for data augmentation, as data augmentation was performed prior to model training. Since data augmentation is only applied during the training phase, it does not affect the inference

**Table 2**
Classification accuracy of different models under 9-channel configuration in Dataset I.

| | Data length (s) | Models | | | | | |
|---|---|---|---|---|---|---|---|
| | | LGFCNN | EEGNet | CCNN | AttentCNN | SSVEPNet | SSVEPformer |
| **ACC(%)** | 0.6 | 67.0±11.2 | 55.8±11.1 | 47.7±11.5 | 49.1±17.8 | 42.9±14.2 | 49.9±14.8 |
| | 1 | 80.3±8.9 | 76.4±12.1 | 65.2±15.1 | 64.8±20.6 | 65.5±15.4 | 76.0±15.1 |

**Table 4**
Computational complexity of all DL models on the Dataset I with the 1-s data length.

| Models | Params (k) | Flops (m) | Inference time (ms) | Train time (s) | Test time (ms) | Accuracy(%) |
|---|---|---|---|---|---|---|
| EEGNet | 52.1 | 186.2 | 3.3 | 109.9 | 12.9 | 86.3±9.4 |
| CCNN | 369.6 | 503.0 | 42.1 | 541.6 | 26.2 | 70.0±14.3 |
| AttentCNN | 675.4 | 10,891.5 | 39.2 | 228.5 | 47.0 | 67.7±14.8 |
| SSVEPNet | 34,639.3 | 4892.8 | 1.9 | 100.4 | 14.5 | 66.1±15.4 |
| SSVEPformer | 2330.0 | 414.7 | 9.5 | 149.0 | 16.0 | 76.3±14.7 |
| LGFCNN | 30.0 | 160.3 | 2.1 | 62.9 | 4.0 | 93.1±4.3 |

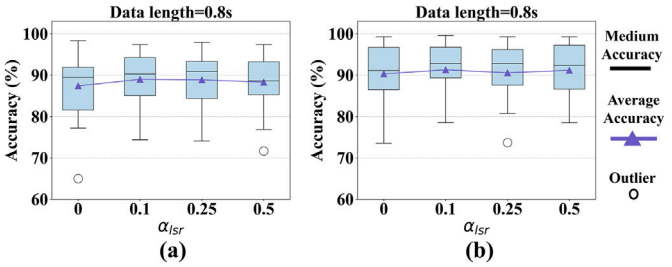



**Fig. 13.** Parameter sensitivity analysis for the loss function regulation factor $\alpha_{lsr}$ for Dataset I (a) and Dataset III (b). The model performance was moderately enhanced when $\alpha_{lsr}$ was set to 0.1.

or test time for individual samples. From the perspective of computation count, the proposed LGFCNN has substantially lower number of parameters and reduced floating-point operations, which is approximately 42 % lower than EEGNet, and dramatically lower than both CCNN and SSVEPformer, illustrating a clear advantage in model compactness. Although its inference time and training time were not the lowest, it still has rapid inference speed and test procedure, indicating the suitability of LGFCNN for real-time BCI applications. Combining the classification performance, LGFCNN underscored its optimal balance between computational efficiency and classification performance (Fig. 13).

## 4. Discussion

### 4.1. Analysis of data augmentation

In recent years, deep learning (DL) methods have provided a new perspective on SSVEP classification. However, the amount of training data is a crucial factor affecting the performance of DL models. Although amplitude scaling has been previously used for data augmentation in EEG and motor imagery studies [46,47], its application and systematic evaluation within SSVEP decoding frameworks remain limited. Most recent SSVEP data augmentation research has focused on strategies such as phase-locked time-shift [11], mask encoding [48], and principal component-based manipulations [49], which often require more complex transformations or additional parameter optimization. In this work, we demonstrate that amplitude scaling—a simple, training-free augmentation method—can be effectively integrated with topology-aware spatial filtering and deep learning models. The simplicity, efficiency, and compatibility of amplitude scaling make it a practical choice for real-world SSVEP applications. As shown in Fig. 11, the model's performance varied significantly with and without the slicing-and-scaling augmentation method. It was demonstrated that a larger amount of data, particularly in shorter data lengths (e.g., 0.4 s), could significantly improve performance by compensating for the limitation of reduced information in shorter segments [11]. To further explore the reason for the improvement, a t-SNE visualization analysis of the learned features from the last layer of the proposed model, as shown in the Fig. 8, was conducted. It is clear that after the data augmentation, there was a clearer clustering among different classes compared to the features before data augmentation. After data augmentation, it showed higher CH-index, higher silhouette coefficient and lower DB-index, reflecting better class cluster separation, better-defined class clusters and higher intra-class compactness, respectively. It could be concluded that the augmented samples might increase the diversity and expose the model to more variations (scale changes) in SSVEP, thereby making it more robust to unseen data [50]. Besides, the scaling factor had no significant impact on performance when the amplitude of the original data was scaled up or down within 20 %, which is consistent with the comparison results presented in [47]. This may be attributed to the fact that the frequency information of SSVEP remains unchanged when its amplitude is scaled up or down within the range of 20 %.

### 4.2. Analysis of the graph-based spatial filter and the proposed model

While most DL models for SSVEP classification incorporate spatial convolution layers to capture electrode relationships, they do not explicitly encode the physical topology of the electrode montage. This study proposes the use of a graph-based spatial filter before feeding data into the DL model, resulting in a significant improvement in accuracy and ITR, as shown in Fig. 11. Besides, while recent SSVEP studies have introduced GNN-based models to extract complex inter-channel dependencies through multiple graph convolutions, these methods rely on dynamically learned adjacency matrices and require substantial parameter optimization. Distinguishing itself from the GNN-based approaches, the proposed graph-based spatial filter is derived directly from the immutable physical layout of the EEG channels, serving as a fixed spatial filter prior without any trainable parameters. The performance results illustrate that embedding a spatial prior based on electrode topology is a highly effective strategy, offering balance between classification performance and model parsimony. Furthermore, the heat maps of the correlation coefficients between EEG channels, as shown in Fig. 10, provide a possible explanation for this improvement by reducing redundant information among channels and highlighting crucial components. As demonstrated in [41], the Laplacian operation might reduce spatial noise and help in identification of sources.

Besides comparison with DL models, comparison with traditional template-based methods was also conducted. As illustrated in Fig. 7, LGFCNN achieved superior performance under the CBATrain protocol but suffered a marked decline with CBFTrain, which can be attributed to the limited training data available for each class, especially given the larger number of stimulus classes. In contrast, eTRCA and TDCA performed better with CBFTrain but dropped markedly under CBATrain, likely due to their phase sensitivity and the difficulty of constructing effective templates from phase-diverse segments. These results demonstrate that deep-learning and template-based methods exhibit complementary strengths: while deep models such as LGFCNN can effectively capture complex temporal dynamics when sufficient phase-diverse data are available, template-based methods retain clear advantages in scenarios with limited training samples.

In terms of the proposed DL neural network, the overall comparison of accuracy and ITR, as depicted in Figs. 4 and 5, demonstrates its significant improvement. Specifically, for the 40-class Dataset I, it improved the ITR by almost 80bits/min when compared with the best baseline model in the case of 0.6 s. More importantly, the number of

parameters in this model was significantly lower than that of other baseline models, demonstrating the advantage of its lightweight structure while maintaining strong feature extraction capability. Ablation studies presented in Table 3 further underline the critical role of the adopted convolutional architecture, specifically the combination of a Conv1D layer followed by a point-wise Conv2D module. One possible interpretation for this performance gain is that Conv1D layer efficiently captures temporal characteristics of the SSVEP signal with fewer parameters, which might also reduce overfitting and improve generalization capability. Furthermore, SSVEP signals primarily consist of periodic waveforms at specific frequencies, making temporal characteristics inherently critical. Therefore,it suggests that emphasizing temporal feature extraction could be more advantageous than purely spatial modules.

### *4.3. Limitations and future work*

Although the proposed method achieved good performance in SSVEP classification, there are still some potential directions for further improvement in practical applications. First, the performance of the proposed model at shorter data lengths, such as 0.2 and 0.4 s, was not satisfactory. Since the graph structure was constructed based on the physical distance between channels, it might overlook the underlying dynamic interrelationships among channels that evolve over time. Therefore, exploring a dynamic graph-based filter could be a promising approach to address the challenges associated with short data lengths. Second, this study primarily focused on intra-subject SSVEP classification, which aligns closely with practical BCI applications emphasizing user-specific optimization scenario. It is acknowledged that cross-subject generalization remains a major challenge due to significant individual variability in EEG signals. While data augmentation increases within-subject data diversity, it does not address the distribution shift between subjects. In future work, advanced transfer learning and domain adaptation approaches could be employed to extend the proposed network for effective inter-subject generalization, thereby further broadening its practical applicability across diverse user populations.

## 5. Conclusion

This study presented a synergistic SSVEP decoding framework, designed to jointly address overfitting under limited data, spatial redundancy across EEG channels, and computational inefficiency. The architecture integrates a slicing-and-scaling data augmentation strategy, a topology-aware graph spatial filter based on the random walk Laplacian, and a lightweight convolutional network for efficient feature extraction and classification. Experimental results on two public datasets and a self-collected dataset validate the effectiveness of the framework, demonstrating superior accuracy and ITR compared to representative baselines. Moreover, the model demonstrates low computational complexity and fast inference speed in the offline evaluation, suggesting its feasibility for scalable and potentially real-time deployment in practical SSVEP-BCI systems. The architectural design, focusing on user-specific decoding, balances model compactness and recognition performance. Future work will explore transfer learning and domain adaptation techniques to extend the framework to cross-subject scenarios and further enhance its generalization across diverse user populations.

## 6. Data and code availability

The Benchmark dataset is available at http://bci.med.tsinghua.edu.cn/download.html. The Beta dataset is available at https://figshare.com/articles/dataset/The_BETA_database/12264401. The self-collected dataset can be obtained from the corresponding author upon request. The code of the proposed model is available at https://github.com/MRBMELMH.

## CRediT authorship contribution statement

**Rui Ma:** Writing – original draft, Visualization, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Yu Cao:** Writing – review & editing, Writing – original draft, Visualization, Validation. **Sheng Quan Xie:** Supervision, Funding acquisition. **Mingming Zhang:** Writing – review & editing, Validation. **Jun Li:** Writing – review & editing, Writing – original draft. **Zhi-Qiang Zhang:** Writing – review & editing, Writing – original draft, Visualization, Methodology, Funding acquisition, Conceptualization.

## Declaration of competing interest

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

## Data availability

Data will be made available upon request.

## References

[1] M. Zhang, J. Wu, J. Song, R. Fu, R. Ma, Y.-C. Jiang, Y.-F. Chen, Decoding coordinated directions of bimanual movements from EEG signals, IEEE Trans. Neural. Syst. Rehabil. Eng. 31 (2023) 248–259.
[2] Y. Zhang, Z. Li, S.Q. Xie, H. Wang, Z. Yu, Z.Q. Zhang, Multi-objective optimization-based high-pass spatial filtering for SSVEP-based brain–computer interfaces, IEEE Trans. Instrum. Meas. 71 (2022) 1–9.
[3] X. Deng, D. Xu, H. Huo, X. Hong, B. Liu, Multi-level hierarchical dynamic graph convolutional networks for motor imagery EEG analysis, Neurocomputing 626 (2025) 129594.
[4] J. Mladenovic, Standardization of protocol design for user training in EEG-based brain-computer interface, J. Neural. Eng. 18 (2021) 011003.
[5] Z. Miao, M. Zhaoa, X. Zhangbc, D. Ming, LMDA-Net: a lightweight multi-dimensional attention network for general EEG-based brain-computer interfaces and interpretability, Neuroimage 276 (2023) 120209.
[6] Z. An, F. Wang, Y. Wen, F. Hu, S. Han, A real-time CNN–BiLSTM-based classifier for patient-centered AR-SSVEP active rehabilitation exoskeleton system, Expert Syst. Appl. 255 (2024) 124706.
[7] L. Shao, L.Y. Zhang, A.N. Belkacem, Y.M. Zhang, X.Q. Chen, J. Li, H.L. Liu, EEG-controlled wall-crawling cleaning robot using SSVEP-based brain-computer interface, J. Healthc. Eng. 2020 (2020) 6968713.
[8] Y. Li, T. Kesavadas, SSVEP-based brain-computer interface for part-picking robotic co-worker, J. Comput. Inf. Sci. Eng. 22 (2) (2022) 021001.
[9] Y. Zhang, K. Qian, S.Q. Xie, C.Y. Shi, J. Li, Z.Q. Zhang, SSVEP-based brain-computer interface controlled robotic platform with velocity modulation, IEEE Trans. Neural. Syst. Rehabil. Eng. 31 (2023) 3448–3458.
[10] D.D. Kapgate, Application of hybrid SSVEP + P300 brain computer interface to control avatar movement in mobile virtual reality gaming environment, Behav. Brain Res. 472 (2024) 115154.
[11] J. Huang, Y. Lv, Z.Q. Zhang, B. Xiong, Q. Wang, B. Wan, P. Yang, Temporally local weighting-based phase-locked time-shift data augmentation method for fast-calibration SSVEP-BCI, IEEE Trans. Neural. Syst. Rehabil. Eng. 32 (2024) 2376–2387.
[12] K. Qin, R. Xu, X.J. He, Y.T. Zhang, X.Y. Wang, A. Cichocki, J. Jin, A time-local transformation steady-state visual evoked potentials recognition framework combined with filter banks, IEEE Trans. Instrum. Meas. 73 (2024) 2533012.
[13] Z.L. Lin, C.S. Zhang, W. Wu, X.R. Gao, Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs, IEEE Trans. Biomed. Eng. 54 (6) (2007) 1172–1176.
[14] X.G. Chen, Y.J. Wang, S.K. Gao, T.P. Jung, X.R. Gao, Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain-computer interface, J. Neural. Eng. 12 (4) (2015) 046008.
[15] Y. Zhang, G.X. Zhou, J. Jin, X.Y. Wang, A. Cichocki, Frequency recognition in SSVEP-based BCI using multiset canonical correlation analysis, Int. J. Neural Syst. 24 (4) (2014) 1450013.
[16] Q.Q. Liu, Y. Jiao, Y.Y. Miao, C.L. Zuo, X.Y. Wang, A. Cichocki, J. Jin, Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset cca, Neurocomputing. 378 (2020) 36–44.
[17] K. Wang, D.H. Zhai, Y.H. Xiong, L.Y. Hu, Y.Q. Xia, An MVMD-CCA recognition algorithm in SSVEP-based BCI and its application in robot control, IEEE Trans. Neural Netw. Learn. Syst. 33 (5) (2022) 2159–2167.
[18] M. Nakanishi, Y.J. Wang, X.G. Chen, Y.T. Wang, X.R. Gao, T.P. Jung, Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis, IEEE Trans. Biomed. Eng. 65 (1) (2018) 104–112.
[19] B.C. Liu, X.G. Chen, N.L. Shi, Y.J. Wang, S.K. Gao, X.R. Gao, Improving the performance of individually calibrated SSVEP-BCI by task- discriminant component analysis, IEEE Trans. Neural. Syst. Rehabil. Eng. 29 (2021) 1998–2007.
[20] J. Chen, et al, A transformer-based deep neural network model for SSVEP classification, Neural Netw. 164 (2023) 521–534.

[21] N. Waytowich, V.J. Lawhern, J.O. Garcia, J. Cummings, J. Faller, P. Sajda, J.M. Vettel, Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials, J. Neural. Eng. 15 (6) (2018) 066031.
[22] W. Ding, J. Shan, B. Fang, C. Wang, F. Sun, X. Li, Filter bank convolutional neural network for short time-window steady-state visual evoked potential classification, IEEE Trans. Neural. Syst. Rehabil. Eng. 29 (2021) 2615–2624.
[23] D. Xu, F. Tang, Y. Li, Q. Zhang, X. Feng, FB-CCNN: a filter bank complex spectrum convolutional neural network with artificial gradient descent optimization, Brain Sci. 13 (5) (2023) 780.
[24] Y. Pan, J. Chen, Y. Zhang, Y. Zhang, An efficient CNN-LSTM network with spectral normalization and label smoothing technologies for SSVEP frequency recognition, J. Neural. Eng. 19 (5) (2022) 056014.
[25] Z. Wang, C.M. Wong, B. Wang, Z. Feng, F. Cong, F. Wan, Compact artificial neural network based on task attention for individual SSVEP recognition with less calibration, IEEE Trans. Neural. Syst. Rehabil. Eng. 31 (2023) 2525–2534.
[26] J. Chen, F. Sun, W. Zhang, S. Zhang, K. Liu, C. Qi, Attention-based multimodal tCNN for classification of steady-state visual evoked potentials and its application to gripper control, IEEE Trans. Neural Netw. Learn. Syst. (2023) 1–9.
[27] D. Klepl, M. Wu, F. He, Graph neural network-based EEG classification: a survey, IEEE Trans. Neural. Syst. Rehabil. Eng. 32 (2024) 493–503.
[28] X.L. Tang, J. Zhang, Y.D. Qi, K. Liu, R. Li, H.M. Wang, A spatial filter temporal graph convolutional network for decoding motor imagery EEG signals, Expert Syst. Appl. 238 (2024) 121915.
[29] K. Yang, Z. Yao, K. Zhang, J. Xu, L. Zhu, S. Cheng, J. Zhang, Automatically extracting and utilizing EEG channel importance based on graph convolutional network for emotion recognition, IEEE J. Biomed. Health Inform. 28 (8) (2024) 4588–4598.
[30] S. Zhang, D. An, J. Liu, J. Chen, Y. Wei, F. Sun, Dynamic decomposition graph convolutional neural network for SSVEP-based brain-computer interface, Neural Netw. 172 (2023) 106075.
[31] S.B. Zhang, D. An, J.C. Liu, Y.G. Wei, F.C. Sun, Unsupervised multi-source variational domain adaptation for inter-subject SSVEP-based BCIs, Expert Syst. Appl. 238 (2024) 122155.
[32] L. Deng, P.R. Li, H.K. Zhang, Q.Y. Zheng, S.H. Liu, X.M. Ding, M.Q. Wang, D.R. Gao, TSMNet: a comprehensive network based on spatio-temporal representations for SSVEP classification, Biomed. Signal Process. Control 105 (2025) 107554.
[33] Y. Liu, W. Dai, Y. Liu, D. Hu, B. Yang, Z. Zhou, An SSVEP-based BCI with 112 targets using frequency spatial multiplexing, J. Neural. Eng. 21 (3) (2024) 036004.
[34] Y. Wang, X. Chen, X. Gao, S. Gao, A benchmark dataset for SSVEP-based brain–computer interfaces, IEEE Trans. Neural. Syst. Rehabil. Eng. 25 (10) (2017) 1746–1752.
[35] B.C. Liu, X.S. Huang, Y.J. Wang, X.G. Chen, X.R. Gao, BETA: a large benchmark database toward SSVEP-BCI application, Front. Neurosci. 14 (2020) 627.
[36] A. Delorme, S. Makeig, EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis, J. Neurosci. Methods 134 (1) (2004) 9–21.
[37] M. Hassan, F. Wendling, Electroencephalography source connectivity: aiming for high resolution of brain networks in time and space, IEEE Signal Process. Mag. 35 (3) (2018) 81–96.
[38] C.M. Michel, M.M. Murray, Towards the utilization of EEG as a brain imaging tool, NeuroImage. 61 (2) (2012) 371–385.
[39] P.L. Nunez, R. Srinivasan, Electric Fields of the Brain: the Neurophysics of EEG, Oxford University Press, Oxford, 2006.
[40] C.E. Tenke, J. Kayser, Generator localization by current source density (CSD): implications of volume conduction and field closure at intracranial and scalp resolutions, Clin. Neurophysiol. 123 (12) (2012) 2328–2345.
[41] D.J. McFarland, The advantages of the surface laplacian in brain-computer interface research, Int. J. Psychophysiol. 97 (3) (2015) 271–276.
[42] C. Carvalhaes, J.A. de Barros, The surface laplacian technique in EEG: theory and methods, Int. J. Psychophysiol. 97 (3) (2015) 174–188.
[43] M. Nakanishi, Y.J. Wang, Y.T. Wang, T.P. Jung, A comparison study of canonical correlation analysis based methods for detecting steady-state visual evoked potentials, PLoS One 10 (2015).
[44] A. Ravi, N.H. Beni, J. Manuel, N. Jiang, Comparing user-dependent and user-independent training of CNN for SSVEP BCI, J. Neural. Eng. 17 (2) (2020) 026028.
[45] R. Wei, C. Hua, J. Chen, D. Mu, J. Zhao, Attention-based multiscale tCNN for SSVEP classification and its application to bionic intelligent soft gripper control, IEEE Trans. Syst. Man Cybern.: Syst. 55 (9) (2025) 6349–6358.
[46] D. Freer, G.-Z. Yang, Data augmentation for self-paced motor imagery classification with C-LSTM, J. Neural. Eng. 17 (1) (2020) 016041.
[47] Z. Wang, S. Li, J. Luo, J. Liu, D. Wu, Channel reflection: knowledge-driven data augmentation for EEG-based brain–computer interfaces, Neural Netw. 176 (2024) 106351.
[48] W. Ding, et al, A novel data augmentation approach using mask encoding for deep learning-based asynchronous SSVEP-BCI, IEEE Trans. Neural. Syst. Rehabil. Eng. 32 (2024) 875–886.
[49] W.L. Ding, A.P. Liu, L.L. Cheng, X. Chen, Data augmentation using masked principal component representation for deep learning-based SSVEP-BCIs, J. Neural. Eng. 22 (3) (2025) 036025.
[50] E. Lashgari, D. Liang, U. Maoz, Data augmentation for deep-learning-based electroencephalography, J. Neurosci. Methods 346 (2020) 108885.

## Author biography

**Rui Ma** received her bachelor's degree from Southern Medical University, Guangzhou, China, in 2018, and her Master' degree from Southern University of Science and Technology, Shenzhen, China, in 2022. Currently, she is pursuing a Ph.D. degree in University of Leeds. Her research focuses on neural signal processing, machine learning, deep learning in BCI systems.

**Yu Cao** received the B.S. degree in automation from the Wuhan University of Technology, Wuhan, China, in 2011, and the M.E. degree in software engineering and the Ph.D. degree in control science and engineering from the Huazhong University of Science and Technology, Wuhan, in 2014 and 2020, respectively. From 2021 to 2024, he was a Postdoctoral Researcher with the School of Artificial Intelligence and Automation, Huazhong University of Science and Technology. He is currently a U.K. Research and Innovation (UKRI) Research Fellow in Rehabilitation Robotics with the School of Electronic and Electrical Engineering, Faculty of Engineering and Physical Sciences, University of Leeds, Leeds, U.K. His current research interests include human–robot interaction control, neuromusculoskeletal modeling, and wearable robotic systems. Dr. Cao received funding from the Marie Skłodowska-Curie Actions, supported by UKRI under the Horizon Europe Guarantee in 2023.

**Sheng Quan Xie** received the Ph.D. degree in mechanical engineering from the University of Canterbury, New Zealand, in 2002. He joined The University of Auckland, New Zealand, in 2003, and became a Chair Professor in (bio)mechatronics in 2011. Since 2017, he has been the Chair of Robotics and Autonomous Systems with the University of Leeds, Leeds, U.K. His current research interests include medical and rehabilitation robots, and advanced robot control. He is an elected fellow of the Institution of Professional Engineers New Zealand, a fellow of American Society of Mechanical Engineers (ASME), and a fellow of the Institution of Mechanical Engineers (IMechE).

**Mingming Zhang** received the Ph.D. degree in mechanical engineering from The University of Auckland, Auckland, New Zealand, in 2015. Since 2015, he has been a Research Fellow and/or a Visiting Research Fellow with the Department of Mechanical Engineering, The University of Auckland. In 2018, he was with the Southern University of Science and Technology as an Associate Professor. His research interests include medical and rehabilitation robotics, haptics control, human–robot physical interaction, and brain–computer interaction.

**Jun Li** received the bachelor's degree in mathematics from Hubei University for Nationalities, in 1995, and the master's degree in computer application from the Huazhong University of Science and Technology, in 1999. He currently works as a Professor with the College of Intelligent Systems Science and Engineering, Hubei Minzu University, Enshi, China. His research interests include computer graphics and image processing, virtual reality technology, deep learning and digital protection of intangible cultural heritage.

**Zhi-Qiang Zhang** received his B.Eng in Computer Science and Technology from Tianjin University in 2005 and Ph.D. in Electrical Engineering from the University of Chinese Academy of Sciences in 2010, he joined Imperial College London as a research associate, where he worked for five and a half years before moving to the University of Leeds in 2016. He currently holds the position of Chair of Biomedical Engineering. His research lies at the intersection of artificial intelligence, robotics, and biomedical engineering.