TBMSCCN: Two-Branch Multi-Scale Convolutional Correlation Network for Steady-State Visual Evoked Potential Classification

Xinjie He, Ian Daly, Wenhao Gu, Yixin Chen, Xiao Wu, Weijie Chen, Xingyu Wang, and Andrzej Cichocki, *Life Fellow, IEEE,* Jing Jin*, *Senior Member, IEEE.*

***Abstract*—In recent years, artificial neural networks have been effectively used to improve the target recognition performance of steady-state visual evoked potential (SSVEP) based Brain-Computer interfaces (BCIs). However, these models require the collection of a large number of calibration trials from users, which typically results in a poor user experience. When fewer calibration trials are acquired this leads to insufficient training of model parameters and weak recognition performance. To tackle these issues, this study proposes a two-branch multi-scale convolutional correlation network (TBMSCCN) in which a correlation network framework is introduced to reduce the model training parameters and prior knowledge of the SSVEP is used to enhance the model representation ability and convergence. First, a multi-scale temporal convolution module is designed to learn local temporal dependencies in a parallel two-branch feature extraction module. Next, a contrastive loss function is constructed in the latent feature space, which can guide the model to learn the intra-class consistent features while speeding up model convergence. Finally, a group convolution module is used as a decision layer to reduce the network parameters, while learning distinguishability features between targets and non-targets. Our offline tests on two public datasets show that proposed TBMSCCN method outperforms TRCA, eTRCA, DNN, Conv-CA and Bi-SiamCA in individual calibration scenarios, which can achieve an average information transform rates (ITRs) of 378.03 ± 139.18 bit/min and 198.92 ± 111.27 bit/min on the "Benchmark" dataset and the "Beta" dataset respectively. Additionally, proposed TBMSCCN method outperform FBCCA, ttCCA, EEGNet, and TST-CFSR in calibration-free scenarios. Furthermore, an online Chinese spelling experiment confirmed the real-world effectiveness of the proposed method. The proposed model has the characteristics of low parameter and strong robustness, which can facilitate the practical engineering application of SSVEP-Based-BCI system. The code is available at https://github.com/xinjieHe123/TBMSCCN.**

***Index Terms*—Brain computer interface (BCI), steady-state visual evoked potential (SSVEP), artificial neural networks, correlation network, multi-scale temporal convolution.**

Manuscript received xxxx; revised xxxx; accepted xxxx. Date of publication xxxx; date of current version xxxx. This work was supported by the Grant National Natural Science Foundation of China under Grant 62176090 and STI 2030-major projects 2022ZD0208900; in part by Shanghai Municipal Science and Technology Major Project under Grant 2021SHZDZX, This research is also supported by Project of Jiangsu Province Science and Technology Plan Special Fund in 2022 (Key research and development plan industry foresight, fundamental research fund for the central universities JKH01231636 and key core technologies) under Grant BE2022064-1. (Corresponding author: Jing jin).

Jing Jin are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China, and also school of Mathematics, East China University of Science and Technology, Shanghai 200237, China (e-mail: jinjingat@gmail.com;).

Xinjie He, Wenhao Gu, Yixin Chen, Xiao Wu, and Xingyu Wang are with the Key Laboratory of Smart Manufacturing in Energy Chemical Process, Ministry of Education, East China University of Science and Technology, Shanghai 200237, China (e-mail: xinjieHe@mail.ecust.edu.cn; 1193489824@qq.com; yixinchen1022@foxmail.com; wuxiao121409@163.com; wjchen827@foxmail.com; xywang@ecust.edu.cn;)

Ian Daly is with the Brain-Computer Interfacing and Neural Engineering Laboratory, School of Computer Science and Electronic Engineering, University of Essex, Wivenhoe Park, Colchester, Essex CO4 3SQ, UK (e-mail: i.daly@essex.ac.uk).

Andrzej Cichocki is with the Systems Research Institute of Polish Academy of Sciences, 01-447bWarsaw, and Nicolaus Copernicus University(UMK), 87-100 Torun, Poland (e-mail: cichockiand@gmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

Digital Object Identifier xxxx.

I. INTRODUCTION

BRAIN computer interfaces (BCIs) can provide a direct communication channel between the human brain and computers or other external control devices [1-3]. Based on the signal acquisition type, BCI techniques can be categorized as invasive, semi-invasive, and non-invasive. Non-invasive BCIs have attracted the widespread attention of researchers due to their inherent safety and portability [4-5]. Depending on the stimuli and neural activity used to control them, non-invasive BCI can be classified into SSVEP-based BCIs [6,7,53], event-related potentials (ERP) based BCIs [8-11], and event-related synchronization or de-synchronization based BCIs [12,13]. Of these BCI categories the SSVEP-BCI has significant advantages including high information transfer rates (ITRs), a strong signal noise ratio (SNR), and vocabulary size. Currently, SSVEP-BCI systems are primarily used in the field of disability assistance and human-computer interaction such as character spelling systems [14], wheelchair control systems [15], smart home systems [16], and human-computer collaboration systems [17-18].

The SSVEP-BCI system may be considered in two parts: the paradigm design and the decoding algorithm. A good decoding algorithm is key to ensure stable and high performances operation of the SSVEP-BCI and is the focus of our work. SSVEP decoding strategies can be classified into unsupervised learning methods and supervised learning

methods. The first unsupervised learning methods deployed for SSVEP-BCIs used power spectrum density analysis (PSDA) to extract the fundamental and octave energy information from the pre-processed EEG at the target frequencies [19]. A correlation analysis method, based on sine-cosine reference templates and spatial filters, was then used to decode the selected SSVEP frequency. Implementations of this approach include canonical correlation analysis (CCA) [20], filter bank CCA (FBCCA) [21], multiple synchronization index (MSI) [22], filter bank MSI (FBMSI) [23], multivariate variational mode decomposition CCA (MVMD-CCA) [24] and spatio-temporal equalization CCA (STE-CCA) [25]. In recent years, online iterative learning strategies have achieved a new breakthrough in unsupervised learning for SSVEP decoding. Wong et al [26] designed an online adaptive canonical correlation analysis (OACCA) method, which can continuously adjust the spatial filter weights in order to learn participant-specific information using the historical sample information. Jin et al [27] proposed an online adaptive canonical correlation analysis method based on spatiotemporal noise estimation. This approach not only removes redundant noise components in the spatial filter, but also learns individual adaptive EEG information. Compared to supervised learning methods, unsupervised learning methods do not require online collection of participant data. Unfortunately, these methods are susceptible to interference from transient spontaneous EEG noise, resulting in low decoding performance.

Supervised learning methods have also been shown to effectively improve the SSVEP decoding performance by collecting a large amount of calibration data from users. Spatial filter-based correlation analysis methods and end-to-end deep learning-based methods are currently the most popular supervised learning methods. Nakanishi et al [28] proposed a task related component analysis (TRCA) and ensemble TRCA (eTRCA) method, which can capture the task-related components of the SSVEP between different samples from the same target stimuli. Subsequently, various versions of the TRCA method have been proposed to further improve the decoding performance of the SSVEP-BCI system. For example, Wang et al [29] proposed a 2D linear discriminant analysis (2D-LDA) method and 2D locally preserved projection (2D-LPP) technique, which can eliminate redundant components from different spatial filters while capture common spatial filter information. Huang et al [30] proposed a neighboring stimuli task-related component analysis (NS-TRCA) method, which can obtain the peripheral visual stimulus information to enhance SSVEP frequency features. Huang et al [32] proposed a phase channel delay alignment task-related component analysis method based on wave transmission theory. Lan et al [33] used source-domain participant-specific EEG signals to enhance the online recognition performance of SSVEPs by maximizing the task-relevant components between participants. Ke et al [34] proposed a periodic repeated component analysis (PRCA) method in which the SSVEP signal was divided into multiple period samples, which were used to train a spatial filter and reference templates. However, existing spatial filter-based correlation analysis methods ignore the nonlinear features in the EEG and there is a need for new SSVEP frequency characterization mechanism that address this.

With the development of artificial neural networks, deep learning-based SSVEP decoding methods have attracted growing attention due to their powerful nonlinear feature extraction capabilities. Guney et al [35] designed an end-to-end deep neural network (DNN) framework consisting of a sub-band combine layer, a spatial convolutional layer, two temporal convolutional blocks, and a fully-connected layer that they reported could achieve an average ITRs of 265.23 bit/min on “Benchmark” dataset. Variations on this method have been, subsequently, proposed to enhance the decoding performance of SSVEP such as task related component analysis network (TRCANet) [36], filter-bank EEG neural network (FB-EEGNet) [37], SSVEP transformer neural network (SSVEPformer) [38] and ensemble deep neural network (Ensemble-DNN) [39]. These types of end-to-end network design frameworks can improve the SSVEP identification performance to some extent, but still have significant limitations. For example, global and fine-tuned training strategies can lead to models that are heavily dependent on the quality of the source data. Consequently, it is difficult to adequately train the model parameters using only a small number of calibration samples from available participants, resulting in model underfitting. The bi-directional correlation network framework is on possible solution to this problem. This framework of models are able to make full use of prior knowledge (sin-cosine reference templates or averaging templates) to supervise model learning, reducing the training parameters and computational complexity of the network. They map training samples and reference templates into a latent space through a two-layer parallel feature extraction module, and then output category probability information through a correlation analysis layer. Examples of models that follow this framework include convolutional correlation analysis (Conv-CA) [40] and bidirectional siamese correlation analysis [41] (Bi-SiamCA). However, these networks achieve worse performance when the size of the calibration data is reduced. To address this issue, it is possible to increase the amount of calibration data using traditional sliding time window (STW) techniques, but this imposes a greater calibration time cost to model training.

Inspired by the traditional correlation template matching method based on spatial filter, this study proposed a lightweight TBMSCCN decoding approach. The approach combines a deep learning framework with a traditional correlation template matching mechanism can reduce the network training parameters while enhancing the nonlinear feature extraction capability of the model. Our innovations are summarized below.

(1) We introduce temporal filters with different convolutional kernel sizes in a parallel two-branch feature extraction module to learn multiscale locally optimal temporal

properties while generating potential feature vectors with stable target frequency characteristics.

(2) We construct a contrastive loss function in the latent feature space, which can guide the model to learn intra-class consistent features, while speeding up model convergence.

(3) We use a group convolution module as a decision layer to reduce the network parameters, while learning distinguishable features between targets and non-targets.

(4) We use two public datasets to validate the effectiveness of our proposed TBMSCCN module in individual calibration scenarios and calibration-free scenarios, respectively.

(5) The proposed model has the characteristics of low parameter and strong robustness, which can facilitate the practical engineering application of SSVEP-Based-BCI system.

II. Materials and Methods

A. TBMSCCN network framework

Our TBMSCCN network contain four sub-modules. Specifically, we use an EEG signal multi-scale convolutional feature extraction module (ES-MSCFEM), an EEG templates multi-scale convolutional feature extraction module (ET – MSCFEM), a correlation analysis layer (CAL), and a correction group convolutional classification module (CGCCM). The structure and parameters of our TBMSCCN network are shown in **Fig.1** and **Table I**.

(1) ES-MSCFEM

The EEG training data can be represented by a four-dimensional tensor: $\mathbf{X}_{\mathbf{EEG}} \in \mathbb{R}^{Nc \times N_s \times N_b \times N_f}$. Where, N_c denotes the number of channels, N_s denotes the number of samples in time, N_b denotes the number of training examples, and N_f denotes the number of target stimulus classes. The SE-MSCFEM module input data shape is $\mathbf{X}_{\mathbf{ES}} \in \mathbb{R}^{1 \times N_s \times N_c}$ and is taken from the EEG data, the total number of training examples is $N = N_b \times N_f$. The ES-MSCFEM module is chiefly used to explore the spatial-temporal target frequency characteristics of the multi-channel EEG signals and map them into latent feature space vectors $\hat{\mathbf{X}}_{\mathbf{ES}} \in \mathbb{R}^{N_s \times 1}$. The module representation mapping relationships are:

$$\hat{\mathbf{X}}_{\mathbf{ES}} = f\left(\mathbf{X}_{ES}\right) \quad (1)$$

where $f(\bullet)$ is the ES-MSCFEM network module in Fig.1, which consists of a spatial-temporal convolution block, a multi-scale temporal convolution block, and a channel convolution module. The first layer is a spatial-temporal convolutional block with 16 filters of 9*9 kernels, which is used to capture multiple spatial-temporal correlation representations between the input channels (N_c = 9) and the local time series (N_s = 9) in the EEG signals. Inspired by the temporal local filter approach [31,42], we use parallel 2D convolutional blocks with different time scales to capture the locally optimal temporal characteristics of the SSVEP signal in the time series from a single channel. In the multi-scale temporal convolution module, the number of temporal filters in each branch is set to 1, 1, 1 and the size of the filter kernel is 9*1, 17*1 and 27*1 respectively. Next, we employ the Concatenate module to splice the EEG signals after each branch and the Mean module to take the grand average in the scale dimension. The third layer is the channel convolution module, which employs a 1 * N_c filter in the spatial dimension to capture the optimal channel combination information about the SSVEP. After these three layers, potential feature vectors with stable SSVEP frequency characteristics are generated.

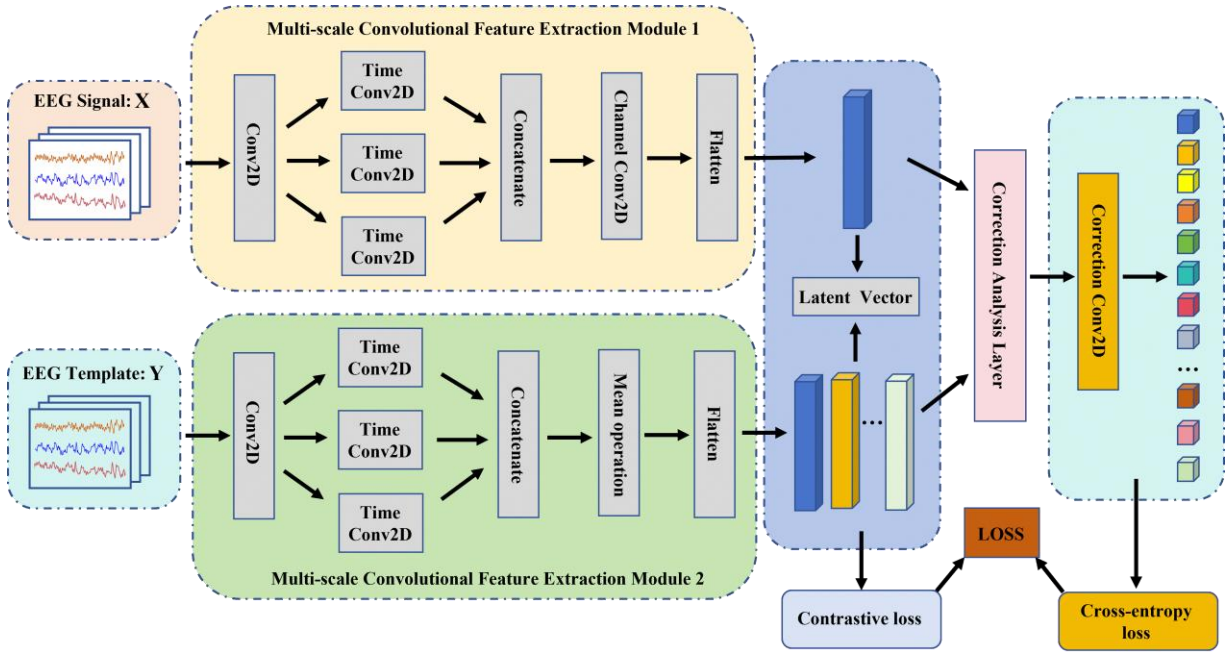


Fig.1 Our proposed TBMSCCN network, which contain four sub-modules: ES-MSCFEM, ET-MSCFEM, CAL, and CGCCM.

(2) ET-MSCFEM

The ET-MSCFEM module utilizes averaged templates as a priori knowledge for guiding model learning. The ET-MSCFEM module's input data shape is $\mathbf{Y_{ET}} \in \mathbb{R}^{N_c \times N_s \times N_f}$, which was constructed by averaging the training samples. The ET -MSCFEM module is used to extract the spatial energy properties at each target frequency and map them into the potential feature space. The module representation mapping relationships are:

$$\hat{\mathbf{Y}}_{\mathbf{ET}} = g\left(\mathbf{Y}_{\mathbf{ET}}\right) \tag{2}$$

where, $\hat{\mathbf{Y}}_{\mathbf{ET}} \in \mathbb{R}^{N_s \times N_f}$ is time series of each target stimulus frequency mapped in latent feature space. The term $g(\bullet)$ denotes the ET-MSCFEM module in **Fig.1**, which consists of two parts: a spatial convolution block and a multi-scale temporal convolution. In the spatial convolution module, we used 40 filters with a kernel size of N_c*1 to assign weights to all channel combinations for a short time for each target stimulus. For the multi-scale temporal convolution module, filters with different kernel sizes are used to capture the optimal temporal local representation information at different target frequencies. The number of temporal filters in each branch is set to 1, 1, 1 and the size of the filter kernel is 9*1, 17*1 and 27*1 respectively. It is worth noting that each convolution operation in the MSCFEM module needs to be performed independently for each target frequency.

Table I TBMSCCN network parameter settings. The channel number of EEG N_c is 9. This network parameter does not change with the length of the time window.

Module	Layer	Kernel	Filter	Parameters	Output
ES-MSCFEM	Input				(Batch, 1, N_s, N_c)
	Conv2D	(9, N_c)	16	(9*N_c + 1) *16	(Batch,16, N_s, N_c)
	Multi-Scale Time Conv2D	(9,1) //(17, 1) //(27, 1)	1/1/1	851	(Batch, 1, N_s, N_c)
	Channel Conv2D	(1, N_c)	1	1*N_c+1	(Batch, 1, N_s, 1)
ET-MSCFEM	Input				(Batch, *Ns*, *Nc*, N_f)
	Conv2D	(N_c,1)	40	(N_c*N_c + 1) *40	(Batch, 40, *Ns*, N_f)
	Multi-Scale Time Conv2D	(9,1) //(17, 1) //(27, 1)	1/1/1	2123	(Batch, 1, *Ns*, N_f)
Correlation Analysis Layer	Input1				(Batch, 1, N_s, 1)
	Input2				(Batch, 1, *Ns*, N_f)
	CAL (self-defined)			0	(Batch, N_f)
CGCCM	Input				(Batch, N_f,1,1)
	Correction Group Conv2D	(1,1)	N_f	(1*1+1) * N_f	(Batch, N_f)

(3) Correlation Analysis Layer

In the correlation analysis layer, the cosine similarity function was defined to learn the category information from different stimulus targets and output the correction feature vector. The latent feature vector of the EEG signal mapped by the ES-MSCFEM module is $\hat{\mathbf{X}}_{\mathbf{ES}} \in \mathbb{R}^{N_s \times 1}$. The latent feature vector of the EEG templates signal mapped by the ET-MSCFEM module is $\hat{\mathbf{Y}}_{\mathbf{ET}} = \left[\hat{\mathbf{Y}}_{\mathbf{ET}}^{1}, \cdots \hat{\mathbf{Y}}_{\mathbf{ET}}^{i}, \cdots, \hat{\mathbf{Y}}_{\mathbf{ET}}^{N_f}\right] \in \mathbb{R}^{N_s \times N_f}$. The cosine similarity at the *i*-th target stimulus frequency is calculated via:

$$\theta_i = \frac{\hat{\mathbf{X}}_{\mathbf{ES}}{}^{\mathrm{T}} \cdot \hat{\mathbf{Y}}_{\mathbf{ET}}^{i}}{\sqrt{\hat{\mathbf{X}}_{\mathbf{ES}}{}^{\mathrm{T}} \hat{\mathbf{X}}_{\mathbf{ES}}} \sqrt{\hat{\mathbf{Y}}_{\mathbf{ET}}^{i\ \mathrm{T}} \hat{\mathbf{Y}}_{\mathbf{ET}}^{i}}} \tag{3}$$

where, T denotes the transpose operation of the matrix. The correlation eigenvector **C** is calculated as:

$$\mathbf{C} = \left[\theta_1, \cdots, \theta_i, \cdots \theta_{N_f}\right] \in \mathbb{R}^{N_f} \tag{4}$$

(4) CGCCM

In the classification layer, a group convolution module was added to reduce the computational complexity of the network, while learning distinguishable correlation features between targets and non-targets. In this module, the number of filters is set to 40, the filter kernel size is set to 1*1, and the output is the probability value of the classification target.

(5) Loss function

He et.al. and Hadsell et.al. [43-44] showed that a contrastive loss function can maximize the consistency of the EEG signal and the EEG reference template in the latent space. Subsequently, in our TBMSCCN network framework, we introduce a contrastive loss module, which makes the similarity between EEG signals and the reference signal increase at the target frequency, while the similarity between EEG signals and the reference signal decreases for the other frequencies. The contrastive loss of the EEG signal belonging to the *m*-th target stimulus is defined as:

$$L_c = -\log \frac{\exp\left[\theta_m / \tau\right]}{\sum_{i=1}^{N_f} \exp\left(\theta_i / \tau\right)} \quad (5)$$

where, $\exp(\bullet)$ denotes exponential operation, θ_i denotes the correlation coefficient value of the *i*-th label corresponding to the training sample. τ denotes the temperature hyperparameter, which was used to control the sensitivity to changes in contrastive loss similarity. The frequency-periodic component of the SSVEP signal is enhanced with increasing time-window length, the value of the contrastive loss similarity metric becomes more reliable, and the value of the temperature hyperparameter τ can be increased adaptively. Based on the above analysis, the adaptive function for setting the value of the temperature hyperparameter τ is

$$\tau = \exp(\alpha t_w + \beta) + \gamma \quad (6)$$

where, t_w denotes the data length of EEG signal, and γ is the regularization term to avoid τ values close to 0. After the output of the classification layer of the TBMSCCN network framework, we introduce the cross-entropy loss for learning the model parameters, which is defined as

$$L_g = -\sum_{i=1}^{N_f} y_i \log(\hat{y}_i) \quad (7)$$

where, y_i is true label, $\hat{y}_i$ is predict label. The loss function L was constructed as

$$L = L_g + \lambda L_c \quad (8)$$

where, λ is the scale control factor, which has a value of 1.

After construction of the loss function, we used the Adam optimizer in the Pytorch environment for network parameter optimization. The ES-MSCFEM and ET-MSCFEM feature extraction modules are independent of each other and the network parameters are not shared. The Pseudo-Code of proposed TBMSCCN as shown in **Algorithm 1**.

Algorithm 1 Pseudo-Code of Proposed **TBMSCCN**

Input: EEG Training trial $\mathbf{X}_{\mathbf{ES}}$, EEG averaged templates $\mathbf{Y}_{\mathbf{ET}}$, Test data $\mathbf{X}^t$

Output: Classification result *f*.

Training phase:

for epoch = 1 to End-epoch **do:**
  Get EEG latent feature space vectors $\hat{\mathbf{X}}_{\mathbf{ES}}$ by equation (1).
  Get EEG template latent feature $\hat{\mathbf{Y}}_{\mathbf{ET}}$ by equation (2).
  Get the correlation eigenvector **C** by equation (3) to (4).
  Get the contrastive loss L_c by equation (5).
  Get the model output value by group convolution module.
  Get the cross-entropy loss L_g by equation (7).
  Get the Loss value L by equation (8).
  Adam optimizer
end
Return TBMSCCN Network.

Testing phase:

for m = 1 to N_f **do:**
  for b = 1 to N_b **do:**
    Get the test data $\mathbf{X}^t$.
    Load the EEG averaged templates $\mathbf{Y}_{\mathbf{ET}}$.
    Load the **TBMSCCN Network.**
    Get the classification result *f* by **TBMSCCN model**
  end
end

B. Dataset description and pre-processing

(1) Dataset description

The "Benchmark" dataset [45] and the "Beta" dataset [46] were used to evaluate the effectiveness of our proposed TBMSCCN network. The "Benchmark" and "Beta" datasets was provided by the Tsinghua University BCI lab. These datasets can be found on the official website of the Tsinghua University BCI lab (http://bci.med.tsinghua.edu.cn/download.html). To record these datasets thirty-five healthy participants were recruited to participate in a 40-target character spelling cue experiment. Each participant's EEG data was recorded via a 64-channel EEG cap with 1000 Hz sample frequency and contained task labelling information. Each character target was encoded by the Joint Frequency and Phase Modulation (JFPM) method with a frequency range of 8-15.8 Hz at 0.2 Hz intervals and a phase range of 0-1.5π at 0.5π intervals. Each participant performed a 6-block experiment with a 40-character cued spelling task. Each trial contained a 0.5s cue time, a 5s stimulus time, and a 0.5s rest time.

The "Beta" dataset [46] is consistent with the stimulus frequencies and phase information used in the "Benchmark" dataset. In this dataset a further seventy healthy participants were recruited to participate in a 40-target character spelling cue experiment within a high electromagnetic noise environment. Each participant performed a 4-block experiment with a 40-character cued spelling experiment. The stimulus duration was 2 s for the first 15 participants and 3 s for the remainder of the participants.

(2) Pre-processing

Raw EEG data was down-sampled to 250Hz. Nine channels in the occipital region (Pz, PO5, PO3, POz, PO4, PO6, O1, Oz, and O2) were selected for SSVEP signal analysis. A Chebyshev I-type bandpass filter with cutoff frequencies from 6 Hz to 90 Hz and cutoff corner frequencies from 4 Hz to 100 Hz was used to remove the interference of low and high frequency noise and to retain the fundamental and harmonic octave information of the SSVEP stimulation frequency. The 50Hz notch filter was used to remove power frequency noise interference. Considering visual latency and cue time, we conducted experiments using EEG data after 0.64s[45] and 0.63s[46], respectively.

(3) Performance evaluation

We use the average recognition accuracy P and ITRs to statistically measure model performance. ITRs can be used to measure the character spelling rate, which is defined as

$$\mathrm{ITR}=\frac{60}{T}\times\left[\log_2\left(N_f\right)+P\log_2\left(P\right)+\left(1-P\right)\log_2\left(\frac{1-P}{N_f-1}\right)\right] \quad (9)$$

where T is the window length, which is equal to the sum of the cue time and the recognition time. N_f denotes the number of stimulus targets and P represents the average recognition accuracy.

C. Comparison experiments setup

To validate the effectiveness of the proposed method in different application scenarios, we implemented comparison experiments with SoTA method in less sample individual calibration scenario and cross subject calibration-free scenario.

(1) Individual calibration scenario

In individual calibration scenario, we use the leave-one-block cross-validation method. The training and test data use the same data length for each time window. We restrict the use of the sliding time window approach to expand the number of training sample sets because it produces a higher calibration cost. For the "Benchmark" dataset, six folds cross validation was used, in which five blocks were used for training and one block was used for testing. For the "Beta" dataset, four folds cross validation was used, in which three blocks were used for training and one block was used for testing.

Five other SoTA methods (TRCA [28], eTRCA [28], DNN [35], Conv-CA [36] and Bi-SiamCA [41]) were used to compare proposed TBMSCCN network.

(2) *Cross subject calibration-free scenario*

In the calibration-free scenario, we employed the leave one-subject-out (LOSO) cross-validation method to assess the generalization performance of the model. Specifically, for the "Benchmark" dataset, the model was trained on data from 34 subjects, with the remaining 1 subject reserved for evaluation. Similarly, for the "Beta" dataset, the model was trained on data from 69 subjects, while the remaining 1 subject was used for testing. To ensure a comprehensive comparison, five SoTA calibration-free SSVEP-BCI algorithms were selected as baselines. The baselines method include FBCCA [21], CSSFT[48],TST-CSFR[47], EEGNet[49] and EEG Conformer[50].

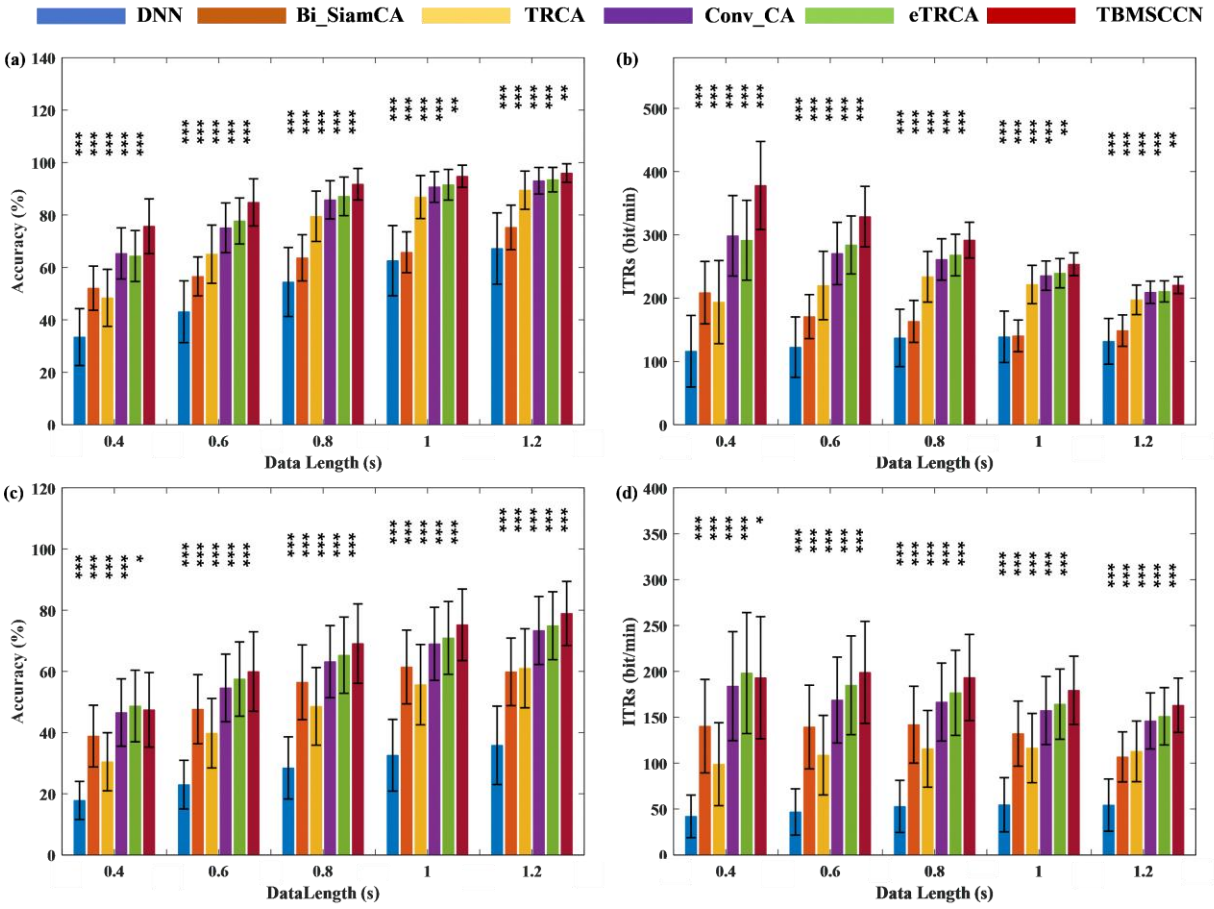


Fig.2 The averaged accuracies and ITRs obtained by the TBMSCCN, eTRCA, Conv_CA, TRCA, Bi_SiamCA, and DNN methods on the "Benchmark" dataset for sub-fig (a), (b) and the "Beta" dataset for sub-fig (c), (d). The data lengths range from 0.4s to 1.2s with a step size of 0.2s. The error bars represent standard deviations. The asterisks indicate significant differences between our proposed method and each of the five other SoTA method obtained via paired t-tests (*p<0.05, **p<0.01, ***p<0.001).

Detailed descriptions of these methods are provided in related references [21,47-50].

III. Experiments Results

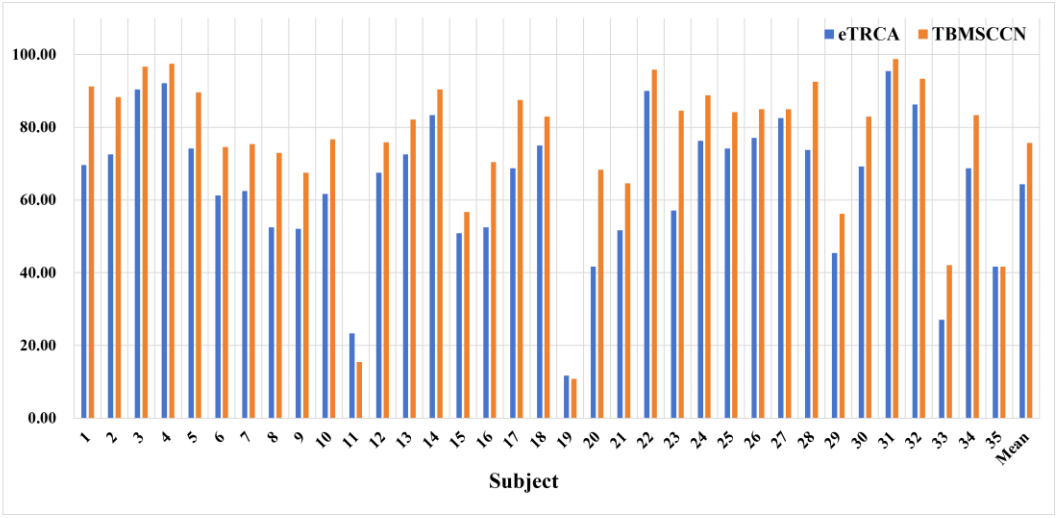
A. Classification Performance in different scenario

(1) Individual calibration scenario: The recognition performance of our proposed TBMSCCN method and the five other SoTA methods on the two public datasets across different time windows are shown in **Fig.2** below.

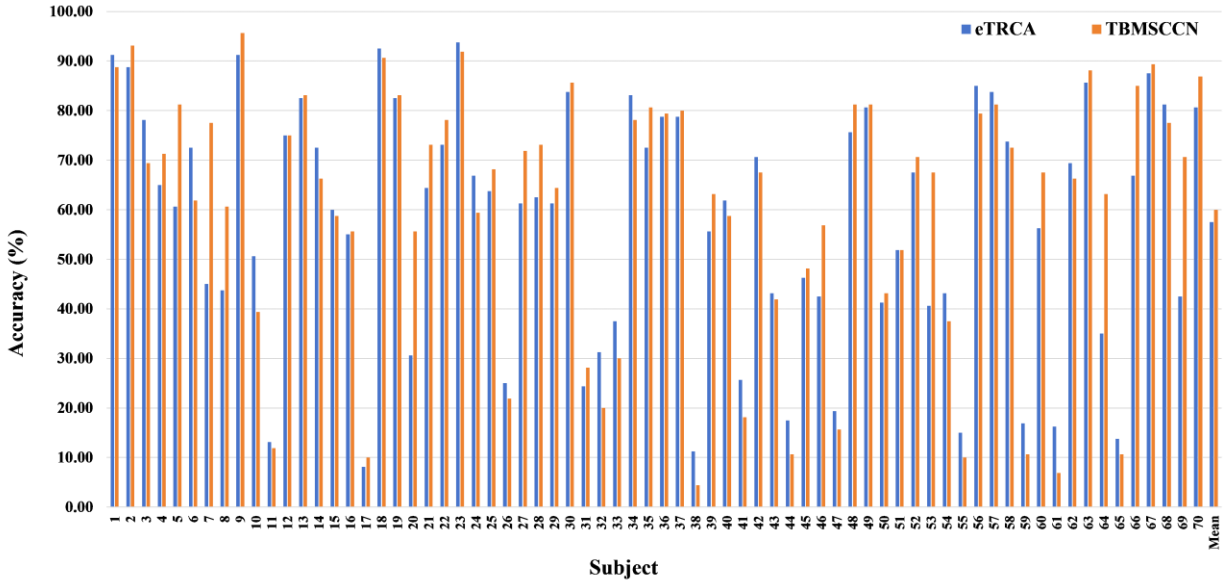
Fig.2 shows that the average recognition accuracy and average ITRs of our proposed method are better than those of the other five SoTA methods on the two public datasets across the different time window lengths. When the time window length is 0.4s, 0.6s, 0.8s, 1.0s, and 1.2s, the average recognition accuracies achieved are 75.70 ± 20.87 %, 84.70 ± 18.00 %, 91.74 ± 12.00 %, 94.78 ± 8.47 %, and 96.00 ± 7.05 % for the "Benchmark" dataset, and 47.43 ± 24.39 %, 59.95 ± 25.98 %, 69.09 ± 25.96 %, 75.22 ± 23.39 %, and 78.94 ± 20.96 % for the "Beta" dataset, respectively.

When the window length is 0.4s, 0.6s, 0.8s, 1.0s and 1.2s, the average ITRs achieved are 378.03 ± 139.18 bit/min, 328.88 ± 95.75 bit/min, 291.73 ± 56.48 bit/min, 253.63 ± 35.99 bit/min, and 220.60 ± 26.42 bit/min for the "Benchmark" dataset, and 193.07 ± 133.08 bit/min, 198.92 ± 111.27 bit/min, 193.29 ± 93.89 bit/min, 179.36 ± 74.39 bit/min, and 163.07 ± 59.18 bit/min for the "Beta" dataset, respectively. When the window length is 0.4s, the highest average ITR can reach 378.03 ± 139.18 bit/min for the "Benchmark" dataset.

Our approach improved the average recognition accuracy and ITR by 11.35%, 27.32%, 42.27%, 10.37%, and 23.6% and



(a) "Benchmark" datasets



(b) "Beta" datasets

Fig.3. The averaged accuracies obtained by TBMSCCN, eTRCA method for each subject in individual calibration scenario. (a) The 35 subjects from "Benchmark" datasets, the data length is 0.4s. (b) The 70 subject from "Beta" datasets, the data length is 0.6s.

86.50 bit/min, 184.31 bit/min, 261.87 bit/min, 79.50 bit/min, and 169.35 bit/min compared to eTRCA TRCA, DNN, Conv-CA, and Bi-SiamCA for the "Benchmark" dataset with a window length of 0.4s. When the window length is 0.6s the highest average ITRs can reach 198.92 ± 111.27 bit/min for the "Beta" datasets. Our approach improved the average recognition accuracy and ITRs by 2.46%, 20.17%, 36.99%, 5.37%, and 12.33%, and 14.08 bit/min, 90.25 bit/min, 152.15 bit/min, 30.18 bit/min, and 59.18 bit/min compared to the eTRCA, TRCA, DNN, Conv-CA and Bi-SiamCA methods for the "Beta" datasets with a window length of 0.6s.

To further evaluate the validity of our proposed method, the two-sided paired t-test was used to assess the significance of the difference between the two methods. **Fig.2** shows the results of these t-tests (in terms of *p*-values) when comparing our TBMSCCN method to the other SoTA approaches on different window lengths for the two datasets. The results are consistently statistically significant ($p < 0.05$).

Among the five SoTA methods, the eTRCA method performed the best. Therefore, we further present the recognition results of our proposed method and the eTRCA method on a single subject. The recognition accuracy over the 35 participants with a window length of 0.4s on the "Benchmark" dataset is shown in **Fig.3 (a).** The recognition accuracy over the 70 participants with a window length of 0.6s on the "Beta" dataset is shown in **Fig.3 (b).** Compared with the eTRCA method, the recognition accuracy for each participant is been significantly improved when using our proposed method.

(2) Cross subject calibration-free scenario: The proposed method TBMSCCN was systematically compared with five calibration-free baseline algorithms (FBCCA, CSSFT, TST-CSFR, EEGNet and EEG Conformer) under different time window length on the "Benchmark" and "Beta" dataset.

Tables II and **III** present the average recognition accuracy and average ITRs of the evaluated methods across varying time window lengths on the two public datasets, respectively. **Table II** and **III** show that the average recognition accuracy and average ITRs of the proposed TBMSCCN were better than those of the other five SoTA methods in two public datasets with different time window lengths. When data length is 0.2s, 0.4s, 0.6s, 0.8s, and 1.0s, the average recognition accuracy can achieve 22.52±14.32%, 37.15±20.90%, 49.21±23.16%, 63.54±20.18 %, and 73.68±24.12 %, for "Benchmark" dataset, and 20.06 ± 10.61 %, 32.77±17.16%, 32.77 ± 17.16 %, 41.31±21.11 %, 54.26±20.95% and 63.04±25.67 % for "Beta" datasets, respectively. When data length is 0.2s, 0.4s, 0.6s, 0.8s, and 1.0s, the average ITRs can achieve 80.69±35.44 bit/min, 116.47±44.62 bit/min, 132.78±38.91 bit/min, 156.28±24.10 bit/min, and 163.12 ± 27.07 bit/min for "Benchmark" dataset, and 67.94±19.87 bit/min, 96.9±31.94 bit/min, 102.07±33.59 bit/min, 123.2 ±26.02bit/min and 128.39±30.35 bit/min for "Beta" datasets, respectively.

When the data length is 1.0 s, the highest average ITRs can reach 163.12±27.07 bit/min for "Benchmark" datasets. Our approach improved the average recognition accuracy and ITRs by 8.76%, 8.18%, 5.79%, 3.79% and 0.7%, and 29.80 bit/min, 27.91 bit/min, 20.00 bit/min, 13.23 bit/min, and 2.49 bit/min compared FBCCA, CSSFT, EEG Conformer, EEGNet and TST-CSFR for "Benchmark" datasets with 1.0s time length. When the data length is 1.0 s, the highest average ITRs can reach 128.39±30.35 bit/min for "Beta" datasets. Our approach improved the average recognition accuracy and ITRs by 6.03%, 5.41%, 8.91%, 7.51% and 1.09%, and 18.80 bit/min, 16.92 bit/min, 27.38 bit/min, 23.24 bit/min, and 3.48 bit/min compared FBCCA, CSSFT, EEG Conformer, EEGNet and TST-CSFR for "Beta" datasets with 1.0s time length.

To further illustrate the validity of the proposed TBMSCCN network, the two-sided paired t-test was used to assess the significant difference between the two methods. **Table II and III** shows that the significance p-values of the TBMSCCN and SoTA approach on different data length of different datasets are less than 0.001.

B. TBMSCCN Network Mode Evaluation

(1) Mode converges evaluation: The models are trained and tested in a runtime environment with model Intel(R) Core

Table II The averaged accuracies obtained by TBMSCCN, FBCCA, CSSFT, EEGConformer, EEGNet and TST-CSFR on two public datasets. The data lengths range from 0.2s to 1.0s with a step of 0.2s. The asterisks indicate significant differences between proposed methods and five other SoTA method obtained by side paired t-tests(*p<0.05,**p<0.01,***p<0.001).

Dataset	Tw(s)	Averaged accuracy (%)					
		TBMSCCN	**FBCCA**	**CSSFT**	**EEG Conformer**	**EEGNet**	**TST-CSFR**
"Benchmark" datasets	0.2	**22.52±14.3**	4.30±1.56 ***	4.71±2.01***	19.37±12.18***	20.88±11.40***	8.11±4.76***
	0.4	**37.15±20.90**	14.30±7.64 ***	17.57±10.39***	34.93±17.56***	36.93±19.15***	25.22±14.51***
	0.6	**49.21±23.16**	32.25±16.89 ***	31.56±14.27***	46.37±21.36***	49.30±20.93***	41.07±19.16***
	0.8	**63.54±20.18**	51.48±20.93***	58.35±19.27***	59.56±22.14***	62.39±22.22***	61.95±20.50***
	1.0	**73.68±24.12**	64.92±21.97***	65.50±23.56***	67.89±23.75***	69.89±22.49***	72.98±25.34***
"Beta" datasets	0.2	**20.06±10.61**	4.94±2.28***	6.04±3.70***	16.75±9.46***	18.42±9.67***	7.56±3.69***
	0.4	**32.77±17.16**	15.67±8.94***	19.16±10.57***	30.48±15.37***	30.84±17.04***	22.71±12.22***
	0.6	**41.31±21.11**	30.87±16.24***	33.39±16.47***	38.65±21.48***	40.27±20.41***	38.33±18.10***
	0.8	**54.26±20.95**	45.07±20.38***	48.67±18.52***	49.24±20.87***	51.53±21.79***	52.13±20.28***
	1	**63.04±25.67**	57.01±24.76***	57.63±26.38***	54.13±23.57***	55.53±24.24***	61.95±20.55***

Table III The averaged ITRs obtained by TBMSCCN, FBCCA, CSSFT, EEGConformer, EEGNet and TST-CSFR on two public datasets. The data lengths range from 0.2s to 1.0s with a step of 0.2s. The asterisks indicate significant differences between proposed methods and five other SoTA method obtained by side paired t-tests(*p<0.05,**p<0.01,***p<0.001).

Dataset	Tw(s)	Averaged ITRs (bit/min)					
		TBMSCCN	FBCCA	CSSFT	EEG Conformer	EEGNet	TST-CSFR
"Benchmark" datasets	0.2	**80.69±35.44**	1.40±0.53***	2.03±0.13***	61.97±25.72***	70.75±22.45***	10.43±2.12***
	0.4	**116.47±44.62**	22.26±5.66***	32.73±11.6***	105.48±32.7***	115.37±38.23***	61.64±22.89***
	0.6	**132.78±38.91**	67.62±22.22***	65.28±16.17***	120.91±33.83***	133.17±32.64***	&99.76±27.92***
	0.8	**156.28±24.10**	112.22±25.7***	136.64±22.21***	141.13±28.34***	151.84±28.52***	&150.16±24.78***
	1	**163.12±27.07**	133.32±23.06***	135.21±26.01***	143.12±26.37***	149.89±24.01	&160.63±29.43***
"Beta" datasets	0.2	**67.94±19.87**	2.51±0.03***	4.86±0.68***	49.07±15.43***	58.35±16.21***	9.02±0.67***
	0.4	**96.9±31.94**	26.99±8.44***	38.99±12.27***	86.09±26.02***	87.76±31.53***	52.52±16.61***
	0.6	**102.07±33.59**	63.82±20.95***	72.53±21.5***	91.79±34.63***	98.01±31.66***	90.58±25.56***
	0.8	**123.2±26.02**	91.98±24.79***	103.83±20.91***	105.76±25.85***	113.6±27.86***	115.68±24.58***
	1	**128.39±30.35**	109.59±28.55***	111.47±31.77***	101.01±26.26***	105.15±27.54***	124.91±20.7***

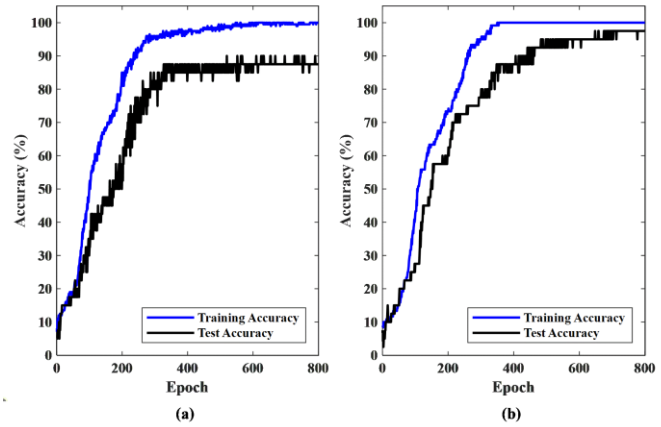


Fig.4 Examples of convergence curves for the training and testing iterations for our proposed TBMSCCN network on the "Benchmark" dataset (a) with a window length of 0.4s participant 1 and for the "Beta" datasets (b) with a window length of 0.6s for participant 2. The training epoch number was set to 800. For the "Benchmark" and "Beta" datasets, the first five or three blocks were used for training and the last block was used for testing.

(TM) i7-12700H CPU and NVIDIA GeForce RTX 3070 Ti Laptop GPU. The training and testing accuracy convergence iteration curves of the proposed TBMSCCN Network are shown below in **Fig. 4**.

Fig.4 (a) and **(b)** show that our model converges within 800 training epochs. Up to epoch 400 the model training and testing accuracies rise rapidly, and the model is able to learn the category frequency features of SSVEP quickly. After epoch 400 the model training and testing accuracy rise more slowly and tend to stabilize.

(2) Mode training parameter: **Table IV** compares the number of training parameters and the maximum averaged ITR of our proposed TBMSCCN network as well as five prominent deep learning methods and two traditional methods in different scenario. We can observe that our approach requires only 7593 training parameters to achieve the highest average ITRs on the "Benchmark" and "Beta" datasets for two

Table IV Comparisons of training parameter number (data length=0.6s, channel number=9), the Maximum averaged accuracies, and the Maximum averaged ITRs on two public datasets.

Method	Trainable parameter	Maximum Averaged Accuracy (%)		Maximum Averaged ITRs (bit/min)	
		"Benchmark" dataset	"Beta" dataset	"Benchmark" dataset	"Beta" dataset
TRCA	360	89.46(1.2s)	61.02(1.2s)	223.83(0.8s)	103.47(0.8s)
eTRCA	360	93.48(1.2s)	74.93(1.2s)	277.55(0.4s)	182.26(0.4s)
DNN	574286	67.21(1.2s)	35.82(1.2s)	138.36(1.0s)	54.61(0.8s)
Conv-CA	6689	93.06(1.2s)	73.37(1.2s)	298.51(0.4s)	183.91(0.4s)
Bi-SiamCA	13692	75.26(1.2s)	59.85(1.2s)	208.67(0.4s)	141.93(0.8s)
TBMSCCN-I	**7593**	**96.76(1.2s)**	**78.94(1.2s)**	**378.03(0.4s)**	**198.92(0.6s)**
TBMSCCN-C	**7593**	**73.68(1.0s)**	**63.04(1.0s)**	**163.12(1.0s)**	**128.39(1.0s)**
EEGNet	1599240	69.89(1.0s)	55.53(1.0s)	151.84(0.8s)	113.60(0.8s)
EEG-Conformer	1814644	67.89(1.0s)	54.13(1.0s)	378.03(0.4s)	105.76(0.8s)

different scenarios. For individual calibration scenario, compared to DNN and Bi-SiamCA, our model (TBMSCCN in individual calibration scenario, TBMSCCN-I) is able to substantially reduce the number of network training parameters while improving the decoding performance of SSVEPs. Compared to the Conv-CA method, our network TBMSCCN-I requires 904 more parameters, while the average maximum ITR is improved by 79.50 bit/min and 15.01 bit/min on the "Benchmark" dataset and "Beta" dataset, respectively. Compared to these two traditional methods, our model does not have a parameter advantage. However, compared to TRCA and eTRCA, our model achieved a maximum ITRs increase of 154.2 bits/min, 100.48 bits/min for "Benchmark" dataset. For "Beta" datasets, the maximum ITRs increase 95.45 bits/min, 16.66 bits/min, respectively. For calibration-free scenario, compared to EEGNet and EEGConformer, our model (TBMSCCN in cross subject transfer scenario, TBMSCCN-C) is able to substantially reduce the number of network training parameters while improving the decoding performance of SSVEPs on two public datasets.

(3) Ablation experiment: In this section, we conducted ablation experiments to further analyze the role of the main modules in the TBMSCCN network. Window lengths of 0.4s and 0.6s were selected for data from the "Benchmark" and "Beta" datasets in individual calibration scenario, respectively. Window lengths of 1.0s and 1.0s were selected for data from the "Benchmark" and "Beta" datasets in cross subject transfer scenarios, respectively. The ablation experiment module is described below. (a) ES-MSTC: The Multi-Scale Time Conv2D layer was replaced by a time Conv2D module in the ES-MSCFEM Module. (b) ET-MSTC: The Multi-Scale Time Conv2D layer was replaced by a time Conv2D module in the ET-MSCFEM Module. (c) MSTC: The Multi-Scale Time Conv2D layers in the two branch modules were both replaced. (d) CGCCM: The CGCCM Module was removed from the TBMSCCN network. (e) CEF: The contrastive loss function was removed from the TBMSCCN network. The results of the ablation experiments are shown in **Table V** and **Table VI** below for individual calibration scenario and cross subject transfer scenarios.

Table V Ablation studies on the two public datasets for individual calibration scenario.

Case	Module ablation					"Benchmark" datasets Tw = 0.4s		"Beta" datasets Tw = 0.6s	
	ES-MSTC	ET-MSTC	MSTC	CGCCM	CEF	Accuracy (%)	ITRs(bit/min)	Accuracy (%)	ITRs(bit/min)
(a)	--	√	√	√	√	73.58±17.83	343.61±133.62	58.25±27.34	173.11±123.56
(b)	√	--	√	√	√	73.21±19.32	340.84±138.83	58.56±29.38	174.56±98.65
(c)	√	√	--	√	√	72.46±21.48	335.25±146.80	56.28±28.64	163.99±134.28
(d)	√	√	√	--	√	67.24±14.29	297.55±122.22	55.21±29.16	159.12±120.54
(e)	√	√	√	√	--	73.68±16.53	344.36±129.26	57.36±21.87	168.97±105.36
(f)	√	√	√	√	√	**75.70±20.87**	**378.03±139.18**	**59.95 ± 25.98**	**198.92±111.27**

Table VI Ablation studies on the two public datasets for cross subject transfer scenario. The time window length is 1.0s for "Benchmark" and "Beta" datasets

Case	Module ablation					"Benchmark" datasets Tw = 1.0s		Beta datasets Tw = 1.0s	
	ES-MSTC	ET-MSTC	MSTC	CGCCM	CEF	Accuracy (%)	ITRs(bit/min)	Accuracy (%)	ITRs(bit/min)
(a)	−	√	√	√	√	71.32±22.38	154.82±23.81	62.41±25.43	126.37±29.87
(b)	√	−	√	√	√	70.51±24.26	152.02±27.34	61.86±26.48	124.62±31.98
(c)	√	√	−	√	√	70.24±25.17	151.09±29.10	60.74±24.54	121.08±28.12
(d)	√	√	√	−	√	68.14±23.84	143.96±26.54	58.34±26.17	113.64±31.35
(e)	√	√	√	√	−	73.12±23.76	161.13±26.38	62.94±24.51	128.07±28.07
(f)	√	√	√	√	√	**73.68±24.12**	**163.12±27.07**	**63.04±25.67**	**128.39±30.35**

Table V and **Table VI** show that the classification accuracy and the averaged ITRs achieved with the TBMSCCN model when all modules were used were significantly higher than in all other cases for both of the public datasets. Therefore, all modules we involved in the TBMSCCN model are verified to be effective in improving the frequency identification performance.

IV. DISCUSSION

A. Comparison of the latest methods

In this section, to further demonstrate the effectiveness of the proposed method, we conducted comparisons experiments with latest approaches in different scenarios. In individual calibration scenarios, the task-discriminant component analysis (TDCA) [55], ePRCA [34] and DDGCNN [59] methods were chosen for experiment comparison. In cross subject transfer scenarios, the SSVEPPoolformer [54] and DG-Conformer [60] approaches were chosen for experiment comparison. The experimental results are recorded as shown in **Table VII** below.

Table VII show proposed TBMSCCN-I method outperforms the TDCA, ePRCA and DDGCNN method with

any time window lengths in individual calibration scenarios. Similarly, proposed TBMSCCN-C method outperforms the SSVEPpoolformer and DG-Conformer method with any time window lengths in cross subject transfer scenarios.

B. Algorithm evaluation for elderly subject

This SSVEP-BCI system is primarily used in the elderly and disabled field. Therefore, it is crucial to evaluate SSVEP-BCI decoding algorithms in elderly populations. In this section, we further evaluate the feasibility of the proposed TBMSCCN algorithm on a large-scale elderly dataset with two different scenarios. The "eldBETA datasets" [57] is briefly described below. This study recruited elderly volunteers with ages greater than 50 years old. One hundred participants (33 males and 67 females) took part in this study. The age of the participants ranged from 52 to 81 with an average of 63.17 ± 6.05 (mean ± standard deviation). A 9-target brain speller of SSVEP-BCI was developed for the elder participants. The 9 targets were aligned in a 3×3 matrix. The stimulation target frequency range is 8 to 12 Hz with 0.5 Hz intervals. For each subject, there are a total of 7 blocks with 63 EEG trials. For each trial, the cue duration is 0.5s, the visual gaze duration is 5s, and the rest time is 0.5s. Details of the dataset are shown in reference [57]. The data preprocessing process is consistent with the description in section B.

Table VII The averaged accuracy and ITRs obtained by **TBMSCCN-I**，**TBMSCCN-C**，**TDCA, ePRCA**，**DDGCNN**，**SSVEPpoolformer and DG-Conformer** on two public datasets for two different scenarios. The data lengths range from 0.6s to 1.0s with a step of 0.2s.

Datasets	Method	Averaged accuracy (%)			Averaged ITRs (bit/min)		
		Tw=0.6s	Tw=0.8s	Tw=1.0s	Tw=0.6s	Tw=0.8s	Tw=1.0s
"Benchmark" datasets	**TBMSCCN-I**	**84.81**	**91.74**	**94.78**	**316.58**	**285.58**	**250.02**
	TDCA [55]	84.25	90.24	94.56	313.07	277.32	248.93
	ePRCA[34]	80.76	89.53	92.46	291.77	273.50	238.82
	DDGCNN [59]	64.80	74.70	83.20	204.78	202.26	198.99
	TBMSCCN-C	**49.21**	**63.54**	**73.68**	**132.78**	**156.28**	**163.12**
	SSVEPpoolformer[54]	40.34	52.62	64.58	96.95	116.15	132.21
	DG-Conformer [60]	48.70	61.35	71.65	130.62	147.87	155.97
"Beta" datasets	**TBMSCCN-I**	**59.95**	**69.09**	**75.22**	**183.61**	**180.39**	**170.15**
	TDCA [55]	59.78	68.64	75.08	182.79	178.53	169.64
	ePRCA[34]	57.84	65.96	71.23	173.54	167.59	155.87
	DDGCNN [59]	42.67	54.36	58.67	155.28	123.56	114.65
	TBMSCCN-C	**41.31**	**54.26**	**63.04**	**102.07**	**123.20**	**128.39**
	SSVEPpoolformer[54]	32.68	43.46	53.78	70.04	86.83	99.99
	DG-Conformer [60]	40.35	50.22	55.74	98.32	109.09	105.78

In individual calibration scenario, the averaged accuracies and ITRs were obtained by TBMSCCN-I, TDCA [55] eTRCA [28], emsTRCA [56], and ePRCA [34] method with 0.6s, 0.8s and 1.0s time window length. In cross subject transfer scenario, the averaged accuracies and ITRs were obtained by TBMSCCN-C, CCA [20], eTRCA [28], CCA + TRCA [58] and CCA + eTRCA [58] method with 0.6s, 0.8s and 1.0s time window length. The experimental results are recorded as shown in **Table VIII** below.

Table VIII The averaged accuracies and ITRs were obtained by **TBMSCCN-I**, **TDCA**, **eTRCA**, **emsTRCA**, and **ePRCA** method on "eldBETA datasets" for individual calibration scenario. The averaged accuracies and ITRs were obtained by **TBMSCCN-C**, **CCA**, **eTRCA**, **CCA+TRCA**, and **CCA + eTRCA** method on "eldBETA datasets" for cross subject transfer scenario. The data length was set to 0.6s, 0.8s and 1.0s.

Scenario	Method	Averaged accuracy (%)			Averaged ITRs (bit/min)		
		Tw=0.6s	Tw=0.8s	Tw=1.0s	Tw=0.6s	Tw=0.8s	Tw=1.0s
Individual calibration scenario	**TBMSCCN-I**	**84.24**	**88.54**	**91.05**	**167.72**	**147.61**	**129.82**
	TDCA [55]	82.45	86.82	90.42	159.99	141.20	127.74
	eTRCA [28]	80.54	84.25	89.54	152.04	132.07	124.88
	emsTRCA [56]	81.46	85.26	89.35	155.84	135.60	124.27
	ePRCA [34]	82.64	87.63	90.68	160.80	144.19	128.59
Cross subject transfer scenario	**TBMSCCN-C**	**52.57**	**56.74**	**62.54**	**60.72**	**56.51**	**57.47**
	CCA [20]	41.73	49.06	54.86	35.81	40.98	43.29
	eTRCA [28]	41.05	46.63	52.27	34.43	36.52	38.92
	CCA+TRCA [58]	46.94	53.78	59.81	47.09	50.26	52.22
	CCA+eTRCA [58]	49.13	56.24	62.27	52.22	55.43	56.94

Table VIII show proposed TBMSCCN-I method outperforms the TDCA, eTRCA, emsTRCA, and ePRCA method with any time window lengths in individual calibration scenarios. When time windows length is 0.6s, the maximum average ITRs can reach 167.72 bits/min. Likewise, proposed TBMSCCN-C method outperforms the CCA, eTRCA, CCA + TRCA, and CCA + eTRCA method with any time window lengths in cross subject tranfer scenarios. When time windows length is 0.6s, the maximum average ITRs can reach 62.54 bits/min.

C. Latent feature spectrum analysis

To further illustrate the feature stability of the model-generated latent vectors, the spectral characteristics of the EEG input signal and the EEG latent vectors are visualized for comparison as shown in **Fig.5** below. The blue solid line represents the spectral curve of the latent feature, the black solid line represents the spectral feature curve of the EEG signal, and the red dashed line indicates the fundamental frequency and harmonic energy of the target frequency.

The black curve indicates that the fundamental and harmonic frequencies of the single EEG signal are obscured by background noise, resulting in unstable target frequency characteristics. When EEG signals are mapped into latent vectors via the ES-MSCFEM feature extraction module, the fundamental frequency and harmonic energy of the target frequency reach peak levels while background noise spectral energy is suppressed. This further validates the stability of the model in generating latent features.

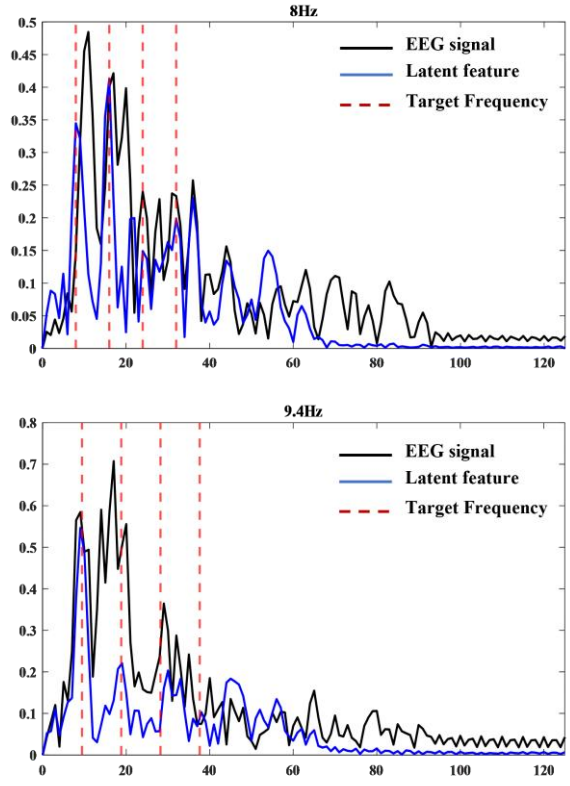


Fig.5 The spectral energy comparison of EEG signals and latent features originates from the 1-st block of the 1-st subject in the "Benchmark" datasets. The time window length is 0.4s.

D. Feature visualization analysis for decision layer

To further illustrate the effectiveness of the CGCCM module, a correlation feature visualization analysis for model decision layer outputs with CGCCM module and without CGCCM module as shown in **Fig.6** below. Decision layer correlation eigenvalues were calculated using 1-st Subject 1-st data block from the "Benchmark" datasets. The blue solid line indicates the correlation coefficient values without CGCCM module, the black solid line indicates the correlation coefficient values with the CGCCM module, and the red dashed line indicates the target stimulation frequency.

Fig.6 shows that the eigenvalues of the model decision layer with the CGCCM module are significantly enhanced at the target frequency, greatly enhancing the model's discrimination ability.

E. STW

In general, STW methods are commonly employed to expand training samples and address the issue of insufficient parameter training in models. **Table IX** compares the recognition performance of the proposed TBMSCCN method and Bi-SiamCA [41] method with STW strategies on two public datasets for individual calibration scenarios. In two public datasets, leave-one-block cross-validation was employed to evaluate model performance. For each subject, the training trial duration was 2s. The test trial time window lengths are set to 0.8s, 1.0s, and 1.2s respectively. We applied the STW method for data augmentation only during the training. The sliding window step size aligns with the time dimension of the test samples.

Table IX show that the proposed TBMSCCN method with the STW strategy outperforms the Bi-SiamCA method with different time windows on two public datasets. **Fig.2** and **Table IX** show that the recognition accuracy of the proposed TBMSCCN method is constrained by the STW strategy. The primary explanation is that proposed TBMSCCN method was a lightweight architecture with fewer parameters. Introducing the STW strategy generated excessive redundant samples, leading to model underfitting. For the "Benchmark" datasets, when the test trial time window length was set to 1.0s, the recognition accuracy of the proposed TBMSCCN network model with the STW strategy decreased by 8.16%. Meanwhile, the user calibration time increased by 240s. When the time window length is 1.0s, the model training time for each epoch under the STW strategy and without the STW strategy is 0.11s and 0.20s, respectively. Training costs were reduced by 45%. Considering system practicality, models with STW strategies increase the calibration time for target subjects.

F. Comparison of training efficiency

In this section, we further compare the training efficiency of different models. **Table X** records the average training time for different models across all single subjects under leave-one block cross-validation. **Table X** show that in the individual calibration scenario, the proposed TBMSCCN method achieves the shortest model training time across all time

windows compared to ConvCA, Bi-SiamCA, and DNN method.

G. Online experiments

In this section, we design a 40-character Chinese spelling system to conduct online algorithm verification. The composition of the Chinese spelling system is illustrated in **Fig.7**. It is mainly composed of a multi-channel EEG acquisition device, an SSVEP frequency modulation interface, a signal processing and recognition module, and a Chinese character output feedback module. The SSVEP frequency modulation interface is consistent with the layout of the "Benchmark" interface, with a stimulus frequency range of 8–15.8 Hz at 0.2 Hz intervals. Ten subjects were recruited from the laboratory, all of whom were thoroughly familiar with the operating procedure of the Chinese speller.

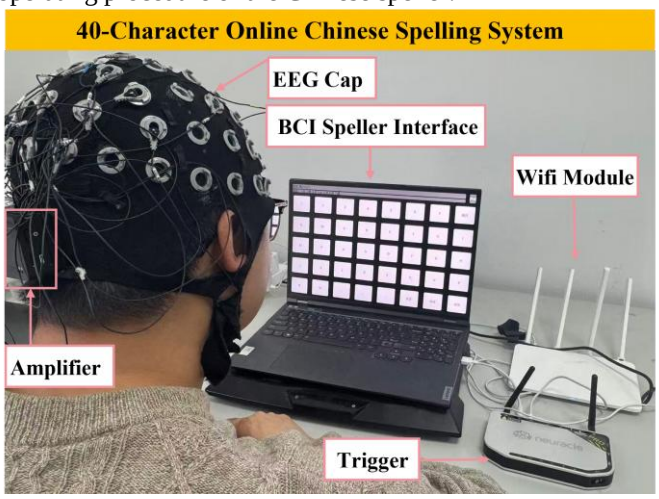


Fig.7. 40 character online Chinese spelling system.

The experimental setup consists of an online training phase and an online testing phase. **Online training phase:** each subject was required to complete 5 blocks of experiments. contains 40 stimulus targets. For each trial, the system prompt duration was set to 0.5 s, the gaze duration was set to 2 s, and the rest interval was set to 0.5 s. Upon completion of online data collection, the relevant data were used to train the TBMSCCN model, which was then deployed to the online testing system.

Online test phase: In the online Chinese spelling task, 10 Chinese words were selected for the online performance test. These words include commonly used oral communication terms, such as {吃饭、喝水、穿衣、熄灯、如厕、开心、难过、疼痛、舒服、谢谢}. Each subject performed free spelling using the associative input method in the order of the task words. The online averaged accuracy, online averaged ITRs, and task completion time for each subject were recorded as shown in **Table XI**.

Table XI demonstrates that the average online character recognition accuracy across all subjects reached 96.77%, and the average ITR can reach 149.04 bit/min in the online Chinese word spelling task. All subjects completed the Chinese character spelling task. The average task completion time is 194.4 s. Therefore, this further confirms that proposed TBMSCCN model can be successfully applied to the SSVEP-based Chinese spelling system.

H. Individual calibration with less sample

Fig.2 and **Table V** show that the recognition performance of the proposed TBMSCCN using the correlation network architecture is significantly better than that of the end-to-end learning DNN network architecture with less calibration trial. On the one hand, an enormous amount of DNN network parameters can lead to insufficient model training. A large number of DNN network parameters can lead to insufficient model training. However, our proposed TBMSCCN network incorporates an average template as a-priori knowledge, which can guide the model to learn the SSVEP frequency features quickly and promote model convergence. The training and testing iteration curves in **Fig.4** effectively verify the convergence of the model.

Work by Nakanishi and colleagues [28] shows that the linear models TRCA and eTRCA are able to enhance SSVEP components by learning optimal channel combinations from training data. However, such methods do not have the capability to perform nonlinear feature extraction for SSVEP signals. Our proposed TBMSCCN network is a non-linear model. It combines CNNs and correlation analysis to provide convolution operations for short EEG signal epochs across multiple channels. Our results suggest that by involving non-linear components, we can obtain a better performing model than through the use of spatial filters. **Fig.3** show the non-linear SSVEP features extracted from most participants are fully exploited and the model identification performance improved compared to the eTRCA method.

The Conv-CA [36] and Bi-SiamCA [41] models are more representative two-branch correlation networks. However, the feature extraction modules and associated parameters of these networks have yet to be explored. **Table V** shows the effectiveness of modules MSTC, CGCCM, and CEF. We introduce the MSTC module in our proposed TBMSCCN network. This module can learn multiscale locally optimal temporal properties while generating potential feature vectors with stable target frequency characteristics. The CEF Module was constructed in the latent feature space, which can guide the model to learn the intra-class consistent features while speeding up model convergence. The CGCCM module is used as a decision layer to reduce the network parameters, while learning distinguishability features between targets and non-targets.

I. Cross subject calibration-free scenario

Table II and **III** show that the recognition performance of the proposed TBMSCCN using the correlation network architecture is significantly better than that of the end to-end learning EEGNet and EEG-Conformer network in calibration-free scenario. On the one hand, the individual variability of EEG signals among different subjects makes it difficult to effectively extract domain-invariant task-relevant components with the stacked training approach of traditional end-to-end models [49-50]. However, proposed TBMSCCN can guide the

model to learn subject-specific discriminative features with the prior knowledge from different subject to enhance the generalization performance of the model. It is evident from the experimental results that the correlation network-based framework demonstrates enhanced generalizability, which substantially improves the performance of SSVEP-based BCI systems in plug-and-play settings.

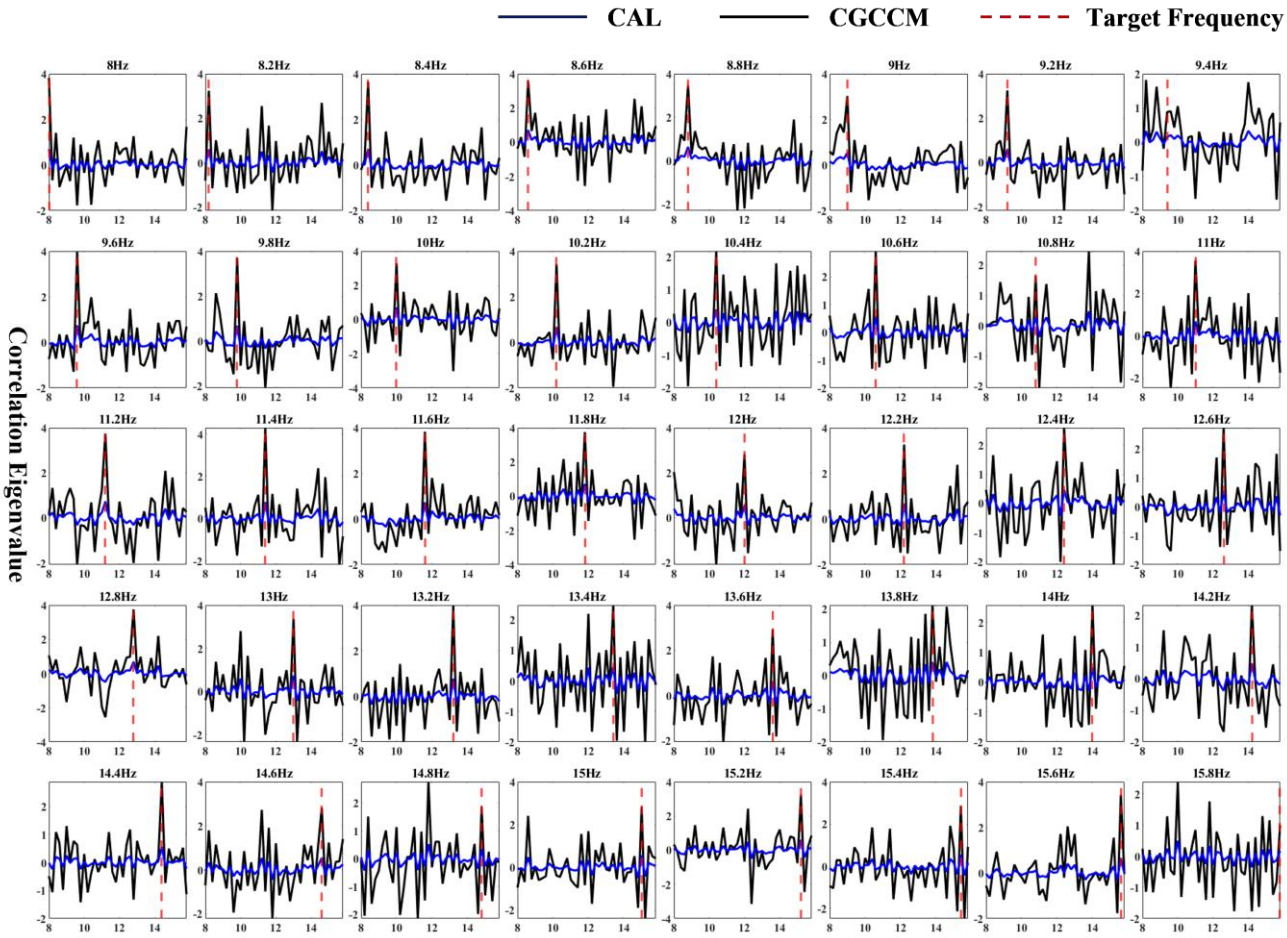


Fig.6 Decision layer correlation eigenvalues visualization with/without CGCCM module. Decision layer correlation eigenvalues were calculated using 1-st Subject 1-st data block from the "Benchmark" datasets in in individual calibration scenarios. The time window length is 0.4s.

Table IX The averaged accuracy and ITRs obtained by **TBMSCCN and Bi-SiamCA** with STW strategy on two public datasets for individual calibration scenarios. The training data time window length is 1.0s, and the test sample data lengths range from 0.8s to 1.2s with a step of 0.2s.

Datasets	Method	Averaged accuracy (%)			Averaged ITRs (bit/min)		
		Tw=0.8s	Tw=1.0s	Tw=1.2s	Tw=0.8s	Tw=1.0s	Tw=1.2s
"Benchmark" datasets	TBMSCCN	81.93	86.62	90.78	235.23	213.00	196.60
	Bi-SiamCA [41]	64.32	69.89	81.67	159.33	149.89	164.14
"Beta" datasets	TBMSCCN	61.09	69.28	73.54	148.46	149.12	139.39
	Bi-SiamCA [41]	53.47	58.43	64.85	120.39	113.91	114.08

J. Limited and Future

Proposed TBMSCCN method still has some limitation. Work by He and Jin et.al [51-52] has shown that the recognition performance of SSVEPs can be further enhanced by utilizing the common information between peripheral stimuli. In the future, it may be necessary to design a stimulus sharing mechanism to enhance the frequency recognition of SSVEPs based on the correlation network framework. Subsequently, inspired by the student-teacher network model, the template enhancement approach and the domain diversity knowledge transfer framework are explored to further enhance the cross-subject generalization performance of the model based on the two-branch correlation network architecture. Finally, to further facilitate the practical engineering application of SSVEP-based BCI systems, we can integrate adaptive dynamic time decision making and asynchronous

discrimination mechanisms into the TBMSCCN framework, thereby enhancing user experience and system flexibility.

Table X The averaged model training time was obtained by TBMSCCN, ConvCA, Bisaim, and DNN method on "Benchmark" dataset for individual calibration scenario with 0.4s, 0.6s, and 0.8s time length.

Method	Averaged model training time (s)		
	T_w=0.4s	T_w=0.6s	T_w=0.8s
TBMSCCN	45.32	57.26	73.52
ConvCA	54.82	67.63	82.28
Bi-SiamCA	66.04	93.47	115.76
DNN	78.56	105.68	126.74

Table XI The average accuracy, average ITRs, and task completion time of a single subject were obtained through a free online spelling experiment with 10 common Chinese vocabulary words.

Subject	Averaged accuracy (%)	Averaged ITRs (bit/min)	Task Completion Time (s)
S01	100.00	159.66	180
S02	95.00	143.14	200
S03	98.91	155.33	184
S04	100.00	159.66	180
S05	98.91	155.33	184
S06	97.87	151.82	188
S07	96.88	148.70	192
S08	90.90	132.03	220
S09	88.14	125.09	236
S10	100.00	159.66	180
Mean	96.77	149.04	194.4

V. Conclusion

In this study, we proposed a TBMSCCN network in which a correlation network framework was introduced to reduce the number of model training parameters and SSVEP prior knowledge was used to enhance the model representation convolution module was designed to learn local temporal dependency features in the parallel two-branch feature extraction module. Next, a contrastive loss function was constructed in the latent feature space, which can guide the model to learn the intra-class consistent features, while speeding up model convergence. Finally, the group convolution module was used as a decision layer to reduce the number of network parameters, while learning distinguishability features between targets and non-targets. Offline tests on two public datasets show that the proposed TBMSCCN outperforms other SoTA methods in two different experiment setting. Online Chinese spelling experiment confirmed the real-world effectiveness of the proposed method.

References

[1] Santhanam G, Ryu S I, Yu B M, et al. A high-performance brain–computer interface[J]. nature, 2006, 442(7099): 195-198.

[2] Chaudhary U, Birbaumer N, Ramos-Murguialday A. Brain–computer interfaces for communication and rehabilitation[J]. Nature Reviews Neurology, 2016, 12(9): 513-525.

[3] Liu Z, Tang J, Gao B, et al. Neural signal analysis with memristor arrays towards high-efficiency brain–machine interfaces[J]. Nature communications, 2020, 11(1): 4234.

[4] Silversmith D B, Abiri R, Hardy N F, et al. Plug-and-play control of a brain–computer interface through neural map stabilization[J]. Nature biotechnology, 2021, 39(3): 326-335.

[5] Pancholi S, Giri A, Jain A, et al. Source aware deep learning framework for hand kinematic reconstruction using EEG signal[J]. IEEE Transactions on Cybernetics, 2022, 53(7): 4094-4106.

[6] Liu Q, Jiao Y, Miao Y, et al. Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset CCA[J]. Neurocomputing, 2020, 378: 36-44.

[7] Kwak N S, Lee S W. Error correction regression framework for enhancing the decoding accuracies of ear-EEG brain–computer interfaces[J]. IEEE transactions on cybernetics, 2019, 50(8): 3654-3667.

[8] Li S, Jin J, Daly I, et al. Feature selection method based on Menger curvature and LDA theory for a P300 brain–computer interface[J]. Journal of Neural Engineering, 2022, 18(6): 066050.

[9] Jin J, Chen Z, Xu R, et al. Developing a novel tactile P300 brain-computer interface with a cheeks-stim paradigm[J]. IEEE Transactions on Biomedical Engineering, 2020, 67(9): 2585-2593.

[10] Jing Jin, Ruitian Xu, Ian Daly, Xueqing Zhao, Xingyu Wang, and Andrzej Cichocki, MOCNN: A Multi-scale Deep Convolutional Neural Network for ERP-based Brain-Computer Interfaces, IEEE Transactions on Cybernetics, 2024, DOI：10.1109/TCYB.2024.3390805.

[11] Cecotti H, Graser A. Convolutional neural networks for P300 detection with application to brain -computer interfaces[J]. IEEE transactions on pattern analysis and machine intelligence, 2010, 33(3): 433-445.

[12] Fumanal-Idocin J, Wang Y K, Lin C T, et al. Motor-imagery-based brain–computer interface using signal derivation and aggregation functions[J]. IEEE Transactions on Cybernetics, 2021, 52(8): 7944-7955.

[13] Cho J H, Jeong J H, Lee S W. NeuroGrasp: Real-time EEG classification of high-level motor imagery tasks using a dual-stage deep learning framework[J]. IEEE Transactions on Cybernetics, 2021, 52(12): 13279-13292.

[14] Xu M, Han J, Wang Y, et al. Implementing over 100 command codes for a high-speed hybrid brain-computer interface using concurrent P300 and SSVEP features[J]. IEEE Transactions on Biomedical Engineering, 2020, 67(11): 3073-3082.

[15] Mai X, Ai J, Ji M, et al. A hybrid BCI combining SSVEP and EOG and its application for continuous wheelchair control[J]. Biomedical Signal Processing and Control, 2024, 88: 105530.

[16] Park S, Ha J, Park J, et al. Brain-controlled, AR-based home automation system using SSVEP-based brain-computer interface and EOG-based eye tracker: A feasibility study for the elderly end User[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 31: 544-553.

[17] Dai W, Liu Y, Lu H, et al. Shared control based on a brain-computer interface for human-multirobot cooperation[J]. IEEE Robotics and Automation Letters, 2021, 6(3): 6123-6130.

[18] Li H, Bi L, Yi J. Sliding-mode nonlinear predictive control of brain-controlled mobile robots[J]. IEEE Transactions on Cybernetics,2020, 52(6): 5419-5431.

[19] Cheng M, Gao X, Gao S, et al. Design and implementation of a brain-computer interface with high transfer rates[J]. IEEE transactions on biomedical engineering, 2002, 49(10): 1181-1186.

[20] Lin Z, Zhang C, Wu W, et al. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs[J]. IEEE transactions on biomedical engineering, 2006, 53(12): 2610-2614.

[21] Chen X, Wang Y, Gao S, et al. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface[J]. Journal of neural engineering, 2015, 12(4): 046008.

[22] Zhang Y, Guo D, Yao D, et al. The extension of multivariate synchronization index method for SSVEP-based BCI[J]. Neurocomputing, 2017, 269: 226-231.

[23] Qin K, Wang R, Zhang Y. Filter bank-driven multivariate synchronization index for training-free SSVEP BCI[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 934-943.

[24] Wang K, Zhai D H, Xiong Y, et al. An MVMD-CCA recognition algorithm in SSVEP-based BCI and its application in robot control[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 33(5): 2159-2167.

[25] Yang C, Han X, Wang Y, et al. A dynamic window recognition algorithm for SSVEP-based brain–computer interfaces using a spatio-temporal equalizer[J]. International journal of neural systems, 2018, 28(10): 1850028.

[26] Wong C M, Wang Z, Nakanishi M, et al. Online adaptation boosts SSVEP-based BCI performance[J]. IEEE Transactions on Biomedical Engineering, 2021, 69(6): 2018-2028.
[27] Jin J, He X, Allison B Z, et al. Leveraging Spatio Temporal Estimation for Online Adaptive Steady State Visual Evoked Potential Recognition[J]. IEEE Transactions on Cognitive and Developmental Systems, 2024. DOI 10.1109/TCDS.2024.3392745.
[28] Nakanishi M, Wang Y, Chen X, et al. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis[J]. IEEE Transactions on Biomedical Engineering, 2017, 65(1): 104-112.
[29] Wang Z, Jin J, Xu R, et al. Efficient spatial filters enhance SSVEP target recognition based on task-related component analysis[J]. IEEE Transactions on Cognitive and Developmental Systems, 2021, 14(3): 1119-1128.
[30] Huang J, Yang P, Xiong B, et al. Incorporating neighboring stimuli data for enhanced SSVEP-based BCIs[J]. IEEE Transactions on Instrumentation and Measurement, 2022, 71: 1-9.
[31] Jin J, Wang Z, Xu R, et al. Robust similarity measurement based on a novel time filter for SSVEPs detection[J]. IEEE Transactions on Neural Networks and Learning Systems, 2021, 34(8): 4096-4105.
[32] Huang J, Yang P, Xiong B, et al. Latency aligning task-related component analysis using wave propagation for enhancing SSVEP-based BCIs[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 30: 851-859.
[33] Lan W, Wang R, He Y, et al. Cross Domain Correlation Maximization for Enhancing the Target Recognition of SSVEP-Based Brain-Computer Interfaces[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023.
[34] Ke Y, Liu S, Ming D. Enhancing SSVEP Identification with Less Individual Calibration Data Using Periodically Repeated Component Analysis[J]. IEEE Transactions on Biomedical Engineering, 2024, 71(4):1319 - 1331.
[35] Guney O B, Oblokulov M, Ozkan H. A deep neural network for ssvep-based brain-computer interfaces[J]. IEEE transactions on biomedical engineering, 2021, 69(2): 932-944.
[36] Deng Y, Sun Q, Wang C, et al. TRCA-Net: using TRCA filters to boost the SSVEP classification with convolutional neural network[J]. Journal of Neural Engineering, 2023, 20(4): 046005.
[37] Yao H, Liu K, Deng X, et al. FB-EEGNet: A fusion neural network across multi-stimulus for SSVEP target detection[J]. Journal of Neuroscience Methods, 2022, 379: 109674.
[38] Chen J, Zhang Y, Pan Y, et al. A transformer-based deep neural network model for SSVEP classification[J]. Neural Networks, 2023, 164: 521-534.
[39] Guney O B, Ozkan H. Transfer learning of an ensemble of DNNs for SSVEP BCI spellers without user-specific training[J]. Journal of Neural Engineering, 2023, 20(1): 016013.
[40] Li Y, Xiang J, Kesavadas T. Convolutional correlation analysis for enhancing the performance of SSVEP-based brain-computer interface[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2020, 28(12): 2681-2690.
[41] Zhang X, Qiu S, Zhang Y, et al. Bidirectional Siamese correlation analysis method for enhancing the detection of SSVEPs[J]. Journal of Neural Engineering, 2022, 19(4): 046027.
[42] Qin K, Xu R, Li S, et al. A Time Local Weighted Transformation Recognition Framework for Steady State Visual Evoked Potentials based Brain Computer Interfaces[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2024.
[43] He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning[C]//Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020: 9729-9738.
[44] Hadsell R, Chopra S, LeCun Y. Dimensionality reduction by learning an invariant mapping[C]//2006 IEEE computer society conference on computer vision and pattern recognition (CVPR'06). IEEE, 2006, 2: 1735-1742.
[45] Wang Y, Chen X, Gao X, et al. A benchmark dataset for SSVEP-based brain–computer interfaces[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2016, 25(10): 1746-1752.
[46] Liu B, Huang X, Wang Y, et al. BETA: A large benchmark database toward SSVEP-BCI application[J]. Frontiers in neuroscience, 2020, 14: 544547.
[47] He X, Allison B Z, Qin K, et al. Leveraging Transfer Superposition Theory for Stable State Visual Evoked Potential Cross-Subject Frequency Recognition[J]. IEEE Transactions on Biomedical Engineering, 2024.
[48] Yan W, Wu Y, Du C, et al. Cross-subject spatial filter transfer method for SSVEP-EEG feature recognition[J]. Journal of neural engineering, 2022, 19(3): 036008.
[49] Lawhern V J, Solon A J, Waytowich N R, et al. EEGNet: a compact convolutional neural network for EEG-based brain–computer interfaces[J]. Journal of neural engineering, 2018, 15(5): 056013.
[50] Song Y, Zheng Q, Liu B, et al. EEG conformer: Convolutional transformer for EEG decoding and visualization[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2022, 31: 710-719.
[51] He X, Allison B Z, Liang W, et al. Leveraging Peripheral Visual Stimuli for Enhanced SSVEP-Based BCIs in Fast Calibration Scenario[J]. IEEE Sensors Journal, 2025.
[52] Jin J, He X, Xu R, et al. Cross Stimulus Transfer Learning Framework Using Common Period Repetition Components for Fast Calibration of SSVEP Based BCIs[J]. IEEE Internet of Things Journal, 2024.
[53] Wang X, Liu A, Cui H, et al. GZSL-Lite: A Lightweight Generalized Zero-Shot Learning Network for SSVEP-Based BCIs[J]. IEEE Transactions on Biomedical Engineering, 2025.
[54] Li C, Liao Z, Cheng Y, et al. SSVEPPoolformer: An Improved Poolformer Model with the Adaptive Denoising Algorithm for SSVEP-EEG Signal Classification[J]. IEEE Transactions on Consumer Electronics, 2025.
[55] Liu B, Chen X, Shi N, et al. Improving the performance of individually calibrated SSVEP-BCI by task-discriminant component analysis[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2021, 29: 1998-2007.
[56] Wong C M, Wan F, Wang B, et al. Learning across multi-stimulus enhances target recognition methods in SSVEP-based BCIs[J]. Journal of neural engineering, 2020, 17(1): 016026.
[57] Liu B, Wang Y, Gao X, et al. eldBETA: a large eldercare-oriented benchmark database of SSVEP-BCI for the aging population[J]. Scientific data, 2022, 9(1): 252.
[58] Shi N, Li X, Liu B, et al. Representative-based cold start for adaptive SSVEP-BCI[J]. IEEE Transactions on Neural Systems and Rehabilitation Engineering, 2023, 31: 1521-1531.
[59] Zhang S, An D, Liu J, et al. Dynamic decomposition graph convolutional neural network for SSVEP-based brain–computer interface[J]. Neural Networks, 2024, 172: 106075.
[60] Liu J, Wang R, Yang Y, et al. Convolutional transformer-based cross subject model for SSVEP-based BCI classification[J]. IEEE Journal of Biomedical and Health Informatics, 2024, 28(11):6581 - 6593.