A Convolutional Block Attention Module and Multi-band Fusion Network for Embedded AR-SSVEP BCI Systems

Hao Zhang
School of Information and Electronics, Beijing Institute of Technology
Beijing, China
ptohka@outlook.com

Ying Sun*
School of Interdisciplinary Science, Beijing Institute of Technology
Beijing, China
suny1417@163.com

Qiaoyi Wang
School of Information and Electronics, Beijing Institute of Technology
Beijing, China
Wqy2316@163.com

Kang Ma
School of Interdisciplinary Science, Beijing Institute of Technology
Beijing, China
kangx.ma@gmail.com

Shuailei Zhang
School of Computer Science and Engineering, Nanyang technological university
Singapore, Singapore
shuailei.zhang@ntu.edu.sg

Feiyang Zhang
School of Instrument Science and Optoelectronic Engineering, Beihang University
Beijing, China
1417zfy@buaa.edu.cn

Chun Hu
School of Interdisciplinary Science, Beijing Institute of Technology
Beijing, China
bit_hc@bit.edu.cn

Dezhi Zheng
School of Interdisciplinary Science, Beijing Institute of Technology
Beijing, China
zhengdezhi@bit.edu.cn

Abstract

Brain-computer interfaces (BCI) based on augmented reality steady-state visual evoked potentials (AR-SSVEP) face critical challenges in mobile environments, including low signal-to-noise ratio (SNR) from dry electrodes and limited computational resources on mobile embedded platforms. To optimize the AR-SSVEP system performance, this study comprehensively considers the stimulus-response coupling mechanism integrating visual optical principles with deep learning-based classification. First, we designed an optimal AR visual stimulation configuration scheme capable of adaptively adjusting key parameters. Second, to address the time-varying non-stationary characteristics of SSVEP and inter-electrode quality variations in dry electrode systems, we propose CBAM-FNet-a lightweight SSVEP detection algorithm that incorporates the Convolutional Block Attention Module (CBAM) with multi-band fusion. The algorithm achieves classification accuracies of 93.84% on benchmark datasets and 74.43% on our self-collected AR-SSVEP dataset, representing performance improvements of up to 22.96% over state-of-the-art methods. Real-time implementation on an embedded unmanned vehicle platform demonstrates 65% control accuracy with an information transfer rate of 50.35 bits/min, validating the practical value of CBAM-FNet in embedded BCI applications and overcoming hardware-imposed performance limitations.

CCS Concepts

• **Human-centered computing → Human computer interaction (HCI)**.

Keywords

Augmented reality; Brain–computer interface; Steady-state visual evoked potentials; Convolutional block attention module

ACM Reference Format:
Hao Zhang, Ying Sun, Qiaoyi Wang, Kang Ma, Shuailei Zhang, Feiyang Zhang, Chun Hu, and Dezhi Zheng. 2025. A Convolutional Block Attention Module and Multi-band Fusion Network for Embedded AR-SSVEP BCI Systems. In *Companion of the 2025 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp C ompanion '25), October 12–16, 2025, Espoo, Finland*. ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3714394.3756273

1 Introduction

Brain-computer interfaces (BCI) have established a novel pathway for communication between the brain and the external environment [4]. BCI based on electroencephalography (EEG) is gaining attention in various fields, including medical rehabilitation [17, 22], education [28], entertainment [19], and smart homes [8, 21]. Steady-state visual evoked potentials (SSVEP) occur when a person's eyes focus on visual stimuli that flicker at a specific frequency, leading to natural oscillations in the brain cortex and generating EEG signals with specific response potentials [2, 30]. SSVEP-BCI is highly effective due to its high signal-to-noise ratios (SNR) and information transfer rate (ITR), resulting in widespread adoption [6, 34].

*Corresponding author.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
UbiComp Companion '25, Espoo, Finland
© 2025 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-1477-1/2025/10
https://doi.org/10.1145/3714394.3756273

However, traditional SSVEP-BCI using LCD displays confines usage to laboratory settings. Recently, researchers have incorporated augmented reality (AR) technology into SSVEP-BCI, forming AR-SSVEP, to enhance the efficiency and comfort of human-machine interaction [1, 10]. However, deploying AR-SSVEP in mobile environments faces significant challenges, including reduced SNR from dry electrodes, computational constraints of embedded platforms, and varying ambient conditions that affect stimulus quality. Previous works have demonstrated the feasibility of embedded SSVEP-BCI systems, such as lightweight electric wheelchairs with hybrid stimulators [16], but achieving robust performance in AR environments with minimal computational resources remains an open challenge.

Research on visual stimulation in AR-SSVEP has been guided by fundamental physics principles governing visual perception and neural oscillations. Researchers have systematically investigated stimulus parameters that include layout [35], color characteristics [5], target density [32], size and contrast optimization [23], and ambient lighting adaptation [31], establishing theoretical foundations for AR-SSVEP stimulus design and demonstrating the importance of systematic optimization for mobile applications.

Efficient detection of SSVEP signals has been a focal point of AR-SSVEP research. Lin et al. introduced Canonical Correlation Analysis (CCA) for SSVEP signal detection [12], followed by Filter Bank CCA (FBCCA) which enhanced performance by incorporating harmonic components [3]. Nakanishi introduced Task-Related Component Analysis (TRCA), achieving remarkable performance by enhancing SNR by removing background EEG activity [18]. Deep learning approaches have further advanced detection capabilities [24], with EEGNet achieving an accuracy of 80% under 12-target conditions [9, 27], and subsequent CNN architectures incorporating various optimization techniques, including LSTM, spectral normalization, and transfer learning strategies [7, 20].Recent works have also explored attention mechanisms to handle temporal non-stationarity and channel-wise variations [13], while metric learning approaches have demonstrated effectiveness in learning robust representations from non-stationary EEG signals [33].

However, the current SSVEP signal detection algorithm has obvious limitations and cannot fully capture the necessary features in the signal, resulting in performance bottlenecks. To solve this problem, this article proposes the convolutional block attention module [29] and multi-band fusion (CBAM-FNet). The algorithm combines the local attention module and can effectively capture key information and features in the signal, thus compensating for the shortcomings of the current algorithm in local perception, especially in a mobile environment.

Under the optimal visual stimulation obtained by test verification, this study proposes a deep learning algorithm that can address the key limitations of BCI signals to enhance the performance of AR-SSVEP to meet the challenges of mobile BCI systems. To verify the effectiveness of the proposed method, we implemented and tested the system on an embedded unmanned vehicle control platform, demonstrating the robustness of the algorithm in a dynamic real environment. The main contributions of this paper are as follows.

- We designed an adaptive AR visual stimulation parameter scheme, including key parameters such as shape, size, and color-brightness interaction, to enhance the visual evoked effects of stimulation targets in AR.
- We proposed a CBAM-FNet method that enhances attention to reliable electrodes and temporal features through the CBAM module, further combined with multi-frequency band fusion technology to improve the ability to capture key SSVEP features.
- We designed and implemented an embedded BCI system based on AR-SSVEP, obtained a self-collected AR-SSVEP dataset to validate the algorithm's effectiveness, and successfully achieved online real-time control of an unmanned vehicle platform.

The remainder of this paper is organized as follows. Section 2 introduces visual stimulus optimization and systematically studies stimulus parameters to improve the quality of SSVEP responses. Section 3 introduces the CBAM-FNet architecture for adaptive signal detection under moving conditions. Section 4 presents the experimental validation, including the description of the dataset, the implementation details, and the performance analysis. Section V outlines the conclusions and proposes future work.

2 Visual Stimulation Optimization

Since AR displays are easily affected by background interference and ambient light changes, the SSVEP signal obtained from the same stimulus display is 30%-50% lower than the SNR of traditional LCDs. In order to optimize the stimulus response and improve the performance of the AR-SSVEP system, we systematically varied three key parameter sets to determine the optimal visual stimulus:

Experiment 1 - Stimulus shape: We tested four geometric shapes of equal area (square, circle, triangle, hexagon) to isolate the effect of shape on the quality of the SSVEP response. Keep the size, color, and lighting conditions of the stimulus consistent.

Experiment 2 - Stimulus size: Using the optimal shape determined in Experiment 1, we evaluated four different sizes of stimuli, ranging from 2.34 ° to 5.85 ° in the field of view, to determine the effect of stimulus size in the AR environment.

Experiment 3 - Interaction between color and brightness: Based on Experiments 1 and 2, we conducted a crossover experiment to test four colors (white, red, green, blue) under three ambient brightness levels (0, 165, 330 luminance) to establish an adaptive color selection strategy for different ambient conditions.

In order to systematically evaluate the effect of stimulus optimization, we collected data from 10 subjects, who subsequently also participated in the self-collected AR-SSVEP dataset collection (described in Section 4.1). The optimization results using various SSVEP detection algorithms are presented in Table 1 and Table 2.

3 CBAM-FNet

The optimized stimulation configuration in the second part provides good visual stimulation. In this chapter, we will introduce CBAM-FNet to address the problems of impedance instability, quality differences between electrodes, and time-varying signal features when mobile AR-SSVEP deployment. This model can adaptively focus on reliable electrodes and discriminative time domain features through a dual attention module and is designed for low-SNR mobile environments.

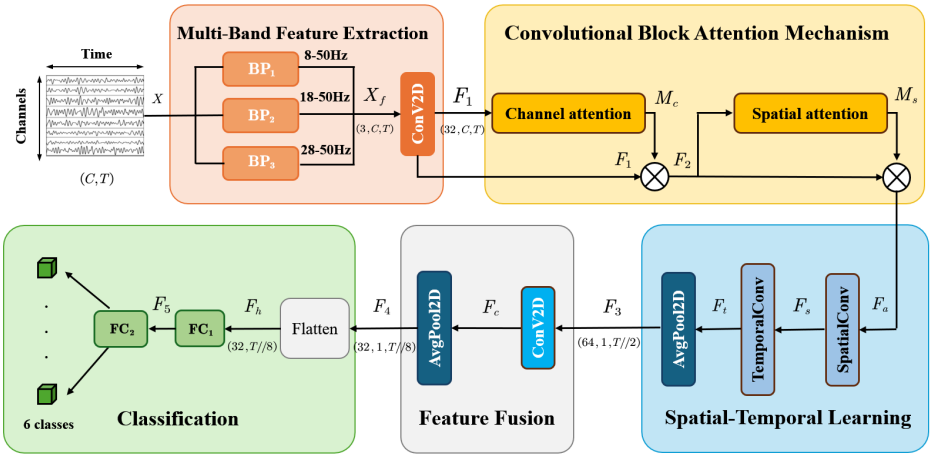


Figure 1: The CBAM-FNet framework

Figure 1 shows the overall framework of CBAM-FNet, which contains five submodules to extract and classify the characteristics of the frequency domain and adapt to the characteristics of the dynamic dry electrode signal.

3.1 Multi-Band Feature Extraction

The input EEG signal undergoes multi-band filtering to extract frequency components across different harmonic ranges. Given the input signal $X \in \mathbb{R}^{C\times T}$, where C represents the number of channels and T denotes time samples, we apply three bandpass filters:

$$X_f = [\mathrm{BP}_1(X), \mathrm{BP}_2(X), \mathrm{BP}_3(X)] \tag{1}$$

where BP_1, BP_2, BP_3 represent bandpass filters with frequency ranges (8-50Hz), (18-50Hz), and (28-50Hz), respectively, which yield $X_f \in \mathbb{R}^{3\times C\times T}$. This decomposition preserves both fundamental frequencies and harmonics, as higher harmonics may have a higher SNR despite lower amplitudes due to lower overall noise levels [11]. Filtered signals are processed through a 2D convolutional layer:

$$F_1 = \mathrm{ELU}(\mathrm{BN}(\mathrm{Conv2D}(X_f, W_1))) \tag{2}$$

where $W_1 \in \mathbb{R}^{32\times 3\times 1\times 64}$ represents the convolutional kernel, BN denotes batch normalization, and ELU is the exponential linear unit activation function with parameter $\alpha = 1.0$:

$$\mathrm{ELU}(x) = \begin{cases} x, & x > 0 \\ \alpha(e^x - 1), & \text{otherwise} \end{cases} \tag{3}$$

3.2 CBAM Module

The CBAM module addresses the inter-electrode quality variations and nonuniform temporal distribution of SSVEP features through sequential channel and spatial attention components.

Channel Attention: Adaptively weights electrode contributions based on signal quality:

$$M_c = \sigma(\mathrm{MLP}(\mathrm{AvgPool}(F_1)) + \mathrm{MLP}(\mathrm{MaxPool}(F_1))) \tag{4}$$

where σ represents the sigmoid function, MLP is a multi-layer perceptron, and the refined features are:

$$F_2 = M_c \otimes F_1 \tag{5}$$

Spatial Attention: Captures discriminative temporal segments:

$$M_s = \sigma(\mathrm{Conv2D}([\mathrm{AvgPool}(F_2); \mathrm{MaxPool}(F_2)])) \tag{6}$$

The final attended features are:

$$F_a = M_s \otimes F_2 \tag{7}$$

3.3 Spatial-Temporal Learning Module

This module employs separable convolutions in depth to extract spatial and temporal features.

$$F_s = \mathrm{BN}(\mathrm{DepthwiseConv2D}(F_a, W_s)) \tag{8}$$

$$F_t = \mathrm{BN}(\mathrm{DepthwiseConv2D}(F_s, W_t)) \tag{9}$$

where $W_s \in \mathbb{R}^{64\times 32\times 8\times 1}$ and $W_t \in \mathbb{R}^{64\times 1\times 1\times 16}$ are the spatial and temporal convolutional kernels, respectively. The features are then processed through activation and pooling:

$$F_3 = \mathrm{Dropout}(\mathrm{AvgPool2D}(\mathrm{ELU}(F_t)), p = 0.2) \tag{10}$$

3.4 Feature Fusion Module

The fusion module integrates information across all frequency bands through a standard convolutional layer.

$$F_c = \mathrm{ELU}(\mathrm{BN}(\mathrm{Conv2D}(F_3, W_f))) \tag{11}$$

$$F_4 = \mathrm{Dropout}(\mathrm{AvgPool2D}(F_c), p = 0.2) \tag{12}$$

where $W_f \in \mathbb{R}^{32\times 64\times 1\times 16}$ facilitates the cross-frequency interaction.

3.5 Classification Module

The final classification is performed through fully connected layers:

$$F_h = \mathrm{Flatten}(F_4) \tag{13}$$

$$F_5 = \mathrm{Dropout}(\mathrm{ELU}(\mathrm{FC}_1(F_h)), p = 0.2) \tag{14}$$

$$y = \mathrm{Softmax}(\mathrm{FC}_2(F_5)) \tag{15}$$

where $\mathrm{FC}_1 : \mathbb{R}^{32\times(T/8)} \to \mathbb{R}^{512}$ and $\mathrm{FC}_2 : \mathbb{R}^{512} \to \mathbb{R}^6$ are fully connected transformations, and the softmax function provides class probabilities:

$$\mathrm{Softmax}(z_i) = \frac{\exp(z_i)}{\sum_{j=1}^{6} \exp(z_j)} \tag{16}$$

3.6 Model Training

The network is trained using cross-entropy loss:

$$L = -\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{6} y_{ij}\log(\hat{y}_{ij}) \tag{17}$$

where n is the batch size, y_{ij} represents the true label and $\hat{y}_{ij}$ denotes the predicted probability for the sample i and class j. The model parameters are optimized using the Adam optimizer with adaptive learning rate scheduling to ensure convergence stability and optimal performance on SSVEP signal classification tasks.

4 Evaluation

4.1 Datasets and Preprocessing

Datasets: To comprehensively validate the proposed CBAM-FNet algorithm, we used two datasets: the Tsinghua benchmark dataset and a self-collected AR-SSVEP dataset.

The Benchmark dataset provides standardized 64-channel EEG data from 35 healthy subjects, originally sampled at 1000 Hz and downsampled to 250 Hz, with frequency-phase joint modulation stimuli ranging from 10.2-15.2 Hz. Each subject completed 6 blocks of 40 trials, each trial lasting 6 s. For consistency with our self-collected AR-SSVEP dataset, we selected 8 occipital channels and 6 stimulus categories from the data of each subject [26].

The self-collected AR-SSVEP dataset contains data on the 8-channel occipital area under the AR display collected from 10 healthy subjects using a dry electrode acquisition device at a sampling rate of 500 Hz. The dataset contains the best stimulus configuration determined in Section 2: a circular medium-size (field of view 4.68°) white stimulus at 0 and 165 luminance, and a red stimulus at 330 luminance. Each condition contains 6 blocks, each block contains 6 trials, the stimulation frequency range is 10-15 Hz, and the interval is 1 Hz. Each trial lasts 7 seconds, including 2 seconds of target presentation, 4 seconds of visual stimulation, and 1 second of stimulus extinction. The experimental process in Section 2 is similar.

Preprocessing: In order to extract effective SSVEP characteristics, the collected EEG signals must be pre-processed. SSVEP signals are mainly concentrated in the range of 8-30 Hz, and the amplitude is relatively weak. Since there is a lot of 50 Hz power line noise related to the human body in EEG, we applied a 5-45 Hz bandpass filter to retain the SSVEP features and remove the noise component. In order to solve the problem of delayed visual response of humans to SSVEP stimulation, we performed signal delay correction to ensure that effective EEG signal fragments are captured. Due to the low SNR caused by the high impedance characteristics of the dry electrode device, we used common average reference (CAR) spatial filtering to enhance the signal. The CAR method eliminates spontaneous brain activity while retaining induced characteristic signals. The CAR formula is as follows:

$$S_i^{CAR} = S_i^{raw} - \frac{1}{n}\sum_{j=1}^{n} S_j^{raw} \tag{18}$$

Where n represents the total number of channels, S_i^{raw} represents the raw signal from channel i. This algorithm can filter out non-spontaneous components and interference in EEG signals, thus making the spatial distribution of EEG features highly concentrated. [14]

In order to meet the neural network training needs and focus on the more important short-term SSVEP detection performance in mobile environments, we applied an overlapping segmentation technique to expand the available data. Both datasets are segmented at 1 second intervals and overlapped at 0.1 second/0.05 second to generate enough training samples while maintaining temporal continuity. This enhancement strategy ensures a fair comparison between all algorithms using the same preprocessing method.

4.2 Baseline and Experiments

Baseline: We compared CBAM-FNet against three established SSVEP detection algorithms: CCA, FBCCA, and EEGNet. CCA serves as a fundamental frequency-domain correlation method, while FBCCA enhances detection by incorporating filter bank decomposition and harmonic information. EEGNet represents the state-of-the-art deep learning approach specifically designed for EEG signal classification tasks.

Experiments: We conducted comprehensive offline evaluations using both the benchmark dataset and the self-collected AR-SSVEP dataset to validate the proposed CBAM-FNet algorithm. The benchmark dataset evaluation evaluates algorithm performance in high-quality standardized EEG data, while the self-collected AR-SSVEP dataset evaluation examines effectiveness in AR environments with dry electrode acquisition systems.

To validate the practical feasibility and real-time control capability of CBAM-FNet in mobile environments, we developed a portable AR-SSVEP unmanned vehicle control system as **Figure 2**. The system comprises four key components as follows: (1) AR Visual Stimulator: The X2 Pro (NED+, CN) with 60Hz refresh rate, featuring a separate head mounted display from the design of the computing unit to reduce head weight and enhance comfort. External computer-based graphics rendering ensures stable visual stimulus signals. The Qt-developed stimulation interface presents a 3×2 array of six stimulus targets with frequencies of 10-15 Hz (1 Hz intervals, ϕ = 0, where brightness values are determined using sinusoidal sampling theory for specific flicker waveforms. [15] (2) Dry Electrode EEG Acquisition Device: 8 dry electrodes placed in the occipital region following an international system of 10-20 (Oz, O1, O2, POz, PO3, PO4, PO5, PO6 sites). The system operates at 500 Hz sampling frequency with an Ag / AgCl ear clip reference electrode with validated stability. (3) Embedded EEG analysis platform: deploy the CBAM-FNet algorithm using the LibTorch neural network inference framework for real-time signal processing and classification. (4) Unmanned Vehicle Microcontroller: Implement vehicle control commands through integration of the x3566 microcontroller with the vehicle remote control system.

The feasibility verification experiment was conducted on two CBAM-FNet subjects who performed well in the initial screening. Subjects performed real-time vehicle control tasks paying visual attention to AR-recognized stimuli to verify the practical applicability of the algorithm in a mobile environment.

For both offline and online experiments, we employed consistent model configuration and training procedures. The datasets

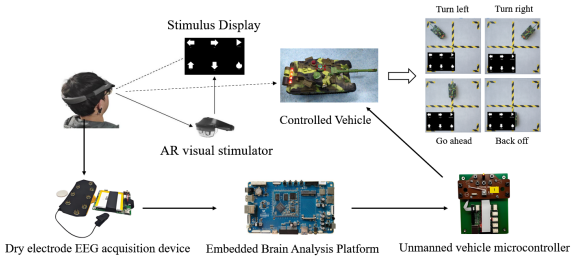


Figure 2: AR-SSVEP unmanned vehicle control system based on embedded platform

were randomly partitioned using six-fold cross-validation. CBAM-FNet was implemented using the PyTorch framework with Adam optimizer (learning rate = 0.0012, batch size = 64) and the cross-entropy loss function. StepLR learning rate adjustment strategy was employed with an early stopping mechanism. All comparative algorithms followed identical training procedures and preprocessing methods to ensure a fair evaluation. Recognition accuracy within 1-second time windows was used as the primary performance metric for offline evaluation, while recognition accuracy and information transfer rate (ITR) were measured for online validation.

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

where N is the number of targets (6), P is the recognition accuracy, and T is the time window in seconds (1s).

4.3 Results

4.3.1 *Stimulus Configuration Optimization.*

The optimal test performance of the experiment in Section 2 is shown in Table 1 and Table 2.

Table 1: Shape and Size Optimization Results for AR-SSVEP Stimuli (FBCCA)

Shape	Accuracy (%)	**Size (FOV)**	Accuracy (%)
Circle	**66.54±18.92**	**4.68°**	**75.53±15.21**
Hexagon	64.81±14.55	5.85°	71.55±20.76
Square	60.58±11.64	3.21°	56.08±10.05
Triangle	60.97±6.49	2.34°	54.09±25.52

Table 2: Color-Brightness Optimization Results for AR-SSVEP Stimuli (FBCCA)

Brightness	White	Red	Green	Blue
0 lum	**72.09±20.50**	46.96±6.21	44.97±17.19	34.13±14.28
165 lum	58.86±12.57	50.27±16.14	37.96±15.21	22.49±15.61
330 lum	56.35±14.29	**58.47±19.58**	37.69±6.74	21.82±6.75

Among algorithms, FBCCA achieved the highest detection accuracy. The optimal stimulus configuration determined by the experiment is circular, moderate in size (field of view 4.68 °), and adaptive in color selection (white at low ambient brightness; red at high ambient brightness). After the optimized configuration, a detection accuracy rate of up to 75.53% was achieved, providing reliable visual stimulus support for the effective verification of the subsequent CBAM-FNet algorithm.

4.3.2 *CBAM-FNet Performance Evaluation.*

Benchmark Dataset Performance: The experimental results of the benchmark dataset demonstrate the superior performance of CBAM-FNet in all evaluation blocks. As shown in **Figure 3(a)**, CBAM-FNet achieved an average detection accuracy of 93.84% within 1s time windows, significantly outperforming the baseline methods: CCA (82.99%), FBCCA (88.96%) and EEGNet (88.46%). With CBAM-FNet showing consistent superiority in the six blocks with accuracy ranging from 93.15% to 94.67%.

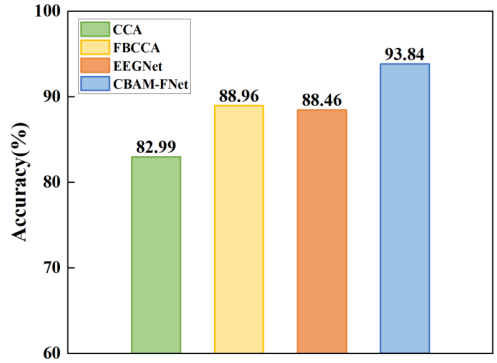
self-collected AR-SSVEP dataset performance: In **Figure 3(b)**, the self-collected dry electrode AR-SSVEP dataset is used for verification, further confirming the robustness of CBAM-FNet under more challenging low SNR conditions. The proposed algorithm achieves an average detection accuracy of 74.43%, significantly outperforming the baseline methods: CCA (45.00%), FBCCA (47.38%) and EEGNet (51.47%), which is 22.96% higher than the best baseline method, demonstrating the effectiveness of the CBAM attention mechanism in dealing with realistic noisy AR-SSVEP environments.

Ablation Study on CBAM Module: To validate the effectiveness of the CBAM attention mechanism, we performed ablation experiments by comparing CBAM-FNet with its base network (without the CBAM module). As shown in Table 3, on the benchmark dataset, the CBAM module improves the average accuracy of 92.35% to 93.84%, representing an absolute improvement of 1.49%. More importantly, on the challenging self-collected AR-SSVEP dataset with lower SNR, the CBAM module demonstrates its critical role by improving the average accuracy from 56.11% to 74.43%, achieving an impressive 18.32% absolute improvement. This significant performance gain under noisy conditions confirms that the CBAM attention mechanism effectively enhances the model's ability to focus on informative features while suppressing noise interference.

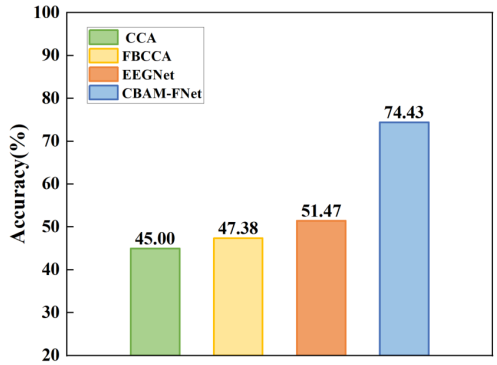
Table 3: Ablation study comparing CBAM-FNet with base network (without CBAM)

Method	Benchmark Acc. (%)	AR-SSVEP Acc. (%)
Base Network	92.35	56.11
CBAM-FNet	**93.84**	**74.43**

Online validation results: The CBAM-FNet algorithm achieved real-time performance on the x3566 embedded platform with inference latency below 100ms per classification, meeting the requirements for online control applications.The feasibility verification results preliminarily in **Table 4** prove that CBAM-FNet has the potential for real-time control through the AR-SSVEP interface. The recognition accuracy of both subjects exceeded 60%, indicating that the system has the ability to generate basic control commands. The verification experiment focused on the implementation of basic vehicle control operations (forward, backward, left turn, right turn, start, ignition).



(a) Accuracy on benchmark dataset



(b) Accuracy on self-collected AR-SSVEP dataset

Figure 3: The performance evaluation results in benchmark and self-collected AR-SSVEP dataset.

Table 4: Online validation results for embedded unmanned vehicle control

Subject	Accuracy (%)		ITR (bit/min)	
	CCA	CBAM-FNet	CCA	CBAM-FNet
1	43.33	**66.67**	16.92	**53.57**
2	36.67	**63.33**	9.98	**47.12**
Mean	40.00	**65.00**	13.45	**50.35**

The current sample is too small (2 subjects), and a comprehensive evaluation of a larger number of subjects is required to achieve stable actual deployment. However, the accuracy level preliminarily verifies the technical feasibility of deploying the algorithm in this paper on the resource-constrained AR-SSVEP embedded mobile control platform, laying the foundation for future experimental verification.

5 Conclusion

This study systematically optimized AR-SSVEP visual stimuli by evaluating the effects of shape, size, color, and brightness on SSVEP response quality. Through comprehensive experiments, we identified optimal stimulus configurations that significantly enhance signal quality in mobile AR environments.

To address the critical problem of low SNR in mobile AR-SSVEP systems, we developed an enhanced detection method incorporating CBAM and multi-band fusion modules. These innovations specifically target SSVEP signal non-stationarity and electrode detachment issues common in dry electrode acquisition. Our experimental validation demonstrates significant performance improvements. In the SSVEP benchmark dataset, the algorithm achieved an accuracy of 93.84%, confirming its fundamental effectiveness. More critically, on our self-collected AR-SSVEP dataset that simulates realistic mobile conditions, we achieved 74.43% accuracy, a substantial 22.96% improvement over the EEGNet algorithm. This performance gain validates the algorithm's robustness in challenging low-SNR environments. We further validated the practical feasibility through an embedded unmanned vehicle platform, achieving 65% average control accuracy and 50.35 bits/minute ITR. Although based on limited samples, these results demonstrate promising real-world applicability under hardware constraints.

Future research directions include: (i) Advanced visual stimulation optimization through high-refresh hardware, personalized subject configuration, and improved display algorithms; (ii) Enhanced algorithmic processing through distillation learning from high-quality SSVEP data, integration of calibration-free frameworks [25], cross-subject transfer learning, and multimodal integration of physiological signals such as laryngeal vibration and cerebral blood oxygen; (iii) Expanded experimental validation with larger subject pools, enriched datasets, and diverse real-world mobile scenarios to facilitate the laboratory-to-application transition.

Acknowledgments

This work was supported by the National Key Research and Development Program of China (Grant No. 2022YFC3300705).

References

[1] Pasquale Arpaia, Egidio De Benedetto, Lucio De Paolis, Giovanni D'Errico, Nicola Donato, and Luigi Duraccio. 2022. Performance enhancement of wearable instrumentation for AR-based SSVEP BCI. *Measurement* 196 (2022), 111188.

[2] Xiaogang Chen, Zhikai Chen, Shangkai Gao, and Xiaorong Gao. 2014. A high-itr ssvep-based bci speller. *Brain-Computer Interfaces* 1, 3-4 (2014), 181–191.

[3] Xiaogang Chen, Yijun Wang, Shangkai Gao, Tzyy-Ping Jung, and Xiaorong Gao. 2015. Filter bank canonical correlation analysis for implementing a high-speed SSVEP-based brain–computer interface. *Journal of neural engineering* 12, 4 (2015), 046008.

[4] Xiaogang Chen, Yijun Wang, Masaki Nakanishi, Xiaorong Gao, Tzyy-Ping Jung, and Shangkai Gao. 2015. High-speed spelling with a noninvasive brain–computer interface. *Proceedings of the national academy of sciences* 112, 44 (2015), E6058–E6067.

[5] Yulin Du and Xincan Zhao. 2022. Visual stimulus color effect on SSVEP-BCI in augmented reality. *Biomedical Signal Processing and Control* 78 (2022), 103906.

[6] Shangkai Gao, Yijun Wang, Xiaorong Gao, and Bo Hong. 2014. Visual and auditory brain–computer interfaces. *IEEE transactions on biomedical engineering* 61, 5 (2014), 1436–1447.

[7] Osman Berke Guney and Huseyin Ozkan. 2023. Transfer learning of an ensemble of DNNs for SSVEP BCI spellers without user-specific training. *Journal of Neural Engineering* 20, 1 (2023), 016013.

[8] Syed Rehan Abbas Jafri, Tehreem Hamid, Rabia Mahmood, Muhammad Asjad Alam, Talha Rafi, Muhammad Zeeshan Ul Haque, and Muhammad Wasim Munir. 2019. Wireless brain computer interface for smart home and medical system. *Wireless Personal Communications* 106 (2019), 2163–2177.

[9] Vernon J Lawhern, Amelia J Solon, Nicholas R Waytowich, Stephen M Gordon, Chou P Hung, and Brent J Lance. 2018. EEGNet: a compact convolutional neural

network for EEG-based brain–computer interfaces. *Journal of neural engineering* 15, 5 (2018), 056013.

[10] Dapeng Li, Aobei Chen, Rui Na, Jiaqi Jing, Chun Hu, Dezhi Zheng, and Shuai Wang. 2023. Microstructure design and optimization of high-sensitivity interdigital capacitive humidity sensor based on uncertainty analysis. *Measurement* 222 (2023), 113599.

[11] Fang-Cheng Lin, John K Zao, Kuan-Chung Tu, Yijun Wang, Yi-Pai Huang, Che-Wei Chuang, Hen-Yuan Kuo, Yu-Yi Chien, Ching-Chi Chou, and Tzyy-Ping Jung. 2012. SNR analysis of high-frequency steady-state visual evoked potentials from the foveal and extrafoveal regions of human retina. In *2012 Annual international conference of the IEEE engineering in medicine and biology society*. IEEE, 1810–1814.

[12] Zhonglin Lin, Changshui Zhang, Wei Wu, and Xiaorong Gao. 2006. Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs. *IEEE transactions on biomedical engineering* 53, 12 (2006), 2610–2614.

[13] Xiaolin Liu, Rongye Shi, Qianxin Hui, Susu Xu, Shuai Wang, Rui Na, Ying Sun, Wenbo Ding, Dezhi Zheng, and Xinlei Chen. 2022. TCACNet: Temporal and channel attention convolutional network for motor imagery classification of EEG-based BCI. *Information Processing & Management* 59, 5 (2022), 103001.

[14] Kip A Ludwig, Rachel M Miriani, Nicholas B Langhals, Michael D Joseph, David J Anderson, and Daryl R Kipke. 2009. Using a common average reference to improve cortical neuron recordings from microelectrode arrays. *Journal of neurophysiology* 101, 3 (2009), 1679–1689.

[15] Nikolay V Manyakov, Nikolay Chumerin, Arne Robben, Adrien Combaz, Marijn Van Vliet, and Marc M Van Hulle. 2013. Sampled sinusoidal stimulation profile and multichannel fuzzy logic classification for monitor-based phase-coded SSVEP brain–computer interfacing. *Journal of neural engineering* 10, 3 (2013), 036011.

[16] Rui Na, Chun Hu, Ying Sun, Shuai Wang, Shuailei Zhang, Mingzhe Han, Wenhan Yin, Jun Zhang, Xinlei Chen, and Dezhi Zheng. 2021. An embedded lightweight SSVEP-BCI electric wheelchair with hybrid stimulator. *Digital Signal Processing* 116 (2021), 103101.

[17] Rui Na, Dezhi Zheng, Ying Sun, Mingzhe Han, Shuai Wang, Shuailei Zhang, Qianxin Hui, Xinlei Chen, Jun Zhang, and Chun Hu. 2021. A wearable low-power collaborative sensing system for high-quality SSVEP-BCI signal acquisition. *IEEE Internet of Things Journal* 9, 10 (2021), 7273–7285.

[18] Masaki Nakanishi, Yijun Wang, Xiaogang Chen, Yu-Te Wang, Xiaorong Gao, and Tzyy-Ping Jung. 2017. Enhancing detection of SSVEPs for a high-speed brain speller using task-related component analysis. *IEEE Transactions on Biomedical Engineering* 65, 1 (2017), 104–112.

[19] Tapsya Nayak, Li-Wei Ko, Tzyy-Ping Jung, and Yufei Huang. 2019. Target classification in a novel SSVEP-RSVP based BCI gaming system. In *2019 IEEE International Conference on Systems, Man and Cybernetics (SMC)*. IEEE, 4194–4198.

[20] Yudong Pan, Jianbo Chen, Yangsong Zhang, and Yu Zhang. 2022. An efficient CNN-LSTM network with spectral normalization and label smoothing technologies for SSVEP frequency recognition. *Journal of neural engineering* 19, 5 (2022), 056014.

[21] Seonghun Park, Ho-Seung Cha, and Chang-Hwan Im. 2019. Development of an online home appliance control system using augmented reality and an SSVEP-based brain–computer interface. *IEEE Access* 7 (2019), 163604–163614.

[22] Chamika Janith Perera, Isira Naotunna, Chameera Sadaruwan, Ranathunga Arachchilage Ruwan Chandra Gopura, and Thilina Dulantha Lalitharatne. 2016. SSVEP based BMI for a meal assistance robot. In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 002295–002300.

[23] Hakim Si-Mohammed, Christian Holz, Andrew Wilson, Hannes Gamper, Adrian Kc Lee, Dimitra Emmanouilidou, Edward Cutrell, and Ivan Tashev. 2023. On the effect of size and contrast of the ssvep visual stimuations on classification accuracy and user-friendliness in virtual reality. In *2023 11th International Winter Conference on Brain-Computer Interface (BCI)*. IEEE, 1–6.

[24] Ying Sun, Wenzheng Ding, Xiaolin Liu, Dezhi Zheng, Xinlei Chen, Qianxin Hui, Rui Na, Shuai Wang, and Shangchun Fan. 2022. Cross-subject fusion based on time-weighting canonical correlation analysis in SSVEP-BCIs. *Measurement* 199 (2022), 111524.

[25] Chenlong Wang, Jiaao Li, Shuailei Zhang, Wenbo Ding, and Xinlei Chen. 2025. Fast SSVEP Detection Using a Calibration-Free EEG Decoding Framework. *arXiv preprint arXiv:2506.01284* (2025).

[26] Yijun Wang, Xiaogang Chen, Xiaorong Gao, and Shangkai Gao. 2016. A benchmark dataset for SSVEP-based brain–computer interfaces. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 25, 10 (2016), 1746–1752.

[27] Nicholas Waytowich, Vernon J Lawhern, Javier O Garcia, Jennifer Cummings, Josef Faller, Paul Sajda, and Jean M Vettel. 2018. Compact convolutional neural networks for classification of asynchronous steady-state visual evoked potentials. *Journal of neural engineering* 15, 6 (2018), 066031.

[28] Chi Man Wong, Qi Tang, Janir Nuno da Cruz, and Feng Wan. 2015. A multi-channel SSVEP-based BCI for computer games with analogue control. In *2015 IEEE international conference on computational intelligence and virtual environments for measurement systems and applications (CIVEMSA)*. IEEE, 1–6.

[29] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.

[30] Ge Zhang, Yan Cui, Yangsong Zhang, Hefei Cao, Guanyu Zhou, Haifeng Shu, Dezhong Yao, Yang Xia, Ke Chen, and Daqing Guo. 2021. Computational exploration of dynamic mechanisms of steady state visual evoked potentials at the whole brain level. *NeuroImage* 237 (2021), 118166.

[31] Rui Zhang, Lijun Cao, Zongxin Xu, Yangsong Zhang, Lipeng Zhang, Yuxia Hu, Mingming Chen, and Dezhong Yao. 2023. Improving AR-SSVEP recognition accuracy under high ambient brightness through iterative learning. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 31 (2023), 1796–1806.

[32] Rui Zhang, Zongxin Xu, Lipeng Zhang, Lijun Cao, Yuxia Hu, Beihan Lu, Li Shi, Dezhong Yao, and Xincan Zhao. 2022. The effect of stimulus number on the recognition accuracy and information transfer rate of SSVEP–BCI in augmented reality. *Journal of Neural Engineering* 19, 3 (2022), 036010.

[33] Shuailei Zhang, Kai Keng Ang, Dezhi Zheng, Qianxin Hui, Xinlei Chen, Yang Li, Ning Tang, Effie Chew, Rosary Yuting Lim, and Cuntai Guan. 2022. Learning eeg representations with weighted convolutional siamese network: A large multi-session post-stroke rehabilitation study. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 30 (2022), 2824–2833.

[34] Yangsong Zhang, Erwei Yin, Fali Li, Yu Zhang, Daqing Guo, Dezhong Yao, and Peng Xu. 2019. Hierarchical feature fusion framework for frequency recognition in SSVEP-based BCIs. *Neural Networks* 119 (2019), 1–9.

[35] Xincan Zhao, Chenyang Liu, Zongxin Xu, Lipeng Zhang, and Rui Zhang. 2020. SSVEP stimulus layout effect on accuracy of brain-computer interfaces in augmented reality glasses. *IEEE Access* 8 (2020), 5990–5998.