Frequency-Aware Class-Incremental Learning for SSVEP-Based Brain–Computer Interfaces

Jiayu An, Huanyu Wu and Dongrui Wu, *Fellow, IEEE*

***Abstract*—Steady-state visual evoked potential (SSVEP)–based brain–computer interfaces (BCIs) have achieved significant progress in recent years. However, most existing SSVEP classification methods are designed under offline settings with a fixed set of stimulus frequencies, limiting their flexibility in practical applications, where SSVEP systems are gradually extended from a small set of stimulus frequencies to denser multi-frequency paradigms. To address this limitation, this paper focuses on class-incremental learning (CIL) for SSVEP-based BCIs and proposes a unified framework to alleviate catastrophic forgetting. Specifically, we propose a frequency-aware rehearsal strategy that exploits the periodic structure of SSVEP signals to improve replay efficiency under limited memory, and feature compactness regularization that enforces intra-class feature consistency across incremental learning stages. Extensive experiments on two public SSVEP datasets demonstrated that the proposed method consistently outperforms representative baselines and effectively mitigates catastrophic forgetting.**

***Index Terms*—class-incremental learning, EEG, SSVEP, BCI**

I. Introduction

A brain–computer interface (BCI) measures and processes neural activities and translates them into commands for external device control [1]. Steady-state visual evoked potential (SSVEP) [2] has attracted considerable attention due to its high information transfer rate and signal-to-noise ratio. SSVEP signals are elicited when a subject focuses on visual stimuli flickering at specific frequencies, producing neural responses that are strongly correlated with the stimulation frequency. This frequency-locked characteristic enables reliable and efficient decoding, making SSVEP widely applied in text input [3], [4], phone dialing [5], [6], and robot control [7], [8].

Existing SSVEP classification methods can be broadly divided into traditional signal processing approaches and deep learning–based methods. Benefiting from the frequency-specific nature of SSVEP, several calibration-free methods have been proposed, such as canonical correlation analysis (CCA) [2] and filter bank CCA (FBCCA) [9]. Methods requiring calibration, including task-related component analysis (TRCA) [10] and task-discriminant component analysis (TDCA) [11], further improve performance by learning subject-specific spatial filters. However, these traditional methods rely heavily on manual features and prior assumptions, and their performance degrades when only short Electroencephalography (EEG) segments are available. To address these limitations, deep learning methods have been introduced to perform end-to-end SSVEP decoding [12]. More recent studies further combine SSVEP decoding with recurrent architectures [13] or transformer-based models [14], achieving improved representation capability.

J. An, H. Wu and D. Wu are with the Ministry of Education Key Laboratory of Image Processing and Intelligent Control, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074 China.

D. Wu is the corresponding author (e-mail: drwu09@gmail.com).

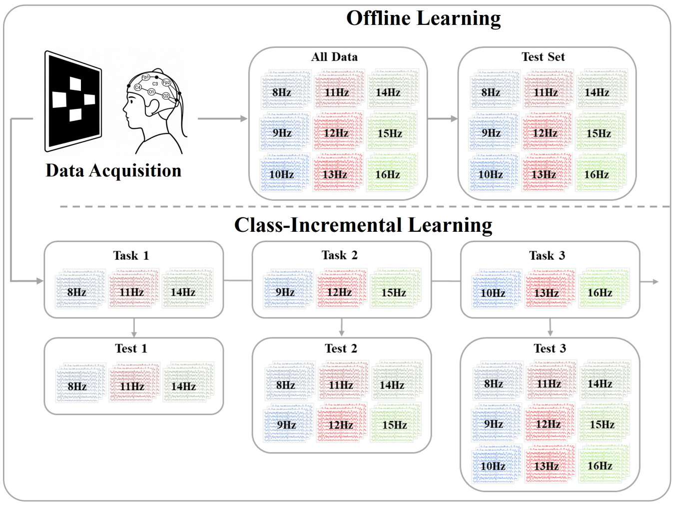

Fig. 1. Comparison between offline learning and CIL.

Despite their effectiveness, existing SSVEP classification methods are primarily designed for offline training with a fixed set of stimulus frequencies, as illustrated in Fig. 1 (top). In practical BCI applications, SSVEP systems often need to continuously accommodate new stimulus frequencies. For example, early SSVEP-based devices typically adopt paradigms with a small number of well-separated frequencies to ensure reliable discrimination, while more advanced systems gradually expand to denser multi-frequency stimulation paradigms to support richer commands and higher information throughput. Since many stimulus frequencies are shared across different devices and paradigms, the ability to incrementally incorporate new frequencies is critical for improving device scalability and general applicability.

Therefore, this paper focuses on class-incremental learning (CIL) [15] in SSVEP-based BCIs, as shown in Fig. 1 (bottom). CIL aims to enable models to continuously learn new classes while preserving performance on previously learned ones, which is challenging due to catastrophic forgetting [16]. In SSVEP-based BCIs, calibration-free methods partially alleviate this issue by constructing new templates according to newly occurred stimulus frequencies, without modifying previously learned representations. In contrast, deep learning–based SSVEP methods rely on updating model parameters when new classes are added, which interferes with existing knowledge and leads to severe performance degradation on previously learned classes. Existing CIL methods mainly address catastrophic forgetting based on regularization, rehearsal, and parameter isolation [17]. However, they are not designed for the frequency-specific characteristics of SSVEP signals, and directly applying them to SSVEP-based BCIs is suboptimal.

To address the above challenges, we propose a CIL framework for SSVEP-based BCIs, which considers the frequency-specific characteristics of SSVEP signals and mitigates catastrophic forgetting. Specifically, we propose a frequency-aware rehearsal (FAR) strategy that augments replay samples by exploiting the periodic structure of SSVEP signals, thereby enhancing the diversity and stability of previously learned frequency representations. In addition, we propose feature compactness regularization (FCR) to encourage intra-class feature consistency across incremental learning stages, further alleviating representation drift.

Our main contributions are:

- We introduce CIL into SSVEP-based BCIs and establish a unified evaluation protocol that enables systematic comparison between traditional SSVEP methods and deep learning–based CIL methods.
- We propose FAR and FCR, which explicitly exploit the frequency-specific characteristics of SSVEP from both data and feature perspectives.
- Extensive experiments on two public SSVEP datasets demonstrated that the proposed method consistently outperforms existing approaches under the CIL setting.

II. Related Works

This section provides a brief overview of SSVEP classification and CIL methods.

A. SSVEP Classification Methods

Existing SSVEP classification methods can be broadly categorized into traditional signal processing–based and deep learning–based methods.

Traditional SSVEP classification methods mainly exploit the frequency-locked characteristics of SSVEP signals. CCA [2] computes the correlation between EEG signals and predefined sinusoidal reference signals at stimulus frequencies and has been widely adopted due to its simplicity and calibration-free nature. Several CCA variants have been further proposed to enhance performance [18]–[20]. To further enhance performance, FBCCA [9] decomposes EEG signals into multiple sub-bands and combines correlation scores across frequency bands, improving robustness and accuracy. Beyond calibration-free approaches, spatial filtering–based methods have been proposed to leverage subject-specific information. TRCA [10] learns spatial filters by maximizing the reproducibility of SSVEP responses across trials, while TDCA [11] further incorporates discriminative information among different stimulus frequencies. Despite their effectiveness, traditional SSVEP methods rely heavily on handcrafted features and their performance often degrades when the signal duration is short.

Deep learning–based methods have been introduced for end-to-end SSVEP decoding. Convolutional neural network (CNN) architectures, such as EEGNet [21] and compact CNN variants [12], [22], automatically learn spatial–temporal features directly from raw EEG signals. More advanced models further incorporate recurrent neural networks, such as long short-term memory (LSTM), to capture temporal dependencies in SSVEP signals [13]. Recently, transformer-based models have also been explored to model long-range dependencies and global temporal patterns [14], demonstrating improved representation capability. While deep learning methods generally achieve superior performance and stronger feature learning capacity, they typically require offline training with a fixed set of stimulus frequencies, which limits their applicability in dynamic SSVEP scenarios.

More recently, EEG foundation models pre-trained on large-scale EEG datasets have been explored to improve representation learning and generalization across tasks [23]. While deep learning and foundation model–based approaches generally achieve superior performance and stronger feature learning capacity, they are still predominantly designed for offline training with a fixed set of stimulus frequencies, which limits their applicability in dynamic SSVEP scenarios.

B. CIL methods

Current CIL approaches that alleviate catastrophic forgetting can be broadly classified into three categories.

Regularization-based approaches mitigate forgetting by introducing additional constraints to restrict updates of parameters that are important for previously learned classes. Elastic Weight Consolidation (EWC) [24] estimates parameter importance using the Fisher information matrix and penalizes changes to critical parameters. Memory Aware Synapses (MAS) [25] evaluates parameter importance based on the sensitivity of model outputs to parameter variations. Learning Without Forgetting (LwF) [26] preserves prior knowledge by constraining the discrepancy between outputs of the old and updated models through knowledge distillation.

Rehearsal-based approaches explicitly store samples from previously learned classes or generate pseudo-samples and replay them during training on new classes. Experience Replay (ER) [27] maintains a memory buffer to jointly train on old and new data. Incremental Classifier and Representation Learning (iCaRL) [28] selects representative exemplars via herding and performs classification using class prototypes. Gradient Episodic Memory (GEM) [29] constrains gradient updates to ensure that the loss on stored samples does not increase.

Parameter isolation–based approaches alleviate forgetting by allocating separate model parameters or subnetworks to different tasks or class groups, thereby reducing interference among learned representations. Despite their success, existing CIL methods are not tailored to the frequency-specific characteristics of SSVEP signals, leading to suboptimal performance in practical incremental learning settings.

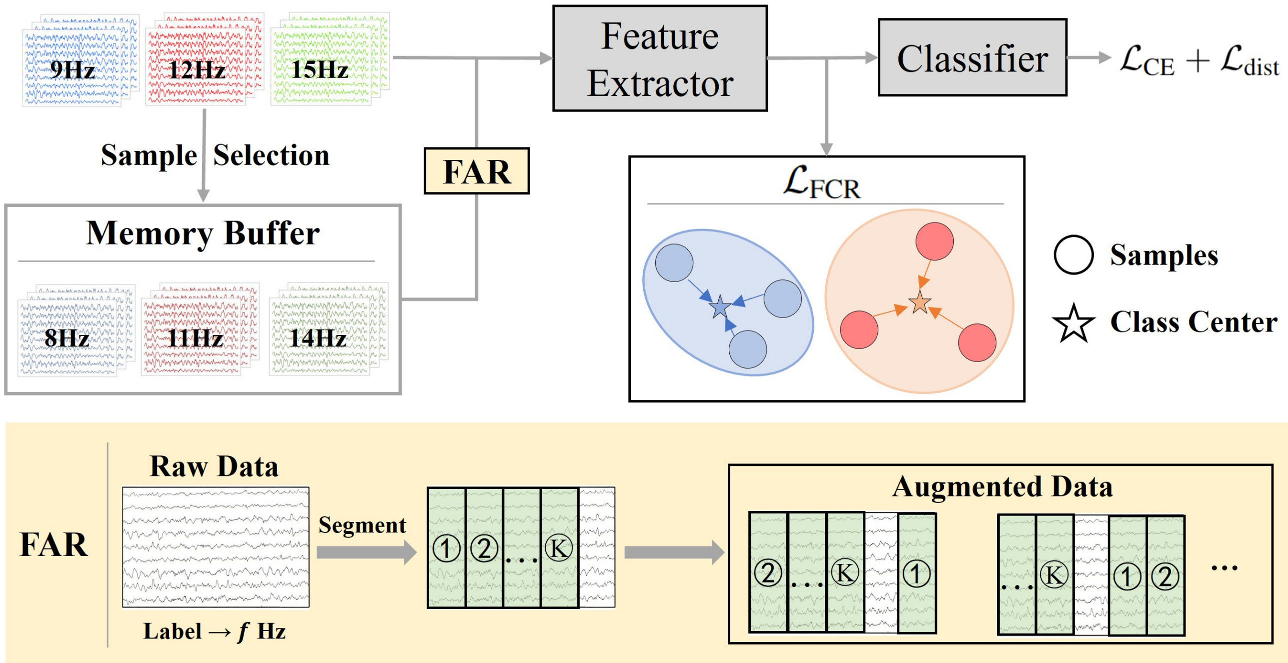

Fig. 2. Overview of our proposed framework for SSVEP-based CIL, which includes a feature extractor and classifier optimized by classification, distillation and FCR loss. FAR exploits the periodic structure of SSVEP signals to augment replay samples via period-level rolling. FCR encourages compact intra-class distributions by minimizing the distance between individual samples and their respective class centers.

III. Method

This section introduces our CIL framework for SSVEP-based BCIs, which aims to mitigate catastrophic forgetting under incremental frequency expansion by exploiting the frequency-specific characteristics of SSVEP signals.

As illustrated in Fig. 2, the proposed method adopts a rehearsal-based CIL pipeline and consists of two key components. FAR improves replay efficiency by leveraging the periodic structure of SSVEP signals, while FCR is employed to stabilize feature representations by enforcing intra-class consistency across incremental learning stages. The details are presented in the following subsections.

A. Problem Setting

Consider CIL for SSVEP-based BCI, where a sequence of T tasks is provided. For the t-th task, a dataset $\mathcal{D}_t = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N_t}$ is available, where $\mathbf{x}_i \in \mathcal{X}$ and $y_i \in \mathcal{Y}$ denote the input sample and its corresponding label, respectively. Across all tasks, there are C classes in total. Each task introduces a disjoint class set $\mathcal{C}_t$, such that $\mathcal{C} = \bigcup_{t=1}^{T} \mathcal{C}_t$ and $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$ for $i \neq j$.

Unlike offline training, where all data are accessible, CIL assumes that only a limited memory buffer $\mathcal{M} = \{(\boldsymbol{x}_i, y_i)\}_{i=1}^{N_s}$ containing samples from previous tasks can be stored. At $t = 1$, the model is trained using $\mathcal{D}_1$. For $t > 1$, training is performed on $\mathcal{D}_t \cup \mathcal{M}$. The goal is to correctly classify newly introduced classes $\mathcal{C}_t$ while maintaining satisfactory performance on previously learned classes $\mathcal{C}_i$ for $i < t$.

B. Unified CIL Framework

Our framework follows the standard CIL paradigm and is inspired by iCaRL [28]. At each incremental stage, the learning process mainly consists of two steps: training on the new task and sample rehearsal.

Specifically, when a new task t arrives, the model is trained using the union of the current task data $\mathcal{D}_t$ and a memory buffer $\mathcal{M}$ that stores a limited number of samples from previously learned classes. The training objective combines a classification loss and a distillation loss to mitigate catastrophic forgetting:

$$\mathcal{L} = \mathcal{L}_{\text{CE}} + \mathcal{L}_{\text{dist}}, \tag{1}$$

where $\mathcal{L}_{\text{CE}}$ denotes the cross-entropy loss for classifying new classes and the distillation loss is defined as

$$\mathcal{L}_{\text{dist}} = \sum_{(x,y)\in\mathcal{M}} \|f_{\text{old}}(\boldsymbol{x}) - f_{\text{new}}(\boldsymbol{x})\|_2^2, \tag{2}$$

where $f(\cdot)$ denotes the output logits of the model and it constrains the outputs of the current model f_{new} to be consistent with the outputs of the previous model f_{old}.

After training on task t, a subset of representative samples from the current task is selected and added to the memory buffer $\mathcal{M}$ under a fixed memory budget. Following iCaRL, we adopt herding [30] to select representative samples for each class.

During inference, instead of using the parametric classification head, we perform classification based on nearest class prototypes [31] in the feature space, where each class prototype is computed as the mean feature representation of its stored samples.

C. FAR

Standard rehearsal strategies typically replay a limited number of stored samples from previous classes. In SSVEP-based BCIs, EEG signals exhibit strong frequency-locked periodic structures, which provide additional exploitable information beyond raw time-series samples. Leveraging this property enables more effective use of a small memory buffer, thereby improving rehearsal efficiency under strict memory constraints.

We propose FAR strategy that exploits the periodic structure of SSVEP signals to generate augmented replay samples. Given an SSVEP sample $\boldsymbol{x}$ corresponding to a stimulus frequency f, the fundamental period length is computed as $P = F_s/f$, where F_s denotes the sampling rate. The signal is then segmented according to this period length.

Based on the period-wise representation, FAR performs period-level rolling augmentation. Given an augmentation number K, K augmented samples are generated by circularly shifting the signal at the period level: the k-th augmented sample is obtained by moving the first k periods to the end of the signal, for $k = 1, \ldots, K$. All augmented samples preserve the original temporal order within each period and inherit the same class label.

By exploiting the intrinsic periodicity of SSVEP signals, FAR enables a small number of stored samples to produce diverse and frequency-consistent replay data. This improves the effectiveness of rehearsal and enhances knowledge retention for previously learned SSVEP classes during incremental learning.

D. FCR

In practical SSVEP-based BCI systems, stimulus frequencies are typically designed to be well separated at the initial stage to ensure high inter-class discriminability. As new classes are introduced, however, their corresponding frequencies are often interpolated between existing ones, leading to reduced inter-class margins across incremental tasks. Although classes within the same task may remain well separated, the reduced discriminability across tasks can aggravate representation drift and catastrophic forgetting.

To address this issue, we propose FCR, which enforces intra-class feature consistency during training. Let $\phi(\cdot)$ denote the feature extractor. At each training epoch, we compute the feature center for each class c using the samples of that class in the current training set $\mathcal{D}_t \cup \mathcal{M}$ as

$$\boldsymbol{\mu}_c = \frac{1}{|\mathcal{S}_c|} \sum_{(x,y)\in\mathcal{S}_c} \phi(\boldsymbol{x}), \tag{3}$$

where $\mathcal{S}_c = \{(\boldsymbol{x}, y) \mid y = c\}$ denotes the set of samples belonging to class c.

Algorithm 1: Class-Incremental Learning for SSVEP with FAR and FCR

Input: Task sequence $\{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_T\}$; memory buffer $|\mathcal{M}|$; sampling rate F_s; backbone network $f(\cdot)$; loss weights λ.
Output: The classification model with class prototypes.

```
Initialize model parameters θ;
Initialize memory buffer M ← ∅;
for t ← 1 to T do
    Construct training set T_t ← D_t ∪ M;
    for each epoch do
        // FAR
        Generate augmented replay samples from M
          by period-level rolling;
        // FCR
        Compute class centers {bmμ_c} using samples
          in T_t;
        // Model update
        Update θ by minimizing
          L = L_CE + L_dist + λL_FCR;
    end
    // Memory update
    Select representative samples from D_t through
      herding under the fixed budget;
    Update memory buffer M;
    Calculate class prototypes using the memory buffer
      M ;
end
```

Based on the class centers, FCR introduces an ℓ_2 regularization term to encourage features to be close to their corresponding class centers:

$$\mathcal{L}_{\text{FCR}} = \sum_{(\boldsymbol{x},y)\in\mathcal{D}_t\cup\mathcal{M}} \|\phi(\boldsymbol{x}) - \boldsymbol{\mu}_y\|_2^2. \tag{4}$$

The proposed regularization is jointly optimized with the classification and distillation objectives during training, promoting compact intra-class feature distributions and reducing representation drift across incremental learning stages.

Algorithm 1 summarizes the overall training procedure of the proposed CIL framework for SSVEP-based BCIs.

IV. Experiments

This section performs experiments to demonstrate the effectiveness of the proposed method.

A. Datasets

We conducted experiments on two public SSVEP datasets, Benchmark [32] and BETA [33].

The Benchmark dataset contains EEG recordings from 35 healthy subjects. In each block, subjects were presented with a 5×8 matrix of 40 visual targets flickering at different frequencies ranging from 8 to 15.8 Hz with a step of 0.2 Hz. EEG signals were recorded using 64 channels. Each trial consists of a visual cue followed by a stimulus period and a short rest interval.

The BETA dataset follows a similar experimental paradigm but includes EEG data from 70 healthy subjects collected in more realistic environments. Visual stimuli were presented in a keyboard-like layout. Compared with Benchmark, BETA exhibits lower signal-to-noise ratios due to unconstrained acquisition conditions, making classification more challenging.

For both datasets, we reorganized the original data into CIL setting with five sequential tasks, where each task introduces a disjoint subset of SSVEP classes. The classes were assigned to tasks following the original frequency label order provided by the datasets, such that new stimulus frequencies are incrementally introduced as learning progresses. EEG segments of 2 seconds were extracted for classification, while the initial cue-related period is discarded, resulting in slightly shorter effective signal durations.

B. Experimental Setup

We adopted SSVEPNet [13] as the backbone network for all experiments. The model was trained using the Adam optimizer with an initial learning rate of 0.001 and a weight decay of 0.001. For each dataset, K-fold cross-validation was performed according to the number of available blocks. Nine occipital EEG channels were used. To construct the memory buffer, we store 5% of samples for each class under a fixed memory budget, resulting in approximately 8 stored samples per class in experiments. λ was set to 0.1.

We evaluated CIL performance using the average accuracy A_t after learning task t and backward transfer (BWT) after learning all tasks, defined as

$$A_t = \frac{1}{t}\sum_{j=1}^{t} a_{t,j}, \tag{5}$$

$$\mathrm{BWT} = \frac{1}{T-1}\sum_{j=1}^{T-1}\left(a_{T,j} - a_{j,j}\right), \tag{6}$$

where $a_{t,j} \in [0,1]$ denotes the classification accuracy on the test set of task j after the model has completed training on task t.

C. Baselines

We compared the proposed method with a diverse set of baselines, covering offline training, traditional SSVEP methods, and representative CIL methods.

- Upper bound and lower bound. Joint training trains a single model using data from all tasks simultaneously and serves as an upper bound. Finetune sequentially trains the model on new tasks and serves as a lower bound.
- Representative calibration-free and calibration-based SSVEP methods, including CCA [2], FBCCA [9], TRCA [10], and TDCA [11], which are widely used in SSVEP classification.
- Representative CIL methods. Regularization-based methods include EWC [24], MAS [25], and LwF [26], which constrain parameter updates to preserve previous knowledge. Replay-based methods include ER [27], BiC [34], A-GEM [35] and iCaRL [28], which mitigate forgetting by replaying stored samples or constraining gradient updates.

TABLE I
EXPERIMENT RESULTS ON BENCHMARK.

Method	After Learning Task					BWT
	1	2	3	4	5	
Joint	90.12	89.79	88.89	89.05	88.25	-3.48
Finetune	88.27	45.83	30.59	23.08	18.54	-72.81
CCA	72.56	71.99	71.09	69.31	69.14	0.48
FBCCA	83.87	78.96	76.79	75.64	74.7	-4.38
TRCA	65.42	69.38	69.56	69.42	68.77	**2.5**
EWC	88.27	46.4	30.93	23.26	18.68	-73.38
MAS	88.27	56.16	39.88	30.22	23.77	-67.91
LWF	88.27	45.83	30.59	23.08	18.54	-72.81
TDCA	72.2	38.66	28.05	24.97	20.93	-54.04
A-GEM	**88.27**	59.07	31.49	24.24	20.92	-71.1
ER	90.12	70.47	60.14	58.05	54.24	-39.19
BiC	87.74	73.51	66.8	53.85	54.01	-25.77
iCaRL	88.33	83.75	80.69	78.92	77.76	-6.81
Ours	87.38	**84.94**	**83.02**	**80.71**	**79.44**	-4.12

D. Main Results

Tables I and II summarize the overall CIL performance on the Benchmark and BETA datasets. Observe that:

- Traditional SSVEP methods exhibit relatively stable performance across incremental tasks. However, calibration-free methods such as CCA and FBCCA perform poorly when only short EEG segments are available, as their correlation-based matching relies on sufficient temporal information for reliable frequency estimation. In addition, CIL incrementally introduces classes while aggregating data from all subjects, rather than performing subject-wise incremental learning. Filter-based methods such as TRCA suffer from degraded filter estimation, and TDCA further experiences performance drops due to its sequential filter construction, which differs significantly from the scenario where all classes are jointly observed.
- Representative CIL methods alleviate catastrophic forgetting to some extent, but forgetting remains substantial as tasks progress. In several cases, their performance is even inferior to traditional SSVEP methods.
- The proposed method consistently achieves superior performance across almost all incremental tasks on both datasets. By explicitly exploiting the frequency-specific characteristics of SSVEP signals and enforcing feature compactness during training, our approach effectively mitigates catastrophic forgetting and maintains strong performance as new classes are continuously introduced.

TABLE II
EXPERIMENT RESULTS ON BETA.

Method	After Learning Task 1	2	3	4	5	BWT
Joint	78.66	70.94	68.32	64.63	63.6	-6.54
Finetune	76.65	36.11	25.16	18.35	15.62	-59.55
CCA	56.34	54.26	51.49	49.76	47.43	0.7
FBCCA	72.9	**67.3**	60.22	57.28	54.59	-14.82
TRCA	64.87	61.01	56.22	52.42	49.88	0.2
EWC	76.07	35.62	25.48	18.08	14.82	-59.21
MAS	76.07	36.41	24.91	16.44	13.43	-57.36
LWF	76.07	36.41	25.31	17.87	15.18	-59.26
TDCA	42.41	18.55	12.43	6.48	5.67	-16.67
A-GEM	76.07	39.82	25.82	17.9	14.77	-58.98
ER	**78.17**	50.22	41.88	36.34	30.83	-45
BiC	77.01	37.39	33.06	30.16	29.95	**3.21**
iCaRL	76.25	66.61	60.79	55.89	52.99	-13.18
Ours	76.3	67.21	**62.69**	**58.42**	**55.84**	-6.94

TABLE III
ABLATION STUDY RESULTS ON THE BENCHMARK DATASET.

FAR	FCR	After Learning Task 1	2	3	4	5	BWT
		88.33	83.75	80.69	78.92	77.76	-6.81
✓		88.15	84.41	82.4	80.59	78.55	-4.7
	✓	87.38	84.79	82.23	79.86	78.34	-7.15
✓	✓	87.38	**84.94**	**83.02**	**80.71**	**79.44**	**-4.12**

E. Ablation Studies

We conducted ablation studies to evaluate the effectiveness of each proposed component, as reported in Table III.

The results show that incorporating either FAR or FCR individually leads to improved performance compared to the baseline CIL framework, indicating that both components contribute to alleviating catastrophic forgetting. FAR consistently enhances performance by improving the effectiveness of replay under limited memory, while preserving previously learned frequency information. Although FCR improves overall classification accuracy by enforcing more compact intra-class feature distributions, it may introduce increased forgetting when applied alone. This is because FCR encourages strong feature clustering based on the current training data, which can bias the representation toward new classes in the absence of sufficiently diverse replay samples. As a result, representations of previously learned classes may drift when replay is limited.

When FAR and FCR are combined, the complementary strengths of the two components effectively improve the performance and mitigate catastrophic forgetting.

F. Effect of FAR

We investigated the effectiveness of FAR from two complementary perspectives.

First, we integrated FAR into existing replay-based CIL methods and evaluated its impact on overall performance. As shown in Fig. 3, incorporating FAR consistently improves performance across different replay-based CIL methods, indicating that FAR is a general and effective enhancement for SSVEP.

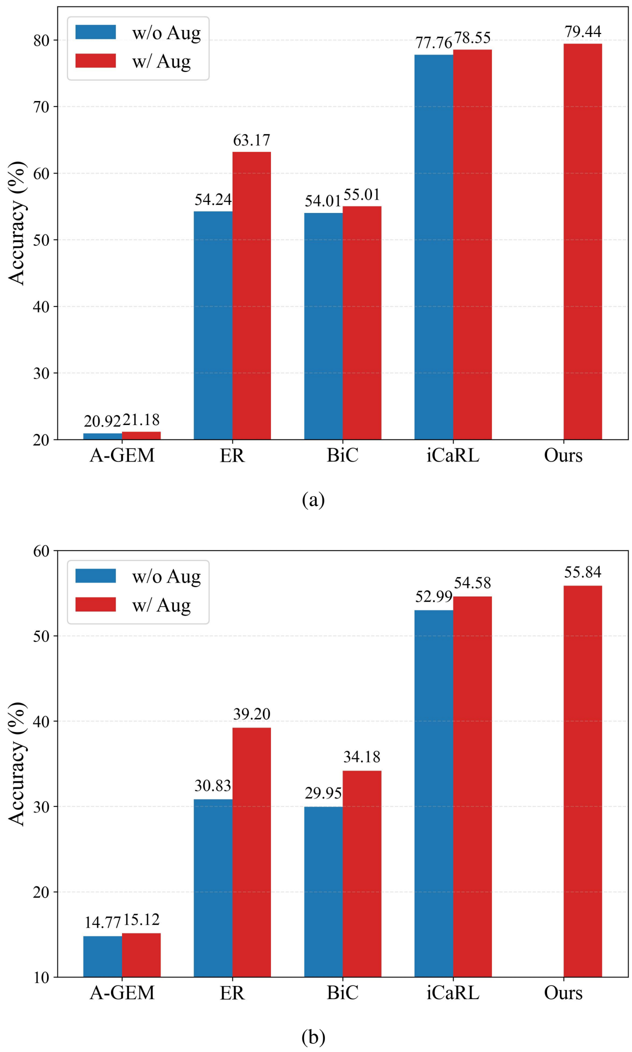

Fig. 3. Comparison of classification accuracy (%) with and without the proposed FAR across different replay-based CIL methods. (a) Benchmark; and (b) BETA.

Second, we compared FAR with several commonly used EEG data augmentation methods, including:

- None, where no data augmentation is applied;
- Alter, which replaces a randomly selected channel with the channel-wise average signal;
- Shuffle, which randomly permutes EEG channels;
- Scaling, which scales randomly selected channels using a factor α sampled from a normal distribution $\mathcal{N}(1, 1)$ and clipped to $0.1 \leq \alpha \leq 2$.

The results were reported in Table IV.

The results show that Noise and Alter provide limited performance gains, while Shuffle and Scaling may even degrade performance due to disruption of spatial channel structures. Unlike these common augmentations that introduce stochastic

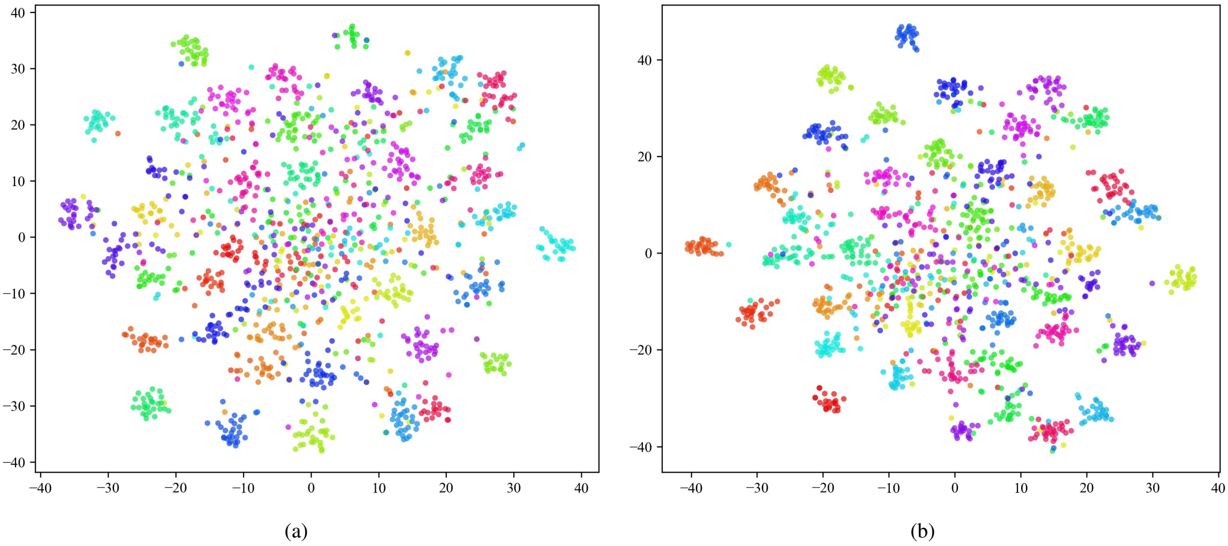

Fig. 4. t-SNE visualization of feature representations (a) without and (b) with FCR.

TABLE IV
COMPARISON OF DIFFERENT DATA AUGMENTATION METHODS.

Method	After Learning Task					BWT
	1	2	3	4	5	
None	**88.33**	83.75	80.69	78.92	77.76	-6.81
Noise	87.38	84.52	82.32	79.91	78.67	-6.98
Alter	87.38	84.88	82	80.18	78.5	-6.56
Shuffle	87.38	83.57	79.96	75.61	70.53	-13.67
Scaling	87.38	84.47	81.61	78.2	74.71	-10.69
FAR (Ours)	87.38	**84.94**	**83.02**	**80.71**	**79.44**	**-4.12**

perturbations, FAR is a domain-knowledge-driven strategy that preserves the intrinsic periodicity and phase-coherence of SSVEP signals. FAR consistently achieves the best performance among all augmentation strategies, demonstrating its effectiveness and robustness by explicitly exploiting the frequency-specific characteristics of SSVEP signals.

G. Effect of FCR

We analyzed the effect of FCR through feature visualization. Fig. 4 presents t-SNE projections of the learned feature representations without and with FCR.

Without FCR, feature distributions of different classes exhibit noticeable overlap, indicating less compact intra-class representations and unclear inter-class boundaries. In contrast, introducing FCR leads to more compact intra-class clusters and clearer separation between different classes. This visualization result demonstrates that FCR effectively enhances feature compactness and discriminability, which is beneficial for reducing representation drift in CIL.

H. Effect of Replay Samples

We analyzed the impact of the number of replay samples stored per class on CIL performance. Specifically, we varied the number of stored samples per class among $\{1, 3, 5, 8, 10\}$ and compared the proposed method with representative baseline approaches, as shown in Fig. 5.

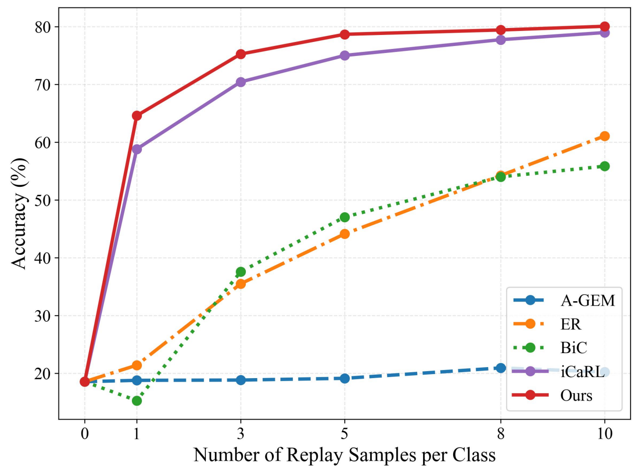

Fig. 5. Performance comparison under different numbers of replay samples per class.

The results show that increasing the number of replay samples generally leads to improved performance for all methods, indicating that larger memory buffers help mitigate catastrophic forgetting. However, the proposed method consistently outperforms baseline methods under all memory settings. Notably, even when only one sample per class is stored, our method achieves strong performance.

V. CONCLUSION

This paper proposes a CIL framework for SSVEP-based BCIs to address the limitation of fixed-class offline training. By explicitly considering the frequency-specific characteristics of SSVEP signals, we propose FAR to enhance replay efficiency and FCR to stabilize feature representations during incremental learning. Experimental results on two public datasets demonstrate that the proposed method consistently outperforms existing approaches and effectively alleviates catastrophic forgetting. This paper provides a first step toward continuous and adaptive SSVEP-based BCIs, and future work will explore more challenging incremental scenarios and real-world online applications.

ACKNOWLEDGEMENT

This research was supported by Brain Science and Brain-like Intelligence Technology-National Science and Technology Major Project 2021ZD0201300.

REFERENCES

[1] F. Lotte, L. Bougrain, A. Cichocki, M. Clerc, M. Congedo, A. Rakotomamonjy, and F. Yger, "A review of classification algorithms for EEG-based brain–computer interfaces: A 10 year update," *Journal of Neural Engineering*, vol. 15, no. 3, p. 031005, 2018.

[2] Z. Lin, C. Zhang, W. Wu, and X. Gao, "Frequency recognition based on canonical correlation analysis for SSVEP-based BCIs," *IEEE Trans. on Biomedical Engineering*, vol. 53, no. 12, pp. 2610–2614, 2006.

[3] E. Yin, Z. Zhou, J. Jiang, Y. Yu, and D. Hu, "A dynamically optimized SSVEP brain–computer interface (BCI) speller," *IEEE Trans. on Biomedical Engineering*, vol. 62, no. 6, pp. 1447–1456, 2015.

[4] X. Chen, Y. Wang, M. Nakanishi, X. Gao, T.-P. Jung, and S. Gao, "High-speed spelling with a noninvasive brain–computer interface," *Proc. of the National Academy of Sciences*, vol. 112, no. 44, 2015.

[5] Y. Wang, R. Wang, X. Gao, B. Hong, and S. Gao, "A practical VEP-based brain-computer interface," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 14, no. 2, pp. 234–240, 2006.

[6] Y.-T. Wang, Y. Wang, and T.-P. Jung, "A cell-phone-based brain–computer interface for communication in daily life," *Journal of Neural Engineering*, vol. 8, no. 2, p. 025018, 2011.

[7] R. Ortner, B. Z. Allison, G. Korisek, H. Gaggl, and G. Pfurtscheller, "An SSVEP BCI to control a hand orthosis for persons with tetraplegia," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 19, no. 1, pp. 1–5, 2011.

[8] X. Chen, B. Zhao, Y. Wang, S. Xu, and X. Gao, "Control of a 7-dof robotic arm system with an SSVEP-based BCI," *Int'l Journal of Neural Systems*, vol. 28, no. 08, p. 1850018, 2018.

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

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

[11] B. Liu, X. Chen, N. Shi, Y. Wang, S. Gao, and X. Gao, "Improving the performance of individually calibrated SSVEP-BCI by task- discriminant component analysis," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 29, pp. 1998–2007, 2021.

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

[13] Y. Pan, J. Chen, Y. Zhang, and Y. Zhang, "An efficient CNN-LSTM network with spectral normalization and label smoothing technologies for SSVEP frequency recognition," *Journal of Neural Engineering*, vol. 19, no. 5, p. 056014, 2022.

[14] J. Chen, Y. Zhang, Y. Pan, P. Xu, and C. Guan, "A transformer-based deep neural network model for SSVEP classification," *Neural Networks*, vol. 164, pp. 521–534, 2023.

[15] D.-W. Zhou, Q.-W. Wang, Z.-H. Qi, H.-J. Ye, D.-C. Zhan, and Z. Liu, "Class-incremental learning: A survey," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 9851–9873, 2024.

[16] S. Kolouri, N. Ketz, X. Zou, J. Krichmar, and P. Pilly, "Attention-based structural-plasticity," *arXiv preprint arXiv:1903.06070*, 2019.

[17] L. Wang, X. Zhang, H. Su, and J. Zhu, "A comprehensive survey of continual learning: Theory, method and application," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 46, no. 8, pp. 5362–5383, 2024.

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

[19] Y. Zhang, G. Zhou, Q. Zhao, A. Onishi, J. Jin, X. Wang, and A. Cichocki, *Multiway Canonical Correlation Analysis for Frequency Components Recognition in SSVEP-Based BCIs*, 2011, pp. 287–295.

[20] A. A. P. Wai, H. Guo, Y. Chi, L. Zhang, X.-S. Hua, and C. Guan, "Optimizing filter-bank canonical correlation analysis for fast response SSVEP brain-computer interface (BCI)," in *Proc. of Int'l Joint Conf. on Neural Networks*, Glasgow, UK, Jul. 2020, pp. 1–8.

[21] V. J. Lawhern, A. J. Solon, N. R. Waytowich, S. M. Gordon, C. P. Hung, and B. J. Lance, "EEGNet: A compact convolutional neural network for EEG-based brain–computer interfaces," *Journal of Neural Engineering*, vol. 15, no. 5, p. 056013, 2018.

[22] J.-L. Chou, Y.-N. Huang, and C.-S. Wei, "A unified brain signal decoder based on multi-branch architecture," in *Proc. of Int'l Joint Conf. on Neural Networks*, Yokohama, Japan, Jun. 2024, pp. 1–8.

[23] D. Liu, Y. Chen, Z. Chen, Z. Cui, Y. Wen, J. An, J. Luo, and D. Wu, "EEG foundation models: Progresses, benchmarking, and open problems," *arXiv preprint arXiv:2601.17883*, 2026.

[24] J. Kirkpatrick, R. Pascanu, N. Rabinowitz, J. Veness, G. Desjardins, A. A. Rusu, K. Milan, J. Quan, T. Ramalho, A. Grabska-Barwinska, D. Hassabis, C. Clopath, D. Kumaran, and R. Hadsell, "Overcoming catastrophic forgetting in neural networks," *Proc. of the National Academy of Sciences*, vol. 114, no. 13, pp. 3521–3526, 2017.

[25] R. Aljundi, F. Babiloni, M. Elhoseiny, M. Rohrbach, and T. Tuytelaars, "Memory aware synapses: Learning what (not) to forget," in *Proc. of the European Conf. on Computer Vision*, Munich, Germany, Sep. 2018.

[26] Z. Li and D. Hoiem, "Learning without forgetting," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 40, no. 12, pp. 2935–2947, 2018.

[27] A. Chaudhry, M. Rohrbach, M. Elhoseiny, T. Ajanthan, P. K. Dokania, P. H. Torr, and M. Ranzato, "On tiny episodic memories in continual learning," *arXiv preprint arXiv:1902.10486*, 2019.

[28] S. A. Rebuffi, A. Kolesnikov, G. Sperl, and C. H. Lampert, "iCaRL: Incremental classifier and representation learning," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, Honolulu, HI, Jul. 2017, pp. 5533–5542.

[29] D. Lopez-Paz and M. A. Ranzato, "Gradient episodic memory for continual learning," *Advances in Neural Information Processing Systems*, vol. 30, no. 10, pp. 6470–6479, 2017.

[30] M. Welling, "Herding dynamical weights to learn," in *Proc. of the Annual Int'l Conf. on Machine Learning*, Miami, FL, Jun. 2009, pp. 1121–1128.

[31] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Distance-based image classification: Generalizing to new classes at near-zero cost," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 35, no. 11, pp. 2624–2637, 2013.

[32] Y. Wang, X. Chen, X. Gao, and S. Gao, "A benchmark dataset for SSVEP-based brain–computer interfaces," *IEEE Trans. on Neural Systems and Rehabilitation Engineering*, vol. 25, no. 10, pp. 1746–1752, 2017.

[33] B. Liu, X. Huang, Y. Wang, X. Chen, and X. Gao, "BETA: A large benchmark database toward SSVEP-BCI application," *Frontiers in Neuroscience*, vol. 14, Jun. 2020.

[34] Y. Wu, Y. Chen, L. Wang, Y. Ye, Z. Liu, Y. Guo, and Y. Fu, "Large scale incremental learning," in *Proc. of IEEE/CVF Conf. on Computer Vision and Pattern Recognition*, Long Beach, CA, Jun. 2019, pp. 374–382.

[35] A. Chaudhry, M. Ranzato, M. Rohrbach, and M. Elhoseiny, "Efficient lifelong learning with A-GEM," in *International Conference on Learning Representations*, New Orleans, LA, May 2019.