

# From Selection to Scheduling: Federated Geometry-Aware Correction Makes Exemplar Replay Work Better under Continual Dynamic Heterogeneity

Zhuang Qi<sup>1</sup>, Ying-Peng Tang<sup>2</sup>, Lei Meng<sup>1,\*</sup>, Guoqing Chao<sup>3</sup>, Lei Wu<sup>1</sup>, Han Yu<sup>2</sup>, Xiangxu Meng<sup>1</sup>

<sup>1</sup>School of Software, Shandong University, China

<sup>2</sup>College of Computing and Data Science, Nanyang Technological University, Singapore

<sup>3</sup>School of Computer Science and Technology, Harbin Institute of Technology, China

z\_qi@mail.sdu.edu.cn, {yingpeng.tang, han.yu}@ntu.edu.sg,  
 {lmeng, i\_lily, mxx}@sdu.edu.cn, guoqingchao@hit.edu.cn

## Abstract

Exemplar replay has become an effective strategy for mitigating catastrophic forgetting in federated continual learning (FCL) by retaining representative samples from past tasks. Existing studies focus on designing sample-importance estimation mechanisms to identify information-rich samples. However, they typically overlook strategies for effectively utilizing the selected exemplars, which limits their performance under continual dynamic heterogeneity across clients and tasks. To address this issue, this paper proposes a Federated gEometry-Aware correcTion method, termed FEAT, which alleviates imbalance-induced representation collapse that drags rare-class features toward frequent classes across clients. Specifically, it consists of two key modules: 1) the Geometric Structure Alignment module performs structural knowledge distillation by aligning the pairwise angular similarities between feature representations and their corresponding Equiangular Tight Frame prototypes, which are fixed and shared across clients to serve as a class-discriminative reference structure. This encourages geometric consistency across tasks and helps mitigate representation drift; 2) the Energy-based Geometric Correction module removes task-irrelevant directional components from feature embeddings, which reduces prediction bias toward majority classes. This improves sensitivity to minority classes and enhances the model's robustness under class-imbalanced distributions. Experimental results show that FEAT outperforms existing methods.

## 1. Introduction

Federated learning (FL) is a distributed machine learning framework that enables multiple clients to collabora-

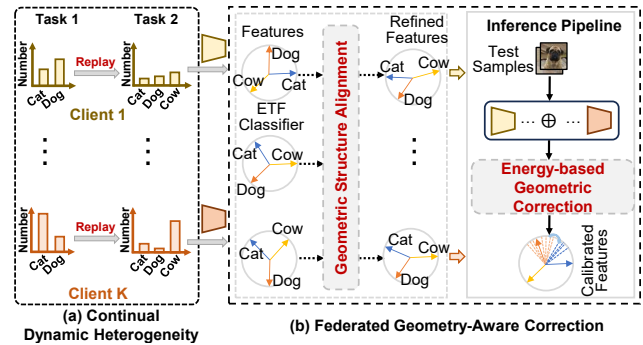


Figure 1. (a) Continual dynamic heterogeneity: clients see different class distributions over time, and each new task is learned with limited replay of past data. (b) Federated geometry-aware calibration: FEAT aligns client feature geometry and applies an energy-based correction at inference to calibrate prediction.

tively train a shared model without sharing their raw data [15, 16, 37, 42, 53, 70]. It enables periodic model parameter exchanges between clients and the server, where the server aggregates local models from clients to iteratively optimize the global model [11, 17, 38, 43, 65, 71]. In real-world FL applications, edge devices are required to handle a continuously evolving sequence of tasks, such as the continual arrival of novel classes, rather than dealing with a static single-task setting [52, 72, 74, 75]. This non-stationary data environment poses a challenge for models to achieve continual learning. Fundamentally, these challenges lie in two aspects: 1) limited local storage hinders the retention of past knowledge; 2) the inherent data heterogeneity across clients, along with the dynamically evolving data distributions, degrades the performance of collaborative learning. These factors lead to severe catastrophic forgetting in federated continual learning (FCL) [31, 60, 73].

To mitigate catastrophic forgetting, existing studies have widely adopted replay-based methods due to their simplic-

\*Corresponding author

ity and effectiveness. Based on the implementation of replay mechanisms, existing methods can be broadly categorized into two types. (1) Generative-based replay methods utilize generative models, such as Generative Adversarial Networks [2, 45, 51], to synthesize pseudo-samples that resemble historical data [50, 57, 66]. These methods train a generator to approximate the distribution of previously encountered data, which can enhance privacy and reduce memory usage [64, 76]. While conceptually appealing, they typically face significant challenges, including the high computational cost associated with training generative models and the potential suboptimal quality of generated samples [64, 66]. (2) In contrast, exemplar-based replay methods offer inherent advantages in terms of knowledge fidelity, as they avoid the challenges associated with training generative models. They typically store a small portion of the data from previous tasks in local memory and replay it during the training of new tasks [5, 29]. Despite the performance advantages demonstrated in existing studies, most of them primarily focus on how to select representative samples, while paying less attention to how to efficiently utilize these limited exemplars under resource constraints [30].

To address these issues, this paper proposes a federated geometry-aware correction method for class-incremental learning, termed FEAT, which mitigates the optimization bias arising from inter-client data heterogeneity and from the imbalance between majority and minority classes. Figure 1 illustrates the main idea of FEAT. Specifically, it is designed with two essential components: (1) the geometric structure alignment module enforces angular consistency between local feature representations and globally shared prototypes by distilling relational geometry in the embedding space. By aligning intra-batch feature correlations with those of fixed Equiangular Tight Frame prototypes, it regularizes the local learning dynamics and fosters structurally consistent representations across clients, which can enhance generalization under heterogeneous task distributions. (2) Moreover, to alleviate task-level data imbalance, the energy-based geometric correction module removes task-irrelevant components from the feature space during inference, which reduces overconfidence in majority classes and enhancing the model’s sensitivity to under-represented classes. By jointly leveraging geometric distillation and debiasing, FEAT effectively harmonizes local and global objectives, yielding improved performance under heterogeneous and imbalanced cases.

Extensive experiments were conducted on three datasets with varying levels of heterogeneity, including performance comparison, ablation study, sensitivity analysis of hyperparameters, and case study on the working mechanism of key modules. The results validate that FEAT effectively enhances the consistency of representation learning across clients and improves model robustness under long-tailed

data distributions. Across all benchmarks, FEAT outperforms seven state-of-the-art methods, showing consistent gains in Top-1 accuracy. In summary, this paper makes the following key contributions:

- This paper reveals two key challenges that persist in exemplar replay-based FCL: the use of replayed data exacerbates inter-client heterogeneity and leads to distributional imbalance between past and current task data.
- This study proposes a method orthogonal to exemplar-replay policies, enabling seamless composition with different strategies without changing selection criteria or memory allocation.
- The results validate that the prototypical angular distillation improves inter-client feature consistency, the normalized de-biasing classifier alleviates class imbalance during inference, jointly addressing these challenges.

## 2. Related Work

### 2.1. Generative Replay for FCL

Generative replay trains models such as Variational Autoencoders (VAEs [13, 49]) or Generative Adversarial Networks (GANs [14]) to approximate the sample distribution of previous tasks, so the learner can revisit pseudo-samples while training new tasks and thus mitigate catastrophic forgetting [1, 35, 46, 47, 64, 66, 69, 73, 76]. In FCL, LANDER uses label text embeddings from a pretrained language model to guide data generation [64], and GenFCIL uses a lightweight server side generator to synthesize feature representations of old classes from shared class information, reducing privacy risks and memory costs by synthesizing rather than storing data [3]. However, training reliable generators often requires substantial computation and diverse data, which may be infeasible for edge devices, and generated samples may be of limited quality, which can weaken their ability to prevent forgetting [12, 64, 76].

### 2.2. Exemplar Replay for FCL

Given the limits of generative models, many studies use exemplar replay, which stores a small set of representative samples to mitigate forgetting. By keeping real data, exemplar replay avoids the compute cost and sample quality issues of generators, offering a simple and faithful way to retain past knowledge [5, 29, 39, 55, 63, 77]. For example, Re-Fed [29] and Re-Fed+ [30] learn a personalized importance model to score local samples and keep those most relevant for reducing forgetting. FedCBDR reconstructs global features of previous tasks in a privacy-preserving way, then guides class-aware and importance-sensitive sampling for balanced replay [55]. Overall, exemplar replay is practical under limited memory [5], but most works emphasize how to select samples while overlooking how to use the chosen exemplars effectively. As a result, chal-

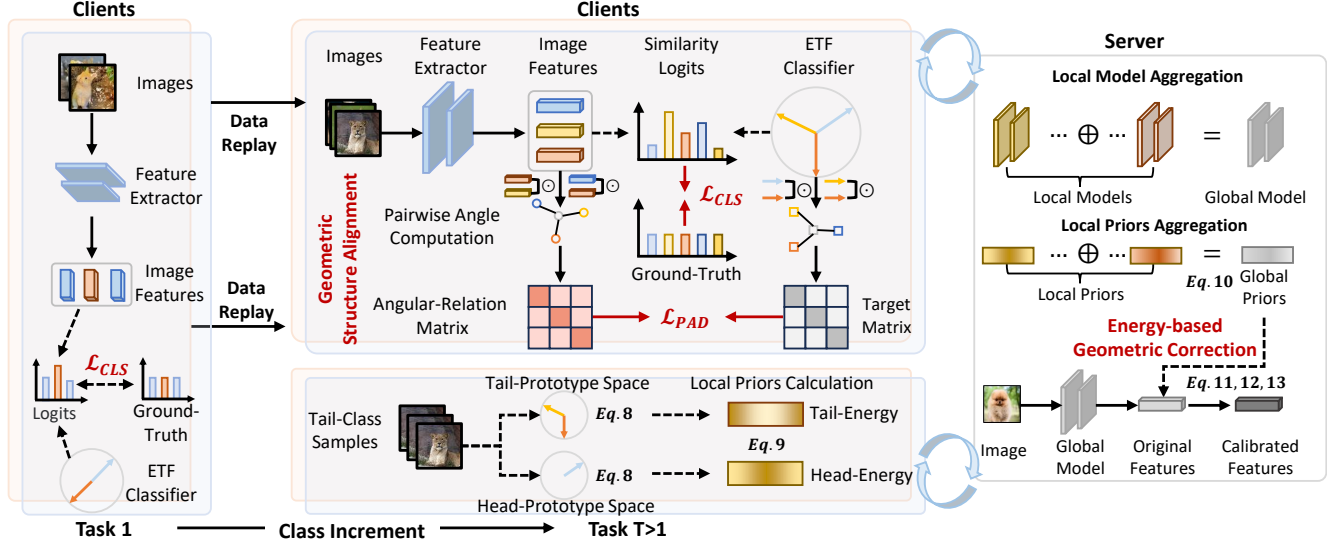


Figure 2. Illustration of the proposed FEAT framework. It contains two main modules: 1) the Geometry Structure Alignment module aligns local representations with global ETF prototypes to alleviate inter-client heterogeneity; and 2) the Energy-based Geometric Correction module removes biased components from the feature space during inference.

lenges such as client heterogeneity and task-level imbalance remain [29], which weakens the performance of the final model [6, 7, 28, 55]. Other approaches also offer insights for mitigating cross-source bias in FL [9, 10, 20–22, 27].

### 2.3. Equiangular Tight Frame in FL

Equiangular Tight Frames (ETF) emerge from Neural Collapse theory, where class means form a simplex ETF and classifier weights align with these means, yielding symmetric margins and stable optimization [41, 48]. In FL, several works fix a global simplex-ETF classifier to lessen classifier bias and partially align client representations under non-IID data [8, 34, 67]. Despite these advances, ETF-based approaches still face feature bias under continual dynamic heterogeneity, as shown in Figures 4 and 5.

## 3. Preliminaries

In federated continual learning (FCL), a central server collaborates with  $K$  distributed clients to progressively train a global model on a sequence of disjoint classification tasks. Each client  $k$  has its own task stream  $\{\mathcal{D}_k^{(1)}, \mathcal{D}_k^{(2)}, \dots, \mathcal{D}_k^{(t)}\}$ , where new tasks continually introduce unseen classes. To prevent forgetting, replay-based strategies allocate a fixed-size memory buffer of capacity  $M$  on each client, which stores up to  $N$  representative samples per previous task. When a new task  $t$  arrives, the client constructs a memory set  $\mathcal{B}_k^{(t-1)}$  by selecting  $N$  samples from each of the earlier tasks, ensuring that the total number of stored samples does not exceed  $M$ . The local training set at round  $t$  is then formed by merging the current data with

the buffered samples, i.e.,  $\text{train}^{(t)} = \mathcal{D}_k^{(t)} \cup \mathcal{B}_k^{(t-1)}$ . Based on these locally constructed datasets, the global model  $\theta_t$  is optimized by minimizing the aggregated training loss across all clients:  $\min_{\theta} \sum_{k=1}^K \sum_{(x,y) \in \text{train}^{(t)}} \mathcal{L}(f_k(x; \theta), y)$ .

## 4. Methodology

This section presents a federated geometry-aware calibration method for data replay-based FCL, which performs debiasing across both client-level data heterogeneity and task-level data imbalance, enabling more efficient utilization of limited replay samples. The framework of the proposed method FEAT is presented in Figure 2 and Algorithm 1.

### 4.1. Geometric Structure Alignment (GSA)

Despite replay helps retain knowledge from previous tasks, it also introduces continually evolving heterogeneity, making it difficult for clients to maintain a consistent representation space. To alleviate this issue, we leverage a unified geometric prior by adopting the Equiangular Tight Frame (ETF) classifier, which encourages globally consistent class directions across clients [58, 79]. However, under imbalanced class distributions, cross-client alignment of tail classes (previous tasks) remains clearly weaker than that of head classes (current task), as shown in Figure 3. Therefore, the GSA module aims to enforce alignment between the angular structure of the learned features and that induced by their ETF prototypes, enabling a more robust and class-balanced representation space. Specifically, at incremental task  $t$ , we adopt a ETF over the currently observed classes  $\mathcal{C}_t$  with size  $C_t = |\mathcal{C}_t|$ . Let  $W_t = [\mathbf{w}_c]_{c \in \mathcal{C}_t} \in \mathbb{R}^{d \times C_t}$  denote

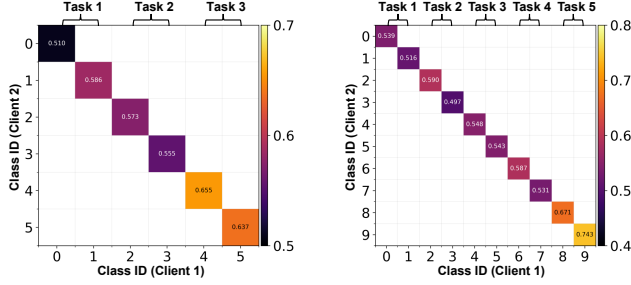


Figure 3. Cross-client feature alignment in an incremental 5-task scenario. At Task 3 (left) and Task 5 (right), cross-client alignment for earlier classes is clearly weaker than for current classes.

the IETF prototypes for these  $C_t$  classes (with  $d \geq C_t$ ),

$$W_t = \sqrt{\frac{C_t}{C_t - 1}} U_t \left( I_{C_t} - \frac{1}{C_t} \mathbf{1}\mathbf{1}^\top \right), \quad (1)$$

where  $U_t \in \mathbb{R}^{d \times C_t}$  satisfies  $U_t^\top U_t = I_{C_t}$ ,  $I_{C_t}$  is the identity, and  $\mathbf{1}$  is the all-ones vector. Importantly, when a new task arrives,  $W_t$  can be reinitialized solely based on the updated class count, without requiring any data. The prototypes have equal  $\ell_2$  norms and identical pair-wise angles:

$$\mathbf{w}_i^\top \mathbf{w}_j = \begin{cases} 1, & i = j, \\ -\frac{1}{C_t - 1}, & i \neq j, \end{cases} \quad i, j \in C_t. \quad (2)$$

Based on the above ETF prototypes, we construct cosine similarity matrices aligned with the current mini-batch. Let  $B$  denote the batch size and  $\mathbf{f}_a$  the feature of the  $a$ -th sample with label  $y_a \in \{1, \dots, C_t\}$ . We define:

$$M_F^{a,b} = \frac{\langle \mathbf{f}_a, \mathbf{f}_b \rangle}{\|\mathbf{f}_a\|_2 \cdot \|\mathbf{f}_b\|_2}, \quad M_P^{a,b} = \frac{\langle \mathbf{w}_{y_a}, \mathbf{w}_{y_b} \rangle}{\|\mathbf{w}_{y_a}\|_2 \cdot \|\mathbf{w}_{y_b}\|_2}, \quad (3)$$

where  $\mathbf{w}_{y_a}$  is the ETF prototype of the class corresponding to  $\mathbf{f}_a$ . Thus, both  $M_F$  and  $M_P$  are  $B \times B$  matrices with the same row/column ordering, which enables direct sample-wise angular distillation. We first normalize each row into a probability distribution with a row-wise softmax:

$$P_F^{a,b} = \text{softmax}_b \left( \frac{M_F^{a,b}}{\tau} \right), \quad P_P^{a,b} = \text{softmax}_b \left( \frac{M_P^{a,b}}{\tau} \right), \quad (4)$$

where  $\tau$  is a temperature parameter. To mitigate class-imbalance (head classes contributing more rows than tail classes), we adopt a class-balanced aggregation. Let  $n_c$  be the number of samples of class  $c$  in the current mini-batch, and  $\mathcal{C}_B = \{c \in \mathcal{C}_t \mid n_c > 0\}$  the set of present classes. The GSA loss is computed by averaging the row-wise KL divergence *per class* and then averaging over classes:

$$\mathcal{L}_{\text{GSA}} = \frac{1}{|\mathcal{C}_B|} \sum_{c \in \mathcal{C}_B} \frac{1}{n_c} \sum_{a: y_a = c} \text{KL}(P_F^{a,\cdot} \parallel P_P^{a,\cdot}). \quad (5)$$

### Algorithm 1 FEAT

- 1: **Initialize:**  $R$ : number of communication rounds;  $K$ : number of clients;  $T$ : number of tasks;  $\mathcal{D}_k$ : local dataset on client  $k$ , including both current and previous tasks;  $\theta_g$ : global model parameters.
- 2: **for** each task  $t = 1$  to  $T$  **do**
- 3:   **for** each communication round  $r = 1$  to  $R$  **do**
- 4:     **for** each client  $k = 1$  to  $K$  **do**
- 5:       Initialize local model parameters:  $\theta_k \leftarrow \theta_g$
- 6:       Select a mini-batch  $\zeta$  from  $\mathcal{D}_k$  and update  $\theta_k$  by Eq. 6 if  $t=1$ , otherwise by Eq. 15.
- 7:       Compute the local statistics  $\{\bar{e}_H^{(T,k)}, \bar{e}_T^{(T,k)}\}$  via Eq. 9.
- 8:     **end for**
- 9:     Aggregate local models via  $\theta_g = \frac{1}{N_C} \sum_k \theta_{(k)}$ .
- 10:    Compute the global statistics  $\{\bar{e}_H^G, \bar{e}_T^G\}$  by Eq. 10.
- 11:   **end for**
- 12:   Replay data in the manner of Re-Fed+ or FedCBDR.
- 13: **end for**
- 14: **Inference:** Employ the EGC module to remove bias via Eqs. 11, 12 and 13.

This class-balanced design strengthens the alignment for tail classes during local client optimization, ensuring they receive sufficient geometric supervision. In addition, we employ a standard classification loss  $\mathcal{L}_{\text{CLS}}$  to ensure the features remain discriminative and well-separated, i.e.,

$$\mathcal{L}_{\text{CLS}} = - \sum_{i=1}^{C_t} y_i \log \left( e^{z_i} / \sum_{j=1}^{C_t} e^{z_j} \right) \quad (6)$$

where  $y_i$  denotes the one-hot encoded ground-truth label for class  $i$ , and  $z_i = \langle \mathbf{f}, \mathbf{w}_i \rangle$  is the similarity score between the feature  $\mathbf{f}$  and the prototype  $\mathbf{w}_i$  of class  $i$ .

### 4.2. Energy-based Geometric Correction (EGC)

Although GSA mitigates cross-client feature misalignment, limited replay leaves a long-tailed distribution that induces a systematic drift of tail features toward head directions. Moreover, we measure this drift with *rank-normalized subspace energies*  $e_H$  and  $e_T$  (Eq. 11) computed via the projectors in Eq. 8. Obviously, a large fraction of tail samples continue to have  $e_H > e_T$ , signaling head-subspace bias (Figure 4). To address this, EGC performs a lightweight inference-time correction.

**Task-wise ETF subspace partitioning.** At incremental task  $t$ , we maintain the ETF prototypes  $W_t = [W_T, W_H] \in \mathbb{R}^{d \times C_t}$  for the currently observed class set  $\mathcal{C}_t$ , where  $W_H = [\mathbf{w}_c]_{c \in \mathcal{C}_H}$  and  $W_T = [\mathbf{w}_c]_{c \in \mathcal{C}_T}$  correspond to head (current) and tail (previous) classes, respectively. Following the simplex-ETF property,

$$\text{rank}(W_H) = |\mathcal{C}_H| - 1, \quad \text{rank}(W_T) = |\mathcal{C}_T| - 1. \quad (7)$$

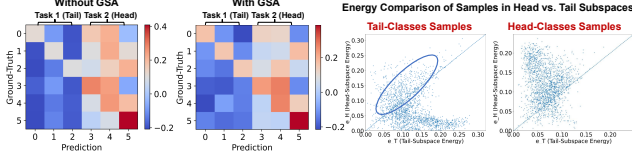


Figure 4. The GSA module reduces the tendency of tail classes to drift toward head-class directions, but tail-class features still show high similarity to head-class prototypes. Moreover, a notable fraction of tail-class samples are aligned with the head-class subspace.

We compute orthogonal projection operators using the Moore–Penrose pseudoinverse ( $\dagger$ ):

$$P_H = W_H(W_H^\top W_H)^\dagger W_H^\top, \quad P_T = W_T(W_T^\top W_T)^\dagger W_T^\top. \quad (8)$$

**Training-time tail priors.** During local client training with exemplar replay, we estimate the typical projection energies of tail samples over the two subspaces. For a normalized feature  $\tilde{f}_a$  from replayed tail data, we maintain Exponential Moving Averages (EMA) of the rank-normalized energies:

$$\begin{aligned} \bar{e}_H^{(T)} &\leftarrow (1 - \rho) \bar{e}_H^{(T)} + \rho \frac{1}{n_T} \sum_{a \in \mathcal{D}_T} \frac{\|P_H \tilde{f}_a\|_2^2}{|\mathcal{C}_H| - 1}, \\ \bar{e}_T^{(T)} &\leftarrow (1 - \rho) \bar{e}_T^{(T)} + \rho \frac{1}{n_T} \sum_{a \in \mathcal{D}_T} \frac{\|P_T \tilde{f}_a\|_2^2}{|\mathcal{C}_T| - 1}. \end{aligned} \quad (9)$$

After each local training round, clients upload only the scalar EMA statistics  $\{\bar{e}_H^{(T,k)}, \bar{e}_T^{(T,k)}\}$  to the server. The server performs a sample-size–weighted aggregation:

$$\bar{e}_H^{(G)} = \frac{1}{\sum_k n_T^{(k)}} \sum_k n_T^{(k)} \bar{e}_H^{(T,k)}, \quad \bar{e}_T^{(G)} = \frac{1}{\sum_k n_T^{(k)}} \sum_k n_T^{(k)} \bar{e}_T^{(T,k)}, \quad (10)$$

where  $k$  indexes clients and  $n_T^{(k)}$  is the number of replayed tail samples used for the local EMA update.

**Inference-time decontamination.** During inference, for any normalized feature  $\tilde{x}$ , we compute rank-normalized projection energies:

$$e_H(\tilde{x}) = \frac{\|P_H \tilde{x}\|_2^2}{|\mathcal{C}_H| - 1}, \quad e_T(\tilde{x}) = \frac{\|P_T \tilde{x}\|_2^2}{|\mathcal{C}_T| - 1}. \quad (11)$$

A confidence gate is derived from the deviation above the global tail prior:

$$g(\tilde{x}) = \max \left\{ \frac{e_H(\tilde{x}) - \bar{e}_H^{(G)}}{e_H(\tilde{x}) + e_T(\tilde{x}) + \varepsilon}, 0 \right\}. \quad (12)$$

The corrected representation suppresses the head-aligned component and enhances the tail-aligned counterpart, followed by  $\ell_2$ -normalization:

$$\tilde{x}' = \tilde{x} - g(\tilde{x})P_H\tilde{x} + g(\tilde{x})P_T\tilde{x}, \quad \tilde{x}' \leftarrow \frac{\tilde{x}'}{\|\tilde{x}'\|_2}. \quad (13)$$

Finally, prediction is made with ETF-based similarities:

$$z_c = (\tilde{x}')^\top \mathbf{w}_c, \quad c \in \mathcal{C}_t. \quad (14)$$

### 4.3. Training Strategy

FEAT adopts a stage-dependent optimization scheme:

- **Initial task ( $t = 1$ ):** Only the classification loss is applied:  $\mathcal{L} = \mathcal{L}_{\text{CLS}}$ .
- **Subsequent tasks ( $t > 1$ ):** Both the classification loss and the GSA loss are employed:

$$\mathcal{L} = \mathcal{L}_{\text{CLS}} + \lambda \cdot \mathcal{L}_{\text{GSA}}, \quad (15)$$

where  $\lambda$  is a balancing hyperparameter.

## 5. Experiments

### 5.1. Experiment Settings

**Datasets.** Following prior work [18, 29, 40, 76], we use three widely used benchmarks: CIFAR10/100 [23], and TinyImageNet-Subset [24]. To emulate non-iid client environments, we partition local datasets using a Dirichlet distribution. Specifically, we consider two client settings (5 and 10 clients) for all datasets, and two task granularities per dataset: CIFAR10 is split into 3 and 5 tasks with  $\beta = \{0.5, 1.0\}$ , while CIFAR100 and TinyImageNet-Subset are split into 5 and 10 tasks with  $\beta = \{0.1, 0.5\}$ .

**Evaluation Metric.** Following prior work [4, 29, 36, 56], we report Top-1 Accuracy, defined as  $N_{\text{correct}}/N_{\text{total}}$ , where  $N_{\text{correct}}$  and  $N_{\text{total}}$  denote the numbers of correct predictions and samples, respectively.

**Implementation Details.** We use ResNet-18 as the backbone for all datasets, and the classifier expands as new classes arrive. Each client trains for 2 local epochs per communication round with batch size 128. We run 100 communication rounds per task using SGD with learning rate 0.04 and weight decay  $1 \times 10^{-5}$ . The exemplar budget per task is 450/300 for CIFAR10 (3/5 tasks), 1000/500 for CIFAR100 (5/10 tasks), and 500/250 for TinyImageNet-Subset (5/10 tasks). For GSA, the temperature  $\tau$  is chosen from  $\{0.07, 0.5\}$  and the weight  $\lambda$  from  $\{0.05, 0.1, 0.5\}$ . For EGC, the EMA decay  $\rho$  is selected from  $\{0.5, 0.7, 0.9\}$ . We set the numerical stability term  $\varepsilon$  to  $10^{-8}$ .

### 5.2. Performance Comparison

To validate the effectiveness of the method FEAT, we compare it with seven state-of-the-art baselines, including FedEWC [19], FedLwF [33], TARGET [76], LANDER [64], FedSSI [32], Re-Fed+ [30], and FedCBDR [55]. To explore the adaptability of FEAT in different frameworks, we construct two hybrid versions: FEAT<sub>R</sub> and FEAT<sub>F</sub>, which correspond to incorporating FEAT into Re-Fed+ and FedCBDR, respectively. Table 1 summarizes the outcomes:

- FEAT<sub>F</sub> consistently achieves the highest Top-1 accuracy across all cases, including various levels of heterogeneity, and different task partitions, demonstrating the robustness of the proposed de-biased framework.

Table 1. Performance comparison between FEAT and baselines across three datasets under varying levels of data heterogeneity ( $\beta$ ) and different numbers of clients ( $K$ ). Results are averaged over three random seeds, with standard deviations reported.

| Method            | CIFAR10                |                        |                        |                        | CIFAR100               |                        |                        |                        | TinyImageNet-Subset    |                        |                        |                        |
|-------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|
|                   | 3 Tasks                |                        | 5 Tasks                |                        | 5 Tasks                |                        | 10 Tasks               |                        | 5 Tasks                |                        | 10 Tasks               |                        |
|                   | $\beta=0.5$            | $\beta=1.0$            | $\beta=0.5$            | $\beta=1.0$            | $\beta=0.1$            | $\beta=0.5$            | $\beta=0.1$            | $\beta=0.5$            | $\beta=0.1$            | $\beta=0.5$            | $\beta=0.1$            | $\beta=0.5$            |
| $K = 5$ Clients   |                        |                        |                        |                        |                        |                        |                        |                        |                        |                        |                        |                        |
| FedEWC            | 40.48 $\pm$ 1.6        | 43.11 $\pm$ 3.1        | 22.16 $\pm$ 2.1        | 29.38 $\pm$ 2.3        | 18.66 $\pm$ 2.6        | 22.64 $\pm$ 4.4        | 10.12 $\pm$ 0.6        | 11.34 $\pm$ 0.6        | 11.52 $\pm$ 1.4        | 14.27 $\pm$ 2.2        | 7.92 $\pm$ 0.7         | 8.85 $\pm$ 0.9         |
| FedLwF            | 56.84 $\pm$ 1.2        | 58.97 $\pm$ 3.1        | 39.21 $\pm$ 1.9        | 52.32 $\pm$ 2.7        | 34.45 $\pm$ 2.0        | 36.34 $\pm$ 2.6        | 19.36 $\pm$ 1.4        | 24.91 $\pm$ 0.9        | 16.78 $\pm$ 0.8        | 21.87 $\pm$ 0.7        | 12.16 $\pm$ 0.6        | 13.11 $\pm$ 0.3        |
| TARGET            | 43.53 $\pm$ 3.8        | 54.87 $\pm$ 4.2        | 36.02 $\pm$ 2.4        | 48.91 $\pm$ 1.6        | 30.62 $\pm$ 2.9        | 32.89 $\pm$ 3.7        | 13.04 $\pm$ 1.1        | 20.57 $\pm$ 0.7        | 17.69 $\pm$ 1.5        | 21.11 $\pm$ 1.3        | 11.98 $\pm$ 0.5        | 15.28 $\pm$ 0.7        |
| LANDER            | 54.41 $\pm$ 2.7        | 59.88 $\pm$ 1.9        | 39.63 $\pm$ 2.0        | 57.55 $\pm$ 2.7        | 43.59 $\pm$ 2.9        | 48.39 $\pm$ 3.3        | 27.36 $\pm$ 1.1        | 32.64 $\pm$ 2.8        | 11.91 $\pm$ 1.2        | 24.77 $\pm$ 2.5        | 12.82 $\pm$ 0.9        | 14.93 $\pm$ 1.2        |
| Re-Fed+           | 52.95 $\pm$ 4.1        | 61.64 $\pm$ 3.9        | 54.15 $\pm$ 2.7        | 58.94 $\pm$ 3.2        | 31.92 $\pm$ 4.3        | 37.68 $\pm$ 2.3        | 28.64 $\pm$ 0.8        | 38.62 $\pm$ 1.7        | 22.89 $\pm$ 0.4        | 26.07 $\pm$ 0.5        | 16.71 $\pm$ 1.3        | 21.02 $\pm$ 1.4        |
| FedSSI            | 51.44 $\pm$ 3.1        | 56.33 $\pm$ 3.1        | 52.26 $\pm$ 2.5        | 56.49 $\pm$ 2.5        | 31.27 $\pm$ 1.3        | 37.63 $\pm$ 1.3        | 26.71 $\pm$ 1.3        | 31.16 $\pm$ 1.3        | 18.56 $\pm$ 1.1        | 21.37 $\pm$ 1.1        | 14.41 $\pm$ 0.9        | 17.22 $\pm$ 0.9        |
| FedCBDR           | 63.32 $\pm$ 1.6        | 65.88 $\pm$ 2.1        | 61.77 $\pm$ 1.5        | 64.79 $\pm$ 1.1        | 45.84 $\pm$ 2.0        | 50.33 $\pm$ 2.3        | 44.52 $\pm$ 1.7        | 45.96 $\pm$ 1.3        | 25.22 $\pm$ 1.1        | 26.38 $\pm$ 1.6        | 19.03 $\pm$ 0.7        | 21.98 $\pm$ 0.5        |
| FEAT <sub>R</sub> | 60.18 $\pm$ 1.4        | 70.28 $\pm$ 1.6        | 60.38 $\pm$ 1.5        | 64.62 $\pm$ 1.6        | 37.14 $\pm$ 1.8        | 42.31 $\pm$ 1.4        | 32.91 $\pm$ 1.3        | 43.34 $\pm$ 1.9        | 23.31 $\pm$ 1.5        | 27.36 $\pm$ 1.6        | 19.26 $\pm$ 1.7        | 22.74 $\pm$ 1.1        |
| FEAT <sub>F</sub> | <b>72.67</b> $\pm$ 1.5 | <b>74.21</b> $\pm$ 2.6 | <b>70.19</b> $\pm$ 1.8 | <b>73.63</b> $\pm$ 2.0 | <b>50.14</b> $\pm$ 1.2 | <b>53.31</b> $\pm$ 0.9 | <b>48.16</b> $\pm$ 1.6 | <b>49.18</b> $\pm$ 0.8 | <b>25.88</b> $\pm$ 1.6 | <b>29.31</b> $\pm$ 0.6 | <b>19.41</b> $\pm$ 1.4 | <b>23.45</b> $\pm$ 0.4 |
| $K = 10$ Clients  |                        |                        |                        |                        |                        |                        |                        |                        |                        |                        |                        |                        |
| FedEWC            | 35.11 $\pm$ 3.5        | 43.32 $\pm$ 2.3        | 22.13 $\pm$ 2.0        | 24.97 $\pm$ 2.8        | 16.95 $\pm$ 0.4        | 17.21 $\pm$ 0.4        | 5.41 $\pm$ 1.2         | 12.28 $\pm$ 0.6        | 10.73 $\pm$ 1.3        | 14.37 $\pm$ 2.4        | 5.42 $\pm$ 1.1         | 6.47 $\pm$ 0.6         |
| FedLwF            | 49.72 $\pm$ 2.0        | 47.51 $\pm$ 3.1        | 41.63 $\pm$ 2.4        | 48.35 $\pm$ 3.9        | 28.76 $\pm$ 1.7        | 36.35 $\pm$ 1.9        | 12.55 $\pm$ 1.3        | 18.91 $\pm$ 0.8        | 17.34 $\pm$ 0.6        | 21.79 $\pm$ 1.7        | 5.28 $\pm$ 1.1         | 11.04 $\pm$ 1.0        |
| TARGET            | 39.20 $\pm$ 2.6        | 40.18 $\pm$ 1.2        | 22.45 $\pm$ 1.9        | 27.79 $\pm$ 1.5        | 17.89 $\pm$ 1.4        | 22.67 $\pm$ 1.7        | 12.72 $\pm$ 1.2        | 15.98 $\pm$ 1.1        | 19.12 $\pm$ 1.6        | 25.88 $\pm$ 0.5        | 6.23 $\pm$ 1.4         | 10.74 $\pm$ 1.0        |
| LANDER            | 42.38 $\pm$ 3.0        | 44.89 $\pm$ 1.9        | 26.63 $\pm$ 4.2        | 31.47 $\pm$ 2.8        | 29.92 $\pm$ 2.5        | 42.83 $\pm$ 2.7        | 11.60 $\pm$ 1.4        | 26.49 $\pm$ 1.5        | 15.68 $\pm$ 2.7        | 21.90 $\pm$ 2.1        | 10.38 $\pm$ 1.4        | 10.68 $\pm$ 1.2        |
| Re-Fed+           | 44.92 $\pm$ 2.2        | 52.97 $\pm$ 3.9        | 37.84 $\pm$ 1.8        | 40.16 $\pm$ 1.6        | 32.77 $\pm$ 1.9        | 35.84 $\pm$ 0.4        | 28.67 $\pm$ 2.2        | 32.74 $\pm$ 0.6        | 20.26 $\pm$ 1.7        | 25.67 $\pm$ 1.8        | 15.46 $\pm$ 1.3        | 19.32 $\pm$ 1.6        |
| FedSSI            | 41.93 $\pm$ 3.1        | 48.27 $\pm$ 3.1        | 38.22 $\pm$ 2.5        | 40.45 $\pm$ 2.5        | 30.56 $\pm$ 1.3        | 35.72 $\pm$ 1.3        | 27.54 $\pm$ 1.3        | 33.16 $\pm$ 1.3        | 16.55 $\pm$ 1.1        | 20.27 $\pm$ 1.1        | 11.32 $\pm$ 0.9        | 14.49 $\pm$ 0.9        |
| FedCBDR           | 59.27 $\pm$ 1.9        | 61.75 $\pm$ 1.1        | 52.00 $\pm$ 2.5        | 59.91 $\pm$ 1.0        | 41.96 $\pm$ 1.5        | 47.54 $\pm$ 1.9        | 37.11 $\pm$ 1.0        | 43.59 $\pm$ 2.4        | 22.95 $\pm$ 0.5        | 25.45 $\pm$ 1.9        | 17.61 $\pm$ 1.1        | 19.54 $\pm$ 1.0        |
| FEAT <sub>R</sub> | 54.62 $\pm$ 1.5        | 62.34 $\pm$ 2.1        | 46.59 $\pm$ 1.3        | 49.53 $\pm$ 1.1        | 35.75 $\pm$ 1.8        | 40.43 $\pm$ 1.6        | 32.38 $\pm$ 1.1        | 37.79 $\pm$ 1.4        | 23.11 $\pm$ 1.9        | 27.15 $\pm$ 1.1        | 18.16 $\pm$ 1.0        | 20.79 $\pm$ 0.9        |
| FEAT <sub>F</sub> | <b>70.63</b> $\pm$ 1.7 | <b>72.89</b> $\pm$ 1.9 | <b>65.48</b> $\pm$ 2.1 | <b>69.76</b> $\pm$ 1.4 | <b>45.72</b> $\pm$ 1.6 | <b>50.36</b> $\pm$ 1.4 | <b>39.36</b> $\pm$ 1.8 | <b>47.81</b> $\pm$ 2.1 | <b>25.15</b> $\pm$ 0.9 | <b>28.85</b> $\pm$ 1.0 | <b>20.49</b> $\pm$ 1.6 | <b>23.17</b> $\pm$ 1.1 |

Table 2. Ablation studies are conducted on CIFAR-10 with 3 tasks under  $\beta \in \{0.5, 1.0\}$  and on CIFAR-100 with 5 tasks under  $\beta \in \{0.1, 0.5\}$ . The number of clients is 5.

| Method       | CIFAR10                |                        | CIFAR100               |                        |
|--------------|------------------------|------------------------|------------------------|------------------------|
|              | $\beta=0.5$            | $\beta=1.0$            | $\beta=0.1$            | $\beta=0.5$            |
| FedCBDR      | 63.32 $\pm$ 1.6        | 65.88 $\pm$ 2.1        | 45.84 $\pm$ 2.0        | 50.33 $\pm$ 2.3        |
| +ETF         | 62.17 $\pm$ 1.6        | 63.66 $\pm$ 2.0        | 44.58 $\pm$ 1.9        | 48.69 $\pm$ 1.2        |
| +ETF+GSA     | 68.54 $\pm$ 1.6        | 70.42 $\pm$ 1.7        | 47.77 $\pm$ 1.2        | 52.23 $\pm$ 1.6        |
| +ETF+EGC     | 69.12 $\pm$ 1.8        | 70.83 $\pm$ 2.1        | 47.16 $\pm$ 1.4        | 51.72 $\pm$ 1.5        |
| +ETF+GSA+EGC | <b>72.67</b> $\pm$ 1.5 | <b>74.21</b> $\pm$ 2.6 | <b>50.14</b> $\pm$ 1.2 | <b>53.31</b> $\pm$ 0.9 |

- Both FEAT<sub>R</sub> and FEAT<sub>F</sub> achieve performance improvements over their corresponding baselines, which validates the importance of the core idea behind our method, which also highlights the plug-and-play nature.
- As the setting becomes more challenging, with stronger heterogeneity and more clients, FEAT's advantage over other methods does not shrink. This indicates that its gains are not tied to easy cases but remain stable when scaled to more complex and realistic FCL scenarios.
- Although generative replay method LANDER achieves competitive performance compared to exemplar replay-based methods in some cases, they rely heavily on a large number of generated samples, which poses potential limitations when deployed in large-scale environments.

### 5.3. Ablation Study

This section aims to investigate the contributions of key components, including the Geometric Structure Alignment (GSA) module and the Energy-based Geometric Correction

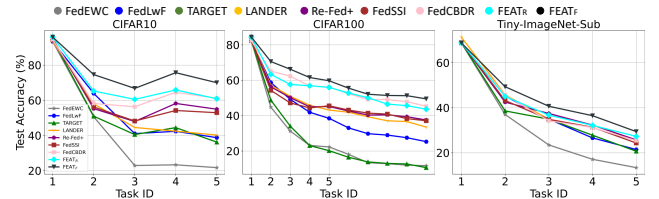


Figure 5. Evaluation on CIFAR-10 (5 tasks), CIFAR-100 (10 tasks), and TinyImageNet-Subset (5 tasks) with 5 clients and heterogeneity  $\beta = 0.5$ . FEAT<sub>F</sub> and FEAT<sub>R</sub> consistently outperform their baselines on all cases.

(EGC) module. The results are presented in Table 3:

- Relying solely on the ETF classifier may not yield performance gains, as severe data imbalance substantially undermines cross-source representation alignment.
- The integration of GSA consistently enhances performance under all datasets and task settings. This improvement stems from GSA's capacity to maintain inter-class geometric consistency, which helps reduce the representation gap among heterogeneous clients.
- By addressing task-level imbalance during inference, the EGC module stabilizes predictions, improves performance, and normalizes decision boundaries, reducing head-task dominance.

### 5.4. Forgetting Curve across Incremental Tasks

This section evaluates FEAT and baselines on three datasets with heterogeneity  $\beta = 0.5$ . As shown in Figure 5, FEAT<sub>F</sub>

and  $\text{FEAT}_R$  consistently outperform their baselines across all increments by mitigating client heterogeneity and task-level imbalance. Moreover,  $\text{FEAT}_F$  attains the highest initial accuracy and shows the slowest, most stable decline, indicating stronger long-term resistance to forgetting.

### 5.5. Communication Cost Analysis

$\text{FEAT}$  retains the same training rounds and aggregation protocol as the baseline, adding only negligible overhead. Per round, each client uploads two scalar EMA statistics ( $\bar{e}_H$  and  $\bar{e}_T$ ), which is trivial compared to model parameters, and the server performs weighted aggregation. GSA runs locally and requires no extra transmission, while EGC is applied at inference and introduces no training-time communication.

### 5.6. Analyzing the Influence of Replay Data Volume

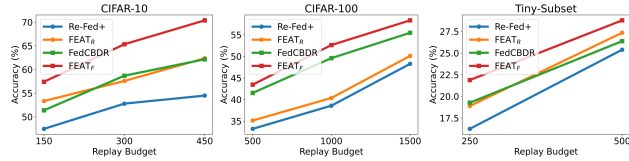


Figure 6. Impact of per-task memory capacity  $M$  on model performance across different datasets, where  $\text{FEAT}_F$  shows the best overall performance.

This section evaluates  $\text{FEAT}_F$  and  $\text{FEAT}_R$  against baselines under different replay budgets on three datasets (5-task split), with replay sizes set to  $\{100, 150, 300\}$  (CIFAR-10),  $\{125, 250, 500\}$  (CIFAR-100), and  $\{250, 500\}$  (TinyImageNet-Subset). As shown in Figure 6,  $\text{FEAT}_F$  and  $\text{FEAT}_R$  consistently achieve performance improvements over their respective baselines across all settings. These results confirm the effectiveness of our design and emphasize the importance of tackling client heterogeneity and task-level imbalance in FCL. Compared to  $\text{FEAT}_R$ ,  $\text{FEAT}_F$  shows more significant performance improvements as the replay budget increases. This can be attributed to the class-wise balanced replay inherited from FedCBDR.

### 5.7. Sensitivity Study of Hyper-parameters

We assess hyperparameter sensitivity on CIFAR-10 (5 tasks,  $\beta=0.5$ , 5 clients), varying  $\lambda \in \{1, 5, 10\}$ ,  $\rho \in \{0.5, 0.7, 0.9\}$ , and  $\tau \in \{0.1, 0.5\}$ . Figure 7 shows stable accuracy with only slight fluctuations, indicating robustness to hyperparameter choices. In general, a smaller  $\lambda$  better exploits geometric structure knowledge, a larger  $\rho$  yields more stable debiased learning, and  $\tau$  has a minor effect. Notably, these configurations consistently outperform its baseline.

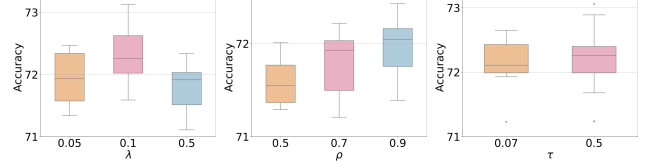


Figure 7. Sensitivity analysis of hyperparameters  $\lambda$ ,  $\rho$ , and  $\tau$ , adjusted over the ranges  $\{0.05, 0.1, 0.5\}$ ,  $\{0.5, 0.7, 0.9\}$ , and  $\{0.07, 0.5\}$ , respectively. For each parameter, the default value is fixed while the other two are varied. It achieves robust performance under a wide range of these parameters.

Table 3. Performance analysis of alternative optimization strategies, including MOON, FedRCL, CLIP2FL, FedFSA, and  $\text{FEAT}$ , applied on top of the FedCBDR baseline. Results are reported as the mean over three runs on three datasets.

| Methods         | CIFAR10       |               | CIFAR100      |               | Tiny-Subset   |               |
|-----------------|---------------|---------------|---------------|---------------|---------------|---------------|
|                 | $\beta = 0.5$ | $\beta = 1.0$ | $\beta = 0.1$ | $\beta = 0.5$ | $\beta = 0.1$ | $\beta = 0.5$ |
| FedCBDR         | 61.77         | 64.79         | 45.84         | 50.33         | 25.22         | 26.38         |
| + MOON          | 62.56         | 66.32         | 45.14         | 49.32         | 24.73         | 26.44         |
| + FedRCL        | 58.74         | 60.21         | 42.19         | 46.46         | 20.27         | 22.46         |
| + CLIP2FL       | 62.33         | 67.97         | 46.19         | 51.11         | 24.18         | 26.67         |
| + FedFSA        | 64.42         | 68.32         | 46.88         | 51.38         | 24.61         | 26.33         |
| + $\text{FEAT}$ | <b>70.19</b>  | <b>73.63</b>  | <b>50.14</b>  | <b>53.31</b>  | <b>25.88</b>  | <b>29.31</b>  |

### 5.8. Performance Analysis of Alternative Strategies

This section evaluates several commonly used enhancements on top of the same baseline (FedCBDR), including MOON [25], FedRCL [59], CLIP2FL [61], FedFSA [54], and  $\text{FEAT}$ . Table 3 reports the mean accuracy over three runs on three datasets under a 5-task setting with 5 clients. Across all cases,  $\text{FEAT}$  achieves the highest accuracy when added to FedCBDR. The gains are especially clear in the more heterogeneous cases (smaller  $\beta$ ), where other strategies provide only moderate or inconsistent improvements, while  $\text{FEAT}$  delivers a consistent boost. This suggests that  $\text{FEAT}$  is effective at handling spatiotemporal drift and cross-client imbalance, rather than only refining local features. In contrast, methods such as MOON and FedFSA offer incremental benefits but do not close the gap.

### 5.9. Final Performance Comparison Across Tasks

Table 4. Per-task and average accuracy (%) of different methods on CIFAR10 and CIFAR100.

|             | CIFAR10      |              |              | CIFAR100     |              |              |              |              |
|-------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
|             | Task 1       | Task 2       | Task 3       | Task 1       | Task 2       | Task 3       | Task 4       | Task 5       |
| FedCBDR     | 43.37        | 33.90        | 96.80        | 43.80        | 34.25        | 32.85        | 45.00        | 83.55        |
| + GSA       | 47.47        | 51.60        | <b>97.92</b> | 48.75        | 36.25        | 36.10        | 52.15        | <b>84.40</b> |
| + EGC       | 53.53        | 53.57        | 97.65        | 50.15        | 38.70        | 39.10        | 51.40        | 82.45        |
| + EGC + GSA | <b>60.83</b> | <b>69.97</b> | 94.28        | <b>50.70</b> | <b>40.65</b> | <b>39.70</b> | <b>53.85</b> | 81.95        |

This section evaluates the final model on each task for CIFAR-10 (3 tasks) and CIFAR-100 (5 tasks) under non-

IID heterogeneity  $\beta=0.5$  with 5 clients. As shown in Table 4, adding GSA and EGC markedly improves final performance on earlier tasks, with the best results when combined. Although accuracy on the last task slightly decreases, it remains competitive, indicating stronger resistance to forgetting and more stable task-wise generalization.

### 5.10. Evaluation under Different Local Epochs

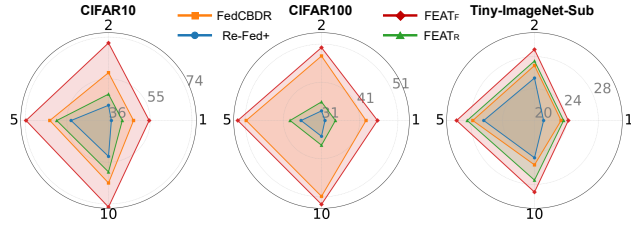


Figure 8. Performance of Re-Fed+ and FedCBDR, with and without FEAT, across local epochs  $\{1, 2, 5, 10\}$  on three datasets. And FEAT brings performance gains to all baselines.

This section evaluates the performance of Re-Fed+ and FedCBDR, with and without FEAT, under varying local epochs  $\{1, 2, 5, 10\}$ . As shown in Figure 8, the versions integrated with FEAT consistently outperform their baselines in all settings. However, the performance improvement does not exhibit a strictly increasing trend with more local updates, indicating that the advantages of GSA and EGC may saturate or become unstable when local training dominates the optimization process. This observation provides valuable insight for future work.

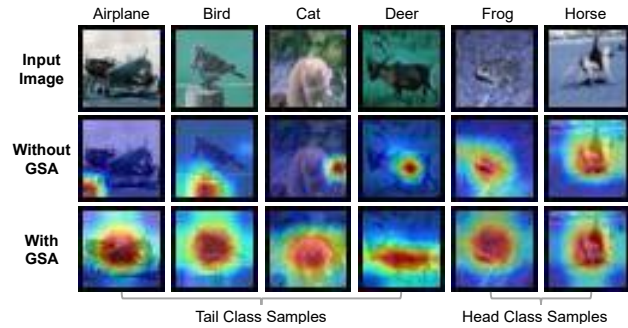


Figure 9. Visualization of visual attention. The GSA module corrects the attention bias of tail classes and strengthens the confidence of attention on head classes.

### 5.11. Comparative Analysis of Visual Attention

Figure 4 shows that GSA reduces tail-to-head drift. This section provides complementary qualitative evidence by visualizing attention maps [26, 44, 62, 68, 78, 80] for tail and head samples. As shown in Figure 9, without GSA, tail classes show diffuse or off-target responses, indicating that

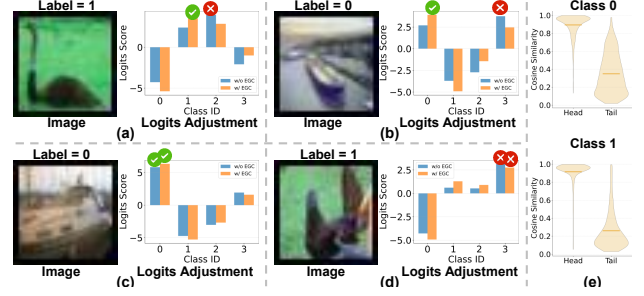


Figure 10. Geometric correction effect of the EGC module on the CIFAR-10 (first two tasks). (a–b) EGC corrects the prediction bias; (c) EGC corrects the prediction bias; (d) it struggles when features deviate too far from the ground truth; (e) The removed component aligns more with the head subspace than the tail subspace for both Class 0 and Class 1.

the model is not attending reliably to the true object. With GSA, attention becomes tighter and aligned with the object region, correcting this bias. For head classes, GSA further sharpens and amplifies the focus around the object, suggesting increased confidence rather than degradation. These results demonstrate the positive impact of the GSA module.

### 5.12. Logits-Level Behavior Analysis

This section further analyzes how the EGC module adjusts logits at inference time. Figure 10 shows four representative CIFAR-10 samples from the first two tasks. Obviously, cases (a) and (b) show that applying EGC amplifies the logit of the target class and suppresses the spurious head-class logits, leading to a correct prediction. Case (c) shows that EGC further enlarges the gap between the ground-truth class and the other classes. Case (d) illustrates a failure case, where the feature has drifted far from the true class manifold; in this situation, the correction is not sufficient to fully recover the correct label, although it still narrows the gap between the ground-truth class and the current top-1 class. In case (e), for tail-class samples, the removed component aligns strongly with the head-class subspace and weakly with the tail subspace, confirming that our correction targets head-attraction drift rather than a uniform shift.

## 6. Conclusions and Future Work

To address the dual challenges of inter-client heterogeneity and task-level data imbalance, we propose FEAT. It consists of two key components: (1) the GSA module that enforces structural consistency between local features and globally shared prototypes, thus promoting inter-client alignment; and (2) the EGC that removes projection bias during inference, improving robustness under class-imbalanced data distributions. Experimental results show that FEAT consistently outperforms existing methods by effectively mitigating inter-client and inter-task bias.

## Acknowledgment

This work is supported in part by the Key Research and Development Program of Shandong Province-Innovation Capability Enhancement Program for Technology-based Small and Medium-sized Enterprises (Grant No. 2024TSGC0667); the Shandong Key Laboratory of Foundational Software (Project No. 11150004040955); the Engineering Research Center of Digital Media Technology, Ministry of Education, China; and the China Scholarship Council.

## References

- [1] Sara Babakniya, Zalan Fabian, Chaoyang He, Mahdi Soltanolkotabi, and Salman Avestimehr. Don't memorize; mimic the past: Federated class incremental learning without episodic memory. *arXiv preprint arXiv:2307.00497*, 2023. 2
- [2] Tanujit Chakraborty, Ujjwal Reddy KS, Shraddha M Naik, et al. Ten years of generative adversarial nets (gans): a survey of the state-of-the-art. *Machine Learning: Science and Technology*, 5(1):011001, 2024. 2
- [3] Yuanlu Chen, Alysia Ziyang Tan, Siwei Feng, Han Yu, Tao Deng, Libang Zhao, and Feng Wu. General federated class-incremental learning with lightweight generative replay. *IEEE Internet of Things Journal*, 2024. 2
- [4] Zitan Chen et al. Class-level structural relation modeling and smoothing for visual representation learning. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 2964–2972, 2023. 5
- [5] Jiahua Dong, Lixu Wang, Zhen Fang, Gan Sun, et al. Federated class-incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10164–10173, 2022. 2
- [6] Jiahua Dong, Duzhen Zhang, Yang Cong, Wei Cong, Henghui Ding, and Dengxin Dai. Federated incremental semantic segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3934–3943, 2023. 3
- [7] Jiahua Dong, Wenqi Liang, Yang Cong, Gan Sun, Lixu Wang, Henghui Ding, Yulun Zhang, and Luc Van Gool. Learning from each other: Generalized federated incremental semantic segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2026. 3
- [8] Ziqing Fan, Jiangchao Yao, Bo Han, Ya Zhang, Yanfeng Wang, et al. Federated learning with bilateral curation for partially class-disjoint data. *Advances in Neural Information Processing Systems*, 36:32006–32019, 2023. 3
- [9] Wei Feng, Sijin Zhou, Yiwen Jiang, and Zongyuan Ge. Prism: Progressive robust learning for open-world continual category discovery. In *The Fourteenth International Conference on Learning Representations*. 3
- [10] Wei Feng, Sijin Zhou, Yiwen Jiang, et al. Neighbor-guided unbiased framework for generalized category discovery in medical image classification. *IEEE Journal of Biomedical and Health Informatics*, 2025. 3
- [11] Lele Fu, Sheng Huang, Yanyi Lai, Tianchi Liao, et al. Beyond federated prototype learning: Learnable semantic anchors with hyperspherical contrast for domain-skewed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 16648–16656, 2025. 1
- [12] Zijian Gao, Kele Xu, et al. Rethinking obscured suboptimality in analytic learning for exemplar-free class-incremental learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(10):1123–1136, 2025. 2
- [13] Laurent Girin, Simon Leglaive, Xiaoyu Bie, Julien Diard, Thomas Hueber, and Xavier Alameda-Pineda. Dynamical variational autoencoders: A comprehensive review. *arXiv preprint arXiv:2008.12595*, 2020. 2
- [14] Ian J Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014. 2
- [15] Ming Hu, Yue Cao, Anran Li, Zhiming Li, et al. Fedmut: Generalized federated learning via stochastic mutation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 12528–12537, 2024. 1
- [16] Ming Hu, Zhihao Yue, Xiaofei Xie, Cheng Chen, et al. Is aggregation the only choice? federated learning via layer-wise model recombination. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1096–1107, 2024. 1
- [17] Ming Hu, Peiheng Zhou, Zhihao Yue, Zhiwei Ling, Yihao Huang, Anran Li, et al. Fedcross: Towards accurate federated learning via multi-model cross-aggregation. In *IEEE International Conference on Data Engineering (ICDE)*, pages 2137–2150. IEEE, 2024. 1
- [18] Sheng Huang, Lele Fu, Fanghua Ye, Tianchi Liao, et al. Soft-consensual federated learning for data heterogeneity via multiple paths. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 5
- [19] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017. 5
- [20] Zhiqiang Kou, Jing Wang, Jiawei Tang, et al. Exploiting multi-label correlation in label distribution learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence*, pages 4326–4334, 2024. 3
- [21] Zhiqiang Kou, Si Qin, Hailin Wang, Jing Wang, Mingkun Xie, et al. Label distribution learning with biased annotations assisted by multi-label learning. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, 2025.
- [22] Zhiqiang Kou, Yucheng Xie, Hailin Wang, et al. Rankmatch: A novel approach to semi-supervised label distribution learning leveraging rank correlation between labels. In *Proceedings of the 39th Conference on Neural Information Processing Systems (NeurIPS 2025)*, 2025. 3
- [23] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 5

- [24] Ya Le and Xuan Yang. Tiny imagenet visual recognition challenge. *CS 231N*, 7(7):3, 2015. 5
- [25] Qinyin Li, Bingsheng He, and Dawn Song. Model-contrastive federated learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 10713–10722, 2021. 7
- [26] Wenhao Li, Qiangchang Wang, Xianjing Meng, et al. Vt-fsl: Bridging vision and text with llms for few-shot learning. *arXiv preprint arXiv:2509.25033*, 2025. 8
- [27] Wenhao Li, Xianjing Meng, Qiangchang Wang, Zhongyi Han, et al. Dvla-rl: Dual-level vision-language alignment with reinforcement learning gating for few-shot learning. *arXiv preprint arXiv:2602.00795*, 2026. 3
- [28] Xiangxian Li, Yuze Zheng, et al. Cross-modal learning using privileged information for long-tailed image classification. *Computational Visual Media*, 10(5):981–992, 2024. 3
- [29] Yichen Li, Qunwei Li, Haozhao Wang, Ruixuan Li, Wenliang Zhong, and Guannan Zhang. Towards efficient replay in federated incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12820–12829, 2024. 2, 3, 5
- [30] Yichen Li, Haozhao Wang, Yining Qi, Wei Liu, and Ruixuan Li. Re-fed+: A better replay strategy for federated incremental learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2, 5
- [31] Yichen Li, Haozhao Wang, Wenchao Xu, Tianzhe Xiao, Hong Liu, et al. Unleashing the power of continual learning on non-centralized devices: A survey. *IEEE Communications Surveys & Tutorials*, 2025. 1
- [32] Yichen Li, Yuying Wang, Haozhao Wang, Yining Qi, Tianzhe Xiao, and Ruixuan Li. Rehearsal-free continual federated learning with synergistic synaptic intelligence. *International Conference on Machine Learning*, 2025. 5
- [33] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017. 5
- [34] Zexi Li, Xinyi Shang, Rui He, Tao Lin, and Chao Wu. No fear of classifier biases: Neural collapse inspired federated learning with synthetic and fixed classifier. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 5319–5329, 2023. 3
- [35] Jinglin Liang, Jin Zhong, Hanlin Gu, et al. Diffusion-driven data replay: A novel approach to combat forgetting in federated class continual learning. In *European Conference on Computer Vision*, pages 303–319. Springer, 2024. 2
- [36] Tianchi Liao, Binghui Xie, Lele Fu, Sheng Huang, Bowen Deng, Chuan Chen, and Zibin Zheng. Federated domain generalization with decision insight matrix. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 5689–5697, 2025. 5
- [37] Junkang Liu, Fanhua Shang, Yuanyuan Liu, Hongying Liu, et al. Fedbcgd: Communication-efficient accelerated block coordinate gradient descent for federated learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 2955–2963, 2024. 1
- [38] Junkang Liu, Yuanyuan Liu, Fanhua Shang, et al. Improving generalization in federated learning with highly heterogeneous data via momentum-based stochastic controlled weight averaging. In *Forty-second International Conference on Machine Learning*, 2025. 1
- [39] Junkang Liu, Fanhua Shang, Kewen Zhu, et al. Fedadamw: A communication-efficient optimizer with convergence and generalization guarantees for federated large models. *arXiv preprint arXiv:2510.27486*, 2025. 2
- [40] Tianhan Liu et al. Cross-training with prototypical distillation for improving the generalization of federated learning. In *2023 IEEE International Conference on Multimedia and Expo (ICME)*, pages 648–653. IEEE, 2023. 5
- [41] Jianfeng Lu and Stefan Steinerberger. Neural collapse under cross-entropy loss. *Applied and Computational Harmonic Analysis*, 59:224–241, 2022. 3
- [42] Yanbiao Ma, Wei Dai, Wenke Huang, and Jiayi Chen. Geometric knowledge-guided localized global distribution alignment for federated learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 20958–20968, 2025. 1
- [43] Lei Meng, Zhuang Qi, Lei Wu, Xiaoyu Du, et al. Improving global generalization and local personalization for federated learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1):76–87, 2024. 1
- [44] Lei Meng, Xiangxian Li, Xiaoshuo Yan, Haokai Ma, et al. Causal inference over visual-semantic-aligned graph for image classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 19449–19457, 2025. 8
- [45] Hojjat Navidan, Parisa Fard Moshiri, Mohammad Nabati, Reza Shahbazian, Seyed Ali Ghorashi, Vahid Shah-Mansouri, and David Windridge. Generative adversarial networks (gans) in networking: A comprehensive survey & evaluation. *Computer Networks*, 194:108149, 2021. 2
- [46] Thinh Nguyen, Khoa D Doan, et al. Overcoming catastrophic forgetting in federated class-incremental learning via federated global twin generator. *arXiv preprint arXiv:2407.11078*, 2024. 2
- [47] Milad Khademi Nori et al. Federated class-incremental learning: A hybrid approach using latent exemplars and data-free techniques to address local and global forgetting. *arXiv preprint arXiv:2501.15356*, 2025. 2
- [48] Vardan Papayan, XY Han, and David L Donoho. Prevalence of neural collapse during the terminal phase of deep learning training. *Proceedings of the National Academy of Sciences*, 117(40):24652–24663, 2020. 3
- [49] Lucas Pinheiro Cinelli, Matheus Araújo Marins, Eduardo Antônio Barros da Silva, and Sérgio Lima Netto. Variational autoencoder. In *Variational methods for machine learning with applications to deep networks*, pages 111–149. Springer, 2021. 2
- [50] Daiqing Qi, Handong Zhao, and Sheng Li. Better generative replay for continual federated learning. *arXiv preprint arXiv:2302.13001*, 2023. 2
- [51] Xin Qi, Meixuan Li, Sijin Zhou, et al. Federated learning for science: A survey on the path to a trustworthy collaboration ecosystem. *Authorea Preprints*, 2025. 2
- [52] Xin Qi, Tao Xu, et al. Federated learning in oncology: bridging artificial intelligence innovation and privacy protection. *Information Fusion*, page 104154, 2026. 1

- [53] Zhuang Qi, Lei Meng, et al. Cross-silo prototypical calibration for federated learning with non-iid data. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3099–3107, 2023. 1
- [54] Zhuang Qi, Lei Meng, et al. Cross-silo feature space alignment for federated learning on clients with imbalanced data. In *The 39th Annual AAAI Conference on Artificial Intelligence (AAAI-25)*, pages 19986–19994, 2025. 7
- [55] Zhuang Qi, Ying-Peng Tang, et al. Class-wise balancing data replay for federated class-incremental learning. In *The Thirty-Ninth Annual Conference on Neural Information Processing Systems*, 2025. 2, 3, 5
- [56] Zhuang Qi, Pan Yu, Lei Meng, et al. Global prompt refinement with non-interfering attention masking for one-shot federated learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. 5
- [57] Xuankun Rong, Jianshu Zhang, Kun He, and Mang Ye. Can: Leveraging clients as navigators for generative replay in federated continual learning. *ICML*, 2025. 2
- [58] Minhyuk Seo, Hyunseo Koh, Wonje Jeung, et al. Learning equi-angular representations for online continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23933–23942, 2024. 3
- [59] Seonguk Seo, Jinkyu Kim, Geeho Kim, and Bohyung Han. Relaxed contrastive learning for federated learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12279–12288, 2024. 7
- [60] Donald Shenaj, Marco Toldo, Alberto Rigon, and Pietro Zanuttigh. Asynchronous federated continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5055–5063, 2023. 1
- [61] Jiangming Shi, Shanshan Zheng, et al. Clip-guided federated learning on heterogeneity and long-tailed data. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 14955–14963, 2024. 7
- [62] Kexuan Shi, Zhuang Qi, et al. Protoconnet: Prototypical augmentation and alignment for open-set few-shot image classification. *Displays*, page 103364, 2026. 8
- [63] Rui Sun, Yumin Zhang, Varun Ojha, Tejal Shah, Haoran Duan, Bo Wei, and Rajiv Ranjan. Exemplar-condensed federated class-incremental learning. *arXiv preprint arXiv:2412.18926*, 2024. 2
- [64] Minh-Tuan Tran, Trung Le, Xuan-May Le, Mehrtash Harandi, and Dinh Phung. Text-enhanced data-free approach for federated class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 23870–23880, 2024. 2, 5
- [65] Haozhao Wang, Haoran Xu, Yichen Li, et al. Fedcda: Federated learning with cross-rounds divergence-aware aggregation. In *The Twelfth International Conference on Learning Representations*, 2023. 1
- [66] Naibo Wang, Yuchen Deng, Wenjie Feng, Jianwei Yin, and See-Kiong Ng. Data-free federated class incremental learning with diffusion-based generative memory. *arXiv preprint arXiv:2405.17457*, 2024. 2
- [67] Xinghao Wu, Jianwei Niu, Xuefeng Liu, Guogang Zhu, Jiayuan Zhang, and Shaojie Tang. Enhancing visual representation with textual semantics: Textual semantics-powered prototypes for heterogeneous federated learning. *arXiv preprint arXiv:2503.13543*, 2025. 3
- [68] Xiaoshuo Yan, Zhaochuan Li, Lei Meng, et al. Empowering vision transformers with multi-scale causal intervention for long-tailed image classification. In *Proceedings of the Thirty-Fourth International Joint Conference on Artificial Intelligence*, pages 6785–6793, 2025. 8
- [69] Xin Yang, Hao Yu, Xin Gao, Hao Wang, Junbo Zhang, and Tianrui Li. Federated continual learning via knowledge fusion: A survey. *IEEE Transactions on Knowledge and Data Engineering*, 36(8):3832–3850, 2024. 2
- [70] Liping Yi, Gang Wang, Xiaoguang Liu, et al. Fedgh: Heterogeneous federated learning with generalized global header. In *Proceedings of the 31st ACM international conference on multimedia*, pages 8686–8696, 2023. 1
- [71] Liping Yi, Han Yu, Chao Ren, Gang Wang, Xiaoxiao Li, et al. Federated model heterogeneous matryoshka representation learning. *Advances in Neural Information Processing Systems*, 37:66431–66454, 2024. 1
- [72] Hao Yu, Xin Yang, Xin Gao, Yan Kang, Hao Wang, Junbo Zhang, and Tianrui Li. Personalized federated continual learning via multi-granularity prompt. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4023–4034, 2024. 1
- [73] Hao Yu, Xin Yang, et al. Overcoming spatial-temporal catastrophic forgetting for federated class-incremental learning. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 5280–5288, 2024. 1, 2
- [74] Hao Yu, Xin Yang, Le Zhang, Hanlin Gu, Tianrui Li, Lixin Fan, and Qiang Yang. Handling spatial-temporal data heterogeneity for federated continual learning via tail anchor. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4874–4883, 2025. 1
- [75] Chengchao Zhang, Fanhua Shang, Hongying Liu, Liang Wan, and Wei Feng. Fedagc: Federated continual learning with asymmetric gradient correction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3841–3850, 2025. 1
- [76] Jie Zhang, Chen Chen, et al. Target: Federated class-continual learning via exemplar-free distillation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4782–4793, 2023. 2, 5
- [77] Jinghua Zhang, Li Liu, Olli Silvén, Matti Pietikäinen, and Dewen Hu. Few-shot class-incremental learning for classification and object detection: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 2
- [78] Jiayuan Zhang, Xuefeng Liu, Jianwei Niu, Shaojie Tang, Haotian Yang, and Xinghao Wu. Causality inspired federated learning for ood generalization. In *Forty-second International Conference on Machine Learning*, 2025. 8
- [79] Yaozong Zheng, Bineng Zhong, Qihua Liang, et al. Decoupled spatio-temporal consistency learning for self-supervised tracking. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 10635–10643, 2025. 3
- [80] Yaozong Zheng, Bineng Zhong, et al. Towards universal modal tracking with online dense temporal token learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2025. 8