ANCHOR THE PAST, CALIBRATE THE PRESENT: MAINTAINING GEOMETRIC CONSISTENCY FOR ONLINE CLASS-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Realistic open environments demand continual recognition of emerging and previously learned categories, motivating online class-incremental learning (OCIL). As pretrained models (PTMs) adapt to new classes, old-class features can drift from historical classifier directions, while the classifier can become misaligned with the updated feature space. Replaying old examples alone does not maintain geometric consistency across these updates. To bridge this gap, we propose Anchoring in Replay (AIR), which contains two complementary parts. Historical geometry anchoring (HGA) aligns old-class replay features with fixed directions from the previous calibrated classifier, while mixed-geometry calibration (MGC) recalibrates the classifier from memory prototypes in the updated feature space. The calibrated directions support current prediction and anchor the next task. Under OCIL protocol, AIR achieves the best final average accuracy (FAA) among the evaluated replay-based methods on four benchmarks across three replay budgets.

1 INTRODUCTION

Realistic open environments continually introduce new categories, challenging traditional recognition models trained for a fixed class set (LeCun et al., 2015; Wang et al., 2024a). Online class-incremental learning (OCIL) addresses this challenge by learning new categories from an evolving stream while retaining recognition of previously learned ones (Aljundi et al., 2019; Mai et al., 2022). Major lines of continual learning (CL) include regularization-based (Kirkpatrick et al., 2017; Bonicelli et al., 2022), constraint-based (Lopez-Paz & Ranzato, 2017; Lai et al., 2025), replay-based (Rolnick et al., 2019; Wang et al., 2024b), and PTM-based methods (Wang et al., 2022b). Replay is widely used in conventional OCIL because of its simplicity and effectiveness (Caccia et al., 2022). Recently, PTMs have also attracted increasing attention in OCIL (Ma'sum et al., 2025). Existing PTM-based work largely explores prompt-based and adapter-based techniques (Wang et al., 2022a; Sun et al., 2025b), which adapt a small set of parameters while retaining most pretrained knowledge.

Replay supports OCIL by retaining a small set of historical examples and revisiting them alongside incoming data (Rolnick et al., 2019; Buzzega et al., 2020). In adapter-based PTMs, prediction depends on two coupled parts: an evolving feature space produced by the frozen backbone with trainable adapters, and a shared classifier defined over all seen classes. Recent CL methods have improved these two parts through adaptation modules and classifier reconstruction (Zhou et al., 2024; 2025). Combining replay with adapter-based PTMs is therefore a natural approach to OCIL, bringing together knowledge retention and lightweight adaptation. However, under a non-IID stream, adapter updates driven by new-class data and limited replay change the representations of old samples, a form of representation drift also studied in replay-based OCIL (Caccia et al., 2022). At the same time, classifier directions are fitted to features observed during these updates, with limited old data available to recalibrate them as the representation changes. The problem is geometric inconsistency across tasks: old-class features can drift away from their historical class directions, while the shared classifier can be misaligned with the updated feature space.

Figure 1 illustrates this inconsistency through the relative positions of features and class directions. In this geometric view, classifier weight vectors determine decision boundaries in feature space (Kansizoglou et al., 2021). Before adaptation, an old-class feature lies in a region assigned to its

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

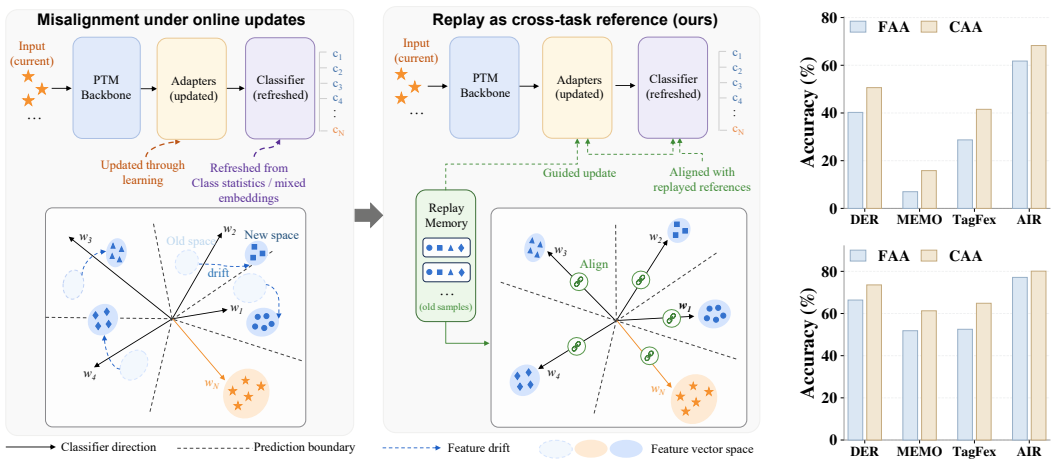

Figure 1: Left: Geometric inconsistency across online updates. Adapter changes can shift old-class features away from historical class directions, while the shared classifier can be misaligned with the updated feature geometry of old and new classes. Our method AIR uses replay as a common reference: HGA anchors old-class features to cached directions during adaptation, and MGC calibrates class directions from the retained memory afterward. The calibrated directions anchor the next task. Right: OCIL results at $M = 5000$ on ImageNet-A (top) and ImageNet-R (bottom).

class. As the adapters change, the feature can move away from the direction that supported its earlier prediction. If the classifier is poorly calibrated to the updated feature positions, the shifted point can fall into a competing class's decision region. The figure thus exposes both aspects of the problem: the loss of alignment with historical directions during adaptation, and the mismatch with the classifier after adaptation. This suggests two complementary objectives: maintain alignment with historical class directions during adaptation and recalibrate the classifier to the updated feature space.

To address this geometric inconsistency, we propose Anchoring in Replay (AIR), which coordinates feature anchoring and classifier calibration through replay. During adaptation, historical geometry anchoring (HGA) uses the previous calibrated classifier as a fixed reference, encouraging old-class replay features to align with historical directions as the adapters learn new classes. After each task, mixed-geometry calibration (MGC) recalibrates the classifier for current prediction using prototypes of the old and new classes represented in memory, thereby aligning it with the updated feature space. The resulting classifier then supplies historical anchors for the next task: the directions calibrated for the present guide subsequent feature learning. AIR thereby uses replay both to anchor features and to update the anchors, with a frozen pretrained backbone, lightweight adapters, and a bounded buffer under strict stream access. Under a unified OCIL protocol, AIR achieves the best FAA among the evaluated replay-based methods on all four benchmarks across three replay budgets, and the best CAA in 10 of the 12 settings.

2 Related Work

2.1 Traditional Online Class-Incremental Learning

Online class-incremental learning has been built largely around replay. Representative foundations include iCaRL (Rebuffi et al., 2017), ER (Rolnick et al., 2019), DER++ (Buzzega et al., 2020), and MIR (Aljundi et al., 2019). Recent OCIL methods keep this replay-centric formulation but refine how replay is sampled or exploited: SCR (Mai et al., 2021) strengthens replay with contrastive objectives, PCR (Lin et al., 2023b) uses proxy supervision, ER-ACE (Caccia et al., 2022), CBA (Wang et al., 2023) and UER (Lin et al., 2023a) address online bias, Onpro (Wei et al., 2023) updates class prototypes online, SER (Bellitto et al., 2024) uses saliency-based feature modulation, and POCL (Wu et al., 2024) and GEC (Lai et al., 2025) revisit the stability-plasticity trade-off from optimization perspectives. However, these traditional OCIL methods are still mostly developed with non-pretrained CNN-based backbones and replay under end-to-end parameter updates. They do not consider how to

exploit the knowledge already retained in pre-trained models to facilitate continual adaptation, and are therefore less commonly adopted in PTM-based continual learning settings.

2.2 PTM-BASED CONTINUAL LEARNING

Pre-trained backbones shift continual learning away from learning representations from scratch and toward lightweight adaptation of features. L2P (Wang et al., 2022b) and DualPrompt (Wang et al., 2022a) introduced prompt-based adaptation, while PROL (Ma'sum et al., 2025) extended prompt learning to streaming continual learning. Other PTM-based class-incremental methods broadened this design space: RanPAC (McDonnell et al., 2023) uses random projections with frozen PTM features, EASE (Zhou et al., 2024) expands subspaces over time, MOS (Sun et al., 2025b) performs model surgery on the pre-trained model, and APER (Zhou et al., 2025) studies adapter-based PTM continual learning with an emphasis on generalizability and adaptivity. DER (Yan et al., 2021), MEMO (Zhou et al., 2023), and TagFex (Zheng et al., 2025) are representative replay-based PTM-based continual learning methods. Our proposed AIR brings replay-based PTM methods into OCIL. Rather than using replay only for rehearsal, it incorporates replay into historical geometry anchoring and mixed-geometry calibration, which helps keep the evolving representation and the shared classifier consistent across tasks.

3 PRELIMINARIES

Problem definition. We consider online class-incremental learning (OCIL), where the stream is partitioned into T tasks,

$$\mathcal{S} = (\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_T), \tag{1}$$

where $\mathcal{D}_t = \{(\mathbf{x}_{t,i}, y_{t,i})\}_{i=1}^{n_t}$ denotes the examples arriving during task t. Its class set $\mathcal{C}_t$ contains only newly introduced categories, and $\mathcal{Y}_t = \bigcup_{\tau=1}^{t} \mathcal{C}_\tau$ is the cumulative seen label set. The learner processes $\mathcal{D}_t$ as K_t successive mini-batches, with each stream example appearing in exactly one incoming batch.

At online step k of task t, the learner receives a current mini-batch $\mathcal{B}_{t,k}$ and, when replay is used, samples a memory mini-batch $\mathcal{R}_{t,k}$ from a replay buffer of capacity M. The two mini-batches are used jointly for one gradient update. Without replay, the update uses only $\mathcal{B}_{t,k}$. Evaluation uses a single shared classifier over all seen classes $\mathcal{Y}_t$, without a task identifier at inference.

Let $\mathbf{\Phi}$ denote the frozen pre-trained encoder parameters and $\mathbf{\Theta}_t$ the adapter parameters after task t. The adapted representation is

$$\mathbf{h}_t(\mathbf{x}) = f(\mathbf{x}; \mathbf{\Phi}, \mathbf{\Theta}_t) \in \mathbb{R}^d. \tag{2}$$

A shared cosine classifier has weights $\mathbf{W}_t = [\mathbf{w}_{t,1}, \ldots, \mathbf{w}_{t,|\mathcal{Y}_t|}]^\top \in \mathbb{R}^{|\mathcal{Y}_t| \times d}$ and a learned scalar σ_t. Its logits are

$$z_{t,c}(\mathbf{x}) = \sigma_t \frac{\mathbf{w}_{t,c}^\top \mathbf{h}_t(\mathbf{x})}{\|\mathbf{w}_{t,c}\|_2 \, \|\mathbf{h}_t(\mathbf{x})\|_2}, \qquad c \in \mathcal{Y}_t, \tag{3}$$

with numerically stabilized normalization in the implementation. The final prediction is $\hat{y}_t(\mathbf{x}) = \arg\max_{c \in \mathcal{Y}_t} z_{t,c}(\mathbf{x})$. The goal is to maintain recognition of all seen classes while updating the adapters and shared classifier within this stream-access constraint.

Adapter tuning. We use bottleneck adapters as the PEFT module in transformer blocks. For block $\ell \in \{1, \ldots, L\}$, let $\mathbf{X}_t^{(\ell)} \in \mathbb{R}^{b \times n \times d}$ be the input token tensor at task t, where b is the mini-batch size, n is the token length, and d is the hidden dimension. The adapter first projects the tokens to a bottleneck of dimension r:

$$\mathbf{Z}_t^{(\ell)} = \mathbf{X}_t^{(\ell)} \mathbf{D}^{(\ell)} \in \mathbb{R}^{b \times n \times r}, \tag{4}$$

where $\mathbf{D}^{(\ell)} \in \mathbb{R}^{d \times r}$ is the down-projection matrix. The bottleneck activations are then mapped back to the hidden dimension,

$$\mathbf{A}_t^{(\ell)} = \phi\left(\mathbf{Z}_t^{(\ell)}\right) \mathbf{U}_{\text{up}}^{(\ell)}, \tag{5}$$

where $\phi(\cdot)$ denotes the element-wise nonlinearity and dropout, and $\mathbf{U}_{\text{up}}^{(\ell)} \in \mathbb{R}^{r \times d}$ is the up-projection matrix. The adapter output is added through the residual path, so adaptation stays in a low-dimensional subspace while the pre-trained backbone remains frozen.

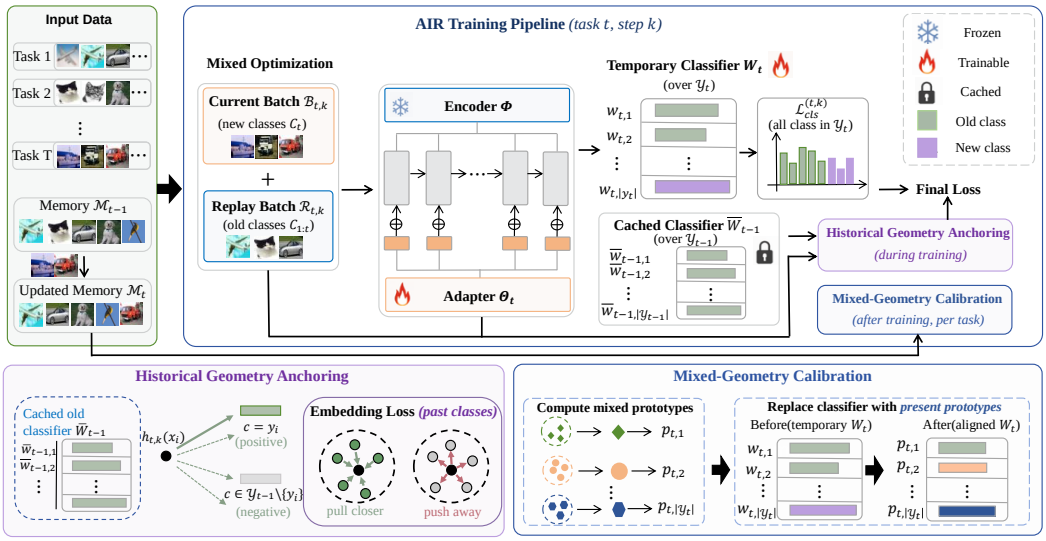

Figure 2: Overview of AIR for task $t > 1$ under OCIL. The previous task's calibrated classifier is cached and held fixed throughout task t. Each incoming batch and a sampled replay batch undergo separate forward passes through the frozen encoder with trainable adapters. One gradient update combines classification on both batches with HGA anchoring earlier-task replay features to the cached class directions, followed by a reservoir update using the incoming examples. After the final mini-batch, MGC recalibrates the classifier from prototypes computed only on $\mathcal{M}_t$ in the final feature space. The calibrated directions anchor task $t + 1$. Appendix A summarizes the full procedure.

4 Method: Anchoring in Replay

4.1 Cross-Task Consistency under Replay

AIR coordinates representation updates and classifier calibration under strict stream access. Let $\mathcal{M}_{t,k}$ denote the replay buffer after update k of task t, with $\mathcal{M}_{t,0} = \mathcal{M}_{t-1}$, $\mathcal{M}_t = \mathcal{M}_{t,K_t}$, and $|\mathcal{M}_{t,k}| \le M$. At update k, we sample a replay mini-batch $\mathcal{R}_{t,k}$ uniformly without replacement within the batch from the live buffer $\mathcal{M}_{t,k-1}$ and form

$$\mathcal{U}_{t,k} = \mathcal{B}_{t,k} \cup \mathcal{R}_{t,k}. \tag{6}$$

The first task uses only incoming examples, so $\mathcal{R}_{1,k} = \varnothing$. From task 2 onward, current and replay examples undergo separate forward passes followed by one joint optimization step. Only after that step do the incoming examples enter the buffer through reservoir sampling (Vitter, 1985). Hence the current batch cannot replay itself in the same update. The live buffer can contain both earlier-task examples and examples already processed in the current task.

This update can change both old-class feature locations and the shared classifier. AIR addresses their consistency at two stages: historical geometry anchoring (HGA) constrains old-class replay features using fixed directions from the previous task, while mixed-geometry calibration (MGC) refreshes classifier directions using the updated buffer after the task ends. Figure 2 shows the resulting data lifecycle. Below, $\mathbf{h}_{t,k}$ and $z_{t,k,c}$ denote the features and logits evaluated immediately before update k; $\mathbf{h}_t$ denotes the final task-t representation used for calibration.

4.2 Historical Geometry Anchoring

At the beginning of task $t > 1$, we cache the previous task's calibrated classifier,

$$\bar{\mathbf{W}}_{t-1} = [\bar{\mathbf{w}}_{t-1,1}, \bar{\mathbf{w}}_{t-1,2}, \ldots, \bar{\mathbf{w}}_{t-1,|\mathcal{Y}_{t-1}|}]^\top \in \mathbb{R}^{|\mathcal{Y}_{t-1}| \times d}. \tag{7}$$

Each $\bar{\mathbf{w}}_{t-1,c}$ is the cached classifier vector for an earlier class c. The cached matrix remains fixed throughout task t, while the current classifier is expanded to $\mathcal{Y}_t$ and trained with the adapters. Since

the cached classifier contains only earlier-task classes, HGA uses the subset

$$\mathcal{R}_{t,k}^{\text{old}} = \{(\mathbf{x}, y) \in \mathcal{R}_{t,k} : y \in \mathcal{Y}_{t-1}\}. \tag{8}$$

Replayed examples from the current task contribute to classification but not to HGA. For an old-class replay example, we compute its current feature's cosine similarity to each cached direction. Let $(\mathbf{x}_i, y_i) \in \mathcal{R}_{t,k}^{\text{old}}$ denote an old-class replay input and its label. The feature vector $\mathbf{h}_{t,k}(\mathbf{x}_i)$ is computed by the frozen encoder with the current adapters immediately before update k of task t. Its cosine similarity to each cached direction is

$$s_{i,c}^{(t,k)} = \frac{\bar{\mathbf{w}}_{t-1,c}^{\top} \mathbf{h}_{t,k}(\mathbf{x}_i)}{\|\bar{\mathbf{w}}_{t-1,c}\|_2 \, \|\mathbf{h}_{t,k}(\mathbf{x}_i)\|_2}, \qquad c \in \mathcal{Y}_{t-1}. \tag{9}$$

The anchoring loss is

$$\mathcal{L}_{\text{emb}}^{(t,k)} = -\frac{1}{|\mathcal{R}_{t,k}^{\text{old}}|} \sum_{(\mathbf{x}_i, y_i) \in \mathcal{R}_{t,k}^{\text{old}}} \log \frac{\exp(s_{i,y_i}^{(t,k)}/\tau_e)}{\sum_{c \in \mathcal{Y}_{t-1}} \exp(s_{i,c}^{(t,k)}/\tau_e)}, \tag{10}$$

where $\tau_e > 0$ is the embedding temperature. We set this loss to zero when $\mathcal{R}_{t,k}^{\text{old}}$ is empty, including the first task. It encourages alignment with the correct historical direction relative to the other old-class directions. Gradients flow through current features to the adapters, but not into $\bar{\mathbf{W}}_{t-1}$. New classes are excluded from the denominator because no historical directions exist for them.

4.3 Mixed-Geometry Calibration

After the final online update $k = K_t$ of task t, MGC recalibrates the classifier using only the final replay buffer $\mathcal{M}_{t,K_t} = \mathcal{M}_t$. With network parameters fixed, we re-encode its retained examples under $\mathbf{h}_t$. For each represented class, define

$$\mathcal{M}_t^c = \{(\mathbf{x}, y) \in \mathcal{M}_t : y = c\}, \qquad \mathbf{p}_{t,c} = \frac{1}{|\mathcal{M}_t^c|} \sum_{(\mathbf{x}, y) \in \mathcal{M}_t^c} \mathbf{h}_t(\mathbf{x}). \tag{11}$$

We replace the corresponding classifier vector directly:

$$\mathbf{w}_{t,c} \leftarrow \mathbf{p}_{t,c}, \qquad c \in \mathcal{Y}_t \text{ with } |\mathcal{M}_t^c| > 0. \tag{12}$$

If a class has no retained examples, its trained classifier vector is left unchanged. Prototypes are means of the unnormalized features; the cosine classifier normalizes weights and features when computing logits. MGC performs no gradient updates and retains the learned logit scale.

The mixed geometry is supplied by retained examples from both the current and previous tasks, all re-encoded in the same final feature space. No discarded stream examples are reopened, and no full-task dataset is traversed for calibration. The calibrated classifier $\mathbf{W}_t$ is used for prediction and cached as the historical reference for task $t + 1$, connecting HGA and MGC across successive tasks.

4.4 Final Objective

At each online update, classification uses all current and replay examples with equal per-example weight:

$$\mathcal{L}_{\text{cls}}^{(t,k)} = -\frac{1}{|\mathcal{U}_{t,k}|} \sum_{(\mathbf{x}, y) \in \mathcal{U}_{t,k}} \log \frac{\exp(z_{t,k,y}(\mathbf{x}))}{\sum_{c \in \mathcal{Y}_t} \exp(z_{t,k,c}(\mathbf{x}))}. \tag{13}$$

The joint objective is

$$\mathcal{L}^{(t,k)} = \mathcal{L}_{\text{cls}}^{(t,k)} + \lambda_{\text{emb}} \mathcal{L}_{\text{emb}}^{(t,k)}, \tag{14}$$

where $\lambda_{\text{emb}} \geq 0$ controls historical anchoring. One backward pass and optimizer step update the adapters, current classifier weights, and logit scale, while the pre-trained encoder and cached classifier remain fixed. Immediately after every mini-batch optimizer step, the current batch updates the reservoir:

$$\mathcal{M}_{t,k} \leftarrow \text{ReservoirUpdate}(\mathcal{M}_{t,k-1}, \mathcal{B}_{t,k}; M). \tag{15}$$

After the task's last batch, memory-only MGC replaces the represented class directions. Thus adaptation and memory insertion operate per mini-batch, whereas calibration and anchor renewal occur at the task boundaries or time to evaluate.

5 EXPERIMENTS

5.1 EXPERIMENTAL SETUP

Benchmarks and protocol. We evaluate OCIL on four benchmarks using the task splits common in recent PTM-based OCIL studies (Ma'sum et al., 2025). Each dataset has 10 tasks: CIFAR-100 introduces 10 classes per task, while CUB-200, ImageNet-A, and ImageNet-R introduce 20. Each training example arrives once in a current mini-batch of 10 samples. From the second task onward, replay methods pair it with 10 samples drawn from a live reservoir of capacity $M \in \{1000, 5000, 10000\}$. Current and replay batches use separate forward passes followed by one joint gradient update; the current batch is then inserted into the reservoir. Further raw-example access is restricted to the bounded memory. Evaluation covers all seen classes without a task identifier. Tables 1 and 2 report means over completed runs; local experiments use three different random seeds.

Implementation details. Our main configuration uses a ViT-B/16 backbone pre-trained on ImageNet-21K. Methods supporting adapters use the same bottleneck adapters with dimension $r = 64$, and replay-based methods use reservoir sampling as the default memory update rule. We tune the learning rate over $\{0.001, 0.005, 0.01, 0.03, 0.05, 0.1\}$ and report the best setting for each method under the one-epoch protocol. For other configurations we adapt the baselines' official implementations. We use PILOT framework Sun et al. (2025a) to reproduce baselines and additionally include replay-augmented L2P-R (Wang et al., 2022b), EASE-R (Zhou et al., 2024), and MOS-R (Sun et al., 2025b) as strong baselines (Appendix F details EASE-R and MOS-R). For AIR's hyperparameters, $\lambda_{\text{emb}} = 4.0$ and $\tau_e = 0.07$ are fixed across datasets and memory budgets.

Experiment design. We organize the evaluation into four parts. First, Table 1 compares AIR with iCaRL (Rebuffi et al., 2017), DER (Yan et al., 2021), MEMO (Zhou et al., 2023), TagFex (Zheng et al., 2025), SCR (Mai et al., 2021), ER-ACE (Caccia et al., 2022), L2P-R (Wang et al., 2022b), EASE-R (Zhou et al., 2024), and MOS-R (Sun et al., 2025b) across the replay-capacity study; EASE-R and MOS-R are reported only at $M = 5000$ and 10000 because missing classes in the reservoir prevent them from completing training at $M = 1000$. Second, Table 2 compares AIR at $M = 5000$ with rehearsal-free PTM-based OCIL methods, including previous literature-reported results and reproduce of APER and EASE. Third, Table 3 isolates the AIR components at $M = 1000$. Finally, Table 4 tests our AIR on a ViT-Tiny backbone against PROL in the same setting and traditional CNN-based baselines.

Metrics. Following recent OCIL work, we report final accuracy after the last task (FAA), cumulative average accuracy over the full trajectory (CAA), and forgetting measure (FM). Higher FAA and CAA are better, while lower FM indicates stronger retention. The detailed definitions of metrics are in Appendix E.

5.2 REPLAY-BASED OCIL COMPARISON

We first compare AIR with replay-based OCIL baselines under the unified three-seed, single-epoch strict OCIL protocol at replay budgets $M \in \{1000, 5000, 10000\}$.

Table 1 shows that AIR achieves the highest mean FAA on all four benchmarks across all three replay budgets, and the highest CAA in 10 of the 12 settings. The two CAA exceptions are CIFAR-100 at $M = 5000$ and 10000, where MOS-R leads by 0.20 and 0.03 percentage points, respectively. Even under the tightest budget $M = 1000$, AIR improves CAA over the strongest reported baseline by 3.84 points on CIFAR-100, 4.01 on ImageNet-R, 13.01 on ImageNet-A, and 5.68 on CUB-200. At $M = 5000$, the CAA/FAA gains over the best baseline for each metric are 0.45/2.46 points on ImageNet-R, 6.43/6.60 on ImageNet-A, and 0.80/0.13 on CUB-200, while FAA on CIFAR-100 is nearly tied with MOS-R (88.94 versus 88.93). The clearest advantage remains on ImageNet-A at $M = 10000$, where the CAA/FAA margins reach 6.54/7.63 points.

AIR also obtains the lowest FM in five settings: CIFAR-100 and ImageNet-R at $M = 1000$, ImageNet-R at $M = 5000$, and CIFAR-100 and ImageNet-R at $M = 10000$. In several remaining settings, the FM gap is small: at $M = 5000$, AIR is 0.48 points above iCaRL on ImageNet-A and 0.36 above MOS-R on CUB-200; at $M = 10000$, the corresponding gaps are 0.20 and 0.19 points.

Table 1: Replay-based OCIL comparison at three replay capacities, reporting mean FAA/CAA/FM (%). EASE-R and MOS-R do not work at $M = 1000$ because missing classes in the reservoir prevent them from completing training.

Budget	Method	CIFAR-100			ImageNet-R			ImageNet-A			CUB-200		
		FAA	CAA	FM	FAA	CAA	FM	FAA	CAA	FM	FAA	CAA	FM
1000	iCaRL	56.14	66.31	16.28	25.11	38.80	9.33	17.66	26.91	**6.12**	32.18	46.31	**4.83**
	DER	78.92	87.27	19.51	61.47	71.00	26.20	41.28	50.37	26.03	66.37	78.43	27.10
	SCR	77.92	85.22	16.02	56.59	69.33	18.57	24.71	34.94	15.03	67.91	79.31	13.82
	ER-ACE	65.70	75.00	19.33	54.66	65.17	11.69	21.18	33.50	8.53	65.10	76.30	7.47
	L2P-R	74.35	83.83	23.19	53.12	65.37	33.45	30.72	41.62	34.32	68.61	79.80	26.70
	MEMO	54.40	68.64	33.28	45.23	58.84	30.23	6.87	16.04	25.35	64.38	73.46	24.54
	TagFex	62.49	79.19	35.07	42.75	59.74	40.71	26.38	41.20	34.57	60.77	75.78	32.34
	AIR	**85.62**	**91.11**	**7.29**	**67.33**	**75.01**	**7.28**	**51.70**	**63.38**	11.31	**75.84**	**85.48**	11.75
5000	iCaRL	55.79	66.67	16.38	27.02	39.08	8.68	19.31	28.02	**6.09**	31.57	46.44	5.56
	DER	82.92	89.43	14.42	66.39	73.57	19.85	40.24	50.63	27.78	69.33	79.27	22.66
	SCR	81.70	87.44	11.50	66.42	73.77	12.82	28.48	37.89	14.38	76.84	83.45	9.42
	ER-ACE	74.20	79.24	9.37	59.47	67.90	5.28	23.28	33.71	7.18	66.75	77.57	9.24
	L2P-R	82.05	87.26	13.90	64.63	70.82	19.72	35.42	43.19	29.63	72.80	80.93	20.79
	MEMO	60.18	71.42	26.63	51.84	61.29	22.82	7.00	15.84	24.57	66.28	73.68	21.82
	EASE-R	87.76	91.94	**3.65**	73.79	79.38	5.68	47.84	58.48	11.31	84.05	89.09	5.69
	TagFex	73.84	83.59	22.11	52.52	64.86	29.08	28.72	41.53	31.74	61.20	76.04	31.89
	MOS-R	88.93	**92.70**	4.57	74.67	79.64	4.52	55.17	61.85	8.12	85.96	88.71	**4.74**
	AIR	**88.94**	92.50	4.47	**77.13**	**80.09**	**2.81**	**61.77**	**68.28**	6.57	**86.09**	**89.89**	5.10
10000	iCaRL	57.70	67.24	14.56	27.38	40.81	9.30	19.75	28.07	**5.48**	31.19	46.45	6.01
	DER	84.41	89.81	12.54	67.00	74.10	18.88	40.16	50.95	28.41	69.28	79.79	23.29
	SCR	81.36	87.22	11.58	66.89	73.79	12.06	29.52	38.16	13.90	78.63	84.04	8.43
	ER-ACE	77.76	77.57	6.01	59.85	68.35	5.49	23.17	33.73	7.41	65.96	77.15	7.80
	L2P-R	83.75	87.86	11.64	66.17	71.56	17.91	33.82	43.01	31.75	74.71	81.12	18.98
	MEMO	62.19	72.11	23.86	52.06	61.53	22.88	7.35	15.88	24.46	64.97	73.42	23.60
	EASE-R	88.15	92.20	3.73	73.89	79.43	6.00	47.69	58.47	11.29	84.34	89.23	5.86
	TagFex	73.53	83.63	22.23	51.91	65.06	29.71	29.14	41.57	31.05	62.05	76.33	30.71
	MOS-R	89.04	**92.72**	4.36	75.08	79.84	4.05	54.80	61.82	9.34	86.36	88.87	**4.69**
	AIR	**89.72**	92.69	**3.64**	**77.71**	**80.60**	**3.03**	**62.43**	**68.36**	5.68	**86.73**	**90.15**	4.88

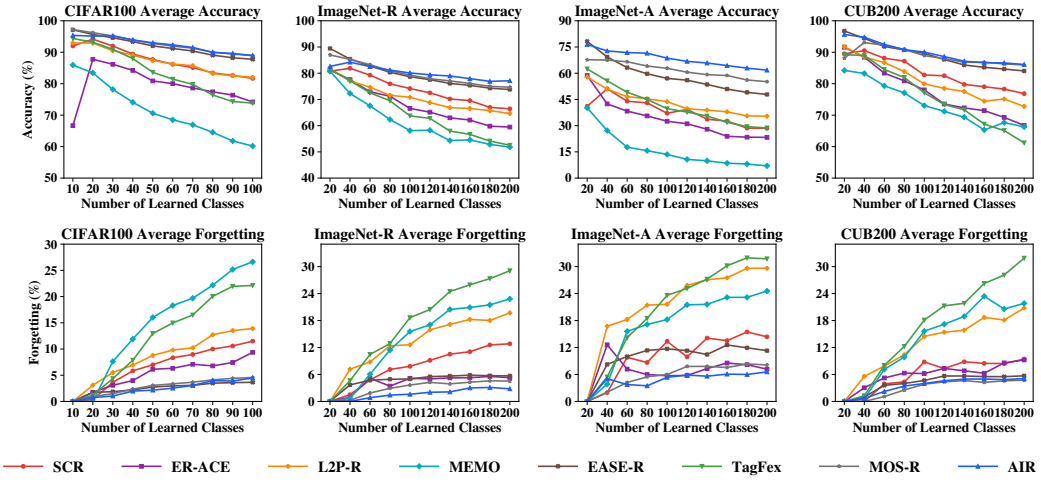

Figure 3: Task-wise OCIL results at $M = 5000$ for eight replay-based methods. Curves report three-seed means of accuracy over all seen classes (top) and average forgetting (bottom).

Importantly, the lower-FM iCaRL results on ImageNet-A come with much lower FAA. The results show that AIR combines low forgetting with substantially stronger final and cumulative accuracy.

The budget sweep further shows that AIR benefits from additional replay memory. As M increases from 1000 to 10000, its CAA rises from 91.11 to 92.69 on CIFAR-100, 75.01 to 80.60 on ImageNet-R, 63.38 to 68.36 on ImageNet-A, and 85.48 to 90.15 on CUB-200. Over the same range, FM decreases from 7.29 to 3.64, 7.28 to 3.03, 11.31 to 5.68, and 11.75 to 4.88, respectively. AIR thus retains the highest FAA across the full budget range while using additional memory to improve cumulative accuracy and overall retention.

Figure 3 complements the aggregate results with task-wise trajectories at $M = 5000$. AIR leads the accuracy curves from the second task on ImageNet-A and from the fourth task on ImageNet-R, while remaining close to the leading methods on CIFAR-100 and CUB-200. It also maintains the lowest forgetting throughout the incremental tasks on ImageNet-R and finishes with the lowest FM among the plotted methods on ImageNet-A. These trajectories show that AIR sustains its accuracy advantage as the class set expands.

5.3 Comparison with Rehearsal-Free PTM-Based OCIL Methods

We next compare AIR at the default replay budget $M = 5000$ with representative rehearsal-free PTM-based OCIL methods.

Table 2: OCIL comparison with rehearsal-free PTM methods in the mean FAA/CAA/FM format (%). RanPAC, RanDumb, L2P, DualPrompt, MOS, and PROL are reported by Ma'sum et al. (2025); APER, EASE and AIR use the same settings and report three-seed means.

Method	CIFAR-100			ImageNet-R			ImageNet-A			CUB-200		
	FAA	CAA	FM	FAA	CAA	FM	FAA	CAA	FM	FAA	CAA	FM
L2P	80.99	86.66	6.71	52.08	55.90	5.67	14.64	21.88	5.35	61.98	71.46	9.81
DualPrompt	82.53	88.06	4.53	61.17	62.73	3.75	20.05	21.57	**1.91**	58.06	49.95	6.23
RanPAC	9.52	27.92	94.52	7.24	18.81	67.27	3.77	14.86	57.76	9.52	27.06	88.98
RanDumb	9.56	27.94	94.51	7.32	18.80	67.67	3.69	14.56	55.15	9.53	27.21	89.69
EASE	83.30	89.01	9.42	70.40	76.82	8.51	48.02	58.26	13.24	75.51	83.28	12.80
APER-adapter	76.32	82.12	7.00	62.98	68.80	6.54	49.22	58.73	10.51	85.31	89.56	5.19
MOS	83.99	89.53	11.34	52.13	67.97	35.72	44.79	55.14	15.50	61.74	77.03	25.25
PROL	86.32	91.35	6.34	73.50	78.13	4.82	47.72	58.96	3.29	72.51	79.89	8.92
AIR ($M = 5000$)	**88.94**	**92.50**	**4.47**	**77.13**	**80.09**	**2.81**	**61.77**	**68.28**	6.57	**86.09**	**89.89**	**5.10**

Table 2 shows that AIR achieves the highest mean FAA and CAA on all four benchmarks. Relative to the strongest reported baseline for each metric, the CAA/FAA gains are 1.15/2.62 points on CIFAR-100, 1.96/3.63 on ImageNet-R, 9.32/12.55 on ImageNet-A, and 0.33/0.78 on CUB-200. The largest improvement occurs on ImageNet-A: AIR reaches 61.77 FAA compared with 49.22 for APER-adapter, the strongest baseline in final accuracy, and 68.28 CAA compared with 58.96 for PROL, the strongest baseline in cumulative accuracy.

The accuracy gains are accompanied by competitive retention. AIR achieves the lowest FM on CIFAR-100, ImageNet-R, and CUB-200. Compared with EASE, it reduces FM from 9.42 to 4.47, 8.51 to 2.81, 13.24 to 6.57, and 12.80 to 5.10 in the same dataset order, while improving both FAA and CAA on all four benchmarks. On ImageNet-A, DualPrompt and PROL report lower FM, but their FAA values are 20.05 and 47.72, respectively, compared with 61.77 for AIR. Together with the matched-budget comparison in Table 1, these results support replay-based geometric coordination as an effective strategy for PTM-based OCIL.

5.4 Component Analysis

We next test whether the two AIR components remain complementary in the tighter replay regime at $M = 1000$. Table 3 compares the direct Adapter+ER endpoint, the two single-component variants, and full AIR with both components enabled.

The direct Adapter+ER endpoint has the lowest CAA and FAA on all four datasets, which confirms that replay with plain adapter tuning is not enough to maintain a stable shared geometry under tight memory. Adding either historical geometry anchoring or mixed-geometry calibration improves this

Table 3: Component analysis at $M = 1000$. The table organizes the ablation by component combinations. The results are three-seed means under the single-epoch OCIL protocol. Each metric cell reports CAA / FAA.

Variant	HGA	MGC	CIFAR-100 CAA / FAA	ImageNet-R CAA / FAA	ImageNet-A CAA / FAA	CUB-200 CAA / FAA
Adapter+ER	×	×	66.23 / 55.88	38.88 / 25.09	26.91 / 17.62	46.26 / 31.96
Adapter+ER + HGA	✓	×	69.80 / 61.85	43.47 / 32.58	31.47 / 23.24	46.76 / 32.91
Adapter+ER + MGC	×	✓	90.34 / 84.93	72.87 / 65.05	54.69 / 44.83	83.10 / 72.17
AIR	✓	✓	**91.11 / 85.62**	**75.01 / 67.33**	**63.38 / 51.70**	**85.48 / 75.84**

endpoint, but neither single-component variant reaches the full AIR design. Notably, the stronger single-component performance of MGC suggests that classifier bias is more severe than the adapters in large PTM-based OCIL, making post-update classifier calibration relatively more effective in this regime. AIR achieves the highest CAA and FAA on every benchmark, improving over the MGC-only variant by 0.77 / 0.69 on CIFAR-100, 2.14 / 2.28 on ImageNet-R, 8.69 / 6.87 on ImageNet-A, and 2.38 / 3.67 on CUB-200 in CAA / FAA.

The components act at different stages: HGA preserves old-class structure during adaptation, while MGC restores classifier consistency after the update. Neither component alone matches the full combination; Appendix B reports the full comparison across replay budgets.

Table 4: Lightweight CIFAR-100 comparison. The baselines are transcribed from the lightweight CIFAR-100 comparison table reported by PROL (Ma'sum et al., 2025).

Method	Mem	Backbone	Size	PTM	FAA
ER	5000	ResNet-18	11.2M	no	38.70
DER++	5000	ResNet-18	11.2M	no	37.62
ERACE	5000	ResNet-18	11.2M	no	39.82
OCM	5000	ResNet-18	11.2M	no	42.01
GSA	5000	ResNet-18	11.2M	no	42.27
OnPro	5000	ResNet-18	11.2M	no	41.47
ERSM	5000	ResNet-18	11.2M	no	47.72
PROL	–	ViT-Tiny	5.6M	yes	64.59
AIR	5000	ViT-Tiny+adapter	5.8M	yes	**71.29**

5.5 Lightweight Experiments

Table 4 tests AIR with a ViT-Tiny backbone, assessing transfer from ViT-B and providing context against lightweight CNN-based OCIL methods. AIR reaches 71.29 FAA, exceeding PROL's 64.59 by 6.70 percentage points with a similarly compact pre-trained backbone. It exceeds ERSM, the strongest CNN baseline in the table, by 23.57 points. This supports AIR's effectiveness across backbone scales.

6 Conclusion

We presented AIR, a replay framework that coordinates representations and classifiers in PTM-based online class-incremental learning. Historical geometry anchoring preserves old-class directions during adapter optimization, and mixed-geometry calibration refreshes the classifier after the update. With standard reservoir replay and bottleneck adapters, AIR achieves the highest mean FAA among the evaluated replay-based methods on all four benchmarks across three replay budgets, and the highest mean CAA in 10 of the 12 settings. These results support using replay as a cross-task reference for maintaining representation–classifier consistency.

AI USE STATEMENT

In this work, we used generative AI tools to assist with translation and method implementation. We have not used generative AI tools to help develop theoretical models or conceptual frameworks, formulate mathematical claims, propose or refine hypotheses, design or provide feedback on the research methodology or experiments, clean or reformat datasets, or interpret results; tasks involving synthetic dataset generation, proving mathematical claims, writing proofs, and qualitative or thematic data analysis are not applicable to this work. Additionally, we used generative AI tools to create or modify scientific figures or images and to edit the paper to improve readability. We have reviewed all AI-assisted work. In particular, the authors reviewed and tested the AI-assisted implementation for correctness, checked AI-assisted figures against the underlying experimental records, and reviewed all translations and language edits. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

The experiments are full reproducible. Section 4 specifies the AIR objective and classifier calibration, and Section 5 describes the benchmarks, task splits, backbone, replay budgets, and evaluation protocol. Appendix A provides the training procedure. Appendix H reports the complete task-wise accuracy and forgetting results underlying the OCIL replay comparison in Table 1. Appendix J reports 95% confidence intervals for Table 1. Additionally, we provide the implementation of our method at https://anonymous.4open.science/r/AIR-B831/.

REFERENCES

Rahaf Aljundi, Eugene Belilovsky, Tinne Tuytelaars, Laurent Charlin, Massimo Caccia, Min Lin, and Lucas Page-Caccia. Online continual learning with maximal interfered retrieval. *Advances in neural information processing systems*, 32, 2019.

Giovanni Bellitto, Federica Proietto Salanitri, Matteo Pennisi, Matteo Boschini, Lorenzo Bonicelli, Angelo Porrello, Simone Calderara, Simone Palazzo, and Concetto Spampinato. Saliency-driven experience replay for continual learning. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang (eds.), *Advances in Neural Information Processing Systems*, pp. 103356–103383. Curran Associates, Inc., 2024.

Lorenzo Bonicelli, Matteo Boschini, Angelo Porrello, Concetto Spampinato, and Simone Calderara. On the effectiveness of lipschitz-driven rehearsal in continual learning. *Advances in Neural Information Processing Systems*, 35:31886–31901, 2022.

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. *Advances in neural information processing systems*, 33:15920–15930, 2020.

Lucas Caccia, Rahaf Aljundi, Nader Asadi, Tinne Tuytelaars, Joelle Pineau, and Eugene Belilovsky. New insights on reducing abrupt representation change in online continual learning. In *ICLR*, 2022.

Ioannis Kansizoglou, Loukas Bampis, and Antonios Gasteratos. Deep feature space: A geometrical perspective. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(10):6823–6838, 2021.

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.

Song Lai, Changyi Ma, Fei Zhu, Zhe Zhao, Xi Lin, Gaofeng Meng, and Qingfu Zhang. Gradient-guided epsilon constraint method for online continual learning. *Advances in Neural Information Processing Systems*, 38:124209–124238, 2025.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE transactions on pattern analysis and machine intelligence*, 40(12):2935–2947, 2017.

Huiwei Lin, Shanshan Feng, Baoquan Zhang, Hongliang Qiao, Xutao Li, and Yunming Ye. Uer: A heuristic bias addressing approach for online continual learning. In *Proceedings of the 31st ACM International Conference on Multimedia*, pp. 96–104, 2023a.

Huiwei Lin, Baoquan Zhang, Shanshan Feng, Xutao Li, and Yunming Ye. Pcr: Proxy-based contrastive replay for online class-incremental continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 24246–24255, 2023b.

David Lopez-Paz and Marc'Aurelio Ranzato. Gradient episodic memory for continual learning. *Advances in neural information processing systems*, 30, 2017.

Zheda Mai, Ruiwen Li, Hyunwoo Kim, and Scott Sanner. Supervised contrastive replay: Revisiting the nearest class mean classifier in online class-incremental continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 3589–3599, June 2021.

Zheda Mai, Ruiwen Li, Jihwan Jeong, David Quispe, Hyunwoo Kim, and Scott Sanner. Online continual learning in image classification: An empirical survey. *Neurocomputing*, 469:28–51, 2022.

M Anwar Ma'sum, Mahardhika Pratama, Savitha Ramasamy, Lin Liu, Habibullah Habibullah, and Ryszard Kowalczyk. Prol: Rehearsal free continual learning in streaming data via prompt online learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 2471–2481, 2025.

Mark D McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, and Anton Van den Hengel. Ranpac: Random projections and pre-trained models for continual learning. *Advances in Neural Information Processing Systems*, 36:12022–12053, 2023.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H Lampert. icarl: Incremental classifier and representation learning. In *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, pp. 2001–2010, 2017.

David Rolnick, Arun Ahuja, Jonathan Schwarz, Timothy Lillicrap, and Gregory Wayne. Experience replay for continual learning. *Advances in neural information processing systems*, 32, 2019.

Hai-Long Sun, Da-Wei Zhou, De-Chuan Zhan, and Han-Jia Ye. Pilot: A pre-trained model-based continual learning toolbox. *SCIENCE CHINA INFORMATION SCIENCES*, 68(4):147101, 2025a.

Hai-Long Sun, Da-Wei Zhou, Hanbin Zhao, Le Gan, De-Chuan Zhan, and Han-Jia Ye. Mos: Model surgery for pre-trained model-based class-incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 20699–20707, 2025b.

Jeffrey S Vitter. Random sampling with a reservoir. *ACM Transactions on Mathematical Software (TOMS)*, 11(1):37–57, 1985.

Liyuan Wang, Xingxing Zhang, Hang Su, and Jun Zhu. A comprehensive survey of continual learning: theory, method and application. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024a.

Maorong Wang, Nicolas Michel, Jiafeng Mao, and Toshihiko Yamasaki. Dealing with synthetic data contamination in online continual learning. *Advances in Neural Information Processing Systems*, 37:51284–51311, 2024b.

Quanziang Wang, Renzhen Wang, Yichen Wu, Xixi Jia, and Deyu Meng. Cba: Improving online continual learning via continual bias adaptor. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19082–19092, October 2023.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, et al. Dualprompt: Complementary prompting for rehearsal-free continual learning. In *European conference on computer vision*, pp. 631–648. Springer, 2022a.

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 139–149, 2022b.

Yujie Wei, Jiaxin Ye, Zhizhong Huang, Junping Zhang, and Hongming Shan. Online prototype learning for online continual learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 18764–18774, 2023.

Yichen Wu, Hong Wang, Peilin Zhao, Yefeng Zheng, Ying Wei, and Long-Kai Huang. Mitigating catastrophic forgetting in online continual learning by modeling previous task interrelations via pareto optimization. In *Forty-first international conference on machine learning*, 2024.

Shipeng Yan, Jiangwei Xie, and Xuming He. Der: Dynamically expandable representation for class incremental learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 3014–3023, 2021.

Bowen Zheng, Da-Wei Zhou, Han-Jia Ye, and De-Chuan Zhan. Task-agnostic guided feature expansion for class-incremental learning. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 10099–10109, 2025.

Da-Wei Zhou, Qi-Wei Wang, Han-Jia Ye, and De-Chuan Zhan. A model or 603 exemplars: Towards memory-efficient class-incremental learning. In *The Eleventh International Conference on Learning Representations*, 2023.

Da-Wei Zhou, Hai-Long Sun, Han-Jia Ye, and De-Chuan Zhan. Expandable subspace ensemble for pre-trained model-based class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 23554–23564, 2024.

Da-Wei Zhou, Zi-Wen Cai, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Revisiting class-incremental learning with pre-trained models: Generalizability and adaptivity are all you need. *International Journal of Computer Vision*, 133(3):1012–1032, 2025.

A Full AIR Training Procedure

Algorithm 1 summarizes the AIR pipeline. At each online task, AIR merges each incoming mini-batch with a replay mini-batch, when replay is used, to form a mixed optimization batch. It then performs the mini-batch update by combining the standard classification loss on all mixed samples with historical geometry anchoring on replayed old-class samples. After each gradient update, AIR refreshes the replay buffer by reservoir sampling. After the final mini-batch, AIR recalibrates the classifier from class prototypes computed on $\mathcal{M}_t$ in the updated feature space. This summary makes explicit where the two AIR components enter the task-level workflow and how they interact within one complete online task.

Algorithm 1 Online training procedure of AIR

Require: Stream batches $\{\mathcal{B}_{t,k}\}$, budget M, frozen encoder $\mathbf{\Phi}$, adapters $\mathbf{\Theta}$, current classifier $\mathbf{W}$ and scale σ

Initialize $\mathcal{M}_0 \leftarrow \varnothing$
for $t = 1$ to T **do**
  Fix historical anchors $\bar{\mathbf{W}}_{t-1} \leftarrow \mathbf{W}_{t-1}$ for $t > 1$
  Expand $\mathbf{W}$ to $\mathcal{Y}_t$; initialize $\mathcal{M}_{t,0} \leftarrow \mathcal{M}_{t-1}$
  for $k = 1$ to K_t **do**
    Receive $\mathcal{B}_{t,k}$ and sample $\mathcal{R}_{t,k}$ from $\mathcal{M}_{t,k-1}$ ($\mathcal{R}_{1,k} = \varnothing$)
    $\mathcal{U}_{t,k} \leftarrow \mathcal{B}_{t,k} \cup \mathcal{R}_{t,k}$
    Compute $\mathcal{L}_{\text{cls}}^{(t,k)}$ on $\mathcal{U}_{t,k}$
    Compute $\mathcal{L}_{\text{emb}}^{(t,k)}$ on $\mathcal{R}_{t,k}^{\text{old}}$ using $\bar{\mathbf{W}}_{t-1}$ (zero if empty)
    $\mathcal{L}^{(t,k)} \leftarrow \mathcal{L}_{\text{cls}}^{(t,k)} + \lambda_{\text{emb}} \mathcal{L}_{\text{emb}}^{(t,k)}$
    Update $\mathbf{\Theta}$, $\mathbf{W}$, σ with one gradient step on $\mathcal{L}^{(t,k)}$
    $\mathcal{M}_{t,k} \leftarrow \text{ReservoirUpdate}(\mathcal{M}_{t,k-1}, \mathcal{B}_{t,k}; M)$
  end for
  Let $\mathcal{M}_t = \mathcal{M}_{t,K_t}$
  for each $c \in \mathcal{Y}_t$ with $\mathcal{M}_t^c \neq \varnothing$ **do**
    Compute prototype $\mathbf{p}_{t,c}$ from $\mathcal{M}_t^c$ using $\mathbf{h}_t$; set $\mathbf{w}_{t,c} \leftarrow \mathbf{p}_{t,c}$
  end for
end for

B Full Component Analysis Across Replay Budgets

Table 5 provides the full three-seed ablation over $M \in \{1000, 5000, 10000\}$, showing the same phenomenon as memory grows: the plain Adapter+ER endpoint continues to perform poorly, and AIR retains a consistent advantage in CAA and FAA over the single-component variants.

Table 5: Complete component analysis across replay budgets. Each block reports three-seed means under the single-epoch OCIL protocol, and each metric cell reports CAA / FAA.

Budget	Variant	HGA	MGC	CIFAR-100 CAA / FAA	ImageNet-R CAA / FAA	ImageNet-A CAA / FAA	CUB-200 CAA / FAA
1000	Adapter+ER	×	×	66.23 / 55.88	38.88 / 25.09	26.91 / 17.62	46.26 / 31.96
	Adapter+ER + HGA	✓	×	69.80 / 61.85	43.47 / 32.58	31.47 / 23.24	46.76 / 32.91
	Adapter+ER + MGC	×	✓	90.34 / 84.93	72.87 / 65.05	54.69 / 44.83	83.10 / 72.17
	AIR	✓	✓	**91.11 / 85.62**	**75.01 / 67.33**	**63.38 / 51.70**	**85.48 / 75.84**
5000	Adapter+ER	×	×	66.67 / 56.16	39.11 / 27.11	28.04 / 19.31	46.42 / 31.49
	Adapter+ER + HGA	✓	×	70.47 / 63.22	45.08 / 37.43	32.17 / 26.99	47.26 / 33.80
	Adapter+ER + MGC	×	✓	91.49 / 87.73	77.23 / 73.43	60.75 / 55.21	88.30 / 82.85
	AIR	✓	✓	**92.50 / 88.94**	**80.09 / 77.13**	**68.28 / 61.77**	**89.89 / 86.09**
10000	Adapter+ER	×	×	67.33 / 57.62	40.88 / 27.31	28.09 / 19.77	46.36 / 31.09
	Adapter+ER + HGA	✓	×	70.57 / 63.31	46.37 / 37.29	32.07 / 26.13	47.54 / 33.69
	Adapter+ER + MGC	×	✓	91.66 / 88.31	77.53 / 74.45	60.64 / 54.95	88.72 / 84.35
	AIR	✓	✓	**92.69 / 89.72**	**80.60 / 77.71**	**68.36 / 62.43**	**90.15 / 86.73**

C RUNTIME COMPARISONS

Table 6 summarizes the computational profile of the evaluated methods at $M = 5000$. Each dataset block reports average training, inference, and end-to-end wall-clock seconds per online task. All experiments are performed on the same machine with a single NVIDIA 4090 GPU.

Table 6: Per-dataset efficiency summary of the $M = 5000$ OCIL runs. Each numeric cell reports wall-clock seconds per online task within the corresponding dataset.

Method	CIFAR-100			CUB-200			ImageNet-A			ImageNet-R		
	Train	Infer	Total	Train	Infer	Total	Train	Infer	Total	Train	Infer	Total
DER	113.66	67.79	181.45	34.46	27.76	62.22	27.42	23.74	51.16	61.89	46.86	108.75
iCaRL	65.66	24.85	90.51	26.49	18.17	44.67	22.85	17.74	40.59	42.46	21.32	63.78
MEMO	38.96	27.71	66.67	22.08	18.47	40.55	20.35	17.63	37.98	29.60	23.07	52.68
TagFex	44.97	24.50	69.47	23.05	18.19	41.24	20.71	17.60	38.32	34.48	21.17	55.65
ER-ACE	57.16	24.94	82.11	25.23	18.47	43.70	22.03	17.78	39.81	38.95	21.86	60.81
SCR	86.14	34.48	120.62	30.91	34.16	65.07	25.69	30.18	55.87	53.92	42.89	96.82
L2P-R	62.45	32.50	94.95	25.52	19.55	45.08	21.90	18.26	40.17	41.12	25.89	67.00
EASE-R	170.84	73.82	244.66	136.99	28.93	165.92	131.89	24.50	156.39	155.77	50.46	206.23
MOS-R	135.72	219.19	354.91	66.27	53.06	119.33	59.75	39.91	99.66	98.09	136.93	235.03
AIR	72.77	25.44	98.21	46.70	18.74	65.44	42.98	18.09	61.07	59.43	22.12	81.55

AIR requires 98.21, 65.44, 61.07, and 81.55 seconds per task on CIFAR-100, CUB-200, ImageNet-A, and ImageNet-R, respectively. Its end-to-end runtime is lower than that of EASE-R and MOS-R on all four datasets.

D SENSITIVITY TO AIR HYPERPARAMETERS

We evaluate the historical-geometry anchoring weight λ_{emb} and embedding temperature τ_e on ImageNet-R under strict OCIL with $M = 5000$. Each sweep varies one hyperparameter while holding the other at its main-experiment setting, with all remaining settings unchanged. Figure 4 reports FAA for the 16 values tested in each sweep.

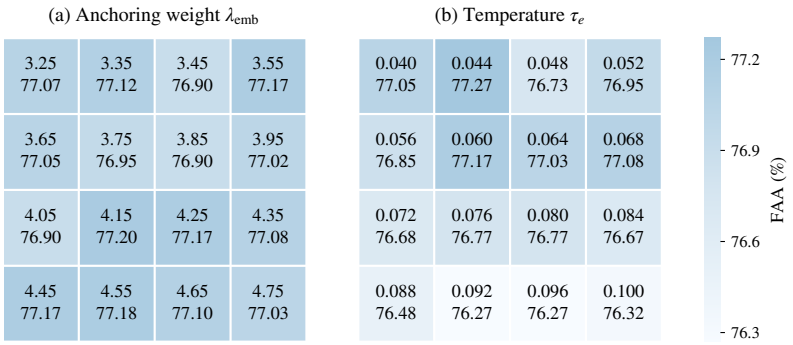

Figure 4: Sensitivity of AIR on ImageNet-R under strict OCIL ($M = 5000$). (a) Varying λ_{emb} with $\tau_e = 0.07$ fixed. (b) Varying τ_e with $\lambda_{\text{emb}} = 4.0$ fixed. The upper and lower numbers in each cell denote the tested value and FAA (%), respectively; both panels share the same FAA color scale. Values increase from left to right and top to bottom.

FAA varies from 76.90% to 77.20% over $\lambda_{\text{emb}} \in [3.25, 4.75]$, a range of 0.30 percentage points. The temperature sweep yields FAA between 76.27% and 77.27%, with the highest value at $\tau_e = 0.044$ and lower accuracy toward the upper end of the tested range. These results suggest that AIR is relatively insensitive to both hyperparameters within the tested ranges. We retain $\lambda_{\text{emb}} = 4.0$ and $\tau_e = 0.07$ for the main experiments.

E Definition of Metrics

We follow the same evaluation protocol used by PROL (Ma'sum et al., 2025) for the three summary metrics reported throughout the paper. Let $a_{i,t}$ denote the classification accuracy on task i after the learner has finished task t, where $1 \leq i \leq t \leq T$ and T is the total number of tasks. The stage-wise average accuracy after task t is

$$\bar{a}_t = \frac{1}{t} \sum_{i=1}^{t} a_{i,t}. \tag{16}$$

Final average accuracy (FAA) is the endpoint of that trajectory,

$$\mathrm{FAA} = \bar{a}_T, \tag{17}$$

while cumulative average accuracy (CAA) averages the same task-wise accuracy over all the tasks,

$$\mathrm{CAA} = \frac{1}{T} \sum_{t=1}^{T} \bar{a}_t. \tag{18}$$

FAA therefore measures the final state after the last task, whereas CAA rewards methods that stay strong throughout the whole stream.

For forgetting, we use the same aggregation as PROL and compute, for each old task, how much its final accuracy drops relative to its best value reached anywhere along the stream. The final forgetting measure is

$$\mathrm{FM} = \frac{1}{T-1} \sum_{i=1}^{T-1} \left(\max_{t \in \{i,\ldots,T\}} a_{i,t} - a_{i,T} \right). \tag{19}$$

In our implementation, this FM is computed directly from the task-grouped accuracy records saved after each online stage. Higher FAA and CAA are better, while lower FM indicates stronger retention.

F Replay Implementations of EASE and MOS

EASE-R. We retain EASE's adapter expansion and reweighted cosine classification (Zhou et al., 2024). Each task trains a new adapter while keeping earlier adapters frozen. To accommodate replay labels, the training proxy classifier spans all seen classes $\mathcal{Y}_t$, and cross-entropy is computed jointly on current and replay examples. After training, retained examples of each new class are encoded through every available adapter to populate that class's classifier blocks. Retained examples of each old class are encoded through the newly trained adapter to populate its new feature block; previously constructed old-class blocks are preserved. These blocks use empirical class means from $\mathcal{M}_t$, replacing similarity-based synthesis with direct replay estimates. Prediction retains EASE's reweighting over the concatenated adapter features.

MOS-R. We retain MOS's adapter merging, orthogonality regularization, classifier correction, and prediction-based adapter routing (Sun et al., 2025b). Each stream update minimizes cross-entropy plus the original orthogonality term on current and replay features, followed by the prescribed adapter merge. At task completion, new-class means and covariances are estimated from retained examples in the resulting adapter space, while earlier class statistics are preserved. Empirical covariances are regularized to be positive definite for Gaussian sampling. From the second task onward, one epoch of classifier correction uses synthetic features sampled from the stored class distributions, with 50 features per seen class and no additional raw-image access. New-class routing prototypes are computed from retained examples using the frozen pretrained features. Inference preserves MOS's prediction-based routing and ensemble, without a supplied task identifier.

Failure at $M = 1000$. Uniform reservoir sampling does not guarantee that every class retains an example. EASE-R requires retained examples of both new and old classes to construct the classifier blocks described above; a missing class therefore leaves a required prototype undefined. MOS-R requires retained examples of each new class to estimate its mean and covariance for classifier correction, as well as its routing prototype. If a new class has no retained examples, these quantities cannot be computed. The $M = 1000$ runs encounter these missing-class failures, and strict

stream access precludes revisiting discarded training examples to recover the required quantities. Consequently, neither method completes the full set of benchmark runs at this capacity, and Table 1 omits all EASE-R and MOS-R results for $M = 1000$.

G LIMITATIONS

AIR has two practical limitations that follow directly from its replay-based design. The first is storage overhead. Even though AIR uses a bounded replay buffer rather than the full historical dataset, it still must retain past samples throughout the online stream. As the replay budget grows, this introduces additional memory cost, which can become restrictive for memory-limited devices.

The second is training-time overhead. Each online update processes not only the current-task samples but also the replayed samples, so the effective batch computation is larger than in rehearsal-free methods. AIR also includes post-update classifier calibration from the mixed old/new feature space, which adds extra computation after task optimization. As a result, AIR improves retention and accuracy at the cost of additional training time relative to simpler approaches.

H TASK-WISE RESULTS FOR THE MAIN EXPERIMENTS

The following tables report three-seed mean accuracy and forgetting (%) for all methods and replay budgets in Table 1. T10 accuracy equals FAA; the last columns report the main-experiment CAA and FM.

Table 7: Task-wise accuracy (%) on CIFAR-100 under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	CAA
1000	iCaRL	96.40	79.28	72.57	65.14	61.40	59.63	58.81	57.53	56.22	56.14	66.31
	DER	97.00	95.52	92.88	90.91	87.09	86.25	84.46	81.25	78.42	78.92	87.27
	SCR	92.23	92.95	90.80	87.60	86.44	83.03	81.59	80.47	79.16	77.92	85.22
	ER-ACE	66.67	86.70	82.36	82.60	77.61	75.86	72.71	70.77	69.03	65.70	75.00
	L2P-R	92.97	92.02	88.51	87.17	84.97	82.18	81.71	79.10	75.32	74.35	83.83
	MEMO	85.93	83.27	77.52	72.29	68.59	64.89	63.23	59.39	56.88	54.40	68.64
	TagFex	94.37	92.68	89.34	85.64	80.75	77.86	74.94	70.16	63.68	62.49	79.19
	AIR	95.27	95.13	94.51	93.31	92.04	90.81	89.54	87.72	87.11	85.62	91.11
5000	iCaRL	96.40	78.92	72.57	65.64	62.57	60.91	59.18	57.66	57.03	55.79	66.67
	DER	97.00	94.85	93.18	91.97	89.36	88.62	87.17	85.34	83.86	82.92	89.43
	SCR	92.00	94.15	91.97	89.43	87.71	86.18	85.14	83.45	82.63	81.70	87.44
	ER-ACE	66.67	87.77	86.17	84.24	80.88	80.02	78.64	77.41	76.37	74.20	79.24
	L2P-R	92.97	92.92	90.48	89.08	87.38	86.26	85.74	83.23	82.50	82.05	87.26
	MEMO	85.93	83.43	78.19	74.08	70.61	68.51	66.94	64.54	61.80	60.18	71.42
	EASE-R	97.13	95.65	94.60	93.38	92.01	91.19	90.39	89.06	88.25	87.76	91.94
	TagFex	94.37	93.18	90.93	87.98	83.63	81.45	79.78	76.40	74.32	73.84	83.59
	MOS-R	97.20	96.23	95.20	93.93	92.79	91.96	91.30	90.00	89.42	88.93	92.70
	AIR	95.37	95.13	95.21	93.91	92.99	92.34	91.52	89.96	89.62	88.94	92.50
10000	iCaRL	96.40	79.17	72.93	66.13	63.12	60.22	60.38	58.55	57.82	57.70	67.24
	DER	97.00	94.72	93.76	91.59	89.75	89.38	87.19	86.02	84.30	84.41	89.81
	SCR	92.00	93.45	92.09	90.49	87.13	86.42	84.67	83.07	81.50	81.36	87.22
	ER-ACE	66.67	64.95	84.06	83.91	81.25	81.78	80.34	77.86	77.16	77.76	77.57
	L2P-R	92.97	92.42	90.89	89.55	87.07	87.30	86.40	84.80	83.49	83.75	87.86
	MEMO	85.93	83.43	78.67	74.33	71.81	68.78	68.08	64.72	63.14	62.19	72.11
	EASE-R	97.13	95.75	94.90	93.59	92.33	91.44	90.70	89.39	88.65	88.15	92.20
	TagFex	94.37	93.18	90.51	87.92	84.01	80.93	80.59	75.85	75.39	73.53	83.63
	MOS-R	97.20	96.22	95.20	94.05	92.73	91.90	91.27	90.19	89.39	89.04	92.72
	AIR	95.37	95.37	94.87	94.04	92.71	92.41	91.96	90.65	89.80	89.72	92.69

Table 8: Task-wise forgetting (%) on CIFAR-100 under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	FM
1000	iCaRL	0.00	2.37	5.17	8.94	10.82	13.18	14.71	16.63	16.99	16.28	16.28
	DER	0.00	3.07	6.10	7.92	12.07	12.53	14.16	17.36	20.30	19.51	19.51
	SCR	0.00	1.47	4.90	7.98	8.73	12.15	13.17	13.68	14.74	16.02	16.02
	ER-ACE	0.00	2.67	6.75	4.74	9.79	10.68	14.16	15.07	16.32	19.33	19.33
	L2P-R	0.00	4.10	8.35	9.64	11.77	15.00	15.27	18.06	22.28	23.19	23.19
	MEMO	0.00	0.23	7.55	14.31	18.44	22.57	24.23	28.21	30.85	33.28	33.28
	TagFex	0.00	1.87	6.88	11.10	16.63	19.65	22.33	27.41	34.27	35.07	35.07
	AIR	0.00	0.73	1.30	2.18	2.68	3.64	4.44	5.56	5.83	7.29	7.29
5000	iCaRL	0.00	2.13	4.73	8.02	9.04	11.53	14.11	16.30	15.84	16.38	16.38
	DER	0.00	4.17	5.22	6.19	8.94	9.28	10.51	12.04	13.50	14.42	14.42
	SCR	0.00	0.37	3.65	5.83	6.99	8.33	8.96	10.00	10.58	11.50	11.50
	ER-ACE	0.00	1.73	3.08	3.98	6.14	6.32	7.09	6.78	7.43	9.37	9.37
	L2P-R	0.00	3.10	5.50	6.92	8.78	9.79	10.23	12.71	13.50	13.90	13.90
	MEMO	0.00	0.00	7.63	11.90	16.02	18.29	19.69	22.17	25.18	26.63	26.63
	EASE-R	0.00	1.73	1.87	2.20	2.75	2.97	3.07	3.48	3.58	3.65	3.65
	TagFex	0.00	1.20	4.35	7.86	12.95	14.97	16.49	20.04	21.96	22.11	22.11
	MOS-R	0.00	0.93	1.48	2.36	3.06	3.36	3.67	4.11	4.40	4.57	4.57
	AIR	0.00	0.73	1.03	1.97	2.16	2.53	3.06	3.95	3.90	4.47	4.47
10000	iCaRL	0.00	2.37	4.65	7.69	8.72	12.48	12.92	15.56	15.20	14.56	14.56
	DER	0.00	4.83	4.75	6.81	8.35	8.26	10.33	11.22	12.79	12.54	12.54
	SCR	0.00	0.13	2.48	4.02	7.44	7.81	9.18	10.07	11.52	11.58	11.58
	ER-ACE	0.00	5.60	2.38	2.82	5.29	4.34	5.19	6.50	5.79	6.01	6.01
	L2P-R	0.00	3.13	4.88	6.01	8.75	8.19	8.93	10.42	11.93	11.64	11.64
	MEMO	0.00	0.00	6.43	11.02	13.84	17.23	17.59	21.38	23.06	23.86	23.86
	EASE-R	0.00	1.43	1.63	2.21	2.77	3.13	3.29	3.60	3.67	3.73	3.73
	TagFex	0.00	1.20	5.12	7.98	12.54	15.61	15.57	20.69	20.62	22.23	22.23
	MOS-R	0.00	1.03	1.55	2.22	3.13	3.42	3.73	3.80	4.18	4.36	4.36
	AIR	0.00	0.40	1.07	1.56	2.23	2.08	2.13	3.09	3.63	3.64	3.64

Table 9: Task-wise accuracy (%) on ImageNet-R under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	CAA
1000	iCaRL	80.57	52.94	41.86	37.73	34.26	30.95	30.22	28.25	26.11	25.11	38.80
	DER	83.55	82.74	78.56	74.70	70.16	68.76	65.45	63.78	60.83	61.47	71.00
	SCR	80.82	81.15	78.49	74.24	69.09	67.05	64.93	62.16	58.73	56.59	69.33
	ER-ACE	81.20	78.18	71.93	66.08	64.07	60.98	59.03	58.60	57.00	54.66	65.17
	L2P-R	81.56	77.51	73.23	68.15	64.72	62.60	60.65	57.98	54.21	53.12	65.37
	MEMO	80.80	71.74	67.21	61.82	56.27	55.02	52.28	51.08	46.98	45.23	58.84
	TagFex	81.64	77.59	70.19	65.41	59.36	55.93	50.96	48.56	45.04	42.75	59.74
	AIR	80.99	82.63	80.31	78.30	75.65	73.88	72.43	70.46	68.08	67.33	75.01
5000	iCaRL	80.57	52.95	43.99	36.76	33.47	30.97	30.11	27.78	27.20	27.02	39.08
	DER	83.55	83.70	78.63	75.26	72.51	71.91	68.73	68.70	66.32	66.39	73.57
	SCR	80.76	81.88	79.28	75.95	74.16	72.52	70.18	69.54	66.98	66.42	73.77
	ER-ACE	81.20	77.57	72.92	71.21	66.57	65.14	62.99	62.14	59.80	59.47	67.90
	L2P-R	81.56	77.01	74.67	71.56	70.83	68.79	66.86	66.55	65.74	64.63	70.82
	MEMO	80.80	72.27	67.56	62.36	58.11	58.24	54.31	54.57	52.86	51.84	61.29
	EASE-R	89.44	85.40	82.80	80.35	78.68	77.54	76.10	75.36	74.30	73.79	79.38
	TagFex	81.64	77.14	72.52	69.49	63.74	62.77	57.96	56.66	54.14	52.52	64.86
	MOS-R	86.98	85.27	83.24	80.92	79.27	77.93	77.07	76.10	74.98	74.67	79.64
	AIR	82.61	84.21	82.53	81.18	80.10	79.39	78.95	77.94	76.89	77.13	80.09
10000	iCaRL	80.57	53.36	44.92	38.68	37.30	34.05	33.07	30.55	28.26	27.38	40.81
	DER	83.55	83.37	80.18	75.85	71.64	72.90	69.53	69.90	67.12	67.00	74.10
	SCR	80.76	82.86	79.08	75.82	73.66	71.85	70.81	69.05	67.12	66.89	73.79
	ER-ACE	81.20	78.51	72.39	70.32	68.19	66.19	63.59	62.52	60.76	59.85	68.35
	L2P-R	81.56	77.66	75.07	73.03	70.50	69.59	68.44	67.60	66.02	66.17	71.56
	MEMO	80.80	72.27	66.82	63.09	58.45	58.72	55.21	54.57	53.32	52.06	61.53
	EASE-R	89.44	85.36	82.77	80.42	78.71	77.53	76.20	75.62	74.37	73.89	79.43
	TagFex	81.64	77.14	73.09	69.10	63.45	62.51	57.75	59.28	54.73	51.91	65.06
	MOS-R	86.98	85.50	83.15	80.82	79.55	78.21	77.17	76.40	75.58	75.08	79.84
	AIR	82.61	84.97	83.08	81.94	80.88	79.77	78.82	78.68	77.52	77.71	80.60

Table 10: Task-wise forgetting (%) on ImageNet-R under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	FM
1000	iCaRL	0.00	2.53	6.78	7.56	6.77	8.47	7.55	7.77	8.38	9.33	9.33
	DER	0.00	3.53	9.41	13.09	17.16	19.18	22.38	24.58	26.68	26.20	26.20
	SCR	0.00	2.26	5.69	8.38	11.44	12.71	13.08	15.49	16.59	18.57	18.57
	ER-ACE	0.00	1.80	6.47	9.14	8.32	10.70	10.66	10.04	9.66	11.69	11.69
	L2P-R	0.00	6.59	11.35	16.88	19.90	23.36	24.73	28.84	31.62	33.45	33.45
	MEMO	0.00	0.44	5.70	11.26	17.46	20.37	22.63	24.39	27.72	30.23	30.23
	TagFex	0.00	4.85	14.13	18.29	23.99	28.46	32.96	36.16	38.13	40.71	40.71
	AIR	0.00	0.00	1.00	1.84	3.07	4.06	4.27	5.69	6.90	7.28	7.28
5000	iCaRL	0.00	2.18	5.12	7.22	7.61	8.18	7.93	9.27	8.50	8.68	8.68
	DER	0.00	1.59	8.80	12.14	14.34	15.45	18.18	18.39	19.80	19.85	19.85
	SCR	0.00	1.50	4.79	7.11	7.82	9.15	10.52	11.06	12.57	12.82	12.82
	ER-ACE	0.00	0.92	5.14	3.42	5.14	5.01	5.24	5.27	5.56	5.28	5.28
	L2P-R	0.00	7.17	8.78	12.37	12.61	15.93	17.14	18.25	18.02	19.72	19.72
	MEMO	0.00	0.68	6.04	11.45	15.56	17.04	20.47	20.92	21.48	22.82	22.82
	EASE-R	0.00	3.68	4.72	4.93	5.00	5.52	5.65	5.84	5.73	5.68	5.68
	TagFex	0.00	4.63	10.45	12.88	18.64	20.45	24.46	25.95	27.33	29.08	29.08
	MOS-R	0.00	0.24	1.88	2.91	3.65	4.24	3.91	4.28	4.58	4.52	4.52
	AIR	0.00	0.00	0.83	1.41	1.60	2.04	2.12	3.01	3.11	2.81	2.81
10000	iCaRL	0.00	2.04	4.27	6.63	4.98	7.57	7.28	8.33	8.69	9.30	9.30
	DER	0.00	2.25	6.78	11.13	15.46	14.34	17.27	17.00	18.92	18.88	18.88
	SCR	0.00	0.73	5.91	7.43	8.32	9.97	9.93	11.53	12.13	12.06	12.06
	ER-ACE	0.00	1.30	5.97	4.99	4.48	5.35	5.50	5.26	5.31	5.49	5.49
	L2P-R	0.00	6.45	8.53	10.16	12.61	14.68	15.22	16.75	17.58	17.91	17.91
	MEMO	0.00	0.68	6.93	10.68	15.22	16.34	19.34	20.84	20.93	22.88	22.88
	EASE-R	0.00	3.55	4.52	4.52	4.67	5.08	5.41	5.72	5.88	6.00	6.00
	TagFex	0.00	4.63	9.64	13.18	18.91	20.57	24.99	23.15	26.85	29.71	29.71
	MOS-R	0.00	0.00	2.25	3.17	3.25	3.79	3.74	4.00	3.94	4.05	4.05
	AIR	0.00	0.00	1.36	1.54	1.94	2.60	2.81	3.00	3.25	3.03	3.03

Table 11: Task-wise accuracy (%) on ImageNet-A under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	CAA
1000	iCaRL	59.66	36.94	29.55	25.77	23.12	20.86	19.44	18.43	17.64	17.66	26.91
	DER	67.70	61.87	59.07	51.31	49.99	48.01	42.37	43.24	38.90	41.28	50.37
	SCR	41.07	52.15	44.96	38.68	35.09	33.09	29.42	25.85	24.38	24.71	34.94
	ER-ACE	58.66	40.12	41.07	36.79	29.99	30.01	28.26	24.98	23.98	21.18	33.50
	L2P-R	57.83	52.54	46.76	44.19	41.53	39.28	36.36	34.00	32.94	30.72	41.62
	MEMO	40.19	27.19	19.21	16.48	12.87	11.24	8.79	9.79	7.81	6.87	16.04
	TagFex	62.47	55.65	49.42	45.61	40.38	37.46	34.08	32.06	28.54	26.38	41.20
	AIR	76.47	73.65	69.48	67.60	64.63	61.42	58.00	57.27	53.57	51.70	63.38
5000	iCaRL	59.66	37.36	29.14	25.44	23.22	23.14	21.65	20.93	20.39	19.31	28.02
	DER	67.70	59.39	55.05	54.91	53.28	48.74	44.94	42.45	39.63	40.24	50.63
	SCR	41.07	51.05	44.01	43.02	37.08	39.28	33.68	32.63	28.58	28.48	37.89
	ER-ACE	58.66	42.45	38.31	35.57	32.53	31.06	27.94	23.91	23.41	23.28	33.71
	L2P-R	57.83	50.89	46.64	45.36	43.73	39.66	38.88	37.92	35.57	35.42	43.19
	MEMO	40.19	27.19	17.67	15.68	13.54	10.72	9.86	8.54	8.03	7.00	15.84
	EASE-R	78.19	69.29	63.22	59.69	57.03	56.01	53.56	50.93	49.05	47.84	58.48
	TagFex	62.47	55.65	49.11	45.01	39.71	37.79	35.50	31.91	29.40	28.72	41.53
	MOS-R	67.70	67.52	66.56	64.11	62.81	60.59	59.18	58.68	56.13	55.17	61.85
	AIR	76.47	72.76	71.83	71.44	68.63	66.79	65.84	64.38	62.92	61.77	68.28
10000	iCaRL	59.66	37.36	29.14	25.44	23.22	23.14	21.65	20.93	20.39	19.75	28.07
	DER	67.70	59.39	55.05	54.91	53.28	48.74	44.94	42.45	42.89	40.16	50.95
	SCR	41.07	51.05	44.01	43.02	37.08	39.28	33.68	32.63	30.25	29.52	38.16
	ER-ACE	58.66	42.45	38.31	35.57	32.53	31.06	27.94	23.91	23.68	23.17	33.73
	L2P-R	57.83	50.89	46.64	45.36	43.73	39.66	38.88	37.92	35.39	33.82	43.01
	MEMO	40.19	27.19	17.67	15.68	13.54	10.72	9.86	8.54	8.03	7.35	15.88
	EASE-R	78.19	69.29	63.22	59.69	57.03	56.01	53.56	50.93	49.05	47.69	58.47
	TagFex	62.47	55.65	49.11	45.01	39.71	37.79	35.50	31.91	29.40	29.14	41.57
	MOS-R	67.70	67.52	66.56	64.11	62.81	60.59	59.18	58.68	56.28	54.80	61.82
	AIR	76.47	72.76	71.83	71.44	68.63	66.79	65.84	64.38	62.99	62.43	68.36

Table 12: Task-wise forgetting (%) on ImageNet-A under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	FM
1000	iCaRL	0.00	8.12	7.03	7.19	6.95	7.13	6.69	6.75	6.59	6.12	6.12
	DER	0.00	7.91	10.38	21.63	19.91	21.64	27.59	25.65	28.89	26.03	26.03
	SCR	0.00	0.20	9.95	13.86	14.74	14.41	15.91	18.01	16.79	15.03	15.03
	ER-ACE	0.00	16.63	5.12	6.63	9.96	7.26	8.36	8.79	7.68	8.53	8.53
	L2P-R	0.00	14.05	20.33	22.11	23.64	25.43	29.23	31.32	32.13	34.32	34.32
	MEMO	0.00	3.84	12.37	15.87	18.40	21.22	23.38	22.30	24.03	25.35	25.35
	TagFex	0.00	5.45	13.13	18.23	22.94	25.89	29.21	31.17	33.58	34.57	34.57
	AIR	0.00	4.82	6.69	6.71	7.52	9.52	11.08	9.65	10.86	11.31	11.31
5000	iCaRL	0.00	7.30	5.46	7.14	5.75	5.64	5.44	5.10	5.48	6.09	6.09
	DER	0.00	13.11	15.91	17.59	18.54	22.93	26.12	28.42	29.60	27.78	27.78
	SCR	0.00	1.96	9.81	8.66	13.37	9.92	14.10	13.52	15.46	14.38	14.38
	ER-ACE	0.00	12.60	7.19	5.98	5.72	5.77	7.31	8.55	8.17	7.18	7.18
	L2P-R	0.00	16.73	18.24	21.44	21.62	25.80	27.07	27.51	29.62	29.63	29.63
	MEMO	0.00	3.84	15.60	17.13	18.22	21.50	21.60	23.16	23.16	24.57	24.57
	EASE-R	0.00	8.23	10.00	11.35	11.70	11.30	10.45	12.56	11.94	11.31	11.31
	TagFex	0.00	5.45	14.19	18.48	23.61	25.14	27.22	30.19	31.95	31.74	31.74
	MOS-R	0.00	2.06	4.21	5.48	5.98	7.79	7.80	7.54	8.34	8.12	8.12
	AIR	0.00	5.06	3.76	3.48	5.37	5.88	5.62	6.08	6.00	6.57	6.57
10000	iCaRL	0.00	7.30	5.46	7.14	5.75	5.64	5.44	5.10	5.49	5.48	5.48
	DER	0.00	13.11	15.91	17.59	18.54	22.93	26.12	28.42	25.78	28.41	28.41
	SCR	0.00	1.96	9.81	8.66	13.37	9.92	14.10	13.52	13.89	13.90	13.90
	ER-ACE	0.00	12.60	7.19	5.98	5.72	5.77	7.31	8.55	8.20	7.41	7.41
	L2P-R	0.00	16.73	18.24	21.44	21.62	25.80	27.07	27.51	29.86	31.75	31.75
	MEMO	0.00	3.84	15.60	17.13	18.22	21.50	21.60	23.16	23.16	24.46	24.46
	EASE-R	0.00	8.23	10.00	11.35	11.70	11.30	10.45	12.56	11.93	11.29	11.29
	TagFex	0.00	5.45	14.19	18.48	23.61	25.14	27.22	30.19	31.95	31.05	31.05
	MOS-R	0.00	2.06	4.21	5.48	5.98	7.79	7.80	7.54	8.39	9.34	9.34
	AIR	0.00	5.06	3.76	3.48	5.37	5.88	5.62	6.08	5.87	5.68	5.68

Table 13: Task-wise accuracy (%) on CUB-200 under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	CAA
1000	iCaRL	85.28	63.55	52.48	44.82	40.89	39.37	35.98	34.59	33.92	32.18	46.31
	DER	90.65	88.76	84.59	82.96	80.66	76.16	74.99	69.67	69.48	66.37	78.43
	SCR	89.49	88.90	85.89	84.16	79.84	78.62	74.72	72.33	71.20	67.91	79.31
	ER-ACE	91.60	88.69	82.02	78.60	78.62	74.03	69.13	69.58	65.66	65.10	76.30
	L2P-R	91.78	91.55	82.76	83.59	79.71	76.99	75.72	74.52	72.83	68.61	79.80
	MEMO	84.22	83.27	80.76	77.12	75.18	71.39	67.28	65.52	65.45	64.38	73.46
	TagFex	89.21	88.96	84.04	82.41	79.20	73.83	69.41	65.45	64.54	60.77	75.78
	AIR	95.66	94.56	90.90	88.51	86.56	84.21	80.86	79.94	77.76	75.84	85.48
5000	iCaRL	85.28	61.16	54.19	48.67	40.79	38.98	36.12	34.48	33.17	31.57	46.44
	DER	90.65	90.13	86.61	83.20	80.92	78.15	72.20	72.01	69.54	69.33	79.27
	SCR	89.49	90.51	88.11	87.18	82.77	82.52	79.78	79.04	78.29	76.84	83.45
	ER-ACE	91.60	88.42	83.36	80.91	78.03	73.55	72.31	71.45	69.28	66.75	77.57
	L2P-R	91.78	88.51	86.72	83.88	79.96	78.52	77.50	74.44	75.13	72.80	80.93
	MEMO	84.22	83.27	79.33	77.08	73.04	71.22	69.36	65.32	67.64	66.28	73.68
	EASE-R	96.75	94.43	91.82	90.74	89.63	87.70	85.95	85.20	84.65	84.05	89.09
	TagFex	89.21	88.96	84.51	81.99	77.13	73.39	71.71	67.16	65.10	61.20	76.04
	MOS-R	88.11	93.08	92.20	90.89	88.93	87.98	86.83	86.84	86.31	85.96	88.71
	AIR	95.66	94.77	92.48	90.84	90.03	88.63	87.15	86.66	86.60	86.09	89.89
10000	iCaRL	85.28	61.16	54.19	48.67	40.79	38.55	36.18	34.69	33.82	31.19	46.45
	DER	90.65	90.13	86.61	83.20	80.92	78.49	74.09	72.47	72.06	69.28	79.79
	SCR	89.49	90.51	88.11	87.18	82.77	82.29	80.58	81.23	79.58	78.63	84.04
	ER-ACE	91.60	88.42	83.36	80.91	78.03	72.10	72.41	70.34	68.34	65.96	77.15
	L2P-R	91.78	88.51	86.72	83.88	79.96	78.60	77.51	74.38	75.16	74.71	81.12
	MEMO	84.22	83.27	79.33	77.08	73.04	71.22	68.60	66.21	66.30	64.97	73.42
	EASE-R	96.75	94.43	91.82	90.74	89.63	87.77	86.17	85.51	85.12	84.34	89.23
	TagFex	89.21	88.96	84.51	81.99	77.13	73.39	71.22	69.61	65.28	62.05	76.33
	MOS-R	88.11	93.08	92.20	90.89	88.93	88.92	87.11	86.87	86.26	86.36	88.87
	AIR	95.66	94.77	92.48	90.84	90.03	89.35	87.65	86.99	87.03	86.73	90.15

Table 14: Task-wise forgetting (%) on CUB-200 under strict OCIL.

Budget	Method	T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	FM
1000	iCaRL	0.00	0.69	1.34	2.49	3.41	3.41	4.69	4.37	4.37	4.83	4.83
	DER	0.00	3.30	9.23	11.10	14.45	18.22	18.93	24.22	24.23	27.10	27.10
	SCR	0.00	0.27	4.06	5.55	10.01	9.76	11.29	12.02	11.54	13.82	13.82
	ER-ACE	0.00	3.35	7.42	7.07	4.10	6.86	8.18	6.03	8.33	7.47	7.47
	L2P-R	0.00	2.85	14.17	12.21	16.31	18.68	19.23	20.22	22.23	26.70	26.70
	MEMO	0.00	0.59	5.78	10.38	12.72	17.26	21.43	23.35	23.60	24.54	24.54
	TagFex	0.00	1.26	9.39	11.39	15.53	20.96	24.57	27.46	28.43	32.34	32.34
	AIR	0.00	0.68	4.14	5.80	7.54	8.49	10.32	9.59	10.71	11.75	11.75
5000	iCaRL	0.00	1.11	0.61	1.71	4.26	4.50	5.15	4.78	4.53	5.56	5.56
	DER	0.00	1.80	7.49	11.78	13.42	15.27	20.99	20.79	23.24	22.66	22.66
	SCR	0.00	0.00	3.92	4.35	8.79	7.33	8.82	8.47	8.43	9.42	9.42
	ER-ACE	0.00	3.04	5.23	6.31	6.20	7.34	6.81	6.28	8.58	9.24	9.24
	L2P-R	0.00	5.59	7.86	10.39	14.48	15.41	15.85	18.69	18.13	20.79	20.79
	MEMO	0.00	0.59	7.14	9.85	15.63	17.18	18.92	23.40	20.58	21.82	21.82
	EASE-R	0.00	0.99	3.58	3.99	4.67	5.66	5.66	5.53	5.52	5.69	5.69
	TagFex	0.00	1.26	8.02	12.25	18.11	21.27	21.84	26.24	28.13	31.89	31.89
	MOS-R	0.00	0.00	1.08	2.54	3.84	4.36	4.64	4.23	4.58	4.74	4.74
	AIR	0.00	0.68	2.22	3.41	4.00	4.61	4.98	4.99	4.85	5.10	5.10
10000	iCaRL	0.00	1.11	0.61	1.71	4.26	4.97	5.26	4.76	4.26	6.01	6.01
	DER	0.00	1.80	7.49	11.78	13.42	15.12	19.59	20.43	20.46	23.29	23.29
	SCR	0.00	0.00	3.92	4.35	8.79	7.70	8.27	6.93	7.92	8.43	8.43
	ER-ACE	0.00	3.04	5.23	6.31	6.20	9.72	5.82	6.67	6.31	7.80	7.80
	L2P-R	0.00	5.59	7.86	10.39	14.48	15.32	16.27	19.52	18.65	18.98	18.98
	MEMO	0.00	0.59	7.14	9.85	15.63	17.18	19.91	22.01	21.98	23.60	23.60
	EASE-R	0.00	0.99	3.58	3.99	4.67	5.49	5.47	5.47	5.57	5.86	5.86
	TagFex	0.00	1.26	8.02	12.25	18.11	21.27	22.06	23.22	27.82	30.71	30.71
	MOS-R	0.00	0.00	1.08	2.54	3.84	3.62	4.35	4.50	5.18	4.69	4.69
	AIR	0.00	0.68	2.22	3.41	4.00	4.08	4.95	4.97	4.74	4.88	4.88

I Comparison of MGC-Based Variants

We compare MGC + DER++ (Buzzega et al., 2020), MGC + LwF (Li & Hoiem, 2017), MGC + PCR Lin et al. (2023b), and AIR under strict OCIL at $M = 5000$, using an input batch size of 10 and a replay batch size of 10. MGC + PCR denotes the AIR variant with HGA replaced by the PCR loss while retaining MGC, rather than standalone PCR. All results are averaged over three seeds.

Table 15: Strict OCIL comparison of MGC-based variants and AIR at $M = 5000$, reporting three-seed mean FAA/CAA/FM (%). Higher FAA/CAA and lower FM are better. Bold indicates the best value within each dataset and metric.

Method	CIFAR-100			ImageNet-R			ImageNet-A			CUB-200		
	FAA ↑	CAA ↑	FM ↓	FAA ↑	CAA ↑	FM ↓	FAA ↑	CAA ↑	FM ↓	FAA ↑	CAA ↑	FM ↓
MGC + DER++	87.15	91.42	6.26	72.55	77.02	4.83	55.17	61.40	6.43	83.67	89.52	6.04
MGC + LwF	87.38	91.76	5.84	73.02	77.18	4.43	57.21	62.25	**5.54**	84.05	89.55	5.64
MGC + PCR	88.92	**92.83**	5.75	76.77	**81.03**	6.49	60.06	66.43	8.13	85.04	89.46	5.51
AIR	**88.94**	92.50	**4.47**	**77.13**	80.09	**2.81**	**61.77**	**68.28**	6.57	**86.09**	**89.89**	**5.10**

AIR has the highest reported FAA on all four datasets. MGC + PCR has higher CAA on CIFAR-100 and ImageNet-R, whereas AIR has higher CAA on ImageNet-A and CUB-200. AIR has the lowest FM on three datasets.

J Confidence Intervals for the Main Replay Comparison

Tables 16 and 17 supplement Table 1 with two-sided 95% Student-t confidence intervals for the three-run means.

Table 16: Mean ± 95% confidence-interval half-width (%) on CIFAR-100 and ImageNet-R. Methods and replay budgets match Table 1.

Budget	Method	CIFAR-100			ImageNet-R		
		FAA	CAA	FM	FAA	CAA	FM
1000	iCaRL	56.14 ± 3.50	66.31 ± 5.59	16.28 ± 4.15	25.11 ± 8.00	38.80 ± 7.78	9.33 ± 1.22
	DER	78.92 ± 2.56	87.27 ± 1.74	19.51 ± 2.41	61.47 ± 2.91	71.00 ± 0.57	26.20 ± 2.76
	SCR	77.92 ± 2.27	85.22 ± 1.30	16.02 ± 2.14	56.59 ± 4.95	69.33 ± 1.33	18.57 ± 2.68
	ER-ACE	65.70 ± 4.87	75.00 ± 14.91	19.33 ± 5.47	54.66 ± 2.08	65.17 ± 1.67	11.69 ± 3.90
	L2P-R	74.35 ± 4.71	83.83 ± 0.74	23.19 ± 5.40	53.12 ± 3.32	65.37 ± 1.36	33.45 ± 4.75
	MEMO	54.40 ± 6.88	68.64 ± 4.50	33.28 ± 4.22	45.23 ± 9.49	58.84 ± 5.02	30.23 ± 7.10
	TagFex	62.49 ± 5.37	79.19 ± 0.52	35.07 ± 6.68	42.75 ± 3.50	59.74 ± 3.01	40.71 ± 2.93
	AIR	85.62 ± 0.82	91.11 ± 0.98	7.29 ± 0.80	67.33 ± 0.77	75.01 ± 1.25	7.28 ± 3.01
5000	iCaRL	55.79 ± 6.06	66.67 ± 6.09	16.38 ± 5.84	27.02 ± 2.29	39.08 ± 6.78	8.68 ± 4.79
	DER	82.92 ± 2.09	89.43 ± 0.50	14.42 ± 1.71	66.39 ± 2.09	73.57 ± 0.57	19.85 ± 2.06
	SCR	81.70 ± 1.72	87.44 ± 0.87	11.50 ± 3.28	66.42 ± 3.34	73.77 ± 2.40	12.82 ± 2.69
	ER-ACE	74.20 ± 4.78	79.24 ± 12.25	9.37 ± 8.13	59.47 ± 2.70	67.90 ± 0.54	5.28 ± 1.83
	L2P-R	82.05 ± 0.66	87.26 ± 1.52	13.90 ± 0.49	64.63 ± 1.17	70.82 ± 0.77	19.72 ± 1.33
	MEMO	60.18 ± 6.53	71.42 ± 5.66	26.63 ± 3.68	51.84 ± 8.82	61.29 ± 5.12	22.82 ± 6.93
	EASE-R	87.76 ± 0.47	91.94 ± 0.33	3.65 ± 1.41	73.79 ± 0.34	79.38 ± 0.50	5.68 ± 1.50
	TagFex	73.84 ± 3.63	83.59 ± 1.89	22.11 ± 4.89	52.52 ± 3.06	64.86 ± 3.03	29.08 ± 2.78
	MOS-R	88.93 ± 0.85	92.70 ± 1.05	4.57 ± 0.30	74.67 ± 3.11	79.64 ± 2.39	4.52 ± 0.22
	AIR	88.94 ± 1.07	92.50 ± 1.38	4.47 ± 2.17	77.13 ± 1.04	80.09 ± 1.70	2.81 ± 1.10
10000	iCaRL	57.70 ± 3.33	67.24 ± 4.72	14.56 ± 4.25	27.38 ± 3.80	40.81 ± 8.50	9.30 ± 5.19
	DER	84.41 ± 1.52	89.81 ± 0.57	12.54 ± 1.44	67.00 ± 1.57	74.10 ± 1.44	18.88 ± 1.81
	SCR	81.36 ± 1.50	87.22 ± 0.88	11.58 ± 2.37	66.89 ± 2.23	73.79 ± 0.98	12.06 ± 2.73
	ER-ACE	77.76 ± 1.96	77.57 ± 23.77	6.01 ± 4.44	59.85 ± 5.07	68.35 ± 3.13	5.49 ± 3.89
	L2P-R	83.75 ± 1.48	87.86 ± 0.71	11.64 ± 1.15	66.17 ± 2.95	71.56 ± 0.93	17.91 ± 3.61
	MEMO	62.19 ± 6.24	72.11 ± 5.32	23.86 ± 5.49	52.06 ± 8.99	61.53 ± 5.32	22.88 ± 6.51
	EASE-R	88.15 ± 0.83	92.20 ± 0.78	3.73 ± 1.73	73.89 ± 1.13	79.43 ± 1.42	6.00 ± 1.26
	TagFex	73.53 ± 5.37	83.63 ± 2.11	22.23 ± 5.34	51.91 ± 3.15	65.06 ± 2.51	29.71 ± 5.09
	MOS-R	89.04 ± 0.69	92.72 ± 0.89	4.36 ± 0.37	75.08 ± 2.65	79.84 ± 2.38	4.05 ± 0.97
	AIR	89.72 ± 0.64	92.69 ± 0.63	3.64 ± 1.23	77.71 ± 0.77	80.60 ± 0.57	3.03 ± 1.96

Table 17: Mean ± 95% confidence-interval half-width (%) on ImageNet-A and CUB-200. Methods and replay budgets match Table 1.

Budget	Method	ImageNet-A			CUB-200		
		FAA	CAA	FM	FAA	CAA	FM
1000	iCaRL	17.66 ± 6.58	26.91 ± 5.29	6.12 ± 0.20	32.18 ± 1.09	46.31 ± 2.51	4.83 ± 0.82
	DER	41.28 ± 1.74	50.37 ± 2.53	26.03 ± 6.83	66.37 ± 3.63	78.43 ± 3.68	27.10 ± 4.47
	SCR	24.71 ± 2.78	34.94 ± 2.70	15.03 ± 8.99	67.91 ± 3.38	79.31 ± 5.69	13.82 ± 3.47
	ER-ACE	21.18 ± 3.47	33.50 ± 4.75	8.53 ± 3.76	65.10 ± 4.02	76.30 ± 3.03	7.47 ± 3.12
	L2P-R	30.72 ± 3.96	41.62 ± 0.55	34.32 ± 3.53	68.61 ± 3.95	79.80 ± 1.63	26.70 ± 5.49
	MEMO	6.87 ± 1.24	16.04 ± 6.73	25.35 ± 8.35	64.38 ± 14.36	73.46 ± 13.61	24.54 ± 6.04
	TagFex	26.38 ± 3.88	41.20 ± 3.48	34.57 ± 6.81	60.77 ± 8.10	75.78 ± 3.08	32.34 ± 11.00
	AIR	51.70 ± 4.91	63.38 ± 4.45	11.31 ± 2.62	75.84 ± 3.43	85.48 ± 3.32	11.75 ± 2.57
5000	iCaRL	19.31 ± 4.27	28.02 ± 1.79	6.09 ± 1.12	31.57 ± 4.57	46.44 ± 4.07	5.56 ± 4.97
	DER	40.24 ± 3.73	50.63 ± 2.04	27.78 ± 1.86	69.33 ± 6.01	79.27 ± 4.72	22.66 ± 5.69
	SCR	28.48 ± 6.80	37.89 ± 2.73	14.38 ± 1.79	76.84 ± 4.38	83.45 ± 3.62	9.42 ± 5.48
	ER-ACE	23.28 ± 3.40	33.71 ± 7.14	7.18 ± 5.15	66.75 ± 7.25	77.57 ± 5.06	9.24 ± 5.06
	L2P-R	35.42 ± 4.35	43.19 ± 1.15	29.63 ± 1.53	72.80 ± 2.05	80.93 ± 2.03	20.79 ± 3.91
	MEMO	7.00 ± 2.78	15.84 ± 6.93	24.57 ± 6.71	66.28 ± 11.65	73.68 ± 12.94	21.82 ± 7.75
	EASE-R	47.84 ± 2.50	58.48 ± 2.93	11.31 ± 8.27	84.05 ± 0.49	89.09 ± 2.57	5.69 ± 0.59
	TagFex	28.72 ± 5.56	41.53 ± 2.86	31.74 ± 5.96	61.20 ± 5.32	76.04 ± 1.96	31.89 ± 9.04
	MOS-R	55.17 ± 6.69	61.85 ± 3.71	8.12 ± 4.59	85.96 ± 5.76	88.71 ± 5.48	4.74 ± 1.52
	AIR	61.77 ± 1.39	68.28 ± 3.50	6.57 ± 1.06	86.09 ± 0.75	89.89 ± 2.89	5.10 ± 0.62
10000	iCaRL	19.75 ± 2.38	28.07 ± 1.89	5.48 ± 0.15	31.19 ± 4.25	46.45 ± 3.33	6.01 ± 5.47
	DER	40.16 ± 3.43	50.95 ± 1.12	28.41 ± 8.87	69.28 ± 6.19	79.79 ± 3.83	23.29 ± 7.65
	SCR	29.52 ± 4.97	38.16 ± 2.25	13.90 ± 5.63	78.63 ± 2.12	84.04 ± 4.12	8.43 ± 2.42
	ER-ACE	23.17 ± 1.46	33.73 ± 7.08	7.41 ± 4.42	65.96 ± 2.64	77.15 ± 3.98	7.80 ± 2.78
	L2P-R	33.82 ± 4.08	43.01 ± 1.23	31.75 ± 0.99	74.71 ± 4.01	81.12 ± 1.67	18.98 ± 3.06
	MEMO	7.35 ± 4.99	15.88 ± 7.13	24.46 ± 4.50	64.97 ± 5.49	73.42 ± 12.05	23.60 ± 0.87
	EASE-R	47.69 ± 2.22	58.47 ± 2.98	11.29 ± 8.05	84.34 ± 0.11	89.23 ± 2.59	5.86 ± 1.36
	TagFex	29.14 ± 3.25	41.57 ± 2.88	31.05 ± 5.02	62.05 ± 4.55	76.33 ± 1.34	30.71 ± 9.37
	MOS-R	54.80 ± 6.76	61.82 ± 3.92	9.34 ± 1.41	86.36 ± 6.36	88.87 ± 5.45	4.69 ± 0.94
	AIR	62.43 ± 0.34	68.36 ± 3.44	5.68 ± 1.62	86.73 ± 0.93	90.15 ± 2.70	4.88 ± 1.46