INTERFERENCE-AWARE CLASSIFIER EXPANSION FOR EXEMPLAR-FREE CLASS-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Exemplar-free class-incremental learning (EFCIL) requires a model to continually expand its decision space while preserving previously acquired knowledge without access to old training samples. Existing methods improve the features, statistics, or structures used to construct the expanded decision space, but comparatively less attention has been paid to the initial geometry of newly introduced trainable classifier weights. We argue that uncontrolled initialization can induce old–new classifier interference at each incremental transition, whereas enforcing a predefined geometry throughout training may restrict subsequent adaptation. We propose Interference-Aware Classifier Expansion (IACE), which conditions classifier expansion without imposing a persistent geometric constraint. IACE introduces Orthogonal Classifier Initialization (OCI) to initialize newly added classifier weights orthogonally to existing classifier directions. Orthogonality is applied only at initialization and released during subsequent optimization, providing a controlled starting geometry while allowing the classifier to adapt freely to incoming data. To preserve prior knowledge during this adaptation, IACE also employs Adaptive-Temperature Knowledge Distillation (ATKD), which adjusts the distillation temperature based on the teacher response at each incremental stage. OCI controls how new classifier directions enter the existing decision space, while ATKD calibrates the old-knowledge signal that guides subsequent learning. Extensive experiments on standard EFCIL benchmarks demonstrate that IACE achieves state-of-the-art performance across different incremental settings.

1 INTRODUCTION

Exemplar-free class-incremental learning (EFCIL) aims to continuously acquire new classes without retaining training samples from previous tasks Zhu et al. (2022); Magistri et al. (2024). Without old exemplars, each incremental task must learn new classes from current data while preserving knowledge acquired from previous tasks. Meanwhile, the decision space must expand as the label space grows. EFCIL therefore requires the model to repeatedly incorporate new class decisions into an already learned space while retaining old knowledge, making it particularly vulnerable to representation drift, decision bias, and catastrophic forgetting.

We consider the common setting of an expanding decision space, where the classifier or decision rule grows as new classes arrive. This setting naturally accommodates a growing label space without requiring the complete future class set to be specified in advance Pernici et al. (2020). Recent EFCIL methods have developed different mechanisms to support such expansion. FeTrIL reconstructs old-class pseudo-features to train a classifier over all observed classes Petit et al. (2023); DPCR estimates semantic shift and reconstructs the classifier from calibrated class statistics He et al. (2025); structure-preserving approaches maintain discriminative class relations as new classes are incorporated Zhang et al. (2026); and transport-based methods align historical class statistics with the evolving representation Xu & Krawczyk (2026a;b). These methods substantially improve the features, statistics, or structures used to construct the expanded decision space.

When expansion is implemented by introducing trainable classifier weights for new classes, however, a preceding factor remains comparatively less studied: their initialization relative to the existing classifier. Independent initialization does not explicitly control the realized relation between newly introduced and existing classifier directions, and may therefore cause undesirable old–new classifier

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.

interference at the beginning of optimization. This issue is particularly important in EFCIL because old samples are unavailable to directly correct such interference. We therefore ask: How can newly introduced classifier directions be initialized to reduce old–new classifier interference while remaining free to adapt during subsequent optimization?

Our key insight is to control the transient geometry of classifier expansion rather than prescribe its final geometry. Persistent geometric constraints can explicitly organize the class space, but also restrict how that space may evolve as new classes are learned. Instead, we use geometry only to establish a favorable initial relation between new and existing classifier directions, and release this constraint once optimization begins. Geometry therefore serves as an initialization prior: it controls where classifier expansion starts, rather than where it must converge.

Controlling the starting geometry addresses only new-class integration. During subsequent learning, old knowledge must still be transferred from the previous model in the absence of old exemplars. Knowledge distillation provides this complementary signal, but its effectiveness depends on the teacher response. Prior EFCIL studies have shown that teacher confidence can substantially affect knowledge transfer Gao et al. (2024), while recent methods have further explored adaptive distillation and dynamic-temperature strategies Luo et al. (2025). Our goal is therefore not to introduce dynamic temperature itself. Instead, we adapt the distillation temperature according to the teacher response observed at the current incremental stage, providing a complementary mechanism for calibrating old-knowledge transfer. We consequently view each task transition as involving two coupled requirements: new-class integration and old-knowledge transfer.

Based on this motivation, we propose Interference-Aware Classifier Expansion (IACE). For new-class integration, IACE introduces Orthogonal Classifier Initialization (OCI), which initializes newly added classifier weights orthogonally to the existing classifier directions. Crucially, orthogonality is imposed only at initialization and is released during subsequent optimization, allowing the classifier geometry to evolve according to the data. For old-knowledge transfer, IACE employs Adaptive-Temperature Knowledge Distillation (ATKD), which adjusts the distillation temperature according to the teacher response at the current incremental stage. OCI therefore controls the transient geometry from which new classifier directions begin learning, while ATKD calibrates the old-knowledge signal that guides subsequent optimization. IACE conditions both sides of the incremental transition without imposing a persistent geometric constraint on the learned classifier.

Our contributions are summarized as follows:

- We revisit EFCIL with an expanding decision space and highlight the transient initialization geometry of newly introduced trainable classifier weights as an important factor affecting old–new classifier interference during incremental learning.
- We propose Interference-Aware Classifier Expansion (IACE), which combines Orthogonal Classifier Initialization (OCI) for low-interference new-class integration with Adaptive-Temperature Knowledge Distillation (ATKD) for teacher-response-adaptive old-knowledge transfer, without enforcing orthogonality during subsequent optimization.
- Extensive experiments on standard EFCIL benchmarks demonstrate that IACE consistently surpasses state-of-the-art methods across different incremental settings.

2 Related Work

Exemplar-free class-incremental learning. Without old training examples, EFCIL methods preserve earlier knowledge through a previous model or compact class-level information. Distillation and feature regularization constrain changes to old predictions or representations (Li & Hoiem, 2018; Zhu et al., 2022; Magistri et al., 2024). Prototype and distribution methods retain summaries that support old-class recognition (Zhu et al., 2021; Goswami et al., 2023). Other approaches rely on frozen representations or analytic classifier updates (Petit et al., 2023; Zhuang et al., 2022; 2024).

When the feature extractor continues to adapt, statistics from earlier tasks must remain useful in the new feature space. SDC estimates prototype shifts, ADC uses adversarial samples to estimate old-class drift, and LDC learns a mapping between successive representations (Yu et al., 2020; Goswami et al., 2024; Gomez-Villa et al., 2024). DPCR estimates semantic shift through dual projection, while Dual-E separates global and local shifts (He et al., 2025; Xu et al., 2026). BiCyc and GATF further

align retained class information with evolving representations (Xu & Krawczyk, 2026a;b). These methods improve the information used to preserve old decisions as the feature space changes.

Classifier expansion and geometry. As the label space grows, the classifier or decision rule must accommodate incoming classes. FeTrIL trains a classifier using reconstructed old-class features (Petit et al., 2023); DPCR reconstructs one from transported class statistics (He et al., 2025). Structure-preserving methods also maintain discriminative relations among classes during expansion (Zhang et al., 2026). These choices shape the expanded decision space.

Classifier geometry can also be specified directly. Pre-allocated fixed classifiers reserve directions for future classes (Pernici et al., 2020), while ProNC progressively extends geometric targets as classes arrive (Wang et al., 2026). In a growing classifier, however, the initial orientation of newly introduced weights relative to existing weights also determines their starting interaction. Independent initialization does not explicitly control this old–new relation, and old examples are unavailable to correct overlap directly. OCI uses current-class prototypes and the retained classifier span to initialize new directions orthogonally. This establishes a task-local starting geometry without prescribing the geometry of the reconstructed classifier.

Knowledge distillation. Classifier expansion also requires old-knowledge transfer while learning from current-task inputs. A frozen previous model supplies this signal through distillation without access to old examples (Hinton et al., 2015; Li & Hoiem, 2018; He et al., 2025). The usefulness of its soft predictions depends on their concentration: overly confident or diffuse responses can change the supervision received by the student. Prior EFCIL work studies teacher confidence (Gao et al., 2024), and DynLearn explores task-adaptive distillation and temperature scheduling (Luo et al., 2025).

ATKD uses the frozen previous model's responses to current-task inputs as feedback at each task boundary. It adjusts the scale of the reconstructed cosine classifier, thereby setting the effective distillation temperature for the following task. This teacher-response-based adjustment complements OCI's control of the new-classifier starting geometry at the same transition.

3 METHOD

3.1 PRELIMINARY: EXEMPLAR-FREE CLASS-INCREMENTAL LEARNING

In exemplar-free class-incremental learning (EFCIL), task $t \in \{1, \dots, T\}$ introduces a disjoint class set $\mathcal{C}_t$. At task t, training uses only $\mathcal{D}^t = \{(x_i, y_i) \mid y_i \in \mathcal{C}_t\}$, while the model must recognize all classes observed so far without task identifiers.

Let $f_t(x) \in \mathbb{R}^d$ denote the feature extractor and $W_t \in \mathbb{R}^{N_t \times d}$ the classifier over the N_t seen classes. Following the expandable-classifier setting, new classifier rows are added as new classes arrive ($t \geq 2$):

$$W_t = \begin{bmatrix} W_{t-1} \\ W_{\text{new}} \end{bmatrix}. \tag{1}$$

Incremental learning commonly combines new-class classification and old-knowledge distillation:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{KD}}, \tag{2}$$

where $\mathcal{L}_{\text{cls}}$ is cross-entropy on the incoming classes and λ weights distillation.

For knowledge distillation, the frozen model from task $t-1$ serves as the teacher. Let $z^{t-1}(x)$ and $z^t(x)$ denote teacher and student logits over the old classes. A conventional form with temperature τ is

$$\mathcal{L}_{\text{KD}} = \mathbb{E}_{x_i} D_{\text{KL}} \left(\text{softmax}\left(\frac{z^{t-1}(x_i)}{\tau} \right) \middle\| \text{softmax}\left(\frac{z^t(x_i)}{\tau} \right) \right). \tag{3}$$

For $t = 1$, no teacher is available, and the model is trained only with $\mathcal{L}_{\text{cls}}$.

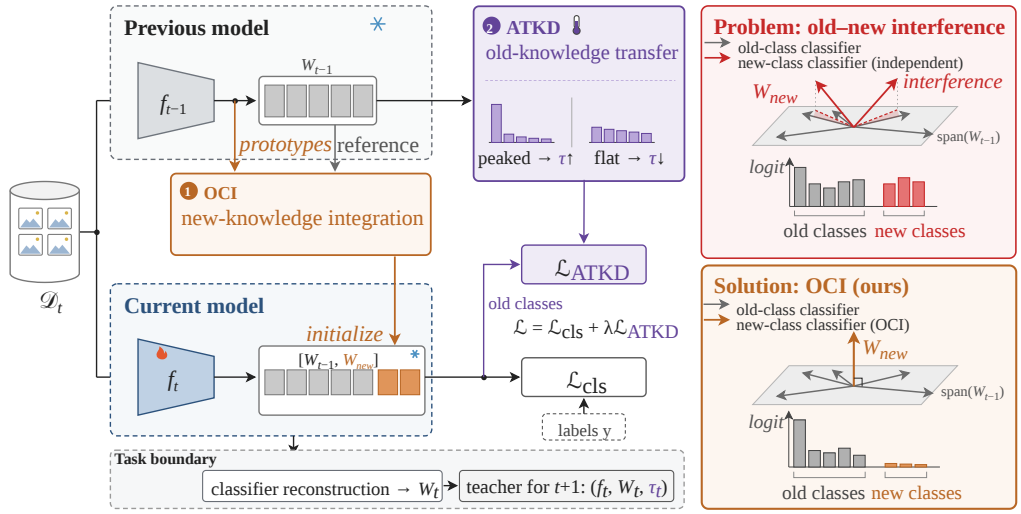


Figure 1: An overview of IACE. (a) OCI initializes incoming classifier directions from current-class prototypes and the retained classifier; ATKD uses the frozen previous model's responses to adjust the next teacher's effective distillation temperature at the task boundary. (b) OCI removes components along the old-classifier span to reduce old–new overlap at initialization.

3.2 Interference-Aware Classifier Expansion

The overview of the proposed IACE method is illustrated in Fig 1 Each incremental task involves two coupled requirements: integrating new classes and preserving old knowledge. IACE conditions the two corresponding supervision signals. Orthogonal Classifier Initialization (OCI) provides a task-local classifier geometry for new-class integration, while Adaptive-Temperature Knowledge Distillation (ATKD) calibrates the teacher signal for old-knowledge transfer.

For task $t \geq 2$, the previous model is frozen. OCI constructs the expanded classifier, while the old-class distillation temperature τ_t is inherited from the preceding task boundary. The classifier directions remain fixed during task-t training, and only f_t is optimized:

$$\mathcal{L} = \mathcal{L}_{\text{cls}} + \lambda \mathcal{L}_{\text{ATKD}}. \tag{4}$$

The classification term is cross-entropy over only incoming-class logits and current labels. The distillation term uses only old-class logits, softened by the fixed factor $\tau = 2$; ATKD changes their effective temperature through the classifier scale. At the task boundary, the frozen previous model is evaluated on current inputs, the classifier is reconstructed from updated class statistics, and ATKD sets the scale of the next teacher.

Figure 1 summarizes this task transition and the classifier relation targeted by OCI.

3.3 Orthogonal Classifier Initialization

Independent initialization does not explicitly account for the classifier geometry established by previous classes. OCI instead constructs new-class directions from current-class semantics while removing their components along the retained classifier directions.

Let W_{t-1} denote the retained classifier. The intended old–new orthogonality is

$$W_{\text{new}}(W_{t-1})^\top = 0. \tag{5}$$

For each incoming class at $t \geq 2$, we average its features extracted by the frozen f_{t-1} and ℓ_2-normalize the mean. Stacking these prototypes gives $P \in \mathbb{R}^{|\mathcal{C}_t| \times d}$.

Let the columns of Q form an orthonormal basis for the row space of W_{t-1}. We remove the old-classifier components:

$$P_\perp = P - PQQ^\top. \tag{6}$$

Thus, $P_\perp$ lies in the orthogonal complement of the retained classifier. OCI takes the polar factor of the projected prototypes. With the thin SVD

$$P_\perp = U\Sigma V^\top, \qquad W_{\text{new}} = UV^\top, \tag{7}$$

The resulting rows are mutually orthonormal and satisfy Eq. (5) when $P_\perp$ has full row rank. The same SVD construction is used otherwise, without an exact orthogonality guarantee.

The resulting W_{new} is appended using Eq. (1). During task t, all classifier directions remain fixed and only f_t is optimized. OCI therefore provides a task-local geometric reference. At the task boundary, classifier reconstruction replaces this temporary geometry with one estimated from the updated class statistics.

3.4 Adaptive-Temperature Knowledge Distillation

ATKD calibrates old-knowledge transfer according to the frozen teacher's response. Since teacher responses vary across incremental stages, a fixed temperature may provide inconsistent supervision. ATKD therefore measures the teacher-response concentration at task t and uses it to determine the temperature for task $t+1$.

We first estimate the effective number of old classes from their retained prototypes. Let $\boldsymbol{\mu}_c$ denote the normalized prototype of old class c. We compute

$$\bar{s}_t = \frac{1}{N_{t-1}(N_{t-1}-1)} \sum_{c \neq c'} \boldsymbol{\mu}_c^\top \boldsymbol{\mu}_{c'}, \qquad k_t = \frac{1}{1-\bar{s}_t}, \tag{8}$$

where k_t is bounded to a valid range. A larger similarity corresponds to a larger effective set of competing old classes.

We then evaluate the frozen teacher on current-task inputs. Let $p_{i,(j)}$ denote the j-th largest probability from the teacher output over the N_{t-1} old classes, and let $m = \lfloor k_t \rfloor$. The teacher-response concentration is

$$r_t = \mathbb{E}_{x_i}\left[\sum_{j=1}^{m} p_{i,(j)} + (k_t - m)p_{i,(m+1)}\right]. \tag{9}$$

Thus, r_t measures the probability mass assigned to the effective top-k_t old classes.

Given a target concentration ρ_t, ATKD determines the temperature for the next task as

$$\tau_{t+1} = \tau_t \frac{\psi_t(r_t)}{\psi_t(\rho_t)}, \qquad \psi_t(u) = \log \frac{u/k_t}{(1-u)/(N_{t-1}-k_t)}. \tag{10}$$

Here, ρ_t denotes the target response concentration, i.e., the desired probability mass assigned to the effective top-k_t old classes. A larger ρ_t encourages a more concentrated teacher response, whereas a smaller ρ_t encourages a softer distribution over old classes. ATKD compares the observed concentration r_t with ρ_t and adjusts the temperature accordingly for the next incremental task. Under a two-level approximation of the teacher response, $\psi_t(u)$ is proportional to inverse temperature. Therefore, when the teacher response is more concentrated than the target, ATKD increases the temperature to soften the next teacher, and vice versa. The update is bounded for stability, and the resulting τ_{t+1} is used for old-class distillation in task $t+1$.

4 Experiments

4.1 Experimental Setting

Datasets and protocols. We evaluate IACE under the exemplar-free Class-IL protocol on CIFAR-100 (Krizhevsky, 2009), TinyImageNet (Deng et al., 2009), ImageNet-100 (Deng et al., 2009), CUB-200 (Wah et al., 2011), and ImageNet-1K (Deng et al., 2009). For each dataset, classes are evenly partitioned into disjoint tasks, including the initial task. Only current-task images are available during training, and task identities are unavailable at inference.

CIFAR-100, TinyImageNet, and ImageNet-100 are evaluated with 10 and 20 tasks, while ImageNet-1K uses 10 tasks. For CUB-200, we evaluate a five-task cold-start protocol and 10- and 20-task

protocols with supervised ImageNet-1K pre-training. Unless otherwise specified, we follow the cold-start setting of Magistri et al. (2024), where the feature extractor is randomly initialized.

Implementation details. All methods in the main comparison use ResNet-18. IACE results are averaged over five class orders, except ImageNet-100 with 20 tasks, where three orders are used, and ImageNet-1K, where one order is used due to computational cost. The distillation hyperparameters are selected on CIFAR-100 and then kept fixed across the remaining benchmarks.

Evaluation metrics. Let A_t denote accuracy over all classes observed after task t. We report final accuracy

$$A_{\text{last}} = A_T, \tag{11}$$

and average incremental accuracy

$$A_{\text{inc}} = \frac{1}{T}\sum_{t=1}^{T} A_t. \tag{12}$$

For controlled analyses, we additionally report F_{last}, defined as the average decrease from each previous task's best accuracy to its accuracy after the final task. Higher A_{last} and A_{inc} and lower F_{last} are preferred.

4.2 Comparison with State-of-the-Art Methods

We compare IACE with representative EFCIL methods, including regularization-based approaches such as LwF (Li & Hoiem, 2018), prototype- and statistics-based methods such as PASS (Zhu et al., 2021), FeCAM (Goswami et al., 2023), EFC (Magistri et al., 2024), and AdaGauss (Rypeść et al., 2024), analytic-learning methods such as ACIL (Zhuang et al., 2022) and DS-AL (Zhuang et al., 2024), and drift-compensation or feature-alignment methods such as SDC (Yu et al., 2020), ADC (Goswami et al., 2024), LDC (Gomez-Villa et al., 2024), DPCR (He et al., 2025), BiCyc (Xu & Krawczyk, 2026a), and GATF (Xu & Krawczyk, 2026b). The reference results for CIFAR-100, TinyImageNet, and ImageNet-100 are compiled primarily from the comparison reported by He et al. (2025), supplemented by the results of Rypeść et al. (2024), Xu & Krawczyk (2026a), and Xu & Krawczyk (2026b). For CUB-200, the five-task random-initialization baselines follow He et al. (2025), whereas the pre-trained 10- and 20-task baselines follow Xu & Krawczyk (2026b) and Xu & Krawczyk (2026a). ImageNet-1K baselines follow He et al. (2025). We include published results only when the dataset, task partition, and initialization setting match our protocol. Since class orders and random seeds may differ across sources, these comparisons are unpaired.

CIFAR-100 and TinyImageNet. Tab. 1 reports the results on CIFAR-100 and TinyImageNet. IACE achieves the highest reported final and average incremental accuracy under all four protocols. Relative to the strongest reported baseline, IACE improves final accuracy by 1.92% and 0.38% on CIFAR-100 with 10 and 20 tasks, and by 1.44% and 0.26% on TinyImageNet, respectively. The gains persist when the number of task transitions doubles, indicating that IACE remains effective as classifier expansion becomes more frequent. This result is consistent with our motivation: OCI provides a controlled geometric reference for new-class integration, while ATKD regulates old-knowledge transfer across incremental transitions.

CUB-200. Tab. 1 evaluates IACE under two backbone initialization strategies on CUB-200. In the cold-start setting with random initialization, IACE achieves the highest reported final accuracy of 30.82%, demonstrating that its benefit does not rely on prior visual knowledge. With ImageNet pre-training, IACE remains consistently effective, obtaining the highest reported final accuracy under both the 10- and 20-task protocols. In particular, with 10 tasks, IACE reaches 58.33% final accuracy and 66.11% average incremental accuracy, outperforming the strongest baseline by 4.63% and 2.11%, respectively. With 20 tasks, it achieves the best final accuracy of 46.35%, although its A_{inc} is 3.10% lower than EFC.

Overall, IACE performs competitively under both randomly initialized and ImageNet-pre-trained backbones, showing that its effectiveness is not tied to a specific initialization strategy. The results suggest that the proposed classifier-expansion mechanism can benefit both representation learning from scratch and incremental adaptation from a strong pre-trained representation. Since the random-

Table 1: Class-IL results (%) on CIFAR-100 and TinyImageNet. IACE is averaged over five class orders. The best result in each column is bold.

Method	CIFAR-100				TinyImageNet			
	$T = 10$		$T = 20$		$T = 10$		$T = 20$	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
LwF (Li & Hoiem, 2018)	42.67	58.52	36.34	51.52	26.99	42.92	18.80	33.05
SDC (Yu et al., 2020)	42.32	58.34	32.10	48.68	23.86	40.66	13.45	29.71
PASS (Zhu et al., 2021)	44.47	55.88	28.48	42.65	23.89	36.82	12.50	25.38
ACIL (Zhuang et al., 2022)	35.53	50.53	27.22	39.58	26.10	41.86	21.40	33.60
FeCAM (Goswami et al., 2023)	34.82	49.14	25.77	41.21	29.83	42.19	22.69	34.48
EFC (Magistri et al., 2024)	43.50	58.10	32.40	47.00	34.50	47.90	28.40	42.10
AdaGauss (Rypeść et al., 2024)	46.80	60.90	37.90	54.40	32.90	45.80	27.50	39.50
DS-AL (Zhuang et al., 2024)	36.83	51.47	28.90	40.37	27.01	40.10	21.86	33.55
ADC (Goswami et al., 2024)	46.80	62.05	34.69	52.16	32.90	46.93	20.69	36.14
LDC (Gomez-Villa et al., 2024)	46.60	61.67	36.76	53.06	33.74	47.37	24.49	38.04
DPCR (He et al., 2025)	49.98	63.21	38.98	54.31	36.40	47.65	26.54	38.09
BiCyc (Xu & Krawczyk, 2026a)	50.60	63.20	41.50	56.50	35.40	49.10	30.20	44.20
GATF (Xu & Krawczyk, 2026b)	52.12	64.40	43.00	56.60	37.20	50.60	31.50	45.10
IACE (ours)	**54.04**	**65.93**	**43.38**	**56.76**	**38.64**	**50.98**	**31.76**	**45.75**

Table 2: Class-IL results (%) on CUB-200 with random and ImageNet pre-trained initialization. Random-initialization baselines use one class order; IACE is averaged over five class orders. The best result in each column is bold.

Method	Randomly		Pre-trained			
	$T = 5$		$T = 10$		$T = 20$	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
LwF (Li & Hoiem, 2018)	25.40	36.38	30.40	46.10	19.40	34.70
SDC (Yu et al., 2020)	24.24	36.00	50.30	60.50	27.90	40.10
ACIL (Zhuang et al., 2022)	21.14	33.14	–	–	–	–
EFC (Magistri et al., 2024)	–	–	51.00	63.30	46.10	**59.30**
AdaGauss (Rypeść et al., 2024)	–	–	52.90	63.40	45.00	57.00
DS-AL (Zhuang et al., 2024)	21.28	32.36	–	–	–	–
ADC (Goswami et al., 2024)	28.84	39.44	49.50	58.80	35.40	48.30
LDC (Gomez-Villa et al., 2024)	28.70	39.09	47.50	55.70	27.20	39.80
DPCR (He et al., 2025)	29.51	39.44	–	–	–	–
BiCyc (Xu & Krawczyk, 2026a)	–	–	53.70	64.00	43.70	55.90
GATF (Xu & Krawczyk, 2026b)	–	–	52.40	63.10	43.70	55.80
IACE (ours)	**30.82**	**39.65**	**58.33**	**66.11**	**46.35**	56.20

and pre-trained settings use different task partitions, however, their absolute accuracies should not be directly compared to quantify the effect of pre-training itself.

ImageNet-100 and ImageNet-1K. In Tab. 3, IACE improves the strongest baseline by 1.40% and 1.25% in final accuracy on ImageNet-100 with 10 and 20 tasks, respectively. Under the 20-task protocol, it also achieves the highest average incremental accuracy. With 10 tasks, its A_{inc} is 0.64% below DPCR, indicating that its advantage is more pronounced in the final expanded model than across all intermediate stages.

On ImageNet-1K, IACE improves DPCR by 1.10% in final accuracy and 1.42% in average incremental accuracy. Although this experiment uses a single class order, the gain at 1,000 classes suggests that the proposed transition mechanism remains effective as the decision space scales.

Across the ten evaluated protocols, IACE achieves the highest reported A_{last} in every setting and the highest A_{inc} in eight. This consistency across task granularities, initialization regimes, and label spaces ranging from 100 to 1,000 classes supports the broad applicability of IACE.

Table 3: Class-IL accuracy (%) on ImageNet-100 and ImageNet-1K. IACE averages five and three class orders for the 10- and 20-task ImageNet-100 splits, respectively. ImageNet-1K uses 10 tasks and one class order. The best result in each column is bold.

Method	ImageNet-100				ImageNet-1K	
	$T = 10$		$T = 20$		$T = 10$	
	A_{last}	A_{inc}	A_{last}	A_{inc}	A_{last}	A_{inc}
LwF (Li & Hoiem, 2018)	42.25	61.23	30.11	50.40	22.01	42.40
ACIL (Zhuang et al., 2022)	44.61	59.77	33.05	48.58	32.28	46.61
EFC (Magistri et al., 2024)	50.90	61.30	38.60	50.50	–	–
AdaGauss (Rypeść et al., 2024)	51.10	65.00	42.60	57.40	–	–
DS-AL (Zhuang et al., 2024)	45.55	60.56	34.10	49.38	33.67	48.84
ADC (Goswami et al., 2024)	46.69	65.60	32.21	52.36	31.34	50.95
LDC (Gomez-Villa et al., 2024)	49.98	67.47	34.87	54.84	35.15	53.88
DPCR (He et al., 2025)	52.16	**67.51**	38.35	57.22	35.49	54.22
BiCyc (Xu & Krawczyk, 2026a)	52.70	66.80	43.80	58.20	–	–
GATF (Xu & Krawczyk, 2026b)	53.20	67.40	44.70	59.10	–	–
IACE (ours)	**54.60**	66.87	**45.95**	**59.51**	**36.59**	**55.64**

Table 4: Component ablation on CIFAR-100 (%), averaged over five class orders.

Components		$T = 10$			$T = 20$		
OCI	ATKD	$A_{\text{last}} \uparrow$	$A_{\text{inc}} \uparrow$	$F_{\text{last}} \downarrow$	$A_{\text{last}} \uparrow$	$A_{\text{inc}} \uparrow$	$F_{\text{last}} \downarrow$
×	×	49.80	63.23	20.92	38.72	54.91	34.12
✓	×	53.19	64.92	16.77	41.67	56.48	29.01
×	✓	52.46	64.19	18.99	40.93	55.72	30.76
✓	✓	**54.04**	**65.93**	**15.17**	**43.38**	**56.76**	**26.84**

Table 5: Final accuracy (%) under fixed or trainable classifier directions, averaged over five class orders. CIFAR-100 and TinyImageNet use 10 tasks; CUB-200 uses five cold-start tasks.

	CIFAR-100		CUB-200		TinyImageNet	
Initialization	Fixed	Trainable	Fixed	Trainable	Fixed	Trainable
Random	52.39	49.85	27.89	25.98	35.99	34.07
OCI	**54.16**	**50.33**	**30.67**	**26.34**	**38.64**	**35.28**

4.3 Ablation and Mechanism Analysis

We examine the two mechanisms underlying IACE: controlling the geometry of newly introduced classifier directions and calibrating old-knowledge transfer.

Complementarity of OCI and ATKD. Tab. 4 evaluates OCI and ATKD under identical class orders and optimization settings. Both components independently improve the baseline. With 10 tasks, OCI increases final accuracy from 49.80% to 53.19% and reduces forgetting from 20.92% to 16.77%, while ATKD improves final accuracy to 52.46% and reduces forgetting to 18.99%. The same trend holds with 20 tasks, where OCI and ATKD improve final accuracy by 2.95% and 2.21%, respectively.

Combining OCI and ATKD gives the best performance under both task partitions. Relative to the baseline, IACE improves final accuracy by 4.24% and 4.66% and reduces forgetting by 5.75% and 7.28% for 10 and 20 tasks, respectively. These results support their complementary roles: OCI improves new-class integration by controlling the geometry introduced during classifier expansion, while ATKD strengthens the preservation of old knowledge during representation adaptation.

What makes OCI effective? Tab. 5 further analyzes the construction and fixation of the incoming-class directions. Across CIFAR-100, CUB-200, and TinyImageNet, fixed classifier directions consistently outperform their trainable counterparts for every initialization strategy, indicating that preserving the geometry established at classifier expansion benefits subsequent representation learning.

OCI also consistently outperforms the other fixed-direction variants. Compared with random orthonormal initialization, it improves final accuracy by 0.57% on CIFAR-100, 2.11% on CUB-200, and 1.82% on TinyImageNet. This shows that orthogonality alone is insufficient. By constructing incoming-class directions from class prototypes while removing their components in the retained classifier subspace, OCI combines semantic guidance with explicit old–new geometric separation. The results therefore support the proposed design of controlling how new classifier directions are introduced into the existing decision space.

5 Conclusion

We studied EFCIL from the perspective of classifier expansion, where each incremental transition must integrate new classes while preserving old knowledge without previous samples. To address these coupled requirements, we proposed Interference-Aware Classifier Expansion (IACE). OCI constructs prototype-guided incoming-class directions while reducing interference with the retained classifier geometry, whereas ATKD calibrates old-knowledge transfer from the frozen teacher response. At the task boundary, the classifier is reconstructed from updated class statistics.

Experiments on five benchmarks validate this design. IACE achieves the highest reported final accuracy across all evaluated protocols, while OCI and ATKD provide complementary gains. Further analysis shows that OCI outperforms generic orthogonal initialization and that ATKD improves retention over fixed and class-count-based temperature rules. These results support jointly controlling new-class integration and old-knowledge transfer for effective classifier expansion.

AI use statement

AI tools were used solely for grammar checking and to improve the clarity and readability of the manuscript. All scientific content, methodology, experimental design, results, analyses, and conclusions were developed and verified by the authors. The authors reviewed and approved the final manuscript and take full responsibility for its content.

REFERENCES

Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. ImageNet: A large-scale hierarchical image database. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 248–255, 2009.

Zijian Gao, Kele Xu, Huiping Zhuang, Li Liu, Xinjun Mao, Bo Ding, Dawei Feng, and Huaimin Wang. Less confidence, less forgetting: Learning with a humbler teacher in exemplar-free class-incremental learning. *Neural Networks*, 179:106513, 2024. doi: 10.1016/j.neunet.2024.106513.

Alex Gomez-Villa, Dipam Goswami, Kai Wang, Andrew D. Bagdanov, Bartlomiej Twardowski, and Joost van de Weijer. Exemplar-free continual representation learning via learnable drift compensation. In *European Conference on Computer Vision (ECCV)*, 2024.

Dipam Goswami, Yuyang Liu, Bartlomiej Twardowski, and Joost van de Weijer. FeCAM: Exploiting the heterogeneity of class distributions in exemplar-free continual learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023.

Dipam Goswami, Albin Soutif-Cormerais, Yuyang Liu, Sandesh Kamath, Bartlomiej Twardowski, and Joost van de Weijer. Resurrecting old classes with new data for exemplar-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28525–28534, 2024.

Run He, Di Fang, Yicheng Xu, Yawen Cui, Ming Li, Cen Chen, Ziqian Zeng, and Huiping Zhuang. Semantic shift estimation via dual-projection and classifier reconstruction for exemplar-free class-incremental learning. In *Proceedings of the 42nd International Conference on Machine Learning (ICML)*, volume 267 of *Proceedings of Machine Learning Research*. PMLR, 2025.

Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531*, 2015. URL https://arxiv.org/abs/1503.02531.

Alex Krizhevsky. Learning multiple layers of features from tiny images. Technical report, University of Toronto, 2009.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018.

Xuandi Luo, Huaidong Zhang, Yehan Zhang, and Shengfeng He. DynLearn: Steering task-specific learning dynamics in exemplar-free class-incremental learning. SSRN preprint, 2025. URL https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5933543.

Simone Magistri, Tomaso Trinci, Joost van de Weijer, Andrew Bagdanov, et al. Elastic feature consolidation for cold start exemplar-free incremental learning. In *International Conference on Learning Representations*, volume 2024, pp. 31909–31931, 2024.

Federico Pernici, Matteo Bruni, Claudio Baecchi, Francesco Turchini, and Alberto Del Bimbo. Class-incremental learning with pre-allocated fixed classifiers. In *25th International Conference on Pattern Recognition (ICPR)*, pp. 6259–6266, 2020. doi: 10.1109/ICPR48806.2021.9413299.

Grégoire Petit, Adrian Popescu, Hugo Schindler, David Picard, and Bertrand Delezoide. FeTrIL: Feature translation for exemplar-free class-incremental learning. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 3911–3920, 2023.

Grzegorz Rypeść, Sebastian Cygert, Tomasz Trzciński, and Bartłomiej Twardowski. Task-recency bias strikes back: Adapting covariances in exemplar-free class incremental learning. *Advances in Neural Information Processing Systems*, 37:63268–63289, 2024.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The Caltech-UCSD birds-200-2011 dataset. Technical Report CNS-TR-2011-001, Caltech, 2011.

Zheng Wang, Wanhao Yu, Li Yang, and Sen Lin. Rethinking continual learning with progressive neural collapse. In *International Conference on Learning Representations*, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/hash/4ed3645a042ec800a474a6a312575c5e-Abstract-Conference.html.

Fankang Xu, Lu Jin, Yanpeng Sun, Shiyu Xuan, and Zechao Li. Dual-estimator: Decoupling global and local semantic shift for drift compensation in class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10799–10809, 2026.

Hongye Xu and Bartosz Krawczyk. Two-way is better than one: Bidirectional alignment with cycle consistency for exemplar-free class-incremental learning. In *International Conference on Learning Representations (ICLR)*, 2026a. URL https://arxiv.org/abs/2606.05675.

Hongye Xu and Bartosz Krawczyk. Geometry-anchored transport framework for exemplar-free class-incremental learning. In *European Conference on Computer Vision (ECCV)*, 2026b.

Lu Yu, Bartlomiej Twardowski, Xialei Liu, Luis Herranz, Kai Wang, Yongmei Cheng, Shangling Jui, and Joost van de Weijer. Semantic drift compensation for class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 6982–6991, 2020.

Xin Zhang, Liang Bai, Guanchao Wang, and Xian Yang. Exemplar-free class incremental learning via preserving class-discriminative structure. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 17979–17988, 2026.

Fei Zhu, Xu-Yao Zhang, Chuang Wang, Fei Yin, and Cheng-Lin Liu. PASS: Prototype augmentation and self-supervision for incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 16243–16252, 2021.

Kai Zhu, Wei Zhai, Yang Cao, Jiebo Luo, and Zheng-Jun Zha. Self-sustaining representation expansion for non-exemplar class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9296–9305, 2022.

Huiping Zhuang, Zhenyu Weng, Hongxin Wei, Renchunzi Xie, Kim-Chuan Toh, and Zhiping Huang. Analytic class-incremental learning with absolute memorization and privacy protection. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 11572–11585, 2022.

Huiping Zhuang, Run He, Kai Tong, Ziqian Zeng, Cen Chen, and Zhiping Lin. DS-AL: A dual-stream analytic learning for exemplar-free class-incremental learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 17237–17244, 2024.