RELORA-CL: FROM REUSABLE UPDATES TO RELIABLE REPRESENTATION MAINTENANCE IN CLASS-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Low-Rank Adaptation (LoRA) enables parameter-efficient class-incremental learning, but successive updates shift historical class representations and render stored prototypes outdated. Without historical inputs, correcting this mismatch requires estimating their displacement from current-task samples. We investigate how continual updates can accommodate new classes while keeping historical representation drift amenable to estimation. Our analysis reveals depth-dependent functional reuse: early-task input subspaces remain more sufficient for shallow-block updates, whereas deeper blocks increasingly require additional directions. Motivated by this observation, we propose reuse-aware LoRA (ReLoRA-CL), which preserves learned input subspaces in shallow blocks and expands complementary directions in deeper blocks. Semantic drift compensation and balanced classifier alignment subsequently maintain historical class statistics and decision boundaries. We formulate prototype migration as finite-sample drift estimation and derive an error bound involving local drift regularity, support mismatch, and residual uncertainty. Under additional sensitivity assumptions, we connect drift regularity to consecutive-task parameter changes, providing a conditional rationale for structured adaptation. ReLoRA-CL improves final accuracy over the strongest baseline by 3.58, 1.64, and 4.53 percentage points on ImageNet-R with 20 tasks, CIFAR-100 with 10 tasks, and ImageNet-A with 10 tasks, respectively. Drift diagnostics further show lower local drift variation and prototype displacement estimation error. The code is available at: https://anonymous.4open.science/r/ReLoRA-8C3A/

1 INTRODUCTION

Class-incremental learning (CIL) aims to acquire new classes sequentially while retaining knowledge of previously learned classes (McCloskey & Cohen, 1989; French, 1999; Kirkpatrick et al., 2017; De Lange et al., 2022; Zhou et al., 2024b). With large pre-trained models, parameter-efficient fine-tuning offers a practical alternative to costly full-parameter adaptation (Zhou et al., 2024a; 2025). In particular, Low-Rank Adaptation (LoRA) learns compact weight updates while keeping the pre-trained backbone frozen (Hu et al., 2022). However, freezing the backbone does not freeze its adapted representations: successive low-rank updates can change the features of historical classes.

This representation drift creates a difficulty in exemplar-free CIL. Historical class prototypes become outdated as the feature extractor evolves, yet the original samples are unavailable for recomputing them. Existing methods address this mismatch through semantic drift compensation and classifier alignment (Yu et al., 2020; Zhu et al., 2021; Zhang et al., 2023a; Li et al., 2026). When compensation relies on current-task samples, its accuracy depends on whether their observed displacements are informative about historical classes. Consequently, effective continual adaptation requires not only learning discriminative features for new classes, but also maintaining representation changes that can be estimated from the available data.

Recent studies organize low-rank updates through parameter update constraints, shared factors and task-specific components (Liang & Li, 2024; Wu et al., 2025; He et al., 2025; Jia et al., 2026). CL-LoRA assigns shared and task-specific adapters to different network depths (He et al., 2025), while SD-LoRA decouples direction and magnitude learning, freezing previous directions while adapting

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.

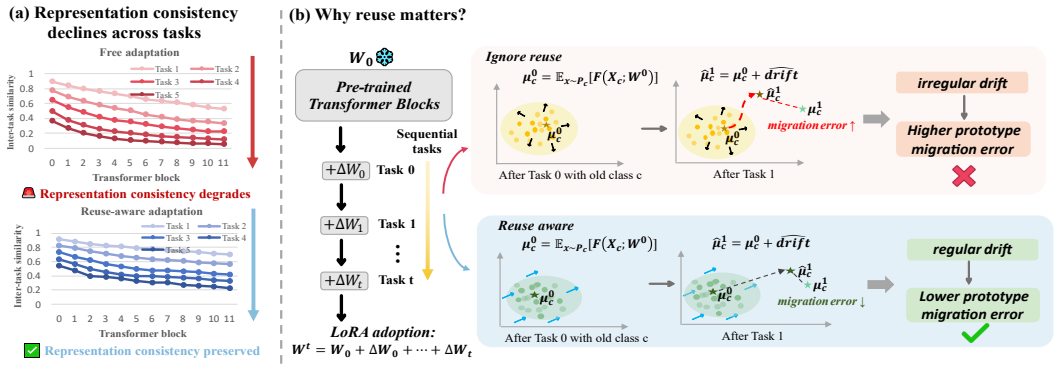


Figure 1: **Update reuse and historical-class maintenance.** (a) Consecutive-task representation cosine similarity under standard (free) and reuse-aware adaptation. (b) Conceptual connection between update reuse, representation drift, and historical prototype migration. The reliability of migration depends on both drift regularity and the support provided by current-task samples.

their magnitudes (Wu et al., 2025). These advances establish the value of structured adaptation, but do not by themselves determine how the resulting updates affect historical prototype migration. We provide a broader discussion of parameter-efficient continual learning and update-subspace reuse in App. A. We therefore investigate a complementary question: *how can continual updates accommodate new classes while keeping historical representation drift amenable to estimation?*

We approach this question by examining update reuse across Transformer depth. An analysis of standard continual LoRA on ImageNet-R reveals a depth-dependent pattern: shallow blocks exhibit stronger cross-task functional reuse, whereas deeper blocks show increasing demand for additional directions. Functional reconstruction further suggests that the Task-0 input subspace remains more sufficient for later updates in shallow blocks. These observations motivate preserving this subspace where it remains useful and expanding adaptation capacity elsewhere. Figure 1 illustrates the associated representation-maintenance problem.

Based on these observations, we propose Reuse-aware LoRA (ReLoRA-CL), which combines depth-dependent adaptation with historical-class maintenance. Reusable Subspace Adaptation (RSA) fixes the Task-0 input-side factors in shallow blocks and learns output-side contributions that recombine the retained directions. Complementary Direction Expansion (CDE) introduces new low-rank components in deeper blocks, orthogonalizing their input directions against stored components. This allocation is motivated by the observed reuse pattern. After adaptation, semantic drift compensation migrates historical prototypes, and balanced pseudo-features support classifier alignment.

To examine the connection between adaptation and maintenance, we formulate prototype migration as finite-sample drift estimation. Under a shared conditional drift assumption, we derive an error bound involving local drift regularity, support mismatch, and residual uncertainty. Additional sensitivity assumptions relate drift regularity to consecutive-task parameter changes, providing a conditional rationale for restricting unnecessary update variation. The analysis does not guarantee that subspace preservation alone improves migration; it motivates evaluating update structure, drift regularity, and prototype estimation together.

Contributions. First, we characterize depth-dependent functional reuse in continual LoRA and examine the sufficiency of early-task input subspaces. Second, we develop ReLoRA-CL, combining shallow subspace preservation, deep complementary expansion, and historical-statistics maintenance. Third, we provide a conditional theoretical analysis connecting parameter changes, drift regularity, and prototype migration error. Finally, experiments on ImageNet-R, CIFAR-100, and ImageNet-A demonstrate improved classification performance, while drift diagnostics provide evidence consistent with the proposed maintenance perspective.

2 Preliminaries

Exemplar-free class-incremental learning. We consider a sequence of tasks $\{\mathcal{D}_t\}_{t=0}^{T-1}$, where $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^{n_t}$ contains samples from a class set $\mathcal{C}_t$, with $\mathcal{C}_t \cap \mathcal{C}_k = \emptyset$ for $k < t$. At stage t, only current-task inputs are available for training; model parameters and historical class statistics

may be retained, but previous input samples are not stored. Let $\mathcal{C}_{\leq t} = \bigcup_{k=0}^{t} \mathcal{C}_k$. The adapted feature extractor is $F_t : \mathcal{X} \to \mathbb{R}^d$, and the classifier is $W_t \in \mathbb{R}^{|\mathcal{C}_{\leq t}| \times d}$. Inference is task-agnostic and predicts $\arg\max_{c \in \mathcal{C}_{\leq t}} [W_t F_t(x)]_c$.

The pre-trained backbone remains frozen, while LoRA modifies selected projections through low-rank updates. For an adapted projection in Transformer block l, let U_l^t denote its complete update after task t, with the fixed LoRA scaling absorbed into the factors. For $t \geq 1$, define $G_l^t = U_l^t - U_l^{t-1}$. Whereas U_l^t represents the accumulated adaptation, G_l^t describes the parameter change relevant to the transition $F_{t-1} \to F_t$.

Historical prototype migration. For an old class $c \in \mathcal{C}_{<t} := \mathcal{C}_{\leq t-1}$ with input distribution P_c, its true prototype is $\mu_c^t = \mathbb{E}_{X \sim P_c}[F_t(X)]$. Its displacement between consecutive stages is $d_c^t = \mu_c^t - \mu_c^{t-1} = \mathbb{E}_{X \sim P_c}\left[F_t(X) - F_{t-1}(X)\right]$. Since historical inputs are unavailable, we estimate this displacement from current-task samples. For each x_i, let $z_i = F_{t-1}(x_i)$ and $\delta_i = F_t(x_i) - F_{t-1}(x_i)$. Given a stored prototype $\widehat{\mu}_c^{t-1}$, consider

$$\widehat{d}_c^t = \sum_{i=1}^{n_t} \omega_{ic} \delta_i, \qquad \widehat{\mu}_c^t = \widehat{\mu}_c^{t-1} + \widehat{d}_c^t, \tag{1}$$

where $\omega_{ic} \geq 0$ and $\sum_i \omega_{ic} = 1$. The weights depend on pre-update features and stored statistics; their specific construction is given in Sec. 3. Migration therefore requires transferring observed drift from current classes to historical classes.

Conditional analysis of drift estimation. We characterize this transfer under a shared conditional drift model. Here, "pre-update features" refers to current inputs encoded by F_{t-1}, rather than stored historical examples.

Assumption 1 (Shared conditional drift)**.** *For fixed extractors F_{t-1} and F_t, write $F_t(X) - F_{t-1}(X) = g_t(Z) + \xi$, where $Z = F_{t-1}(X)$. A common L_t-Lipschitz function g_t applies on the union of the relevant old- and current-class feature supports, and $\mathbb{E}[\xi \mid Z, c] = 0$ for each relevant class c. Conditioned on the extractors, pre-update features $\{z_i\}$, and stored statistics, the observed residuals $\{\xi_i\}$ have zero mean, are pairwise uncorrelated, and satisfy $\mathbb{E}\|\xi_i\|_2^2 \leq \sigma_t^2$. The interpolation weights are determined by the same conditioning information.*

This assumption requires the conditional mean drift to transfer across classes; it does not follow from low-rank adaptation alone. The conditional residual requirements also need justification when the samples used to estimate drift have participated in training F_t.

Proposition 1 (Prototype displacement estimation)**.** *Under Assumption 1, let $Z_c = F_{t-1}(X)$ for an independent $X \sim P_c$. Define $r_c^t = \sum_i \omega_{ic} \mathbb{E}\|z_i - Z_c\|_2$ and $n_{\mathrm{eff},c}^t = (\sum_i \omega_{ic}^2)^{-1}$. Conditioned on the extractors, pre-update features, and stored statistics, the estimator in Eq. 1 satisfies*

$$\left(\mathbb{E}\|\widehat{d}_c^t - d_c^t\|_2^2\right)^{1/2} \leq L_t r_c^t + \frac{\sigma_t}{\sqrt{n_{\mathrm{eff},c}^t}}. \tag{2}$$

The first term couples drift regularity with the distance between current observations and the historical class distribution. The second captures residual uncertainty under the interpolation weights. In particular,

$$r_c^t \leq \sum_i \omega_{ic} \|z_i - \widehat{\mu}_c^{t-1}\|_2 + \|\widehat{\mu}_c^{t-1} - \mu_c^{t-1}\|_2 + \mathbb{E}\|Z_c - \mu_c^{t-1}\|_2. \tag{3}$$

Thus, distant current classes, accumulated prototype error, and within-class spread can limit migration even for a regular drift field. The bound is explanatory rather than directly computable, because the historical distribution is unavailable.

Connecting drift regularity to parameter changes. Suppose the conditional mean drift admits a local representation

$$g_t(z) = \sum_l \mathcal{J}_{l,t}(z)[G_l^t] + e_t(z), \tag{4}$$

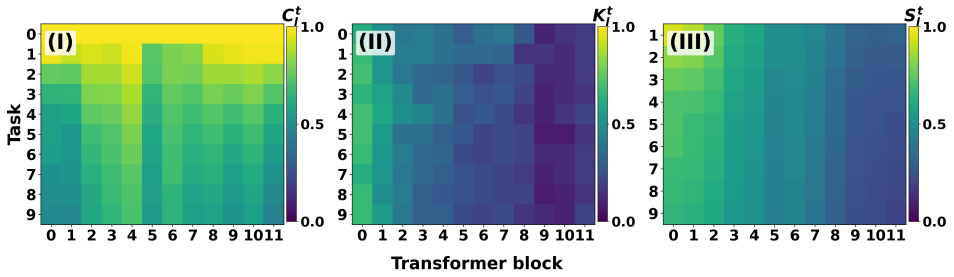


Figure 2: **Depth-dependent update reuse in continual LoRA.** Using standard continual LoRA on ImageNet-R ($T = 10$), we measure (I) update alignment C_l^t, (II) update compactness K_l^t, and (III) their combined score S_l^t. These diagnostics characterize functional consistency and concentration; subspace sufficiency is further examined through reconstruction in App. C.

where $\mathcal{J}_{l,t}(z)$ is a linear parameter-to-feature sensitivity operator and e_t is a remainder. Assume, on the relevant feature region, that

$$\left\| \left(\mathcal{J}_{l,t}(z) - \mathcal{J}_{l,t}(z') \right)[H] \right\|_2 \leq \gamma_{l,t} \|H\|_F \|z - z'\|_2 \tag{5}$$

for all compatible H, and that e_t is ε_t-Lipschitz. Then

$$\mathrm{Lip}(g_t) \leq C_t + \varepsilon_t, \qquad C_t = \sum_l \gamma_{l,t} \|G_l^t\|_F. \tag{6}$$

These are additional sensitivity and remainder assumptions, not automatic consequences of a first-order expansion. Together with Eq. 2, they identify parameter variation as one factor controlling an upper bound on migration error, alongside support and residual uncertainty.

Proposition 2 (Conditional projection comparison)**.** *Let P_l be the orthogonal projector onto* $\mathrm{row}(A_l^0)$. *For fixed reference updates $\{G_l\}$, define $\widetilde{G}_l = G_l P_l$ in selected blocks $\mathcal{H}$ and $\widetilde{G}_l = G_l$ elsewhere. For common nonnegative sensitivities $\{\gamma_l\}$,*

$$\sum_l \gamma_l \|\widetilde{G}_l\|_F \leq \sum_l \gamma_l \|G_l\|_F. \tag{7}$$

With a fixed input factor, consecutive updates satisfy $G_l^t = (\bar{B}_l^t - \bar{B}_l^{t-1}) A_l^0$ and hence $G_l^t = G_l^t P_l$. This excludes off-subspace variation but does not constrain the magnitude of the remaining update. Moreover, an optimized constrained update need not equal the projection of a freely optimized one. Prop. 2 therefore motivates subspace preservation conditionally; it does not establish that ReLoRA-CL necessarily has smaller drift or migration error than free LoRA. The analysis concerns one-stage mean estimation. Errors may accumulate across stages, and covariance evolution and classification accuracy require separate evaluation. App. B provides the proofs and accumulated-error discussion.

3 Method

ReLoRA-CL combines structured parameter adaptation with historical-class maintenance. The analysis in Sec. 2 identifies drift regularity as one factor affecting prototype migration, alongside sample support and residual uncertainty. This motivates controlling how adaptation changes representations, while retaining sufficient capacity for new classes. We first examine the reuse of previously learned updates, then use the observed depth-dependent pattern to organize shallow subspace preservation and deep direction expansion. Semantic drift compensation (SDC) and classifier alignment (CA) subsequently maintain historical class statistics and decision boundaries.

3.1 Characterizing Depth-Dependent Update Reuse

A fixed update subspace is useful only while it remains expressive for later tasks. Preserving an insufficient subspace can restrict learning, whereas continually replacing reusable directions introduces additional parameter variation. To examine this distinction, we analyze standard LoRA on ImageNet-R ($T = 10$), allowing both low-rank factors to change across tasks.

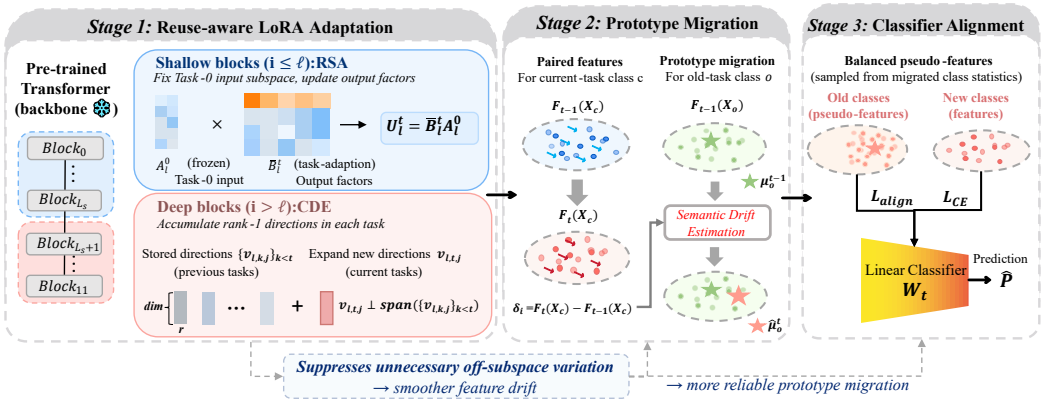


Figure 3: **Overview of ReLoRA-CL.** (1) RSA preserves learned input-side subspaces in shallow blocks, while CDE adds complementary directions in deeper blocks. (2) Paired features of current-task samples estimate historical prototype displacements. (3) Balanced pseudo-features support classifier alignment. The dashed connection indicates the conditional relationship between parameter changes, drift regularity, and migration error.

Empirical characterization. Let $\delta_l^{(t)}(x)$ denote the LoRA-induced feature update at block l after task t. We measure update alignment by $C_l^t = \mathbb{E}_{x\sim\mathcal{D}_t}[\cos(\delta_l^{(0)}(x), \delta_l^{(t)}(x))]$, evaluating both updates on the same current-task inputs. This measures functional agreement with the Task-0 update, rather than direct overlap between input-side subspaces.

For the sampled update-response matrix, let u_l^t be the smallest rank retaining 95% of its squared singular-value energy. We define compactness as $K_l^t = (D - u_l^t)/D$, where D is the response dimension, and combine the diagnostics as $S_l^t = \max(C_l^t, 0)K_l^t$. Directional consistency alone does not characterize structural concentration, and compactness alone does not indicate cross-task reuse. We therefore treat S_l^t as a descriptive score and complement it with the functional reconstruction analysis in App. C.

Figure 2 shows stronger functional reuse in shallow blocks and decreasing agreement in deeper blocks as tasks accumulate. The reconstruction results further suggest that the Task-0 input subspace remains more sufficient for shallow-block updates. Together, these observations motivate different adaptation strategies across Transformer depth and task progression.

Conditional subspace interpretation. Following the shared-optimum perspective of SD-LoRA (Wu et al., 2025), suppose task-optimal updates satisfy $\|\Delta W_{l,t}^\star - \Delta W_l^\star\|_{\mathrm{op}} \le \epsilon_l$ for a common reference $\Delta W_l^\star$. Let P_l^0 project onto $\mathrm{row}(A_l^0)$ and $[\Delta W_l^\star]_r$ be a best rank-r approximation. Define the subspace mismatch $\xi_l = \|[\Delta W_l^\star]_r(I - P_l^0)\|_{\mathrm{op}}$. The triangle inequality gives $\|\Delta W_{l,t}^\star(I - P_l^0)\|_{\mathrm{op}} \le \epsilon_l + \sigma_{r+1}(\Delta W_l^\star) + \xi_l$.

Thus, retaining the Task-0 subspace is appropriate when task variation, spectral truncation error, and subspace mismatch are small. Low rank alone does not ensure these conditions. The empirical analysis motivates subspace preservation in shallow blocks, while weaker reconstruction in deeper blocks motivates additional directions.

3.2 Depth-Dependent Subspace Adaptation

Guided by the observed reuse pattern, we apply RSA to blocks $l \le L_s$ and CDE to blocks $l > L_s$. The boundary L_s is a fixed hyperparameter, rather than an online decision based on S_l^t. For each projection, the complete update is $U_l^t = U_{l,\mathrm{RSA}}^t$ in shallow blocks and $U_l^t = U_{l,\mathrm{CDE}}^t$ in deeper blocks.

Reusable Subspace Adaptation (RSA). For a local input h, a LoRA branch produces $B_l A_l h$: A_l selects input directions, while B_l maps their responses to the output. When later tasks remain well represented by the learned input subspace, adapting both factors provides flexibility that may be unnecessary. RSA therefore learns A_l^0 on Task 0 and fixes it thereafter, while allowing output-side

adaptation:

$$U^t_{l,\mathrm{RSA}} = \bar{B}^t_l A^0_l = \left(\sum_{k=0}^{t} a^{(t)}_{l,k} B_{l,k} \right) A^0_l, \qquad l \le L_s. \tag{8}$$

At task t, historical factors $B_{l,k}$ for $k < t$ are frozen, while a new factor $B_{l,t}$ and the input-independent mixture coefficients $\{a^{(t)}_{l,k}\}^t_{k=0}$ are optimized. This separates the addition of a new output mapping from the reweighting of stored mappings.

The motivation is expressivity within a retained subspace. If $\mathrm{row}(A^t_l) \subseteq \mathrm{row}(A^0_l)$, then $A^t_l = Q^t_l A^0_l$ for some Q^t_l, and hence $B_{l,t}A^t_l = (B_{l,t}Q^t_l)A^0_l$. Input-side recombination can therefore be absorbed into the output factor without changing the represented update. The task-wise decomposition in Eq. 8 is an optimization parameterization; its expressivity is equivalent to that of the merged factor $\bar{B}^t_l$.

With A^0_l fixed, consecutive updates remain in its row subspace, excluding off-subspace parameter variation. This connects RSA to the conditional projection analysis in Sec. 2, but does not bound the magnitude of output-side changes or guarantee invariant historical features.

Complementary Direction Expansion (CDE). In deeper blocks, restricting all future updates to an early subspace can prevent the model from capturing new discriminative directions. CDE instead adds task-specific low-rank components while retaining previously learned ones: $U^t_{l,\mathrm{CDE}} = \sum^t_{k=0} \sum^r_{j=1} b^{(t)}_{l,k,j} u_{l,k,j} v^\top_{l,k,j}$, $l > L_s$. Here, each $u_{l,k,j}v^\top_{l,k,j}$ is a rank-1 update, and $b^{(t)}_{l,k,j}$ controls its contribution. The term $k = 0$ represents the initial-task components.

For a new task, we orthogonalize its input factors $\{v_{l,t,j}\}^r_{j=1}$ against the stored input directions and mutually within the new set. This removes overlap with the existing input span at the orthogonalization step, directing added capacity toward complementary directions. Only the new factors and their coefficients are optimized; historical factors remain frozen, with $b^{(t)}_{l,k,j} = b^{(k)}_{l,k,j}$ for $k < t$. Complementarity is defined within each deep projection, not relative to shallow-layer subspaces. It also does not imply zero interference on historical inputs, which motivates the subsequent statistics compensation.

Current-task optimization. We expand W_t for the new classes and optimize the active adaptation parameters and classifier using $\mathcal{L}^t_{\mathrm{CE}} = -n_t^{-1} \sum_{(x_i,y_i)\in\mathcal{D}_t} \log[\mathrm{softmax}(W_t F_t(x_i))]_{y_i}$, where the softmax covers all seen classes. The backbone remains frozen throughout training; after Task 0, the shallow input factors and historical adaptation factors are also fixed. Although all seen classes enter the loss normalization, only current classes provide training examples. Consequently, adaptation can both shift historical representations and bias the classifier toward new classes. We address these two effects separately below.

3.3 Historical-Statistics Compensation

Estimating historical displacement. Subspace preservation constrains parameter updates but does not eliminate representation drift. Directly reusing old prototypes can therefore leave them misaligned with the current extractor. Following SDC (Yu et al., 2020), we estimate their displacement from paired current-task features.

Before adaptation, we retain F_{t-1}. After adaptation, we compute $z_i = F_{t-1}(x_i)$ and $\delta_i = F_t(x_i) - F_{t-1}(x_i)$ using the same inputs and a consistent feature-extraction protocol. For historical class c, define $q_{ic} = -\|z_i - \widehat{\mu}^{t-1}_c\|^2_2/(2\tau^2_c)$, where $\tau_c > 0$ is a bandwidth. The normalized weights are $\omega_{ic} = \exp(q_{ic} - m_c)/\sum_j \exp(q_{jc} - m_c)$, with $m_c = \max_j q_{jc}$. We then migrate the prototype by

$$\widehat{d}^t_c = \sum_i \omega_{ic}\delta_i, \qquad \widehat{\mu}^t_c = \widehat{\mu}^{t-1}_c + \widehat{d}^t_c. \tag{9}$$

Distances are measured in the pre-update feature space so that current features and stored prototypes use the same reference representation, up to accumulated prototype error. Nearby observations receive larger weights because their displacements are more informative under locally regular drift. The bandwidth balances localization against effective sample size, as reflected in Eq. 2. Subtracting m_c improves numerical stability but cannot compensate for inadequate support from current classes. New-class prototypes are computed directly from post-adaptation features.

Balanced classifier alignment. Migrating means addresses feature-space mismatch, but does not remove classifier bias caused by current-class-only training. Following statistical classifier alignment (Zhang et al., 2023a), we approximate each class distribution using its mean and covariance and give every seen class equal representation during recalibration.

At class arrival, we estimate a regularized positive-definite covariance Σ_c from its post-adaptation features. For historical classes, we retain this covariance and migrate only the mean, giving a translation-only approximation. For each class, we sample $\widetilde{z}_{c,m} = \widehat{\mu}_c^t + L_c\epsilon_{c,m}$, where $L_cL_c^\top = \Sigma_c$ and $\epsilon_{c,m} \sim \mathcal{N}(0, I)$. With F_t frozen, we optimize

$$\mathcal{L}_{\text{align}}^t = -\frac{1}{M|\mathcal{C}_{\le t}|} \sum_{c\in\mathcal{C}_{\le t}} \sum_{m=1}^{M} \log[\text{softmax}(W_t\widetilde{z}_{c,m})]_c. \tag{10}$$

Balanced sampling removes class-frequency imbalance in this surrogate objective.

3.4 Training Procedure and Resource Costs

Task 0 initializes the adaptation modules and class statistics. For each subsequent task, ReLoRA-CL optimizes $\mathcal{L}_{\text{CE}}^t$, extracts paired features for historical prototype migration, and performs classifier alignment after updating statistics. Inference uses a single extractor and classifier without task identity or adapter selection. Storage grows with retained adaptation factors and class statistics: full covariance storage requires $O(|\mathcal{C}_{\le t}|d^2)$, and drift estimation temporarily stores the previous extractor. Input-independent updates can be merged for inference, while RSA and CDE retain historical factors and input directions during training.

4 Experiments

We evaluate ReLoRA-CL from three perspectives: classification performance across incremental protocols, the contributions of adaptation and historical-statistics maintenance, and the relationship between representation drift and prototype migration. We additionally examine the resource costs of the complete framework.

4.1 Experimental Setup

Benchmarks and protocols. We test on ImageNet-R (Hendrycks et al., 2021a), CIFAR-100 (Krizhevsky, 2009), and ImageNet-A (Hendrycks et al., 2021b). ImageNet-R contains 200 classes and is evaluated with $T \in \{5, 10, 20\}$ sequential tasks; CIFAR-100 and ImageNet-A use $T = 10$. All experiments follow exemplar-free CIL: training accesses only current-task inputs, while model parameters and class statistics may be retained. Evaluation is task-agnostic over all classes observed so far.

Implementation. We use ViT-B/16 (Dosovitskiy et al., 2021), pretrained on ImageNet-21K and fine-tuned on ImageNet-1K (Deng et al., 2009). The backbone is frozen, and LoRA is applied to the query and value projections. Unless otherwise stated, the shallow/deep boundary is $L_s = 4$ and the LoRA rank is $r = 10$. We optimize with Adam (Kingma & Ba, 2015) using an initial learning rate of 8×10^{-3} and report means over five runs.

Baselines and metrics. We compare with Full Fine-Tuning, L2P (Wang et al., 2022b), Dual-Prompt (Wang et al., 2022a), CODA-Prompt (Smith et al., 2023), HiDe-Prompt (Wang et al., 2023), InfLoRA (Liang & Li, 2024), SD-LoRA (Wu et al., 2025), FM-LoRA (Yu et al., 2025), FS-LoRA (Hou et al., 2026), DGS (Li et al., 2026), and GOD (Jia et al., 2026). We use published baseline results under the corresponding protocols, except for DGS, which we reproduce using its public implementation with the same backbone and data splits. Let a_t denote accuracy over all seen classes after stage t. We report final accuracy $\text{ACC} = a_{T-1}$ and average anytime accuracy $\text{AAA} = T^{-1}\sum_{t=0}^{T-1} a_t$. All accuracy improvements below are measured in percentage points.

4.2 Comparison with Existing Methods

Table 1 shows that ReLoRA-CL achieves the highest ACC and AAA among the compared methods in all five settings. On ImageNet-R, its ACC improvements over the strongest baseline in each protocol

Table 1: Classification performance under different CIL protocols. ACC measures final accuracy and AAA averages accuracy across stages. ReLoRA-CL results are means over five runs. Higher values are better; the best result in each column is bold.

Method	ImageNet-R						CIFAR-100		ImageNet-A	
	$T=5$		$T=10$		$T=20$		$T=10$		$T=10$	
	ACC↑	AAA↑	ACC↑	AAA↑	ACC↑	AAA↑	ACC↑	AAA↑	ACC↑	AAA↑
Full Fine-Tuning	64.92	75.57	60.57	72.31	49.95	65.32	69.49	80.35	16.31	30.04
L2P	73.04	76.94	71.26	76.13	68.97	74.16	83.18	87.69	42.94	51.40
DualPrompt	69.99	72.24	68.22	73.81	65.23	71.30	81.48	86.41	45.49	54.68
CODA-Prompt	76.63	80.30	74.05	78.14	69.38	73.95	86.31	90.67	45.36	57.03
HiDe-Prompt	74.77	78.15	74.65	78.46	73.59	77.93	86.45	91.34	42.70	56.32
InfLoRA	76.95	81.81	74.75	80.67	69.89	76.68	86.75	91.72	49.20	60.92
SD-LoRA	79.15	83.01	77.34	82.04	75.26	80.22	88.01	92.54	55.96	64.95
FM-LoRA	78.93	82.37	77.88	82.59	76.47	80.97	88.37	92.66	57.12	66.08
FS-LoRA	79.97	83.95	76.63	82.50	75.67	81.82	87.20	93.06	55.45	65.31
DGS	80.35	84.20	77.80	82.92	74.85	83.85	86.75	91.73	57.81	65.01
GOD	80.72	85.08	77.43	83.34	73.45	80.83	88.11	92.87	56.22	63.23
ReLoRA-CL (Ours)	**82.33**	**85.45**	**81.25**	**85.37**	**80.05**	**84.37**	**90.01**	**93.86**	**62.34**	**71.29**

Table 2: Component analysis across CIL protocols. The full ReLoRA-CL framework combines RSA-CDE with SDC-CA. Higher ACC and AAA are better.

Configuration	ImageNet-R						CIFAR-100		ImageNet-A	
	$T=5$		$T=10$		$T=20$		$T=10$		$T=10$	
	ACC↑	AAA↑	ACC↑	AAA↑	ACC↑	AAA↑	ACC↑	AAA↑	ACC↑	AAA↑
w/ SDC-CA	77.88	83.05	75.73	82.42	74.97	81.25	75.78	87.15	60.50	69.75
w/ RSA	79.63	84.03	78.92	84.94	76.75	82.20	87.29	92.22	58.39	68.09
w/ CDE	80.45	84.61	78.47	83.70	77.68	82.37	86.98	92.09	58.85	67.84
w/ RSA-CDE	79.50	84.01	78.90	83.93	77.85	83.07	88.03	92.66	57.08	67.71
ReLoRA-CL (Ours)	**82.33**	**85.45**	**81.25**	**85.37**	**80.05**	**84.37**	**90.01**	**93.86**	**62.34**	**71.29**

are 1.61, 3.37, and 3.58 points for $T = 5$, 10, and 20, respectively. As the number of tasks increases from 5 to 20, ReLoRA-CL's final accuracy decreases by 2.28 points, compared with 3.89 points for SD-LoRA. These results indicate that the framework remains effective under finer task partitions and more adaptation stages. The gains also extend beyond ImageNet-R. On CIFAR-100, ReLoRA-CL improves ACC over the strongest listed baseline from 88.37% to 90.01%. On ImageNet-A, it achieves 62.34% ACC and 71.29% AAA, exceeding the respective best baseline results by 4.53 and 5.21 points.

4.3 Component Analysis

Historical-statistics maintenance. Table 2 shows that adding SDC-CA to the RSA-CDE configuration improves ACC in every evaluated setting. The gains are 2.83, 2.35, and 2.20 points on ImageNet-R with $T = 5$, 10, and 20, respectively, and 1.98 and 5.26 points on CIFAR-100 and ImageNet-A. For ImageNet-R ($T = 10$), AAA also increases from 83.93% to 85.37%. These comparisons support the benefit of maintaining historical statistics and recalibrating the classifier after structured adaptation.

Adaptation structure. The effectiveness of RSA and CDE varies across different protocols, indicating their complementary roles under different adaptation demands. Although their combination without SDC-CA does not consistently outperform individual components, the complete ReLoRA-CL framework achieves the best ACC and AAA in all settings. These results validate the benefit of jointly modeling adaptation structure and historical-statistics maintenance.

Layer-boundary sensitivity. App. D evaluates different values of L_s on ImageNet-R and CIFAR-100 ($T = 10$). The reported stability across the tested partitions suggests that performance is not narrowly dependent on a single boundary. This analysis concerns the fixed layer allocation used by ReLoRA-CL; the boundary is not selected online from the reuse score.

Efficiency results. Figure 4 reports the accuracy–resource trade-off. Compared with SD-LoRA, ReLoRA-CL improves ACC by 3.91 points with moderate additional parameter growth. It also achieves lower time and memory costs under the reported setup. Compared with CDE, ReLoRA-CL obtains higher accuracy with shorter task time, at the cost of more trainable parameters. These

Table 3: Classification results and drift diagnostics for SD-LoRA and ReLoRA-CL. Lower Drift Var and PDE indicate smaller observed local drift variation and prototype displacement estimation error.

Metric	ImageNet-R						CIFAR-100	
	$T = 5$		$T = 10$		$T = 20$		$T = 10$	
	SD-LoRA	ReLoRA-CL	SD-LoRA	ReLoRA-CL	SD-LoRA	ReLoRA-CL	SD-LoRA	ReLoRA-CL
ACC ↑	79.15	**82.33**	77.34	**81.25**	75.26	**80.05**	88.01	**90.01**
AAA ↑	83.01	**85.45**	82.04	**85.37**	80.22	**84.37**	92.54	**93.86**
Drift Var ↓	0.2648	**0.1848**	0.1296	**0.1132**	0.1862	**0.1344**	0.2211	**0.2119**
PDE ↓	3.2381	**2.1692**	1.3025	**1.2238**	0.9916	**0.8647**	1.5278	**1.3965**

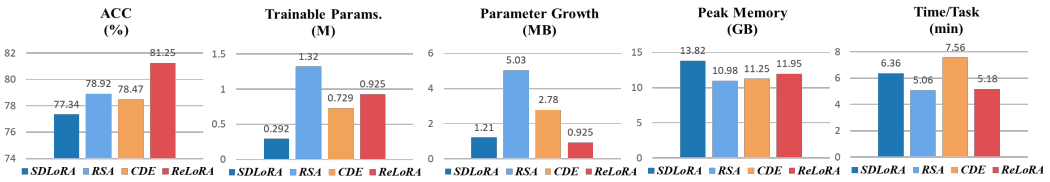


Figure 4: Performance-efficiency trade-off on ImageNet-R with $T = 10$. The metrics compare ReLoRA-CL(ReLoRA) with representative LoRA-based adaptation methods in terms of accuracy and computational cost.

results suggest that parameter count alone does not fully characterize the practical cost of continual adaptation.

4.4 Drift Regularity and Prototype Migration

We next examine whether the classification gains are accompanied by more regular observed drift and lower prototype displacement estimation error. For transition $t-1 \to t$, let $z_i = F_{t-1}(x_i)$ and $\delta_i = F_t(x_i) - F_{t-1}(x_i)$. We measure local drift variation as $V^{(t)}_{\text{drift}} = |\mathcal{N}_t|^{-1} \sum_{(i,j)\in\mathcal{N}_t} \|\delta_i - \delta_j\|_2 / (\|z_i - z_j\|_2 + \epsilon)$, where $\mathcal{N}_t$ is a k-NN graph in the pre-update feature space and $\epsilon > 0$ ensures numerical stability. This statistic measures average local variation on sampled neighbors, rather than a global Lipschitz constant.

Prototype displacement estimation error (PDE) is defined as $E^{(t)}_{\text{PDE}} = \big(|\mathcal{C}_{<t}|^{-1} \sum_{c\in\mathcal{C}_{<t}} \|\widehat{d}^t_c - d^t_c\|_2^2\big)^{1/2}$, where $d^t_c = \mathbb{E}_{X\sim P_c}[F_t(X) - F_{t-1}(X)]$ is the oracle class displacement. PDE is an evaluation diagnostic, not a quantity available during exemplar-free training.

As shown in Table 3, ReLoRA-CL has lower drift variation and PDE in all four settings. On ImageNet-R ($T = 20$), drift variation decreases from 0.1862 to 0.1344, while PDE decreases from 0.9916 to 0.8647, corresponding to reductions of 27.82% and 12.80%, respectively. The improvements are smaller on CIFAR-100, indicating that their magnitude depends on the evaluation setting.

These observations are consistent with the conditional analysis in Sec. 2: more regular drift can facilitate prototype migration when sample support and residual uncertainty are favorable. They do not isolate drift regularity as the sole cause of improved accuracy. In particular, absolute PDE also depends on displacement magnitude and feature scale; normalized errors and a no-migration reference would further distinguish easier estimation from smaller overall drift.

5 Conclusion

In this work, we study sequential LoRA updates in exemplar-free class-incremental learning and identify a depth-dependent reuse pattern: shallow blocks retain stronger reuse of previous update subspaces, whereas deeper blocks require additional adaptation capacity through complementary directions. Based on this observation, we propose ReLoRA-CL, a preserve-or-expand strategy with historical-statistics maintenance for reliable prototype migration. We further formulate migration as drift estimation and show that structured updates can promote more regular representation changes. Experiments on ImageNet-R, CIFAR-100, and ImageNet-A demonstrate improved continual adaptation performance with lower drift variation and prototype displacement error, highlighting heterogeneous update reuse as an effective principle for continual low-rank adaptation.

REFERENCES

Chang Che, Ziqi Wang, Pengwan Yang, Cheems Wang, Hui Ma, and Zenglin Shi. LoRA in LoRA: Towards parameter-efficient architecture expansion for continual visual instruction tuning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 19978–19986, 2026.

Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory G. Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3366–3385, 2022.

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.

Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *International Conference on Learning Representations (ICLR)*, 2021.

Robert M. French. Catastrophic forgetting in connectionist networks. *Trends in Cognitive Sciences*, 3 (4):128–135, 1999.

Qiankun Gao, Chen Zhao, Yifan Sun, Teng Xi, Gang Zhang, Bernard Ghanem, and Jian Zhang. A unified continual learning framework with general parameter-efficient tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11483–11493, 2023.

Jiangpeng He, Zhihao Duan, and Fengqing Zhu. CL-LoRA: Continual low-rank adaptation for rehearsal-free class-incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30534–30544, 2025.

Dan Hendrycks, Steven Basart, Norman Mu, Saurav Kadavath, Frank Wang, Evan Dorundo, Rahul Desai, Tyler Zhu, Samyak Parajuli, Mike Guo, Dawn Song, Jacob Steinhardt, and Justin Gilmer. The many faces of robustness: A critical analysis of out-of-distribution generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8340–8349, 2021a.

Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15262–15271, 2021b.

Yuting Hou, Xuefei Tong, and Kedian Mu. FS-LoRA: Fast and slow low-rank adaptation for class incremental learning. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 3706–3710, 2026. doi: 10.1109/ICASSP55912.2026.11463466.

Neil Houlsby, Andrei Giurgiu, Stanislaw Jastrzebski, Bruna Morrone, Quentin de Laroussilhe, Andrea Gesmundo, Mona Attariyan, and Sylvain Gelly. Parameter-efficient transfer learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pp. 2790–2799, 2019.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. LoRA: Low-rank adaptation of large language models. In *International Conference on Learning Representations (ICLR)*, 2022.

Wangwang Jia, Zijian Gao, Tianjiao Wan, Yuan Cao, Yong Dou, and Kele Xu. Geometry-driven OOD detectors are class-incremental learners. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 34638–34649, 2026.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

Dawid Jan Kopiczko, Tijmen Blankevoort, and Yuki M. Asano. VeRA: Vector-based random matrix adaptation. In *International Conference on Learning Representations (ICLR)*, 2024.

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

Brian Lester, Rami Al-Rfou, and Noah Constant. The power of scale for parameter-efficient prompt tuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pp. 3045–3059, 2021.

Kai Li, Jiafeng Li, Lianghua He, and Ying Wen. DGS: Dual gradient and semantic-shift guided low-rank adaptation for class incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 32267–32277, 2026.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, pp. 4582–4597, 2021.

Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018. doi: 10.1109/TPAMI.2017.2773081.

Yan-Shuo Liang and Wu-Jun Li. InfLoRA: Interference-free low-rank adaptation for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23638–23647, 2024.

Shih-Yang Liu, Chien-Yi Wang, Hongxu Yin, Pavlo Molchanov, Yu-Chiang Frank Wang, Kwang-Ting Cheng, and Min-Hung Chen. DoRA: Weight-decomposed low-rank adaptation. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32100–32121, 2024.

Xuan Liu and Xiaobin Chang. Lora subtraction for drift-resistant space in exemplar-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15308–15318, 2025.

Michael McCloskey and Neal J. Cohen. Catastrophic interference in connectionist networks: The sequential learning problem. In *Psychology of Learning and Motivation*, volume 24, pp. 109–165. Academic Press, 1989.

Mark D. McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, and Anton van den Hengel. RanPAC: Random projections and pre-trained models for continual learning. In *Advances in Neural Information Processing Systems*, volume 36, 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 (CVPR)*, pp. 2001–2010, 2017.

James Seale Smith, Leonid Karlinsky, Vyshnavi Gutta, Paola Cascante-Bonilla, Donghyun Kim, Assaf Arbelle, Rameswar Panda, Rogerio Feris, and Zsolt Kira. CODA-Prompt: Continual decomposed attention-based prompting for rehearsal-free continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 11909–11919, 2023.

Liyuan Wang, Jingyi Xie, Xingxing Zhang, Mingyi Huang, Hang Su, and Jun Zhu. Hierarchical decomposition of prompt-based continual learning: Rethinking obscured sub-optimality. In *Advances in Neural Information Processing Systems*, volume 36, pp. 69054–69076, 2023.

Zifeng Wang, Zizhao Zhang, Sayna Ebrahimi, Ruoxi Sun, Han Zhang, Chen-Yu Lee, Xiaoqi Ren, Guolong Su, Vincent Perot, and Jennifer Dy. DualPrompt: Complementary prompting for rehearsal-free continual learning. In *European Conference on Computer Vision (ECCV)*, pp. 631–648, 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 (CVPR)*, pp. 139–149, 2022b.

Yichen Wu, Hongming Piao, Long-Kai Huang, Renzhen Wang, Wanhua Li, Hanspeter Pfister, Deyu Meng, Kede Ma, and Ying Wei. SD-LoRA: Scalable decoupled low-rank adaptation for class incremental learning. In *International Conference on Learning Representations (ICLR)*, 2025.

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.

Xiaobing Yu, Jin Yang, Xiao Wu, Peijie Qiu, and Xiaofeng Liu. FM-LoRA: Factorized low-rank meta-prompting for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 6463–6472, 2025.

Gengwei Zhang, Liyuan Wang, Guoliang Kang, Ling Chen, and Yunchao Wei. SLCA: Slow learner with classifier alignment for continual learning on a pre-trained model. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 19148–19158, 2023a.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *International Conference on Learning Representations (ICLR)*, 2023b.

Da-Wei Zhou, Hai-Long Sun, Jingyi Ning, Han-Jia Ye, and De-Chuan Zhan. Continual learning with Pre-Trained Models: A survey. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 8363–8371, 2024a.

Da-Wei Zhou, Qi-Wei Wang, Zhi-Hong Qi, Han-Jia Ye, De-Chuan Zhan, and Ziwei Liu. Class-Incremental Learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(12):9851–9873, 2024b.

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. doi: 10.1007/s11263-024-02218-0.

Fei Zhu, Xu-Yao Zhang, Chuang Wang, Fei Yin, and Cheng-Lin Liu. Prototype augmentation and self-supervision for incremental learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5871–5880, 2021.

Hao Zhu, Yifei Zhang, Junhao Dong, and Piotr Koniusz. BiLoRA: Almost-orthogonal parameter spaces for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25613–25622, 2025.

A RELATED WORK

A.1 PARAMETER-EFFICIENT CONTINUAL LEARNING

Continual learning aims to acquire knowledge from sequential tasks while mitigating catastrophic forgetting (McCloskey & Cohen, 1989; French, 1999; Kirkpatrick et al., 2017; Li & Hoiem, 2018; Rebuffi et al., 2017; De Lange et al., 2022; Zhou et al., 2024b). With the emergence of large pre-trained models (McDonnell et al., 2023; Zhou et al., 2025), parameter-efficient fine-tuning (PEFT) has become an effective paradigm for continual adaptation without updating the full backbone (Houlsby et al., 2019; Lester et al., 2021; Li & Liang, 2021; Zhou et al., 2024a). Among PEFT approaches, LoRA is particularly attractive for rehearsal-free class-incremental learning because it expresses model adaptation through compact low-rank updates (Hu et al., 2022; Gao et al., 2023; Liang & Li, 2024; Wu et al., 2025; He et al., 2025; Zhu et al., 2025; Liu & Chang, 2025).

Recent methods exploit the structure of low-rank updates to mitigate cross-task interference. InfLoRA constrains task-specific update directions (Liang & Li, 2024), while SD-LoRA identifies stable adaptation directions across incremental stages (Wu et al., 2025). Other PEFT studies investigate rank allocation, low-rank parameterization, and directional decomposition (Zhang et al., 2023b; Kopiczko et al., 2024; Liu et al., 2024). Collectively, these studies show that the geometry of low-rank adaptation is central to the stability–plasticity trade-off.

Our work studies a different question within this geometric view: *how much of an update subspace learned at an early stage remains functionally reusable by later tasks, and whether this reusability is uniform across network depth*. We characterize this behavior through cross-task update alignment, subspace compactness, and functional reconstruction, revealing a depth-dependent transition from persistent reuse to increasing demand for complementary directions. This observation motivates treating subspace preservation and subspace expansion as different responses to empirically distinct reuse regimes.

A.2 From Parameter Sharing to Update-Subspace Reuse

A related line of work organizes continual adaptation through shared and task-dependent parameters (He et al., 2025; Yu et al., 2025; Che et al., 2026; Hou et al., 2026; Jia et al., 2026). For example, LiLoRA employs a task-invariant input-side matrix to facilitate parameter sharing across tasks (Che et al., 2026), while CL-LoRA assigns shared and task-specific adaptation modules to different Transformer depths (He et al., 2025). These approaches motivate parameter sharing as an effective mechanism for continual adaptation.

ReLoRA-CL is instead formulated around the geometry of *learned update subspaces*. The relevant distinction is not whether parameters are labeled as shared or task-specific, but whether directions acquired from previous tasks remain sufficient to represent subsequent functional updates. In blocks exhibiting persistent reuse, Reusable Subspace Adaptation (RSA) retains the Task-0 learned input subspace and represents later adaptation through output-side recombination. This preserves the reusable span while allowing task-dependent coefficients to change. In blocks where reconstruction within the retained span progressively deteriorates, Complementary Direction Expansion (CDE) introduces directions outside the stored span and explicitly orthogonalizes them against previously acquired directions, while keeping historical directions fixed.

This yields a preserve-or-expand mechanism tied to measured subspace reusability: RSA exploits an existing learned span when it remains expressive, whereas CDE enlarges that span when reuse becomes insufficient. Accordingly, the methodological distinction lies in using the *cross-task sufficiency of learned update subspaces* to organize continual adaptation, rather than assigning sharing solely through parameter identity or module ownership. This formulation also enables us to study how the resulting update geometry affects representation drift and the maintenance of historical class statistics.

B Proofs and Scope of the Drift Analysis

B.1 Proof of Prop. 1

Condition on the extractors, pre-update features, and stored statistics under Assumption 1. The old-class zero-mean residual condition gives $d_c^t = \mathbb{E} g_t(Z_c)$. Decompose

$$\widehat{d}_c^t - d_c^t = \underbrace{\sum_i \omega_{ic}(g_t(z_i) - \mathbb{E} g_t(Z_c))}_{b_c} + \underbrace{\sum_i \omega_{ic}\xi_i}_{\eta_c}. \tag{11}$$

Since weights are nonnegative and sum to one, Lipschitz continuity yields

$$\|b_c\|_2 \le \sum_i \omega_{ic} \mathbb{E} \|g_t(z_i) - g_t(Z_c)\|_2 \le L_t r_c^t. \tag{12}$$

Conditional zero means and uncorrelated residuals imply

$$\mathbb{E} \|\eta_c\|_2^2 = \sum_i \omega_{ic}^2 \mathbb{E} \|\xi_i\|_2^2 \le \frac{\sigma_t^2}{n_{\mathrm{eff},c}^t}. \tag{13}$$

The triangle inequality in L_2 proves Eq. 2. Applying the ordinary triangle inequality to $z_i - Z_c = (z_i - \widehat{\mu}_c^{t-1}) + (\widehat{\mu}_c^{t-1} - \mu_c^{t-1}) + (\mu_c^{t-1} - Z_c)$ proves Eq. 3. The support quantity involves the unavailable old-class distribution, making this an explanatory bound rather than a directly computable certificate during exemplar-free training.

B.2 Derivation of the Local Regularity Bound

Subtract Eq. 4 at two feature locations and apply Eq. 5:

$$\begin{aligned}\|g_t(z) - g_t(z')\|_2 &\le \sum_l \left\|(\mathcal{J}_{l,t}(z) - \mathcal{J}_{l,t}(z'))[G_l^t]\right\|_2 + \|e_t(z) - e_t(z')\|_2 \\ &\le \left(\sum_l \gamma_{l,t} \left\|G_l^t\right\|_F + \varepsilon_t\right) \|z - z'\|_2 . \end{aligned} \tag{14}$$

This establishes Eq. 6. Substituting this Lipschitz bound into Eq. 2 controls its bias term by $(C_t+\varepsilon_t)r_c^t$. Comparing algorithms still requires accounting for support, residual variance, sensitivities, and remainders.

B.3 Proof of Prop. 2

An orthogonal projector satisfies $P_l = P_l^\top = P_l^2$, so $G_l P_l$ and $G_l(I - P_l)$ are orthogonal in the Frobenius inner product:

$$\|G_l\|_F^2 = \|G_l P_l\|_F^2 + \|G_l(I - P_l)\|_F^2 . \tag{15}$$

Thus $\|G_l P_l\|_F \le \|G_l\|_F$. Multiply by $\gamma_l \ge 0$ and sum, leaving other blocks unchanged, to obtain Eq. 7. Strict reduction requires a nonzero off-subspace component in a projected block with positive sensitivity.

B.4 Accumulated Error and Interpretation

Let $e_c^t = \widehat{\mu}_c^t - \mu_c^t$. Since $\mu_c^t = \mu_c^{t-1} + d_c^t$, migration gives the exact recursion

$$e_c^t = e_c^{t-1} + \widehat{d}_c^t - d_c^t. \tag{16}$$

Errors can accumulate from class arrival onward; the one-stage result does not give a uniform guarantee over arbitrary task sequences. The analysis also concerns means, not covariance preservation or classifier accuracy. These distinctions motivate separate measurements of consecutive-task update norms, local drift variation, oracle prototype displacement error, and classification performance. Historical samples used for oracle diagnostics must remain outside training and model selection.

C Validation of the Reusable Subspace Assumption

We validate the reusable subspace assumption on ImageNet-R with 10 sequential tasks ($T = 10$). The proposed Reusable Subspace Adaptation (RSA) relies on the observation that later-task LoRA updates in shallow Transformer blocks largely reuse the Task-0 learned input subspace. To further examine this property, we measure whether the input-side factors learned by later tasks can be reconstructed from the fixed invariant basis.

Specifically, let A_{RSA} denote the fixed Task-0 input factor and A_t denote the input factor learned by task t. If both factors span the same r-dimensional subspace, there exists an invertible transformation matrix Q_t such that $A_t = Q_t A_{\mathrm{RSA}}$. Under this condition, the corresponding LoRA update can be rewritten as

$$B_t A_t = B_t Q_t A_{\mathrm{RSA}} = \widetilde{B}_t A_{\mathrm{RSA}}, \tag{17}$$

which indicates that optimizing the input factor does not introduce new representational capacity when the underlying subspace remains unchanged. The additional task-specific variation can instead be absorbed by the output-side factor.

To empirically examine this property, we compare the functional similarity between the original LoRA update using the freely optimized input factor A_t and its reconstruction using the fixed invariant basis A_{RSA}. Figure 5 reports the reconstruction similarity across incremental tasks and Transformer layers under the ImageNet-R ($T = 10$) setting.

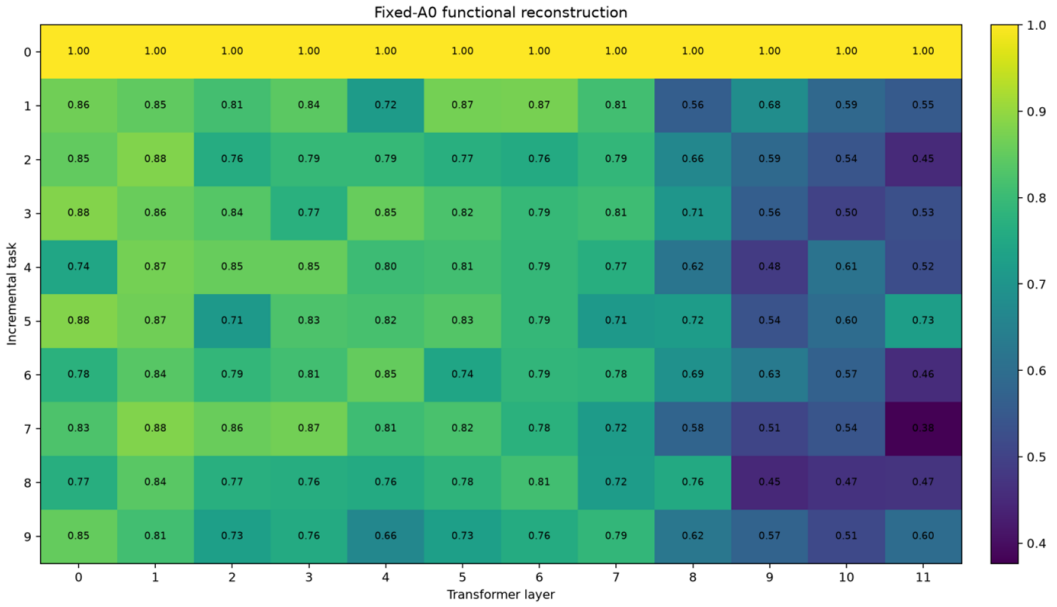


Figure 5: Functional reconstruction similarity between freely optimized LoRA updates and the fixed Task-0 invariant basis across incremental tasks and Transformer layers. Higher similarity indicates stronger reuse of the invariant subspace.

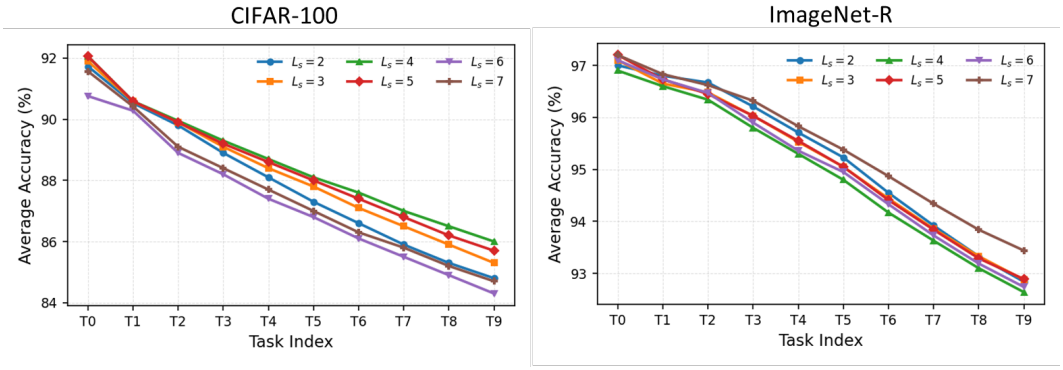


Figure 6: Sensitivity analysis of the layer-wise decomposition boundary L_s in ReLoRA-CL on ImageNet-R and CIFAR-100 with 10 sequential tasks.

The results show that shallow Transformer blocks maintain consistently high reconstruction similarity, indicating that later-task updates largely remain within the invariant subspace learned from Task 0. In contrast, deeper blocks exhibit lower similarity as task-specific variations accumulate. This observation supports our layer-wise decomposition strategy: shallow blocks are suitable for RSA, while deeper blocks require additional adaptation capacity through CDE.

D Sensitivity Analysis of Layer-wise Decomposition

The proposed ReLoRA-CL framework employs a layer-wise decomposition that assigns shallow Transformer blocks to Reusable Subspace Adaptation (RSA) and deeper blocks to Complementary Direction Expansion (CDE). The boundary layer L_s determines the transition between these two adaptation strategies.

To evaluate the sensitivity to this design choice, we vary L_s under ImageNet-R and CIFAR-100 with 10 sequential tasks. Different L_s values represent different trade-offs between preserving reusable subspaces and introducing complementary directions.

As shown in Fig. 6, ReLoRA-CL maintains comparable performance across a wide range of layer partitions on both benchmarks, indicating robustness to the choice of L_s. Although different datasets favor slightly different partitions, no single boundary dominates across all settings. These results suggest that the effectiveness of ReLoRA-CL comes from the layer-wise adaptation principle rather than a specific manually selected split.

This analysis further supports our observation that Transformer blocks exhibit heterogeneous update reuse patterns, motivating different adaptation strategies for different layers.

AI USE STATEMENT

Generative AI tools were used only for limited assistance with writing clarity, grammar checking, and formatting refinement. They were not used for designing the methodology, conducting experiments, performing theoretical analysis, or generating scientific claims. All AI-assisted content was reviewed and revised by the authors, who take full responsibility for the final manuscript.

ETHICS STATEMENT

This work does not involve human subjects, sensitive personal data, or potentially harmful applications. The proposed method is intended for research purposes, and we have carefully considered issues related to research integrity, reproducibility, and responsible use of machine learning techniques.

REPRODUCIBILITY STATEMENT

To facilitate reproducibility, we provide the anonymous source code at: `https://anonymous.4open.science/r/ReLoRA-8C3A/`. The implementation details, experimental settings, and additional analyses are provided in the main paper and supplementary materials. We release the code to enable independent verification and further research on our approach.