VD-LoRA: Adaptive Reuse of Low-Rank Directions for Continual Learning

Luqiong Ding[1,2*], Jiayao Tan[3*], Chenggong Ni[1,2], Fuyuan Hu[1,4,5†], and Fan Lyu[6†]

[1] School of Electronics and Information Engineering, Suzhou University of Science and Technology, China
[2] Suzhou Key Laboratory of Embodied Intelligent Agents for Cooperative Perception and Advanced Control, China
[3] School of Artificial Intelligence, Tianjin University, China
[4] Jiangsu Industrial Intelligent and Low-carbon Technology Engineering Center, China
[5] Suzhou Key Laboratory of Intelligent and Low-carbon Technology Application, China
[6] Computer Vision Center, Universitat Autònoma de Barcelona, Spain
{quliongding,cgn}@post.usts.edu.cn, jiayaotan@tju.edu.cn, fuyuanhu@mail.usts.edu.cn, fanlyu@cvc.uab.cat

Abstract. Class-incremental learning (CIL) aims to continually acquire new classes while preserving previously learned knowledge, posing the well-known stability–plasticity dilemma. With the emergence of large pre-trained models, parameter-efficient fine-tuning methods such as LoRA have become a practical paradigm for rehearsal-free CIL by restricting adaptation to low-rank weight updates. Recent LoRA-based approaches often mitigate cross-task interference by enforcing orthogonality between task updates. However, under strict rank budgets and long task sequences, uniformly protecting previously used directions can become overly conservative, progressively shrinking the effective update space and limiting plasticity. To address this limitation, we propose Variational Direction-aware LoRA (VD-LoRA), a Bayesian framework that enables adaptive consolidation within the low-rank update space. VD-LoRA maintains a recursive variational posterior over LoRA directions to estimate direction-wise uncertainty, using posterior precision to adaptively control protection strength. Based on this uncertainty signal, we further design a direction selection mechanism that favors updates that are both task-relevant and safe to adapt. This enables selective reuse of low-risk directions while preserving critical ones, maintaining a task-aligned update subspace under fixed rank constraints. Extensive experiments on multiple rehearsal-free CIL benchmarks demonstrate that VD-LoRA consistently outperforms existing LoRA-based methods while improving long-horizon stability and plasticity. Our code is available at https://github.com/LuQiongD/VD-LoRA.

*Equal contribution.
†Corresponding authors.

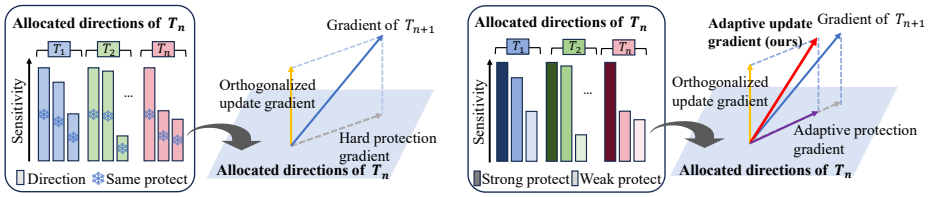
Keywords: Continual Learning · Class-Incremental Learning · Low-Rank Adaptation · Bayesian Inference · Uncertainty Estimation

1 Introduction

Class-incremental learning (CIL) [5, 20, 25, 29, 30, 33] learns new classes sequentially while retaining knowledge acquired from previous tasks. A key challenge in CIL is the stability-plasticity dilemma [7, 14, 16, 32, 40], where the model must both preserve knowledge from previous tasks and maintain sufficient plasticity for new ones. In the era of large-scale pre-trained models (PTMs) [1, 3], parameter-efficient fine-tuning (PEFT) [8, 9, 21, 31, 38] updates only a small number of task-related parameters, such as adapters [12], low-rank adaptation (LoRA) [13], or prompt tuning [18], while keeping the pre-trained backbone largely frozen, thereby reducing interference and storage overhead. Specifically, LoRA parameterizes task adaptation as explicit low-rank updates in weight space, providing a structured update space in which cross-task sharing and interference control can be naturally imposed. Consequently, LoRA-based CIL has rapidly emerged as an important research direction, and recent methods have explored diverse ways to organize, protect, and reuse low-rank updates across tasks in order to better balance stability and plasticity [9, 21, 24, 34, 41, 43].

Among these methods, orthogonality-based LoRA [21, 34, 35, 43] is designed to mitigate cross-task interference by *enforcing orthogonality*, encouraging the LoRA update for a new task to avoid directions already occupied by previous tasks. Under the common fixed-basis parameterization, the low-rank update can be decomposed into column-wise components, and we refer to each basis coordinate, equivalently each column-wise update component, as an update direction [35, 43]. However, under strict rank budgets and long task sequences, orthogonality alone can be overly conservative. Existing methods [21, 34, 35, 43] typically treat previously allocated update directions as permanently protected, without distinguishing truly sensitive directions from those that can be safely reused. Consequently, once a direction is occupied, it is subject to persistent hard protection, progressively shrinking the effective update space. As tasks accumulate, the rank-r subspace becomes increasingly misaligned with task-relevant gradients, weakening plasticity.

As illustrated in Fig. 1, existing orthogonality-based methods implicitly assume that all previously allocated directions are equally critical and should be strictly preserved. However, this uniform consolidation strategy ignores the heterogeneity of direction-wise sensitivity, causing unnecessary restrictions on future updates. *We argue that orthogonality should be applied adaptively rather than uniformly.* While distinct tasks should be separated when alternative directions exist, previously used directions should remain reusable when their interference risk is low. The key challenge is therefore to estimate, at a direction-wise level, which components require strong protection and which can be safely adapted, so as to preserve stability without sacrificing plasticity under a fixed rank budget.



(a) Static uniform protection (b) Uncertainty adaptive protection

Fig. 1: Under strict rank budgets, existing orthogonality-based methods protect previously allocated directions in a static and uniform manner, which reduces interference but progressively shrinks the effective update space and weakens plasticity. Our method instead uses recursive uncertainty to assign direction-wise protection strength, preserving sensitive directions while allowing safe reuse of less critical ones, so that the resulting update better aligns with the current task gradient.

To address this limitation, we propose Variational Direction-aware LoRA (VD-LoRA), a rehearsal-free framework that enables adaptive, direction-wise consolidation in low-rank space. Instead of uniformly protecting all previously used directions, VD-LoRA estimates their sensitivity and adjusts protection strength accordingly. Specifically, we introduce a recursive variational posterior over LoRA directions to model their task-wise uncertainty. The posterior precision serves as a direction-wise consolidation strength: highly confident directions are protected more strongly, while uncertain ones remain available for selective reuse. Building on this uncertainty estimate, we further design an uncertainty-aware direction selection mechanism that favors directions which are both informative for the current task and sufficiently safe to update. In this way, VD-LoRA jointly determines where to update and how strongly to protect, preserving orthogonal decoupling when beneficial while avoiding unnecessary exclusion of reusable directions. As a result, the effective low-rank subspace remains better aligned with task-relevant gradients under a fixed rank budget, improving plasticity without sacrificing stability over long task sequences. In summary, we enable the model to adaptively adjust the protection strength for old knowledge as tasks progress, so that it can mitigate forgetting without progressively shrinking the learning space for new tasks. Extensive experiments on multiple rehearsal-free class-incremental benchmarks show that our method outperforms existing LoRA-based approaches.

Our contributions are summarized as follows.

(1) We introduce a direction-wise adaptive consolidation framework for LoRA-based class-incremental learning, replacing uniform hard orthogonality with uncertainty-calibrated soft protection in the low-rank space.
(2) We formulate a recursive variational posterior over LoRA directions to quantify consolidation strength and derive an uncertainty-aware selection criterion; together, they determine which directions to update and how strongly to protect them under strict rank constraints.

(3) We further introduce stability regularization by applying uncertainty-scaled perturbations to unselected directions, thereby improving the model's robustness over long task sequences.

2 Related Work

2.1 LoRA-based CIL

PEFT has been widely used in CIL to adapt large PTMs while keeping most backbone parameters frozen. Prompt-based approaches [18, 31, 37, 38] learn task prompts or a prompt pool, but they typically require maintaining prompts and selecting them at inference, and sequential updates in a shared prompt space can still induce cross-task interference, leading to storage growth or additional inference overhead in rehearsal-free settings. In contrast, LoRA [13] adapts transformers through low-rank updates in weight space and avoids maintaining a prompt pool or prompt routing at inference. Recent LoRA-based CIL methods [9, 21, 22, 24, 35, 39, 41, 43] mainly study how to organize and protect adapters across tasks. CL-LoRA [9] and SD-LoRA [41] improve stability by freezing or reweighting previous adapters with task-specific composition, while LaST-LoRA [22] and LoRA Subtraction [24] address long-horizon drift by tracking adapter changes or compensating drift via difference-based updates. Orthogonality-based methods [21, 35, 43] further reduce interference by separating update directions or allocating disjoint subspaces. However, under strict rank budgets and long task sequences, deterministic orthogonality with persistent hard protection can become overly conservative, progressively shrinking the effective update space. Moreover, existing approaches lack direction-wise uncertainty estimation to distinguish truly sensitive directions from safely reusable ones. This limitation motivates our direction-wise Bayesian formulation for adaptive consolidation in low-rank space.

2.2 Bayesian inference in CL

Bayesian continual learning [2, 4, 6, 19, 27] mitigates forgetting by modeling parameters as distributions and propagating posterior uncertainty across tasks. In variational or Laplace-based formulations [2, 4, 6], the posterior from the previous task serves as a prior for the next, and uncertainty naturally determines how strongly each parameter should be preserved. However, existing Bayesian CL methods primarily operate on full model weights or coarse layer-wise scales, with limited exploration in highly constrained low-rank adapter spaces. Although recent works [26, 36] introduce uncertainty modeling for LoRA parameters, recursive posterior estimation at the level of individual low-rank directions remains underexplored. We address this gap by performing variational Bayesian inference directly in the LoRA direction space, enabling direction-wise adaptive protection and principled reuse under strict rank budgets.

3 Preliminary: Orthogonality-Based LoRA in CIL

Task Formulation. In class-incremental learning (CIL), the model is trained on a sequence of tasks $\{T_t\}_{t=1}^{N}$, where each task introduces new classes and past training data are not revisited under the rehearsal-free constraint. The goal is to learn new tasks while mitigating catastrophic forgetting [23, 28].

Orthogonality-Based LoRA in CIL. Following existing LoRA-based CIL methods [21, 34, 35, 43], we adopt LoRA to adapt a frozen pretrained backbone by adding a low-rank update to a weight matrix:

$$W_t = W_0 + \Delta W_t, \qquad \Delta W_t = B_t A_t, \tag{1}$$

where only the low-rank factor B_t is trainable.

To reduce inter-task interference, orthogonality-based methods allocate task updates into disjoint coordinate subspaces. Concretely, they introduce a fixed dictionary of basis vectors $\mathcal{E} = \{e_i\}_{i=1}^{d_{\text{in}}}$ and activate a size-r index set S_t per task. Under this formulation, the update can be written as a column-wise decomposition:

$$\Delta W_t \;=\; \sum_{i \in S_t} b_{t,i}\, e_i^{\top}, \tag{2}$$

where each e_i represents a **low-rank update direction** (i.e., a column-basis component in the update space) and $b_{t,i}$ is its coefficient vector. For any two distinct tasks $s \neq t$, let S_s and ΔW_s be defined analogously. Strict orthogonality between tasks is enforced by requiring $S_s \cap S_t = \emptyset$, which yields $\langle \Delta W_s, \Delta W_t \rangle_F = 0$ and isolates task-specific knowledge into non-overlapping subspaces [35, 43].

Limitation of Strict Orthogonality. While effective at reducing interference, strict disjoint allocation becomes overly conservative under tight rank budgets and long task sequences. As tasks accumulate, the available update directions are rapidly exhausted, forcing later tasks to adapt within an increasingly restricted residual subspace. This often leads to gradient–subspace mismatch and under-adaptation, weakening plasticity. Moreover, the binary "*once-used, always-frozen*" rule ignores that different directions have heterogeneous sensitivity, unnecessarily preventing safe reuse and accelerating capacity collapse. These limitations motivate a direction-wise, adaptive protection mechanism that preserves sensitive directions while allowing low-risk reuse.

4 Methodology

To overcome the plasticity limitation caused by gradient–subspace mismatch under strict orthogonality, we propose **Variational Direction-aware Low-Rank Adaptation (VD-LoRA)**, which replaces rigid geometric exclusion with an uncertainty-aware probabilistic framework for flexible direction reuse while preserving the decoupling benefits of orthogonal directions. An overview of the proposed framework is shown in Figure 2. Specifically, we first introduce *Variational Modeling of Direction Uncertainty*, where LoRA direction coefficients are

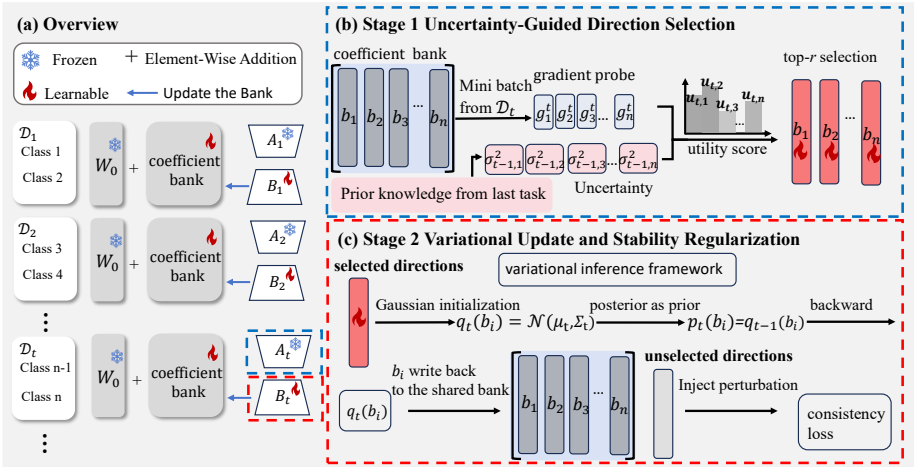


Fig. 2: Overview of the proposed VD-LoRA. Our method performs rehearsal-free continual learning on a frozen backbone with a shared LoRA column bank. Before training task t, we select r update directions via online probing to construct A_t. During training, we conduct variational updates on the selected columns under a recursive prior and commit the updated statistics back to the bank for reuse in subsequent tasks. We additionally regularize the model against uncertainty-scaled perturbations along unselected directions to improve robustness to future updates.

modeled with recursive variational posteriors so that the posterior from the previous task serves as the prior for the next task, enabling the model to track the historical uncertainty of each direction. Based on this uncertainty signal, we then perform *Uncertainty-Guided Direction Selection*, a lightweight gradient probing mechanism that ranks candidate directions by balancing their potential adaptation gain against the risk of interfering with highly consolidated knowledge. To further maintain structural stability, we apply *Stability Regularization for Unselected Directions*, where unselected directions remain frozen in the shared dictionary and their uncertainty is used to construct perturbations, while the selected directions are optimized under these perturbations to improve robustness to potential future updates. Finally, posterior predictive averaging is employed during inference to improve predictive robustness and calibration. Together, these components construct a task-aligned low-rank update space that enhances plasticity while preserving previously acquired knowledge over long task sequences.

4.1 Variational Modeling of Direction Uncertainty

Strict orthogonality freezes a direction once it is used, implicitly assuming that all previously activated directions are equally sensitive. In practice, however, different directions contribute unequally to previously learned knowledge. Some directions encode strongly consolidated information and should be protected more carefully, while others remain relatively flexible and can still be reused. To

capture this heterogeneity, we model LoRA direction coefficients probabilistically and interpret the posterior variance as a direction-wise uncertainty signal.

Consider a LoRA-injected layer with pretrained weight W_0. Using a fixed direction dictionary $\{e_i\}$, the LoRA update can be expressed as a direction-wise decomposition

$$W_t = W_0 + \sum_{i=1}^{d_{\text{in}}} b_i e_i^\top, \tag{3}$$

where b_i denotes the coefficient vector associated with direction e_i. Each direction therefore represents an independent component of the low-rank update.

To estimate how sensitive each direction is to future updates, we place a variational posterior on the coefficient of every direction:

$$q_t(b_i) = \mathcal{N}(\mu_{t,i}, \text{diag}(\sigma_{t,i}^2)), i \in S_t, \tag{4}$$

where S_t denotes the set of active directions for task t (determined in Section 4.2). Directions outside S_t retain their previously stored posterior statistics. For the first task, since all directions share the same prior variance, *the initial selection S_1 depends only on the gradient probing scores*, $\text{diag}(\sigma_{t,i}^2)$ denotes a diagonal covariance matrix constructed from the variance vector $\sigma_{t,i}^2$, and $\mu_{t,i}$ represents the expected contribution of direction i, while $\sigma_{t,i}^2$ quantifies the uncertainty associated with that direction. Intuitively, directions with smaller variance are more strongly consolidated by previous tasks and should therefore be protected from large updates, whereas directions with larger variance remain more adaptable and can tolerate reuse.

To accumulate knowledge across tasks, each direction maintains a recursive prior. Specifically, the posterior obtained at the most recent activation of a direction is stored and reused as its prior in later tasks. Denote this prior as $p_t(b_i)$ and $p_1(b_i) = \mathcal{N}(0, I)$. Under a fixed rank budget, only a subset of directions will be updated for each task. The training objective is

$$\mathcal{L}_{\text{v}} = \mathbb{E}_{q_t}[\mathcal{L}_{\text{task}}(W_t; \mathcal{D}_t)] + \beta \sum_{i \in S_t} \text{KL}(q_t(b_i) \,\|\, p_t(b_i)), \tag{5}$$

where $\mathcal{L}_{\text{task}}$ denotes the task loss (Cross-Entropy) and β balances stability and plasticity. *Optimizing Eq. (5) updates the variational posterior $q_t(b_i)$*, which is then stored and used as the prior for future tasks. Since both posterior and prior are Gaussian, the KL term admits a closed-form expression:

$$\text{KL}(q \,\|\, p) = \frac{1}{2} \sum_k \left[\log \frac{\sigma_{p,k}^2}{\sigma_k^2} + \frac{\sigma_k^2 + (\mu_k - \mu_{p,k})^2}{\sigma_{p,k}^2} - 1 \right]. \tag{6}$$

This formulation provides a direction-wise uncertainty estimate. *Directions with small prior variance are strongly consolidated and thus penalized more heavily by the KL term, while directions with larger uncertainty remain more flexible.* At the beginning of task t, the inherited posterior variances $\sigma_{t-1,i}^2$ provide the uncertainty signals used to determine which directions should be activated. The derivation of the Gaussian KL expression and complete details of the variational formulation are provided in the supplementary material.

4.2 Uncertainty-Guided Direction Selection

Given the direction-wise uncertainty estimated in Section 4.1, we now determine which directions should be activated for the current task under the rank budget r. An ideal direction should both help reduce the current task loss and pose minimal risk to previously consolidated knowledge.

Combining Uncertainty with Gradient Signals. To measure the usefulness of each direction, we estimate the gradient of the task loss with respect to its coefficient. Let $g_{t,i}$ denote this gradient. The uncertainty $\sigma^2_{t-1,i}$ inherited from previous tasks indicates how safely the direction can be reused. Combining these two signals, we define a utility score:

$$u_{t,i} = g_{t,i}^\top \operatorname{diag}\Big(\big(\sigma^2_{t-1,i} + \varepsilon\big)^\gamma\Big)\, g_{t,i}, \tag{7}$$

where ε is a small constant for numerical stability and γ controls the contribution of posterior uncertainty to direction selection. This score is large when a direction both aligns with the current gradient and retains sufficient uncertainty. Conversely, directions with very small variance have been strongly consolidated and thus receive lower utility.

Direction Selection. After computing the utility scores for all directions, we select the top-r directions to form the active set:

$$S_t = \text{Top-}r\left(\{u_{t,i}\}_{i=1}^{d_{\text{in}}}\right). \tag{8}$$

In practice, the gradients $g_{t,i}$ are estimated using a small number of mini-batches from the current task before training begins. Once S_t is determined, only the variational parameters of these directions are updated according to the objective in Eq. (5), while all other directions remain fixed. A detailed derivation of the utility score, together with controlled experiments validating its design, is provided in the supplementary material.

4.3 Stabilizing Unselected Directions

The uncertainty estimated in Section 4.1 indicates how sensitive each direction is to future updates. While the recursive prior constrains the directions selected for the current task in Section 4.2, the remaining directions ($i \notin S_t$) may still be activated by future tasks. If the current model is overly sensitive to these inactive directions, subsequent updates may cause abrupt shifts of the decision boundary. To improve robustness to such future updates, we introduce a stability regularization based on uncertainty-scaled perturbations.

Specifically, we simulate potential future updates by injecting perturbations associated with the unselected directions. For a layer input feature h, we construct a perturbed feature:

$$\tilde{h} = h + \rho \sum_{i \notin S_t} \sigma_{t-1,i} \odot \epsilon_i, \tag{9}$$

where $\epsilon_i \sim \mathcal{N}(0, I)$, $\odot$ denotes element-wise multiplication, and ρ controls the perturbation magnitude. The perturbation scale is proportional to the uncertainty $\sigma_{t-1,i}$ obtained from the variational posterior in Section 4.1, so directions with larger residual uncertainty induce stronger perturbations. This encourages the model to remain robust to updates in directions that are more likely to be reused.

Let p and $\tilde{p}$ denote the model predictions from the original and perturbed features, respectively. We enforce prediction consistency using a KL divergence:

$$\mathcal{L}_{\mathrm{c}} = \mathrm{KL}(\mathrm{sg}(p) \,\|\, \tilde{p}) \,, \tag{10}$$

where $\mathrm{sg}(\cdot)$ denotes the stop-gradient operation. The overall objective becomes:

$$\mathcal{L} = \mathcal{L}_{\mathrm{v}} + \lambda \, \mathcal{L}_{\mathrm{c}}, \tag{11}$$

where $\mathcal{L}_{\mathrm{v}}$ is the variational objective defined in Section 4.1, and λ controls the strength of the stability regularization.

Summary of Workflow. For each task, VD-LoRA follows a three-stage procedure. First, a lightweight probing stage evaluates gradient demand and directional uncertainty to rank candidate directions and select an active subset under the rank budget. Second, the model performs variational adaptation by updating only the selected directions, while the remaining directions remain fixed and their uncertainty is used to construct perturbations for consistency regularization. Finally, the shared bank stores the updated posterior statistics, and predictions are obtained via Monte Carlo averaging to produce uncertainty-aware outputs across the task sequence.

5 Experiments

5.1 Experimental Setup

Datasets. We evaluate our method on several widely used continual learning benchmarks, including CIFAR-100 [17], ImageNet-R [10], ImageNet-A [11], and VTAB [42]. CIFAR-100 contains 100 classes of natural images and serves as a standard benchmark for class-incremental learning. ImageNet-R comprises 200 ImageNet classes rendered in diverse artistic styles such as cartoons, sketches, and paintings, and is widely used to evaluate PEFT-based continual learning under appearance shifts. ImageNet-A consists of 200 natural adversarial categories that are intentionally difficult for standard classifiers, making it suitable for evaluating robustness under stronger distribution shifts. VTAB is a collection of 19 visual classification tasks spanning natural, specialized, and structured domains. For class-incremental evaluation, we report results under multiple task granularities, including 5, 10, and 20 tasks, while keeping the class set and training protocol unchanged.

Architectural and Training Details. All methods are implemented in PyTorch and follow a protocol similar to previous LoRA-based continual learning

Table 1: Comparison on ImageNet-R with varying numbers of tasks N.

	ImageNet-R ($N = 5$)		*ImageNet-R* ($N = 10$)		*ImageNet-R* ($N = 20$)	
Method	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**
L2P	64.13 (±0.78)	68.66 (±0.41)	62.54 (±0.24)	67.98 (±0.27)	57.92 (±0.28)	64.57 (±0.29)
DualPrompt	67.88 (±0.17)	71.16 (±0.31)	65.41 (±0.52)	69.39 (±0.43)	61.00 (±0.72)	65.80 (±0.67)
CODA-Prompt	73.09 (±0.83)	76.91 (±0.21)	71.47 (±0.35)	75.82 (±0.29)	67.28 (±0.30)	72.34 (±0.17)
CL-LoRA	79.88 (±0.52)	84.98 (±0.49)	78.94 (±0.54)	84.49 (±0.61)	76.11 (±0.24)	83.23 (±0.92)
SD-LoRA	78.65 (±0.23)	82.98 (±0.44)	77.01 (±0.43)	81.67 (±0.18)	75.26 (±0.36)	80.03 (±0.55)
PLAN	77.79 (±0.24)	81.93 (±0.63)	75.25 (±0.42)	80.41 (±0.56)	71.06 (±0.42)	77.93 (±0.56)
LoRA-DRS	76.72 (±0.42)	81.97 (±0.64)	74.84 (±0.72)	81.76 (±0.29)	74.64 (±0.73)	80.99 (±0.98)
BiLoRA	78.32 (±0.37)	83.31 (±0.52)	77.95 (±0.14)	81.52 (±0.26)	72.41 (±0.65)	79.28 (±0.76)
InfLoRA	76.88 (±0.25)	82.04 (±0.31)	74.78 (±0.42)	80.67 (±0.33)	70.09 (±0.82)	76.87 (±0.82)
VD-LoRA (Ours)	**80.22** (±0.52)	**85.77** (±0.38)	**79.55** (±0.39)	**84.97** (±1.23)	**77.53** (±0.35)	**83.90** (±0.77)

Table 2: Comparison on CIFAR-100, ImageNet-A and VTAB.

	CIFAR-100		*ImageNet-A*		*VTAB*	
Method	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**
L2P	83.31 (±0.52)	88.20 (±0.45)	40.26 (±0.25)	49.83 (±0.78)	72.26 (±0.75)	77.83 (±0.68)
DualPrompt	84.51 (±0.21)	90.02 (±0.08)	41.26 (±0.34)	53.84 (±0.25)	80.26 (±0.15)	83.83 (±0.28)
CODA-Prompt	86.65 (±0.17)	91.29 (±0.34)	42.58 (±0.33)	53.68 (±0.44)	81.26 (±0.25)	84.23 (±0.57)
CL-LoRA	86.86 (±0.36)	91.65 (±0.63)	59.14 (±0.39)	69.03 (±1.45)	93.66 (±0.29)	94.13 (±0.48)
SD-LoRA	87.92 (±0.41)	90.62 (±1.17)	56.59 (±0.62)	66.34 (±1.80)	93.26 (±0.41)	93.83 (±0.57)
PLAN	87.54 (±0.31)	92.21 (±0.35)	—	—	—	—
LoRA-DRS	88.52 (±0.37)	92.01 (±0.16)	59.05 (±0.23)	69.49 (±1.73)	89.45 (±0.41)	92.65 (±0.62)
BiLoRA	86.02 (±0.46)	90.55 (±0.41)	52.22 (±0.73)	63.34 (±0.66)	88.53 (±0.53)	90.77 (±0.92)
InfLoRA	86.31 (±0.88)	91.22 (±0.56)	47.40 (±1.01)	61.47 (±0.74)	87.44 (±3.09)	88.75 (±1.26)
VD-LoRA (Ours)	**89.69** (±0.31)	**92.72** (±0.11)	**60.27** (±0.88)	**70.18** (±1.24)	**94.11** (±0.22)	**94.45** (±0.63)

works [21]. We adopt a ViT-B/16 backbone pre-trained on ImageNet-21K and further fine-tuned on ImageNet-1K before the CIL phase. During training, we use the Adam optimizer [15] with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, and a batch size of 128. Each task is trained for 50 epochs on ImageNet-R, 50 epochs on ImageNet-A, and 20 epochs on CIFAR-100. Following prior LoRA-based methods, LoRA modules are inserted into the key and value projections of the ViT attention blocks. Unless otherwise specified, we use rank $r = 10$ and report the mean and standard deviation over five independent runs.

Baselines. We compare VD-LoRA with several representative continual learning methods, including L2P [38], DualPrompt [37], CODA-Prompt [31], InfLoRA [21], SD-LoRA [41], CL-LoRA [9], BiLoRA [43], LoRA-DRS [24], and PLAN [35]. For all approaches, we reproduce the reported methods under our experimental setup to ensure a fair and direct comparison.

Evaluation Metrics. We report the final accuracy (Acc) after the last task and the average anytime accuracy (AAA), computed by averaging the accuracy over all post-task evaluations. To assess predictive calibration and uncertainty quality, we additionally report negative log-likelihood (NLL) and expected calibration error (ECE). Specifically, given a test set $\{(x_n, y_n)\}_{n=1}^{N}$ and the predicted probability $P_\theta(y_n \mid x_n)$ assigned to the ground-truth label, we compute the neg-

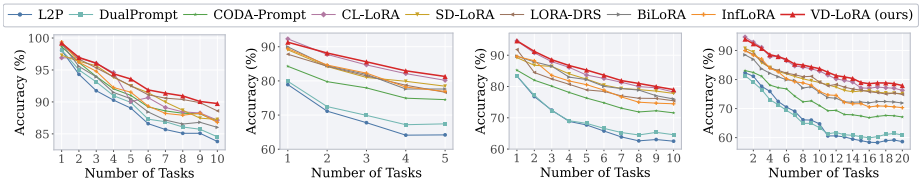


Fig. 3: Variation of the performance of different methods during the learning of ImageNet-R and CIFAR-100.

ative log-likelihood (NLL) and expected calibration error (ECE) as follows:

$$\mathrm{NLL} = -\frac{1}{N}\sum_{n=1}^{N} \log P_\theta(y_n \mid x_n), \tag{12}$$

$$\mathrm{ECE} = \sum_{m=1}^{M} \frac{|B_m|}{N} \left|\mathrm{acc}(B_m) - \mathrm{conf}(B_m)\right|, \tag{13}$$

where B_m denotes the m-th confidence bin, and $\mathrm{acc}(B_m)$ and $\mathrm{conf}(B_m)$ denote the average accuracy and confidence within that bin, respectively.

5.2 Main Results on Benchmarks

Table 1 reports the main results on ImageNet-R under different task settings, while Table 2 summarizes the results on CIFAR-100, ImageNet-A, and VTAB. Across datasets and task granularities, VD-LoRA consistently outperforms prompt-based baselines and deterministic LoRA baselines in terms of Acc and AAA. Under a strict low-rank budget, these results indicate that direction-level Bayesian tracking can allocate adaptation directions while regulating inter-task interference through uncertainty-aware protection. The advantage of VD-LoRA becomes more pronounced as the task sequence length increases. Figure 3 further shows the accuracy evolution over the task sequence on ImageNet-R and CIFAR-100. Compared with prior methods, VD-LoRA exhibits smaller performance drops when encountering new tasks and maintains a higher accuracy trajectory in later stages. Overall, the results support that selectively reusing empirically safe directions, while prioritizing directions that are most beneficial to the current task, improves plasticity under strict low-rank constraints without sacrificing stability.

5.3 Detailed Analysis

To directly quantify how much task-relevant gradient energy remains expressible in the effective update space, we define a gradient-alignment plasticity ratio. Let g_t denote the gradient of the current task loss with respect to the direction

Table 3: Comparison under different adapter ranks r on ImageNet-R with $T = 20$.

	r=1		r=5		r=10		r=16		r=30		r=50	
Method	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**	**Acc↑**	**AAA↑**
CL-LoRA	75.02	80.55	75.53	81.09	76.28	82.03	76.83	82.38	76.62	82.27	75.13	81.33
SD-LoRA	73.33	80.42	75.13	80.75	75.88	82.28	75.55	82.01	75.63	81.60	76.12	81.71
LoRA-DRS	72.97	78.02	74.68	79.36	75.60	80.10	75.85	80.86	75.25	80.49	72.33	80.00
BiLoRA	71.83	77.01	72.07	78.03	72.32	78.84	72.43	76.85	71.92	75.86	71.93	75.88
InfLoRA	69.52	75.81	70.58	76.71	70.32	77.11	70.68	77.61	69.10	76.56	65.68	73.83
VD-LoRA (Ours)	**75.92**	**81.01**	**76.12**	**81.52**	**77.53**	**83.90**	**77.01**	**82.74**	**76.88**	**82.56**	**76.34**	**81.96**

coefficients. Let P_{old} be the projector onto the protected subspace induced by previously consolidated knowledge, and let Π_{S_t} be the orthogonal projector onto the set of update-allowed directions selected for task t. We then measure the fraction of gradient energy that remains after applying both constraints as

$$\eta^2(t) \;=\; \frac{\left\|\Pi_{S_t}\left(I - P_{\text{old}}\right) g_t\right\|_2^2}{\|g_t\|_2^2}. \tag{14}$$

Here, $\eta^2(t) \in [0, 1]$ measures the fraction of current-task gradient energy that can still be expressed within the effective update space after accounting for both old-knowledge protection and the selected update directions.

Directional Plasticity and Effective Low-Rank Capacity. To examine whether the effective low-rank update space remains aligned with current-task gradients over long task sequences, we track the directional plasticity ratio $\eta^2(t)$ across tasks. Figure 4 reports $\eta^2(t)$ on ImageNet-R with $T = 10$ under different rank budgets. Compared with the orthogonality-based baselines InfLoRA and BiLoRA, whose $\eta^2(t)$ decreases rapidly as tasks accumulate, VD-LoRA maintains a substantially higher ratio throughout the sequence. This indicates that VD-LoRA preserves a more task-aligned effective update space under fixed rank constraints by avoiding strongly consolidated directions while retaining directions useful for the current task. To remove the influence of each method's direction selection, we further provide an oracle-controlled diagnostic in the supplementary material.

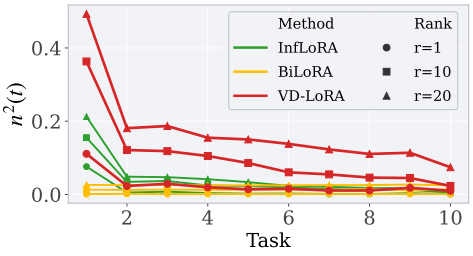


Fig. 4: Directional plasticity ratio across tasks and rank budgets on ImageNet-R with $T = 10$.

Rank Budget Study. We vary the adapter rank r under the ImageNet-R setting with 20 tasks, while keeping all other configurations unchanged. We report the average new-task accuracy, defined as the mean of the post-task accuracies measured immediately after finishing training on each task, evaluated on that

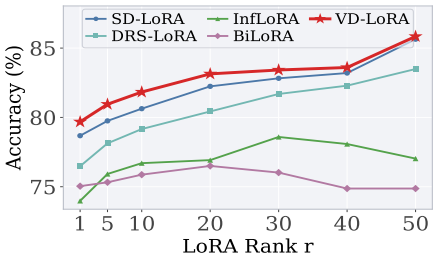


Fig. 5: Average new-task accuracy under different rank budgets.

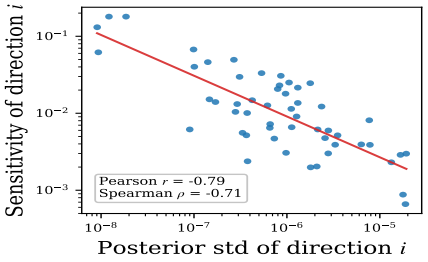


Fig. 6: Relationship between posterior standard deviation and old-task sensitivity across all activated directions.

task. As shown in Fig. 5, increasing r within a certain range improves the average new-task accuracy for all methods. Under small rank budgets, baselines based on hard subspace separation are more prone to under-adaptation, whereas VD-LoRA achieves higher average new-task accuracy across all r, indicating stronger plasticity when capacity is scarce. As reported in Table 3, VD-LoRA also remains competitive under different rank settings, with the most pronounced gains in the low-rank regime.

Full-Population Direction Sensitivity. To examine whether posterior uncertainty reflects direction sensitivity, we analyze all activated directions with learned posterior statistics on CIFAR-100 after task 8. We freeze the model and perturb only the posterior mean of direction i by a fixed symmetric offset $\pm\delta$, while keeping all other directions unchanged. We define its sensitivity as

$$s_i = \frac{1}{2}\left[L_{\text{old}}(\mu_i^+) - L_{\text{old}}(\mu)\right] + \frac{1}{2}\left[L_{\text{old}}(\mu_i^-) - L_{\text{old}}(\mu)\right]. \tag{15}$$

Here, $\mu_i^{\pm}$ denotes the posterior means obtained by perturbing only direction i by $\pm\delta$, while μ denotes the unperturbed posterior means. L_{old} denotes the average classification loss on the old-task validation set. Figure 6 shows a strong negative correlation between posterior standard deviation and direction sensitivity. Directions with smaller posterior uncertainty are more sensitive to the same perturbation and should therefore receive stronger protection.

Calibration and Uncertainty. We evaluate whether direction-level Bayesian tracking improves probability calibration, as measured by ECE and NLL. Table 4 reports ECE and NLL on ImageNet-R with $T = 10$ using the final model evaluated over all seen classes. VD-LoRA achieves lower ECE and NLL than deterministic LoRA baselines, indicating more reliable confidence estimates in addition to higher accuracy.

Ablation Study of Components. We conduct a progressive ablation study on ImageNet-R to examine the contributions of recursive KL regularization (RKL), uncertainty-guided direction selection (SEL), and stability regularization (STR). The first row excludes all three components. We then sequentially add RKL, SEL, and STR. When SEL is disabled, basis directions are selected randomly.

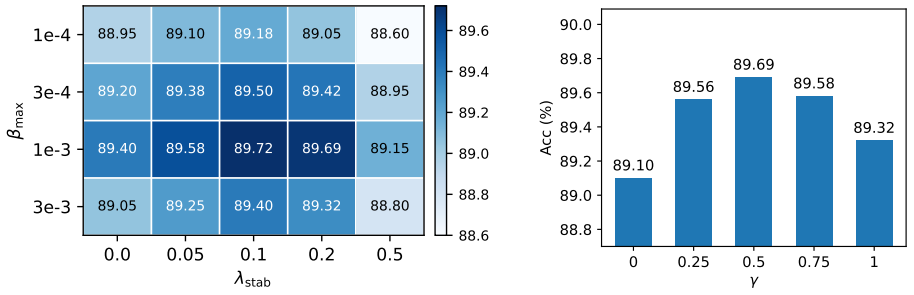
Table 4: Calibration on ImageNet-R.

Method	ECE↓	NLL↓
CL-LoRA	0.0204	0.8457
LoRA-DRS	0.0196	1.1672
BiLoRA	0.0238	1.1347
InfLoRA	0.0256	1.1271
VD-LoRA	**0.0120**	**0.7001**

Table 5: Ablation on three components.

RKL	SEL	STR	Tasks		
			5	**10**	**20**
✗	✗	✗	80.22	76.54	68.95
✓	✗	✗	83.30	80.20	73.80
✓	✓	✗	85.10	83.60	82.60
✓	✓	✓	**85.77**	**84.97**	**83.90**

As shown in Table 5, adding RKL consistently improves AAA by protecting previously consolidated knowledge. Incorporating SEL yields further gains by jointly considering current-task gradient demand and posterior uncertainty when allocating the limited rank capacity. Finally, STR provides additional improvements by encouraging robustness to uncertainty-scaled perturbations associated with the unselected directions. The consistent gains obtained when progressively introducing these components demonstrate their complementary contributions to VD-LoRA.



(a) Ablation study on β and λ. **(b)** Ablation study on γ.

Fig. 7: Hyperparameter sensitivity of VD-LoRA on CIFAR-100.

Hyperparameter Sensitivity. We evaluate the robustness of VD-LoRA to three key hyperparameters on CIFAR-100 under the standard class-incremental setting, while fixing all other configurations. We vary the uncertainty-weighting exponent in direction selection $\gamma \in \{0, 0.25, 0.5, 0.75, 1.0\}$, the maximum weight of the recursive KL term $\beta_{\max} \in \{10^{-4}, 3{\times}10^{-4}, 10^{-3}, 3{\times}10^{-3}\}$, and the stability regularization weight $\lambda_{\text{stab}} \in \{0, 0.05, 0.1, 0.2, 0.5\}$, and report results in Figure 7. Overall, the performance varies smoothly across these ranges, indicating that the gains do not rely on delicate tuning.

Additional controlled comparisons are provided in the supplementary material, including a BLoB-style variational-A design and a deterministic gradient-only LoRA variant without recursive Bayesian protection. These comparisons further isolate the effects of fixed direction identities, recursive posterior con-

solidation, and uncertainty-weighted direction selection. A detailed efficiency analysis, covering training time, inference latency, peak memory, and storage overhead, is also provided in the supplementary material.

6 Conclusion

We presented VD-LoRA, a Bayesian framework for rehearsal-free class-incremental learning with large pre-trained models under strict rank budgets. The key idea is to enable direction-wise adaptive consolidation in the low-rank update space, replacing uniform hard protection with uncertainty-calibrated soft protection. By maintaining recursive uncertainty estimates over LoRA directions, our approach selectively reuses low-risk directions while preserving critical ones, allowing the effective update subspace to remain aligned with task-relevant gradients throughout long task sequences. Extensive experiments on standard class-incremental benchmarks demonstrate that VD-LoRA consistently improves both accuracy and robustness over existing LoRA-based approaches, particularly in long-horizon settings where the update space becomes increasingly constrained. A promising direction for future work is to extend direction-wise Bayesian consolidation to online or task-free continual adaptation and combine it with adaptive rank allocation, enabling the model to determine both how many directions are needed and which historical directions can be safely reused under tighter computational and memory budgets.

Acknowledgements

Our work was supported by the National Natural Science Foundation of China under Grant Nos. 62476189 and 62406323; the Suzhou Key Core Technology "List Hanging Marshal" Project under Grant No. SYG2024149; the European Union's Horizon Europe research and innovation programme under Grant Agreement No. 101214398 (ELLIOT); and project PID2022-143257NB-I00, funded by MCIN/AEI/10.13039/501100011033 and FEDER.

References

1. Awais, M., Naseer, M., Khan, S., Anwer, R.M., Cholakkal, H., Shah, M., Yang, M.H., Khan, F.S.: Foundation models defining a new era in vision: a survey and outlook. IEEE Transactions on Pattern Analysis and Machine Intelligence **47**(4), 2245–2264 (2025)
2. Benjamin, A.S., Pehle, C., Daruwalla, K.: Continual learning with the neural tangent ensemble. In: Advances in Neural Information Processing Systems. vol. 37, pp. 58816–58840 (2024)
3. Bommasani, R., Hudson, D.A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M.S., Bohg, J., Bosselut, A., Brunskill, E., et al.: On the opportunities and risks of foundation models. arXiv preprint arXiv:2108.07258 (2021)

4. Carvalho Melo, L., Abate, A., Gal, Y.: Temporal-difference variational continual learning. In: Advances in Neural Information Processing Systems. vol. 38, pp. 142354–142396 (2025)
5. Castro, F.M., Marín-Jiménez, M.J., Guil, N., Schmid, C., Alahari, K.: End-to-end incremental learning. In: Proceedings of the European conference on computer vision. pp. 233–248 (2018)
6. Dai, H., Chauhan, J.: Continual generalized category discovery: Learning and forgetting from a bayesian perspective. arXiv preprint arXiv:2507.17382 (2025)
7. Du, K., Zhou, Y., Lyu, F., Li, Y., Xie, J., Shen, Y., Hu, F., Liu, G.: Rebalancing multi-label class-incremental learning. In: Proceedings of the AAAI conference on artificial intelligence. vol. 39, pp. 16372–16380 (2025)
8. Gao, Q., Zhao, C., Sun, Y., Xi, T., Zhang, G., Ghanem, B., Zhang, J.: A unified continual learning framework with general parameter-efficient tuning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 11483–11493 (2023)
9. He, J., Duan, Z., Zhu, F.: Cl-lora: Continual low-rank adaptation for rehearsal-free class-incremental learning. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 30534–30544 (2025)
10. Hendrycks, D., Basart, S., Mu, N., Kadavath, S., Wang, F., Dorundo, E., Desai, R., Zhu, T., Parajuli, S., Guo, M., et al.: The many faces of robustness: A critical analysis of out-of-distribution generalization. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 8340–8349 (2021)
11. Hendrycks, D., Zhao, K., Basart, S., Steinhardt, J., Song, D.: Natural adversarial examples. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 15262–15271 (2021)
12. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International conference on machine learning. pp. 2790–2799. PMLR (2019)
13. Hu, E.J., Shen, Y., Wallis, P., Allen-Zhu, Z., Li, Y., Wang, S., Wang, L., Chen, W.: LoRA: Low-rank adaptation of large language models. In: International Conference on Learning Representations (2022)
14. Kim, D., Han, B.: On the stability-plasticity dilemma of class-incremental learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 20196–20204 (2023)
15. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980 (2014)
16. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017)
17. Krizhevsky, A.: Learning multiple layers of features from tiny images. Tech. rep., University of Toronto (2009)
18. Lester, B., Al-Rfou, R., Constant, N.: 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)
19. Li, M., Lu, Y., Dai, Q., Huang, S., Ding, Y., Lu, H.: BECAME: Bayesian continual learning with adaptive model merging. In: Proceedings of the 42nd International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 267, pp. 35481–35501. PMLR (2025)
20. Li, Z., Hoiem, D.: Learning without forgetting. IEEE Transactions on Pattern Analysis and Machine Intelligence **40**(12), 2935–2947 (2018)

21. Liang, Y.S., Li, W.J.: Inflora: Interference-free low-rank adaptation for continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 23638–23647 (2024)
22. Liu, S., Xuan, L.K., Zhu, Y.: Last-lora: Adaptive knowledge reuse and latent subspace tracking for continual learning. In: Proceedings of the 2nd International Workshop on Continual Learning meets Multimodal Foundation Models: Fundamentals and Advances. pp. 1–7 (2025)
23. Liu, T., Lyu, F., Ni, C., Zhang, Z., Hu, F., Wang, L.: Daa: Amplifying unknown discrepancy for test-time discovery. In: Advances in Neural Information Processing Systems (2025)
24. Liu, X., Chang, X.: Lora subtraction for drift-resistant space in exemplar-free continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15308–15318 (2025)
25. Lyu, F., Cheng, G., Liu, D., Zhao, L., Zhang, Z., Hu, F., Wang, L.: Mitigating catastrophic forgetting in online continual learning with dual-margin contrastive replay. IEEE Transactions on Circuits and Systems for Video Technology **36**(4), 4433–4449 (2026)
26. Meo, C., Sycheva, K., Goyal, A., Dauwels, J.: Bayesian-lora: Lora based parameter efficient fine-tuning using optimal quantization levels and rank values trough differentiable bayesian gates. arXiv preprint arXiv:2406.13046 (2024)
27. Nguyen, C.V., Li, Y., Bui, T.D., Turner, R.E.: Variational continual learning. arXiv preprint arXiv:1710.10628 (2017)
28. Ni, C., Lyu, F., Tan, J., Hu, F., Yao, R., Zhou, T.: Maintaining consistent inter-class topology in continual test-time adaptation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 15319–15328 (2025)
29. Parisi, G.I., Kemker, R., Part, J.L., Kanan, C., Wermter, S.: Continual lifelong learning with neural networks: A review. Neural networks **113**, 54–71 (2019)
30. Shi, H., Xu, Z., Wang, H., Qin, W., Wang, W., Wang, Y., Wang, Z., Ebrahimi, S., Wang, H.: Continual learning of large language models: A comprehensive survey. ACM Computing Surveys (2025)
31. Smith, J.S., Karlinsky, L., Gutta, V., Cascante-Bonilla, P., Kim, D., Arbelle, A., Panda, R., Feris, R., Kira, Z.: 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. pp. 11909–11919 (2023)
32. Tan, J., Lyu, F., Liu, T., Hu, F., Feng, W.: Towards dynamic modality alignment in multimodal continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 39911–39921 (2026)
33. Wang, L., Zhang, X., Su, H., Zhu, J.: A comprehensive survey of continual learning: Theory, method and application. IEEE transactions on pattern analysis and machine intelligence **46**(8), 5362–5383 (2024)
34. Wang, X., Chen, T., Ge, Q., Xia, H., Bao, R., Zheng, R., Zhang, Q., Gui, T., Huang, X.: Orthogonal subspace learning for language model continual learning. In: Findings of the Association for Computational Linguistics: EMNLP 2023. pp. 10658–10671 (2023)
35. Wang, X., Zhuang, Z., Zhang, Y.: Plan: Proactive low-rank allocation for continual learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 2909–2918 (2025)
36. Wang, Y., Shi, H., Han, L., Metaxas, D., Wang, H.: Blob: Bayesian low-rank adaptation by backpropagation for large language models. In: Advances in neural information processing systems. vol. 37, pp. 67758–67794 (2024)

37. Wang, Z., Zhang, Z., Ebrahimi, S., Sun, R., Zhang, H., Lee, C.Y., Ren, X., Su, G., Perot, V., Dy, J., et al.: Dualprompt: Complementary prompting for rehearsal-free continual learning. In: European conference on computer vision. pp. 631–648. Springer (2022)
38. Wang, Z., Zhang, Z., Lee, C.Y., Zhang, H., Sun, R., Ren, X., Su, G., Perot, V., Dy, J., Pfister, T.: Learning to prompt for continual learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 139–149 (2022)
39. Wei, X., Li, G., Marculescu, R.: Online-lora: Task-free online continual learning via low rank adaptation. In: 2025 IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 6634–6645. IEEE (2025)
40. Wu, G., Gong, S., Li, P.: Striking a balance between stability and plasticity for class-incremental learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1124–1133 (2021)
41. Wu, Y., Piao, H., Huang, L.K., Wang, R., Li, W., Pfister, H., Meng, D., Ma, K., Wei, Y.: SD-LoRA: Scalable decoupled low-rank adaptation for class incremental learning. In: International Conference on Learning Representations (2025)
42. Zhai, X., Puigcerver, J., Kolesnikov, A., Ruyssen, P., Riquelme, C., Lucic, M., Djolonga, J., Pinto, A.S., Neumann, M., Dosovitskiy, A., et al.: A large-scale study of representation learning with the visual task adaptation benchmark. arXiv preprint arXiv:1910.04867 (2019)
43. Zhu, H., Zhang, Y., Dong, J., Koniusz, P.: Bilora: almost-orthogonal parameter spaces for continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 25613–25622 (2025)