BILATERAL SPECTRAL LOW-RANK ADAPTATION FOR CLASS-INCREMENTAL LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Merging sequential low-rank updates into a shared residual enables LoRA-based class-incremental learning (CIL) without retaining task-specific adapters, but introduces severe cross-task interference. Prior methods mitigate forgetting through subspace or gradient constraints, while recent spectral approaches mainly regulate task-local adaptation spectra, leaving the joint input–output spectral structure of the accumulated residual underexplored. In this paper, we conduct a spectral analysis of the accumulated residual to diagnose this interference. We reveal that (i) incoming updates increasingly overlap with historical left and right singular subspaces; and (ii) bilateral-overlap components are substantially more interference-prone than one-sided or complementary components. Based on these observations, we introduce **SF-LoRA**, a bilateral spectral filtering framework that achieves a better stability-plasticity trade-off. SF-LoRA constructs a compact spectral memory through energy-density-guided truncation, preserving informative modes without unnecessarily sacrificing plastic capacity. At each optimization step, SF-LoRA applies a closed-form proximal correction that selectively attenuates interference-prone bilateral components while granting new tasks full access to complementary orthogonal spaces. Across ImageNet-R, ImageNet-A, CIFAR-100, and CUB-200, SF-LoRA establishes a new state of the art, lifting last accuracy by up to 1.60 percentage points while reducing forgetting without retaining task-specific adapters.

1 INTRODUCTION

Class-Incremental Learning (CIL) requires models to sequentially master non-stationary distributions without eroding previously consolidated knowledge (Kirkpatrick et al., 2017; Lopez-Paz & Ranzato, 2017; Parisi et al., 2019). In the era of foundation models, rich pre-trained representations provide an expressive inductive bias (Radford et al., 2021; Steiner et al., 2022), shifting the core challenge of CIL from training representations from scratch to steering pre-existing features via parameter-efficient fine-tuning (PEFT) (Houlsby et al., 2019; Han et al., 2024; Gao et al., 2023). Among PEFT strategies, Low-Rank Adaptation (LoRA) (Hu et al., 2022) is widely adopted for its minimal compute footprint. However, current LoRA-based continual learners predominantly isolate each task within an independent adapter, incurring a memory footprint that scales linearly with the task sequence and often demanding test-time task identification (Zhou et al., 2024; Fukuda et al., 2025). A conceptually appealing alternative is sequential residual consolidation, accumulating incoming low-rank updates into a single, unified residual. Yet, this efficiency introduces a fundamental conflict: each low-rank update is optimized in isolation for the current task distribution, while naive accumulation projects novel gradients onto historical coordinate systems. This unconstrained superposition corrupts earlier task-specific representations, inducing severe catastrophic forgetting. Resolving this trade-off requires uncovering the precise geometric structure of cross-task interference in low-rank weight updates.

To mitigate cross-task interference under parameter-efficient constraints, existing literature relies on either modular parameter isolation (Wang et al., 2023; Li et al., 2025; He et al., 2025) or directional null-space projection (Liang & Li, 2024; Chen et al., 2026; Zheng et al., 2026). Although effective, these formulations implicitly treat parameter subspaces as uniformly sensitive, overlooking the spectral anisotropy of neural representations, where task-critical knowledge is concentrated along a sparse set of dominant singular directions (Wang et al., 2025; Sharma et al., 2024). A nascent

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.

line of work addresses this limitation by regularizing the spectral landscape of low-rank adaptation. Specifically, EBLoRA (Gu et al., 2026) balances singular components within task-specific updates to resolve internal spectral collapse; E^2-LoRA (Li et al., 2026a) exploits the low-rank structure of feature drift by concentrating adaptation energy into leading ranks; and SCLoRA (Wi & Park, 2026) bounds singular-value growth against the spectral envelope of the pre-trained weights. While these approaches demonstrate the diagnostic power of spectral analysis, these frameworks adopt an atomized view of adaptation: they calibrate individual updates in isolation, measure drift unilaterally in output space, or anchor adaptation strictly to the static backbone. Consequently, they remain blind to the compounding spectral geometry that emerges when successive updates are additively consolidated into a shared residual. In this continual accumulation regime, cross-task interference is fundamentally governed by a four-subspace interaction: a new update induces forgetting if its input-receptive (row) space is activated by historical feature distributions and its output-expressive (column) space simultaneously projects non-orthogonally onto the historical representation space. The failure to jointly regularize this bilateral spectral geometry remains the primary barrier to robust residual consolidation.

To address this challenge, this work investigates catastrophic forgetting through the lens of bilateral spectral interference with the accumulated residual. Rather than regulating individual updates in isolation, we examine how incoming parameter trajectories algebraically intersect with the joint left- and right-singular subspaces of the historical accumulation. Our empirical analysis uncovers a crucial disparity: while incoming updates naturally drift toward historical singular directions, bilateral overlap, concurrent alignment with both input (row) and output (column) historical subspaces, induces disproportionately severe normalized forgetting compared to benign, unilateral overlap. Motivated by this diagnostic insight, we introduce **SF-LoRA (Spectral-Filtered LoRA)**, a framework that constructs a compact historical spectral surrogate via energy-density-guided truncation and applies a closed-form proximal correction to incoming adapter updates. Instead of enforcing rigid null-space exclusions that undermine plasticity, SF-LoRA acts as a graduated bilateral filter: it aggressively suppresses catastrophic bilateral overlap, gently modulates unilateral alignment to permit beneficial feature reuse, and leaves orthogonal subspaces unconstrained. Across ImageNet-R, ImageNet-A, CIFAR-100, and CUB-200, SF-LoRA consistently improves accuracy and retention over strong PEFT-based CIL baselines, improving accuracy by up to 1.60 percentage points without retaining task-specific adapters.

The main contributions of this work are summarized as follows:

- **Bilateral Interference Discovery:** We identify bilateral spectral interference in sequentially merged LoRAs, showing that concurrent input–output overlap progressively accumulates during continual learning and exhibits substantially stronger normalized forgetting than one-sided or complementary overlap.
- **Principled Filtering (SF-LoRA):** We propose SF-LoRA, which combines energy-density spectral truncation with a closed-form proximal filter. It selectively dampens destructive bilateral interference while preserving unilateral feature reuse, avoiding the plasticity loss of rigid null-space projections.

2 Related Work

Continual Learning. Continual learning aims to acquire knowledge from sequential tasks while mitigating catastrophic forgetting. Classical approaches include replay, regularization or distillation, parameter isolation, and prompt-based adaptation. Replay methods preserve historical information through stored examples or model outputs, such as GEM (Lopez-Paz & Ranzato, 2017), iCaRL (Rebuffi et al., 2017), and DER (Buzzega et al., 2020). Regularization and distillation methods constrain changes to prior knowledge, e.g., EWC (Kirkpatrick et al., 2017), Learning without Forgetting (Li & Hoiem, 2018), and LUCIR (Hou et al., 2019). Parameter-isolation methods reserve task-specific capacity, e.g., PackNet (Mallya & Lazebnik, 2018). Prompt-based methods such as L2P (Wang et al., 2022b), DualPrompt (Wang et al., 2022a), CODA-Prompt (Smith et al., 2023), and VPT-NSP (Lu et al., 2024) adapt frozen backbones with lightweight or constrained prompts. These approaches preserve prior knowledge primarily at the data, output, parameter, or prompt level. Our work instead studies structured interference arising when sequential low-rank updates are merged into a shared residual.

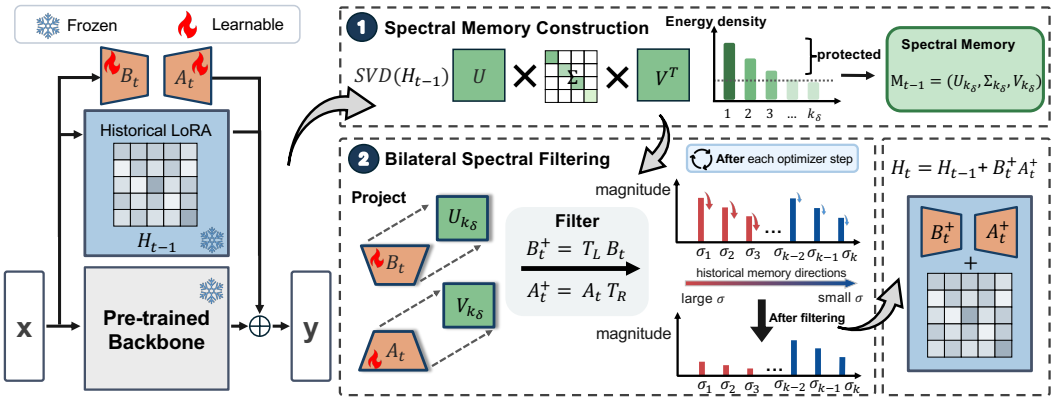


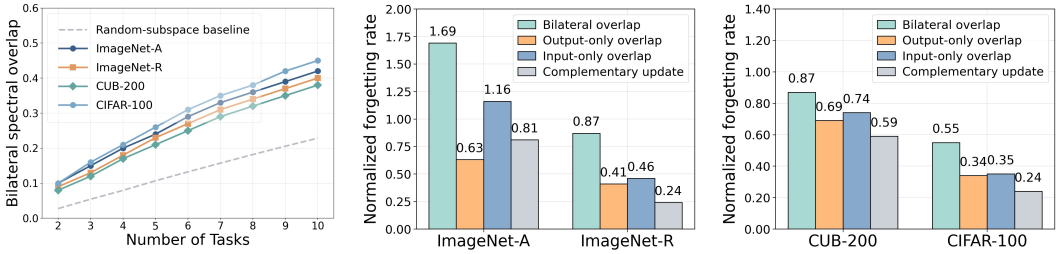
Figure 1: Overview of SF-LoRA. SF-LoRA first applies SVD to the accumulated LoRA residual and constructs a compact spectral memory $(U_{(k_\delta)}, \Sigma_{(k_\delta)}, V_{(k_\delta)})$ through energy-density-guided truncation. The retained left and right singular directions then define importance-weighted filters for the new LoRA factors B_t and A_t, respectively, yielding stronger suppression when an update overlaps with historical directions on both sides while leaving complementary directions largely unconstrained. The filtered update is finally merged into the backbone.

Parameter-Efficient and LoRA-Based Continual Learning. PEFT adapts pretrained models by optimizing a small number of additional parameters while keeping the backbone frozen (Liu et al., 2024; 2022). Among PEFT methods, LoRA is particularly attractive for continual learning because its low-rank updates can be efficiently trained and merged into the backbone (Houlsby et al., 2019; Li & Liang, 2021; Hu et al., 2022). Existing LoRA-based continual learning methods mainly mitigate interference through subspace constraints, gradient regulation, or adapter consolidation. O-LoRA (Wang et al., 2023) enforces approximately orthogonal task subspaces, InfLoRA (Liang & Li, 2024) constructs interference-aware adaptation spaces, and KeepLoRA (Luo et al., 2026) regulates new gradients through historical residual spaces. More recent methods, including BiLoRA (Zhu et al., 2025), SplitLoRA (Qiu et al., 2026), SD-LoRA (Wu et al., 2025b), and LoDA (He et al., 2026), further improve stability–plasticity through parameter-space isolation, gradient-space decomposition, decoupled adaptation, or shared/task-specific subspace modeling. Other approaches improve scalability through adapter organization or merging, such as LoRA-Mixer (Li et al., 2026b), GLAM (Testa et al., 2026), and Merge before Forget (Qiao & Mahdavi, 2025). In contrast, SF-LoRA treats the accumulated merged residual as a unified representation of historical adaptation and regulates incoming updates according to its spectral structure.

Spectral Structure of Low-Rank Adaptation. Singular value decomposition naturally represents a low-rank update as $\Delta W = \sum_i \sigma_i u_i v_i^\top$, where u_i, v_i, and σ_i characterize its output directions, input directions, and spectral strengths. Existing PEFT methods exploit spectral structure for rank allocation, initialization, or structured adaptation. AdaLoRA (Zhang et al., 2023), PiSSA (Meng et al., 2024), MiLoRA (Wang et al., 2025), and SVFT (Lingam et al., 2024) leverage singular directions or spectral information of pretrained weights or activations to construct or allocate low-rank adaptation. Recent spectral studies of LoRA and continual adaptation further investigate spectral properties of LoRA. EBLoRA (Gu et al., 2026) analyzes spectral imbalance within task-specific updates, E^2-LoRA (Li et al., 2026a) studies the low-rank energy structure of output feature drift, and SCLoRA regulates LoRA singular values according to the pretrained weight spectrum (Wi & Park, 2026). These methods mainly focus on the current adaptation, induced feature drift, or its relation to pretrained weights. In contrast, SF-LoRA models the accumulated residual $H_{t-1} = \sum_{\tau<t} \Delta W_\tau$ and explicitly captures its joint left–right spectral structure, enabling importance-weighted bilateral filtering and compact energy-density-guided spectral memory.

3 Methodology

In this section, we first formulate continual LoRA merging and analyze the bilateral spectral interference induced by accumulated residuals. We then introduce SF-LoRA, which combines importance-



(a) (b) (c)

Figure 2: Bilateral spectral overlap analysis. (a) Evolution of bilateral spectral overlap energy between current LoRA updates and accumulated historical residuals across incremental tasks. (b) and (c) Comparison of normalized forgetting rate (NFR) among four update regions across four datasets. The bilateral-overlap region exhibits substantially stronger interference than the other regions.

weighted bilateral spectral filtering with energy-density-guided spectral memory construction. Fig. 1 provides an overview of the proposed framework.

3.1 Preliminaries

Problem Formulation. In CIL, the model sequentially learns N tasks $\mathcal{S} = \{\mathcal{T}_1, \ldots, \mathcal{T}_N\}$, where each task dataset $\mathcal{D}_t = \{(\mathbf{x}_i^t, y_i^t)\}_{i=1}^{N_t}$ contains input-label pairs with a distinct label space $\mathcal{C}_t$ satisfying $\mathcal{C}_t \cap \mathcal{C}_{t'} = \emptyset$ for $t \neq t'$. During training on task $\mathcal{T}_t$, only $\mathcal{D}_t$ is accessible. The goal of CIL is to learn a unified model that performs well on all observed classes without accessing previous data or task identities during inference.

LoRA Merging for Class-Incremental Learning. LoRA parameterizes the update of a linear layer $\mathbf{W}_0 \in \mathbb{R}^{d_{\text{out}} \times d_{\text{in}}}$ as $\Delta\mathbf{W} = \mathbf{BA}$, where $\mathbf{B} \in \mathbb{R}^{d_{\text{out}} \times r}$, $\mathbf{A} \in \mathbb{R}^{r \times d_{\text{in}}}$, and $r \ll \min(d_{\text{in}}, d_{\text{out}})$. After tasks $1{:}t-1$,

$$\mathbf{H}_{t-1} = \sum_{\tau<t} \mathbf{B}_\tau \mathbf{A}_\tau, \qquad \mathbf{W}_{t-1} = \mathbf{W}_0 + \mathbf{H}_{t-1}.$$

The residual $\mathbf{H}_{t-1}$ need not be stored explicitly and can be recovered as $\mathbf{W}_{t-1} - \mathbf{W}_0$ when required. During task $\mathcal{T}_t$, $\mathbf{W}_{t-1}$ is frozen and a temporary LoRA module $(\mathbf{B}_t, \mathbf{A}_t)$ is optimized:

$$\mathbf{Y} = (\mathbf{W}_{t-1} + \mathbf{B}_t\mathbf{A}_t)\mathbf{X}. \tag{1}$$

After training, the update is merged into $\mathbf{W}_t = \mathbf{W}_{t-1} + \mathbf{B}_t\mathbf{A}_t$, and the temporary LoRA factors are discarded.

3.2 Motivation and Observations: Bilateral Spectral Interference

Spectral supports of historical adaptations. Before task t, all previously learned LoRA updates have been merged into the accumulated residual

$$\mathbf{H}_{t-1} = \sum_{\tau<t} \Delta\mathbf{W}_\tau = \mathbf{U}\Sigma\mathbf{V}^\top, \tag{2}$$

where $R = \text{rank}(\mathbf{H}_{t-1})$, $\mathbf{U} = [u_1, \ldots, u_R]$ and $\mathbf{V} = [v_1, \ldots, v_R]$ span the historical output- and input-side spectral subspaces, respectively, and $\Sigma = \text{diag}(\sigma_1, \ldots, \sigma_R)$ with $\sigma_1 \geq \cdots \geq \sigma_R > 0$ characterizes the spectral strength of these directions. For a linear transformation, the right singular subspace identifies input directions used by the accumulated adaptation, while the left singular subspace identifies its output directions.

Bilateral decomposition. Let $\mathbf{P}_U = \mathbf{U}\mathbf{U}^\top$ and $\mathbf{P}_V = \mathbf{V}\mathbf{V}^\top$ denote the projectors onto the historical output- and input-side spectral supports, respectively. Any current update $\Delta\mathbf{W}_t$ can be uniquely decomposed according to the historical spectral supports as

$$\Delta\mathbf{W}_t = \underbrace{\mathbf{P}_U\Delta\mathbf{W}_t\mathbf{P}_V}_{\text{bilateral overlap}} + \underbrace{\mathbf{P}_U\Delta\mathbf{W}_t(\mathbf{I}-\mathbf{P}_V)}_{\text{output-only overlap}} + \underbrace{(\mathbf{I}-\mathbf{P}_U)\Delta\mathbf{W}_t\mathbf{P}_V}_{\text{input-only overlap}} + \underbrace{(\mathbf{I}-\mathbf{P}_U)\Delta\mathbf{W}_t(\mathbf{I}-\mathbf{P}_V)}_{\text{complementary update}}. \tag{3}$$

We refer to the bilateral overlap component as bilateral spectral overlap, since both its input and output directions are defined by the historical singular subspaces. The bilateral spectral overlap component simultaneously lies in both the historical input-side and output-side spectral supports, whereas the two one-sided components interact with only one side. The complementary component lies outside both supports.

Observation 1: Incoming updates progressively compound historical bilateral overlap. As shown in Fig. 2a, the fraction of current-update energy that lies simultaneously in the historical output- and input-side spectral supports progressively increases as more tasks are learned. This quantity remains consistently above a rank-matched random-subspace baseline, suggesting that new updates increasingly reuse historical spectral directions rather than occupying arbitrary subspaces. Here, bilateral spectral overlap is measured as the ratio of the squared Frobenius norm of the bilateral component in Eq. 3 to that of the full current update. We report the average ratio across all LoRA-adapted layers.

Observation 2: Bilateral overlap disproportionately drives catastrophic forgetting. We next empirically test whether the four spectral regions differ in their interference with previously learned classes. For each region, we apply its original layer-wise components jointly across all adapted layers and normalize the resulting old-class accuracy drop by their joint Frobenius norm. As shown in Fig. 2b and 2c, bilateral overlap consistently exhibits a higher normalized forgetting rate than the one-sided and complementary regions across all four benchmarks. The formal definition of NFR is provided in Appendix A.3.1.

Together, these observations show that current updates increasingly align with historical spectral directions, while simultaneous input–output overlap exhibits a higher normalized forgetting rate than the other regions. This motivates the bilateral regulation introduced in Sec. 3.3, which controls alignment on each side, including one-sided overlap, and applies additional suppression to updates that overlap with both historical supports.

3.3 Importance-Weighted Bilateral Spectral Filtering

The observations above suggest that continual adaptation should regulate historical spectral directions from both input and output sides, without excluding them entirely from future updates. We therefore introduce an importance-weighted bilateral spectral filter. For clarity, we first derive the filter over the complete historical spectral support; in practice, the compact spectral memory constructed in Sec. 3.4 is used.

Historical spectral directions do not contribute equally to the accumulated adaptation. We assign each singular mode an importance weight

$$\alpha_i = \left(\frac{\sigma_i}{\sigma_1}\right)^p, \qquad \mathbf{\Lambda} = \mathrm{diag}(\alpha_1, \ldots, \alpha_R), \tag{4}$$

where p controls how strongly protection is concentrated on dominant historical modes. Since $\|\sigma_i u_i v_i^\top\|_F^2 = \sigma_i^2$, $p = 2$ makes α_i proportional to the spectral energy of mode i relative to the dominant mode. We use the same importance profile for the historical left and right singular subspaces.

At optimization step s of task t, let η_s denote the learning rate of the LoRA parameters, and let $\widehat{\Delta\mathbf{W}}_{t,s} = \widehat{\mathbf{B}}_{t,s}\widehat{\mathbf{A}}_{t,s}$ be the update obtained after the standard task-loss step. We set the effective filtering strength to $\gamma_s = 2\eta_s\lambda$, where $\lambda \geq 0$ is a fixed coefficient shared by both spectral sides. This couples the correction strength to the preceding gradient step, so that the filtering strength follows the optimization schedule. We then define

$$\begin{aligned}\Delta\mathbf{W}_{t,s}^{+} = \arg\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \widehat{\Delta\mathbf{W}}_{t,s}\|_F^2 + \frac{\gamma_s}{2}\|\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{X}\|_F^2 + \frac{\gamma_s}{2}\|\mathbf{X}\mathbf{V}\mathbf{\Lambda}^{1/2}\|_F^2 \\ + \frac{\gamma_s^2}{2}\|\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{X}\mathbf{V}\mathbf{\Lambda}^{1/2}\|_F^2.\end{aligned} \tag{5}$$

The first term keeps the corrected update close to the current-task solution. The next two terms penalize overlap with the historical output- and input-side subspaces, while the last term acts on their bilateral intersection:

$$\|\mathbf{\Lambda}^{1/2}\mathbf{U}^\top\mathbf{X}\mathbf{V}\mathbf{\Lambda}^{1/2}\|_F^2 = \sum_{i,j}\alpha_i\alpha_j|u_i^\top\mathbf{X}v_j|^2. \tag{6}$$

The coefficient γ_s^2 is not an additional hyperparameter. It follows from composing the two one-sided contractions,

$$(1+\gamma_s\alpha_i)(1+\gamma_s\alpha_j) = 1+\gamma_s\alpha_i+\gamma_s\alpha_j+\gamma_s^2\alpha_i\alpha_j, \tag{7}$$

which leads to the factorized bilateral filter below.

Proposition 3.1 (Closed-form bilateral filtering)**.** *For any $\gamma_s \geq 0$, the objective in Eq. 5 admits the following closed-form solution:*

$$\Delta\mathbf{W}_{t,s}^+ = \mathbf{T}_{L,s}\widehat{\Delta\mathbf{W}}_{t,s}\mathbf{T}_{R,s}, \tag{8}$$

where

$$\mathbf{T}_{L,s} = \left(\mathbf{I}+\gamma_s\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\right)^{-1}, \qquad \mathbf{T}_{R,s} = \left(\mathbf{I}+\gamma_s\mathbf{V}\mathbf{\Lambda}\mathbf{V}^\top\right)^{-1}. \tag{9}$$

For $\widehat{\Delta\mathbf{W}}_{t,s} = \widehat{\mathbf{B}}_{t,s}\widehat{\mathbf{A}}_{t,s}$, the closed-form solution can be implemented exactly on the LoRA factors without materializing the dense update:

$$\mathbf{B}_{t,s}^+ = \mathbf{T}_{L,s}\widehat{\mathbf{B}}_{t,s}, \qquad \mathbf{A}_{t,s}^+ = \widehat{\mathbf{A}}_{t,s}\mathbf{T}_{R,s}. \tag{10}$$

The proof is provided in Appendix A.1.

Using the eigendecomposition structure of the spectral memory, these operations can be efficiently implemented as direction-dependent shrinkage without explicitly constructing large matrix inverses. Specifically, the importance-aware filter assigns each historical spectral direction an adaptive shrinkage coefficient

$$\kappa_{i,s} = \frac{1}{1+\gamma_s\alpha_i}. \tag{11}$$

Since α_i measures the relative importance of historical mode i, larger-energy modes obtain smaller shrinkage coefficients and therefore receive stronger protection. Moreover, γ_s controls the overall filtering strength according to the learning-rate schedule. Since the columns of $\mathbf{U}$ and $\mathbf{V}$ are orthonormal, the filtering operators can be equivalently written as

$$\mathbf{T}_{L,s} = \mathbf{I}-\mathbf{U}\,\mathrm{diag}(1-\kappa_{1,s},\ldots,1-\kappa_{R,s})\mathbf{U}^\top, \mathbf{T}_{R,s} = \mathbf{I}-\mathbf{V}\,\mathrm{diag}(1-\kappa_{1,s},\ldots,1-\kappa_{R,s})\mathbf{V}^\top. \tag{12}$$

Thus, only the scalar coefficients $\{\kappa_{i,s}\}$ need to be updated with the learning-rate schedule. A larger α_i gives stronger attenuation of new updates along more important historical directions. Consequently, bilateral components are scaled by $\kappa_{i,s}\kappa_{j,s}$, one-sided components by a single shrinkage factor, and complementary components remain unchanged. A formal derivation is given in Appendix A.2.

3.4 Energy-Density-Guided Spectral Memory Construction

The importance weighting in Sec. 3.3 determines how strongly historical spectral directions are protected, but not how many directions should be retained. As tasks accumulate, H_{t-1} may develop a long tail of weak singular modes. Retaining these modes merely to satisfy a cumulative-energy target can unnecessarily enlarge the protected subspace and restrict new-task adaptation. We therefore use spectral energy density to construct a compact historical memory.

Let σ_i denote the i-th singular value of $\mathbf{H}_{t-1}$ in Eq. 2. We define its normalized spectral energy as

$$e_i = \frac{\sigma_i^2}{\sum_{j=1}^R \sigma_j^2}, \tag{13}$$

which measures the contribution of the i-th singular direction to the total historical adaptation energy. We then define the average energy of the leading k directions as

$$D(k) = \frac{1}{k}\sum_{i=1}^k e_i, \qquad k_\delta = \max\{k : D(k) \geq \delta\}, \quad 0 \leq \delta \leq D(1). \tag{14}$$

Table 1: Last and average accuracy (%) on ImageNet-R under 5-, 10-, and 20-task class-incremental splits. Results are reported as mean ± standard deviation over three seeds.

Method	Venue	ImageNet-R(5 Tasks)		ImageNet-R(10 Tasks)		ImageNet-R(20 Tasks)	
		$\mathcal{A}_{Last}$ (↑)	$\mathcal{A}_{Avg}$ (↑)	$\mathcal{A}_{Last}$ (↑)	$\mathcal{A}_{Avg}$ (↑)	$\mathcal{A}_{Last}$ (↑)	$\mathcal{A}_{Avg}$ (↑)
L2P	CVPR'22	70.83 ± 0.58	78.34 ± 0.47	72.34 ± 0.17	77.36 ± 0.64	69.64 ± 0.42	75.28 ± 0.57
DualPrompt	ECCV'22	73.05 ± 0.50	79.47 ± 0.40	69.10 ± 0.62	74.28 ± 0.66	66.01 ± 0.58	72.45 ± 0.37
CODAPrompt	CVPR'23	74.91 ± 0.33	79.25 ± 0.53	73.31 ± 0.50	78.47 ± 0.53	69.96 ± 0.50	75.34 ± 0.85
InfLoRA	CVPR'24	77.52 ± 0.37	82.01 ± 0.12	74.75 ± 0.64	80.67 ± 0.55	71.01 ± 0.45	77.28 ± 0.45
SD-LoRA	ICLR'25	79.15 ± 0.20	83.01 ± 0.42	77.34 ± 0.35	82.04 ± 0.24	75.26 ± 0.37	80.22 ± 0.72
Bi-LoRA	CVPR'25	78.32 ± 0.37	83.31 ± 0.52	77.95 ± 0.14	81.52 ± 0.26	72.41 ± 0.65	79.28 ± 0.76
LoRA-P&M	NeurIPS'25	81.47 ± 0.56	85.96 ± 0.52	79.95 ± 0.18	85.29 ± 0.93	76.37 ± 0.09	82.77 ± 0.31
CoSO	NeurIPS'25	82.10 ± 0.13	86.38 ± 0.07	81.10 ± 0.39	85.56 ± 0.13	78.19 ± 0.28	83.69 ± 0.12
VD-LoRA	ECCV'26	80.22 ± 0.52	85.77 ± 0.38	79.55 ± 0.39	84.97 ± 1.23	77.53 ± 0.35	83.90 ± 0.77
SplitLoRA	ICLR'26	81.92 ± 0.29	85.83 ± 0.55	81.00 ± 0.17	85.84 ± 0.62	78.82 ± 0.28	84.57 ± 0.44
LoDA	ICML'26	83.37 ± 0.14	87.46 ± 0.41	81.93 ± 0.20	86.90 ± 0.40	78.96 ± 0.23	84.94 ± 0.17
SF-LoRA	**-**	$\mathbf{83.66 \pm 0.19}$	$\mathbf{87.87 \pm 0.35}$	$\mathbf{82.48 \pm 0.17}$	$\mathbf{87.64 \pm 0.42}$	$\mathbf{79.45 \pm 0.22}$	$\mathbf{85.53 \pm 0.15}$

Table 2: Last and average accuracy (%) on three benchmark datasets (10 tasks) are reported. Results are reported as mean ± standard deviation over three seeds.

Method	Venue	ImageNet-A		CUB-200		CIFAR-100	
		$\mathcal{A}_{Last}$ (↑)	$\mathcal{A}_{Avg}$ (↑)	$\mathcal{A}_{Last}$ (↑)	$\mathcal{A}_{Avg}$ (↑)	$\mathcal{A}_{Last}$ (↑)	$\mathcal{A}_{Avg}$ (↑)
L2P	CVPR'22	44.04 ± 0.93	51.24 ± 2.26	67.02 ± 1.90	79.62 ± 1.60	84.06 ± 0.88	88.26 ± 1.34
DualPrompt	ECCV'22	53.19 ± 0.74	64.59 ± 0.08	68.48 ± 0.47	80.59 ± 1.50	86.93 ± 0.24	91.13 ± 0.32
CODAPrompt	CVPR'23	52.08 ± 0.12	63.92 ± 0.12	77.23 ± 1.12	81.90 ± 0.85	83.21 ± 3.39	87.71 ± 3.17
InfLoRA	CVPR'24	49.20 ± 1.12	60.92 ± 0.61	70.82 ± 0.23	81.39 ± 0.14	86.75 ± 0.35	91.72 ± 0.15
SD-LoRA	ICLR'25	55.96 ± 0.73	64.95 ± 1.63	77.48 ± 0.20	85.59 ± 0.44	88.01 ± 0.31	92.54 ± 0.18
Bi-LoRA	CVPR'25	52.22 ± 0.46	63.34 ± 0.66	-	-	87.46 ± 0.76	92.50 ± 0.62
LoRA-P&M	NeurIPS'25	56.57 ± 0.78	65.35 ± 1.81	78.29 ± 0.50	83.39 ± 0.61	88.45 ± 0.35	92.89 ± 1.13
CoSO	NeurIPS'25	57.27 ± 0.26	71.27 ± 0.21	79.27 ± 0.41	84.60 ± 0.92	85.21 ± 0.14	90.94 ± 0.11
VD-LoRA	ECCV'26	60.27 ± 0.88	70.18 ± 1.24	-	-	89.69 ± 0.31	92.72 ± 0.11
SplitLoRA	ICLR'26	62.28 ± 0.79	71.24 ± 1.79	81.28 ± 0.79	88.75 ± 0.69	90.33 ± 0.73	93.70 ± 0.32
LoDA	ICML'26	62.59 ± 0.64	70.87 ± 1.61	81.74 ± 0.78	89.35 ± 0.98	90.47 ± 0.06	93.46 ± 1.42
SF-LoRA	**-**	$\mathbf{64.19 \pm 0.65}$	$\mathbf{73.61 \pm 1.45}$	$\mathbf{82.40 \pm 0.17}$	$\mathbf{90.23 \pm 0.62}$	$\mathbf{90.67 \pm 0.29}$	$\mathbf{93.74 \pm 0.55}$

Since $e_1 \geq e_2 \geq \cdots \geq e_R$, $D(k)$ is non-increasing in k. The condition $\delta \leq D(1)$ guarantees that the feasible set is non-empty, so k_δ defines a well-defined maximal leading spectral prefix. Unlike cumulative-energy truncation, the density criterion adapts to the concentration of the historical spectrum. For a rapidly decaying spectrum, it may retain more directions than a cumulative-energy threshold, but the additional tail modes receive negligible protection because their spectral weights $\alpha_i = (\sigma_i/\sigma_1)^p$ are close to zero. For a more diffuse spectrum, cumulative-energy truncation may retain many moderately energetic directions to reach its target energy, whereas the density criterion stops once their average energy falls below δ, avoiding unnecessarily broad protection. The retained components form the compact spectral memory

$$\mathcal{M}_{t-1} = \left(\mathbf{U}_{(k_\delta)}, \mathbf{\Sigma}_{(k_\delta)}, \mathbf{V}_{(k_\delta)}\right). \tag{15}$$

A single threshold δ is used for all LoRA-adapted weight matrices within a dataset, while k_δ adapts to the spectrum of each matrix. Thus, energy-density truncation controls the breadth of historical protection, whereas the spectral weights in Sec. 3.3 control its direction-wise strength.

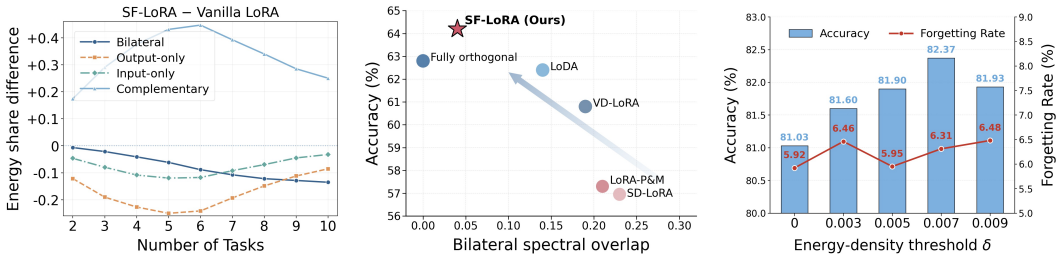
4 Experiments

4.1 Experimental Settings

Datasets and Evaluation Metrics. We conduct experiments on four widely used class-incremental learning benchmarks: ImageNet-R (Hendrycks et al., 2021a), ImageNet-A (Hendrycks et al., 2021b), CIFAR-100 (Krizhevsky et al., 2009), and CUB-200 (Wah et al., 2011). ImageNet-R and ImageNet-A contain 200 classes and provide challenging distribution-shifted image recognition settings, while CIFAR-100 consists of 100 object categories and CUB-200 contains 200 fine-grained bird categories. Following (Wu et al., 2025b), we use Last Accuracy ($\mathcal{A}_{\text{Last}}$) and Average Accuracy

Table 3: Ablation analysis of the proposed SF-LoRA.

Training Strategy	ImageNet-A (10 Tasks)			ImageNet-R (10 Tasks)		
	$\mathcal{A}_{Last}\uparrow$	$\mathcal{A}_{Avg}\uparrow$	FR $\downarrow$	$\mathcal{A}_{Last}\uparrow$	$\mathcal{A}_{Avg}\uparrow$	FR $\downarrow$
Vanilla LoRA	$58.91_{\pm 0.78}$	$70.63_{\pm 1.67}$	$18.84_{\pm 0.68}$	$72.35_{\pm 0.64}$	$81.09_{\pm 0.58}$	$15.83_{\pm 0.87}$
Right-only	$63.46_{\pm 0.61}$	$73.17_{\pm 1.39}$	$12.34_{\pm 0.42}$	$80.78_{\pm 0.28}$	$86.04_{\pm 0.39}$	$8.28_{\pm 0.91}$
Left-only	$62.21_{\pm 0.73}$	$72.37_{\pm 1.53}$	$14.92_{\pm 0.34}$	$78.62_{\pm 0.35}$	$85.03_{\pm 0.62}$	$12.25_{\pm 0.87}$
SF-LoRA	$\mathbf{64.19}_{\pm 0.65}$	$\mathbf{73.61}_{\pm 1.45}$	$\mathbf{9.97}_{\pm 0.41}$	$\mathbf{82.48}_{\pm 0.17}$	$\mathbf{87.64}_{\pm 0.42}$	$\mathbf{6.02}_{\pm 0.54}$



(a) (b) (c)

Figure 3: Mechanism and ablation analyses. (a) Difference in regional update-energy share between SF-LoRA and Vanilla LoRA on ImageNet-R (10 tasks); positive values indicate a larger energy share under SF-LoRA. (b) Bilateral spectral overlap versus final accuracy on ImageNet-A (10 tasks), including our hard bilateral-orthogonal ablation ("Fully orthogonal"). (c) Sensitivity to the energy-density threshold δ on ImageNet-R (10 tasks).

($\mathcal{A}_{\text{Avg}}$) as the primary evaluation metrics. For ablation and mechanism analyses, we additionally report Forgetting Rate (FR) and region-wise energy-share differences between SF-LoRA and Vanilla LoRA. Formal definitions are provided in Appendix A.3.1.

Baselines and Implementation Details. We compare SF-LoRA with representative state-of-the-art PEFT-based methods: L2P (Wang et al., 2022b), DualPrompt (Wang et al., 2022a), CODA-Prompt (Smith et al., 2023), InfLoRA (Liang & Li, 2024), SD-LoRA (Wu et al., 2025b), Bi-LoRA (Zhu et al., 2025), LoRA-P&M (Qiu et al., 2025), CoSO (Cheng et al., 2025), VD-LoRA (Ding et al., 2026), SplitLoRA (Qiu et al., 2026), and LoDA (He et al., 2026). To align our evaluation with prior PEFT-based CIL studies (Liang & Li, 2024; Wu et al., 2025b; He et al., 2026), we adopt the pretrained ViT setting to facilitate comparison with existing methods. Specifically, following (He et al., 2026), we use ViT-B/16 (Dosovitskiy et al., 2021) pretrained on ImageNet-21K (Russakovsky et al., 2015) with rank-32 LoRA modules inserted into all attention blocks. Results are reported as mean $\pm$ standard deviation over three random seeds. We use the same spectral hyperparameters, $\delta = 0.007$ and $\lambda = 150$, across all datasets. More setup details are provided in Appendix A.3.3.

4.2 Experimental Results

Comparison with State-of-the-Art Methods. Tables 1 and 2 summarize the main results across different datasets and task-sequence lengths. SF-LoRA achieves the best mean $\mathcal{A}_{Last}$ and $\mathcal{A}_{Avg}$ in all reported settings. The improvements are particularly pronounced on ImageNet-A, where SF-LoRA reaches 64.19 $\mathcal{A}_{Last}$ and 73.61 $\mathcal{A}_{Avg}$, outperforming the strongest competing results by 1.60 and 2.34 percentage points, respectively. On ImageNet-R, the advantage remains consistent across 5-, 10-, and 20-task splits, indicating that the performance gain persists as the incremental sequence becomes longer. SF-LoRA also achieves the best mean performance on CUB-200 and CIFAR-100. Overall, these results demonstrate that SF-LoRA provides consistently strong performance across datasets and task granularities while retaining a single merged model without task-specific adapters.

Computation and Storage. Table 4 reports the training-time computation and storage overhead on 20-task ImageNet-R. The reported GFLOPs include the SVD for spectral-memory construction as well as the subsequent spectral filtering. SF-LoRA requires 17.83 training GFLOPs and introduces no additional task-level forward pass. Its 14.1 MB extra storage corresponds to the retained spectral

memory $(U_{(k_\delta)}, \Sigma_{(k_\delta)}, V_{(k_\delta)})$. The accumulated residual $\mathbf{H}_{t-1}$ is not stored separately; when constructing the spectral memory, it is recovered on demand as $\mathbf{H}_{t-1} = \mathbf{W}_{t-1} - \mathbf{W}_0$ from the merged weights and the pretrained checkpoint, and discarded after decomposition. Thus, SF-LoRA avoids retaining either task-specific LoRA adapters or an additional dense historical residual.

4.3 Ablation and Mechanism Analysis

Effect and mechanism of bilateral spectral filtering. Table 3 shows that both one-sided variants improve over Vanilla LoRA, while the full bilateral filter yields the best accuracy and forgetting performance. Fig. 3a provides a complementary view of how SF-LoRA changes the spectral composition of the update. As tasks accumulate, the bilateral energy-share difference becomes increasingly negative, consistent with the stronger two-sided contraction $\kappa_i\kappa_j$. The one-sided differences decrease initially but later move back toward zero, reflecting weaker single-sided regulation and continued reuse of historical directions. Accordingly, the complementary difference peaks in the middle of the task sequence and then declines as the one-sided gaps narrow.

Soft regulation versus hard exclusion. Fig. 3b compares bilateral spectral overlap and final accuracy across several LoRA-based adaptation strategies. Methods with larger overlap, including VD-LoRA, LoRA-P&M, and SD-LoRA, generally exhibit lower accuracy, while LoDA and SF-LoRA occupy lower-overlap, higher-accuracy regions. However, completely eliminating overlap is not optimal: the *Fully orthogonal* variant projects updates onto the complements of both historical spectral subspaces and achieves zero overlap, yet performs worse than SF-LoRA. SF-LoRA instead retains a small amount of spectral reuse while achieving the highest accuracy, suggesting that historical directions should be selectively attenuated rather than entirely excluded. This supports the role of importance-weighted soft filtering in balancing historical protection and new-task plasticity.

Table 4: Training computation and storage overhead on ImageNet-R (20 tasks).

Method	Extra Fwd	Train GFLOPs	Extra Storage
InfLoRA	2/task	17.67	11.3 MB
SD-LoRA	None/0	20.51	14.1 MB
VPT-NSP	1/task	17.72	28.4 MB
LoDA	3/task	18.06	27.0 MB
SF-LoRA	None/0	17.83	14.1 MB

Table 5: Effect of the spectral weighting exponent p on ImageNet-R with 5 and 10 tasks.

Method	5 tasks		10 tasks	
	$\mathcal{A}_{Last}\uparrow$	FR $\downarrow$	$\mathcal{A}_{Last}\uparrow$	FR $\downarrow$
p=0	82.87	6.59	81.48	7.24
p=1	82.92	6.82	82.00	6.72
p=2	**83.66**	**6.53**	**82.48**	**6.02**

Effect of spectral truncation and weighting. Fig. 3c varies the energy-density threshold δ. At $\delta = 0$, no directions are truncated; spectral weighting and bilateral filtering remain active. Moderate truncation improves accuracy with small changes in forgetting, suggesting that excluding weak modes can ease constraints on new-task learning. Table 5 examines the weighting exponent p. The energy-based choice $p = 2$ yields the highest final accuracy and lowest FR among the tested values on both 5- and 10-task ImageNet-R. We use $p = 2$ in the main experiments.

5 Conclusion

In this work, we study forgetting in continually merged LoRA updates through bilateral spectral interference. We identify the accumulated residual as a structured historical memory whose input- and output-side spectral geometry jointly determines cross-task interference. Accordingly, we propose SF-LoRA, which combines importance-weighted bilateral spectral filtering with energy-density-guided spectral memory construction. SF-LoRA selectively attenuates interference-prone overlap with historical spectral directions while preserving complementary directions for new-task adaptation, providing a way to balance stability and plasticity without retaining task-specific adapters.

Our current evaluation focuses on LoRA-adapted vision transformers for image classification. Extending SF-LoRA to broader model architectures, adaptation modules, and application domains would further assess the generality of bilateral spectral filtering. In addition, spectral-memory construction requires an SVD of the accumulated residual once per task; although its overhead is modest in our current setting, the cost may become more pronounced for substantially larger adapted matrices. Developing more scalable or approximate spectral-memory construction strategies therefore represents a promising direction for future work.

AI USE STATEMENT

Generative AI tools were used during the preparation of this manuscript to assist with writing refinement, grammar checking, improving the readability of technical explanations, and providing feedback on the organization of figures and experimental presentations. AI tools were also used as a discussion aid for exploring alternative formulations and improving the clarity of the manuscript. No generative AI tools were used to generate the core scientific contributions of this work, including the proposed SF-LoRA framework, theoretical motivation, mathematical derivations, experimental methodology, or implementation of the proposed approach. All experimental results, analyses, and conclusions were independently verified by the authors. The authors remain fully responsible for all contents of this submission, including the correctness of the methodology, claims, experimental results, and final manuscript.

REPRODUCIBILITY STATEMENT

We provide sufficient implementation and evaluation details to facilitate reproduction of our results. The formulation and closed-form optimization of SF-LoRA are described in Sec. 3, with theoretical derivations provided in Appendices A.1–A.2. The complete training procedure is summarized in Algorithm 1. We report the class-incremental learning protocol, evaluation metrics, backbone and LoRA configuration, optimizer and training schedule, random-seed protocol, and dataset-specific hyperparameters in Appendix A.3. We further report the computational and storage overhead of SF-LoRA and provide supplementary task-wise and hyperparameter analyses.

REFERENCES

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and SIMONE CALDERARA. Dark experience for general continual learning: a strong, simple baseline. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 15920–15930. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/b704ea2c39778f07c617f6b7ce480e9e-Paper.pdf.

Cheng Chen, Pengpeng Zeng, Yuyu Guo, Lianli Gao, Heng Tao Shen, and Jingkuan Song. JANUS-LORA: A balanced low-rank adaptation for continual learning. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=XJ5r19iN9G.

Quan Cheng, Yuanyu Wan, Lingyu Wu, Chenping Hou, and Lijun Zhang. Continuous subspace optimization for continual learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=iLYV4iIC0c.

Luqiong Ding, Jiayao Tan, Chenggong Ni, Fuyuan Hu, and Fan Lyu. Vd-lora: Adaptive reuse of low-rank directions for continual learning. In *European Conference on Computer Vision*, pp. 1–18. Springer, 2026.

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*, 2021. URL https://openreview.net/forum?id=YicbFdNTTy.

Takuma Fukuda, Hiroshi Kera, and Kazuhiko Kawamoto. Adapter merging with centroid prototype mapping for scalable class-incremental learning. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4884–4893, 2025. doi: 10.1109/CVPR52734.2025.00460.

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 *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 11449–11459, 2023. doi: 10.1109/ICCV51070.2023.01055.

Hao Gu, Mao-Lin Luo, Zi-Hao Zhou, Han-Chen Zhang, Min-Ling Zhang, and Tong Wei. Spectral imbalance causes forgetting in low-rank continual adaptation. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=kqE6GjpQTn.

Zeyu Han, Chao Gao, Jinyang Liu, Jeff Zhang, and Sai Qian Zhang. Parameter-efficient fine-tuning for large models: A comprehensive survey. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=lIsCS8b6zj.

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

Lingfeng He, De Cheng, Huaijie Wang, Xi Yang, Nannan Wang, and Xinbo Gao. Task-driven subspace decomposition for knowledge sharing and isolation in loRA-based continual learning. In *Forty-third International Conference on Machine Learning*, 2026. URL https://openreview.net/forum?id=UJJfkcJ7K9.

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 *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8320–8329, 2021a. doi: 10.1109/ICCV48922.2021.00823.

Dan Hendrycks, Kevin Zhao, Steven Basart, Jacob Steinhardt, and Dawn Song. Natural adversarial examples. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15257–15266, 2021b. doi: 10.1109/CVPR46437.2021.01501.

Saihui Hou, Xinyu Pan, Chen Change Loy, Zilei Wang, and Dahua Lin. Learning a unified classifier incrementally via rebalancing. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 831–839, 2019. doi: 10.1109/CVPR.2019.00092.

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 *International conference on machine learning*, pp. 2790–2799. PMLR, 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*, 2022. URL https://openreview.net/forum?id=nZeVKeeFYf9.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, et al. Overcoming catastrophic forgetting in neural networks. *Proceedings of the national academy of sciences*, 114(13):3521–3526, 2017.

Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

Jiashuo Li, Shaokun Wang, Bo Qian, Yuhang He, Xing Wei, Qiang Wang, and Yihong Gong. Dynamic integration of task-specific adapters for class incremental learning. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30545–50555, 2025. doi: 10.1109/CVPR52734.2025.02844.

Longhua Li, Lei Qi, Qi Tian, and Xin Geng. Energy-structured low-rank adaptation for continual learning. In *Forty-third International Conference on Machine Learning*, 2026a. URL https://openreview.net/forum?id=4OP391X7Qm.

Wenbing Li, Zikai Song, Hang Zhou, Junqing Yu, Yunyao Zhang, and Wei Yang. LoRA-mixer: Coordinate modular loRA experts through serial attention routing. In *The Fourteenth International Conference on Learning Representations*, 2026b. URL https://openreview.net/forum?id=GMP1S4R6Ke.

Xiang Lisa Li and Percy Liang. Prefix-tuning: Optimizing continuous prompts for generation, 2021. URL https://arxiv.org/abs/2101.00190.

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 *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23638–23647, 2024. doi: 10.1109/CVPR52733.2024.02231.

Vijay Lingam, Atula Tejaswi Neerkaje, Aditya Vavre, Aneesh Shetty, Gautham Krishna Gudur, Joydeep Ghosh, Eunsol Choi, Alex Dimakis, Aleksandar Bojchevski, and sujay sanghavi. SVFT: Parameter-efficient fine-tuning with singular vectors. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=uS0PwIBzC0.

Haokun Liu, Derek Tam, Muqeeth Mohammed, Jay Mohta, Tenghao Huang, Mohit Bansal, and Colin Raffel. Few-shot parameter-efficient fine-tuning is better and cheaper than in-context learning. In Alice H. Oh, Alekh Agarwal, Danielle Belgrave, and Kyunghyun Cho (eds.), *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=rBCvMG-JsPd.

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 Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp (eds.), *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pp. 32100–32121. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/liu24bn.html.

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

Yue Lu, Shizhou Zhang, De Cheng, Yinghui Xing, Nannan Wang, PENG WANG, and Yanning Zhang. Visual prompt tuning in null space for continual learning. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=8pRemr5kEi.

Mao-Lin Luo, Zi-Hao Zhou, Yi-Lin Zhang, Yuanyu Wan, Min-Ling Zhang, and Tong Wei. KeeploRA: Continual learning with residual gradient adaptation. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=T3Vc5fkTzV.

Arun Mallya and Svetlana Lazebnik. Packnet: Adding multiple tasks to a single network by iterative pruning. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 7765–7773, 2018. doi: 10.1109/CVPR.2018.00810.

Fanxu Meng, Zhaohui Wang, and Muhan Zhang. PiSSA: Principal singular values and singular vectors adaptation of large language models. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=6ZBHIEtdP4.

German I Parisi, Ronald Kemker, Jose L Part, Christopher Kanan, and Stefan Wermter. Continual lifelong learning with neural networks: A review. *Neural networks*, 113:54–71, 2019.

Fuli Qiao and Mehrdad Mahdavi. Merge before forget: A single LoRA continual learning via continual merging, 2025. URL https://arxiv.org/abs/2512.23017.

Haomiao Qiu, Miao Zhang, Ziyue Qiao, and Liqiang Nie. Train with perturbation, infer after merging: A two-stage framework for continual learning. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=lqm1qJ43Sw.

Haomiao Qiu, Miao Zhang, Ziyue Qiao, Weili Guan, Min Zhang, and Liqiang Nie. SplitloRA: Balancing stability and plasticity in continual learning through gradient space splitting. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=Zm1hjXxRQV.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5533–5542, 2017. doi: 10.1109/CVPR.2017.587.

Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, et al. Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115(3):211–252, 2015.

Pratyusha Sharma, Jordan T. Ash, and Dipendra Misra. The truth is in there: Improving reasoning in language models with layer-selective rank reduction. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=ozX92bu8VA.

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 *2023 IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, pp. 11909–11919. IEEE, 2023.

Andreas Peter Steiner, Alexander Kolesnikov, Xiaohua Zhai, Ross Wightman, Jakob Uszkoreit, and Lucas Beyer. How to train your vit? data, augmentation, and regularization in vision transformers. *Transactions on Machine Learning Research*, 2022. ISSN 2835-8856. URL https://openreview.net/forum?id=4nPswr1KcP.

Irene Testa, Luigi Quarantiello, Eric Nuertey Coleman, Samrat Mukherjee, Julio Hurtado, and Vincenzo Lomonaco. GLAM: Efficient continual learning at scale via grouped LoRA adapter merging, 2026. URL https://arxiv.org/abs/2509.13211.

Catherine Wah, Steve Branson, Peter Welinder, Pietro Perona, and Serge Belongie. The caltech-ucsd birds-200-2011 dataset. 2011.

Hanqing Wang, Yixia Li, Shuo Wang, Guanhua Chen, and Yun Chen. MiLoRA: Harnessing minor singular components for parameter-efficient LLM finetuning. In Luis Chiruzzo, Alan Ritter, and Lu Wang (eds.), *Proceedings of the 2025 Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*, pp. 4823–4836, Albuquerque, New Mexico, April 2025. Association for Computational Linguistics. ISBN 979-8-89176-189-6. doi: 10.18653/v1/2025.naacl-long.248. URL https://aclanthology.org/2025.naacl-long.248/.

Xiao Wang, Tianze Chen, Qiming Ge, Han Xia, Rong Bao, Rui Zheng, Qi Zhang, Tao Gui, and Xuan-Jing Huang. Orthogonal subspace learning for language model continual learning. In *Findings of the Association for Computational Linguistics: EMNLP 2023*, pp. 10658–10671, 2023.

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

Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 139–149, 2022b. doi: 10.1109/CVPR52688.2022.00024.

Hyowon Wi and Noseong Park. Can spectral-clipping enable better learning while forgetting less for low-rank adaptation? In Maria Liakata, Viviane P. Moreira, Jiajun Zhang, and David Jurgens (eds.), *Proceedings of the 64th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pp. 25708–25734, San Diego, California, United States, July 2026. Association for Computational Linguistics. ISBN 979-8-89176-390-6. doi: 10.18653/v1/2026.acl-long.1179. URL https://aclanthology.org/2026.acl-long.1179/.

Fangwen Wu, Lechao Cheng, Shengeng Tang, Xiaofeng Zhu, Chaowei Fang, Dingwen Zhang, and Meng Wang. Navigating semantic drift in task-agnostic class-incremental learning. In *Forty-second International Conference on Machine Learning*, 2025a. URL https://openreview.net/forum?id=M6L7Eaw9BW.

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 *The Thirteenth International Conference on Learning Representations*, 2025b. URL https://openreview.net/forum?id=5U1rlpX68A.

Qingru Zhang, Minshuo Chen, Alexander Bukharin, Pengcheng He, Yu Cheng, Weizhu Chen, and Tuo Zhao. Adaptive budget allocation for parameter-efficient fine-tuning. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=lq62uWRJjiY.

Yaoyue Zheng, Yin Zhang, Joost van de Weijer, Gido M van de Ven, Shaoyi Du, Xuetao Zhang, and Zhiqiang Tian. Revisiting weight regularization for low-rank continual learning. In *The Fourteenth International Conference on Learning Representations*, 2026. URL https://openreview.net/forum?id=pZj2DhfaVD.

Da-Wei Zhou, Hai-Long Sun, Han-Jia Ye, and De-Chuan Zhan. Expandable subspace ensemble for pre-trained model-based class-incremental learning. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23554–23564, 2024. doi: 10.1109/CVPR52733.2024.02223.

Hao Zhu, Yifei Zhang, Junhao Dong, and Piotr Koniusz. Bilora: Almost-orthogonal parameter spaces for continual learning. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 25613–25622, 2025. doi: 10.1109/CVPR52734.2025.02385.

A Appendix

A.1 Proof of Proposition 3.1

Proof. For simplicity, we omit the task and step indices and write $\gamma_s = \gamma \geq 0$ and $\widehat{\Delta W}_{t,s} = \widehat{\Delta W}$. Let

$$S_U = U\Lambda U^\top, \qquad S_V = V\Lambda V^\top.$$

Since $\Lambda \succeq 0$, both S_U and S_V are positive semidefinite. The objective in Eq. 5 can be written as

$$J(X) = \frac{1}{2}\|X - \widehat{\Delta W}\|_F^2 + \frac{\gamma}{2}\|S_U^{1/2}X\|_F^2 + \frac{\gamma}{2}\|XS_V^{1/2}\|_F^2 + \frac{\gamma^2}{2}\|S_U^{1/2}XS_V^{1/2}\|_F^2.$$

The first term is strictly convex, and the remaining terms are convex. Thus, J is strictly convex, and its unique minimizer satisfies

$$\nabla_X J = X - \widehat{\Delta W} + \gamma S_U X + \gamma X S_V + \gamma^2 S_U X S_V = 0.$$

Rearranging and factorizing gives

$$\widehat{\Delta W} = (I + \gamma S_U)X(I + \gamma S_V).$$

Since $\gamma \geq 0$ and $S_U, S_V \succeq 0$, both $I + \gamma S_U$ and $I + \gamma S_V$ are positive definite and therefore invertible. Hence,

$$X = (I + \gamma S_U)^{-1}\widehat{\Delta W}(I + \gamma S_V)^{-1} = T_L\widehat{\Delta W}T_R,$$

where

$$T_L = (I + \gamma U\Lambda U^\top)^{-1}, \qquad T_R = (I + \gamma V\Lambda V^\top)^{-1}.$$

This is exactly the solution in Eq. 8.

Factorized implementation. For $\widehat{\Delta W} = \widehat{B}\widehat{A}$, where the factors have inner dimension r, the solution satisfies

$$\Delta W^+ = T_L\widehat{B}\widehat{A}T_R = (T_L\widehat{B})(\widehat{A}T_R).$$

Thus, the filter is implemented exactly on the LoRA factors without materializing the dense update. Since T_L and T_R are invertible for finite $\gamma \geq 0$,

$$\operatorname{rank}(\Delta W^+) = \operatorname{rank}(\widehat{\Delta W}) \leq r.$$

Therefore, filtering preserves the rank of the current LoRA update. □

A.2 Differential Spectral Suppression

Proof. Consider a historical left singular vector u_i.

Because

$$U\Lambda U^\top u_i = \alpha_i u_i,$$

we have

$$T_L u_i = (I + \gamma U\Lambda U^\top)^{-1}u_i = \frac{1}{1 + \gamma\alpha_i}u_i.$$

Define

$$\kappa_i = \frac{1}{1 + \gamma\alpha_i}.$$

Similarly,

$$v_j^\top T_R = \kappa_j v_j^\top.$$

Therefore, for a bilateral component,

$$T_L(u_i v_j^\top) T_R = \kappa_i \kappa_j u_i v_j^\top.$$

For a one-sided component,

$$T_L(u_i b^\top) T_R = \kappa_i u_i b^\top,$$

where b is orthogonal to the historical right subspace. Similarly,

$$T_L(a v_j^\top) T_R = \kappa_j a v_j^\top.$$

For a complementary component,

$$T_L(a b^\top) T_R = a b^\top.$$

Since

$$0 < \kappa_i \leq 1,$$

we have

$$\kappa_i \kappa_j \leq \min(\kappa_i, \kappa_j).$$

Hence bilateral overlap receives the strongest suppression, one-sided overlap receives weaker suppression, and complementary directions remain unchanged. □

A.3 Experiment Details

A.3.1 Evaluation Metrics

We denote by A_{ij} the classification accuracy on task i after completing training on task j, where $1 \leq i \leq j \leq T$. We report Last Accuracy ($\mathcal{A}_{Last}$), Average Accuracy ($\mathcal{A}_{Avg}$), and Forgetting Rate (FR).

Last Accuracy. Last Accuracy measures the overall classification performance after the model has completed the entire task sequence. It is defined as the mean accuracy over all learned tasks at the final incremental stage:

$$\mathcal{A}_{Last} = \frac{1}{T} \sum_{i=1}^{T} \mathcal{A}_{iT}.$$

A higher $\mathcal{A}_{\text{Last}}$ indicates that the final model retains stronger performance over the complete set of learned classes. This metric therefore summarizes the final stability–plasticity outcome after all incremental updates have been incorporated.

Average Accuracy. Average Accuracy evaluates performance throughout the full continual learning process rather than only at the final stage. At each incremental stage j, we first average the accuracies over all tasks observed up to that point, and then average these values across all stages:

$$\mathcal{A}_{Avg} = \frac{1}{T} \sum_{j=1}^{T} \frac{1}{j} \sum_{i=1}^{j} \mathcal{A}_{ij}.$$

Compared with $\mathcal{A}_{\text{Last}}$, $\mathcal{A}_{\text{Avg}}$ captures the model's performance trajectory during continual learning and reflects whether high accuracy is maintained consistently as new tasks are introduced.

Forgetting Rate. Forgetting Rate quantifies the loss of previously acquired knowledge caused by subsequent task learning. For each task i, we compare its best accuracy achieved after it is learned with its accuracy after the final task:

$$\mathrm{FR} = \frac{1}{T-1} \sum_{i=1}^{T-1} \left(\max_{j \in \{i,\dots,T\}} A_{ij} - A_{iT} \right).$$

The final task is excluded because no subsequent task is learned after it. A lower FR indicates better retention of previously learned knowledge, whereas a larger value corresponds to stronger catastrophic forgetting. Together with the two accuracy metrics, FR provides a complementary measure of the stability of the continual learner.

Normalized Forgetting Rate (NFR). To compare the interference induced by different spectral regions while accounting for their update magnitudes, we normalize the resulting forgetting by the joint Frobenius norm of the corresponding regional components across all LoRA-adapted layers.

For region r, we first apply its original layer-wise components $\{\Delta W_t^{(\ell,r)}\}_{\ell=1}^{L}$ simultaneously to the model. Their joint magnitude is defined as

$$N_t^{(r)} = \left(\sum_{\ell=1}^{L} \left\| \Delta W_t^{(\ell,r)} \right\|_F^2 \right)^{1/2}.$$

Let A_{t-1}^{old} denote the accuracy on previously learned classes before applying the regional update, and $\widetilde{A}_{t,r}^{\mathrm{old}}$ the accuracy after applying it. The normalized forgetting rate is

$$F_t^{(r)} = \frac{A_{t-1}^{\mathrm{old}} - \widetilde{A}_{t,r}^{\mathrm{old}}}{N_t^{(r)}}.$$

We report its average over incremental tasks:

$$F_{\mathrm{NFR}}^{(r)} = \frac{1}{T-1} \sum_{t=2}^{T} F_t^{(r)}.$$

Importantly, normalization is performed post hoc: the regional updates are not rescaled before evaluation.

Region-wise Energy-Share Difference. For the ℓ-th LoRA-adapted layer, let $\Delta W_t^{(\ell,r)}$ denote the component of the current update in region $r \in \{\mathrm{bi}, \mathrm{out}, \mathrm{in}, \mathrm{comp}\}$ according to Eq. (3). We define its normalized energy share as

$$e_{t,r}^{(\ell)} = \frac{\|\Delta W_t^{(\ell,r)}\|_F^2}{\|\Delta W_t^{(\ell)}\|_F^2}, \qquad e_{t,r} = \frac{1}{L} \sum_{\ell=1}^{L} e_{t,r}^{(\ell)}.$$

For Fig. 3a, we report the difference between SF-LoRA and Vanilla LoRA:

$$\Delta e_{t,r} = e_{t,r}^{\mathrm{SF}} - e_{t,r}^{\mathrm{Vanilla}}.$$

Positive values indicate that SF-LoRA assigns a larger fraction of update energy to region r, while negative values indicate a smaller fraction. Because the four regions form an orthogonal decomposition, $\sum_r e_{t,r} = 1$ and consequently $\sum_r \Delta e_{t,r} = 0$.

A.3.2 Optimization Algorithm

Algorithm 1 summarizes the training procedure. At each step, the base optimizer first updates the current-task LoRA factors and classifier head, after which SF-LoRA applies the bilateral spectral correction to the updated LoRA factors. Spectral memory is constructed once per task. Here, g_t denotes the classifier head for task t, and layer indices are omitted for clarity.

Algorithm 1 SF-LoRA for Class-Incremental Learning

Require: Pretrained weights W_0, task sequence $\{\mathcal{D}_t\}_{t=1}^{T}$, hyperparameters λ, δ, p

Initialize the deployed weights as W_0
for $t = 1, \ldots, T$ **do**
  Initialize (B_t, A_t), classifier head g_t, and a new optimizer
  Freeze W_{t-1} and previous heads $\{g_\tau\}_{\tau<t}$
  if $t > 1$ **then**
    Recover $H_{t-1} = W_{t-1} - W_0$ and compute $H_{t-1} = U\Sigma V^\top$
    Determine k_δ and retain $(U, \Sigma, V) \leftarrow (U^{(k_\delta)}, \Sigma^{(k_\delta)}, V^{(k_\delta)})$
    Compute $\alpha_i = (\sigma_i/\sigma_1)^p$, $i = 1, \ldots, k_\delta$
  end if
  for each optimization step s **do**
    Update (B_t, A_t, g_t) and optimizer states using $\mathcal{L}_t$
    if $t > 1$ **then**
      $\gamma_s \leftarrow 2\eta_s\lambda, \quad \kappa_{i,s} \leftarrow (1 + \gamma_s\alpha_i)^{-1}$
      $D_s \leftarrow \mathrm{diag}(1 - \kappa_{1,s}, \ldots, 1 - \kappa_{k_\delta,s})$
      $B_t \leftarrow B_t - UD_s(U^\top B_t)$
      $A_t \leftarrow A_t - (A_tV)D_sV^\top$
    end if
  end for
  $W_t \leftarrow W_{t-1} + B_tA_t$
  Retain and freeze g_t; discard (B_t, A_t) and the optimizer
end for

A.3.3 Implementation Details

We utilized a single NVIDIA V100 GPU for all experimental evaluations. For the network architecture, we adopt ViT-B/16 (Dosovitskiy et al., 2021) with 12 transformer blocks pretrained on ImageNet-21K (Russakovsky et al., 2015) as our backbone architecture across all experiments. Consistent with prior works (He et al., 2026; Wu et al., 2025a), we set LoRA rank to $r = 32$, and integrate the LoRA architecture into all attention blocks. The model was optimized using the SGD optimizer with a Cosine Annealing scheduler. We train each task for 20 epochs with a batch size of 48. Following the standard protocol, we report the mean and standard deviation over three runs with different random seeds. Unless otherwise specified, we use $p = 2$ for spectral importance weighting. We present the detailed hyperparameters of all benchmarks in Table 6.

Following prior CIL protocols (He et al., 2026; Qiu et al., 2026; Cheng et al., 2025), each incremental task introduces a bias-free linear classifier head for the newly arriving classes. The current head is optimized jointly with the LoRA parameters, while previous heads are frozen. At inference, both features and classifier weights are ℓ_2-normalized, and logits from all learned heads are concatenated for prediction over all observed classes. Thus, all task heads operate in the same normalized cosine-logit space without head-specific bias or weight-norm scaling. Class-order shuffling is enabled: within each random seed, the resulting class order and task partition are fixed and shared across all compared methods, whereas different seeds induce different class orders. Training uses only data from the current task without replay. Classifier heads are retained for inference under the common CIL protocol. Their storage is identical across compared methods and is therefore not counted as method-specific auxiliary storage in Table 4.

Table 6: Hyperparameter settings for different datasets.

Hyperparameter	ImageNet-A	ImageNet-R	CIFAR-100	CUB-200
Learning rate	0.01	0.01	0.01	0.005
Classifier learning rate	0.05	0.05	0.05	0.005

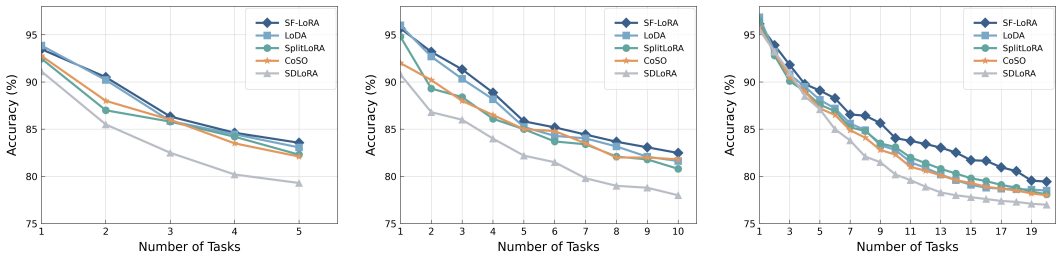
A.4 Supplementary Experiments

A.4.1 Comparison Results

Fig. 4 reports the performance trajectories on ImageNet-R under 5-, 10-, and 20-task settings, complementing the endpoint results in Table 1 with stage-wise comparisons throughout continual learning.

Table 7 further reports results on 20-task ImageNet-A. SF-LoRA achieves 56.30 $\mathcal{A}_{\text{Last}}$ and 66.91 $\mathcal{A}_{\text{Avg}}$, compared with 55.74 and 65.54 for LoDA, the strongest baseline in this setting. The corresponding gains are 0.56 and 1.37 percentage points, showing that the advantage of SF-LoRA persists under a longer incremental sequence.

Comparison with Recent Spectral Methods. Recent spectral continual-adaptation methods, including EBLoRA, E^2-LoRA, and SCLoRA, study spectral properties of low-rank adaptation from different perspectives. EBLoRA focuses on spectral imbalance within task-local updates, E^2-LoRA analyzes the low-rank energy structure of feature drift, and SCLoRA regulates LoRA spectra with respect to pretrained weights. These spectral objects differ from the accumulated bilateral residual studied by SF-LoRA. Our main empirical comparison therefore follows the common ViT-based class-incremental protocol used by prior PEFT-based CIL methods, while Sec. 2 provides a methodological comparison with these recent spectral approaches.



(a) 5 Tasks (b) 10 Tasks (c) 20 Tasks

Figure 4: Variation of the performance of different methods during the learning of ImageNet-R.

Table 7: Results (%) on ImageNet-A (20 Tasks). Results are reported as mean ± standard deviation over three independent runs.

Method	ImageNet-A (20 Tasks)	
	$\mathcal{A}_{Last}\uparrow$	$\mathcal{A}_{Avg}\uparrow$
L2P	$40.48_{\pm 1.78}$	$49.62_{\pm 1.46}$
DualPrompt	$42.28_{\pm 1.94}$	$53.39_{\pm 1.64}$
CODAPrompt	$44.62_{\pm 1.92}$	$54.86_{\pm 0.50}$
InfLoRA	$46.81_{\pm 2.30}$	$57.13_{\pm 0.95}$
SD-LoRA	$49.12_{\pm 1.78}$	$59.63_{\pm 1.57}$
LoRA-P&M	$52.27_{\pm 1.84}$	$62.15_{\pm 1.33}$
LoDA	$55.74_{\pm 1.89}$	$65.54_{\pm 2.43}$
SF-LoRA	$\mathbf{56.30}_{\pm 1.14}$	$\mathbf{66.91}_{\pm 0.96}$

A.4.2 Task-wise Analysis of Regional Forgetting

To complement the averaged region-wise results in Fig. 2b and 2c, we further examine how the interference induced by different spectral regions evolves across individual incremental tasks. Fig. 5 reports the task-wise normalized forgetting rates on all four benchmarks. Following the definition in Appendix A.3.1, all layer-wise components of each region are applied simultaneously at their original magnitudes, and the resulting forgetting is normalized post hoc by their joint Frobenius norm.

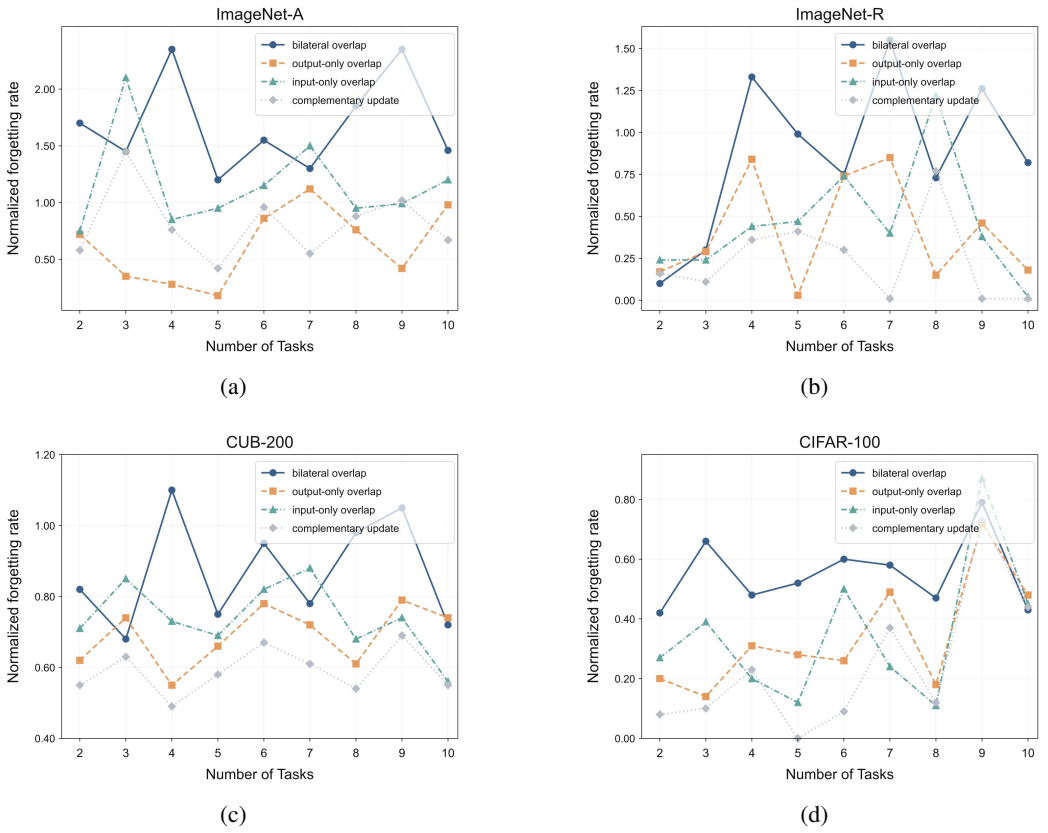


Figure 5: Task-wise normalized forgetting rates of the four spectral regions on ImageNet-A, ImageNet-R, CUB-200, and CIFAR-100. NFR is computed using the joint Frobenius norm across all adapted layers. The bilateral-overlap component generally exhibits stronger interference throughout the incremental sequence.

Across datasets, the bilateral-overlap component generally exhibits the highest NFR over the incremental sequence. Although occasional crossings occur at individual tasks due to task-specific variation, the bilateral component remains consistently more interference-prone overall than the one-sided and complementary components.

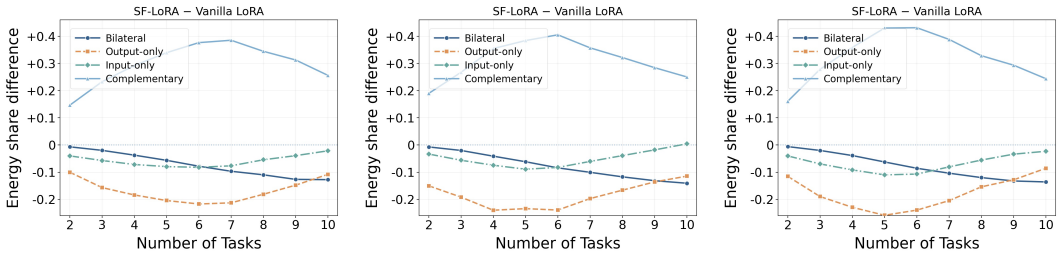
A.4.3 Cross-Dataset Analysis of Regional Energy Redistribution

To complement the region-wise energy analysis on ImageNet-R in Fig. 3a, we further examine how SF-LoRA redistributes update energy on ImageNet-A, CUB-200, and CIFAR-100. Fig. 6 reports the difference in regional energy share between SF-LoRA and Vanilla LoRA across incremental tasks, where positive values indicate a larger energy share under SF-LoRA.

A consistent pattern emerges across all three datasets. As tasks accumulate, the bilateral-overlap component receives a progressively smaller fraction of the update energy, while the complementary component gains substantially more energy. The one-sided components are also generally reduced, reflecting the joint regulation of the historical input- and output-side spectral supports. Together with Fig. 3a, these results show that SF-LoRA consistently redistributes adaptation away from historically overlapping spectral regions and toward complementary directions across different CIL benchmarks.

A.4.4 Hyperparameter Analysis

Complementing the analyses of spectral weighting and truncation in Table 5 and Fig. 3c, we examine the sensitivity to the filtering coefficient λ on 10-task ImageNet-R. This coefficient controls the



(a) ImageNet-A (b) CUB-200 (c) CIFAR-100

Figure 6: Region-wise energy-share difference (SF-LoRA−Vanilla LoRA) across incremental tasks on (a) ImageNet-A, (b) CUB-200, and (c) CIFAR-100. Positive values indicate a larger energy share under SF-LoRA.

overall filtering strength through $\gamma_s = 2\eta_s\lambda$: at a fixed learning rate, a larger λ produces stronger shrinkage along protected historical directions.

Table 8 reports results for $\lambda \in \{50, 100, 150, 200, 250, 300\}$. Performance improves as λ increases from 50 to 150, with $\mathcal{A}_{Last}$ increasing from 81.53 to 82.48 and FR decreasing from 6.83 to 6.02. Further increasing λ does not yield consistent gains and can reduce accuracy or increase forgetting. Overall, performance remains stable over a reasonably broad range of λ, with the shared setting $\lambda = 150$ providing a favorable stability–plasticity trade-off.

Table 8: Sensitivity to the filtering coefficient λ on ImageNet-R (10 tasks).

	$\lambda = 50$	$\lambda = 100$	$\lambda = 150$	$\lambda = 200$	$\lambda = 250$	$\lambda = 300$
$\mathcal{A}_{Last}$	81.53	81.63	**82.48**	81.87	81.75	82.21
$\mathcal{A}_{Avg}$	87.08	86.88	**87.64**	87.02	86.96	87.42
FR	6.83	6.25	**6.02**	6.04	6.58	6.21