EASE: FEDERATED MULTIMODAL UNLEARNING VIA ENTANGLEMENT-AWARE ANCHOR CLOSURE

Anonymous authors
Paper under double-blind review

ABSTRACT

Federated Multimodal Unlearning (FMU) eliminates the influence of a compromised client, a target class, or specified samples from a jointly trained multimodal model without accessing raw data. FMU faces significant challenges due to the entanglement of modalities and client gradient subspaces. Current methods cannot break the cross-modal reconstruction pipeline caused by bilinear coupling, and they do not separate unlearn-exclusive update directions from those shared with remaining clients. To address these, we first introduce an Anchor Principle to formalize bilinear cross-modal coupling, principal-angle subspace entanglement, and ongoing federated updates. We then design EASE, an **E**ntanglement-**A**ware **S**ubspace **E**xcision framework. Within this framework, both the visual and language branches are updated simultaneously to prevent either branch from independently reconstructing the cross-modal alignment. Cosine–Sine decomposition is then employed to filter update directions unique to the unlearning process while preserving those shared among the remaining clients. In addition, a direction-selective unlearn lock is applied to prevent subsequent training from reverting toward the removed directions, thereby reinforcing the unlearning effect. Across three datasets, three backbone architectures, and three unlearning scenarios, EASE most consistently narrows the deviation from the retrain reference on both forget and retain metrics among *nine* baselines. For example, on Flickr30K with CLIP-B/32 under client unlearning, EASE matches the retrain reference within 0.2 R@1 points on the forget set while staying within 4.2 R@1 points on the retain set.

1 INTRODUCTION

Federated Learning (FL) (McMahan et al., 2017; Li et al., 2020) allows multiple clients to jointly train a shared model while keeping their raw data private, establishing itself as a key paradigm in distributed machine learning (Kairouz et al., 2021; Yang et al., 2019). Recent advances integrate FL with multimodal models through parameter-efficient Low-Rank Adaptation (LoRA) (Hu et al., 2022) adapters, forming Federated Multimodal Learning (FML) (Chen et al., 2024; Yu et al., 2023). When deployed in regulated domains, FML inherits the right to be forgotten (Voigt & von dem Bussche, 2017): once a client withdraws, a target concept becomes obsolete, or harmful content surfaces, the system must approximate the retain-only retrain model without centralizing the removed pairs, a task referred to as Federated Unlearning (Liu et al., 2021; Halimi et al., 2022; Wang et al., 2022). We review related work in Appendix A.

As illustrated in Figure 1, while federated unlearning has drawn growing interest (Liu et al., 2021; Halimi et al., 2022; Wu et al., 2022; Gao et al., 2024), extending it to multimodal models remains open: client-update-only access, the bilinear coupling of the alignment loss, and the continued use of FedAvg jointly create failure modes absent from single-modality unlearning. We use an anchor to denote a residual mechanism that keeps a forgotten image-text alignment reconstructable after an unlearning request, and we identify three such mechanisms in this regime, a failure pattern we call the Anchor Principle for federated multimodal contrastive unlearning. The first arises because joint embedding training entangles forgotten knowledge across modalities via bilinear similarity, so the gradient with respect to one modality depends on the embeddings produced by the other. Excising parameters from one branch alone leaves the other as a cross-modal anchor that drives the alignment gradient back toward the original pairing, raising a first question: ***I) How can we close this Modality***

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.

Anchor while keeping each branch's representation faithful to the retained pairs? Compounding this, federated training entangles client gradient subspaces along a continuous spectrum, mixing directions exclusive to the forget set with directions shared with retained clients. Existing methods treat the forget set's contribution as an indivisible block, conflating removable anchors with retain support, and as we confirm in Section 2, they consistently suffer a forget–retain trade-off, motivating a second question: ***II) How can we excise only the removable Unique-Subspace Anchors while sparing the retain-support directions shared with retained clients?***

Even after the first two anchors are closed during unlearning, each client's local SGD follows an alignment gradient with non-zero components along the removed directions. Without a persistent safeguard, every subsequent round of federated training rewrites a fraction of the cut anchors back and gradually undoes the unlearning, raising a third question: ***III) How can we prevent Temporal Re-anchoring across continued rounds while leaving directions outside the forget channel free to recover retained knowledge?***

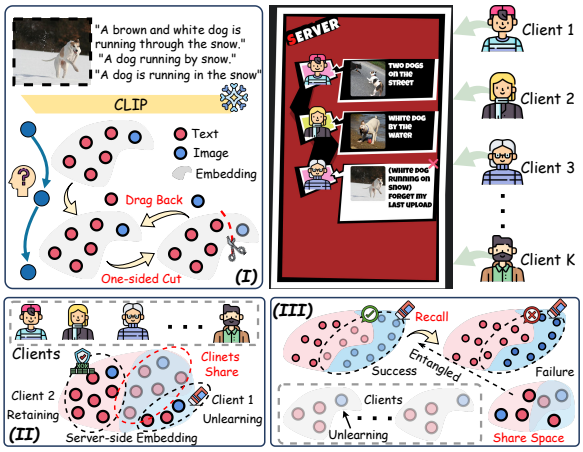


Figure 1: **Problem Illustration.** Three residual anchors in federated multimodal unlearning. **I)** Modality Anchor: the intact branch preserves forgotten alignments after one-sided excision. **II)** Unique-Subspace Anchor: forget-exclusive directions mix with retain support across client subspaces. **III)** Temporal Re-anchoring: continued aggregation rewrites the removed directions during subsequent FedAvg training rounds.

We address all three challenges with EASE: an **E**ntanglement-**A**ware **S**ubspace **E**xcision framework for federated multimodal unlearning, built on lightweight LoRA adapters and projection operators over a frozen multimodal backbone. For challenge **I)**, we establish **Bilateral Knowledge Excision (BKE)** to close the Modality Anchor by displacing the image and text branches simultaneously, so the bilinear similarity can no longer reconstruct the removed pairing through the untouched modality. For challenge **II)**, we develop **Gradient Subspace Decomposition (GSD)**. Reusing client updates already transmitted during FedAvg, the server extracts per-client subspace bases via SVD and measures their pairwise principal angles; directions nearly orthogonal to the retained subspace are treated as forget-exclusive anchors, while directions shared with retained clients are preserved as retain support. For challenge **III)**, we devise **Projection with Forget Lock (PFL)**. At each unlearning round, the server projects the parameter displacement onto the complement of the unique subspace, while a client-side Forget Lock penalizes drift back along the identified unique directions during local SGD, leaving the orthogonal complement unconstrained for retained knowledge recovery. The projection closes the anchor at the unlearning round, and the lock keeps it closed, with drift bounded inversely by the lock strength. Our main contributions are:

❶ ***Anchor Principle.*** We identify and formalize an Anchor Principle for federated multimodal contrastive unlearning: forgotten knowledge persists through three principal anchor locations in this regime, the Modality Anchor (bilinear cross-modal coupling), the Unique-Subspace Anchor (principal-angle subspace entanglement), and Temporal Re-anchoring (residual alignment gradient field). The principle is falsifiable: it predicts that closing any strict subset of the three leaves a measurable reconstruction path, and our anchor-by-anchor ablations confirm this prediction through the alignment residual ρ across client, class, and sample unlearning.

❷ ***Anchor Closure Framework.*** We introduce EASE, realizing the Anchor Principle through three coordinated mechanisms: BKE shuts the Modality Anchor by displacing both branches simultaneously, a step unavailable to single-modality unlearning; GSD replaces block-level erasure of the forget subspace with per-direction disposal via the Cosine–Sine decomposition of the forget–retain subspace pair; and PFL prevents Temporal Re-anchoring with a server-side projection and a client-side penalty supported only on the identified unique directions, in contrast to full-space importance regularization. We back the framework with geometric guarantees on unique-subspace excision and retain-direction preservation.

❸ ***Empirical Verification.*** Experiments on Flickr30K, COCO, and TextCaps with three multimodal backbones across three unlearning scenarios show that EASE most consistently narrows the gap to the retrain reference on both forget and retain sides among nine federated and centralized baselines, with retrieval, alignment-residual, and low-FPR LiRA diagnostics covering utility, anchor closure, and targeted membership leakage.

2 PRELIMINARIES

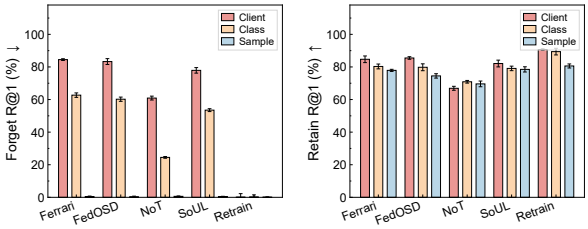
Federated multimodal learning. We consider a federated system with one central server and K clients. Each client k holds a private multimodal dataset $\mathcal{D}_k$ of N_k image-text pairs (x_v, x_t), where subscripts v and t denote the visual and textual modalities throughout, and data never leaves the client. The model consists of frozen pretrained encoders and lightweight trainable modules, namely Low-Rank Adaptation (LoRA) adapters and projectors. We denote the visual and language trainable parameters by w_v and w_t, and write $w = [w_v;\ w_t]$ for the joint parameter vector. A complete notation table covering all symbols used in this paper is provided in Appendix B.1.

Contrastive alignment objective. The visual and language encoders map each input into a shared embedding space, producing m-dimensional representations z_v and z_t. The system is trained via Federated Averaging (FedAvg) with the symmetric InfoNCE alignment loss:

$$\mathcal{L}_a(w;\,\mathcal{D}) = -\frac{1}{2N}\sum_{i=1}^{N}\left[\log\frac{\exp(z_{i,v}^{\top}z_{i,t}/\gamma)}{\sum_{j=1}^{N}\exp(z_{i,v}^{\top}z_{j,t}/\gamma)} + \log\frac{\exp(z_{i,v}^{\top}z_{i,t}/\gamma)}{\sum_{j=1}^{N}\exp(z_{j,v}^{\top}z_{i,t}/\gamma)}\right], \tag{1}$$

where γ is the temperature and $\mathcal{D} = \bigcup_{k=0}^{K-1}\mathcal{D}_k$; the two terms correspond to the visual-to-text and text-to-visual retrieval directions. A key structural property of this loss is that the gradient with respect to one modality depends on the embeddings of the other: $\nabla_{w_v}\mathcal{L}_a$ depends on all textual embeddings $z_{j,t}$, and vice versa; this cross-modal coupling drives the first anchor channel of our principle, the Modality Anchor (Section 3.2).

Unlearning problem. Given an unlearning request, we partition $\mathcal{D}$ into a forget set $\mathcal{D}_f$ and a retain set $\mathcal{D}_r = \mathcal{D} \setminus \mathcal{D}_f$. We consider three practical scenarios: client unlearning, where one or more clients request complete data withdrawal; class unlearning, where all samples matching a target class must be forgotten across the federation; and sample unlearning, where designated samples across all clients must be removed. We instantiate class unlearning with annotation-grounded pseudo-classes derived from dataset-provided entity annotations (Flickr30K Entities) and WordNet-aligned concept categories (COCO, TextCaps); see Appendix B.5. We let w_0 denote the shared federated initialization, the pretrained backbone with freshly initialized trainable modules that has never trained on $\mathcal{D}_f$; w_n the deployed model after standard federated training; and $\tilde{w}$ the retrain model obtained by rerunning the full training pipeline on $\mathcal{D}_r$ from scratch.



(a) Unlearning completeness. (b) Retention integrity.

Figure 2: Forget and retain trade-off of federated unlearning baselines on CLIP-B32 with Flickr30K.

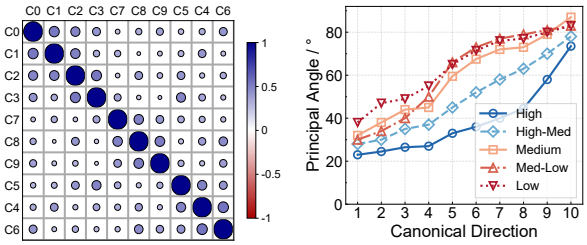
Desiderata. The goal is to produce an unlearned model w^* satisfying three criteria: (i) unlearning completeness, the retrieval behavior of w^* on $\mathcal{D}_f$ matches that of $\tilde{w}$; (ii) retention integrity, the performance of w^* on $\mathcal{D}_r$ matches that of $\tilde{w}$; and (iii) efficiency, the cost is substantially lower than full retraining. The retrain reference fixes the deletion scope: directions supported by both forget and retain data are re-learnable from $\mathcal{D}_r$ alone and persist in $\tilde{w}$ itself, so the target of removal is the forget-exclusive component of the forget data's influence, the part $\tilde{w}$ would not contain. Existing methods instead erase that influence from w as an indivisible block, which cannot separate forget-exclusive directions, the removable anchors, from directions shared with retain, the retain support, and leaves untouched the cross-modal coupling of Section 3.2. These challenges motivate the entanglement-aware excision framework in Section 3.

Observation 1. Existing federated unlearning methods exhibit a forget–retain trade-off on CLIP-B32 with Flickr30K, indicating that at least one anchor channel remains open in every baseline.

We evaluate four federated unlearning baselines, Ferrari (Gu et al., 2024), FedOSD (Pan et al., 2025a), NoT (Khalil et al., 2025), and SoUL (Jia et al., 2024), on CLIP-B32 fine-tuned on Flickr30K under three unlearning scenarios. We report Forget R@1, lower indicating more thorough erasure, and Retain R@1, higher indicating better preservation. Figure 2b shows that most baselines preserve Retain R@1 close to the Retrain reference, suggesting retained knowledge survives the unlearning procedure. Figure 2a, however, reveals a sharp failure mode: only sample unlearning is effective, while client and class requests leave the forget set largely intact. NoT is the exception that erases more aggressively, but at the cost of a lower Retain, while the remaining baselines under-erase. These observations expose a tension in current federated multimodal unlearning, motivating a finer treatment of how forget-side anchors and retain support coexist within client gradient updates.

Observation 2. Client gradient subspaces in federated multimodal training are partially entangled, and their overlap admits a continuous principal-angle spectrum from retain-aligned support to forget-exclusive anchors, with angles rising monotonically from shared to unique directions.

The trade-off in Observation 1 stems from the entangled structure of client gradient updates: the forget gradient subspace mixes removable anchors (forget-exclusive directions) with retain support (directions shared with retained clients), and blind erasure conflates the two. To validate this, we train a CLIP-B/32 projector on Flickr30K distributed across 10 clients via FedAvg for 30 rounds, collect per-round parameter updates, and extract each client's top-10 update directions via SVD. As shown in Figure 3a, pairwise subspace similarity varies widely with an off-diagonal mean of 0.49, confirming that client gradient subspaces are neither fully shared nor fully orthogonal. Figure 3b decomposes five representative pairs by their principal angles: in all cases, the angles increase monotonically from small values for shared directions to large values for unique directions. These findings motivate an entanglement-aware decomposition that erases only the forget-exclusive Unique-Subspace Anchors while leaving the retain support intact.



(a) Subspace similarity. (b) Principal angle spectrum.

Figure 3: Gradient subspace entanglement in federated multimodal training on Flickr30K with 10 clients.

3 Methodology

3.1 Framework Overview

Motivated by Observations 1–2, our framework closes the three anchor channels in turn (Figure 4): **I)** Bilateral Knowledge Excision (BKE, Section 3.2) shuts the Modality Anchor by displacing both visual and language branches simultaneously; **II)** Gradient Subspace Decomposition (GSD, Section 3.3) separates removable Unique-Subspace Anchors from retain support via Cosine–Sine decomposition of the forget–retain subspace pair; **III)** Projection with Forget Lock (PFL, Section 3.4) prevents Temporal Re-anchoring through a server-side projection and a direction-selective client-side penalty.

3.2 Closing the Modality Anchor (BKE)

The Modality Anchor. In multimodal models with a bilinear similarity (e.g., InfoNCE), each modality's gradient depends on the other modality's embedding. Excising parameter directions on one branch alone leaves the unmodified branch as a memory of the original alignment, and the alignment gradient exploits this memory as an anchor that reconstructs the erased pairing. BKE suppresses this Modality Anchor by displacing both branches simultaneously, as formalized in Appendix C.8.

Anchor mechanism. The cross-modal alignment depends on both the visual and language parameters through the bilinear similarity in Eq. (1). To see why unilateral excision fails, consider a

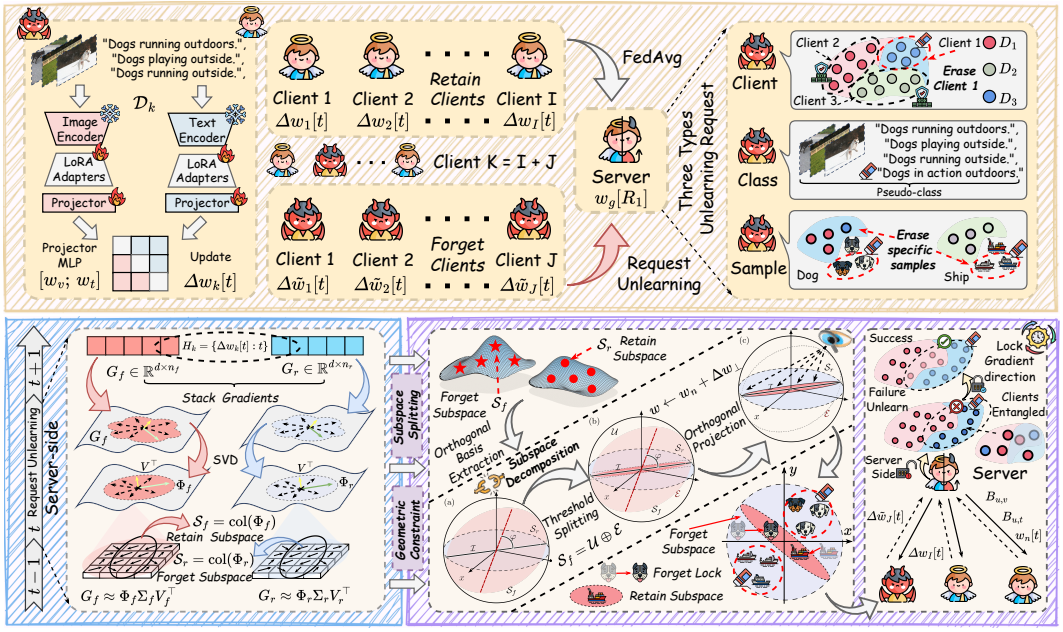


Figure 4: **Overview of EASE.** BKE displaces both visual and language branches simultaneously to close the cross-modal reconstruction channel. GSD separates forget-exclusive directions from retain support via principal-angle decomposition. PFL projects the server displacement onto the unique subspace and locks each client's drift along it during continued training.

forget-set pair $(x_{v,i}, x_{t,i})$. The InfoNCE gradient with respect to w_v decomposes as:

$$\nabla_{w_v} \mathcal{L}_a = \frac{1}{\gamma} J_v(w_v)^\top \left(\sum_{j=1}^{N} \frac{\exp\!\left(z_{i,v}^\top z_{j,t}/\gamma\right)}{\sum_l \exp\!\left(z_{i,v}^\top z_{t,l}/\gamma\right)} z_{j,t}(w_t) \;-\; z_{i,t}(w_t) \right), \tag{2}$$

where J_v denotes the Jacobian $\partial z_v / \partial w_v$. This cross-modal dependence is the physical seat of the Modality Anchor: if we remove forget-unique directions from w_v only, leaving w_t at the pre-unlearning value $w_{n,t}$, the visual representation is displaced, yet the textual embedding $z_{i,t}(w_{n,t})$ still encodes the original pairing. Since this intact embedding dominates the softmax weights in the contrastive loss, the gradient evaluated at the excised point reduces to:

$$\nabla_{w_v} \mathcal{L}_a \big|_{w_v^*,\, w_{n,t}} \;\propto\; J_v(w_v^*)^\top \, z_{i,t}(w_{n,t}). \tag{3}$$

The unchanged textual embedding $z_{i,t}(w_{n,t})$ is the Modality Anchor: it encodes the semantic content of the erased pair and continuously drives the alignment gradient to re-align w_v with the original pairing during retraining.

Alignment residual. We quantify the open Modality Anchor by the normalized gap between the unlearned model w^* and the retrain reference $\tilde{w}$:

$$\rho = \frac{\left| z_v(w_v^*)^\top z_t(w_t^*) \;-\; z_v(\tilde{w}_v)^\top z_t(\tilde{w}_t) \right|}{\left| z_v(w_{n,v})^\top z_t(w_{n,t}) \;-\; z_v(\tilde{w}_v)^\top z_t(\tilde{w}_t) \right| + \epsilon}, \tag{4}$$

with w_n the pre-unlearning model and ϵ a stability constant. $\rho \approx 0$ indicates the anchor is closed; $\rho \approx 1$ that it remains fully open.

Comparison of excision strategies. **(i)** Visual-only excision on w_v leaves $z_t(w_{n,t})$ as the Modality Anchor that re-pulls z_v toward the forgotten alignment, yielding $\rho > 0$; **(ii)** by symmetry, language-only excision also yields $\rho > 0$; **(iii)** bilateral excision (ours) displaces both embeddings, and with GSD and PFL holding the unique directions closed across rounds, $\rho \approx 0$. Appendix C.8 formalizes this contrast under Assumptions A1–A4 (Appendix C.1).

3.3 CLOSING THE UNIQUE-SUBSPACE ANCHOR (GSD)

Removable anchors versus retain support. In federated non-IID training, the forget client's parameter updates share directions with retained clients along a continuous principal-angle spectrum, so BKE's bilateral displacement must be complemented by decomposing the forget subspace into removable Unique-Subspace Anchors (forget-exclusive directions, safe to erase) and retain support (directions shared with retain, must preserve).

Gradient history collection. Over the initial R_1 FedAvg rounds, each client's per-round update $\Delta w_k[t]$ (its post-local-SGD parameter delta relative to the broadcast global model) is already transmitted to the server; caching it as a per-client history H_k follows the provenance setting standard in federated unlearning (Liu et al., 2021; Cao et al., 2023; Gu et al., 2024), adds no communication beyond FedAvg, and stores only LoRA and projector deltas, a small fraction of backbone size (Appendix B.8). The full history is consumed only by client unlearning; sample and class unlearning depend on a single request-specific round in which clients holding matching data upload forget-specific gradients (Appendix B.5). The global model w_n after R_1 rounds is the deployed pre-unlearning model; the excision and the Forget Lock instead reference the initialization w_0 (Section 2), whose trainable modules never trained on $\mathcal{D}_f$.

Forget and retain subspace extraction. We construct a forget gradient matrix $G_f \in \mathbb{R}^{d\times n_f}$ by column-stacking the updates associated with the forget set $\mathcal{D}_f$. The construction of G_f depends on the unlearning scenario. For client unlearning, G_f collects the full history of updates from the requesting client:

$$G_f = \big[\Delta w_{k^*}[t_1],\ \ldots,\ \Delta w_{k^*}[t_{|H_{k^*}|}]\big], \tag{5}$$

where $t_1, \ldots, t_{|H_{k^*}|}$ are the rounds in which k^* participated. For sample and class unlearning, the forget targets are generally distributed across multiple clients. The server broadcasts a target descriptor (a set of sample identifiers for sample unlearning, or a textual descriptor for class unlearning), and every client that holds matching data performs one epoch of training on its local forget subset and uploads the resulting gradient; G_f is column-stacked from these per-client contributions. A retain gradient matrix G_r is constructed analogously from each client's remaining non-forget data (see Appendix B.5 for details).

We apply Singular Value Decomposition (SVD) to G_f and retain the top-p left singular vectors that capture at least a fraction τ_e of the total energy:

$$p = \min\Big\{j \in \{1,\ldots,\mathrm{rank}(G_f)\}\ :\ \frac{\sum_{i=1}^{j}\sigma_i^2}{\|G_f\|_F^2}\ \geq\ \tau_e\Big\}, \tag{6}$$

This energy-based criterion adapts p to the spectral decay of G_f. The resulting forget subspace basis is $\Phi_f \in \mathbb{R}^{d\times p}$, whose columns are the top-p left singular vectors of G_f. The same procedure applied to G_r yields the retain subspace basis $\Phi_r \in \mathbb{R}^{d\times q}$. We denote the column spaces of Φ_f and Φ_r by $\mathcal{S}_f$ and $\mathcal{S}_r$, respectively.

Entanglement spectrum. Not all directions inside $\mathcal{S}_f$ behave the same way: a direction lying entirely within $\mathcal{S}_r$ is fully retain-aligned and serves as retain support, since erasing it destroys retained knowledge; a direction orthogonal to $\mathcal{S}_r$ is forget-exclusive and constitutes a removable anchor. To quantify this overlap for every forget-side canonical direction, we introduce the entanglement spectrum of $\mathcal{S}_f$ with respect to $\mathcal{S}_r$. We compute the cross-correlation matrix $M = \Phi_f^\top \Phi_r$ and take its full SVD:

$$M = U_f\,\Sigma\,V_f^\top,\quad U_f \in \mathbb{R}^{p\times p},\quad V_f \in \mathbb{R}^{q\times q},\quad r = \min(p,q), \tag{7}$$

where the rectangular diagonal of Σ has entries $\cos\varphi_1, \ldots, \cos\varphi_r$, the cosines of the principal angles. If $p > q$, the remaining $p - q$ columns of U_f span directions in $\mathcal{S}_f$ that are orthogonal to $\mathcal{S}_r$; we assign them entanglement coefficient 0. Thus for every $i \in \{1,\ldots,p\}$ we define $\kappa_i = \cos\varphi_i$ for $i \leq r$ and $\kappa_i = 0$ for $i > r$. The rotated basis vectors $c_i = (\Phi_f U_f)_{\cdot,i}$ are the canonical directions of $\mathcal{S}_f$ relative to $\mathcal{S}_r$: a coefficient near 1 indicates retain-aligned (retain support); near 0 indicates forget-exclusive (removable anchor). We partition using a threshold $\delta \in [0,1]$ into unique (anchor) and entangled (support) basis matrices, where a larger δ classifies more directions as removable anchors; we analyze the sensitivity to this threshold in Section 4:

$$B_u = \big[c_i\big]_{\kappa_i \leq \delta} \in \mathbb{R}^{d\times|\mathcal{U}|},\quad B_e = \big[c_i\big]_{\kappa_i > \delta} \in \mathbb{R}^{d\times|\mathcal{E}|}. \tag{8}$$

The unique subspace $\mathcal{U} = \mathrm{col}(B_u)$ spans forget-exclusive removable anchors (collectively Unique-Subspace Anchor) and the entangled subspace $\mathcal{E} = \mathrm{col}(B_e)$ spans retain support, with $\mathcal{S}_f = \mathcal{U} \oplus \mathcal{E}$. The decomposition is intrinsic to the subspace pair $(\mathcal{S}_f, \mathcal{S}_r)$, which distinguishes anchor identification from single-subspace projection methods that erase $\mathcal{S}_f$ as a block. Applying it independently to each LoRA layer and projector within both modalities yields per-group bases $B^l_{u,\mu}$ and $B^l_{e,\mu}$; the theorems in Section 3.4 apply per block, and global guarantees follow by summing across blocks.

3.4 Preventing Temporal Re-anchoring (PFL)

Temporal Re-anchoring. A one-time projection excises the Unique-Subspace Anchor at $t{=}0$, but retain-data SGD still has nonzero components $g_u(t) := \Pi_{u,\mu} \nabla_{w_\mu} \mathcal{L}_a(w(t))$ along $\mathcal{U}_\mu$, since the unique subspace is only partially orthogonal to the retain gradient field. Letting $d(t) := \Pi_{u,\mu}(w_\mu(t) - w_{0,\mu})$, a single SGD step under the combined objective gives the recurrence

$$d(t{+}1) \;=\; (1 - 2\eta_{\mathrm{lr}}\alpha)\, d(t) \;-\; \eta_{\mathrm{lr}}\, g_u(t), \tag{9}$$

where $g_u(t)$ is the residual alignment gradient surviving along $\mathcal{U}_\mu$ after BKE has suppressed the Modality Anchor's leading component. Without persistent constraint ($\alpha = 0$), $g_u(t)$ accumulates across rounds and rewrites the cut anchors, the third anchor channel.

Direction-selective lock. PFL constrains drift along $\mathcal{U}_\mu$ while leaving $\mathcal{U}_\mu^\perp$ unconstrained, an asymmetry that neither per-round server projection alone nor axis-aligned importance regularizers can enforce (Appendix B.12). The construction pairs a server-side projection that pulls each global modality $\mu \in \{v, t\}$ back to $\mathcal{M}_\mu$ with a client-side quadratic penalty supported on $\mathcal{U}_\mu$:

$$w_{g,\mu} \;\leftarrow\; w_{g,\mu} - \Pi_{u,\mu}(w_{g,\mu} - w_{0,\mu}), \quad \Pi_{u,\mu} := B_{u,\mu} B_{u,\mu}^\top, \tag{10}$$

$$\mathcal{L}_f(w) = \sum_{\mu \in \{v,t\}} \left\| B_{u,\mu}^\top (w_\mu - w_{0,\mu}) \right\|_2^2. \tag{11}$$

Each client trains on the retained data $\mathcal{D}'_k = \mathcal{D}_k \setminus \mathcal{D}_f$ with the combined objective

$$\mathcal{L}(w;\, \mathcal{D}'_k) = \mathcal{L}_a(w;\, \mathcal{D}'_k) + \alpha \cdot \mathcal{L}_f(w), \tag{12}$$

where $\alpha > 0$ is the lock strength. The recurrence Eq. (9) then becomes a contraction with factor $1 - 2\eta_{\mathrm{lr}}\alpha < 1$ (Theorem C.11), bounding unique-subspace drift by $G_u/(2\alpha)$ regardless of round count, while $\mathcal{U}_\mu^\perp$ (including the entangled retain support from GSD) stays unconstrained so retained knowledge recovers. The bases $B_{u,v}$, $B_{u,t}$, and reference w_0 are broadcast with the global model, so $\mathcal{L}_f$ is local with no extra communication. After the excision step, the server runs R_2 stabilization rounds without projection; the lock remains active (Algorithm 1, Appendix B.13).

$\mathcal{L}_f$ acts on exactly the basis GSD returns, so the two thresholds form a pair of complementary controls: δ sets how much of the entangled spectrum is treated as removable, and α sets how firmly those directions stay closed under continued training, with $\mathcal{U}_\mu^\perp$ left free for recovery (Appendices C.4–C.7; δ and α sensitivity in Section 4; robustness of the estimated $\mathcal{U}$ in Appendix B.8).

What the theorems certify. Theorems C.8–C.11 certify the geometry of anchor closure: projection removes the full displacement along the identified $\mathcal{U}_\mu$, collateral energy on $\mathcal{S}_r$ stays within $\delta^2 \|\Delta_r\|_2^2$, and the Forget Lock contracts unique-subspace drift to $G_u/(2\alpha)$. Behavioral equivalence to the retrain reference is measured directly through ρ, raincloud distributions, and LiRA TPR (Section 4, Appendices B.11, B.10, E, E.1).

4 Experiments

4.1 Experimental Setup

Datasets. We evaluate on three benchmark image-text datasets: Flickr30K (Young et al., 2014), MS COCO (Lin et al., 2014), and TextCaps (Sidorov et al., 2020), covering standard caption retrieval and scene-text retrieval. We use three backbones: CLIP-B/32 (Radford et al., 2021), CLIP-L/14, and SigLIP (Zhai et al., 2023). Dataset splits are detailed in Appendix B.2; implementation details and hyperparameters are in Appendix B.5. Class unlearning targets the coarse pseudo-classes defined in

Section 2; construction details are in Appendix B.5. Unless noted otherwise, reported numbers are means over three seeds (42, 1234, 2024).

Counterparts. We compare our method against nine baselines from two categories. Federated unlearning methods: (1) **FedEraser** (Liu et al., 2021), (2) **FedRecover** (Cao et al., 2023), (3) **Ferrari** (Gu et al., 2024), (4) **FedOSD** (Pan et al., 2025a), (5) **NoT** (Khalil et al., 2025), (6) **SoUL** (Jia et al., 2024), (7) **FFMU** (Che et al., 2023), (8) **FUSED** (Zhong et al., 2025); centralized unlearning adapted to the federated setting: (9) **GradAscent** (Thudi et al., 2022); and the retrain reference $\tilde{w}$. Detailed descriptions of all counterparts are provided in Appendix B.3.

Evaluation metrics. We evaluate four metrics: Recall@k (Karpathy & Fei-Fei, 2015) with $k \in \{1, 5, 10\}$ for cross-modal retrieval, the likelihood-ratio attack true-positive rate at 1% false-positive rate (LiRA TPR@FPR=1%) (Carlini et al., 2022) as the privacy audit, a shadow-model membership inference attack (MIA) (Shokri et al., 2017) score reported alongside it for comparability with prior federated unlearning work, and communication cost (Kairouz et al., 2021) in megabytes. Full metric definitions are in Appendix B.4. With this setup, three research questions structure the remainder of the section: **Q1** (Superiority), **Q2** (Effectiveness), and **Q3** (Sensitivity); Appendix D adds compute-matched retrains.

4.2 Superiority (Q1)

For **Q1**, we test whether w^* matches the retrain reference $\tilde{w}$ on the forget set $\bar{\mathcal{D}}_f$ and the retain set $\bar{\mathcal{D}}_r$. Table 1 gives the comparison on Flickr30K with CLIP-B/32; eight dataset–backbone combinations are in Appendix F. Figure 5 summarizes per-metric performance across scenarios.

Table 1: Main comparison on Flickr30K with CLIP-B/32 across three unlearning scenarios. The top sub-table reports forget-side metrics (F-$R@k$) and membership inference (MIA); the bottom sub-table reports retain-side metrics (R-$R@k$) and communication cost ($Comm.$, MB). For F-$R@k$ and R-$R@k$, parenthetical values give the absolute gap to the retrain reference with (↑green) closer to ideal and (↓red) farther; for MIA, (val) is $|\text{MIA} - 50|$ and smaller is better; for $Comm.$, it is the absolute difference to retrain. Best in **red**, runner-up in blue among baselines and our method; Tied scores share a rank; *Retrain* does not participate in ranking.

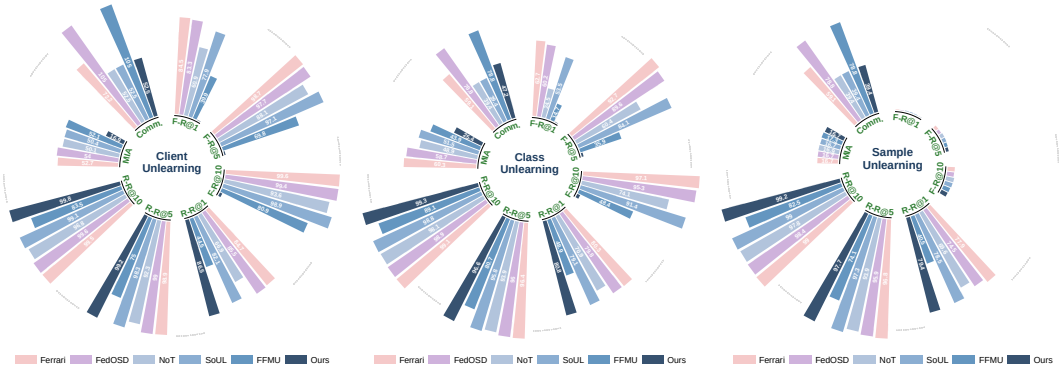
Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-$R@1$ ↓	F-$R@5$ ↓	F-$R@10$ ↓	MIA	F-$R@1$ ↓	F-$R@5$ ↓	F-$R@10$ ↓	MIA	F-$R@1$ ↓	F-$R@5$ ↓	F-$R@10$ ↓	MIA
FedEraser	69.8 (↓69.7)	90.9 (↓89.4)	95.2 (↓92.9)	48.9 (1.1)	38.8 (↓38.6)	68.2 (↓67.5)	79.2 (↓77.8)	64.0 (14.0)	0.3 (↓0.2)	2.4 (↓1.3)	5.8 (↓3.4)	16.7 (33.3)
FedRecover	57.9 (↓57.8)	81.2 (↓79.7)	87.7 (↓85.4)	47.2 (2.8)	41.6 (↓41.4)	72.0 (↓71.3)	83.5 (↓82.1)	41.7 (8.3)	0.5 (↓0.4)	2.8 (↓1.7)	6.2 (↓3.8)	16.9 (33.1)
Ferrari	84.5 (↓84.4)	98.7 (↓97.2)	99.6 (↓97.3)	52.7 (2.7)	62.7 (↓62.5)	92.7 (↓92.0)	97.1 (↓95.7)	60.3 (10.3)	0.4 (↓0.3)	2.5 (↓1.4)	5.9 (↓3.5)	16.7 (33.3)
FedOSD	83.3 (↓83.2)	97.7 (↓96.2)	99.4 (↓97.1)	54.0 (4.0)	60.2 (↓60.0)	89.6 (↓88.9)	95.3 (↓93.9)	58.7 (8.7)	0.3 (↓0.2)	2.4 (↓1.3)	5.7 (↓3.3)	16.7 (33.3)
NoT	60.9 (↓60.8)	88.1 (↓86.6)	93.6 (↓91.3)	**50.1** (0.1)	24.5 (↓24.3)	60.4 (↓59.7)	74.1 (↓72.7)	**48.9** (1.1)	0.6 (↓0.5)	3.0 (↓1.9)	6.5 (↓4.1)	16.9 (33.1)
SoUL	77.9 (↓77.8)	97.1 (↓95.6)	98.9 (↓96.6)	50.4 (0.4)	53.5 (↓53.3)	84.1 (↓83.4)	91.4 (↓90.0)	51.5 (1.5)	0.5 (↓0.4)	2.9 (↓1.8)	6.3 (↓3.9)	16.7 (33.3)
FFMU	39.9 (↓39.8)	69.8 (↓68.3)	80.9 (↓78.6)	52.1 (2.1)	14.7 (↓14.5)	35.9 (↓35.2)	48.4 (↓47.0)	43.9 (6.1)	0.3 (↓0.2)	2.5 (↓1.4)	5.8 (↓3.4)	**17.3** (32.7)
FUSED	82.1 (↓82.0)	97.5 (↓96.0)	99.4 (↓97.1)	52.1 (2.1)	56.5 (↓56.3)	87.3 (↓86.6)	93.9 (↓92.5)	58.9 (8.9)	0.4 (↓0.3)	2.7 (↓1.6)	6.0 (↓3.6)	16.7 (33.3)
GradAscent	1.8 (↓1.7)	7.2 (↓5.7)	14.1 (↓11.8)	54.5 (4.5)	15.9 (↓15.7)	39.4 (↓38.7)	53.7 (↓52.3)	46.5 (3.5)	0.3 (↓0.2)	2.4 (↓1.3)	5.7 (↓3.3)	16.8 (33.2)
EASE	**0.3** (↓0.2)	**1.0** (↑0.5)	**2.2** (↑0.1)	16.9 (33.1)	**0.3** (↓0.1)	**1.5** (↓0.8)	**2.6** (↓1.2)	25.4 (24.6)	**0.1**	**2.2** (↓1.1)	**5.5** (↓3.1)	16.7 (33.3)
Retrain	0.1	1.5	2.3	16.7	0.2	0.7	1.4	24.2	0.1	1.1	2.4	16.7

Method	R-$R@1$ ↑	R-$R@5$ ↑	R-$R@10$ ↑	$Comm.$ ↓	R-$R@1$ ↑	R-$R@5$ ↑	R-$R@10$ ↑	$Comm.$ ↓	R-$R@1$ ↑	R-$R@5$ ↑	R-$R@10$ ↑	$Comm.$ ↓
FedEraser	85.1 (↓5.6)	98.7 (↓1.2)	99.5 (↓0.5)	262.5	78.8 (↓10.6)	96.0 (↓3.9)	98.9 (↓1.1)	196.9	78.3 (↓2.3)	96.9 (↓0.8)	99.0 (↓0.4)	196.9
FedRecover	73.6 (↓17.1)	94.9 (↓5.0)	97.2 (↓2.8)	**47.2** (215.3)	76.8 (↓12.6)	93.0 (↓6.9)	94.6 (↓5.4)	**35.4** (161.5)	58.8 (↓21.8)	81.6 (↓16.1)	87.4 (↓12.0)	**35.4** (161.5)
Ferrari	84.7 (↓6.0)	98.9 (↓1.0)	99.5 (↓0.5)	72.2 (190.3)	80.3 (↓9.1)	96.4 (↓3.5)	99.1 (↓0.9)	59.1 (137.8)	77.9 (↓2.7)	96.8 (↓0.9)	99.0 (↓0.4)	59.1 (137.8)
FedOSD	85.5 (↓5.2)	99.0 (↓0.9)	99.6 (↓0.4)	105.0 (157.5)	79.8 (↓9.6)	96.0 (↓3.9)	98.9 (↓1.1)	78.8 (118.1)	74.5 (↓6.1)	95.9 (↓1.8)	98.4 (↓1.0)	78.8 (118.1)
NoT	66.9 (↓23.8)	92.3 (↓7.6)	96.8 (↓3.2)	52.5 (210.0)	70.9 (↓18.5)	93.9 (↓6.0)	98.1 (↓1.9)	39.4 (157.5)	69.6 (↓11.0)	93.9 (↓3.8)	97.8 (↓1.6)	39.4 (157.5)
SoUL	82.1 (↓8.6)	98.3 (↓1.6)	99.1 (↓0.9)	52.5 (210.0)	79.1 (↓10.3)	95.8 (↓4.1)	98.8 (↓1.2)	39.4 (157.5)	78.5 (↓2.1)	97.3 (↓0.4)	99.0 (↓0.4)	39.4 (157.5)
FFMU	44.5 (↓46.2)	75.0 (↓24.9)	83.5 (↓16.5)	105.0 (157.5)	48.9 (↓40.5)	80.7 (↓19.2)	89.1 (↓10.9)	78.8 (118.1)	40.5 (↓40.1)	74.1 (↓23.6)	82.5 (↓16.9)	78.8 (118.1)
FUSED	86.2 (↓4.5)	99.1 (↓0.8)	99.7 (↓0.3)	105.0 (157.5)	80.5 (↓8.9)	96.4 (↓3.5)	99.2 (↓0.8)	78.8 (118.1)	79.0 (↓1.6)	97.5 (↓0.2)	99.1 (↓0.3)	78.8 (118.1)
GradAscent	1.6 (↓89.1)	7.8 (↓92.1)	13.1 (↓86.9)	72.2 (190.3)	36.7 (↓52.7)	66.2 (↓33.7)	79.6 (↓20.4)	59.1 (137.8)	76.6 (↓4.0)	96.0 (↓1.7)	98.7 (↓0.7)	59.1 (137.8)
EASE	**86.5** (↓4.2)	**99.2** (↓0.7)	**99.8** (↓0.2)	52.5 (210.0)	**80.8** (↓8.6)	**96.6** (↓3.3)	**99.3** (↓0.7)	47.2 (149.7)	**79.4** (↓1.2)	**97.7**	**99.2** (↓0.2)	39.4 (157.5)
Retrain	90.7	99.9	100.0	262.5	89.4	99.9	100.0	196.9	80.6	97.7	99.4	196.9

Across nine dataset/backbone combinations in Table 1 and Appendix F, EASE most consistently narrows both F-$R@k$ and R-$R@k$ to retrain. GradAscent erases thoroughly but collapses retain by conflating anchors with support; Ferrari and FUSED preserve retain but under-erase by leaving the Modality Anchor open. Appendices B.11 and B.9 examine the embedding distributions and anchor signals, respectively.

4.3 Effectiveness (Q2)

For **Q2**, each ablation row of Table 2 disables exactly one component (marked ✗) of the full method on CLIP-B/32 with Flickr30K across all three unlearning scenarios. GSD is the most load-bearing



(a) Client Unlearning (b) Class Unlearning (c) Sample Unlearning

Figure 5: Per-metric comparison on CLIP-B/32 / Flickr30K under three unlearning scenarios; each sector is one metric and bars extend outward with its value. Our method achieves the lowest Forget R@k with competitive Retain R@k.

Table 2: Ablation on CLIP-B/32 / Flickr30K across three unlearning scenarios (mean over three seeds). Top: forget-side metrics and MIA; bottom: retain-side metrics and alignment residual ρ (Section 3.2; ρ=0 matches retrain). Parentheticals give the absolute gap to retrain, (↑green) closer and (↓red) farther; for MIA, (val) is $|\mathrm{MIA} - 50|$ (smaller is better), with best/runner-up in red/blue and ties sharing a rank. ✓ enabled, ✗ disabled; †visual-only, ‡text-only.

Components			Client Unlearning				Class Unlearning				Sample Unlearning			
BKE	GSD	PFL	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
✓	✓	✓	0.4 (↓0.3)	1.2 (↓0.3)	2.1 (↑0.2)	16.8 (33.2)	0.3 (↓0.1)	1.4 (↓0.7)	2.7 (↓1.3)	25.1 (24.9)	0.1	2.0 (↓0.9)	5.3 (↓2.9)	16.8 (33.2)
✗†	✓	✓	1.3 (↓1.2)	2.7 (↓1.2)	3.9 (↓1.6)	18.4 (31.6)	1.0 (↓0.8)	2.4 (↓1.7)	3.8 (↓2.4)	26.3 (23.7)	0.8 (↓0.7)	3.0 (↓1.9)	6.1 (↓3.7)	17.6 (32.4)
✗‡	✓	✓	1.8 (↓1.7)	3.4 (↓1.9)	5.0 (↓2.7)	**19.1** (30.9)	1.6 (↓1.4)	2.9 (↓2.2)	4.6 (↓3.2)	**26.8** (23.2)	1.1 (↓1.0)	3.5 (↓2.4)	6.7 (↓4.3)	**18.0** (32.0)
✓	✗	✓	0.2 (↓0.1)	0.8 (↑0.7)	2.0 (↑0.3)	17.2 (32.8)	0.2	1.1 (↓0.4)	1.8 (↓0.4)	24.6 (25.4)	0.0 (↓0.1)	1.6 (↓0.5)	4.7 (↓2.3)	17.1 (32.9)
✓	✓	✗	0.7 (↓0.6)	1.9 (↓0.4)	3.5 (↓1.2)	17.0 (33.0)	0.5 (↓0.3)	2.0 (↓1.3)	3.5 (↓2.1)	25.9 (24.1)	0.4 (↓0.3)	2.6 (↓1.5)	6.2 (↓3.8)	17.0 (33.0)
Retrain			0.1	1.5	2.3	16.7	0.2	0.7	1.4	24.2	0.1	1.1	2.4	16.7
BKE	**GSD**	**PFL**	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	ρ ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	ρ ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	ρ ↓
✓	✓	✓	86.3 (↓4.4)	99.1 (↓0.8)	99.8 (↓0.2)	0.08 (↓0.08)	80.6 (↓8.8)	96.7 (↓3.2)	99.2 (↓0.8)	0.07 (↓0.07)	79.5 (↓1.1)	97.7	99.2 (↓0.2)	0.06 (↓0.06)
✗†	✓	✓	83.4 (↓7.3)	97.9 (↓2.0)	99.1 (↓0.9)	0.23 (↓0.23)	78.2 (↓11.2)	95.3 (↓4.6)	98.5 (↓1.5)	0.21 (↓0.21)	77.8 (↓2.8)	96.9 (↓0.8)	98.7 (↓0.7)	0.19 (↓0.19)
✗‡	✓	✓	81.8 (↓8.9)	97.6 (↓2.3)	98.7 (↓1.3)	0.29 (↓0.29)	77.6 (↓11.8)	94.5 (↓5.4)	98.1 (↓1.9)	0.26 (↓0.26)	76.9 (↓3.7)	96.5 (↓1.2)	98.4 (↓1.0)	0.23 (↓0.23)
✓	✗	✓	12.8 (↓77.9)	31.1 (↓68.8)	52.6 (↓47.4)	0.10 (↓0.10)	5.2 (↓84.2)	18.4 (↓81.5)	39.8 (↓60.2)	0.12 (↓0.12)	7.1 (↓73.5)	25.3 (↓72.4)	46.9 (↓52.5)	0.11 (↓0.11)
✓	✓	✗	88.1 (↓2.6)	99.3 (↓0.6)	99.7 (↓0.3)	0.15 (↓0.15)	82.8 (↓6.6)	97.1 (↓2.8)	99.2 (↓0.8)	0.13 (↓0.13)	80.2 (↓0.4)	97.6 (↓0.1)	99.0 (↓0.4)	0.12 (↓0.12)
Retrain			90.7	99.9	100.0	0	89.4	99.9	100.0	0	80.6	97.7	99.4	0

component for utility: removing it applies excision to the undecomposed forget subspace, so forget-side metrics remain close to retrain but retain support is over-erased, dropping client R-R@1 from 86.3 to 12.8 (Theorem C.9). Removing BKE reopens the Modality Anchor at the embedding level: under client unlearning, ρ rises from 0.08 to 0.23 with visual-only excision and 0.29 with text-only excision, alongside a retain loss in every scenario. Removing the Forget Lock trades durability for a small retain margin: retain rises slightly while ρ roughly doubles. Only the full method balances forget closure, retain preservation, and low residual alignment; Appendices B.7 and B.14 add keep-one-component variants and VLM/MLLM transfer.

4.4 Sensitivity (Q3)

We sweep δ, α, K, and β (Figures 6–8, Appendix B.6). Both thresholds act as design controls rather than tuned constants: raising δ from 0.1 to 0.9 more than halves client F-R@10 (6.2 to 2.8) while client R-R@1 moves 4.4 points, and every threshold at or above 0.5 selects the same subspace (Table 6). The robustness studies of Appendix B.8 run at the default δ and α with no retuning.

5 Conclusion

We presented EASE, which combines bilateral excision, subspace decomposition, and a direction-selective Forget Lock under the Anchor Principle. Across datasets, backbones, and unlearning scenarios, EASE narrows forget- and retain-side gaps to retrain.

AI USE STATEMENT

Generative AI tools were used in this work to review the research methodology, to refine how the research questions are stated, and to help interpret experimental results. They were not used to propose the research hypotheses, to design the experiments, to create synthetic data sets, to clean or reformat the datasets, to build the theoretical framework, to state any mathematical claim or supply a critical step in its proof, to implement the method or any part of the evaluation pipeline, or to translate text; qualitative and thematic data analysis does not arise in this work. Beyond the above, we used generative AI tools to retrieve and summarize related literature, to draft the notation table of the appendix, to typeset tables in LaTeX, to edit the manuscript for clarity, and to verify and format the bibliography. We have reviewed all AI-assisted work. Every reported number originates in an experiment log from the authors' own runs and was checked against that log wherever it appears in the paper; each assumption, theorem and proof step was written and verified by the authors against the claim it supports; and every bibliography entry was confirmed field by field against the publisher record, CrossRef or DBLP, with no field completed from model output. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REFERENCES

Rahaf Aljundi, Francesca Babiloni, Mohamed Elhoseiny, Marcus Rohrbach, and Tinne Tuytelaars. Memory Aware Synapses: Learning What (not) to Forget. In *Proceedings of the European Conference on Computer Vision*, pp. 144–161, 2018. doi: 10.1007/978-3-030-01219-9_9.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-VL Technical Report. arXiv preprint arXiv:2502.13923, 2025.

Lucas Bourtoule, Varun Chandrasekaran, Christopher A. Choquette-Choo, Hengrui Jia, Adelin Travers, Baiwu Zhang, David Lie, and Nicolas Papernot. Machine Unlearning. In *IEEE Symposium on Security and Privacy*, pp. 141–159, 2021. doi: 10.1109/SP40001.2021.00019.

Xiaoyu Cao, Jinyuan Jia, Zaixi Zhang, and Neil Zhenqiang Gong. FedRecover: Recovering from Poisoning Attacks in Federated Learning using Historical Information. In *IEEE Symposium on Security and Privacy*, pp. 1366–1383, 2023. doi: 10.1109/SP46215.2023.10179336.

Nicholas Carlini, Steve Chien, Milad Nasr, Shuang Song, Andreas Terzis, and Florian Tramèr. Membership Inference Attacks From First Principles. In *IEEE Symposium on Security and Privacy*, pp. 1897–1914, 2022. doi: 10.1109/SP46214.2022.9833649.

Tianshi Che, Yang Zhou, Zijie Zhang, Lingjuan Lyu, Ji Liu, Da Yan, Dejing Dou, and Jun Huan. Fast Federated Machine Unlearning with Nonlinear Functional Theory. In *Proceedings of the International Conference on Machine Learning*, 2023.

Haokun Chen, Yao Zhang, Denis Krompass, Jindong Gu, and Volker Tresp. FedDAT: An Approach for Foundation Model Finetuning in Multi-Modal Heterogeneous Federated Learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 11285–11293, 2024. doi: 10.1609/aaai.v38i10.29007.

Xinlei Chen, Hao Fang, Tsung-Yi Lin, Ramakrishna Vedantam, Saurabh Gupta, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO Captions: Data Collection and Evaluation Server. arXiv preprint arXiv:1504.00325, 2015.

Mehrdad Farajtabar, Navid Azizan, Alex Mott, and Ang Li. Orthogonal Gradient Descent for Continual Learning. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 3762–3773, 2020.

XiaoHua Feng, Chaochao Chen, Yuyuan Li, and Zibin Lin. Fine-grained Pluggable Gradient Ascent for Knowledge Unlearning in Language Models. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 10141–10155, 2024. doi: 10.18653/v1/2024.emnlp-main.566.

Jack Foster, Stefan Schoepf, and Alexandra Brintrup. Fast Machine Unlearning without Retraining through Selective Synaptic Dampening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 12043–12051, 2024. doi: 10.1609/aaai.v38i11.29092.

Chaohao Fu, Weijia Jia, and Na Ruan. Client-Free Federated Unlearning via Training Reconstruction with Anchor Subspace Calibration. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 9281–9285, 2024. doi: 10.1109/icassp48485.2024.10447085.

Xiangshan Gao, Xingjun Ma, Jingyi Wang, Youcheng Sun, Bo Li, Shouling Ji, Peng Cheng, and Jiming Chen. VeriFi: Towards Verifiable Federated Unlearning. *IEEE Transactions on Dependable and Secure Computing*, 21(6):5720–5736, 2024. doi: 10.1109/TDSC.2024.3382321.

Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal Sunshine of the Spotless Net: Selective Forgetting in Deep Networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 9301–9309, 2020. doi: 10.1109/CVPR42600.2020.00932.

Hanlin Gu, Win Kent Ong, Chee Seng Chan, and Lixin Fan. Ferrari: Federated Feature Unlearning via Optimizing Feature Sensitivity. In *Advances in Neural Information Processing Systems*, 2024.

Anisa Halimi, Swanand Kadhe, Ambrish Rawat, and Nathalie Baracaldo. Federated Unlearning: How to Efficiently Erase a Client in FL? arXiv preprint arXiv:2207.05521, 2022.

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.

Jinghan Jia, Yihua Zhang, Yimeng Zhang, Jiancheng Liu, Bharat Runwal, James Diffenderfer, Bhavya Kailkhura, and Sijia Liu. SOUL: Unlocking the Power of Second-Order Optimization for LLM Unlearning. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing*, pp. 4276–4292, 2024. doi: 10.18653/v1/2024.emnlp-main.245.

Peter Kairouz, H. Brendan McMahan, Brendan Avent, Aurélien Bellet, Mehdi Bennis, Arjun Nitin Bhagoji, Kallista Bonawitz, Zachary Charles, Graham Cormode, Rachel Cummings, et al. Advances and Open Problems in Federated Learning. *Foundations and Trends in Machine Learning*, 14(1–2):1–210, 2021. doi: 10.1561/2200000083.

Andrej Karpathy and Li Fei-Fei. Deep Visual-Semantic Alignments for Generating Image Descriptions. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 3128–3137, 2015. doi: 10.1109/CVPR.2015.7298932.

Yasser H. Khalil, Leo Brunswic, Soufiane Lamghari, Xu Li, Mahdi Beitollahi, and Xi Chen. NoT: Federated Unlearning via Weight Negation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 25759–25769, 2025. doi: 10.1109/CVPR52734.2025.02399.

Hyunjune Kim, Sangyong Lee, and Simon S. Woo. Layer Attack Unlearning: Fast and Accurate Machine Unlearning via Layer Level Attack and Knowledge Distillation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pp. 21241–21248, 2024. doi: 10.1609/aaai.v38i19.30118.

James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017. doi: 10.1073/pnas.1611835114.

Pang Wei Koh and Percy Liang. Understanding Black-box Predictions via Influence Functions. In *Proceedings of the International Conference on Machine Learning*, pp. 1885–1894, 2017.

Meghdad Kurmanji, Peter Triantafillou, Jamie Hayes, and Eleni Triantafillou. Towards Unbounded Machine Unlearning. In *Advances in Neural Information Processing Systems*, volume 36, pp. 1957–1987, 2023.

Jiaqi Li, Chuanyi Zhang, Miaozeng Du, Hui Zhang, Yongrui Chen, Qianshan Wei, Junfeng Fang, Ruipeng Wang, Sheng Bi, and Guilin Qi. Forget the Token and Pixel: Rethinking Gradient Ascent for Concept Unlearning in Multimodal Generative Models. In *Findings of the Association for Computational Linguistics: ACL*, pp. 12179–12200, 2025. doi: 10.18653/v1/2025.findings-acl.630.

Junnan Li, Dongxu Li, Caiming Xiong, and Steven C. H. Hoi. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. In *Proceedings of the International Conference on Machine Learning*, pp. 12888–12900, 2022.

Junnan Li, Dongxu Li, Silvio Savarese, and Steven C. H. Hoi. BLIP-2: Bootstrapping Language-Image Pre-Training with Frozen Image Encoders and Large Language Models. In *Proceedings of the International Conference on Machine Learning*, pp. 19730–19742, 2023.

Kunhao Li, Wenhao Li, Di Wu, Lei Yang, Jun Bai, Ju Jia, and Jason Xue. Cross-Modal Unlearning via Influential Neuron Path Editing in Multimodal Large Language Models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 35589–35597, 2026. doi: 10.1609/aaai.v40i42.40870.

Tian Li, Anit Kumar Sahu, Ameet Talwalkar, and Virginia Smith. Federated Learning: Challenges, Methods, and Future Directions. *IEEE Signal Processing Magazine*, 37(3):50–60, 2020. doi: 10.1109/MSP.2020.2975749.

Wei Li, Tao Feng, Hangjie Yuan, Ang Bian, Guodong Du, Sixin Liang, Jianhong Gan, and Ziwei Liu. UniGrad-FS: Unified Gradient Projection With Flatter Sharpness for Continual Learning. *IEEE Transactions on Industrial Informatics*, 20(12):13873–13882, 2024. doi: 10.1109/tii.2024.3435499.

Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: Common Objects in Context. In *Proceedings of the European Conference on Computer Vision*, pp. 740–755, 2014. doi: 10.1007/978-3-319-10602-1_48.

Gaoyang Liu, Xiaoqiang Ma, Yang Yang, Chen Wang, and Jiangchuan Liu. FedEraser: Enabling Efficient Client-Level Data Removal from Federated Learning Models. In *IEEE/ACM International Symposium on Quality of Service*, pp. 1–10, 2021. doi: 10.1109/IWQOS52092.2021.9521274.

Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. Improved Baselines with Visual Instruction Tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 26286–26296, 2024. doi: 10.1109/CVPR52733.2024.02484.

Yi Lu, Shu Li, and Yurong Qian. Unlearning in Cross-Modal Retrieval via Prior-Prototype Guided Partitioned Dampening. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 40, pp. 7583–7590, 2026. doi: 10.1609/aaai.v40i9.37699.

Brendan McMahan, Eider Moore, Daniel Ramage, Seth Hampson, and Blaise Agüera y Arcas. Communication-Efficient Learning of Deep Networks from Decentralized Data. In *Proceedings of the International Conference on Artificial Intelligence and Statistics*, pp. 1273–1282, 2017.

Zibin Pan, Zhichao Wang, Chi Li, Kaiyan Zheng, Boqi Wang, Xiaoying Tang, and Junhua Zhao. Federated Unlearning with Gradient Descent and Conflict Mitigation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pp. 19804–19812, 2025a. doi: 10.1609/aaai.v39i19.34181.

Zijie Pan, Zuobin Ying, Yajie Wang, Chuan Zhang, Weiting Zhang, Wanlei Zhou, and Liehuang Zhu. Feature-Based Machine Unlearning for Vertical Federated Learning in IoT Networks. *IEEE Transactions on Mobile Computing*, 24(6):5031–5044, 2025b. doi: 10.1109/tmc.2025.3530529.

Samuele Poppi, Tobia Poppi, Federico Cocchi, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Safe-CLIP: Removing NSFW Concepts from Vision-and-Language Models. In *Proceedings of the European Conference on Computer Vision*, pp. 340–356, 2024a. doi: 10.1007/978-3-031-73668-1_20.

Samuele Poppi, Sara Sarto, Marcella Cornia, Lorenzo Baraldi, and Rita Cucchiara. Unlearning Vision Transformers Without Retaining Data via Low-Rank Decompositions. In *Proceedings of the International Conference on Pattern Recognition*, pp. 147–163, 2024b. doi: 10.1007/978-3-031-78122-3_10.

Benliu Qiu, Heqian Qiu, Haitao Wen, Lanxiao Wang, Yu Dai, Fanman Meng, Qingbo Wu, and Hongliang Li. Geodesic-Aligned Gradient Projection for Continual Task Learning. *IEEE Transactions on Image Processing*, 34:1995–2007, 2025. doi: 10.1109/tip.2025.3551139.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the International Conference on Machine Learning*, pp. 8748–8763, 2021.

Gobinda Saha, Isha Garg, and Kaushik Roy. Gradient Projection Memory for Continual Learning. In *International Conference on Learning Representations*, 2021.

Reza Shokri, Marco Stronati, Congzheng Song, and Vitaly Shmatikov. Membership Inference Attacks Against Machine Learning Models. In *IEEE Symposium on Security and Privacy*, pp. 3–18, 2017. doi: 10.1109/SP.2017.41.

Oleksii Sidorov, Ronghang Hu, Marcus Rohrbach, and Amanpreet Singh. TextCaps: A Dataset for Image Captioning with Reading Comprehension. In *Proceedings of the European Conference on Computer Vision*, pp. 742–758, 2020. doi: 10.1007/978-3-030-58536-5_44.

Uri Stern, Eli Corn, and Daphna Weinshall. Forget Me Not: Fighting Local Overfitting With Knowledge Fusion and Distillation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48 (5):5004–5015, 2025. doi: 10.1109/tpami.2025.3647862.

Anvith Thudi, Hengrui Jia, Ilia Shumailov, and Nicolas Papernot. On the Necessity of Auditable Algorithmic Definitions for Machine Unlearning. In *USENIX Security Symposium*, pp. 4007–4022, 2022.

Ramakrishna Vedantam, C. Lawrence Zitnick, and Devi Parikh. CIDEr: Consensus-Based Image Description Evaluation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 4566–4575, 2015.

Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer International Publishing, 2017. doi: 10.1007/978-3-319-57959-7.

Junxiao Wang, Song Guo, Xin Xie, and Heng Qi. Federated Unlearning via Class-Discriminative Pruning. In *Proceedings of the ACM Web Conference*, pp. 622–632, 2022. doi: 10.1145/3485447.3512222.

Pengfei Wang, Wei Song, Heng Qi, Changjun Zhou, Fuliang Li, Yong Wang, Peng Sun, and Qiang Zhang. Server-Initiated Federated Unlearning to Eliminate Impacts of Low-Quality Data. *IEEE Transactions on Services Computing*, 17(3):1196–1211, 2024a. doi: 10.1109/tsc.2024.3355188.

Shipeng Wang, Xiaorong Li, Jian Sun, and Zongben Xu. Training Networks in Null Space of Feature Covariance for Continual Learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 184–193, 2021. doi: 10.1109/CVPR46437.2021.00025.

Weiqi Wang, Chenhan Zhang, Zhiyi Tian, and Shui Yu. FedU: Federated Unlearning via User-Side Influence Approximation Forgetting. *IEEE Transactions on Dependable and Secure Computing*, 22(3):2550–2562, 2024b. doi: 10.1109/tdsc.2024.3520614.

Zhenwei Wang, Pengfei Wang, Guangjie Han, Jianxin Zhang, Qing Lin, Muhammad Ameen, and Qiang Zhang. Inverse Feature Consistency Federated Unlearning for Vision-Language Model. *IEEE Transactions on Mobile Computing*, 25(4):5267–5284, 2025. doi: 10.1109/tmc.2025.3629294.

Zichen Wang, Xiangshan Gao, Cong Wang, Peng Cheng, and Jiming Chen. Efficient Vertical Federated Unlearning via Fast Retraining. *ACM Transactions on Internet Technology*, 24(2):1–22, 2024c. doi: 10.1145/3657290.

Chen Wu, Sencun Zhu, and Prasenjit Mitra. Federated Unlearning with Knowledge Distillation. arXiv preprint arXiv:2201.09441, 2022.

Heng Xu, Tianqing Zhu, Lefeng Zhang, Wanlei Zhou, and Philip S. Yu. Update Selective Parameters: Federated Machine Unlearning Based on Model Explanation. *IEEE Transactions on Big Data*, 11 (2):524–539, 2024. doi: 10.1109/tbdata.2024.3409947.

Enneng Yang, Li Shen, Zhenyi Wang, Shiwei Liu, Guibing Guo, Xingwei Wang, and Dacheng Tao. Revisiting Flatness-Aware Optimization in Continual Learning With Orthogonal Gradient Projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(5):3895–3907, 2025. doi: 10.1109/tpami.2025.3539019.

Qiang Yang, Yang Liu, Tianjian Chen, and Yongxin Tong. Federated Machine Learning: Concept and Applications. *ACM Transactions on Intelligent Systems and Technology*, 10(2):12:1–12:19, 2019. doi: 10.1145/3298981.

Peter Young, Alice Lai, Micah Hodosh, and Julia Hockenmaier. From image descriptions to visual denotations: New similarity metrics for semantic inference over event descriptions. *Transactions of the Association for Computational Linguistics*, 2:67–78, 2014. doi: 10.1162/tacl_a_00166.

Lang Yu, Lina Ge, Guanghui Wang, Jianghao Yin, Qin Chen, Jie Zhou, and Liang He. PI-Fed: Continual Federated Learning With Parameter-Level Importance Aggregation. *IEEE Internet of Things Journal*, 11(22):37187–37199, 2024. doi: 10.1109/jiot.2024.3440029.

Qiying Yu, Yang Liu, Yimu Wang, Ke Xu, and Jingjing Liu. Multimodal Federated Learning via Contrastive Representation Ensemble. In *International Conference on Learning Representations*, 2023.

Xiaohan Yuan, Jiqiang Liu, Bin Wang, Guorong Chen, Xiangrui Xu, Junyong Wang, Tao Li, and Wei Wang. FedEditor: Efficient and Effective Federated Unlearning in Cooperative Intelligent Transportation Systems. *IEEE Transactions on Information Forensics and Security*, 20:6560–6575, 2025. doi: 10.1109/tifs.2025.3583231.

Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual Learning Through Synaptic Intelligence. In *Proceedings of the International Conference on Machine Learning*, pp. 3987–3995, 2017.

Xiaohua Zhai, Basil Mustafa, Alexander Kolesnikov, and Lucas Beyer. Sigmoid Loss for Language Image Pre-Training. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 11941–11952, 2023. doi: 10.1109/ICCV51070.2023.01100.

Shizhou Zhang, Yue Lu, De Cheng, Yinghui Xing, Nannan Wang, Peng Wang, and Yanning Zhang. VPT-NSP^2++: Importance-Aware Visual Prompt Tuning in Null Space for Continual Learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 48(4):4318–4335, 2025. doi: 10.1109/tpami.2025.3642298.

Yang Zhang, Zhiyu Hu, Yimeng Bai, Jiancan Wu, Qifan Wang, and Fuli Feng. Recommendation Unlearning via Influence Function. *ACM Transactions on Recommender Systems*, 3(2):1–23, 2024. doi: 10.1145/3701763.

Fengda Zhao, Qianyi Xu, Hao Wang, and Dingding Guo. Rapid federated unlearning with tuning parameters based on fisher information matrix. *Applied Intelligence*, 55(15), 2025a. doi: 10.1007/s10489-025-06593-0.

Shuai Zhao, Junying Zhang, Xindi Ma, Qi Jiang, Zhuo Ma, Sheng Gao, Zuobin Ying, and Jianfeng Ma. FedWiper: Federated Unlearning via Universal Adapter. *IEEE Transactions on Information Forensics and Security*, 20:4042–4054, 2025b. doi: 10.1109/tifs.2025.3557671.

Zhengyi Zhong, Weidong Bao, Ji Wang, Shuai Zhang, Jingxuan Zhou, Lingjuan Lyu, and Wei Yang Bryan Lim. Unlearning through Knowledge Overwriting: Reversible Federated Unlearning via Selective Sparse Adapter. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 30661–30670, 2025. doi: 10.1109/CVPR52734.2025.02855.

Changjun Zhou, Chenglin Pan, Minglu Li, and Pengfei Wang. Federated Unlearning With Fast Recovery. *IEEE Transactions on Mobile Computing*, 24(10):9709–9725, 2025. doi: 10.1109/tmc.2025.3563265.

APPENDIX CONTENTS

A Related Work **16**

B Technical Appendix **16**

B.1 Notation Table 16
B.2 Dataset Details 17
B.3 Baseline Descriptions 18
B.4 Evaluation Metrics 19
B.5 Implementation Details 20
B.6 Hyperparameter Sensitivity 20
B.7 Component Combination Ablation 22
B.8 Robustness of the GSD Decomposition 22
B.9 Anchor Diagnostics on Baselines 25
B.10 Relearning Resistance 26
B.11 Visualization of Unlearning Effect 26
B.12 Alternative Lock Designs 27
B.13 Algorithm 28
B.14 Transfer to VLM and MLLM Backbones 28
B.15 Broader Impacts 29

C Proofs of Theoretical Results **30**

C.1 Assumptions 30
C.2 Properties of the Unlearning Projector 31
C.3 Forget Lock Characterization 32
C.4 Proof of Theorem C.8 (Unlearning Completeness) 32
C.5 Proof of Theorem C.9 (Retention Integrity) 34
C.6 Completeness–Integrity Trade-off Analysis 34
C.7 Proof of Theorem C.11 (Unlearning Durability) 35
C.8 Proof of Proposition C.12 (Anchor Reconstruction Tendency under Unilateral Excision) 35

D Compute-Matched Retrain Comparison **38**

E Likelihood-Ratio Attack at Low False-Positive Rate **38**

E.1 Entanglement-Stratified Membership Inference 40

F Full Comparison Results **40**

A Related Work

Federated Unlearning. Federated unlearning reconstructs the influence of a withdrawn client without retraining. Historical reconstruction calibrates remaining clients' updates (Liu et al., 2021; Cao et al., 2023; Gu et al., 2024); curvature- and direction-based approaches localize the forget contribution via second-order signals, exemplified by SoUL (Jia et al., 2024; Zhao et al., 2025a; Wang et al., 2024b), or orthogonal subspace projection, exemplified by FedOSD (Pan et al., 2025a; Fu et al., 2024; Xu et al., 2024); task-vector methods, exemplified by NoT (Khalil et al., 2025; Zhou et al., 2025; Wang et al., 2024a), subtract a directional vector of the forget client's accumulated update; recent extensions reduce cost through nonlinear functional refinement, exemplified by FFMU (Che et al., 2023; Wang et al., 2024c; Yuan et al., 2025), or sparse-adapter overwriting, exemplified by FUSED (Zhong et al., 2025; Zhao et al., 2025b; Pan et al., 2025b); and verification-oriented studies (Gao et al., 2024; Halimi et al., 2022; Wang et al., 2022; Wu et al., 2022) provide complementary auditing tools. However, these methods treat the forget contribution as an indivisible block, assume a single-modality loss, and take a one-shot view of unlearning: none separates forget-exclusive from retain-shared directions inside the forget subspace, and none addresses the cross-modal reconstruction channel or continued-training drift in federated multimodal settings. Their provenance requirements also span a spectrum: historical-reconstruction and subspace methods cache per-round client updates as we do, while task-vector and adapter methods operate from the final model alone; EASE sits scenario-dependently on this spectrum, consuming the full history only for client unlearning and a single request-specific round for sample and class requests (Section 3.3).

Centralized Machine Unlearning. Centralized methods assume a trusted curator with full data access. Exact approaches retrain affected data shards (Bourtoule et al., 2021; Foster et al., 2024; Wang et al., 2024c); approximate approaches edit a trained model via gradient ascent on the forget loss (Thudi et al., 2022; Golatkar et al., 2020; Feng et al., 2024), influence-function subtraction (Koh & Liang, 2017; Zhang et al., 2024; Wang et al., 2024b), distillation from a retain-only teacher (Kurmanji et al., 2023; Kim et al., 2024; Stern et al., 2025), or importance-weighted regularization on parameter coordinates (Kirkpatrick et al., 2017; Zenke et al., 2017; Aljundi et al., 2018). However, none of these methods is designed for unlearning that must remain valid under continued training over heterogeneous clients.

Multimodal Unlearning. A small but growing body of work targets unlearning in multimodal contrastive models. CLIP-specific erasers (Poppi et al., 2024a;b; Wang et al., 2025) remove undesired image-text associations via fine-tuning on a designated forget set, and recent multimodal unlearners (Li et al., 2026; Lu et al., 2026; Li et al., 2025) acknowledge cross-modal coupling by issuing dual-branch updates. However, all these efforts are centralized and assume unrestricted access to raw pairs, so they do not transfer to the federated regime where the server only sees gradient summaries.

Gradient Subspace and Importance-weighted Regularization. Two long-standing families in continual learning are closely related to our subspace and lock components. Gradient subspace methods, exemplified by GPM, Adam-NSCL, and OGD (Saha et al., 2021; Wang et al., 2021; Farajtabar et al., 2020; Yang et al., 2025; Qiu et al., 2025; Li et al., 2024; Zhang et al., 2025), confine new updates to the orthogonal complement of a single past-task subspace, while importance-weighted regularizers, exemplified by EWC and SI (Kirkpatrick et al., 2017; Zenke et al., 2017; Yu et al., 2024), weight a quadratic drift penalty by per-parameter importance scores. However, both families were conceived for centralized continual learning rather than unlearning, and neither separates forget-exclusive directions from retain-shared directions inside a forget subspace nor remains valid under continued federated training.

B Technical Appendix

B.1 Notation Table

We summarize the commonly used notation throughout this paper in Table 3.

Table 3: Notation and Definitions.

Notation	Definition
K	The number of clients.
$\mathcal{D}_k$	The private multimodal dataset for the k-th client.
N_k	The number of image-text pairs for the k-th client.
(x_v, x_t)	An image-text pair in $\mathcal{D}_k$.
$w_v,\ w_t$	The trainable parameters for the visual and language branches.
w	The joint parameter vector $[w_v;\ w_t]$.
$z_v,\ z_t$	The visual and textual embeddings in the shared space.
m	The dimension of the embedding space.
γ	The temperature in the InfoNCE loss.
$\mathcal{L}_a$	The symmetric InfoNCE alignment loss.
$J_v,\ J_t$	The Jacobians $\partial z_v/\partial w_v$ and $\partial z_t/\partial w_t$.
$\mathcal{D}_f$	The forget set whose influence is to be removed.
$\mathcal{D}_r$	The retain set $\mathcal{D} \setminus \mathcal{D}_f$.
w_0	The shared federated initialization; the anchor reference for excision and the Forget Lock.
w_n	The deployed model after standard federated training.
$\tilde{w}$	The retrain model obtained by training from scratch on $\mathcal{D}_r$.
w^*	The unlearned model.
ρ	Alignment residual, computed per forget pair and reported as the median over $\mathcal{D}_f$.
$\Delta w_k[t]$	The parameter update of client k at round t.
H_k	The gradient update history of client k.
$G_f,\ G_r$	The forget and retain gradient matrices.
$\Phi_f,\ \Phi_r$	The orthonormal bases of $\mathcal{S}_f$ and $\mathcal{S}_r$.
$\mathcal{S}_f,\ \mathcal{S}_r$	The forget and retain subspaces.
$p,\ q$	The dimensions of $\mathcal{S}_f$ and $\mathcal{S}_r$.
τ_e	The energy threshold for SVD truncation.
M	The cross-correlation matrix $\Phi_f^\top \Phi_r$.
$\cos\varphi_i$	The entanglement coefficient of the i-th canonical direction.
c_i	The i-th canonical direction of $\mathcal{S}_f$.
δ	The threshold separating removable anchor directions from retain support directions.
$B_u,\ B_e$	The unique (anchor) and entangled (support) basis matrices.
$\mathcal{U},\ \mathcal{E}$	The subspaces spanned by removable anchor directions and retain support directions.
$\Pi_{u,\mu}$	The orthogonal projector onto the unique subspace $\mathcal{U}_\mu$, defined as $B_{u,\mu}(B_{u,\mu})^\top$.
$\mathcal{L}_f$	The Forget Lock regularizer.
α	The Forget Lock regularization strength.
η_{lr}	The learning rate.
G_u	The uniform upper bound on the projected alignment gradient $\|\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a\|_2$.

B.2 Dataset Details

We evaluate on three standard image-text retrieval benchmarks.

Flickr30K [TACL'14] (Young et al., 2014). Flickr30K contains 31,783 images collected from Flickr, each annotated with five human-written English captions, for a total of 158,915 image-text pairs. Following the Karpathy split (Karpathy & Fei-Fei, 2015), we use 29,000 images for training, 1,014 for validation, and 1,000 for the test set used to compute Recall@k; the remaining 769 images are not assigned to any split and are omitted from our evaluation, consistent with prior retrieval work using this split. The captions cover diverse everyday scenes, making Flickr30K a standard benchmark for general-purpose cross-modal retrieval.

MS COCO Captions [ECCV'14] (Lin et al., 2014; Chen et al., 2015). MS COCO is a large-scale dataset comprising 123,287 images of common objects in context, each paired with five human-

written captions. We again adopt the Karpathy split, using $113{,}287$ training images, $5{,}000$ validation images, and the $5{,}000$-image test split for retrieval evaluation. Compared with Flickr30K, MSCOCO covers a broader object taxonomy and a wider caption style, and stresses retrieval on a larger candidate pool.

TextCaps [ECCV'20] (Sidorov et al., 2020). TextCaps extends the MSCOCO-style caption format with scene-text understanding: each of the $28{,}408$ images contains visible textual content (signs, logos, product labels), and captions are written to reference this text where appropriate. We use the official splits of $21{,}953$ training, $3{,}166$ validation, and $3{,}289$ test images. TextCaps complements the other two datasets by stressing fine-grained alignment between language and visual text, a regime where forget-set leakage is particularly sensitive to bilateral closure of the Modality Anchor.

Federated partition. For every dataset, training images (together with their captions) are distributed across K clients via a Dirichlet partition with concentration parameter β over pseudo-cluster labels obtained by KMeans on concatenated visual-textual embeddings of the pretrained backbone (Section B.5). The default configuration is $K{=}50$ and $\beta{=}0.5$, yielding moderately non-IID client shards; we sweep both parameters in Figure 7. Forget sets are constructed on top of this partition according to the unlearning scenario, as described in Appendix B.5.

B.3 Baseline Descriptions

This section provides a comprehensive overview of the nine counterparts employed in our experiments. We group them by methodological category.

Federated unlearning methods.

- **FedEraser** [IWQOS'21] (Liu et al., 2021). FedEraser is one of the earliest federated unlearning methods. It reconstructs an approximation of the retrained model by combining cached historical parameter updates from the remaining clients, replaying them through a calibrated aggregation procedure. This avoids the cost of full retraining while producing a model that behaves similarly to one never trained on the forget client's data.
- **FedRecover** [S&P'23] (Cao et al., 2023). FedRecover originally targets recovery from poisoning attacks but is naturally adapted to federated unlearning. It stores per-round gradient and model information, then uses this history to estimate what the global model would have looked like if the forget client had been absent, applying a low-cost correction instead of rerunning training from scratch.
- **Ferrari** [ICML'24] (Gu et al., 2024). Ferrari performs rapid approximate retraining guided by cached client gradients. The server constructs a compact representation of each client's contribution and selectively reverses the forget client's influence, achieving significantly lower wall-clock time than retrain while preserving retrieval accuracy on the retain set.
- **FedOSD** [AAAI'25] (Pan et al., 2025a). FedOSD applies orthogonal subspace decomposition to client update directions. It identifies the subspace spanned by the forget client's updates and projects the global model away from that subspace, selectively removing the forget component. The method is close in spirit to ours, but operates on single-modality models and does not separate removable anchor directions from retain support directions inside the forget subspace.
- **NoT** [CVPR'25] (Khalil et al., 2025). NoT treats each client's historical contribution as a task vector and performs unlearning by subtracting the forget client's task vector from the global model. This yields aggressive erasure on the forget set but frequently damages retained knowledge, since the task vector subtraction is applied uniformly without separating shared and unique components.
- **SoUL** [EMNLP'24] (Jia et al., 2024). SoUL employs second-order curvature information to localize the forget contribution. Using approximate Hessian or Fisher-information-based signals, it estimates which parameters are most relevant to the forget data and perturbs only those. While more targeted than gradient-based baselines, it incurs nontrivial computational cost due to second-order estimation.
- **FFMU** [ICML'23] (Che et al., 2023). FFMU accelerates federated unlearning by modeling local unlearning models as output functions of a Nemytskii operator and leveraging nonlinear functional

analysis to refine them. It achieves competitive unlearning quality with reduced computation, but operates on single-modality models and does not address the Modality Anchor present in multimodal architectures.

- **FUSED** [CVPR'25] (Zhong et al., 2025). FUSED performs reversible federated unlearning by training independent sparse adapters that overwrite forget-specific knowledge. It identifies the most affected layers through sensitivity analysis and retrains only sparse adapters on those layers, reducing unlearning cost while keeping the process reversible. Like other federated unlearning methods, it operates on single-modality models.

Centralized unlearning methods adapted to the federated setting.

- **GradAscent** [USENIX Security'22] (Thudi et al., 2022). Gradient ascent is a simple and widely used unlearning baseline that performs gradient updates on the forget set using the negative of the original training loss. We adapt it to the federated setting by running localized gradient ascent on the forget client followed by standard federated aggregation on the remaining clients. Although straightforward to implement, it often overshoots and degrades retain-set performance.

B.4 Evaluation Metrics

We report four groups of metrics to capture unlearning completeness, retention integrity, privacy leakage, and efficiency.

Retrieval accuracy (Recall@k) [CVPR'15] (Karpathy & Fei-Fei, 2015). Recall@k is the standard metric for cross-modal retrieval. Given a query from one modality, it measures whether the ground-truth match in the other modality is ranked within the top-k retrieved items. We report Recall@k for $k \in \{1, 5, 10\}$, averaged over both image-to-text and text-to-image directions. Three splits are evaluated: (i) the forget set $\bar{\mathcal{D}}_f$ for unlearning completeness, where lower values indicate that the model has forgotten the target pairs; (ii) the retain set $\bar{\mathcal{D}}_r$ for retention integrity, where higher values indicate that retained knowledge is preserved; and (iii) the held-out test set for generalization, where higher values indicate that unlearning does not harm out-of-sample performance.

Membership inference attack (MIA) [S&P'17] (Shokri et al., 2017). We adopt a shadow-model MIA protocol: shadow models are trained on disjoint splits with known member/non-member labels, and their per-sample loss and logit statistics train a binary attack classifier. The classifier is then applied to the target unlearned model. We retain the reported MIA percentage and use 50% as the aggregate comparison target. The ranking score is $d_{50} = |\mathrm{MIA} - 50|$, so a smaller distance is better on either side of 50%. MIA parentheticals report d_{50}; best and runner-up colors follow its two smallest distinct values, with ties sharing a rank and Retrain excluded from ranking. This convention is separate from the retrain-based retrieval gaps and from the low-FPR LiRA metric below.

LiRA true-positive rate at 1% false-positive rate [S&P'22] (Carlini et al., 2022). An aggregate MIA score can hide worst-case leakage because an attacker usually cares about confidently identifying a small fraction of training members rather than performing well on average. We therefore also report the likelihood-ratio attack (LiRA) of Carlini et al. evaluated on the same shadow-model pool. For each candidate sample, LiRA computes a per-example likelihood-ratio statistic between two Gaussian distributions fitted over shadow-model logits (one for models that contain the sample, one for models that do not); the threshold is calibrated so that the false-positive rate on held-out non-members equals 1%, and we report the resulting true-positive rate ($\mathrm{TPR@FPR}{=}1\%$) on the forget set. Values near the 1% random-guess line indicate that forget samples are close to non-members under a targeted low-FPR attacker, while larger values indicate residual memorization that the aggregate MIA score can underestimate.

Communication cost (Kairouz et al., 2021). We measure the total payload transmitted between clients and the server during the unlearning procedure, in megabytes. This includes model broadcasts, client updates, and any auxiliary tensors such as unique-subspace bases $B_{u,v}$ and $B_{u,t}$. Lower values indicate lower bandwidth overhead for deployment.

B.5 Implementation Details

Model architecture. We use frozen pretrained multimodal encoders in bf16 as the backbone. Two types of trainable modules are added in fp32: (1) LoRA adapters inserted into the query and value projections of each attention layer (`q_proj`, `v_proj` for HuggingFace CLIP/SigLIP; `out_proj`, `c_fc` for OpenCLIP), with dropout 0.1; and (2) two-layer MLP projectors (one per modality, ReLU activation, no bias) that map each encoder's output to a shared 256-dimensional embedding space.

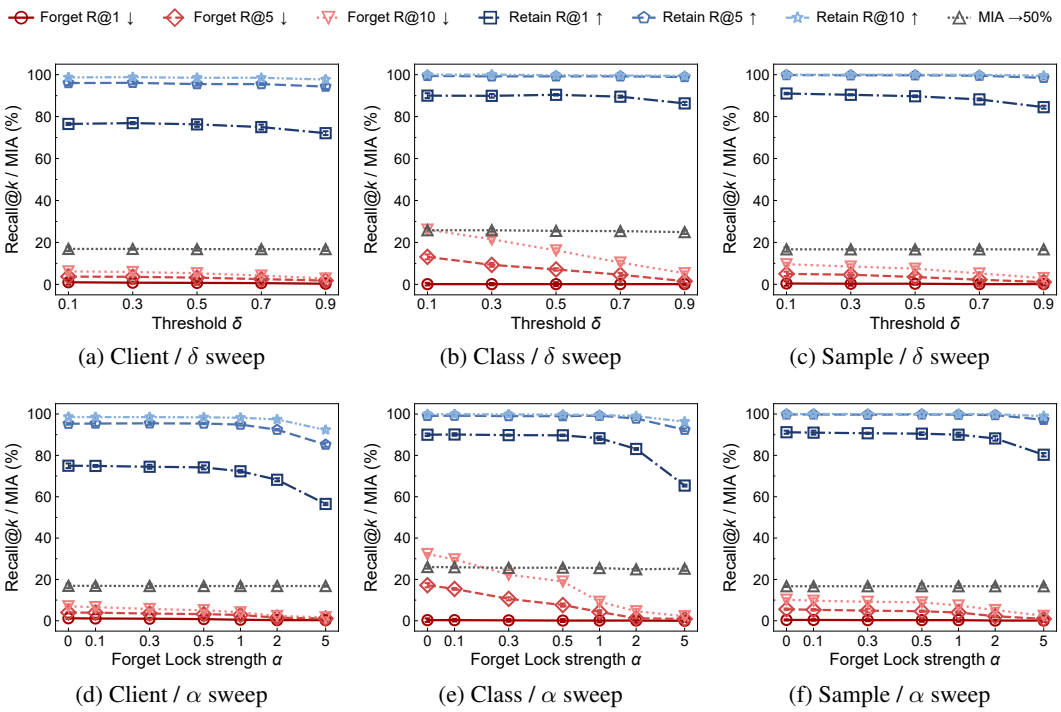
Gradient delta construction. The per-round update $\Delta w_k[t]$ is computed differently for the two module types. For the projectors, we directly compute the parameter difference between the local and global states. For LoRA layers, each adapter contributes a low-rank modification BA to the frozen weight; we compute the effective weight delta as the difference in the low-rank product before and after local training, rather than differencing B and A separately, since only the product has a well-defined interpretation in the full parameter space. These deltas are then column-stacked to form the gradient matrices G_f and G_r (Section 3.3).

Unlearning scenario setup. For client unlearning, a designated client k^* withdraws from the federation; all its historical updates form G_f, and k^* does not participate in subsequent rounds. For sample unlearning, the server specifies a global forget set $\mathcal{D}_f$ whose samples typically span multiple clients (we partition $\mathcal{D}_f$ uniformly across clients at data-splitting time); each client with overlapping samples performs one epoch of training on its local forget subset and uploads the resulting gradient, which is column-stacked into G_f, and every client subsequently continues training on $\mathcal{D}_k \setminus \mathcal{D}_f$. For class unlearning, we construct each annotation-grounded pseudo-class from categorical concepts already provided with the dataset, namely Flickr30K Entities entity categories on Flickr30K and WordNet-aligned concept categories on MS COCO and TextCaps, rather than from unsupervised clustering, so the forget request maps to an interpretable categorical concept at coarser granularity than ImageNet-1K-style fine-grained classes. Concretely, for each dataset we designate one such category as the target concept and instantiate the pseudo-class as the pairs whose captions or scene text reference it, materialized through a small set of stem-level keywords on the caption with light visual filtering to remove off-topic pairs. On Flickr30K (Young et al., 2014), the target concept is the *dog* category, instantiated by captions whose noun phrases include {dog, dogs, puppy, etc.}. On MS COCO Captions (Lin et al., 2014), the target concept is the *kitchen* category, instantiated by captions whose noun phrases include {kitchen, oven, stove, microwave, refrigerator, etc.}. On TextCaps (Sidorov et al., 2020), where each image carries scene text, the target concept is the *shop sign* category, instantiated by captions whose noun phrases include {store, shop, sign, etc.}. Pseudo-class samples are distributed uniformly across clients at data-splitting time. The server then specifies the target pseudo-class by its textual descriptor; every client that holds overlapping samples contributes a per-client gradient to form G_f, and all clients exclude the target-pseudo-class samples from subsequent training.

Training setup. We train with Adam at a constant learning rate η_{lr} of $1\text{e}{-4}$, local batch size 128, and 3 local epochs per federated round. The federation has 50 clients with data distributed via a Dirichlet partition of concentration β at 0.5, and the InfoNCE temperature γ takes value 0.07. The protocol mirrors the natural lifecycle of an unlearning request: 40 rounds of standard FedAvg training (R_1) produce the deployed model w_n; an unlearning request then triggers Bilateral Knowledge Excision (Phase II); 10 subsequent Phase III rounds (R_2) lock in the excision under continued training. Across these stages, the entanglement threshold δ takes value 0.95 for the client and class scenarios and 0.7 for the sample scenario, the SVD energy threshold τ_e takes value 0.99, and the Forget Lock strength α takes value 2.0 during Phase II and 1.0 during Phase III. LoRA adapters use rank 16 with scaling factor 8. For experiments reporting three-seed statistics, the seeds are 42, 1234, and 2024; single-seed diagnostics and transfer studies are identified separately. Experiments are run on $4\times$ NVIDIA H100 GPUs with PyTorch $2.6.0$ (CUDA 12.4) and PEFT $0.18.1$.

B.6 Hyperparameter Sensitivity

We sweep the entanglement threshold δ and the Forget Lock strength α on Flickr30K with CLIP-B/32 across the three unlearning scenarios. Figure 6 compares the resulting retrieval and membership-inference values across configurations.



(a) Client / δ sweep (b) Class / δ sweep (c) Sample / δ sweep

(d) Client / α sweep (e) Class / α sweep (f) Sample / α sweep

Figure 6: Sensitivity of EASE to the entanglement threshold δ (top row) and the Forget Lock strength α (bottom row) on Flickr30K with CLIP-B/32 across three unlearning scenarios.

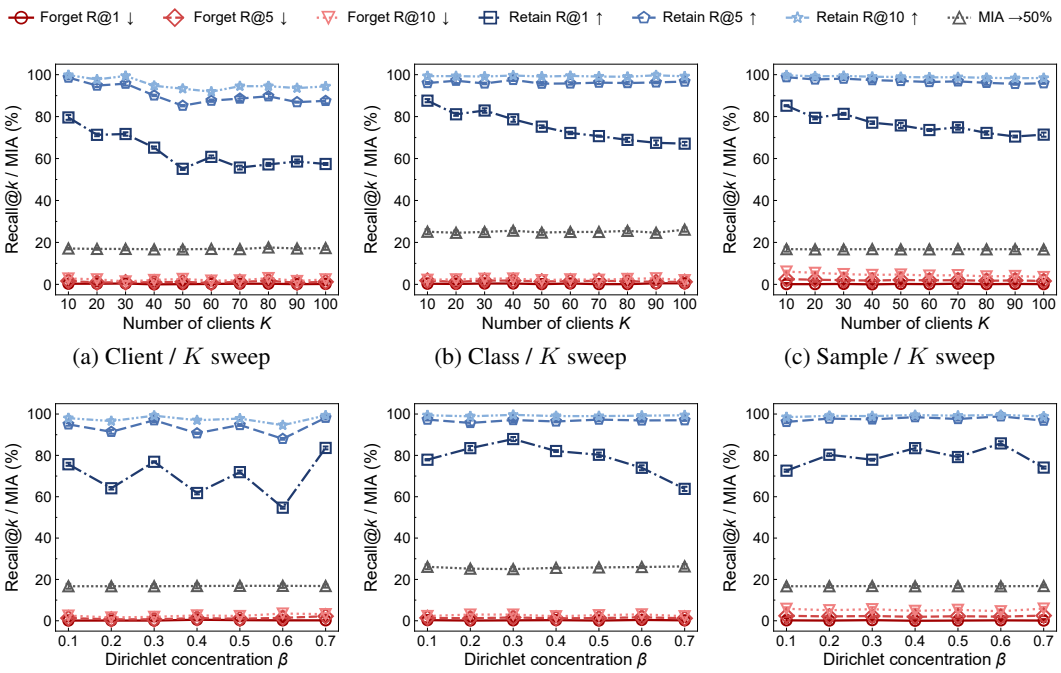
Findings on δ (top row). The δ sweep reduces forget recall as more directions are assigned to the excision subspace, while the retain cost varies by scenario. On CLIP-B/32, client F-R@10 decreases from 6.2 to 2.8 as δ increases from 0.1 to 0.9, and client R-R@1 decreases from 76.5 to 72.1. Class F-R@10 changes from 26.1 to 5.3 over the same range. Enlarging the removable subspace therefore buys most of the forget-side reduction for a retain movement of 4.4 points over the whole range.

Findings on α (bottom row). The Forget Lock strength spans the same spectrum with a single knob. On CLIP-B/32, increasing α from 0 to 5 changes client F-R@10 from 7.0 to 1.7 and client R-R@1 from 75.0 to 56.5; class R-R@1 changes from 90.0 to 65.4. The default α sits in the interior of this range, where forget recall is already low and retain utility is preserved, and it is the value used throughout the comparisons.

Findings on K and β (federated configuration). Forget recall stays low across every displayed K and β setting, so anchor closure holds independently of the federation configuration. Retain retrieval follows the partition instead: in the client β sweep R-R@1 ranges from 54.7 at $\beta = 0.6$ to 83.6 at $\beta = 0.7$, and it declines as the federation grows, the ordinary behaviour of non-IID federated training that applies to every method in the comparison.

Backbone transfer. Figure 8 repeats the δ and α sweeps with CLIP-L/14. The higher-rank forget metrics in the class scenario differ substantially from the client scenario: class F-R@10 is 74.8 at $\delta = 0.1$ and 11.8 at $\delta = 0.9$, and changes from 82.7 at $\alpha = 0$ to 5.6 at $\alpha = 5$. Class F-R@1 is much smaller, ranging from 3.9 to 1.0 in the δ sweep. The class scenario therefore reaches the same rank-10 suppression at a larger threshold than the client scenario, which is what the per-scenario default reflects.

Class scenario asymmetry on CLIP-L/14. The class panels show a gradual reduction in higher-rank forget retrieval. At $\delta = 0.7$, class F-R@5 and F-R@10 remain 13.4 and 32.0; at $\delta = 0.9$ they are 4.9 and 11.8. A larger threshold therefore reduces residual retrieval without making every metric collapse at $\delta = 0.7$. This matches class requests distributing alignment over more directions than



(a) Client / K sweep (b) Class / K sweep (c) Sample / K sweep

(d) Client / β sweep (e) Class / β sweep (f) Sample / β sweep

Figure 7: Sensitivity of EASE to the federation size K (top row) and the Dirichlet concentration β (bottom row) on Flickr30K with CLIP-B/32, evaluated under three unlearning scenarios.

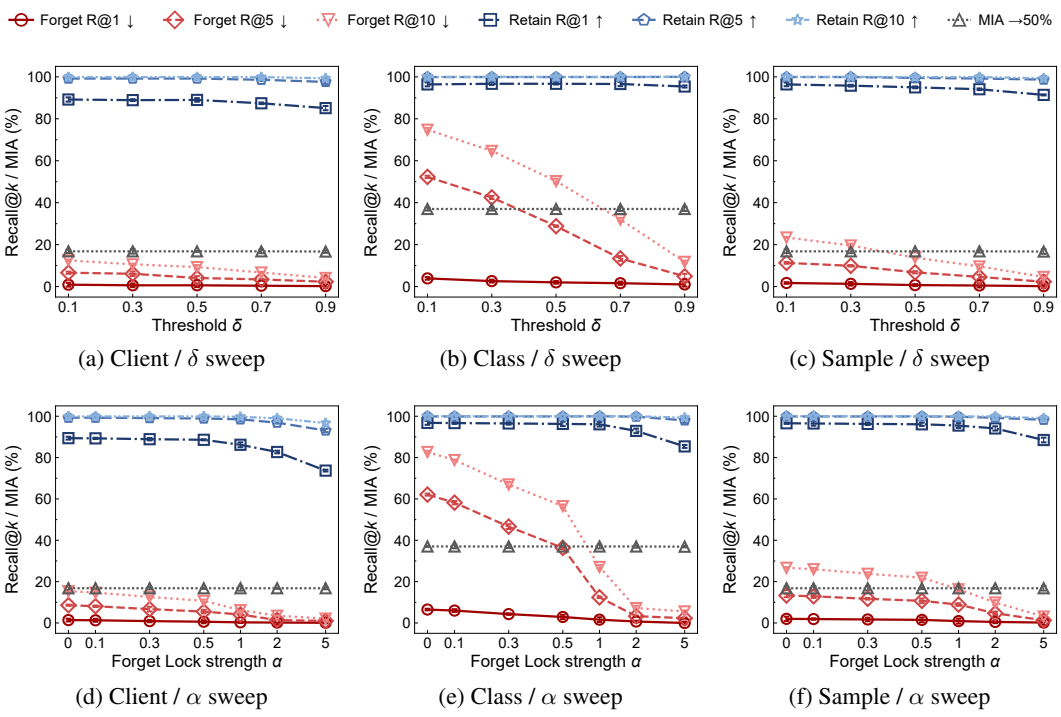
individual client or sample requests, which is why the threshold enters the pipeline as a per-scenario control.

B.7 Component Combination Ablation

Table 2 removes one component at a time and measures each component's marginal necessity: how much the full method loses when that component is disabled. Table 4 extends this to component combinations, marking enabled components with ✓: rows with two enabled components reproduce the leave-one-out variants of Table 2, while rows with a single enabled component measure standalone contribution, which the leave-one-out design cannot expose when components overlap in function. A disabled component is replaced by the same fallback as in Table 2: disabling BKE falls back to single-branch excision (✗† visual-only, ✗‡ text-only), disabling GSD applies the excision to the undecomposed block-level forget subspace, and disabling PFL continues FedAvg without projection or lock. All runs follow the protocol of Table 2: CLIP-B/32 on Flickr30K, three unlearning scenarios, mean over three seeds.

B.8 Robustness of the GSD Decomposition

This section establishes how far the unique subspace $\mathcal{U}$ can be perturbed before the behaviour of EASE changes, stress-testing the estimate from both ends: on the input side we degrade the cached history before GSD runs, and on the output side we deliberately contaminate the identified subspace after GSD runs. All runs use the paper's default δ and α with no retuning, on Flickr30K with CLIP-B/32.



(a) Client / δ sweep (b) Class / δ sweep (c) Sample / δ sweep

(d) Client / α sweep (e) Class / α sweep (f) Sample / α sweep

Figure 8: Sensitivity of EASE to the entanglement threshold δ (top row) and the Forget Lock strength α (bottom row) on Flickr30K with CLIP-L/14 across three unlearning scenarios.

Table 4: Component combination ablation on CLIP-B/32 with Flickr30K across three unlearning scenarios (mean over three seeds). ✓ marks an enabled component; a disabled component uses the fallback of the corresponding *w/o* row in Table 2 (†visual-only, ‡text-only excision). Top: forget-side metrics and MIA; bottom: retain-side metrics and alignment residual ρ. For F-$R@k$, R-$R@k$, ρ, parentheticals give absolute gap to retrain with (↑green) closer and (↓red) farther; for MIA, (val) is $|\text{MIA} - 50|$, with smaller distance ranked better and ties sharing a rank. Best in **red**, runner-up in blue among variants. Two-component rows and the full method repeat Table 2 for reference.

Components			Client Unlearning				Class Unlearning				Sample Unlearning			
BKE	GSD	PFL	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
✓	✗	✗	0.5 (↓0.4)	1.1 (↑0.4)	2.7 (↓0.4)	17.6 (32.4)	0.4 (↓0.2)	1.8 (↓1.1)	3.0 (↓1.6)	25.8 (24.2)	0.2 (↓0.1)	2.1 (↓1.0)	5.0 (↓2.6)	17.3 (32.7)
✗†	✓	✗	8.9 (↓8.8)	19.1 (↓17.6)	29.8 (↓27.5)	**25.1** (24.9)	7.4 (↓7.2)	17.1 (↓16.4)	25.4 (↓24.0)	**31.7** (18.3)	5.9 (↓5.8)	14.6 (↓13.5)	24.7 (↓22.3)	**22.4** (27.6)
✗†	✗	✓	3.6 (↓3.5)	8.6 (↓7.1)	16.1 (↓13.8)	20.6 (29.4)	3.1 (↓2.9)	7.2 (↓6.5)	13.9 (↓12.5)	29.1 (20.9)	2.4 (↓2.3)	7.1 (↓6.0)	12.6 (↓10.2)	19.3 (30.7)
✗†	✓	✓	1.3 (↓1.2)	2.7 (↓1.2)	3.9 (↓1.6)	18.4 (31.6)	1.0 (↓0.8)	2.4 (↓1.7)	3.8 (↓2.4)	26.3 (23.7)	0.8 (↓0.7)	3.0 (↓1.9)	6.1 (↓3.7)	17.6 (32.4)
✗‡	✓	✓	1.8 (↓1.7)	3.4 (↓1.9)	5.0 (↓2.7)	19.1 (30.9)	1.6 (↓1.4)	2.9 (↓2.2)	4.6 (↓3.2)	26.8 (23.2)	1.1 (↓1.0)	3.5 (↓2.4)	6.7 (↓4.3)	18.0 (32.0)
✓	✗	✓	**0.2** (↓0.1)	**0.8** (↑0.7)	**2.0** (↑0.3)	17.2 (32.8)	**0.2**	**1.1** (↓0.4)	**1.8** (↓0.4)	24.6 (25.4)	**0.0** (↓0.1)	**1.6** (↓0.5)	**4.7** (↓2.3)	17.1 (32.9)
✓	✓	✗	0.7 (↓0.6)	1.9 (↓0.4)	3.5 (↓1.2)	17.0 (33.0)	0.5 (↓0.3)	2.0 (↓1.3)	3.5 (↓2.1)	25.9 (24.1)	0.4 (↓0.3)	2.6 (↓1.5)	6.2 (↓3.8)	17.0 (33.0)
✓	✓	✓	0.4 (↓0.3)	1.2 (↓0.3)	2.1 (↑0.2)	16.8 (33.2)	0.3 (↓0.1)	1.4 (↓0.7)	2.7 (↓1.3)	25.1 (24.9)	0.1	2.0 (↓0.9)	5.3 (↓2.9)	16.8 (33.2)
Retrain			0.1	1.5	2.3	16.7	0.2	0.7	1.4	24.2	0.1	1.1	2.4	16.7
BKE	**GSD**	**PFL**	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*ρ* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*ρ* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*ρ* ↓
✓	✗	✗	9.2 (↓81.5)	24.3 (↓75.6)	45.6 (↓54.4)	0.17 (↓0.17)	3.9 (↓85.5)	14.8 (↓85.1)	35.2 (↓64.8)	0.19 (↓0.19)	5.1 (↓75.5)	20.9 (↓76.8)	42.3 (↓57.1)	0.18 (↓0.18)
✗†	✓	✗	82.9 (↓7.8)	97.1 (↓2.8)	98.5 (↓1.5)	0.47 (↓0.47)	77.5 (↓11.9)	94.4 (↓5.5)	97.7 (↓2.3)	0.42 (↓0.42)	77.0 (↓3.6)	96.1 (↓1.6)	98.1 (↓1.3)	0.40 (↓0.40)
✗†	✗	✓	28.9 (↓61.8)	55.8 (↓44.1)	73.2 (↓26.8)	0.32 (↓0.32)	24.4 (↓65.0)	52.1 (↓47.8)	69.8 (↓30.2)	0.30 (↓0.30)	30.8 (↓49.8)	60.9 (↓36.8)	76.5 (↓22.9)	0.28 (↓0.28)
✗†	✓	✓	83.4 (↓7.3)	97.9 (↓2.0)	99.1 (↓0.9)	0.23 (↓0.23)	78.2 (↓11.2)	95.3 (↓4.6)	98.5 (↓1.5)	0.21 (↓0.21)	77.8 (↓2.8)	96.9 (↓0.8)	98.7 (↓0.7)	0.19 (↓0.19)
✗‡	✓	✓	81.8 (↓8.9)	97.6 (↓2.3)	98.7 (↓1.3)	0.29 (↓0.29)	77.6 (↓11.8)	94.5 (↓5.4)	98.1 (↓1.9)	0.26 (↓0.26)	76.9 (↓3.7)	96.5 (↓1.2)	98.4 (↓1.0)	0.23 (↓0.23)
✓	✗	✓	12.8 (↓77.9)	31.1 (↓68.8)	52.6 (↓47.4)	0.10 (↓0.10)	5.2 (↓84.2)	18.4 (↓81.5)	39.8 (↓60.2)	0.12 (↓0.12)	7.1 (↓73.5)	25.3 (↓72.4)	46.9 (↓52.5)	0.11 (↓0.11)
✓	✓	✗	**88.1** (↓2.6)	**99.3** (↓0.6)	99.7 (↓0.3)	0.15 (↓0.15)	**82.8** (↓6.6)	**97.1** (↓2.8)	99.2 (↓0.8)	0.13 (↓0.13)	**80.2** (↓0.4)	97.6 (↓0.1)	99.0 (↓0.4)	0.12 (↓0.12)
✓	✓	✓	86.3 (↓4.4)	99.1 (↓0.8)	**99.8** (↓0.2)	**0.08** (↓0.08)	80.6 (↓8.8)	96.7 (↓3.2)	**99.2** (↓0.8)	**0.07** (↓0.07)	79.5 (↓1.1)	**97.7**	**99.2** (↓0.2)	**0.06** (↓0.06)
Retrain			90.7	99.9	100.0	0	89.4	99.9	100.0	0	80.6	97.7	99.4	0

History degradation. We degrade the cached update history along four axes: truncation to the last 20, 10, 5, or 1 rounds; round subsampling at 50%, 25%, or 10%; per-client sketches of rank 16, 8, or 4; and relative Gaussian noise at 5%, 10%, or 20%. Table 5 reports the principal-angle deviation from the full-history subspace $\mathcal{U}_0$ and the retain-side change in all three scenarios. Its full-history reference follows the full-method row of Table 2; the same reference is used in Table 6 and Figure 10b. For the W=20, p=50%, and 5%-noise settings, the deviation stays below five degrees and the retain drop is at most 0.7 R-R@1. More severe degradation incurs larger costs. The worst listed condition, 20% relative noise, deviates 21.7 degrees and costs 4.7, 3.7, and 3.7 R-R@1 under client, class, and sample unlearning. The rank-8 sketch uses 20% of the storage while keeping the retain drop within 1.7 R-R@1 across scenarios.

Table 5: GSD under degraded update histories on Flickr30K with CLIP-B/32. $\angle(\mathcal{U}, \mathcal{U}_0)$ is the deviation from the full-history subspace; parentheses give the change from the full-history run.

Axis	Setting	Storage	$\angle(\mathcal{U}, \mathcal{U}_0)$ (deg)	R-R@1 ↑ Client	Class	Sample
Full history (40 rounds)		100%	0.0	86.3	80.6	79.5
Window	W=20	50%	2.7	85.9 (↓0.4)	80.3 (↓0.3)	79.1 (↓0.4)
	W=10	25%	4.3	85.8 (↓0.5)	80.0 (↓0.6)	78.9 (↓0.6)
	W=5	12.5%	6.1	85.6 (↓0.7)	79.6 (↓1.0)	78.6 (↓0.9)
	W=1 (final)	2.5%	8.7	85.5 (↓0.8)	79.3 (↓1.3)	78.4 (↓1.1)
Subsample	p=50%	50%	2.5	86.0 (↓0.3)	80.2 (↓0.4)	79.0 (↓0.5)
	p=25%	25%	7.2	84.8 (↓1.5)	79.1 (↓1.5)	78.2 (↓1.3)
	p=10%	10%	13.6	83.9 (↓2.4)	77.7 (↓2.9)	76.8 (↓2.7)
Sketch	r=16	40%	3.9	85.7 (↓0.6)	80.0 (↓0.6)	78.7 (↓0.8)
	r=8	20%	6.8	84.7 (↓1.6)	78.9 (↓1.7)	78.2 (↓1.3)
	r=4	10%	12.1	83.8 (↓2.5)	77.4 (↓3.2)	76.9 (↓2.6)
Noise	σ=5%	100%	4.1	85.9 (↓0.4)	80.2 (↓0.4)	78.8 (↓0.7)
	σ=10%	100%	10.9	84.1 (↓2.2)	79.0 (↓1.6)	77.9 (↓1.6)
	σ=20%	100%	21.7	81.6 (↓4.7)	76.9 (↓3.7)	75.8 (↓3.7)
Retrain		–	–	90.7	89.4	80.6

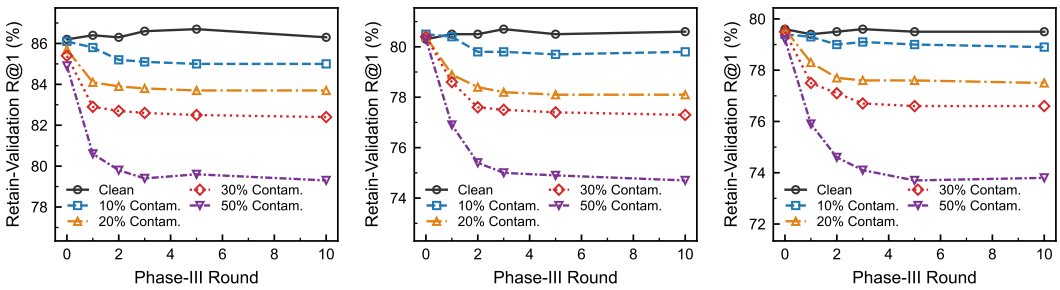
Table 6: Stability of $\mathcal{U}$ on Flickr30K with CLIP-B/32. Top: mean pairwise angle across three optimization seeds with fixed request, partition, and initialization; $\dim \mathcal{U}$ gives the range across scenarios, modalities, and seeds. N/A denotes an empty-subspace comparison. † marks the operating thresholds. Bottom: retain R@1 after ten contamination rounds; parentheses give the drop from the clean run.

Sweep	Value	$\dim \mathcal{U}$	$\angle_{\text{seed}}(\mathcal{U})$ (deg) ↓ Client	Class	Sample
Threshold δ	0.1	0–41	*N/A*	43.84	61.83
	0.3	5–50	67.26	40.73	56.56
	0.5	5–50	64.07	40.52	56.40
	0.7†	5–50	64.07	40.52	56.40
	0.9	5–50	64.07	40.52	56.40
	0.95†	5–50	64.07	40.52	56.40
Dirichlet β	0.1	7–50	65.32	42.62	56.35
	0.3	8–50	67.29	41.16	56.82
	0.5	5–50	64.07	40.52	56.40
	0.7	8–50	71.08	39.92	56.75

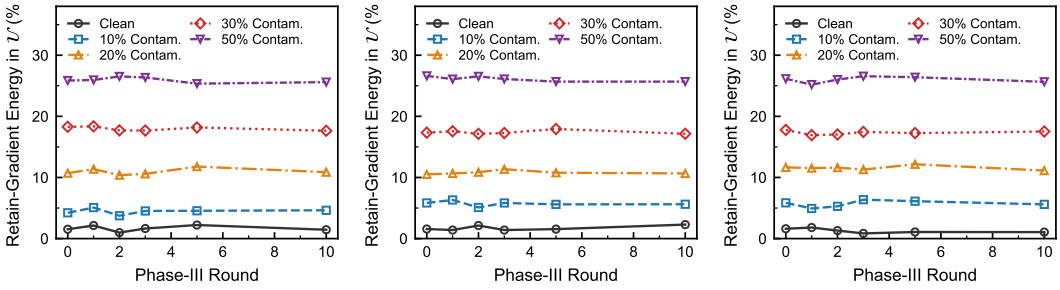
Injected rate	Onset round	R-R@1 ↑ (round 10) Client	Class	Sample
Clean	–	86.3	80.6	79.5
10%	2	85.0 (↓1.3)	79.8 (↓0.8)	78.9 (↓0.6)
20%	1	83.7 (↓2.6)	78.1 (↓2.5)	77.5 (↓2.0)
30%	1	82.4 (↓3.9)	77.3 (↓3.3)	76.6 (↓2.9)
50%	1	79.3 (↓7.0)	74.7 (↓5.9)	73.8 (↓5.7)

Unique-subspace contamination. The CS decomposition assigns every direction of the forget subspace a principal-angle cosine to the retain subspace and classifies a direction as entangled when this cosine exceeds δ. We rank the entangled directions by this cosine and move the top-ranked ones into $\mathcal{U}$ until they account for 10/20/30/50% of its dimension; the enlarged $\mathcal{U}$ then passes through the remaining pipeline. This stress condition targets directions with high retain-side overlap. The round-10 retain cost grows with the injected rate: client drops are 1.3/2.6/3.9/7.0 R-R@1, class drops are 0.8/2.5/3.3/5.9, and sample drops are 0.6/2.0/2.9/5.7, as reported in Table 6 (bottom). Figure 9 tracks the corresponding retain-validation trajectories. The listed onset is round 2 at 10% contamination and round 1 at 20%, 30%, and 50%. The curves approach their endpoints after the early decline, with smaller round-to-round fluctuations thereafter. The bottom row tracks the share of retain-gradient energy inside $\mathcal{U}$ on retain-validation batches. This share remains low in the clean condition and higher at larger contamination rates throughout the displayed rounds. Together, the two rows associate stronger overlap with the penalized directions with a larger retain-side cost; they do not report forget-side or MIA changes.

Client heterogeneity. The Dirichlet β sweep in Figure 7 shows that forget recall stays low across the whole heterogeneity range, with retain R@1 tracking the difficulty of the partition (54.7–83.6 on client) as non-IID data ordinarily dictates in federated training. Table 6 (top) reports the geometry behind that behaviour. The subspace the pipeline selects is identical for every threshold at or above $\delta = 0.5$, so the operating value $\delta = 0.95$ sits inside a flat region rather than on a tuned point, and across the heterogeneity sweep the class and sample angles move by at most 2.7 and 0.5 degrees. The dimension range 5–50 reflects that $\mathcal{U}$ is sized by the request rather than fixed in advance.



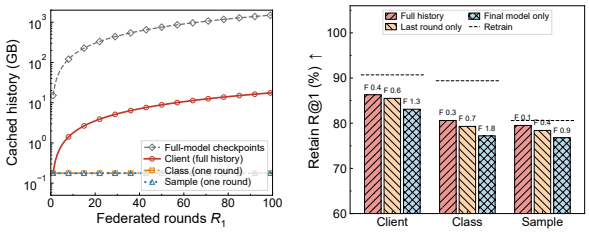
(a) Client unlearning (b) Class unlearning (c) Sample unlearning



(d) Client, energy in $\mathcal{U}$ (e) Class, energy in $\mathcal{U}$ (f) Sample, energy in $\mathcal{U}$

Figure 9: Phase-III rounds under contamination of $\mathcal{U}$ on Flickr30K with CLIP-B/32. Top: server-side retain-validation R@1. Bottom: share of the retain-gradient energy that falls inside $\mathcal{U}$, the quantity the Forget Lock suppresses.

History storage and the final-model fallback. The cached history stores only the LoRA and projector deltas: 1.77M parameters per client per round, or 3.54 MB in fp16. Using decimal MB and GB, 40 rounds at $K{=}50$ occupy 7.08 GB, about 1.2% of the 605 GB required by 302.5 MB full-model checkpoints per client-round (Figure 10a). Only client unlearning consumes the full history; class and sample unlearning read a single request-specific round, so their footprint stays flat in R_1. Figure 10b uses the full-method reference of Table 2. For client unlearning, the last-round fallback changes retain R@1 from 86.3 to 85.5 and forget R@1 from 0.4 to 0.6: a 0.8 decrease and a 0.2 increase, respectively. Its class and sample retain drops are 1.3 and 1.1, matching the $W{=}1$ row of Table 5. The final-model-only variant rebuilds the forget subspace from gradients probed at the deployed model.



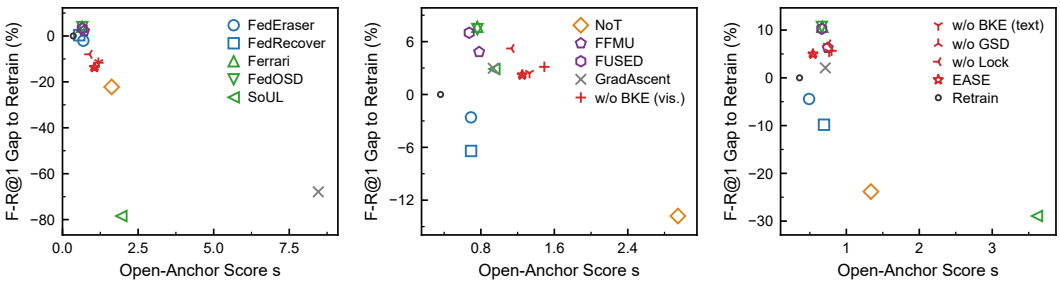
(a) Cached-history volume. (b) Fallback retain R@1.

Figure 10: Storage footprint of the cached update history and the cost of running without it, Flickr30K with CLIP-B/32, K of 50 clients.

B.9 Anchor Diagnostics on Baselines

The Anchor Principle makes a prediction about methods it was not designed around: a method's forget-side gap to the retrain reference should be governed by the reconstruction channels it leaves open, whichever method it is. The diagnostic compares nine baselines and four ablation variants, measuring one signal per anchor. The Modality Anchor signal is the alignment residual ρ of Section 3.2. The Unique-Subspace Anchor signal r_u is the fraction of the forget-side displacement from w_0 that still lies inside $\mathcal{U}$ after unlearning,

$$r_u := \frac{\|\Pi_u(w - w_0)\|_2}{\|\Pi_f(w - w_0)\|_2}, \tag{13}$$



(a) Client unlearning (b) Class unlearning (c) Sample unlearning

Figure 11: Anchor diagnostics on Flickr30K with CLIP-B/32, seed 42; legend entries are shared across panels. Each panel shows the raw score s and signed F-R@1 difference to retrain for nine baselines, four ablations, EASE, and Retrain. Scores are not clipped or calibrated. Empty entangled subspaces make EASE and w/o GSD coincide.

where Π_f projects onto the forget subspace $\mathcal{S}_f$ and $\mathcal{U}$ is the subspace GSD identifies from the same history for every method. The Drift Anchor signal d_u is the displacement along $\mathcal{U}$ accumulated over the R_2 stabilization rounds, normalized by the original forget displacement $\|\Pi_u(w_n - w_0)\|_2$. We use the raw signal values, whose mean is the open-anchor score s; the score is not constrained to $[0, 1]$. Figure 11 plots s against the F-R@1 gap to retrain for all methods in the three scenarios on Flickr30K with CLIP-B/32; these diagnostic checkpoints are separate from those underlying Table 1. The diagnostic examines whether these raw signals covary with the retrieval gap. The relation is scenario dependent: the class panel does not exhibit a simple monotone ordering across methods, and the signed gap distinguishes higher from lower forget recall relative to retrain. These panels therefore do not by themselves establish a universal ranking of methods by anchor score.

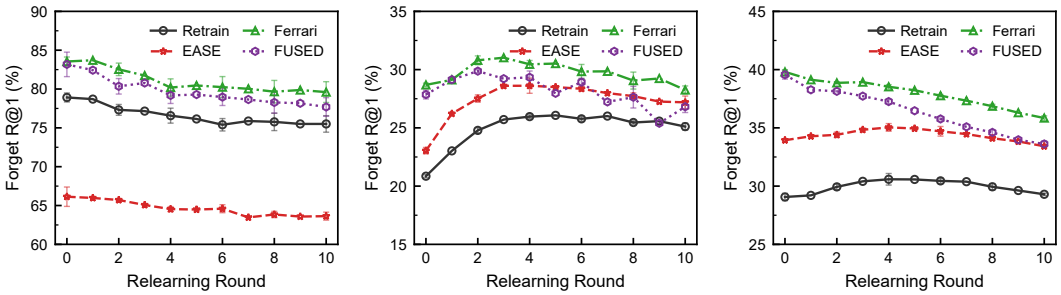
B.10 Relearning Resistance

A relearning diagnostic examines how forget-set retrieval changes after further exposure to forget data. Starting from each method's unlearned checkpoint, an attacker holding 5% of $\mathcal{D}_f$ fine-tunes for ten local rounds at the Phase-III learning rate with the Forget Lock disabled. We evaluate F-R@1 using the complete forget-set gallery after each round. Figure 12 compares EASE, Retrain, Ferrari, and FUSED on Flickr30K with CLIP-B/32. The trajectories differ across scenarios: client EASE changes from 66.13 to 63.63, while Retrain changes from 78.91 to 75.51; class EASE rises from 23.04 to 27.19, alongside Retrain's 20.85 to 25.11; sample EASE changes from 33.94 to 33.42, while Retrain remains near 29. In all three scenarios the EASE trajectory moves in the same direction and by a comparable amount as the retrain reference: -2.50 against Retrain's -3.40 on client, $+4.15$ against $+4.26$ on class, and -0.52 against a flat Retrain curve near 29 on sample. Ten rounds of attacker fine-tuning with the lock disabled therefore reproduce the retrain reference's own response to the same exposure rather than reopening the forget set.

B.11 Visualization of Unlearning Effect

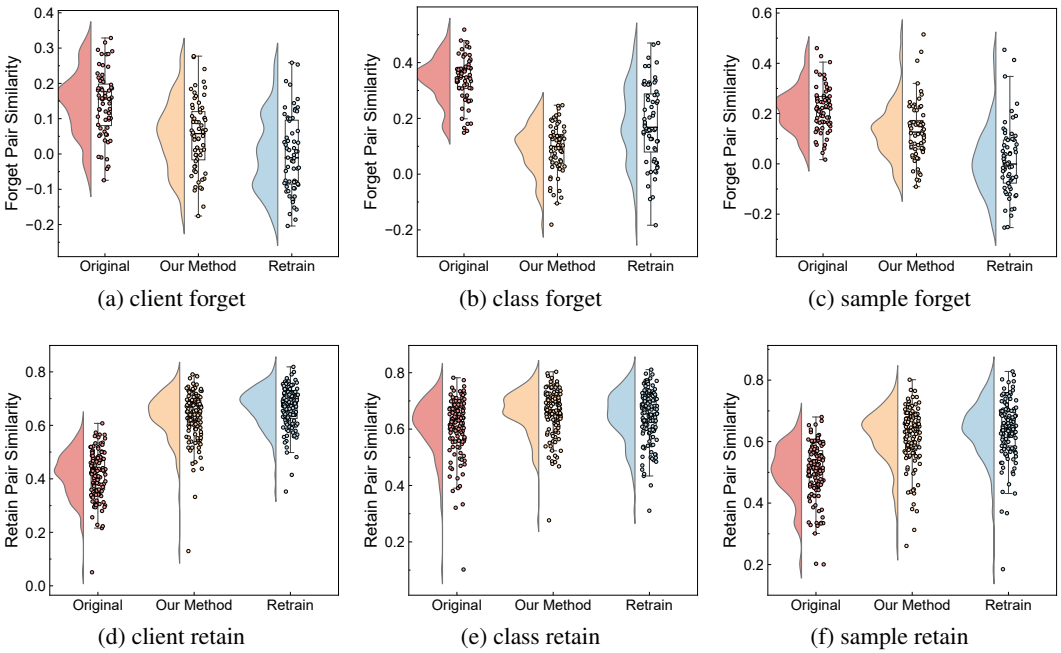
We visualize per-pair cosine similarity distributions under the original model w_n, our method w^*, and the retrain reference $\tilde{w}$ in Figure 13. For each image-text pair, we compute $\cos(z_v^i, z_t^i)$ between the visual and textual embeddings on the shared projector output; higher values indicate stronger cross-modal alignment, and an unlearned model is expected to produce similarities close to those of the retrain reference on forget pairs and close to the original on retain pairs. We report these per-pair similarities for three unlearning scenarios and two data splits. Each panel combines a half-violin, a strip plot, and a box plot, so the reader sees the overall shape, the individual pair values, and the median with quartiles in a single view.

Two patterns emerge. First, on forget pairs, the original model places most pairs at high similarity, reflecting the alignment learned during standard federated training; our method shifts this distribution leftward toward the retrain reference across client, class, and sample scenarios, meaning that the cross-modal alignment between forget image-text pairs is reduced toward the level of a model trained



(a) Client unlearning (b) Class unlearning (c) Sample unlearning

Figure 12: Forget R@1 under ten rounds of relearning on 5% of the forget set, Flickr30K with CLIP-B/32. The complete forget-set gallery is evaluated at each round. Curves and error bars show the mean and sample standard deviation over three seeds. This diagnostic uses separate checkpoints trained for the relearning experiment.



(a) client forget (b) class forget (c) sample forget

(d) client retain (e) class retain (f) sample retain

Figure 13: Per-pair image–text cosine similarity under the original, our method, and retrain. Top: forget pairs, where our method shifts toward the retrain reference. Bottom: retain pairs, where the three distributions overlap.

from scratch without those pairs. Second, on retain pairs, the distribution under our method remains close to the retrain reference, indicating that bilateral projection and the Forget Lock preserve much of the unrelated knowledge. These two patterns provide a direct visual counterpart to the quantitative results in Section 4.2 and Section 4.3: the Modality Anchor closure moves forget pairs toward retrain levels, and the direction-selective Forget Lock keeps retain pairs close to the retrain reference.

B.12 Alternative Lock Designs

This section justifies the direction-selective Forget Lock used in PFL (Section 3.4) by ruling out two natural alternatives. **(i) Per-round server projection alone.** Re-projecting the global model onto the affine submanifold $\mathcal{M}_\mu = w_{0,\mu} + \mathcal{U}_\mu^\perp$ at every round closes $d(t)$ at the start of each round but leaves the within-round trajectory unconstrained: clients still drift along $\mathcal{U}_\mu$ during their local SGD,

and the drift is recovered by aggregation only after the damage is done. **(ii) Parameter-importance regularizers (EWC, SI, MAS).** These penalize drift on axis-aligned parameter coordinates weighted by importance scores. The unique subspace $\mathcal{U}_\mu$ is generally an oblique subspace of parameter space, not aligned with any coordinate axis, so an axis-aligned penalty cannot express the constraint that drift is bounded only along $\mathcal{U}_\mu$: tuning the penalty large enough to suppress $g_u(t)$ over-constrains $\mathcal{U}_\mu^\perp$ and crushes retain support, while tuning it small leaves $g_u(t)$ unbounded. PFL's quadratic penalty $\mathcal{L}_f$ is supported on $\mathcal{U}_\mu$ and therefore enforces the asymmetry that neither alternative can express.

B.13 Algorithm

The three unlearning scenarios share the pipeline in Algorithm 1; they differ only in how G_f, G_r are constructed (Section 3.3) and in the participant and data configurations across phases (Section 3.4).

Algorithm 1 EASE: **E**ntanglement-**A**ware **S**ubspace **E**xcision

Input: Deployed model w_n, anchor reference w_0, client gradient histories $\{H_k\}_{k=1}^K$, forget specification, thresholds δ, τ_e, lock strength α, stabilization rounds R_2.

▷ *Phase I: Gradient Subspace Decomposition*
Construct G_f, G_r from $\{H_k\}$ based on unlearning scenario ▷ Eq. (5)
for each modality $\mu \in \{v, t\}$ **do**
  $\Phi_{f,\mu} \leftarrow$ SVD of $G_{f,\mu}$, retain directions covering τ_e energy ▷ Eq. (6)
  $\Phi_{r,\mu} \leftarrow$ SVD of $G_{r,\mu}$, retain directions covering τ_e energy
  Compute $M_\mu \leftarrow \Phi_{f,\mu}^\top \Phi_{r,\mu}$; SVD to get principal angles $\{\varphi_i\}$ ▷ Eq. (7)
  Partition into removable anchor and retain support: $B_{u,\mu} \leftarrow [c_i]_{\cos\varphi_i \le \delta}$; $\Pi_{u,\mu} \leftarrow B_{u,\mu} B_{u,\mu}^\top$ ▷ Eq. (8)
end for
▷ *Phase II: Bilateral Knowledge Excision*
$w_g \leftarrow w_n$ ▷ Start from the deployed model
for $\mu \in \{v, t\}$ **do**
  $w_{g,\mu} \leftarrow w_{g,\mu} - \Pi_{u,\mu}(w_{g,\mu} - w_{0,\mu})$ ▷ Remove anchor component, Eq. (10)
end for
▷ *Phase III: Stabilization under Forget Lock*
for round $t = 1, 2, \ldots, R_2$ **do**
  Broadcast w_g, unique bases $\{B_{u,\mu}\}$, anchor reference w_0 to clients
  for each client k **in parallel do**
    Compute alignment loss $\mathcal{L}_a(w; \mathcal{D}'_k)$ on retain data ▷ Eq. (1)
    Compute Forget Lock penalty $\mathcal{L}_f(w) = \sum_\mu \|B_{u,\mu}^\top (w_\mu - w_{0,\mu})\|_2^2$ ▷ Eq. (11)
    Update w_k via local optimizer step on $\mathcal{L}_a + \alpha \cdot \mathcal{L}_f$ ▷ Lock suppresses Temporal Re-anchoring, Eq. (12)
    Upload w_k to server
  end for
  $w_g \leftarrow \frac{1}{K} \sum_{k=1}^K w_k$ ▷ Server aggregation
end for
return Unlearned model $w^* \leftarrow w_g$

B.14 Transfer to VLM and MLLM Backbones

The main experiments use dual-encoder backbones whose cross-modal interface is a single bilinear similarity. We probe transfer to four attention-fusion architectures, two fusion-encoder VLMs (BLIP (Li et al., 2022), BLIP-2 (Li et al., 2023)) and two MLLMs (Qwen2.5-VL-7B (Bai et al., 2025), LLaVA-1.5-7B (Liu et al., 2024)), so that each class contributes two structurally distinct interfaces. In every case BKE excises both sides of the cross-modal interface: on BLIP the ITC head and the text branch's cross-attention, on BLIP-2 the querying transformer's cross-attention and its text branch with the ITC/ITM heads, and on the two MLLMs the LoRA emitting the projected image tokens and the LLM attention LoRA consuming them. ρ is read from the ITM logit on the VLMs and as the length-normalized log-likelihood of the paired caption on the MLLMs; GSD and PFL apply unchanged. The method-comparison studies cover client, class, and sample unlearning on Flickr30K

with frozen backbones against the baselines of Table 1, scored against each architecture's own retrain reference. BLIP and BLIP-2 keep the budget of 50 clients and 40 rounds, while the MLLMs use 10 clients and 15 rounds with retrieval scored by conditional log-likelihood ranking and retain-set captioning by CIDEr (Vedantam et al., 2015). All transfer studies are single-seed runs. The method comparison is reported with the full comparison results in Appendix F (Tables 18–21); Table 7 repeats the client-unlearning component ablation of Table 2 under the same transfer protocol.

Table 7: Component ablation under the transfer protocol: client unlearning on Flickr30K with frozen backbones (single seed). Parentheticals give absolute gap to each architecture's own retrain reference with (↑green) closer and (↓red) farther. Best in **red**, runner-up in blue among variants of the same architecture. ✓ enabled, ✗ disabled; †visual-only, ‡text-only excision. CIDEr is reported for all backbones on the same captioning evaluation pipeline.

BKE	GSD	PFL	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	ρ ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*CIDEr* ↑
BLIP-ITM										
✓	✓	✓	1.4 (↓1.2)	2.5 (↓1.4)	3.6 (↓1.4)	0.18 (↓0.14)	86.7 (↓0.7)	98.7 (↓0.5)	99.3 (↓0.5)	84.6 (↓1.1)
✗†	✓	✓	2.7 (↓2.5)	3.8 (↓2.7)	5.4 (↓3.2)	0.32 (↓0.28)	84.5 (↓2.9)	98.1 (↓1.1)	98.5 (↓1.3)	82.9 (↓2.8)
✗‡	✓	✓	3.3 (↓3.1)	4.6 (↓3.5)	6.1 (↓3.9)	0.37 (↓0.33)	83.7 (↓3.7)	97.5 (↓1.7)	98.1 (↓1.7)	81.6 (↓4.1)
✓	✗	✓	**0.5** (↓0.3)	**1.4** (↓0.3)	**2.5** (↓0.3)	**0.15** (↓0.11)	18.9 (↓68.5)	45.3 (↓53.9)	62.7 (↓37.1)	58.8 (↓26.9)
✓	✓	✗	1.9 (↓1.7)	2.9 (↓1.8)	4.4 (↓2.2)	0.26 (↓0.22)	**87.0** (↓0.4)	**99.0** (↓0.2)	**99.4** (↓0.4)	**85.3** (↓0.4)
Retrain			0.2	1.1	2.2	0.04	87.4	99.2	99.8	85.7
BLIP-2										
✓	✓	✓	3.6 (↓2.2)	4.6 (↓2.1)	5.0 (↓1.6)	0.27 (↓0.19)	83.2 (↓2.6)	97.4 (↓0.9)	98.4 (↓0.7)	82.4 (↓4.2)
✗†	✓	✓	4.9 (↓3.5)	5.8 (↓3.3)	6.9 (↓3.5)	0.40 (↓0.32)	81.6 (↓4.2)	96.3 (↓2.0)	97.6 (↓1.5)	79.8 (↓6.8)
✗‡	✓	✓	5.8 (↓4.4)	6.7 (↓4.2)	7.5 (↓4.1)	0.45 (↓0.37)	80.7 (↓5.1)	95.9 (↓2.4)	97.0 (↓2.1)	78.9 (↓7.7)
✓	✗	✓	**1.8** (↓0.4)	**2.8** (↓0.3)	**3.8** (↓0.4)	**0.22** (↓0.14)	16.1 (↓69.7)	43.2 (↓55.1)	59.1 (↓40.0)	57.4 (↓29.2)
✓	✓	✗	4.1 (↓2.7)	4.9 (↓2.4)	5.7 (↓2.3)	0.33 (↓0.25)	**84.5** (↓1.3)	**97.9** (↓0.4)	**98.6** (↓0.5)	**84.6** (↓2.0)
Retrain			1.4	2.5	3.4	0.08	85.8	98.3	99.1	86.6
Qwen2.5-VL-7B										
✓	✓	✓	7.9 (↓2.2)	9.7 (↓2.1)	10.9 (↓2.5)	0.43 (↓0.12)	77.8 (↓1.7)	92.0 (↓0.8)	93.0 (↓1.2)	52.4 (↓10.0)
✗†	✓	✓	10.2 (↓4.5)	12.7 (↓5.1)	13.1 (↓4.7)	0.56 (↓0.25)	75.0 (↓4.5)	90.7 (↓2.1)	92.1 (↓2.1)	49.5 (↓12.9)
✗‡	✓	✓	12.4 (↓6.7)	14.1 (↓6.5)	15.3 (↓6.9)	0.62 (↓0.31)	73.7 (↓5.8)	90.0 (↓2.8)	91.2 (↓3.0)	48.3 (↓14.1)
✓	✗	✓	**6.1** (↓0.4)	**8.1** (↓0.5)	**8.8** (↓0.4)	**0.40** (↓0.09)	13.3 (↓66.2)	34.0 (↓58.8)	55.1 (↓39.1)	28.9 (↓33.5)
✓	✓	✗	8.6 (↓2.9)	10.3 (↓2.7)	11.8 (↓3.4)	0.49 (↓0.18)	**78.3** (↓1.2)	**92.1** (↓0.7)	**93.6** (↓0.6)	**53.7** (↓8.7)
Retrain			5.7	7.6	8.4	0.31	79.5	92.8	94.2	62.4
LLaVA-1.5-7B										
✓	✓	✓	10.4 (↓2.6)	12.1 (↓3.0)	12.8 (↓2.3)	0.48 (↓0.14)	75.5 (↓2.4)	89.8 (↓1.0)	91.2 (↓0.9)	49.5 (↓9.7)
✗†	✓	✓	13.2 (↓5.4)	14.9 (↓5.8)	16.0 (↓5.5)	0.61 (↓0.27)	72.6 (↓5.3)	87.5 (↓3.3)	89.4 (↓2.7)	46.5 (↓12.7)
✗‡	✓	✓	14.7 (↓6.9)	16.8 (↓7.7)	17.4 (↓6.9)	0.67 (↓0.33)	71.3 (↓6.6)	86.7 (↓4.1)	88.8 (↓3.3)	45.1 (↓14.1)
✓	✗	✓	**8.4** (↓0.6)	**9.9** (↓0.8)	**11.1** (↓0.6)	**0.44** (↓0.10)	11.2 (↓66.7)	31.4 (↓59.4)	50.9 (↓41.2)	24.7 (↓34.5)
✓	✓	✗	11.0 (↓3.2)	12.9 (↓3.8)	13.7 (↓3.2)	0.53 (↓0.19)	**76.4** (↓1.5)	**90.1** (↓0.7)	**91.6** (↓0.5)	**50.6** (↓8.6)
Retrain			7.8	9.1	10.5	0.34	77.9	90.8	92.1	59.2

Table 7 shows each component carrying the same duty on the transfer architectures as on CLIP. GSD is what prevents over-erasure: the undecomposed forget subspace can drive F-$R@k$ and ρ close to the retrain reference, but it also erases retain support, collapsing R-$R@1$ to $18.9/16.1$ on the VLMs and $13.3/11.2$ on the MLLMs. Single-branch BKE ablations leave a larger residual alignment signal and a moderate retain loss, especially for text-side excision. Removing PFL mostly raises the residual drift while preserving utility slightly better under the default lock strength. Thus, the transfer ablation points to GSD as the load-bearing component for avoiding over-erasure, while BKE and PFL mainly shape residual closure and durability.

B.15 Broader Impacts

EASE lowers the cost of honoring data-removal requests in federated multimodal systems, supporting compliance with regulations such as GDPR. We release no new dataset or pretrained model, so direct deployment risk is limited. Two indirect concerns remain: unlearning can be misused to remove evidence of unwanted training data, which deployments should mitigate by logging requests and the corresponding parameter changes for audit; and applying EASE outside dual-encoder retrieval without re-establishing the anchor structure may leave residual reconstruction paths and give false assurance of removal.

C PROOFS OF THEORETICAL RESULTS

Throughout this section, we use standard properties of orthogonal projectors (idempotency, self-adjointness, range/kernel characterization), quadratic forms induced by positive semidefinite matrices, and the inner product differentiation rule. The entanglement coefficients introduced in Section 3.3 are the cosines of the principal angles between subspaces: $\kappa_i = \cos\varphi_i$ for $i \le r = \min(p, q)$ and $\kappa_i = 0$ for $i > r$, with $\kappa_1 \ge \cdots \ge \kappa_p$ inherited from the singular-value ordering. We write c_i for the forget-side canonical directions defined in Section 3.3 and v_j for their retain-side counterparts, the columns of $\Phi_r V_f$; $\{c_i\}_{i=1}^p$ is an orthonormal basis of $\mathcal{S}_f$ and $\{v_j\}_{j=1}^q$ an orthonormal basis of $\mathcal{S}_r$. The index set $\mathcal{I}_u(\delta) = \{i : \kappa_i \le \delta\}$ collects the removable anchor directions at threshold δ. All excision and lock statements reference the initialization w_0 (Section 2); the deployed model w_n enters only through gradients and embeddings evaluated at it.

C.1 ASSUMPTIONS

The theorems in Section 3 rely on the following assumptions; each theorem statement indicates the subset it uses.

Assumption C.1 (A1: Smoothness and bounded Jacobians)**.** The alignment loss $\mathcal{L}_a(w; \mathcal{D})$ is L-smooth in w, i.e., $\|\nabla\mathcal{L}_a(w) - \nabla\mathcal{L}_a(w')\|_2 \le L\|w - w'\|_2$ for every w, w' encountered during training. The encoder Jacobians $J_\mu = \partial z_\mu / \partial w_\mu$ are bounded in operator norm, $\|J_\mu(w_\mu)\|_{\mathrm{op}} \le M$ for both $\mu \in \{v, t\}$ and all w_μ within the training region.

Assumption C.2 (A2: Local linearization)**.** There exists a radius $r > 0$ around the pre-unlearning point w_n such that, for $\|w_\mu - w_{n,\mu}\|_2 \le r$, the encoder embedding admits a first-order Taylor expansion $z_\mu(w_\mu) = z_\mu(w_{n,\mu}) + J_\mu(w_{n,\mu})(w_\mu - w_{n,\mu}) + O(\|w_\mu - w_{n,\mu}\|_2^2)$ with a uniformly bounded second-order residual. Equivalently, the Jacobian is Lipschitz inside the radius: $\|J_\mu(w_\mu) - J_\mu(w_{n,\mu})\|_{\mathrm{op}} \le M_2 \|w_\mu - w_{n,\mu}\|_2$ for a constant M_2.

Assumption C.3 (A3: Embedding-level anchor decomposition under bilateral excision)**.** At the bilateral excised point $w^* = (w_v^*, w_t^*)$ with $\Pi_{u,\mu}(w_\mu^* - w_{0,\mu}) = 0$ for both $\mu \in \{v, t\}$, write $\xi_\mu := w_\mu^* - w_{n,\mu}$ for the excision displacement and $\|\xi\|_2^2 := \sum_\mu \|\xi_\mu\|_2^2$. Define the converged-training residual $\epsilon_{\mathrm{tr}} := \max_\mu \|\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_n)\|_2$, the per-sample embedding shifts $\delta z_{\mu,i} := z_{\mu,i}(w_\mu^*) - z_{\mu,i}(w_{n,\mu})$ with batch maxima $\bar\delta_\mu := \max_i \|\delta z_{\mu,i}\|_2$, the bilateral cross-shift $\kappa_c := \max_i \|\delta z_{v,i} - \delta z_{t,i}\|_2$, and the Jacobian-drift term $\epsilon_J := 2M_2\|\xi\|_2/\gamma$ with M_2 from A2. We assume the residual Modality Anchor signal admits the decomposition

$$\left\|\Pi_{u,\mu}\,\nabla_{w_\mu}\mathcal{L}_a(w_v^*, w_t^*)\right\|_2 \;\le\; \epsilon_{\mathrm{tr}} + L_a\,\kappa_c + \epsilon_J, \qquad \mu \in \{v, t\}, \tag{14}$$

with the closed-form constants

$$L_a := M\,L_e, \qquad L_e \;\le\; \frac{2(1+\gamma)}{\gamma^2}, \tag{15}$$

where L_e is the Lipschitz constant of the InfoNCE embedding gradient with respect to the batch embeddings (derived in the proof of Lemma C.4) and M is the Jacobian operator-norm bound from A1, so the anchor sensitivity L_a carries embedding shifts into parameter-gradient space. Lemma C.4 derives the same inequality from A1–A2 alone with κ_c replaced by the larger quantity $\bar\delta_v + \bar\delta_t$. The κ_c form is the additional content of A3: it encodes that bilateral excision near a converged checkpoint shifts the two modalities by nearly the same vector, so that only the mismatch between the shifts, not their magnitude, feeds the residual anchor gradient. The co-direction property reflects that both branches are excised from the same converged checkpoint along their own unique subspaces, and it has a directly measurable counterpart: the alignment residual ρ of Section 3.2, which bilateral excision drives to 0.08 under client unlearning where one-sided excision leaves it at 0.23–0.29 (Table 2).

Lemma C.4 (Conservative anchor bound from A1–A2)**.** *Under Assumptions A1–A2 alone, at the bilateral excised point $w^* = (w_v^*, w_t^*)$ as above,*

$$\left\|\Pi_{u,\mu}\,\nabla_{w_\mu}\mathcal{L}_a(w_v^*, w_t^*)\right\|_2 \;\le\; \epsilon_{\mathrm{tr}} + L_a\left(\bar\delta_v + \bar\delta_t\right) + \epsilon_J, \qquad \mu \in \{v, t\}, \tag{16}$$

with the constants of Eq. (15)*.*

Proof. By the chain rule, $\nabla_{w_\mu}\mathcal{L}_a(w) = J_\mu(w_\mu)^\top \nabla_{z_\mu}\mathcal{L}_a\big(z(w)\big)$, where $z(w)$ collects the batch embeddings of both modalities. Write $z^* := z(w^*)$ and $z_n := z(w_n)$. Adding and subtracting, and applying the projector,

$$\begin{aligned}\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w^*) \;&= \Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_n) \;+\; \Pi_{u,\mu}J_\mu(w_\mu^*)^\top\big[\nabla_{z_\mu}\mathcal{L}_a(z^*) - \nabla_{z_\mu}\mathcal{L}_a(z_n)\big] \\ &\quad+\; \Pi_{u,\mu}\big[J_\mu(w_\mu^*) - J_\mu(w_{n,\mu})\big]^\top\nabla_{z_\mu}\mathcal{L}_a(z_n).\end{aligned}$$

The first summand has norm at most ϵ_{tr} by definition. For the second, fix the anchor index i and write the InfoNCE embedding gradient of Eq. (2) as $\nabla_{z_\mu}\mathcal{L}_a = \frac{1}{\gamma}\big(\sum_j \sigma_j z_{\bar\mu,j} - z_{\bar\mu,i}\big)$ with softmax weights σ_j over the logits $z_\mu^\top z_{\bar\mu,j}/\gamma$. For unit-norm embeddings, $\partial\sigma_j/\partial z_\mu = \frac{1}{\gamma}\sigma_j\big(z_{\bar\mu,j} - \sum_l \sigma_l z_{\bar\mu,l}\big)$, so a perturbation of z_μ changes $\sum_j \sigma_j z_{\bar\mu,j}$ by at most $\frac{2}{\gamma}\bar\delta_\mu$; a perturbation of the batch $\{z_{\bar\mu,j}\}$ changes the same sum by at most $\bar\delta_{\bar\mu}$ through the embeddings directly and by at most $\frac{2}{\gamma}\bar\delta_{\bar\mu}$ through the weights, and moves $z_{\bar\mu,i}$ by at most $\bar\delta_{\bar\mu}$. With the $1/\gamma$ prefactor,

$$\big\|\nabla_{z_\mu}\mathcal{L}_a(z^*) - \nabla_{z_\mu}\mathcal{L}_a(z_n)\big\|_2 \;\le\; \frac{2}{\gamma^2}\,\bar\delta_\mu + \frac{2+2\gamma}{\gamma^2}\,\bar\delta_{\bar\mu} \;\le\; L_e\big(\bar\delta_v + \bar\delta_t\big), \qquad L_e := \frac{2(1+\gamma)}{\gamma^2}.$$

Since $\|J_\mu(w_\mu^*)^\top\|_{\mathrm{op}} \le M$ by A1 and the projector is a contraction, the second summand has norm at most $ML_e(\bar\delta_v + \bar\delta_t)$, which is $L_a(\bar\delta_v + \bar\delta_t)$. For the third summand, A2 gives $\|J_\mu(w_\mu^*) - J_\mu(w_{n,\mu})\|_{\mathrm{op}} \le M_2\|\xi_\mu\|_2$, and the embedding gradient at z_n is bounded by $\|\nabla_{z_\mu}\mathcal{L}_a(z_n)\|_2 \le \frac{1}{\gamma}\big(\sum_j \sigma_j\|z_{\bar\mu,j}\|_2 + \|z_{\bar\mu,i}\|_2\big) \le 2/\gamma$ for unit-norm embeddings. Hence the third summand has norm at most $2M_2\|\xi_\mu\|_2/\gamma \le \epsilon_J$. Summing the three bounds gives Eq. (16). □

The lemma is weaker than A3 because $\kappa_c \le \bar\delta_v + \bar\delta_t$ by the triangle inequality, with equality only for anti-parallel shifts of equal norm. The lemma is symmetric between unilateral and bilateral excision, since a one-sided excision simply sets one of $\bar\delta_v$, $\bar\delta_t$ to zero, so A3 sharpens it by exactly the amount that the two modality shifts agree in direction — the agreement that bilateral excision supplies and a one-sided excision cannot. Proposition C.12 (iii) states the resulting suppression.

Assumption C.5 (A4: Directional alignment for unilateral dynamics)**.** For the unilateral excision case, the retain-objective gradient along the unique subspace remains aligned with its initial direction over k local SGD steps within the linearization radius of Assumption A2. Formally, writing $g(k) := -\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_\mu(k), w_{n,\bar\mu})$ and $d(k) := \Pi_{u,\mu}(w_\mu(k) - w_{0,\mu})$, with g_0 the norm of $g(0)$ and e the unit vector $g(0)/g_0$, we assume

$$\langle g(k),\, e\rangle \;\ge\; g_0 - 2L\,\langle d(k),\, e\rangle. \tag{17}$$

A4 is a local alignment condition of the kind used in linearized analyses of subspace-constrained SGD. It is not implied by L-smoothness alone: smoothness bounds $\|g(k) - g(0)\|_2$ by L times the full parameter displacement, whereas A4 charges the loss of alignment only to the component of the drift along e. It is required only for the multi-step lower bound in Proposition C.12 (ii); claims (i), (iii) and the one-step lower bound do not use it. The condition is restricted to the linearization radius of A2.

The completeness and retention results (Theorems C.8, C.9) are purely geometric and require no assumptions beyond standard projector properties. The durability result (Theorem C.11) uses only the bounded projected-gradient condition stated in its hypothesis. The anchor reconstruction result (Proposition C.12) uses A1–A2 for claim (i) and the one-step version of (ii), additionally A4 for the multi-step recurrence in (ii), and A3's decomposition (Eq. 14) for the bilateral suppression bound in (iii). Lemma C.4 records the corresponding bound available from A1–A2 alone, which Proposition C.12 (iii) refines.

C.2 Properties of the Unlearning Projector

Proposition C.6 (Properties of the Unlearning Projector)**.** *Let $\Pi_{u,\mu} = B_{u,\mu}(B_{u,\mu})^\top \in \mathbb{R}^{d_\mu\times d_\mu}$ denote the orthogonal projector onto $\mathcal{U}_\mu = \mathrm{col}(B_{u,\mu})$. Then $\Pi_{u,\mu}$ is idempotent ($\Pi_{u,\mu}^2 = \Pi_{u,\mu}$) and self-adjoint ($\Pi_{u,\mu}^\top = \Pi_{u,\mu}$). Its range is $\mathrm{im}(\Pi_{u,\mu}) = \mathcal{U}_\mu$ and its kernel is $\ker(\Pi_{u,\mu}) = \mathcal{U}_\mu^\perp$. The complementary projector preserves all directions outside $\mathcal{U}_\mu$: $(I_{d_\mu} - \Pi_{u,\mu})v = v$ for every $v \in \mathcal{U}_\mu^\perp$.*

Proof. Let $B_{u,\mu} \in \mathbb{R}^{d_\mu \times p_u}$ with orthonormal columns, i.e., $(B_{u,\mu})^\top B_{u,\mu} = I_{p_u}$, and $\Pi_{u,\mu} = B_{u,\mu}(B_{u,\mu})^\top$. For idempotency, we compute

$$\begin{aligned}\Pi_{u,\mu}^2 &= B_{u,\mu}(B_{u,\mu})^\top \cdot B_{u,\mu}(B_{u,\mu})^\top \\ &= B_{u,\mu}\big[(B_{u,\mu})^\top B_{u,\mu}\big](B_{u,\mu})^\top \\ &= B_{u,\mu}\, I_{p_u}\, (B_{u,\mu})^\top = \Pi_{u,\mu},\end{aligned} \tag{18}$$

Self-adjointness follows directly:

$$\Pi_{u,\mu}^\top = \big(B_{u,\mu}(B_{u,\mu})^\top\big)^\top = B_{u,\mu}(B_{u,\mu})^\top = \Pi_{u,\mu}. \tag{19}$$

For the range, take any $v = B_{u,\mu}\alpha \in \mathcal{U}_\mu$. Then

$$\Pi_{u,\mu}\, v = B_{u,\mu}\big[(B_{u,\mu})^\top B_{u,\mu}\big]\alpha = B_{u,\mu}\, I_{p_u}\, \alpha = B_{u,\mu}\alpha = v, \tag{20}$$

so $\mathcal{U}_\mu \subseteq \mathrm{im}(\Pi_{u,\mu})$. Conversely, for any $w \in \mathbb{R}^{d_\mu}$, $\Pi_{u,\mu}\, w = B_{u,\mu}\big((B_{u,\mu})^\top w\big) \in \mathcal{U}_\mu$, giving $\mathrm{im}(\Pi_{u,\mu}) = \mathcal{U}_\mu$. For the kernel, the equivalence chain

$$\Pi_{u,\mu}\, v = 0 \iff B_{u,\mu}(B_{u,\mu})^\top v = 0 \iff (B_{u,\mu})^\top v = 0 \iff v \perp \mathcal{U}_\mu \tag{21}$$

establishes $\ker(\Pi_{u,\mu}) = \mathcal{U}_\mu^\perp$. Finally, for any $v \in \mathcal{U}_\mu^\perp$,

$$(I_{d_\mu} - \Pi_{u,\mu})\, v = v - \Pi_{u,\mu}\, v = v - 0 = v, \tag{22}$$

so the complementary projector preserves all directions outside $\mathcal{U}_\mu$. Since $\mathcal{E}_\mu \subseteq \mathcal{U}_\mu^\perp$ by orthogonality of the principal-angle coordinate system, this includes all retain support and retain-only directions. □

C.3 Forget Lock Characterization

Proposition C.7 (Forget Lock Characterization)**.** *Let $\Pi_{u,\mu} = B_{u,\mu}(B_{u,\mu})^\top$ be the orthogonal projector onto $\mathcal{U}_\mu$. Then $\mathcal{L}_f(w) = 0$ if and only if $\Pi_{u,\mu}(w_\mu - w_{0,\mu}) = 0$ for both $\mu \in \{v, t\}$. Moreover, $\nabla_{w_\mu}\mathcal{L}_f = 2\Pi_{u,\mu}(w_\mu - w_{0,\mu})$, which vanishes for displacements orthogonal to $\mathcal{U}_\mu$ and points back toward $w_{0,\mu}$ along $\mathcal{U}_\mu$ otherwise.*

Proof. Since the two modalities contribute independently, it suffices to analyze a single modality μ. Let $\Delta_\mu = w_\mu - w_{0,\mu}$. The μ-th summand of $\mathcal{L}_f$ can be expanded as

$$\begin{aligned}\ell_\mu &= \Delta_\mu^\top \Pi_{u,\mu}\, \Delta_\mu \\ &= \Delta_\mu^\top B_{u,\mu}(B_{u,\mu})^\top \Delta_\mu \\ &= \big\|(B_{u,\mu})^\top \Delta_\mu\big\|_2^2 \;\geq\; 0.\end{aligned} \tag{23}$$

Therefore $\ell_\mu = 0$ if and only if

$$(B_{u,\mu})^\top \Delta_\mu = 0 \iff \Delta_\mu \in \mathcal{U}_\mu^\perp \iff \Pi_{u,\mu}\, \Delta_\mu = 0. \tag{24}$$

Summing over both modalities, $\mathcal{L}_f(w) = 0$ iff $\Pi_{u,\mu}(w_\mu - w_{0,\mu}) = 0$ for both $\mu \in \{v, t\}$. Since $\Pi_{u,\mu}$ is symmetric and $w_{0,\mu}$ is constant, the standard identity $\nabla_x(x^\top A\, x) = 2A\, x$ gives

$$\begin{aligned}\nabla_{w_\mu}\mathcal{L}_f &= \nabla_{w_\mu}\big[\Delta_\mu^\top \Pi_{u,\mu}\, \Delta_\mu\big] \\ &= 2\,\Pi_{u,\mu}\, \Delta_\mu \\ &= 2\,\Pi_{u,\mu}\, (w_\mu - w_{0,\mu}).\end{aligned} \tag{25}$$

This gradient lies in $\mathcal{U}_\mu = \mathrm{im}(\Pi_{u,\mu})$ and vanishes iff $\Delta_\mu \in \mathcal{U}_\mu^\perp$. □

C.4 Proof of Theorem C.8 (Unlearning Completeness)

Theorem C.8 (Unlearning Completeness)**.** *Let w_μ^* denote the parameters obtained by projecting the deployed model via Eq.* (10) *and $\Delta_\mu = w_{n,\mu} - w_{0,\mu}$ the training displacement. Then:*

(i) Exact erasure: $\Pi_{u,\mu}(w^*_\mu - w_{0,\mu}) = 0$*, i.e., the excised parameters carry zero displacement from the initialization along* $\mathcal{U}_\mu$*.*

(ii) Energy removal: The energy removed equals $\|\Pi_{u,\mu}\Delta_\mu\|_2^2$*. For any* $\Delta_\mu \in \mathcal{S}_f$ *expressed in the canonical basis as* $\Delta_\mu = \sum_{i=1}^p \alpha_i c_i$*, this equals* $\sum_{i:\kappa_i \le \delta} \alpha_i^2$*.*

(iii) Forget energy ratio: $\eta_f(\delta) = \|B_{u,\mu}^\top G_{f,\mu}\|_F^2 / \|G_{f,\mu}\|_F^2$ *is monotonically non-decreasing in* δ *and satisfies* $\eta_f(1) \ge \tau_e$ *(equality only when the truncation exactly meets the energy threshold).*

Proof. We work with a single modality μ throughout; the argument applies identically to both branches. Let $\Delta_\mu = w_{n,\mu} - w_{0,\mu}$ denote the training displacement of the deployed model.

We begin by establishing exact erasure. By Eq. (10), the projected parameters satisfy $w^*_\mu = w_{n,\mu} - \Pi_{u,\mu}\Delta_\mu$. Subtracting $w_{0,\mu}$ from both sides and applying $\Pi_{u,\mu}$:

$$\begin{aligned}\Pi_{u,\mu}(w^*_\mu - w_{0,\mu}) &= \Pi_{u,\mu}\big[\Delta_\mu - \Pi_{u,\mu}\Delta_\mu\big] \\ &= \Pi_{u,\mu}\left(I_{d_\mu} - \Pi_{u,\mu}\right)\Delta_\mu \\ &= \left(\Pi_{u,\mu} - \Pi_{u,\mu}^2\right)\Delta_\mu \\ &= \left(\Pi_{u,\mu} - \Pi_{u,\mu}\right)\Delta_\mu = 0,\end{aligned} \tag{26}$$

where the last line uses idempotency $\Pi_{u,\mu}^2 = \Pi_{u,\mu}$ (Proposition C.6).

Next, we quantify the energy removed by the projection. The squared norm of the removed component is

$$\begin{aligned}\|\Delta_\mu - (w^*_\mu - w_{0,\mu})\|_2^2 &= \|\Delta_\mu - (I_{d_\mu} - \Pi_{u,\mu})\Delta_\mu\|_2^2 \\ &= \|\Pi_{u,\mu}\,\Delta_\mu\|_2^2.\end{aligned} \tag{27}$$

Now suppose $\Delta_\mu \in \mathcal{S}_f$ and write $\Delta_\mu = \sum_{i=1}^p \alpha_i\, c_i$ in the canonical basis $\{c_i\}_{i=1}^p = \{(\Phi_f U_f)_{\cdot,i}\}_{i=1}^p$. Since $\{c_i\}$ are orthonormal and $\Pi_{u,\mu}$ projects onto $\mathrm{span}\{c_i : \kappa_i \le \delta\}$, we have $\Pi_{u,\mu}\, c_i = c_i$ if $\kappa_i \le \delta$ and $\Pi_{u,\mu}\, c_i = 0$ otherwise. Therefore

$$\begin{aligned}\|\Pi_{u,\mu}\,\Delta_\mu\|_2^2 &= \Big\|\Pi_{u,\mu}\sum_{i=1}^p \alpha_i\, c_i\Big\|_2^2 \\ &= \Big\|\sum_{i:\,\kappa_i \le \delta} \alpha_i\, c_i\Big\|_2^2 \\ &= \sum_{i:\,\kappa_i \le \delta} \alpha_i^2.\end{aligned} \tag{28}$$

Finally, we analyze the monotonicity of the forget energy ratio $\eta_f(\delta) = \|B_{u,\mu}^\top G_{f,\mu}\|_F^2 / \|G_{f,\mu}\|_F^2$. As δ increases, the index set $\mathcal{I}_u(\delta) = \{i : \kappa_i \le \delta\}$ can only grow, so $B_{u,\mu}$ gains columns and $\|B_{u,\mu}^\top G_{f,\mu}\|_F^2$ is non-decreasing in δ. At $\delta = 1$, all p canonical directions are classified as removable anchor directions, so $B_{u,\mu} = \Phi_f U_f$. The boundary value evaluates as

$$\begin{aligned}\eta_f(1) &= \frac{\|B_{u,\mu}^\top G_{f,\mu}\|_F^2}{\|G_{f,\mu}\|_F^2} = \frac{\|(\Phi_f U_f)^\top G_{f,\mu}\|_F^2}{\|G_{f,\mu}\|_F^2} \\ &= \frac{\|U_f^\top \Phi_f^\top G_{f,\mu}\|_F^2}{\|G_{f,\mu}\|_F^2} = \frac{\|\Phi_f^\top G_{f,\mu}\|_F^2}{\|G_{f,\mu}\|_F^2} = \frac{\sum_{i=1}^p \sigma_i^2}{\|G_{f,\mu}\|_F^2} \ge \tau_e,\end{aligned} \tag{29}$$

where the fourth equality uses the orthogonality of U_f (which preserves the Frobenius norm), and the last inequality follows from the energy criterion (6), which selects p as the smallest index whose cumulative energy already reaches τ_e. Equality holds exactly when the cumulative energy of the top-p directions equals $\tau_e\,\|G_{f,\mu}\|_F^2$. □

C.5 Proof of Theorem C.9 (Retention Integrity)

Theorem C.9 (Retention Integrity)**.** *For any retain displacement $\Delta_r \in \mathcal{S}_r$, the collateral energy is bounded by*

$$\|\Pi_{u,\mu}\,\Delta_r\|_2^2 \;\le\; \delta^2\,\|\Delta_r\|_2^2. \tag{30}$$

Proof. Let $\Delta_r \in \mathcal{S}_r$ with $\|\Delta_r\|_2 = 1$ (the general case follows by homogeneity). Since V_f is a $q \times q$ orthogonal matrix and Φ_r has orthonormal columns, the retain-side canonical directions $v_j = (\Phi_r V_f)_{\cdot,j}$ form an orthonormal basis of $\mathcal{S}_r$, so

$$\Delta_r = \sum_{j=1}^{q} \beta_j\, v_j, \qquad \sum_{j=1}^{q} \beta_j^2 = 1. \tag{31}$$

The inner products between the two canonical bases are the entries of the rectangular diagonal matrix Σ in Eq. (7):

$$\langle c_i, v_j\rangle = (\Phi_f U_f)_{\cdot,i}^{\top}(\Phi_r V_f)_{\cdot,j} = \big(U_f^{\top}\Phi_f^{\top}\Phi_r V_f\big)_{ij} = \big(U_f^{\top} M V_f\big)_{ij} = \Sigma_{ij}, \tag{32}$$

which equals κ_i when $i = j \le r$ and 0 otherwise. Therefore, for each removable anchor direction c_i,

$$\langle c_i, \Delta_r\rangle = \sum_{j=1}^{q} \beta_j\,\Sigma_{ij} = \begin{cases} \beta_i\,\kappa_i, & i \le r, \\ 0, & i > r, \end{cases} \tag{33}$$

and the second case is consistent with $\kappa_i = 0$ for $i > r$. Since $\Pi_{u,\mu}$ is the orthogonal projector onto the span of the orthonormal set $\{c_i : \kappa_i \le \delta\}$,

$$\begin{aligned} \|\Pi_{u,\mu}\,\Delta_r\|_2^2 &= \sum_{i:\,\kappa_i \le \delta} \langle c_i, \Delta_r\rangle^2 = \sum_{i \le r:\,\kappa_i \le \delta} \beta_i^2\,\kappa_i^2 \\ &\le \delta^2 \sum_{i \le r:\,\kappa_i \le \delta} \beta_i^2 \;\le\; \delta^2 \sum_{j=1}^{q} \beta_j^2 \;=\; \delta^2\,\|\Delta_r\|_2^2, \end{aligned} \tag{34}$$

where the first inequality uses $\kappa_i \le \delta$ on the summation set and the second drops non-negative terms. □

C.6 Completeness–Integrity Trade-off Analysis

The threshold δ simultaneously controls the unlearning completeness ratio $\eta_f(\delta)$ and the retention damage bound $\eta_r(\delta)$. We formalize their joint behavior.

Proposition C.10 (Monotonicity and Pareto Trade-off)**.** *Let $\eta_f(\delta) = \|B_{u,\mu}^{\top} G_{f,\mu}\|_F^2 / \|G_{f,\mu}\|_F^2$ and $\eta_r(\delta) = \max_{\Delta_r \in \mathcal{S}_r \setminus \{0\}} \|\Pi_{u,\mu}\Delta_r\|_2^2 / \|\Delta_r\|_2^2$. Then both $\eta_f(\delta)$ and $\eta_r(\delta)$ are monotonically non-decreasing in $\delta \in [0,1]$. Moreover, for any target unlearning ratio $\eta_f^* \in (0, \tau_e]$, the smallest threshold achieving $\eta_f(\delta) \ge \eta_f^*$ is $\delta^* = \kappa_{k^*}$, where $k^* = \max\{k : \eta_f(\kappa_k) \ge \eta_f^*\}$ and $\varphi_1 \le \cdots \le \varphi_p$ are the principal angles in non-decreasing order. This δ^* simultaneously minimizes the retention damage bound $(\delta^*)^2$.*

Proof. We first establish monotonicity. For η_f, as δ increases, the index set $\mathcal{I}_u(\delta) = \{i : \kappa_i \le \delta\}$ can only grow. Since $B_{u,\mu}$ is formed by collecting the columns of $\Phi_f U_f$ indexed by $\mathcal{I}_u(\delta)$, the projection $B_{u,\mu}^{\top} G_{f,\mu}$ captures a non-decreasing share of the forget gradient energy:

$$\delta' > \delta \implies \mathcal{I}_u(\delta) \subseteq \mathcal{I}_u(\delta') \implies \|B_{u,\mu}(\delta)^{\top} G_{f,\mu}\|_F^2 \;\le\; \|B_{u,\mu}(\delta')^{\top} G_{f,\mu}\|_F^2, \tag{35}$$

so $\eta_f(\delta)$ is non-decreasing. For η_r: by Theorem C.9, $\eta_r(\delta) \le \delta^2$, and the bound δ^2 is strictly increasing. More precisely,

$$\eta_r(\delta) = \max_{i \in \mathcal{I}_u(\delta)} \kappa_i^2, \tag{36}$$

which is non-decreasing since $\mathcal{I}_u(\delta)$ grows with δ.

Turning to the optimal threshold, we observe that η_f is a step function of δ that jumps only at the coefficient values $\kappa_1 \ge \cdots \ge \kappa_p$, so setting the threshold to κ_k selects $\mathcal{I}_u(\kappa_k) \supseteq \{k, \ldots, p\}$, with

equality when κ_k is not tied with κ_{k-1}, and $\eta_f(\kappa_k)$ is non-increasing in k. The feasible indices $\{k : \eta_f(\kappa_k) \geq \eta_f^*\}$ therefore form an initial segment $\{1, \ldots, k^*\}$ with

$$\delta^* = \kappa_{k^*}, \quad k^* = \max\{k : \eta_f(\kappa_k) \geq \eta_f^*\}, \tag{37}$$

and κ_{k^*} is the smallest feasible cosine. For any threshold below κ_{k^*}, the index set shrinks to $\{k^*+1, \ldots, p\}$ or further and the target is missed by maximality of k^*, so δ^* is the smallest δ satisfying the constraint. Since $\eta_r \leq \delta^2$ is increasing in δ, this δ^* simultaneously minimizes the retention damage bound $(\delta^*)^2$. □

C.7 Proof of Theorem C.11 (Unlearning Durability)

Theorem C.11 (Unlearning Durability). *Suppose $\|\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a\|_2 \leq G_u$ for all w encountered during training and $\eta_{\text{lr}} < 1/(2\alpha)$. The Temporal Re-anchoring drift after t steps from the projected point obeys*

$$\|\Pi_{u,\mu}(w_\mu(t) - w_{0,\mu})\|_2 \;\leq\; \frac{G_u}{2\alpha}\left[1 - (1 - 2\eta_{\text{lr}}\alpha)^t\right] \;\leq\; \frac{G_u}{2\alpha}. \tag{38}$$

Proof. Fix modality μ and let $d(t) = \Pi_{u,\mu}(w_\mu(t) - w_{0,\mu})$ denote the unique-subspace displacement at step t. A single SGD step under the combined objective (12) updates:

$$w_\mu(t{+}1) = w_\mu(t) - \eta_{\text{lr}}\big[\nabla_{w_\mu}\mathcal{L}_a(w(t)) + 2\alpha\,\Pi_{u,\mu}(w_\mu(t) - w_{0,\mu})\big]. \tag{39}$$

Applying $\Pi_{u,\mu}$ to both sides and writing $g_u(t) = \Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w(t))$:

$$\begin{aligned} d(t{+}1) &= \Pi_{u,\mu}\big[w_\mu(t{+}1) - w_{0,\mu}\big] \\ &= \Pi_{u,\mu}(w_\mu(t) - w_{0,\mu}) - \eta_{\text{lr}}\,\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a - 2\eta_{\text{lr}}\alpha\,\Pi_{u,\mu}^2(w_\mu(t) - w_{0,\mu}) \\ &= d(t) - \eta_{\text{lr}}\,g_u(t) - 2\eta_{\text{lr}}\alpha\,d(t) \\ &= (1 - 2\eta_{\text{lr}}\alpha)\,d(t) - \eta_{\text{lr}}\,g_u(t), \end{aligned} \tag{40}$$

where $\|g_u(t)\|_2 \leq G_u$ by assumption. Taking norms and applying the triangle inequality:

$$\|d(t{+}1)\|_2 \;\leq\; (1 - 2\eta_{\text{lr}}\alpha)\,\|d(t)\|_2 + \eta_{\text{lr}}\,G_u. \tag{41}$$

Since $\eta_{\text{lr}} < 1/(2\alpha)$, the contraction factor $\zeta = 1 - 2\eta_{\text{lr}}\alpha \in (0, 1)$. Starting from $\|d(0)\|_2 = 0$ (exact erasure by Theorem C.8), unrolling the recurrence gives:

$$\begin{aligned} \|d(t)\|_2 &\leq \eta_{\text{lr}}G_u \sum_{s=0}^{t-1}\zeta^s \\ &= \frac{\eta_{\text{lr}}G_u\,(1 - \zeta^t)}{1 - \zeta} \\ &= \frac{G_u}{2\alpha}\left[1 - (1 - 2\eta_{\text{lr}}\alpha)^t\right]. \end{aligned} \tag{42}$$

Since $(1 - 2\eta_{\text{lr}}\alpha)^t \geq 0$, the bound is at most $G_u/(2\alpha)$. The bound survives federated aggregation: FedAvg replaces the client displacements $d_k(t)$ by a convex combination $\sum_k p_k\,d_k(t)$, and each client resumes from the same aggregate, so the aggregate satisfies the same recurrence with $g_u(t)$ replaced by $\sum_k p_k\,g_{u,k}(t)$, whose norm is again at most G_u. □

C.8 Proof of Proposition C.12 (Anchor Reconstruction Tendency under Unilateral Excision)

Proposition C.12 (Anchor Reconstruction Tendency under Unilateral Excision). *Under Assumptions A1–A4 (Appendix C.1), suppose excision is applied to modality μ only, producing w_μ^* with $\Pi_{u,\mu}(w_\mu^* - w_{0,\mu}) = 0$, while $w_{\bar{\mu}}$ is held at $w_{n,\bar{\mu}}$. Define the Modality Anchor signal along the unique subspace at the excised point as $g_0 := \|\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_\mu^*, w_{n,\bar{\mu}})\|_2$. Then:*

- *(i) Anchor survives (A1–A2): $g_0 > 0$ whenever the InfoNCE residual r_μ assembled from the intact embeddings $z_{\bar{\mu}}(w_{n,\bar{\mu}})$ (Eq. 47) has a non-zero component along $J_\mu(w_\mu^*)\mathcal{U}_\mu$, the image of the unique subspace under the Jacobian.*

(ii) Drift lower-bound tendency (A1–A2, A4): let the client run local SGD on the combined objective (12) *with lock strength $\alpha \geq 0$ ($\alpha = 0$ is plain retain training) from the excised point, and let $d(k) := \Pi_{u,\mu}(w_\mu(k) - w_{0,\mu})$ be the drift away from the erased state, so that $d(0) = 0$. One step gives $\|d(1)\|_2 = \eta_{\mathrm{lr}}\, g_0$. Under the alignment condition A4, for step $\eta_{\mathrm{lr}} < 1/(2(L+\alpha))$ and k steps within the linearization radius,*

$$\|d(k)\|_2 \;\geq\; \frac{g_0}{2(L+\alpha)}\Big[1 - \big(1 - 2\eta_{\mathrm{lr}}(L+\alpha)\big)^k\Big], \tag{43}$$

which grows monotonically and approaches $g_0/(2(L+\alpha))$ geometrically. The lock shrinks this floor in proportion to α, and the floor stays proportional to g_0, which is the quantity bilateral excision reduces.

(iii) Bilateral suppression (A3): applying excision to both $\mu \in \{v, t\}$ simultaneously and writing $g_0^{\mathrm{bi}} := \|\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_v^*, w_t^*)\|_2$ for the anchor signal at the bilateral point, A3 (Eq. 14) gives

$$g_0^{\mathrm{bi}} \;\leq\; \epsilon_{\mathrm{tr}} + L_a\,\kappa_c + \epsilon_J, \tag{44}$$

which is small whenever the federated checkpoint converged on retain pairs (ϵ_{tr} small), the bilateral excision induces near-symmetric embedding shifts (κ_c small), and the excision displacement is small relative to the linearization radius (ϵ_J small). With g_0 replaced by g_0^{bi}, the floor in (43) *is suppressed to the same scale.*

Proof. We analyze the gradient of the alignment loss $\mathcal{L}_a$ with respect to the parameters of a single modality, and show that unilateral unlearning leaves a Modality Anchor signal from the unmodified branch that drives ρ away from zero.

The cross-modal similarity $s(w) = z_{i,v}(w_v)^\top z_{j,t}(w_t)$ is an inner product of two vector-valued functions depending on disjoint parameter blocks. Let $J_v(w_v) = \partial z_{i,v}/\partial w_v \in \mathbb{R}^{m\times d_v}$ and $J_t(w_t) = \partial z_{j,t}/\partial w_t \in \mathbb{R}^{m\times d_t}$ denote the Jacobians. By the inner product differentiation rule:

$$\begin{aligned}\nabla_{w_v} s(w) &= \nabla_{w_v}\big[z_{i,v}(w_v)^\top z_{j,t}(w_t)\big] \\ &= J_v(w_v)^\top\, z_{j,t}(w_t),\end{aligned} \tag{45}$$

where the cross term vanishes because $z_{j,t}$ does not depend on w_v. By symmetry:

$$\nabla_{w_t} s(w) = J_t(w_t)^\top\, z_{i,v}(w_v). \tag{46}$$

Now consider the unilateral case where only w_v is unlearned. Suppose the projection is applied to w_v only, producing w_v^*, while the language parameters remain at $w_{n,t}$. At the unlearned point $w^* = (w_v^*, w_{n,t})$, the InfoNCE gradient of Eq. (2) evaluates to

$$\nabla_{w_v}\mathcal{L}_a \;=\; \frac{1}{\gamma}\,J_v(w_v^*)^\top\, r_v, \qquad r_v \;:=\; \sum_{j=1}^{N}\sigma_j\, z_{j,t}(w_{n,t}) \;-\; z_{i,t}(w_{n,t}), \tag{47}$$

where σ_j are the softmax weights of the contrastive loss. The critical observation is that the residual r_v is assembled entirely from textual embeddings computed under the pre-unlearning parameters $w_{n,t}$, which encode the full memory of the forgotten cross-modal alignment. This residual acts as the Modality Anchor: during retraining, the gradient in Eq. (47) continuously pushes w_v toward directions that re-align $z_{i,v}$ with the unchanged textual embeddings, thereby reconstructing the forgotten pairing and driving ρ away from zero.

By an identical argument with the roles of v and t exchanged, if only w_t is unlearned to w_t^* while w_v remains at $w_{n,v}$, the alignment gradient with respect to w_t at the unlearned point $w^* = (w_{n,v}, w_t^*)$ satisfies

$$\nabla_{w_t}\mathcal{L}_a \;=\; \frac{1}{\gamma}\,J_t(w_t^*)^\top\, r_t, \qquad r_t \;:=\; \sum_{j=1}^{N}\sigma_j\, z_{j,v}(w_{n,v}) \;-\; z_{i,v}(w_{n,v}). \tag{48}$$

Here r_t is assembled from the original visual embeddings, which serve as the Modality Anchor that pulls the language branch back toward the erased alignment during retraining, again yielding $\rho > 0$.

Now suppose both modalities are unlearned simultaneously, producing $w^* = (w_v^*, w_t^*)$. The projection in Eq. (10) removes the unique-subspace components of the training displacement from both w_v and w_t, displacing both embeddings away from the original alignment. At the bilateral unlearned point, the alignment gradients become

$$\nabla_{w_v}\mathcal{L}_a \;=\; \frac{1}{\gamma}\, J_v(w_v^*)^\top\, r_v^*, \qquad r_v^* \;:=\; \sum_{j=1}^{N} \sigma_j\, z_{j,t}(w_t^*) \;-\; z_{i,t}(w_t^*), \tag{49}$$

$$\nabla_{w_t}\mathcal{L}_a \;=\; \frac{1}{\gamma}\, J_t(w_t^*)^\top\, r_t^*, \qquad r_t^* \;:=\; \sum_{j=1}^{N} \sigma_j\, z_{j,v}(w_v^*) \;-\; z_{i,v}(w_v^*). \tag{50}$$

Crucially, both residuals are now assembled from displaced embeddings that no longer encode the original cross-modal alignment. To formalize this, we verify that the projection removes the forget-specific components from both branches. For the visual side, Theorem C.8 gives:

$$\Pi_{u,v}\,(w_v^* - w_{0,v}) = 0, \tag{51}$$

and the Forget Lock (Proposition C.7) maintains this condition throughout retraining. An analogous condition holds for the language side:

$$\Pi_{u,t}\,(w_t^* - w_{0,t}) = 0. \tag{52}$$

Since the forget-specific knowledge resides in $\mathcal{U}_v$ and $\mathcal{U}_t$, Eqs. (51)–(52) guarantee that along the forget-specific directions both branches coincide with the initialization w_0, which never trained on $\mathcal{D}_f$. When only modality μ is excised, the residual entering the gradient is built from the intact embeddings of the unmodified branch, as in Eq. (47), and encodes the forgotten alignment as a fixed Modality Anchor. Since g_0 equals $\frac{1}{\gamma}\|\Pi_{u,\mu}J_\mu(w_\mu^*)^\top r_\mu\|_2$, and $\Pi_{u,\mu}J_\mu(w_\mu^*)^\top r_\mu$ vanishes exactly when r_μ is orthogonal to the image $J_\mu(w_\mu^*)\mathcal{U}_\mu$, one has $g_0 > 0$ whenever r_μ has a non-zero component along $J_\mu(w_\mu^*)\mathcal{U}_\mu$, establishing claim (i). When both modalities are excised simultaneously, the residual is instead assembled from the displaced embeddings, as in Eqs. (49)–(50). Since $\Pi_{u,\bar{\mu}}(w_{\bar{\mu}}^* - w_{0,\bar{\mu}}) = 0$, along the unique subspace the $\bar{\mu}$ branch coincides with the initialization that never trained on $\mathcal{D}_f$, so the embeddings entering the residual no longer carry the forgotten pairing, removing the dominant Modality Anchor signal along unique directions. By A3 (Eq. 14), the residual anchor gradient at the bilateral excised point satisfies $g_0^{\mathrm{bi}} \le \epsilon_{\mathrm{tr}} + L_a\,\kappa_c + \epsilon_J$, establishing claim (iii). The structural reason this bound is small in practice is that bilateral excision displaces $z_v(w_v^*)$ and $z_t(w_t^*)$ by approximately equal amounts (small κ_c), so the InfoNCE alignment gradient does not regrow the original pairing direction.

Drift lower-bound tendency (claim (ii)). For the unilateral case, let $d(k) = \Pi_{u,\mu}(w_\mu(k) - w_{0,\mu})$ denote the unique-subspace drift after k local SGD steps on the combined objective (12) with lock strength $\alpha \ge 0$; $d(0) = 0$ by the projection. Applying $\Pi_{u,\mu}$ to the SGD update, exactly as in Eq. (40), gives

$$d(k{+}1) \;=\; (1 - 2\eta_{\mathrm{lr}}\alpha)\, d(k) + \eta_{\mathrm{lr}}\, g(k), \qquad g(k) := -\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_\mu(k), w_{n,\bar{\mu}}), \tag{53}$$

where the sign convention makes $g(k)$ the projected descent direction. At $k = 0$, $g(0) = -\Pi_{u,\mu}\nabla_{w_\mu}\mathcal{L}_a(w_\mu^*, w_{n,\bar{\mu}})$ with $\|g(0)\|_2 = g_0$, and since $d(0) = 0$ the lock term vanishes, giving the one-step drift

$$\|d(1)\|_2 \;=\; \eta_{\mathrm{lr}}\,\|g(0)\|_2 \;=\; \eta_{\mathrm{lr}}\, g_0, \tag{54}$$

which requires only A1–A2.

For $k \ge 2$, we invoke Assumption A4 (directional alignment). Write e for the unit vector $g(0)/g_0$ and $a(k)$ for the component $\langle d(k), e\rangle$. Projecting Eq. (53) onto e and applying the alignment condition Eq. (17) gives

$$a(k{+}1) \;\ge\; (1 - 2\eta_{\mathrm{lr}}\alpha)\, a(k) + \eta_{\mathrm{lr}}\big(g_0 - 2L\, a(k)\big) \;=\; \big(1 - 2\eta_{\mathrm{lr}}(L+\alpha)\big)\, a(k) + \eta_{\mathrm{lr}}\, g_0. \tag{55}$$

Since $\eta_{\mathrm{lr}} < 1/(2(L+\alpha))$, the factor $\zeta' := 1 - 2\eta_{\mathrm{lr}}(L+\alpha)$ lies in $(0,1)$, so the right-hand side is non-decreasing in $a(k)$ and induction from $a(0) = 0$ unrolls the recurrence:

$$a(k) \;\ge\; \eta_{\mathrm{lr}}\, g_0 \sum_{s=0}^{k-1} (\zeta')^s \;=\; \frac{\eta_{\mathrm{lr}}\, g_0\big(1 - (\zeta')^k\big)}{1 - \zeta'} \;=\; \frac{g_0}{2(L+\alpha)}\Big[1 - \big(1 - 2\eta_{\mathrm{lr}}(L+\alpha)\big)^k\Big]. \tag{56}$$

By the Cauchy–Schwarz inequality, $\|d(k)\|_2 \geq \langle d(k), e \rangle$, which yields Eq. (43) in the proposition statement. The drift grows monotonically in k and approaches $g_0/(2(L+\alpha))$ geometrically. Setting $\alpha = 0$ recovers plain retain training with floor $g_0/(2L)$; a positive α lowers the floor to $g_0/(2(L+\alpha))$ but leaves it proportional to g_0, which is why the lock alone does not repair unilateral excision. Under bilateral excision, A3 (Eq. 14) bounds the anchor signal by $\epsilon_{\text{tr}} + L_a\,\kappa_c + \epsilon_J$, so the floor in Eq. (43) is suppressed to the same scale, removing the systematic reconstruction pressure present in the unilateral case. □

D Compute-Matched Retrain Comparison

A natural reviewer concern is whether the efficiency gains of EASE simply reflect a smaller training budget relative to retrain. To isolate budget from algorithmic contribution, we compare our method against retrain early-stopped at 25%, 50%, and 75% of the full federated round count, alongside the full-budget retrain reference. The communication cost $Comm.$ directly reflects the compute budget consumed, since per-round payload is constant within each scenario. Results on Flickr30K with CLIP-B/32 are reported in Table 8.

Table 8: Compute-matched retrain comparison on Flickr30K with CLIP-B/32. Retrain-α% denotes early-stopping retrain executed for α% of the full communication rounds; its budget is reflected by $Comm.$ The top sub-table reports forget-side metrics (F-$R@k$) and membership inference (MIA); the bottom sub-table reports retain-side metrics (R-$R@k$) and communication cost ($Comm.$, MB). Parenthetical values give the absolute gap to the full-budget retrain reference with (↑green) closer to ideal and (↓red) farther; for MIA, (val) is $|\text{MIA} - 50|$ (smaller is better); for $Comm.$, it is the absolute difference to full-budget retrain. MIA ranks use distance to 50%, with tied values sharing a rank. *Retrain (Full)* is the reference and does not participate in ranking.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-$R@1$ ↓	F-$R@5$ ↓	F-$R@10$ ↓	MIA	F-$R@1$ ↓	F-$R@5$ ↓	F-$R@10$ ↓	MIA	F-$R@1$ ↓	F-$R@5$ ↓	F-$R@10$ ↓	MIA
Retrain-25%	0.0 (↑0.1)	0.4 (↑1.1)	0.9 (↑1.4)	12.3 (37.7)	0.1 (↑0.1)	0.4 (↑0.3)	0.9 (↑0.5)	19.5 (30.5)	0.0 (↑0.1)	0.3 (↑0.8)	0.8 (↑1.6)	14.8 (35.2)
Retrain-50%	0.0 (↑0.1)	0.8 (↑0.7)	1.6 (↑0.7)	15.2 (34.8)	0.2	0.6 (↑0.1)	1.1 (↑0.3)	22.8 (27.2)	0.0 (↑0.1)	0.6 (↑0.5)	1.6 (↑0.8)	16.0 (34.0)
Retrain-75%	0.1	1.2 (↑0.3)	1.9 (↑0.4)	16.4 (33.6)	0.2	0.7	1.3 (↑0.1)	23.9 (26.1)	0.1	0.9 (↑0.2)	2.1 (↑0.3)	16.5 (33.5)
EASE	0.3 (↓0.2)	1.0 (↑0.5)	2.2 (↑0.1)	**16.9** (33.1)	0.3 (↓0.1)	1.5 (↓0.8)	2.6 (↓1.2)	**25.4** (24.6)	0.1	2.2 (↓1.1)	5.5 (↓3.1)	**16.7** (33.3)
Retrain (Full)	0.1	1.5	2.3	16.7	0.2	0.7	1.4	24.2	0.1	1.1	2.4	16.7

Method	R-$R@1$ ↑	R-$R@5$ ↑	R-$R@10$ ↑	$Comm.$ ↓	R-$R@1$ ↑	R-$R@5$ ↑	R-$R@10$ ↑	$Comm.$ ↓	R-$R@1$ ↑	R-$R@5$ ↑	R-$R@10$ ↑	$Comm.$ ↓
Retrain-25%	62.3 (↓28.4)	88.5 (↓11.4)	94.7 (↓5.3)	65.6 (196.9)	61.7 (↓27.7)	87.5 (↓12.4)	93.9 (↓6.1)	49.2 (147.7)	54.2 (↓26.4)	83.1 (↓14.6)	90.5 (↓8.9)	49.2 (147.7)
Retrain-50%	79.8 (↓10.9)	97.1 (↓2.8)	99.0 (↓1.0)	131.3 (131.2)	78.0 (↓11.4)	95.6 (↓4.3)	98.3 (↓1.7)	98.5 (98.4)	69.2 (↓11.4)	92.6 (↓5.1)	96.8 (↓2.6)	98.5 (98.4)
Retrain-75%	87.1 (↓3.6)	99.3 (↓0.6)	99.8 (↓0.2)	196.9 (65.6)	85.1 (↓4.3)	98.5 (↓1.4)	99.5 (↓0.5)	147.7 (49.2)	76.8 (↓3.8)	96.4 (↓1.3)	98.5 (↓0.9)	147.7 (49.2)
EASE	86.5 (↓4.2)	99.2 (↓0.7)	99.8 (↓0.2)	52.5 (210.0)	80.8 (↓8.6)	96.6 (↓3.3)	99.3 (↓0.7)	47.2 (149.7)	79.4 (↓1.2)	97.7	99.2 (↓0.2)	39.4 (157.5)
Retrain (Full)	90.7	99.9	100.0	262.5	89.4	99.9	100.0	196.9	80.6	97.7	99.4	196.9

EASE uses less communication than Retrain-25% in every scenario (Table 8) yet improves R-$R@1$ by +24.2, +19.1, and +25.2 points on the client, class, and sample scenarios, respectively, while keeping forget-side metrics near the full-budget retrain reference. Retrain-25% can also produce low forget recall, but its retain collapse indicates undertraining rather than targeted unlearning. EASE reaches retain quality close to Retrain-75% at only 20–25% of the full communication budget.

E Likelihood-Ratio Attack at Low False-Positive Rate

Figure 14 compares LiRA TPR@FPR=1%, defined in Appendix B.4, across all nine (dataset, backbone) combinations and three unlearning scenarios.

Cross-dataset pattern. Across all nine combinations, EASE stays within 1.5 points of *Retrain* on every scenario, placing it consistently at the bottom of each subplot. In contrast, forget-reconstruction baselines such as Ferrari, FedOSD, and FUSED sit in the 25–42% range on the client and class scenarios, suggesting that their aggregate MIA values in Section 4.2 underestimate how much forget-side signal a targeted attacker can still recover. GradAscent produces the largest low-FPR leakage in the client scenario, consistent with the known instability of gradient-reversal updates on tightly coupled modalities.

Backbone sensitivity. The relative ordering is stable across CLIP-B/32, CLIP-L/14, and SigLIP; absolute values shift upward for larger encoders (CLIP-L/14) due to greater capacity for memorization, but the gap between EASE and baselines widens rather than shrinks. This is consistent with bilateral

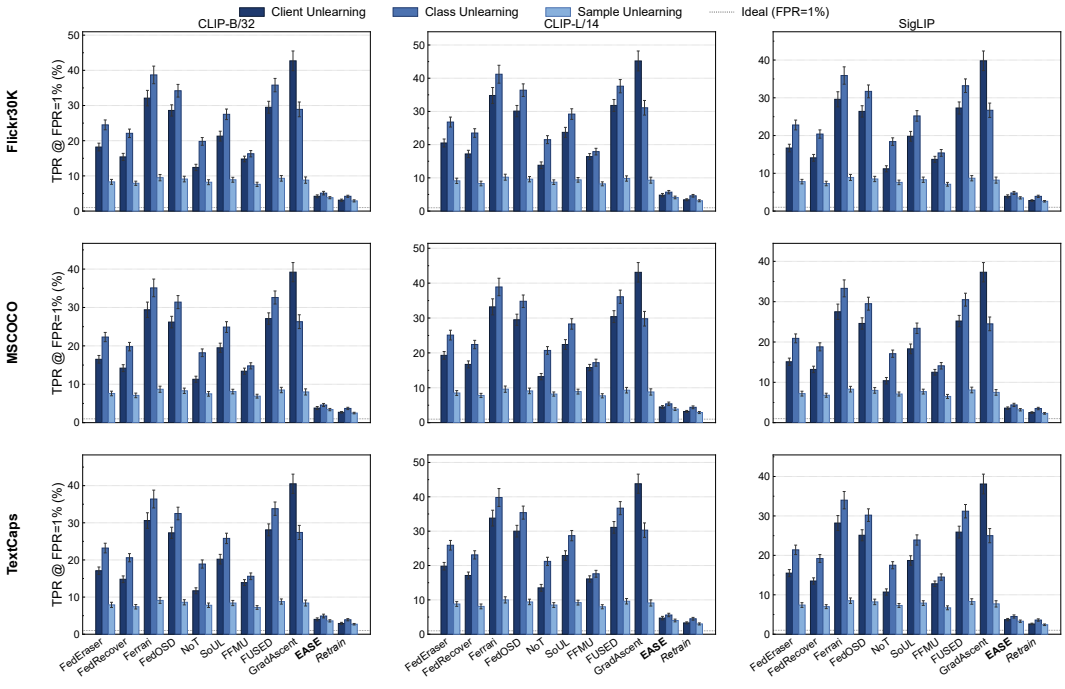


Figure 14: $\mathrm{TPR@FPR} = 1\%$ across three datasets (rows: Flickr30K, MSCOCO, TextCaps) and three backbones (columns: CLIP-B/32, CLIP-L/14, SigLIP). Lower is better; the dotted line marks the 1% random-guess reference for non-member-like behavior. *Retrain* is the non-participating reference.

excision targeting the Modality Anchor that scales with encoder capacity, while per-modality baselines do not close this anchor regardless of backbone size.

Scenario dependence. The sample scenario produces uniformly lower TPR than client and class, because sample forget sets are sparsely distributed across clients and carry a weaker attacker signature in aggregate. Even so, our method approaches *Retrain* most closely in this regime, indicating that gains from bilateral excision persist even where the residual leakage is already small.

E.1 Entanglement-Stratified Membership Inference

The aggregate audit above treats the forget set as a whole; this analysis asks whether leakage concentrates on the samples whose influence overlaps most with retained data. Each forget sample receives an entanglement score $s \in [0, 1]$, its normalized gradient projection onto the entangled subspace $\mathcal{E}$, and the forget set is partitioned into four fixed score intervals; under client unlearning on Flickr30K with CLIP-B/32 the bins contain 809, 971, 668, and 452 samples. The per-sample LiRA attack of Appendix B.4 is then evaluated within each bin, with the 1% false-positive threshold calibrated against that bin's own non-member pool, so the rates in Table 9 are not directly comparable to the single-threshold aggregate in Figure 14.

Table 9: Entanglement-stratified LiRA TPR@FPR=1% under client unlearning. Columns give the four entanglement-score bins and the bin-weighted mean; parentheses give the per-bin gap to *Retrain*.

	Backbone	Method	$[0, .25)$	$[.25, .5)$	$[.5, .75)$	$[.75, 1]$	W-mean
Flickr30K	CLIP-B/32	Retrain	1.1	1.2	1.3	1.5	1.2
		EASE	1.7 (+0.6)	2.0 (+0.8)	2.2 (+0.9)	2.5 (+1.0)	2.0 (+0.8)
	CLIP-L/14	Retrain	1.6	1.4	1.6	2.0	1.6
		EASE	13.0 (+11.4)	14.9 (+13.5)	15.7 (+14.1)	16.4 (+14.4)	14.5 (+12.9)
	SigLIP	Retrain	0.3	0.5	0.4	1.1	0.5
		EASE	17.5 (+17.2)	18.5 (+18.0)	18.3 (+17.9)	20.0 (+18.9)	18.4 (+17.9)
MSCOCO	CLIP-B/32	Retrain	0.8	1.0	1.0	1.7	1.0
		EASE	8.0 (+7.2)	8.5 (+7.5)	8.0 (+7.0)	10.3 (+8.6)	8.5 (+7.5)
	CLIP-L/14	Retrain	1.1	1.3	1.4	1.8	1.4
		EASE	19.0 (+17.9)	20.4 (+19.1)	21.5 (+20.1)	23.0 (+21.2)	21.2 (+19.8)
	SigLIP	Retrain	0.8	0.9	1.3	1.4	1.0
		EASE	22.9 (+22.1)	25.3 (+24.4)	24.5 (+23.2)	26.7 (+25.3)	24.8 (+23.8)
TextCaps	CLIP-B/32	Retrain	2.1	2.2	2.2	2.3	2.2
		EASE	16.4 (+14.3)	16.9 (+14.7)	18.5 (+16.3)	19.4 (+17.1)	17.4 (+15.2)
	CLIP-L/14	Retrain	0.5	0.9	1.1	1.3	0.9
		EASE	27.7 (+27.2)	28.4 (+27.5)	28.9 (+27.8)	31.1 (+29.8)	28.6 (+27.7)
	SigLIP	Retrain	0.9	1.2	1.3	1.4	1.1
		EASE	36.9 (+36.0)	37.7 (+36.5)	38.6 (+37.3)	40.6 (+39.2)	37.7 (+36.6)

On Flickr30K with CLIP-B/32, *Retrain* rises from 1.1 to 1.5 across the four bins, while EASE rises from 1.7 to 2.5. The corresponding gaps are $+0.6$, $+0.8$, $+0.9$, and $+1.0$. What this stratification is built to test is whether the most entangled samples form a leakage hot spot once the retain-side directions are preserved, and across all nine dataset–backbone pairs they do not: the EASE rate rises by at most $1.5\times$ from the least- to the most-entangled bin, and in six of the nine pairs that rise is smaller than the one *Retrain* itself shows over the same bins. Leakage therefore follows the entanglement profile of the retrain reference instead of concentrating where excision and retention compete, which is the behaviour the unique-subspace decomposition is designed to produce.

F Full Comparison Results

We report the remaining eight (dataset, backbone) combinations here, extending Table 1 in the main body. Each combination follows the same two-part layout and ranking convention: the top sub-table covers forget-side metrics (F-$R@k$) and MIA, and the bottom sub-table covers retain-side metrics (R-$R@k$) and communication cost. Tables 18–21 extend the comparison to the four transfer architectures of Appendix B.14 (BLIP, BLIP-2, Qwen2.5-VL-7B, LLaVA-1.5-7B) across client, class, and sample unlearning, following the protocol and metric conventions defined there.

Table 10: Main comparison on Flickr30K with CLIP-L/14 across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	88.5 (↓88.4)	97.3 (↓96.2)	99.2 (↓97.2)	52.8 (2.8)	79.8 (↓79.6)	95.5 (↓95.0)	97.7 (↓96.2)	57.6 (7.6)	**0.0**	0.9 (↑0.5)	**1.7** (↑0.3)	16.7 (33.3)
FedRecover	75.3 (↓75.2)	91.5 (↓90.4)	94.5 (↓92.5)	54.5 (4.5)	69.7 (↓69.5)	87.0 (↓86.5)	91.8 (↓90.3)	59.6 (9.6)	0.0	1.0 (↑0.4)	2.1 (↓0.1)	16.7 (33.3)
Ferrari	95.9 (↓95.8)	99.9 (↓98.8)	100.0 (↓98.0)	59.5 (9.5)	94.5 (↓94.3)	99.9 (↓99.4)	100.0 (↓98.5)	63.7 (13.7)	0.0	1.0 (↑0.4)	2.4 (↓0.4)	16.7 (33.3)
FedOSD	95.7 (↓95.6)	100.0 (↓98.9)	100.0 (↓98.0)	57.6 (7.6)	93.2 (↓93.0)	100.0 (↓99.5)	100.0 (↓98.5)	61.5 (11.5)	0.3 (↓0.3)	1.8 (↓0.4)	2.8 (↓0.8)	16.7 (33.3)
NoT	85.4 (↓85.3)	98.8 (↓97.7)	99.4 (↓97.4)	51.2 (1.2)	76.4 (↓76.2)	96.9 (↓96.4)	98.8 (↓97.3)	55.9 (5.9)	0.1 (↓0.1)	1.3 (↑0.1)	3.3 (↓1.3)	16.7 (33.3)
SoUL	93.6 (↓93.5)	99.8 (↓98.7)	100.0 (↓98.0)	53.6 (3.6)	91.2 (↓91.0)	99.4 (↓98.9)	99.9 (↓98.4)	59.7 (9.7)	0.1 (↓0.1)	0.9 (↑0.5)	2.0	16.7 (33.3)
FFMU	65.6 (↓65.5)	90.3 (↓89.2)	94.6 (↓92.6)	54.5 (4.5)	46.8 (↓46.6)	81.3 (↓80.8)	90.5 (↓89.0)	**52.3** (2.3)	0.0	**0.6** (↑0.8)	1.7 (↑0.3)	**16.8** (33.2)
FUSED	94.6 (↓94.5)	99.8 (↓98.7)	100.0 (↓98.0)	59.2 (9.2)	92.9 (↓92.7)	99.8 (↓99.3)	100.0 (↓98.5)	63.5 (13.5)	0.2 (↓0.2)	0.8 (↑0.6)	2.1 (↓0.1)	16.7 (33.3)
GradAscent	6.4 (↓6.3)	18.1 (↓17.0)	28.0 (↓26.0)	**50.0** (0.0)	2.2 (↓2.0)	9.2 (↓8.7)	15.2 (↓13.7)	47.1 (2.9)	0.0	1.0 (↑0.4)	1.9 (↑0.1)	16.7 (33.3)
EASE	**0.2** (↓0.1)	**1.1**	**2.3** (↓0.3)	16.7 (33.3)	**0.1** (↑0.1)	**1.4** (↓0.9)	**4.0** (↓2.5)	16.8 (33.2)	0.4 (↓0.4)	4.0 (↓2.6)	7.5 (↓5.5)	16.7 (33.3)
Retrain	0.1	1.1	2.0	16.7	0.2	0.5	1.5	16.8	0.0	1.4	2.0	16.7

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	94.4 (↓5.2)	99.3 (↓0.7)	99.5 (↓0.5)	537.5	90.4 (↓6.5)	99.1 (↓0.9)	99.7 (↓0.3)	403.1	91.3 (↓7.4)	99.2 (↓0.8)	99.9 (↓0.1)	403.1
FedRecover	91.9 (↓7.7)	98.3 (↓1.7)	99.4 (↓0.6)	**96.8** (440.7)	89.8 (↓7.1)	98.5 (↓1.5)	99.3 (↓0.7)	**72.6** (330.5)	72.2 (↓26.5)	87.4 (↓12.6)	91.4 (↓8.6)	**72.6** (330.5)
Ferrari	94.1 (↓5.5)	99.2 (↓0.8)	99.6 (↓0.4)	147.8 (389.7)	91.2 (↓5.7)	99.2 (↓0.8)	99.8 (↓0.2)	120.9 (282.2)	92.1 (↓6.6)	99.1 (↓0.9)	99.8 (↓0.2)	120.9 (282.2)
FedOSD	93.7 (↓5.9)	99.0 (↓1.0)	99.6 (↓0.4)	215.0 (322.5)	91.5 (↓5.4)	99.4 (↓0.6)	99.8 (↓0.2)	161.2 (241.9)	92.7 (↓6.0)	99.5 (↓0.5)	99.7 (↓0.3)	161.2 (241.9)
NoT	92.5 (↓7.1)	98.7 (↓1.3)	99.4 (↓0.6)	107.5 (430.0)	85.3 (↓11.6)	98.7 (↓1.3)	99.5 (↓0.5)	80.6 (322.5)	90.5 (↓8.2)	98.8 (↓1.2)	99.6 (↓0.4)	80.6 (322.5)
SoUL	93.3 (↓6.3)	99.1 (↓0.9)	99.6 (↓0.4)	107.5 (430.0)	90.8 (↓6.1)	99.1 (↓0.9)	99.7 (↓0.3)	80.6 (322.5)	91.7 (↓7.0)	99.2 (↓0.8)	99.8 (↓0.2)	80.6 (322.5)
FFMU	55.0 (↓44.6)	82.6 (↓17.4)	90.3 (↓9.7)	215.0 (322.5)	63.6 (↓33.3)	92.2 (↓7.8)	99.3 (↓0.7)	161.2 (241.9)	69.0 (↓29.7)	91.5 (↓8.5)	95.2 (↓4.8)	161.2 (241.9)
FUSED	94.1 (↓5.5)	99.1 (↓0.9)	99.7 (↓0.3)	215.0 (322.5)	90.8 (↓6.1)	99.0 (↓1.0)	99.8 (↓0.2)	161.2 (241.9)	91.7 (↓7.0)	99.4 (↓0.6)	99.9 (↓0.1)	161.2 (241.9)
GradAscent	7.1 (↓92.5)	19.5 (↓80.5)	28.4 (↓71.6)	147.8 (389.7)	3.6 (↓93.3)	12.1 (↓87.9)	21.6 (↓78.4)	120.9 (282.2)	69.9 (↓28.8)	89.2 (↓10.8)	94.9 (↓5.1)	120.9 (282.2)
EASE	**95.1** (↓4.5)	**99.5** (↓0.5)	**99.8** (↓0.2)	107.5 (430.0)	**92.2** (↓4.7)	**99.5** (↓0.5)	**99.9** (↓0.1)	96.8 (306.3)	**93.1** (↓5.6)	**99.6** (↓0.4)	**100.0**	80.6 (322.5)
Retrain	99.6	100.0	100.0	537.5	96.9	100.0	100.0	403.1	98.7	100.0	100.0	403.1

Table 11: Main comparison on Flickr30K with SigLIP across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	82.1 (↓81.7)	93.9 (↓92.6)	97.0 (↓94.9)	51.2 (1.2)	57.4 (↓57.3)	82.6 (↓81.4)	91.2 (↓89.1)	64.0 (14.0)	0.1 (↓0.1)	0.7 (↑0.2)	1.6 (↑0.1)	16.7 (33.3)
FedRecover	73.1 (↓72.7)	88.6 (↓87.3)	92.6 (↓90.5)	48.1 (1.9)	51.2 (↓51.1)	78.6 (↓77.4)	85.6 (↓83.5)	**47.3** (2.7)	0.3 (↓0.3)	**0.4** (↑0.5)	1.4 (↑0.3)	**16.8** (33.2)
Ferrari	93.1 (↓92.7)	99.8 (↓98.5)	100.0 (↓97.9)	54.9 (4.9)	77.4 (↓77.3)	97.9 (↓96.7)	99.6 (↓97.5)	67.8 (17.8)	0.1 (↓0.1)	0.9	1.3 (↑0.4)	16.7 (33.3)
FedOSD	90.4 (↓90.0)	99.3 (↓98.0)	99.9 (↓97.8)	52.4 (2.4)	75.9 (↓75.8)	96.6 (↓95.4)	99.5 (↓97.4)	66.5 (16.5)	0.2 (↓0.2)	0.8 (↑0.1)	1.8 (↓0.1)	16.7 (33.3)
NoT	77.6 (↓77.2)	95.7 (↓94.4)	98.1 (↓96.0)	**49.9** (0.1)	53.5 (↓53.4)	84.7 (↓83.5)	93.1 (↓91.0)	58.5 (8.5)	0.3 (↓0.3)	0.7 (↑0.2)	1.6 (↑0.1)	16.7 (33.3)
SoUL	89.7 (↓89.3)	98.4 (↓97.1)	99.5 (↓97.4)	52.8 (2.8)	70.3 (↓70.2)	94.7 (↓93.5)	98.4 (↓96.3)	62.5 (12.5)	0.2 (↓0.2)	0.6 (↑0.3)	1.7	16.7 (33.3)
FFMU	62.8 (↓62.4)	87.9 (↓86.6)	92.5 (↓90.4)	**49.9** (0.1)	37.5 (↓37.4)	68.4 (↓67.2)	80.3 (↓78.2)	55.8 (5.8)	0.2 (↓0.2)	1.3 (↓0.4)	1.9 (↓0.2)	**16.8** (33.2)
FUSED	93.3 (↓92.9)	99.8 (↓98.5)	100.0 (↓97.9)	54.0 (4.0)	74.8 (↓74.7)	96.1 (↓94.9)	98.6 (↓96.5)	67.2 (17.2)	**0.0**	1.0 (↓0.1)	1.6 (↑0.1)	16.7 (33.3)
GradAscent	91.1 (↓90.7)	99.4 (↓98.1)	99.7 (↓97.6)	54.5 (4.5)	73.3 (↓73.2)	95.8 (↓94.6)	99.1 (↓97.0)	62.3 (12.3)	0.1 (↓0.1)	0.9	**1.2** (↑0.5)	16.7 (33.3)
EASE	**0.6** (↓0.2)	**1.8** (↓0.5)	**3.1** (↓1.0)	16.7 (33.3)	**0.1**	**0.9** (↑0.3)	**2.4** (↓0.3)	18.7 (31.3)	0.8 (↓0.8)	3.7 (↓2.8)	9.2 (↓7.5)	**16.8** (33.2)
Retrain	0.4	1.3	2.1	16.7	0.1	1.2	2.1	18.3	0.0	0.9	1.7	16.7

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	90.1 (↓9.1)	99.1 (↓0.9)	99.7 (↓0.3)	475.0	89.4 (↓8.4)	98.7 (↓1.3)	99.2 (↓0.8)	356.2	92.4 (↓5.9)	99.3 (↓0.7)	99.8 (↓0.2)	356.2
FedRecover	87.9 (↓11.3)	98.2 (↓1.8)	99.3 (↓0.7)	**85.5** (389.5)	86.1 (↓11.7)	97.9 (↓2.1)	98.8 (↓1.2)	**64.1** (292.1)	68.0 (↓30.3)	86.2 (↓13.8)	90.5 (↓9.5)	**64.1** (292.1)
Ferrari	90.9 (↓8.3)	99.1 (↓0.9)	99.8 (↓0.2)	130.6 (344.4)	89.1 (↓8.7)	98.6 (↓1.4)	99.2 (↓0.8)	106.9 (249.3)	91.6 (↓6.7)	99.5 (↓0.5)	99.8 (↓0.2)	106.9 (249.3)
FedOSD	91.2 (↓8.0)	99.3 (↓0.7)	99.8 (↓0.2)	190.0 (285.0)	88.3 (↓9.5)	98.7 (↓1.3)	99.1 (↓0.9)	142.5 (213.7)	89.7 (↓8.6)	99.5 (↓0.5)	99.6 (↓0.4)	142.5 (213.7)
NoT	82.7 (↓16.5)	98.6 (↓1.4)	99.5 (↓0.5)	95.0 (380.0)	87.5 (↓10.3)	98.1 (↓1.9)	99.0 (↓1.0)	71.2 (285.0)	82.0 (↓16.3)	99.0 (↓1.0)	99.5 (↓0.5)	71.2 (285.0)
SoUL	90.5 (↓8.7)	99.0 (↓1.0)	99.7 (↓0.3)	95.0 (380.0)	88.7 (↓9.1)	98.5 (↓1.5)	99.2 (↓0.8)	71.2 (285.0)	92.0 (↓6.3)	99.4 (↓0.6)	99.7 (↓0.3)	71.2 (285.0)
FFMU	62.3 (↓36.9)	87.4 (↓12.6)	92.0 (↓8.0)	190.0 (285.0)	67.5 (↓30.3)	90.7 (↓9.3)	98.8 (↓1.2)	142.5 (213.7)	66.1 (↓32.2)	89.0 (↓11.0)	93.5 (↓6.5)	142.5 (213.7)
FUSED	90.5 (↓8.7)	98.9 (↓1.1)	99.8 (↓0.2)	190.0 (285.0)	89.1 (↓8.7)	98.4 (↓1.6)	99.3 (↓0.7)	142.5 (213.7)	93.3 (↓5.0)	99.7 (↓0.3)	99.8 (↓0.2)	142.5 (213.7)
GradAscent	88.7 (↓10.5)	98.4 (↓1.6)	99.4 (↓0.6)	130.6 (344.4)	86.9 (↓10.9)	97.9 (↓2.1)	98.9 (↓1.1)	106.9 (249.3)	90.2 (↓8.1)	98.8 (↓1.2)	99.4 (↓0.6)	106.9 (249.3)
EASE	**91.9** (↓7.3)	**99.4** (↓0.6)	**99.9** (↓0.1)	95.0 (380.0)	**90.1** (↓7.7)	**98.9** (↓1.1)	**99.4** (↓0.6)	85.5 (270.7)	**93.4** (↓4.9)	**99.8** (↓0.2)	**99.9** (↓0.1)	71.2 (285.0)
Retrain	99.2	100.0	100.0	475.0	97.8	100.0	100.0	356.2	98.3	100.0	100.0	356.2

Table 12: Main comparison on MSCOCO with CLIP-B/32 across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	39.5 (↓39.5)	69.9 (↓69.2)	81.4 (↓79.9)	**50.1** (0.1)	38.5 (↓38.4)	68.7 (↓68.0)	82.5 (↓80.6)	53.2 (3.2)	**0.1** (↓0.1)	**0.7** (↓0.2)	1.8 (↓1.0)	40.9 (9.1)
FedRecover	32.6 (↓32.6)	58.4 (↓57.7)	71.4 (↓69.9)	**49.9** (0.1)	42.5 (↓42.4)	74.0 (↓73.3)	83.2 (↓81.3)	**49.7** (0.3)	0.6 (↓0.6)	1.0 (↓0.5)	1.5 (↓0.7)	**41.1** (8.9)
Ferrari	54.2 (↓54.2)	85.0 (↓84.3)	92.5 (↓91.0)	50.8 (0.8)	56.4 (↓56.3)	86.4 (↓85.7)	94.6 (↓92.7)	55.7 (5.7)	0.3 (↓0.3)	1.0 (↓0.5)	1.6 (↓0.8)	40.9 (9.1)
FedOSD	49.4 (↓49.4)	80.9 (↓80.2)	89.9 (↓88.4)	51.0 (1.0)	55.6 (↓55.5)	85.4 (↓84.7)	93.5 (↓91.6)	55.1 (5.1)	0.2 (↓0.2)	0.9 (↓0.4)	1.8 (↓1.0)	40.9 (9.1)
NoT	39.6 (↓39.6)	70.4 (↓69.7)	82.5 (↓81.0)	50.2 (0.2)	35.1 (↓35.0)	68.4 (↓67.7)	78.8 (↓76.9)	47.9 (2.1)	0.4 (↓0.4)	1.1 (↓0.6)	1.5 (↓0.7)	41.0 (9.0)
SoUL	46.3 (↓46.3)	78.0 (↓77.3)	87.7 (↓86.2)	50.7 (0.7)	54.4 (↓54.3)	84.0 (↓83.3)	92.2 (↓90.3)	54.1 (4.1)	0.1 (↓0.1)	1.0 (↓0.5)	1.5 (↓0.7)	40.9 (9.1)
FFMU	29.6 (↓29.6)	59.3 (↓58.6)	72.2 (↓70.7)	50.4 (0.4)	27.3 (↓27.2)	55.9 (↓55.2)	70.0 (↓68.1)	49.3 (0.7)	0.3 (↓0.3)	1.0 (↓0.5)	**1.2** (↓0.4)	40.9 (9.1)
FUSED	49.2 (↓49.2)	79.9 (↓79.2)	89.9 (↓88.4)	50.5 (0.5)	55.3 (↓55.2)	85.0 (↓84.3)	93.3 (↓91.4)	55.2 (5.2)	0.2 (↓0.2)	0.8 (↓0.3)	1.4 (↓0.6)	40.9 (9.1)
GradAscent	51.6 (↓51.6)	82.0 (↓81.3)	91.4 (↓89.9)	50.9 (0.9)	50.6 (↓50.5)	79.9 (↓79.2)	88.2 (↓86.3)	51.3 (1.3)	0.4 (↓0.4)	0.9 (↓0.4)	1.6 (↓0.8)	40.9 (9.1)
EASE	**0.1** (↓0.1)	**0.5** (↑0.2)	**1.4** (↑0.1)	40.9 (9.1)	**0.1**	**0.9** (↓0.2)	**2.3** (↓0.4)	41.1 (8.9)	0.3 (↓0.3)	1.1 (↓0.6)	2.5 (↓1.7)	40.9 (9.1)
Retrain	0.0	0.7	1.5	40.9	0.1	0.7	1.9	41.1	0.0	0.5	0.8	40.9

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	59.8 (↓10.7)	90.1 (↓5.1)	95.5 (↓3.6)	262.5	65.1 (↓10.7)	95.1 (↓3.3)	98.4 (↓1.6)	196.9	49.9 (↓1.7)	84.6 (↓1.9)	93.2 (↓1.4)	196.9
FedRecover	56.8 (↓13.7)	89.7 (↓5.5)	95.3 (↓3.8)	**47.2** (215.3)	62.1 (↓13.7)	94.4 (↓4.0)	98.2 (↓1.8)	**35.4** (161.5)	27.4 (↓24.2)	53.3 (↓33.2)	64.4 (↓30.2)	**35.4** (161.5)
Ferrari	59.8 (↓10.7)	90.1 (↓5.1)	95.5 (↓3.6)	72.2 (190.3)	64.3 (↓11.5)	95.1 (↓3.3)	98.4 (↓1.6)	59.1 (137.8)	48.8 (↓2.8)	84.8 (↓1.7)	93.2 (↓1.4)	59.1 (137.8)
FedOSD	60.1 (↓10.4)	90.2 (↓5.0)	95.6 (↓3.5)	105.0 (157.5)	65.7 (↓10.1)	95.0 (↓3.4)	98.5 (↓1.5)	78.8 (118.1)	43.5 (↓8.1)	76.8 (↓9.7)	87.1 (↓7.5)	78.8 (118.1)
NoT	47.8 (↓22.7)	81.5 (↓13.7)	91.9 (↓7.2)	52.5 (210.0)	63.5 (↓12.3)	94.6 (↓3.8)	98.2 (↓1.8)	39.4 (157.5)	38.0 (↓13.6)	68.6 (↓17.9)	81.3 (↓13.3)	39.4 (157.5)
SoUL	59.4 (↓11.1)	90.0 (↓5.2)	95.5 (↓3.6)	52.5 (210.0)	64.7 (↓11.1)	95.0 (↓3.4)	98.3 (↓1.7)	39.4 (157.5)	44.9 (↓6.7)	77.1 (↓9.4)	88.7 (↓5.9)	39.4 (157.5)
FFMU	33.7 (↓36.8)	67.2 (↓28.0)	79.7 (↓19.4)	105.0 (157.5)	51.4 (↓24.4)	86.4 (↓12.0)	98.0 (↓2.0)	78.8 (118.1)	25.6 (↓26.0)	50.9 (↓35.6)	66.6 (↓28.0)	78.8 (118.1)
FUSED	59.0 (↓11.5)	90.3 (↓4.9)	95.4 (↓3.7)	105.0 (157.5)	65.4 (↓10.4)	95.2 (↓3.2)	98.4 (↓1.6)	78.8 (118.1)	49.2 (↓2.4)	84.9 (↓1.6)	93.3 (↓1.3)	78.8 (118.1)
GradAscent	57.6 (↓12.9)	89.4 (↓5.8)	95.1 (↓4.0)	72.2 (190.3)	62.9 (↓12.9)	94.4 (↓4.0)	98.1 (↓1.9)	59.1 (137.8)	47.4 (↓4.2)	84.1 (↓2.4)	92.9 (↓1.7)	59.1 (137.8)
EASE	**60.8** (↓9.7)	**90.4** (↓4.8)	**95.7** (↓3.4)	52.5 (210.0)	**66.1** (↓9.7)	**95.4** (↓3.0)	**98.6** (↓1.4)	47.2 (149.7)	**50.6** (↓1.0)	**85.1** (↓1.4)	**93.4** (↓1.2)	39.4 (157.5)
Retrain	70.5	95.2	99.1	262.5	75.8	98.4	100.0	196.9	51.6	86.5	94.6	196.9

Table 13: Main comparison on MSCOCO with CLIP-L/14 across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	30.9 (↓30.6)	55.6 (↓54.7)	67.3 (↓65.7)	49.1 (0.9)	36.0 (↓35.9)	61.8 (↓61.2)	74.7 (↓72.5)	51.2 (1.2)	**0.1**	**0.6** (↑0.5)	1.5 (↑0.5)	40.9 (9.1)
FedRecover	24.8 (↓24.5)	49.0 (↓48.1)	59.3 (↓57.7)	51.3 (1.3)	40.9 (↓40.8)	69.5 (↓68.9)	77.4 (↓75.2)	46.7 (3.3)	0.1	0.9 (↑0.2)	1.9 (↑0.1)	40.9 (9.1)
Ferrari	49.9 (↓49.6)	82.0 (↓81.1)	91.4 (↓89.8)	51.9 (1.9)	59.9 (↓59.8)	86.2 (↓85.6)	91.8 (↓89.6)	52.3 (2.3)	0.1	0.9 (↑0.2)	2.0	40.9 (9.1)
FedOSD	44.3 (↓44.0)	74.8 (↓73.9)	86.4 (↓84.8)	50.7 (0.7)	57.9 (↓57.8)	85.4 (↓84.8)	91.8 (↓89.6)	52.1 (2.1)	0.2 (↓0.1)	0.9 (↑0.2)	2.3 (↓0.3)	**41.0** (9.0)
NoT	34.6 (↓34.3)	64.8 (↓63.9)	78.7 (↓77.1)	**49.9** (0.1)	35.2 (↓35.1)	62.4 (↓61.8)	73.3 (↓71.1)	48.0 (2.0)	0.3 (↓0.2)	0.7 (↑0.4)	**1.4** (↑0.6)	40.9 (9.1)
SoUL	42.4 (↓42.1)	72.9 (↓72.0)	85.5 (↓83.9)	50.3 (0.3)	54.9 (↓54.8)	83.2 (↓82.6)	89.2 (↓87.0)	51.3 (1.3)	0.2 (↓0.1)	1.1	1.6 (↑0.4)	40.9 (9.1)
FFMU	27.8 (↓27.5)	52.8 (↓51.9)	64.3 (↓62.7)	48.8 (1.2)	22.5 (↓22.4)	48.4 (↓47.8)	60.4 (↓58.2)	49.0 (1.0)	0.1	1.4 (↓0.3)	2.7 (↓0.7)	40.9 (9.1)
FUSED	44.6 (↓44.3)	76.0 (↓75.1)	87.2 (↓85.6)	50.5 (0.5)	55.8 (↓55.7)	83.8 (↓83.2)	91.0 (↓88.8)	52.0 (2.0)	0.2 (↓0.1)	1.0 (↑0.1)	2.0	40.9 (9.1)
GradAscent	43.1 (↓42.8)	74.1 (↓73.2)	85.8 (↓84.2)	50.8 (0.8)	50.0 (↓49.9)	80.3 (↓79.7)	87.7 (↓85.5)	**50.2** (0.2)	0.1	0.8 (↑0.3)	2.1 (↓0.1)	40.9 (9.1)
EASE	**0.4** (↓0.1)	**1.4** (↓0.5)	**2.1** (↓0.5)	40.9 (9.1)	**0.3** (↓0.2)	**1.2** (↓0.6)	**2.6** (↓0.4)	41.1 (8.9)	0.1	1.8 (↓0.7)	3.8 (↓1.8)	40.9 (9.1)
Retrain	0.3	0.9	1.6	40.9	0.1	0.6	2.2	41.1	0.1	1.1	2.0	40.9

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	57.8 (↓14.9)	88.9 (↓7.8)	93.8 (↓5.5)	537.5	77.2 (↓10.3)	98.4 (↓1.3)	99.7 (↓0.3)	403.1	53.1 (↓7.9)	82.5 (↓7.7)	91.9 (↓4.7)	403.1
FedRecover	55.6 (↓17.1)	88.4 (↓8.3)	93.6 (↓5.7)	**96.8** (440.7)	74.7 (↓12.8)	97.6 (↓2.1)	99.2 (↓0.8)	**72.6** (330.5)	22.5 (↓38.5)	42.8 (↓47.4)	55.2 (↓41.4)	**72.6** (330.5)
Ferrari	57.4 (↓15.3)	88.6 (↓8.1)	93.8 (↓5.5)	147.8 (389.7)	76.9 (↓10.6)	98.3 (↓1.4)	99.5 (↓0.5)	120.9 (282.2)	52.8 (↓8.2)	82.3 (↓7.9)	91.8 (↓4.8)	120.9 (282.2)
FedOSD	58.4 (↓14.3)	88.8 (↓7.9)	93.9 (↓5.4)	215.0 (322.5)	77.5 (↓10.0)	98.4 (↓1.3)	99.6 (↓0.4)	161.2 (241.9)	40.8 (↓20.2)	73.1 (↓17.1)	85.1 (↓11.5)	161.2 (241.9)
NoT	50.8 (↓21.9)	81.6 (↓15.1)	90.4 (↓8.9)	107.5 (430.0)	75.3 (↓12.2)	97.8 (↓1.9)	99.4 (↓0.6)	80.6 (322.5)	33.3 (↓27.7)	67.4 (↓22.8)	80.7 (↓15.9)	80.6 (322.5)
SoUL	57.4 (↓15.3)	88.7 (↓8.0)	93.8 (↓5.5)	107.5 (430.0)	76.5 (↓11.0)	98.2 (↓1.5)	99.6 (↓0.4)	80.6 (322.5)	44.2 (↓16.8)	76.5 (↓13.7)	88.2 (↓8.4)	80.6 (322.5)
FFMU	36.8 (↓35.9)	68.6 (↓28.1)	79.4 (↓19.9)	215.0 (322.5)	64.3 (↓23.2)	93.1 (↓6.6)	99.2 (↓0.8)	161.2 (241.9)	25.0 (↓36.0)	51.4 (↓38.8)	64.5 (↓32.1)	161.2 (241.9)
FUSED	57.0 (↓15.7)	89.0 (↓7.7)	93.7 (↓5.6)	215.0 (322.5)	76.1 (↓11.4)	98.1 (↓1.6)	99.6 (↓0.4)	161.2 (241.9)	52.0 (↓9.0)	82.6 (↓7.6)	91.9 (↓4.7)	161.2 (241.9)
GradAscent	54.8 (↓17.9)	88.1 (↓8.6)	93.5 (↓5.8)	147.8 (389.7)	73.9 (↓13.6)	97.6 (↓2.1)	99.3 (↓0.7)	120.9 (282.2)	42.8 (↓18.2)	73.4 (↓16.8)	85.7 (↓10.9)	120.9 (282.2)
EASE	**58.8** (↓13.9)	**89.1** (↓7.6)	**94.0** (↓5.3)	107.5 (430.0)	**77.9** (↓9.6)	**98.6** (↓1.1)	**99.8** (↓0.2)	96.8 (306.3)	**53.8** (↓7.2)	**82.7** (↓7.5)	**92.0** (↓4.6)	80.6 (322.5)
Retrain	72.7	96.7	99.3	537.5	87.5	99.7	100.0	403.1	61.0	90.2	96.6	403.1

Table 14: Main comparison on MSCOCO with SigLIP across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	56.7 (↓56.5)	87.9 (↓87.3)	92.3 (↓90.7)	50.7 (0.7)	39.2 (↓39.2)	71.2 (↓70.7)	81.4 (↓80.4)	55.8 (5.8)	0.3 (↓0.1)	0.8 (↑0.4)	**1.1** (↑1.3)	40.9 (9.1)
FedRecover	52.0 (↓51.8)	79.0 (↓78.4)	86.9 (↓85.3)	**49.6** (0.4)	48.1 (↓48.1)	75.1 (↓74.6)	84.2 (↓83.2)	53.5 (3.5)	0.2	1.1 (↑0.1)	2.3 (↑0.1)	41.2 (8.8)
Ferrari	69.7 (↓69.5)	93.5 (↓92.9)	97.3 (↓95.7)	52.6 (2.6)	66.7 (↓66.7)	93.2 (↓92.7)	97.7 (↓96.7)	56.4 (6.4)	**0.1** (↑0.1)	0.8 (↑0.4)	1.9 (↑0.5)	40.5 (9.5)
FedOSD	66.8 (↓66.6)	92.8 (↓92.2)	96.9 (↓95.3)	52.1 (2.1)	62.1 (↓62.1)	90.7 (↓90.2)	96.8 (↓95.8)	56.2 (6.2)	0.1 (↑0.1)	0.9 (↑0.3)	2.1 (↑0.3)	41.0 (9.0)
NoT	44.2 (↓44.0)	74.4 (↓73.8)	83.8 (↓82.2)	51.4 (1.4)	49.7 (↓49.7)	80.9 (↓80.4)	89.3 (↓88.3)	51.6 (1.6)	0.2	1.0 (↑0.2)	2.0 (↑0.4)	42.1 (7.9)
SoUL	63.0 (↓62.8)	87.7 (↓87.1)	93.8 (↓92.2)	51.5 (1.5)	63.1 (↓63.1)	89.8 (↓89.3)	95.5 (↓94.5)	53.9 (3.9)	0.1 (↑0.1)	**0.7** (↑0.5)	1.8 (↑0.6)	40.3 (9.7)
FFMU	28.5 (↓28.3)	54.9 (↓54.3)	67.3 (↓65.7)	**50.4** (0.4)	42.0 (↓42.0)	70.3 (↓69.8)	80.7 (↓79.7)	**50.2** (0.2)	0.3 (↓0.1)	1.2	2.5 (↓0.1)	**43.5** (6.5)
FUSED	69.0 (↓68.8)	92.7 (↓92.1)	96.9 (↓95.3)	52.5 (2.5)	64.7 (↓64.7)	91.5 (↓91.0)	96.4 (↓95.4)	56.0 (6.0)	0.1 (↑0.1)	0.8 (↑0.4)	1.7 (↑0.7)	40.7 (9.3)
GradAscent	65.4 (↓65.2)	92.7 (↓92.1)	97.0 (↓95.4)	51.1 (1.1)	12.7 (↓12.7)	30.6 (↓30.1)	43.3 (↓42.3)	51.2 (1.2)	0.1 (↑0.1)	0.7 (↑0.5)	1.6 (↑0.8)	40.9 (9.1)
EASE	**0.0** (↑0.2)	**0.6**	**2.2** (↓0.6)	40.9 (9.1)	**0.0**	**0.7** (↓0.2)	**1.7** (↓0.7)	40.9 (9.1)	0.6 (↓0.4)	2.6 (↓1.4)	6.1 (↓3.7)	40.9 (9.1)
Retrain	0.2	0.6	1.6	40.9	0.0	0.5	1.0	41.0	0.2	1.2	2.4	40.9

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	67.7 (↓4.8)	96.3 (↓1.9)	98.9 (↓0.6)	475.0	74.7 (↓11.4)	97.5 (↓1.9)	99.6 (↓0.4)	356.2	66.7 (↓1.5)	92.5	96.9 (↓1.2)	356.2
FedRecover	66.0 (↓6.5)	95.8 (↓2.4)	98.7 (↓0.8)	142.5 (332.5)	72.5 (↓13.6)	97.2 (↓2.2)	99.2 (↓0.8)	**64.1** (292.1)	64.1 (↓4.1)	91.7 (↓0.8)	96.7 (↓1.4)	**53.8** (302.4)
Ferrari	67.4 (↓5.1)	96.1 (↓2.1)	98.9 (↓0.6)	130.6 (344.4)	74.3 (↓11.8)	97.6 (↓1.8)	99.5 (↓0.5)	106.9 (249.3)	66.3 (↓1.9)	92.4 (↓0.1)	96.9 (↓1.2)	71.3 (284.9)
FedOSD	58.6 (↓13.9)	91.1 (↓7.1)	99.0 (↓0.5)	190.0 (285.0)	75.3 (↓10.8)	97.4 (↓2.0)	99.5 (↓0.5)	142.5 (213.7)	67.4 (↓0.8)	92.2 (↓0.3)	97.0 (↓1.1)	142.5 (213.7)
NoT	51.7 (↓20.8)	79.5 (↓18.7)	88.1 (↓11.4)	**71.3** (403.7)	73.1 (↓13.0)	97.1 (↓2.3)	99.3 (↓0.7)	71.2 (285.0)	55.1 (↓13.1)	82.3 (↓10.2)	90.7 (↓7.4)	71.3 (284.9)
SoUL	67.0 (↓5.5)	91.6 (↓6.6)	98.9 (↓0.6)	71.3 (403.7)	74.3 (↓11.8)	97.5 (↓1.9)	99.4 (↓0.6)	71.2 (285.0)	66.7 (↓1.5)	92.3 (↓0.2)	96.8 (↓1.3)	71.3 (284.9)
FFMU	32.4 (↓40.1)	59.2 (↓39.0)	70.7 (↓28.8)	142.5 (332.5)	64.1 (↓22.0)	96.7 (↓2.7)	99.1 (↓0.9)	142.5 (213.7)	32.8 (↓35.4)	60.5 (↓32.0)	73.2 (↓24.9)	142.5 (213.7)
FUSED	67.0 (↓5.5)	96.2 (↓2.0)	99.0 (↓0.5)	190.0 (285.0)	73.9 (↓12.2)	97.8 (↓1.6)	99.6 (↓0.4)	142.5 (213.7)	67.7 (↓0.5)	92.4 (↓0.1)	97.0 (↓1.1)	142.5 (213.7)
GradAscent	65.2 (↓7.3)	95.3 (↓2.9)	98.5 (↓1.0)	130.6 (344.4)	46.3 (↓39.8)	84.7 (↓14.7)	94.8 (↓5.2)	106.9 (249.3)	64.9 (↓3.3)	91.7 (↓0.8)	96.6 (↓1.5)	106.9 (249.3)
EASE	**68.4** (↓4.1)	**96.5** (↓1.7)	**99.1** (↓0.4)	95.0 (380.0)	**75.7** (↓10.4)	**97.9** (↓1.5)	**99.7** (↓0.3)	85.5 (270.7)	**68.1** (↓0.1)	**92.7** (↑0.2)	**97.1** (↓1.0)	71.2 (285.0)
Retrain	72.5	98.2	99.5	475.0	86.1	99.4	100.0	356.2	68.2	92.5	98.1	356.2

Table 15: Main comparison on TextCaps with CLIP-B/32 across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	60.3 (↓60.1)	85.2 (↓84.7)	89.6 (↓88.1)	63.0 (13.0)	22.0 (↓22.0)	44.3 (↓42.4)	53.5 (↓50.3)	55.3 (5.3)	0.1 (↓0.1)	0.7 (↑0.5)	2.4 (↓0.4)	51.3 (1.3)
FedRecover	43.7 (↓43.5)	70.6 (↓70.1)	78.7 (↓77.2)	62.5 (12.5)	18.8 (↓18.8)	43.5 (↓41.6)	53.8 (↓50.6)	54.9 (4.9)	0.1 (↓0.1)	0.8 (↑0.4)	1.7 (↑0.3)	**50.0** (0.0)
Ferrari	83.3 (↓83.1)	98.2 (↓97.7)	98.8 (↓97.3)	63.2 (13.2)	76.6 (↓76.6)	95.5 (↓93.6)	99.1 (↓95.9)	56.6 (6.6)	0.1 (↓0.1)	0.7 (↑0.5)	1.8 (↑0.2)	49.8 (0.2)
FedOSD	80.4 (↓80.2)	97.4 (↓96.9)	98.7 (↓97.2)	63.1 (13.1)	43.2 (↓43.2)	72.8 (↓70.9)	85.6 (↓82.4)	55.3 (5.3)	0.1 (↓0.1)	1.0 (↑0.2)	1.7 (↑0.3)	50.8 (0.8)
NoT	43.0 (↓42.8)	77.3 (↓76.8)	86.5 (↓85.0)	63.2 (13.2)	11.1 (↓11.1)	25.5 (↓23.6)	38.6 (↓35.4)	55.3 (5.3)	0.1 (↓0.1)	1.0 (↑0.2)	2.5 (↓0.5)	47.9 (2.1)
SoUL	73.5 (↓73.3)	93.9 (↓93.4)	97.8 (↓96.3)	63.2 (13.2)	30.7 (↓30.7)	61.7 (↓59.8)	73.6 (↓70.4)	55.3 (5.3)	**0.0**	0.8 (↑0.4)	1.8 (↑0.2)	49.4 (0.6)
FFMU	19.5 (↓19.3)	51.0 (↓50.5)	65.3 (↓63.8)	62.9 (12.9)	3.5 (↓3.5)	14.7 (↓12.8)	22.3 (↓19.1)	55.3 (5.3)	0.3 (↓0.3)	1.0 (↑0.2)	1.6 (↑0.4)	45.4 (4.6)
FUSED	78.3 (↓78.1)	97.1 (↓96.9)	98.4 (↓96.9)	63.2 (13.2)	40.8 (↓40.8)	72.0 (↓70.1)	85.3 (↓82.1)	55.3 (5.3)	0.1 (↓0.1)	**0.6** (↑0.6)	**1.5** (↑0.5)	49.9 (0.1)
GradAscent	1.7 (↓1.5)	8.6 (↓8.1)	14.7 (↓13.2)	60.9 (10.9)	39.9 (↓39.9)	71.5 (↓69.6)	81.3 (↓78.1)	56.6 (6.6)	0.2 (↓0.2)	0.8 (↑0.4)	2.2 (↓0.2)	49.9 (0.1)
EASE	**0.2**	**0.9** (↓0.4)	**1.9** (↓0.4)	**51.9** (1.9)	**0.0**	**1.7** (↑0.2)	**3.0** (↑0.2)	**53.5** (3.5)	0.2 (↓0.2)	1.4 (↓0.2)	2.3 (↓0.3)	49.2 (0.8)
Retrain	0.2	0.5	1.5	49.3	0.0	1.9	3.2	51.7	0.0	1.2	2.0	48.7

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	91.6 (↓4.9)	98.9 (↓1.1)	99.4 (↓0.6)	262.5	82.3 (↓10.0)	94.6 (↓5.4)	96.5 (↓3.5)	196.9	84.4 (↓2.4)	97.8 (↓1.2)	99.4 (↓0.4)	196.9
FedRecover	66.3 (↓30.2)	91.9 (↓8.1)	99.1 (↓0.9)	**47.2** (215.3)	71.1 (↓21.2)	92.5 (↓7.5)	96.5 (↓3.5)	**35.4** (161.5)	42.9 (↓43.9)	69.5 (↓29.5)	77.0 (↓22.8)	**35.4** (161.5)
Ferrari	79.8 (↓16.7)	98.9 (↓1.1)	99.4 (↓0.6)	72.2 (190.3)	85.7 (↓6.6)	99.1 (↓0.9)	99.6 (↓0.4)	59.1 (137.8)	83.7 (↓3.1)	98.0 (↓1.0)	99.2 (↓0.6)	59.1 (137.8)
FedOSD	64.9 (↓31.6)	90.7 (↓9.3)	99.5 (↓0.5)	105.0 (157.5)	86.5 (↓5.8)	98.9 (↓1.1)	99.5 (↓0.5)	78.8 (118.1)	83.3 (↓3.5)	97.9 (↓1.1)	99.3 (↓0.5)	78.8 (118.1)
NoT	31.1 (↓65.4)	63.2 (↓36.8)	78.0 (↓22.0)	52.5 (210.0)	51.7 (↓40.6)	86.9 (↓13.1)	96.2 (↓3.8)	39.4 (157.5)	44.3 (↓42.5)	80.2 (↓18.8)	89.5 (↓10.3)	39.4 (157.5)
SoUL	66.5 (↓30.0)	91.9 (↓8.1)	96.4 (↓3.6)	52.5 (210.0)	86.1 (↓6.2)	99.0 (↓1.0)	99.6 (↓0.4)	39.4 (157.5)	83.7 (↓3.1)	97.9 (↓1.1)	99.3 (↓0.5)	39.4 (157.5)
FFMU	15.1 (↓81.4)	38.4 (↓61.6)	54.4 (↓45.6)	105.0 (157.5)	27.0 (↓65.3)	61.9 (↓38.1)	76.5 (↓23.5)	78.8 (118.1)	21.3 (↓65.5)	52.1 (↓46.9)	67.6 (↓32.2)	78.8 (118.1)
FUSED	90.5 (↓6.0)	98.8 (↓1.2)	99.5 (↓0.5)	105.0 (157.5)	86.1 (↓6.2)	99.1 (↓0.9)	99.7 (↓0.3)	78.8 (118.1)	84.7 (↓2.1)	98.2 (↓0.8)	99.3 (↓0.5)	78.8 (118.1)
GradAscent	1.9 (↓94.6)	6.8 (↓93.2)	13.2 (↓86.8)	72.2 (190.3)	62.5 (↓29.8)	89.8 (↓10.2)	95.1 (↓4.9)	59.1 (137.8)	3.1 (↓83.7)	10.0 (↓89.0)	17.3 (↓82.5)	59.1 (137.8)
EASE	**92.3** (↓4.2)	**99.2** (↓0.8)	**99.6** (↓0.4)	52.5 (210.0)	**87.5** (↓4.8)	**99.4** (↓0.6)	**99.8** (↓0.2)	47.2 (149.7)	**85.1** (↓1.7)	**98.3** (↓0.7)	**99.5** (↓0.3)	39.4 (157.5)
Retrain	96.5	100.0	100.0	262.5	92.3	100.0	100.0	196.9	86.8	99.0	99.8	196.9

Table 16: Main comparison on TextCaps with CLIP-L/14 across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	66.4 (↓66.4)	85.0 (↓84.3)	89.5 (↓87.8)	63.1 (13.1)	54.9 (↓54.6)	75.6 (↓74.0)	82.2 (↓79.9)	63.2 (13.2)	**0.0** (↑0.3)	**0.7** (↑0.9)	**1.4** (↑0.8)	49.1 (0.9)
FedRecover	60.7 (↓60.7)	80.4 (↓79.7)	86.0 (↓84.3)	62.5 (12.5)	36.0 (↓35.7)	59.8 (↓58.2)	67.0 (↓64.7)	62.8 (12.8)	0.3	0.7 (↑0.9)	1.5 (↑0.7)	47.6 (2.4)
Ferrari	92.7 (↓92.7)	99.7 (↓99.0)	100.0 (↓98.3)	63.2 (13.2)	76.8 (↓76.5)	95.1 (↓93.5)	98.3 (↓96.0)	63.2 (13.2)	0.4 (↓0.1)	1.1 (↑0.5)	2.1 (↑0.1)	49.7 (0.3)
FedOSD	91.2 (↓91.2)	99.5 (↓98.8)	100.0 (↓98.3)	63.2 (13.2)	73.0 (↓72.7)	94.2 (↓92.6)	97.6 (↓95.3)	63.2 (13.2)	0.4 (↓0.1)	1.3 (↑0.3)	2.9 (↓0.7)	**50.2** (0.2)
NoT	51.5 (↓51.5)	80.7 (↓80.0)	89.9 (↓88.2)	63.2 (13.2)	28.4 (↓28.1)	53.4 (↓51.8)	65.7 (↓63.4)	63.2 (13.2)	0.0 (↑0.3)	0.8 (↑0.8)	1.7 (↑0.5)	48.5 (1.5)
SoUL	84.4 (↓84.4)	96.9 (↓96.2)	98.8 (↓97.1)	63.2 (13.2)	64.4 (↓64.1)	86.8 (↓85.2)	93.7 (↓91.4)	63.2 (13.2)	0.6 (↓0.3)	1.4 (↑0.2)	2.3 (↓0.1)	50.4 (0.4)
FFMU	24.2 (↓24.2)	54.3 (↓53.6)	68.1 (↓66.4)	63.2 (13.2)	11.0 (↓10.7)	26.7 (↓25.1)	36.9 (↓34.6)	63.1 (13.1)	0.2 (↑0.1)	1.3 (↑0.3)	2.3 (↓0.1)	48.5 (1.5)
FUSED	89.4 (↓89.4)	98.6 (↓97.9)	99.9 (↓98.2)	63.2 (13.2)	70.2 (↓69.9)	92.2 (↓90.6)	96.5 (↓94.2)	63.2 (13.2)	0.4 (↓0.1)	1.2 (↑0.4)	2.4 (↓0.2)	49.7 (0.3)
GradAscent	0.5 (↓0.5)	3.9 (↓3.2)	7.5 (↓5.8)	63.0 (13.0)	**0.7** (↓0.4)	2.7 (↓1.1)	5.3 (↓3.0)	**61.7** (11.7)	0.0 (↑0.3)	0.8 (↑0.8)	1.7 (↑0.5)	47.0 (3.0)
EASE	**0.3** (↓0.3)	**1.9** (↓1.2)	**3.9** (↓2.2)	**53.3** (3.3)	0.7 (↓0.4)	**1.8** (↓0.2)	**3.4** (↓1.1)	62.0 (12.0)	0.5 (↓0.2)	1.9 (↓0.3)	2.6 (↓0.4)	52.8 (2.8)
Retrain	0.0	0.7	1.7	50.3	0.3	1.6	2.3	61.3	0.3	1.6	2.2	52.2

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	90.9 (↓5.1)	99.1 (↓0.9)	99.6 (↓0.4)	537.5	87.6 (↓10.9)	98.4 (↓1.6)	99.3 (↓0.7)	403.1	91.1 (↓2.8)	98.8 (↓0.7)	99.7 (↓0.2)	403.1
FedRecover	87.6 (↓8.4)	98.3 (↓1.7)	99.3 (↓0.7)	**96.8** (440.7)	84.6 (↓13.9)	97.6 (↓2.4)	99.1 (↓0.9)	**72.6** (330.5)	63.1 (↓30.8)	82.4 (↓17.1)	87.6 (↓12.3)	**72.6** (330.5)
Ferrari	90.2 (↓5.8)	99.0 (↓1.0)	99.6 (↓0.4)	147.8 (389.7)	87.6 (↓10.9)	98.5 (↓1.5)	99.2 (↓0.8)	120.9 (282.2)	91.1 (↓2.8)	98.9 (↓0.6)	99.6 (↓0.3)	120.9 (282.2)
FedOSD	90.9 (↓5.1)	98.8 (↓1.2)	99.5 (↓0.5)	215.0 (322.5)	86.8 (↓11.7)	98.3 (↓1.7)	99.3 (↓0.7)	161.2 (241.9)	91.4 (↓2.5)	99.1 (↓0.4)	99.7 (↓0.2)	161.2 (241.9)
NoT	63.6 (↓32.4)	92.2 (↓7.8)	99.4 (↓0.6)	107.5 (430.0)	75.6 (↓22.9)	98.0 (↓2.0)	99.1 (↓0.9)	80.6 (322.5)	62.5 (↓31.4)	89.7 (↓9.8)	96.3 (↓3.6)	80.6 (322.5)
SoUL	90.2 (↓5.8)	98.9 (↓1.1)	99.6 (↓0.4)	107.5 (430.0)	87.2 (↓11.3)	98.4 (↓1.6)	99.3 (↓0.7)	80.6 (322.5)	90.7 (↓3.2)	98.7 (↓0.8)	99.6 (↓0.3)	80.6 (322.5)
FFMU	36.9 (↓59.1)	68.9 (↓31.1)	81.2 (↓18.8)	215.0 (322.5)	49.3 (↓49.2)	82.1 (↓17.9)	90.1 (↓9.9)	161.2 (241.9)	39.7 (↓54.2)	73.1 (↓26.4)	83.3 (↓16.6)	161.2 (241.9)
FUSED	90.6 (↓5.4)	99.2 (↓0.8)	99.6 (↓0.4)	215.0 (322.5)	87.2 (↓11.3)	98.7 (↓1.3)	99.4 (↓0.6)	161.2 (241.9)	90.3 (↓3.6)	99.0 (↓0.5)	99.5 (↓0.4)	161.2 (241.9)
GradAscent	0.7 (↓95.3)	4.0 (↓96.0)	7.2 (↓92.8)	147.8 (389.7)	1.1 (↓97.4)	4.2 (↓95.8)	10.7 (↓89.3)	120.9 (282.2)	9.9 (↓84.0)	26.1 (↓73.4)	40.9 (↓59.0)	120.9 (282.2)
EASE	**91.6** (↓4.4)	**99.3** (↓0.7)	**99.8** (↓0.2)	107.5 (430.0)	**88.6** (↓9.9)	**98.8** (↓1.2)	**99.5** (↓0.5)	96.8 (306.3)	**92.1** (↓1.8)	**99.2** (↓0.3)	**99.8** (↓0.1)	80.6 (322.5)
Retrain	96.0	100.0	100.0	537.5	98.5	100.0	100.0	403.1	93.9	99.5	99.9	403.1

Table 17: Main comparison on TextCaps with SigLIP across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	*MIA*
FedEraser	67.5 (↓67.5)	83.8 (↓82.9)	90.0 (↓88.1)	63.1 (13.1)	66.1 (↓66.0)	83.4 (↓82.0)	87.4 (↓84.8)	61.4 (11.4)	0.3	1.0 (↓0.2)	1.8 (↑0.2)	50.4 (0.4)
FedRecover	56.3 (↓56.3)	72.0 (↓71.1)	78.7 (↓76.8)	62.9 (12.9)	57.1 (↓57.0)	74.6 (↓73.2)	79.5 (↓76.9)	60.1 (10.1)	0.2 (↑0.1)	1.6 (↓0.8)	2.3 (↓0.3)	48.2 (1.8)
Ferrari	94.5 (↓94.5)	99.7 (↓98.8)	100.0 (↓98.1)	63.2 (13.2)	96.8 (↓96.7)	99.9 (↓98.5)	100.0 (↓97.4)	61.4 (11.4)	0.3	1.1 (↓0.3)	1.8 (↑0.2)	**50.0** (0.0)
FedOSD	92.5 (↓92.5)	99.7 (↓98.8)	100.0 (↓98.1)	63.2 (13.2)	95.3 (↓95.2)	99.6 (↓98.2)	100.0 (↓97.4)	61.4 (11.4)	**0.0** (↑0.3)	0.9 (↓0.1)	2.1 (↓0.1)	**50.0** (0.0)
NoT	54.7 (↓54.7)	81.3 (↓80.4)	88.7 (↓86.8)	63.2 (13.2)	49.6 (↓49.5)	79.7 (↓78.3)	87.1 (↓84.5)	61.4 (11.4)	0.0 (↑0.3)	0.9 (↓0.1)	1.7 (↑0.3)	49.6 (0.4)
SoUL	83.5 (↓83.5)	95.7 (↓94.8)	98.8 (↓96.9)	63.2 (13.2)	86.3 (↓86.2)	97.9 (↓96.5)	99.2 (↓96.6)	61.4 (11.4)	0.3	1.0 (↓0.2)	1.9 (↑0.1)	50.4 (0.4)
FFMU	35.9 (↓35.9)	64.5 (↓63.6)	75.5 (↓73.6)	63.1 (13.1)	29.4 (↓29.3)	57.0 (↓55.6)	68.8 (↓66.2)	61.3 (11.3)	0.3	**0.6** (↑0.2)	1.3 (↑0.7)	49.9 (0.1)
FUSED	91.5 (↓91.5)	99.2 (↓98.3)	99.8 (↓97.9)	63.2 (13.2)	94.7 (↓94.6)	99.6 (↓98.2)	100.0 (↓97.4)	61.4 (11.4)	0.1 (↑0.2)	1.0 (↓0.2)	1.6 (↑0.4)	49.6 (0.4)
GradAscent	84.0 (↓84.0)	96.8 (↓95.9)	98.5 (↓96.6)	63.2 (13.2)	86.3 (↓86.2)	97.2 (↓95.8)	98.7 (↓96.1)	61.4 (11.4)	0.0 (↑0.3)	0.7 (↑0.1)	**1.2** (↑0.8)	**50.0** (0.0)
EASE	**0.0**	**1.1** (↓0.2)	**2.3** (↓0.4)	**52.9** (2.9)	**0.0** (↑0.1)	**1.3** (↑0.1)	**2.1** (↑0.5)	**59.0** (9.0)	0.2 (↑0.1)	0.9 (↓0.1)	2.1 (↓0.1)	51.2 (1.2)
Retrain	0.0	0.9	1.9	49.9	0.1	1.4	2.6	55.7	0.3	0.8	2.0	51.0

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*Comm.* ↓
FedEraser	94.5 (↓5.4)	99.4 (↓0.6)	99.8 (↓0.2)	475.0	93.1 (↓5.4)	99.0 (↓1.0)	99.7 (↓0.3)	356.2	97.2 (↓1.5)	99.7 (↓0.3)	99.9 (↓0.1)	356.2
FedRecover	92.0 (↓7.9)	98.4 (↓1.6)	99.3 (↓0.7)	**85.5** (389.5)	92.1 (↓6.4)	98.5 (↓1.5)	99.3 (↓0.7)	**64.1** (292.1)	57.9 (↓40.8)	76.7 (↓23.3)	83.3 (↓16.7)	**64.1** (292.1)
Ferrari	93.4 (↓6.5)	99.3 (↓0.7)	99.6 (↓0.4)	130.6 (344.4)	92.7 (↓5.8)	99.2 (↓0.8)	99.5 (↓0.5)	106.9 (249.3)	96.9 (↓1.8)	99.6 (↓0.4)	99.7 (↓0.3)	106.9 (249.3)
FedOSD	93.8 (↓6.1)	99.1 (↓0.9)	99.8 (↓0.2)	190.0 (285.0)	94.1 (↓4.4)	99.4 (↓0.6)	99.7 (↓0.3)	142.5 (213.7)	96.5 (↓2.2)	99.5 (↓0.5)	99.8 (↓0.2)	142.5 (213.7)
NoT	64.2 (↓35.7)	90.5 (↓9.5)	94.3 (↓5.7)	95.0 (380.0)	73.8 (↓24.7)	98.7 (↓1.3)	99.4 (↓0.6)	71.2 (285.0)	68.6 (↓30.1)	93.5 (↓6.5)	99.6 (↓0.4)	71.2 (285.0)
SoUL	93.8 (↓6.1)	99.2 (↓0.8)	99.7 (↓0.3)	95.0 (380.0)	93.1 (↓5.4)	99.1 (↓0.9)	99.6 (↓0.4)	71.2 (285.0)	96.5 (↓2.2)	99.5 (↓0.5)	99.8 (↓0.2)	71.2 (285.0)
FFMU	42.8 (↓57.1)	71.9 (↓28.1)	81.3 (↓18.7)	190.0 (285.0)	48.9 (↓49.6)	82.2 (↓17.8)	92.9 (↓7.1)	142.5 (213.7)	43.3 (↓55.4)	77.6 (↓22.4)	89.2 (↓10.8)	142.5 (213.7)
FUSED	94.2 (↓5.7)	99.5 (↓0.5)	99.8 (↓0.2)	190.0 (285.0)	94.1 (↓4.4)	99.1 (↓0.9)	99.6 (↓0.4)	142.5 (213.7)	96.1 (↓2.6)	99.4 (↓0.6)	99.9 (↓0.1)	142.5 (213.7)
GradAscent	91.2 (↓8.7)	98.6 (↓1.4)	99.4 (↓0.6)	130.6 (344.4)	90.5 (↓8.0)	98.3 (↓1.7)	99.3 (↓0.7)	106.9 (249.3)	93.9 (↓4.8)	98.9 (↓1.1)	99.5 (↓0.5)	106.9 (249.3)
EASE	**95.2** (↓4.7)	**99.6** (↓0.4)	**99.9** (↓0.1)	95.0 (380.0)	**94.5** (↓4.0)	**99.5** (↓0.5)	**99.8** (↓0.2)	85.5 (270.7)	**97.9** (↓0.8)	**99.9** (↓0.1)	**100.0**	71.2 (285.0)
Retrain	99.9	100.0	100.0	475.0	98.5	100.0	100.0	356.2	98.7	100.0	100.0	356.2

Table 18: Main comparison on Flickr30K with BLIP-ITM across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	F-R@1 ↓	*F-R@5* ↓	*F-R@10* ↓	ρ ↓	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	ρ ↓	*F-R@1* ↓	*F-R@5* ↓	*F-R@10* ↓	ρ ↓
FedEraser	61.8 (↓61.6)	79.9 (↓78.8)	88.9 (↓86.7)	0.65 (↓0.61)	64.2 (↓60.5)	84.3 (↓79.0)	91.6 (↓85.7)	0.67 (↓0.62)	58.9 (↓58.8)	78.3 (↓77.7)	88.4 (↓86.9)	0.58 (↓0.55)
FedRecover	56.4 (↓56.2)	73.3 (↓72.2)	81.7 (↓79.5)	0.57 (↓0.53)	59.5 (↓55.8)	76.8 (↓71.5)	85.1 (↓79.2)	0.64 (↓0.59)	53.4 (↓53.3)	71.2 (↓70.6)	80.1 (↓78.6)	0.55 (↓0.52)
Ferrari	93.9 (↓93.7)	98.6 (↓97.5)	99.8 (↓97.6)	0.69 (↓0.65)	97.9 (↓94.2)	100.0 (↓94.7)	100.0 (↓94.1)	0.67 (↓0.62)	90.2 (↓90.1)	97.7 (↓97.1)	98.7 (↓97.2)	0.64 (↓0.61)
FedOSD	92.5 (↓92.3)	97.6 (↓96.5)	99.0 (↓96.8)	0.61 (↓0.57)	94.8 (↓91.1)	100.0 (↓94.7)	100.0 (↓94.1)	0.66 (↓0.61)	88.5 (↓88.4)	96.1 (↓95.5)	97.7 (↓96.2)	0.57 (↓0.54)
NoT	54.7 (↓54.5)	82.1 (↓81.0)	89.6 (↓87.4)	0.66 (↓0.62)	56.7 (↓53.0)	86.1 (↓80.8)	93.2 (↓87.3)	0.70 (↓0.65)	52.3 (↓52.2)	78.4 (↓77.8)	87.6 (↓86.1)	0.64 (↓0.61)
SoUL	84.8 (↓84.6)	96.9 (↓95.8)	99.7 (↓97.5)	0.66 (↓0.62)	88.4 (↓84.7)	100.0 (↓94.7)	100.0 (↓94.1)	0.71 (↓0.66)	81.9 (↓81.8)	94.3 (↓93.7)	96.6 (↓95.1)	0.62 (↓0.59)
FFMU	34.9 (↓34.7)	65.4 (↓64.3)	76.1 (↓73.9)	0.49 (↓0.45)	40.1 (↓36.4)	67.0 (↓61.7)	79.3 (↓73.4)	0.55 (↓0.50)	33.8 (↓33.7)	61.0 (↓60.4)	75.2 (↓73.7)	0.50 (↓0.47)
FUSED	91.9 (↓91.7)	97.4 (↓96.3)	98.6 (↓96.4)	0.62 (↓0.58)	95.7 (↓92.0)	100.0 (↓94.7)	100.0 (↓94.1)	0.72 (↓0.67)	89.1 (↓89.0)	95.6 (↓95.0)	98.1 (↓96.6)	0.61 (↓0.58)
GradAscent	84.3 (↓84.1)	95.7 (↓94.6)	98.6 (↓96.4)	0.67 (↓0.63)	88.5 (↓84.8)	100.0 (↓94.7)	100.0 (↓94.1)	0.71 (↓0.66)	83.6 (↓83.5)	95.2 (↓94.6)	96.9 (↓95.4)	0.67 (↓0.64)
EASE	**1.5** (↓1.3)	**2.3** (↓1.2)	**3.8** (↓1.6)	**0.19** (↓0.15)	**4.7** (↓1.0)	**5.7** (↓0.4)	**7.0** (↓1.1)	**0.20** (↓0.15)	**1.3** (↓1.2)	**1.9** (↓1.3)	**3.2** (↓1.7)	**0.11** (↓0.08)
Retrain	0.2	1.1	2.2	0.04	3.7	5.3	5.9	0.05	0.1	0.6	1.5	0.03

Method	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*CIDEr* ↑	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*CIDEr* ↑	*R-R@1* ↑	*R-R@5* ↑	*R-R@10* ↑	*CIDEr* ↑
FedEraser	79.0 (↓8.4)	94.9 (↓4.3)	98.0 (↓1.8)	51.05 (↓14.05)	78.6 (↓7.4)	93.0 (↓5.6)	96.7 (↓2.4)	46.86 (↓17.14)	81.9 (↓6.7)	94.7 (↓4.7)	98.7 (↓1.2)	49.38 (↓16.82)
FedRecover	76.6 (↓10.8)	93.6 (↓5.6)	97.4 (↓2.4)	48.19 (↓16.91)	74.8 (↓11.2)	90.6 (↓8.0)	96.0 (↓3.1)	45.54 (↓18.46)	78.6 (↓10.0)	93.8 (↓5.6)	97.8 (↓2.1)	46.26 (↓19.94)
Ferrari	83.1 (↓4.3)	97.0 (↓2.2)	99.2 (↓0.6)	54.14 (↓10.96)	80.8 (↓5.2)	96.6 (↓2.0)	97.8 (↓1.3)	48.50 (↓15.50)	83.9 (↓4.7)	97.5 (↓1.9)	**99.8** (↓0.1)	54.60 (↓11.60)
FedOSD	81.0 (↓6.4)	96.7 (↓2.5)	98.9 (↓0.9)	49.78 (↓15.32)	79.8 (↓6.2)	95.8 (↓2.8)	97.6 (↓1.5)	50.10 (↓13.90)	81.9 (↓6.7)	97.3 (↓2.1)	99.3 (↓0.6)	52.23 (↓13.97)
NoT	77.5 (↓9.9)	92.5 (↓6.7)	96.4 (↓3.4)	49.93 (↓15.17)	74.8 (↓11.2)	91.3 (↓7.3)	95.2 (↓3.9)	46.31 (↓17.69)	78.9 (↓9.7)	95.2 (↓4.2)	98.1 (↓1.8)	50.28 (↓15.92)
SoUL	80.5 (↓6.9)	96.5 (↓2.7)	98.4 (↓1.4)	52.36 (↓12.74)	78.3 (↓7.7)	94.1 (↓4.5)	97.3 (↓1.8)	49.90 (↓14.10)	82.2 (↓6.4)	97.1 (↓2.3)	99.5 (↓0.4)	54.89 (↓11.31)
FFMU	69.8 (↓17.6)	86.3 (↓12.9)	94.6 (↓5.2)	44.22 (↓20.88)	72.6 (↓13.4)	89.3 (↓9.3)	92.2 (↓6.9)	39.30 (↓24.70)	73.6 (↓15.0)	90.4 (↓9.0)	94.0 (↓5.9)	46.10 (↓20.10)
FUSED	82.2 (↓5.2)	96.8 (↓2.4)	98.7 (↓1.1)	51.67 (↓13.43)	78.6 (↓7.4)	96.3 (↓2.3)	97.1 (↓2.0)	50.92 (↓13.08)	83.0 (↓5.6)	96.5 (↓2.9)	99.4 (↓0.5)	52.35 (↓13.85)
GradAscent	79.8 (↓7.6)	94.3 (↓4.9)	98.7 (↓1.1)	45.54 (↓19.56)	76.7 (↓9.3)	93.4 (↓5.2)	95.9 (↓3.2)	42.09 (↓21.91)	81.0 (↓7.6)	96.4 (↓3.0)	98.1 (↓1.8)	51.37 (↓14.83)
EASE	**86.6** (↓0.8)	**98.8** (↓0.4)	**99.2** (↓0.6)	**59.67** (↓5.43)	**84.1** (↓1.9)	**98.3** (↓0.3)	**98.5** (↓0.6)	**59.16** (↓4.84)	**86.6** (↓2.0)	**98.5** (↓0.9)	99.2 (↓0.7)	**62.46** (↓3.74)
Retrain	87.4	99.2	99.8	65.10	86.0	98.6	99.1	64.00	88.6	99.4	99.9	66.20

Table 19: Main comparison on Flickr30K with BLIP-2 across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$
FedEraser	62.4 (↓61.0)	81.1 (↓78.6)	89.9 (↓86.5)	0.62 (↓0.54)	65.3 (↓59.9)	85.7 (↓79.0)	93.3 (↓85.7)	0.68 (↓0.59)	59.1 (↓58.4)	79.7 (↓78.4)	87.1 (↓84.8)	0.58 (↓0.52)
FedRecover	57.4 (↓56.0)	73.7 (↓71.2)	81.1 (↓77.7)	0.58 (↓0.50)	60.0 (↓54.6)	78.1 (↓71.4)	83.8 (↓76.2)	0.63 (↓0.54)	54.5 (↓53.8)	70.9 (↓69.6)	81.2 (↓78.9)	0.54 (↓0.48)
Ferrari	94.9 (↓93.5)	99.5 (↓97.0)	100.0 (↓96.6)	0.67 (↓0.59)	97.5 (↓92.1)	100.0 (↓93.3)	100.0 (↓92.4)	0.69 (↓0.60)	92.7 (↓92.0)	97.3 (↓96.0)	99.2 (↓96.9)	0.65 (↓0.59)
FedOSD	93.2 (↓91.8)	98.6 (↓96.1)	100.0 (↓96.6)	0.67 (↓0.59)	96.8 (↓91.4)	100.0 (↓93.3)	100.0 (↓92.4)	0.72 (↓0.63)	91.3 (↓90.6)	96.2 (↓94.9)	99.7 (↓97.4)	0.64 (↓0.58)
NoT	54.3 (↓52.9)	81.2 (↓78.7)	89.7 (↓86.3)	0.65 (↓0.57)	58.8 (↓53.4)	87.5 (↓80.8)	94.0 (↓86.4)	0.70 (↓0.61)	53.1 (↓52.4)	79.8 (↓78.5)	87.8 (↓85.5)	0.64 (↓0.58)
SoUL	85.6 (↓84.2)	97.1 (↓94.6)	100.0 (↓96.6)	0.65 (↓0.57)	90.4 (↓85.0)	100.0 (↓93.3)	100.0 (↓92.4)	0.69 (↓0.60)	83.4 (↓82.7)	96.3 (↓95.0)	97.8 (↓95.5)	0.60 (↓0.54)
FFMU	38.5 (↓37.1)	65.2 (↓62.7)	77.0 (↓73.6)	0.52 (↓0.44)	41.1 (↓35.7)	69.3 (↓62.6)	79.9 (↓72.3)	0.59 (↓0.50)	36.4 (↓35.7)	62.1 (↓60.8)	76.3 (↓74.0)	0.48 (↓0.42)
FUSED	91.7 (↓90.3)	99.6 (↓97.1)	100.0 (↓96.6)	0.65 (↓0.57)	97.1 (↓91.7)	100.0 (↓93.3)	100.0 (↓92.4)	0.72 (↓0.63)	89.2 (↓88.5)	97.1 (↓95.8)	98.5 (↓96.2)	0.61 (↓0.55)
GradAscent	86.5 (↓85.1)	97.5 (↓95.0)	98.6 (↓95.2)	0.72 (↓0.64)	90.5 (↓85.1)	100.0 (↓93.3)	100.0 (↓92.4)	0.72 (↓0.63)	85.0 (↓84.3)	95.0 (↓93.7)	98.3 (↓96.0)	0.66 (↓0.60)
EASE	**3.7** (↓2.3)	**4.3** (↓1.8)	**4.9** (↓1.5)	**0.26** (↓0.18)	**7.1** (↓1.7)	**8.5** (↓1.8)	**9.3** (↓1.7)	**0.26** (↓0.17)	**2.5** (↓1.8)	**3.1** (↓1.8)	**4.6** (↓2.3)	**0.18** (↓0.12)
Retrain	1.4	2.5	3.4	0.08	5.4	6.7	7.6	0.09	0.7	1.3	2.3	0.06
Method	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$
FedEraser	78.7 (↓7.1)	92.7 (↓5.6)	96.7 (↓2.4)	50.49 (↓13.31)	75.1 (↓9.5)	91.4 (↓6.3)	95.8 (↓2.7)	44.65 (↓18.05)	79.5 (↓7.4)	94.8 (↓3.9)	98.1 (↓1.2)	51.85 (↓12.75)
FedRecover	75.3 (↓10.5)	92.6 (↓5.7)	95.4 (↓3.7)	44.49 (↓19.31)	75.6 (↓9.0)	90.4 (↓7.3)	94.8 (↓3.7)	42.77 (↓19.93)	79.5 (↓7.4)	92.0 (↓6.7)	96.2 (↓3.1)	47.30 (↓17.30)
Ferrari	80.7 (↓5.1)	95.7 (↓2.6)	**98.6** (↓0.5)	53.56 (↓10.24)	78.6 (↓6.0)	94.8 (↓2.9)	97.7 (↓0.8)	49.20 (↓13.50)	80.8 (↓6.1)	97.6 (↓1.1)	**99.4** (↓0.1)	51.59 (↓13.01)
FedOSD	79.2 (↓6.6)	95.8 (↓2.5)	98.1 (↓1.0)	49.28 (↓14.52)	76.9 (↓7.7)	94.3 (↓3.4)	97.2 (↓1.3)	49.79 (↓12.91)	81.3 (↓5.6)	96.1 (↓2.6)	98.7 (↓0.6)	52.10 (↓12.50)
NoT	76.4 (↓9.4)	91.0 (↓7.3)	97.2 (↓1.9)	48.99 (↓14.81)	72.7 (↓11.9)	92.3 (↓5.4)	95.4 (↓3.1)	46.10 (↓16.60)	79.9 (↓7.0)	93.3 (↓5.4)	97.3 (↓2.0)	47.36 (↓17.24)
SoUL	79.3 (↓6.5)	95.7 (↓2.6)	97.7 (↓1.4)	50.30 (↓13.50)	78.1 (↓6.5)	94.2 (↓3.5)	97.0 (↓1.5)	45.91 (↓16.79)	79.8 (↓7.1)	96.1 (↓2.6)	98.6 (↓0.7)	49.77 (↓14.83)
FFMU	69.0 (↓16.8)	84.7 (↓13.6)	94.6 (↓4.5)	40.81 (↓22.99)	65.5 (↓19.1)	88.5 (↓9.2)	93.8 (↓4.7)	35.51 (↓27.19)	72.7 (↓14.2)	87.2 (↓11.5)	95.4 (↓3.9)	41.32 (↓23.28)
FUSED	80.9 (↓4.9)	95.7 (↓2.6)	97.8 (↓1.3)	49.37 (↓14.43)	76.9 (↓7.7)	94.5 (↓3.2)	97.1 (↓1.4)	48.66 (↓14.04)	82.9 (↓4.0)	97.2 (↓1.5)	98.7 (↓0.6)	54.40 (↓10.20)
GradAscent	79.6 (↓6.2)	93.9 (↓4.4)	96.5 (↓2.6)	42.52 (↓21.28)	77.8 (↓6.8)	92.7 (↓5.0)	95.1 (↓3.4)	41.58 (↓21.12)	78.5 (↓8.4)	95.5 (↓3.2)	97.1 (↓2.2)	49.44 (↓15.16)
EASE	**83.4** (↓2.4)	**97.5** (↓0.8)	98.3 (↓0.8)	**57.68** (↓6.12)	**82.7** (↓1.9)	**97.0** (↓0.7)	**98.2** (↓0.3)	**54.35** (↓8.35)	**84.6** (↓2.3)	**98.2** (↓0.5)	98.7 (↓0.6)	**56.12** (↓8.48)
Retrain	85.8	98.3	99.1	63.80	84.6	97.7	98.5	62.70	86.9	98.7	99.3	64.60

Table 20: Main comparison on Flickr30K with Qwen2.5-VL-7B across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$
FedEraser	65.8 (↓60.1)	86.4 (↓78.8)	92.3 (↓83.9)	0.77 (↓0.46)	70.5 (↓60.3)	90.5 (↓78.6)	98.0 (↓85.3)	0.80 (↓0.44)	65.6 (↓61.4)	83.7 (↓78.3)	93.5 (↓87.2)	0.76 (↓0.48)
FedRecover	60.7 (↓55.0)	79.4 (↓71.8)	86.9 (↓78.5)	0.74 (↓0.43)	65.9 (↓55.7)	82.3 (↓70.4)	88.9 (↓76.2)	0.80 (↓0.44)	58.8 (↓54.6)	77.6 (↓72.2)	84.2 (↓77.9)	0.73 (↓0.45)
Ferrari	100.0 (↓94.3)	100.0 (↓92.4)	100.0 (↓91.6)	0.83 (↓0.52)	100.0 (↓89.8)	100.0 (↓88.1)	100.0 (↓87.3)	0.85 (↓0.49)	98.1 (↓93.9)	100.0 (↓94.6)	100.0 (↓93.7)	0.78 (↓0.50)
FedOSD	96.8 (↓91.1)	100.0 (↓92.4)	100.0 (↓91.6)	0.78 (↓0.47)	100.0 (↓89.8)	100.0 (↓88.1)	100.0 (↓87.3)	0.83 (↓0.47)	94.6 (↓90.4)	100.0 (↓94.6)	100.0 (↓93.7)	0.77 (↓0.49)
NoT	60.7 (↓55.0)	88.2 (↓80.6)	92.9 (↓84.5)	0.79 (↓0.48)	64.2 (↓54.0)	92.8 (↓80.9)	98.2 (↓85.5)	0.82 (↓0.46)	58.9 (↓54.7)	86.0 (↓80.6)	92.3 (↓86.0)	0.79 (↓0.51)
SoUL	91.2 (↓85.5)	100.0 (↓92.4)	100.0 (↓91.6)	0.80 (↓0.49)	94.3 (↓84.1)	100.0 (↓88.1)	100.0 (↓87.3)	0.81 (↓0.45)	88.5 (↓84.3)	100.0 (↓94.6)	100.0 (↓93.7)	0.73 (↓0.45)
FFMU	42.6 (↓36.9)	70.2 (↓62.6)	80.5 (↓72.1)	0.65 (↓0.34)	45.2 (↓35.0)	74.6 (↓62.7)	84.5 (↓71.8)	0.72 (↓0.36)	39.3 (↓35.1)	68.2 (↓62.8)	79.4 (↓73.1)	0.67 (↓0.39)
FUSED	97.5 (↓91.8)	100.0 (↓92.4)	100.0 (↓91.6)	0.80 (↓0.49)	100.0 (↓89.8)	100.0 (↓88.1)	100.0 (↓87.3)	0.86 (↓0.50)	95.8 (↓91.6)	100.0 (↓94.6)	100.0 (↓93.7)	0.78 (↓0.50)
GradAscent	92.0 (↓86.3)	100.0 (↓92.4)	100.0 (↓91.6)	0.83 (↓0.52)	93.8 (↓83.6)	100.0 (↓88.1)	100.0 (↓87.3)	0.86 (↓0.50)	88.9 (↓84.7)	100.0 (↓94.6)	100.0 (↓93.7)	0.81 (↓0.53)
EASE	**7.8** (↓2.1)	**9.8** (↓2.2)	**10.7** (↓2.3)	**0.42** (↓0.11)	**11.9** (↓1.7)	**13.3** (↓1.4)	**14.3** (↓1.6)	**0.48** (↓0.12)	**6.3** (↓2.1)	**7.5** (↓2.1)	**8.0** (↓1.7)	**0.39** (↓0.11)
Retrain	5.7	7.6	8.4	0.31	10.2	11.9	12.7	0.36	4.2	5.4	6.3	0.28
Method	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$
FedEraser	69.7 (↓9.8)	86.6 (↓6.2)	89.7 (↓4.5)	40.66 (↓21.04)	70.0 (↓8.2)	84.9 (↓6.7)	89.1 (↓3.9)	40.62 (↓19.68)	73.2 (↓7.5)	87.9 (↓5.3)	91.8 (↓2.7)	43.14 (↓19.66)
FedRecover	69.9 (↓9.6)	83.5 (↓9.3)	90.6 (↓3.6)	42.43 (↓19.27)	66.1 (↓12.1)	82.7 (↓8.9)	87.4 (↓5.6)	40.74 (↓19.56)	71.2 (↓9.5)	87.5 (↓5.7)	91.2 (↓3.3)	39.25 (↓23.55)
Ferrari	71.3 (↓8.2)	89.5 (↓3.3)	92.3 (↓1.9)	49.22 (↓12.48)	71.2 (↓7.0)	86.8 (↓4.8)	90.8 (↓2.2)	45.79 (↓14.51)	74.7 (↓6.0)	90.0 (↓3.2)	92.9 (↓1.6)	45.36 (↓17.44)
FedOSD	72.0 (↓7.5)	88.0 (↓4.8)	91.7 (↓2.5)	46.09 (↓15.61)	70.1 (↓8.1)	85.4 (↓6.2)	89.3 (↓3.7)	41.53 (↓18.77)	73.0 (↓7.7)	90.6 (↓2.6)	91.1 (↓3.4)	47.46 (↓15.34)
NoT	67.3 (↓12.2)	87.2 (↓5.6)	90.6 (↓3.6)	40.43 (↓21.27)	65.8 (↓12.4)	85.0 (↓6.6)	89.2 (↓3.8)	37.41 (↓22.89)	70.2 (↓10.5)	87.7 (↓5.5)	89.6 (↓4.9)	42.41 (↓20.39)
SoUL	70.8 (↓8.7)	87.4 (↓5.4)	90.2 (↓4.0)	45.22 (↓16.48)	67.6 (↓10.6)	86.5 (↓5.1)	88.9 (↓4.1)	43.49 (↓16.81)	72.8 (↓7.9)	90.0 (↓3.2)	92.6 (↓1.9)	47.20 (↓15.60)
FFMU	63.2 (↓16.3)	79.0 (↓13.8)	87.0 (↓7.2)	31.84 (↓29.86)	61.4 (↓16.8)	81.8 (↓9.8)	87.2 (↓5.8)	30.39 (↓29.91)	66.2 (↓14.5)	79.5 (↓13.7)	86.7 (↓7.8)	37.41 (↓25.39)
FUSED	71.7 (↓7.8)	89.3 (↓3.5)	91.9 (↓2.3)	48.08 (↓13.62)	70.6 (↓7.6)	88.3 (↓3.3)	89.8 (↓3.2)	43.90 (↓16.40)	73.0 (↓7.7)	90.7 (↓2.5)	92.3 (↓2.2)	44.86 (↓17.94)
GradAscent	70.8 (↓8.7)	85.3 (↓7.5)	89.4 (↓4.8)	44.36 (↓17.34)	68.2 (↓10.0)	86.5 (↓5.1)	88.4 (↓4.6)	35.27 (↓25.03)	73.1 (↓7.6)	88.3 (↓4.9)	90.8 (↓3.7)	41.90 (↓20.90)
EASE	**77.7** (↓1.8)	**91.9** (↓0.9)	**93.1** (↓1.1)	**52.15** (↓9.55)	**75.4** (↓2.8)	**90.2** (↓1.4)	**92.3** (↓0.7)	**51.01** (↓9.29)	**78.6** (↓2.1)	**91.8** (↓1.4)	**93.6** (↓0.9)	**53.89** (↓8.91)
Retrain	79.5	92.8	94.2	61.70	78.2	91.6	93.0	60.30	80.7	93.2	94.5	62.80

Table 21: Main comparison on Flickr30K with LLaVA-1.5-7B across three unlearning scenarios. Convention follows Table 1.

Method	Client Unlearning				Class Unlearning				Sample Unlearning			
	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$	$F\text{-}R@1\downarrow$	$F\text{-}R@5\downarrow$	$F\text{-}R@10\downarrow$	$\rho\downarrow$
FedEraser	69.6 (↓61.8)	88.9 (↓79.8)	95.0 (↓84.5)	0.82 (↓0.48)	72.6 (↓60.2)	92.0 (↓78.1)	98.6 (↓83.9)	0.81 (↓0.42)	65.8 (↓60.0)	86.3 (↓79.1)	94.9 (↓86.8)	0.80 (↓0.50)
FedRecover	64.9 (↓57.1)	81.4 (↓72.3)	87.3 (↓76.8)	0.76 (↓0.42)	67.1 (↓54.7)	86.0 (↓72.1)	91.5 (↓76.8)	0.83 (↓0.44)	62.4 (↓56.6)	78.9 (↓71.7)	85.9 (↓77.8)	0.70 (↓0.40)
Ferrari	100.0 (↓92.2)	100.0 (↓90.9)	100.0 (↓89.5)	0.82 (↓0.48)	100.0 (↓87.6)	100.0 (↓86.1)	100.0 (↓85.3)	0.85 (↓0.46)	99.1 (↓93.3)	100.0 (↓92.8)	100.0 (↓91.9)	0.80 (↓0.50)
FedOSD	99.0 (↓91.2)	100.0 (↓90.9)	100.0 (↓89.5)	0.82 (↓0.48)	100.0 (↓87.6)	100.0 (↓86.1)	100.0 (↓85.3)	0.85 (↓0.46)	96.4 (↓90.6)	100.0 (↓92.8)	100.0 (↓91.9)	0.81 (↓0.51)
NoT	63.0 (↓55.2)	89.0 (↓79.9)	95.2 (↓84.7)	0.82 (↓0.48)	66.9 (↓54.5)	94.3 (↓80.4)	99.3 (↓84.6)	0.90 (↓0.51)	59.4 (↓53.6)	86.9 (↓79.7)	94.8 (↓86.7)	0.78 (↓0.48)
SoUL	93.8 (↓86.0)	100.0 (↓90.9)	100.0 (↓89.5)	0.79 (↓0.45)	95.3 (↓82.9)	100.0 (↓86.1)	100.0 (↓85.3)	0.86 (↓0.47)	89.3 (↓83.5)	100.0 (↓92.8)	100.0 (↓91.9)	0.77 (↓0.47)
FFMU	44.5 (↓36.7)	72.0 (↓62.9)	81.5 (↓71.0)	0.66 (↓0.32)	47.5 (↓35.1)	78.5 (↓64.6)	86.3 (↓71.6)	0.73 (↓0.34)	41.0 (↓35.2)	70.5 (↓63.3)	81.1 (↓73.0)	0.65 (↓0.35)
FUSED	98.5 (↓90.7)	100.0 (↓90.9)	100.0 (↓89.5)	0.81 (↓0.47)	100.0 (↓87.6)	100.0 (↓86.1)	100.0 (↓85.3)	0.88 (↓0.49)	97.5 (↓91.7)	100.0 (↓92.8)	100.0 (↓91.9)	0.80 (↓0.50)
GradAscent	92.2 (↓84.4)	100.0 (↓90.9)	100.0 (↓89.5)	0.85 (↓0.51)	97.4 (↓85.0)	100.0 (↓86.1)	100.0 (↓85.3)	0.89 (↓0.50)	90.0 (↓84.2)	100.0 (↓92.8)	100.0 (↓91.9)	0.81 (↓0.51)
EASE	**10.3** (↓2.5)	**12.0** (↓2.9)	**12.6** (↓2.1)	**0.47** (↓0.13)	**15.0** (↓2.6)	**15.3** (↓1.4)	**17.3** (↓2.6)	**0.57** (↓0.18)	**7.9** (↓2.1)	**9.5** (↓2.3)	**10.7** (↓2.6)	**0.44** (↓0.14)
Retrain	7.8	9.1	10.5	0.34	12.4	13.9	14.7	0.39	5.8	7.2	8.1	0.30
Method	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$	$R\text{-}R@1\uparrow$	$R\text{-}R@5\uparrow$	$R\text{-}R@10\uparrow$	$CIDEr\uparrow$
FedEraser	68.3 (↓9.6)	84.1 (↓6.7)	88.2 (↓3.9)	39.24 (↓19.56)	66.9 (↓9.5)	81.4 (↓8.1)	87.0 (↓3.8)	39.35 (↓17.85)	68.8 (↓10.4)	84.8 (↓6.5)	88.4 (↓4.2)	40.78 (↓18.72)
FedRecover	65.1 (↓12.8)	83.7 (↓7.1)	86.2 (↓5.9)	35.56 (↓23.24)	65.8 (↓10.6)	79.6 (↓9.9)	84.0 (↓6.8)	34.53 (↓22.67)	67.3 (↓11.9)	82.6 (↓8.7)	89.4 (↓3.2)	37.50 (↓22.00)
Ferrari	70.9 (↓7.0)	86.3 (↓4.5)	89.1 (↓3.0)	43.35 (↓15.45)	68.7 (↓7.7)	85.1 (↓4.4)	88.1 (↓2.7)	41.64 (↓15.56)	71.6 (↓7.6)	88.4 (↓2.9)	89.9 (↓2.7)	**47.87** (↓11.63)
FedOSD	71.5 (↓6.4)	85.6 (↓5.2)	90.2 (↓1.9)	43.89 (↓14.91)	69.4 (↓7.0)	85.9 (↓3.6)	88.2 (↓2.6)	43.16 (↓14.04)	70.5 (↓8.7)	86.2 (↓5.1)	89.4 (↓3.2)	42.25 (↓17.25)
NoT	64.8 (↓13.1)	83.4 (↓7.4)	88.2 (↓3.9)	39.17 (↓19.63)	63.4 (↓13.0)	82.6 (↓6.9)	86.2 (↓4.6)	35.56 (↓21.64)	66.3 (↓12.9)	84.4 (↓6.9)	89.4 (↓3.2)	36.90 (↓22.60)
SoUL	68.7 (↓9.2)	85.7 (↓5.1)	88.4 (↓3.7)	41.74 (↓17.06)	68.8 (↓7.6)	83.9 (↓5.6)	86.6 (↓4.2)	40.79 (↓16.41)	71.7 (↓7.5)	88.1 (↓3.2)	89.2 (↓3.4)	42.87 (↓16.63)
FFMU	63.8 (↓14.1)	75.8 (↓15.0)	83.9 (↓8.2)	30.66 (↓28.14)	60.1 (↓16.3)	79.7 (↓9.8)	83.6 (↓7.2)	32.26 (↓24.94)	64.0 (↓15.2)	81.3 (↓10.0)	84.1 (↓8.5)	37.62 (↓21.88)
FUSED	70.5 (↓7.4)	86.2 (↓4.6)	90.3 (↓1.8)	40.32 (↓18.48)	67.0 (↓9.4)	85.2 (↓4.3)	88.0 (↓2.8)	37.42 (↓19.78)	74.0 (↓5.2)	86.9 (↓4.4)	91.1 (↓1.5)	45.84 (↓13.66)
GradAscent	69.9 (↓8.0)	86.0 (↓4.8)	88.3 (↓3.8)	35.39 (↓23.41)	66.4 (↓10.0)	82.5 (↓7.0)	85.8 (↓5.0)	38.30 (↓18.90)	71.9 (↓7.3)	86.5 (↓4.8)	90.4 (↓2.2)	34.64 (↓24.86)
EASE	**75.6** (↓2.3)	**89.6** (↓1.2)	**91.1** (↓1.0)	**49.42** (↓9.38)	**73.4** (↓3.0)	**88.6** (↓0.9)	**90.0** (↓0.8)	**45.44** (↓11.76)	**76.9** (↓2.3)	**90.1** (↓1.2)	**91.2** (↓1.4)	47.35 (↓12.15)
Retrain	77.9	90.8	92.1	58.80	76.4	89.5	90.8	57.20	79.2	91.3	92.6	59.50