LEARNING WITH UNLEARNING: SYNAPTIC ROLE PARTITIONING FOR SIMULTANEOUS CONTINUAL LEARNING AND MACHINE UNLEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Continual Learning (CL) asks a model to acquire new knowledge without forgetting what it already knows. Machine Unlearning (MU) asks it to destroy specific knowledge without damaging the rest. The challenge is having them in the same system simultaneously; the two objectives pull on the same weights, and an edit made for one harms the other. To address this, we propose *Learning with Unlearning* (LwU), a unified framework that resolves the conflict at the level where it occurs, the individual synapse. LwU scores every weight twice, once against the data to keep and once against the data to delete, and partitions the network into four zones with separate update rules. Weights that only require retention are frozen. Weights that only require deletion are cleared and returned to the network as free capacity. Contested weights take a gradient-ascent step projected orthogonally to the retention gradient. The remaining capacity learns the incoming task by a privileged teacher via self-distillation. Deletion is served post hoc, and no retraining is required. Across both CL and MU jobs, LwU outperforms or matches the strongest baseline. Its real-world practicality has been demonstrated on the Physical Artificial Intelligence (PAI) dataset we conducted. Additionally, in the image generation task and in removing dangerous content, it outperformed the closest baseline, as shown in Fig. 1.



Figure 1: Suppressing nudity in text-to-image generation. Our proposed LwU is compared with the original Stable Diffusion, SalUn, and a random-mask baseline. Each column shares the same prompt across methods. Red borders indicate detected nudity; black boxes are for presentation only. The rest of this set of images is provided in Fig. 3.

1 INTRODUCTION

Neural networks are typically trained under the assumption that all training data is available upfront and remains valid indefinitely. Real deployments violate both parts of this assumption. First, data

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.

arrives sequentially, causing models to overwrite previously acquired knowledge through *catastrophic forgetting* (De Lange et al., 2021). Second, some training data may later need to be removed, requiring the model to eliminate its influence after training (Ginart et al., 2019; Xu et al., 2023). Continual Learning (CL) addresses the first requirement by aiming to acquire new knowledge while preserving previously obtained knowledge (Vahedifar et al., 2026b). Machine Unlearning (MU) addresses the second requirement by handling deletion requests after training is completed, when retraining from scratch is prohibitively expensive (Xu et al., 2023). Consider a Physical Artificial Intelligence (PAI) system that learns sequentially from different surgeons. The system must adapt to each new operating style without losing earlier skills, while also being able to remove a patient's influence months later. By then, however, the parameters shaped by that patient may have been reused by several subsequent tasks. Supporting continual learning and data deletion in isolation is therefore insufficient: a deployed system must handle both throughout its lifetime. CL and MU cannot be combined reliably by simply applying one method after the other. Every unlearning operation changes the starting point of the next continual update, while subsequent learning may reactivate representations that the deletion step was intended to remove. Existing joint CL–MU approaches primarily manage this interaction at the task level with buffers (Chatterjee et al., 2024) or at the module level (Liu et al., 2022). However, retained and forgotten knowledge are often entangled within the same parameters. The central conflict, therefore, arises at the level of individual synapses: a parameter may support information that must be preserved, information that must be removed, or both.

To address the need to integrate learning and unlearning, we introduce *Learning with Unlearning* (LwU), a unified framework that enables both processes within a single architecture. LwU decomposes the network into four functional subnetworks (See Fig.2): *(i) Zone A (Safe Retain)*, which preserves parameters critical for retained data and previously learned tasks; *(ii) Zone B (Pure Forget)*, which isolates parameters associated with data targeted for unlearning; *(iii) Zone C (Conflict)*, where shared parameters are selectively retained or suppressed based on their importance; and *(iv) Zone D (Plastic)*, which remains fully adaptable to accommodate new knowledge. This partition separates preservation, deletion, conflict resolution, and adaptation, rather than forcing all four objectives into a single global update. We evaluate LwU against buffer-free CL methods, we examine whether its frozen and plastic zones improve the balance between learning new tasks and preserving earlier ones. Additionally, we evaluate LwU against established MU methods, to measure whether its edits remove the requested. We also compare LwU with joint CL–MU frameworks over long sequences of interleaved learning and deletion requests. Finally, we apply the same protocol to real haptic recordings from twenty participants in a PAI scenario, where surgeon-specific tasks arrive sequentially, and patient-level deletion requests occur over time. Our contributions are:

- **A unified framework for continual learning and unlearning.** LwU uses a single fixed-capacity model and a synapse-level partition for both objectives. Dual importance scores assign each parameter to one of preservation, deletion, conflict resolution, or adaptation, allowing these operations to be controlled independently.
- **Buffer-free continual learning.** LwU stores no examples from previous tasks. A frozen privileged teacher transfers permitted prior behavior to the plastic parameters, while a persistent causal memory supplies contextual information through confidence-gated pseudo-label statistics.
- **Post hoc, retrain-free unlearning.** LwU handles deletion requests after training and between continual tasks through a single model edit. Reinitializing forget-dominant parameters removes their stored information and restores their capacity, while retain-orthogonal updates reduce the influence of the forget set on contested parameters without a first-order change in the retain loss.

Related work. Due to space constraints, we defer our related work to Appendix A.

2 Learning with Unlearning

Problem formulation. Tasks $T = \{T_k\}_{k=1}^{K}$ arrive in sequence with data $\mathcal{D} = \{\mathcal{D}_k\}_{k=1}^{K}$, where $\mathcal{D}_k = \{(x_i^k, y_i^k)\}_{i=1}^{n_k}$. We want a model $f_\theta = Wh_\theta$, with parameters θ, penultimate feature $h_\theta(x) \in \mathbb{R}^d$ and classifier $W \in \mathbb{R}^{C\times d}$, that performs well on everything it has seen, removes a forget set $\mathcal{D}_f \subset \mathcal{D}$ on request, and does not catastrophically forget $\mathcal{D}_r = \mathcal{D} \setminus \mathcal{D}_f$. Deletion is post hoc: a user supplies $\mathcal{D}_f$ and a small retain subset $\mathcal{D}'_r \subset \mathcal{D}_r$, and nothing else. Buffer-free means we store **no** episodic exemplars from any previous task at any point. We define a partition at the synaptic (weights) level, **not** the neuron level.

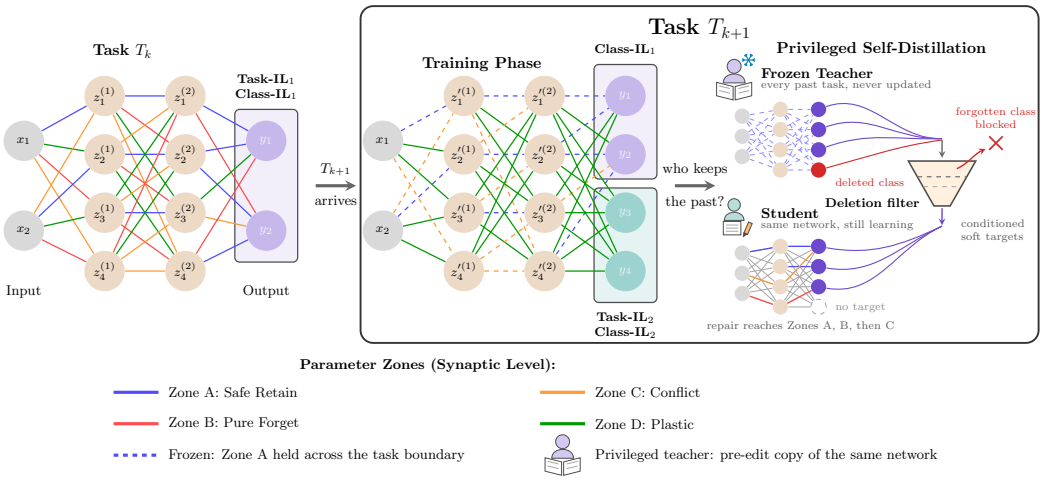


Figure 2: The LwU pipeline. **Left (T_k):** synapses, not neurons, are partitioned by a pair of Fisher scores into Zone A (blue, safe retain), Zone B (red, pure forget), Zone C (orange, conflict), and Zone D (green, plastic). **Right (T_{k+1}):** Zone A is held by the cumulative freeze mask $\mathbf{A}_{k-1}$ (dashed blue), Zone B is reinitialized and returned to the plastic pool, Zone C receives the retain-orthogonal ascent, and Zone D absorbs the new task and its new output classes.

2.1 Zone Identification

Every zone decision answers one question. Does this weight serve the data we keep, or the data we delete? We score each parameter twice, once against each side, using the diagonal empirical Fisher information. For $s \in \{r, f\}$,

$$\phi_s^i = \frac{1}{N_s} \sum_{b \subset \mathcal{D}_s} |b| \left(\frac{\partial \ell_b(\theta)}{\partial \theta_i} \right)^2, \qquad N_s = \sum_{b \subset \mathcal{D}_s} |b|. \tag{1}$$

The score is the expected squared gradient of the log-likelihood, so it measures how sharply the loss on $\mathcal{D}_s$ responds to a perturbation of θ_i.

2.2 Zone Construction

Thresholds. Cutoffs are quantiles of the two score maps, $\tau_r = Q_{q_r}(\phi_r)$ and $\tau_f = Q_{q_f}(\phi_f)$, recomputed once per request. Quantiles rather than absolute cutoffs keep zone sizes stable across architectures and datasets.

Dominance. The two maps live on different scales. The forget set is small and its gradients are large, so comparing ϕ_r^i against ϕ_f^i raw would hand almost everything to the forget side. We therefore compare scores normalized by their own thresholds,

$$\rho_s^i = \phi_s^i / \tau_s, \qquad s \in \{r, f\}, \tag{2}$$

so $\rho_s^i > 1$ exactly when ϕ_s^i clears its quantile. A weight belongs to one side only when its normalized score there exceeds $m \geq 1$ times its normalized score on the other.

Zones. Combining the thresholds with dominance gives

$$\textbf{Safe Retain} \quad \Rightarrow \quad \mathcal{Z}_A = \{i : \rho_r^i > 1, \ \rho_r^i > m\, \rho_f^i\}, \tag{3}$$

$$\textbf{Pure Forget} \quad \Rightarrow \quad \mathcal{Z}_B = \{i : \rho_f^i > 1, \ \rho_f^i > m\, \rho_r^i\}, \tag{4}$$

$$\textbf{Conflict} \quad \Rightarrow \quad \mathcal{Z}_C = \{i : \max(\rho_r^i, \rho_f^i) > 1\} \setminus (\mathcal{Z}_A \cup \mathcal{Z}_B), \tag{5}$$

$$\textbf{Plastic} \quad \Rightarrow \quad \mathcal{Z}_D = \{i : \max(\rho_r^i, \rho_f^i) \leq 1\}. \tag{6}$$

The scores are nonnegative and $m \geq 1$, so no weight can dominate on both sides and $\mathcal{Z}_A \cap \mathcal{Z}_B = \emptyset$; $\mathcal{Z}_C$ is whatever else clears a threshold and $\mathcal{Z}_D$ is what clears neither, so the four sets partition $\theta = \theta_A \cup \theta_B \cup \theta_C \cup \theta_D$. Zone A is what only retention needs, Zone B is what only deletion needs, Zone C is contested, and Zone D is spare capacity.

2.3 ZONE-BASED OPTIMIZATION

A method that runs CL and MU at once must meet three requirements. Hold on to $\mathcal{D}_r$. Remove $\mathcal{D}_f$. Keep learning.

Zone A: retention under deletion. Unlearning is not scored on forget accuracy alone. Retaining accuracy is the other half, and it is the half that collapses when an edit is careless, because an aggressive unlearning step takes the rest of the model down with the deleted class. Zone A is the answer to that half. Its weights are excluded from every destructive update: no reinitialization, no ascent, and no gradient from the incoming task, enforced across task boundaries by the cumulative freeze mask $\mathbf{A}_{k-1}$. The one update Zone A does receive is the distillation repair of Section 2.4, whose target is the pre-edit model itself. Zone A can therefore be pulled back toward prior behavior, never away from it.

Zone B: forgetting, made cheap by the continual setting (catastrophic forgetting). Post-hoc unlearning usually cannot afford a hard edit, so it dampens instead of deleting. The continual setting removes that constraint, because deletion is not the last thing that happens to the model. A next task is already on its way, and Zone B is handed straight back to the plastic pool, so we take the hard edit and clear the weights outright with He initialization (He et al., 2015):

$$\theta_B \leftarrow \mathrm{He}(\theta_B), \qquad \theta_B^{(\text{1-D})} \leftarrow 0. \tag{7}$$

Two forces finish what the reset starts. Training on T_{k+1} writes unrelated features into those slots, so the erased capacity is occupied rather than merely blank. And ordinary catastrophic forgetting pushes whatever residue survives further away from $\mathcal{D}_f$ with each subsequent task; however, this is not guaranteed and can be seen as beneficial forgetting. The open door that makes CL hard is what makes unlearning cheap. No damping coefficient appears, because nothing has to be held in place.

Zone C: the contested weights. These serve both objectives and belong to neither. They have to increase the forget loss without reducing retention performance. With $g_r^C = \nabla_{\theta_C}\ell(\mathcal{D}_r')$ and $g_f^C = \nabla_{\theta_C}\ell(\mathcal{D}_f)$, we project the ascent direction onto the orthogonal complement of the retain gradient,

$$g_u^\perp = g_f^C - \frac{\langle g_f^C, g_r^C\rangle}{\|g_r^C\|^2 + \epsilon}\, g_r^C, \qquad \theta_C^{(s+1)} = \theta_C^{(s)} + \eta_C\, g_u^\perp, \tag{8}$$

and clip the cumulative displacement to $\|\theta_C^{(s)} - \theta_C^{(0)}\| \le \delta_C$, with δ_C proportional to the mean importance in Zone C. The projected step satisfies

$$\langle g_u^\perp, g_r^C\rangle = \frac{\epsilon}{\|g_r^C\|^2 + \epsilon}\, \langle g_f^C, g_r^C\rangle, \tag{9}$$

so it retains only a fraction $\epsilon/(\|g_r^C\|^2 + \epsilon)$ of the raw ascent's first-order effect on the retain loss, and this fraction vanishes as $\epsilon \to 0$. The stabilizer also keeps $\|g_u^\perp\| \le \|g_f^C\|$, so projection can only shrink the step, never amplify it. The guarantee is local: it holds per step, to first order, with respect to the gradient on $\mathcal{D}_r'$, and while the displacement budget is inactive; when the clip binds, the rescaling toward $\theta_C^{(0)}$ can reintroduce a small retain-gradient component. Within these limits, conflict resolution needs no trade-off coefficient between the forget and retain objectives.

Zone D: room to keep learning. Before training on T_k we release the recycled capacity, $\mathbf{A}_{k-1} \leftarrow \mathbf{A}_{k-1} \setminus \mathcal{Z}_B$, and apply the masked update

$$\theta^{(s+1)} = \theta^{(s)} - \eta\, \nabla_\theta \ell(\mathcal{D}_k) \odot (1 - \mathbf{A}_{k-1}). \tag{10}$$

Gradients reach Zones B, C, and D. Because Zone B rejoins the pool with every request, the plastic subspace grows each time something is deleted: forgetting buys plasticity rather than spending it. The three requirements now have owners. $\mathcal{D}_f$ is cleared by Zones B and C and finished off by the overwriting and drift that follow. $\mathcal{D}_r$ is protected by Zone A. New classes are learned in Zone D. What is still missing is whatever keeps the old behavior alive while Zone D writes the new one.

2.4 THE PRIVILEGED TEACHER FOR SELF DISTILLATION

Capacity alone does not give continual learning. Zone D says where new knowledge may be written; it says nothing about what survives the writing. To preserve earlier behavior without replaying earlier

data, the model needs a teacher. We use its previous self. A network trained through T_{k-1} already carries a compressed record of its experience in its parameters. Self-distillation (SD) makes that record a teaching signal: we freeze $\theta^T = \theta^{(k-1)}$ and let it guide the model as its parameters θ change. The teacher's privilege is having learned from data the student can no longer access. Then forcing the student to imitate the teacher's behavior without access to the teacher's privileged knowledge is self-distillation. Teacher and student share the same architecture; their difference is the knowledge each must preserve.

For continual learning, the previous model teaches what it knows. For unlearning, it must teach selectively. We therefore mask the forgotten classes $\mathcal{F}$ by setting their logits to $-\infty$ before softmax. At temperature κ, this gives, for every $c \notin \mathcal{F}$,

$$q_\kappa^{\mathcal{F}}(c \mid x; \theta^0) = \frac{q_\kappa(c \mid x; \theta^0)}{1 - \sum_{c' \in \mathcal{F}} q_\kappa(c' \mid x; \theta^0)} = q_\kappa\big(c \mid x,\ y \notin \mathcal{F}; \theta^0\big), \tag{11}$$

with zero probability for $c \in \mathcal{F}$, where θ^0 is the frozen pre-edit copy. The renormalization has a precise meaning: it conditions the teacher's prediction on the label belonging to a retained class. The teacher still carries its history, but the target preserves only its relative preferences among the remaining classes. Masking removes the deleted classes' output mass, not their training influence: θ^0 was trained on $\mathcal{D}_f$, so its preferences among retained classes still reflect it. For instance and subclass requests the deleted samples carry retained labels, so $\mathcal{F}$ does not grow and the request adds no mask. In both cases distillation is evaluated only on $\mathcal{D}'_r$, so the teacher never supplies a target on $\mathcal{D}_f$ itself; removal is carried by the Zone B reset and the Zone C ascent, and the teacher's role is to repair retained behavior around them. On the supplied retain subset, the student minimizes

$$\mathcal{L}_{\mathrm{SD}} = \kappa^2 D_{\mathrm{KL}}\big(q_\kappa^{\mathcal{F}}(\cdot \mid x; \theta^0) \,\big\|\, p_\kappa(\cdot \mid x; \theta)\big) + \lambda_{\mathrm{CE}}\, \mathrm{CE}\big(p_1(\cdot \mid x; \theta), y\big), \tag{12}$$

where D_{KL} is the Kullback–Leibler divergence and CE is cross-entropy. Repair runs in two passes on $\mathcal{D}'_r$. The first restricts gradients to $\mathcal{Z}_A \cup \mathcal{Z}_B$ and distills from the pre-edit copy θ^0: Zone A repairs collateral damage caused by the edit, while Zone B relearns the retained structure from its fresh initialization. The second, run when $\lambda_C > 0$, updates Zone C alone and distills from θ^R in place of θ^0, the model as repaired by the first pass, so the contested weights settle against a teacher whose Zones A and B already reflect the edit rather than against the pre-edit model. Both teachers are masked to $\mathcal{F}$. Task training reuses the teacher with a different target. After each task, and again after each deletion, the current model is frozen as θ^T; the first task has no teacher and uses cross-entropy alone. While learning T_k, the student minimizes

$$\mathcal{L}_{\mathrm{task}} = \mathrm{CE}\big(p(\cdot \mid x; \theta)_{\mathcal{O}_k}, y\big) + \lambda_T\, \kappa_T^2\, D_{\mathrm{KL}}\big(q_{\kappa_T}(\cdot \mid x; \theta^T)_{\mathcal{O}^{\mathrm{old}}} \,\big\|\, p_{\kappa_T}(\cdot \mid x; \theta)_{\mathcal{O}^{\mathrm{old}}}\big), \tag{13}$$

where a subscript set means the softmax is taken over those logits only, $\mathcal{O}_k = \mathcal{Y}_k \setminus \mathcal{F}$ are the active classes, $\mathcal{O}^{\mathrm{old}} = \mathcal{Y}_{k-1} \setminus \mathcal{F}$ the retained old classes, and gradients are confined to the unfrozen weights $1 - \mathbf{A}_{k-1}$. Distillation therefore covers only old classes; the new classes of T_k are learned from their labels alone. The deleted set $\mathcal{F}$ is never reset at a task boundary: once a class is deleted, it is excluded from every later cross-entropy and distillation target. Self-distillation thus provides the thread connecting learning and unlearning: the previous model teaches the next one, with its targets constrained to what is allowed to remain.

2.5 Test-Time Memorization

The teacher helps knowledge survive changes to the model. Once deployed, however, the parameters are fixed, while the test stream may continue to change. We therefore add a lightweight fast-weight memory: each batch reads what earlier batches have written, then leaves information for those that follow. After editing and repair, we freeze the parameters $\theta^\star$, discard the teacher, and initialize $\mathbf{M}_0 = \mathbf{0} \in \mathbb{R}^{C \times d}$ at the start of each evaluation stream. For a batch X_t, normalized feature keys $K_t = \mathrm{rowNorm}(h_{\theta^\star}(X_t))$ read the existing memory to augment the base logits $Z_t = f_{\theta^\star}(X_t)$:

$$\widehat{Z}_t = Z_t + \beta K_t \mathbf{M}_{t-1}^\top, \tag{14}$$

where β controls the memory contribution. Only after prediction do high-confidence outputs from the base logits, selected using threshold ρ, produce centered pseudo-label targets. These update $\mathbf{M}_t$ through a closed-form least-squares step with momentum, decay, and row-norm projection. This predict-then-write order makes adaptation causal: information from X_t can influence only future

batches. The memory stores aggregate statistics rather than examples, requires no ground-truth labels, and has storage cost independent of stream length. The slow model remains unchanged, and deleted rows are zeroed after every update so that memory adds no logit to forgotten classes. With zero initialization, setting $\eta_M = 0$ recovers the stored slow model's predictions.

3 Results and Discussion

3.1 Continual Learning framework

Datasets and Architecture. We introduced a new dataset, the Robotic Tactile Internet (RTI) dataset, containing haptic tele-manipulation for PAI scenarios, with details explained in Appendix B. Additionally, we used Seq-CIFAR-100 (Krizhevsky & Hinton, 2009) and Seq-Tiny-ImageNet (Le & Yang, 2015). All methods were trained with the Adam optimizer (Kingma & Ba, 2015) on ResNet-18 (He et al., 2016) and evaluated on both Task-Incremental Learning (Task-IL) (Classifying only between the last classes, with task identity given to the head) and Class-Incremental Learning (Class-IL) (Classifying between all classes, without task identity given to the head). We follow the Generalizable Two-phase Evaluation Protocol (GTEP) (Cha & Cho, 2025): hyperparameters are selected on half of the classes and then transferred, without further tuning, to a disjoint set of the remaining evaluation classes; phase results are averaged over 3 task orderings. Previously, the GTEP paper found that the CL method is cheating by selecting hyperparameters from a validation set that comes from the entire dataset. It is sensible, in a way, that the main assumption of CL is that we do not have access to all data and that using future data to tune hyperparameters gives your method an unfair advantage. In other words, tasks in the real world are now known beforehand.

Baselines and metrics. We compare against Stochastic Gradient Descent (SGD) and joint-training lower and upper bounds, respectively, as well as Elastic Weight Consolidation (EWC) (Kirkpatrick et al., 2017), Synaptic Intelligence (SI) (Zenke et al., 2017), Learning without Forgetting (LwF) (Li & Hoiem, 2018), SpaceNet (Sokar et al., 2021), Winning SubNetworks (WSN) (Kang et al., 2022), Neuro-Inspired Stability-Plasticity Adaptation (NISPA) (Gurbuz & Dovrolis, 2022), Prediction Error-based Classification (PEC) (Zajac et al., 2024), Unified Continual Learning and Machine Unlearning (UniCLUN)† (Chatterjee et al., 2024), and Shapley Neuron Value (SNV) (Vahedifar et al., 2026a). WSN is evaluated only in Task-IL, whereas PEC is evaluated only in Class-IL. The symbol † identifies methods designed jointly for continual learning and machine unlearning. We report final average accuracy (ACC), backward transfer (BWT), forward transfer (FWT) and the harmonic plasticity–stability score (PS) as the primary CL metrics; they are defined in Appendix C.1.

3.2 Machine Unlearning framework

Datasets and Architecture. We evaluate class-, subclass-, and instance-level deletion on CIFAR-10 and CIFAR-100 (Krizhevsky & Hinton, 2009), and RTI datasets. All methods were trained separately with two backbones: ResNet-18 and a Convolutional Neural Network (CNN).

Baselines and metrics. We include three reference methods: *Original* (or joint training), the source model trained on the complete dataset $\mathcal{D}_r \cup \mathcal{D}_f$ without unlearning; *Fine-tuning*, which updates the Original model using only the accessible subset of $\mathcal{D}_r$ after removing $\mathcal{D}_f$; and *Retrain*, which trains a new model from scratch using only $\mathcal{D}_r$ and serves as the gold-standard unlearning reference. We further compare LwU with established unlearning methods, including Amnesiac (Graves et al., 2021) (evaluated only for class-level deletion), UNlearning by Selective Impair and Repair (UNSIR) (Tarun et al., 2024), Selective Synaptic Dampening (SSD) (Foster et al., 2024), SCalable Remembering and Unlearning unBound (SCRUB) (Kurmanji et al., 2023), UniCLUN†, Bad Teacher (Chundawat et al., 2023), and Saliency Unlearning (SalUn) (Fan et al., 2024). Every method is selected using validation data under a utility constraint relative to Retrain, and the official test set is evaluated only after selection. We report the absolute forget-accuracy gap to Retrain $|\Delta\mathcal{D}_f|$, retain and test accuracy, membership-inference attack (MIA) performance, predictive and parameter-distribution Kullback–Leibler (KL) divergence, unlearning time, and post-unlearning relearning accuracy; they are defined in Appendix C.2.

Additional experiments are provided in Appendix G for both CL and MU jobs. Each experiment uses a single NVIDIA RTX PRO 6000 Blackwell Max-Q GPU for both CL and MU jobs.

Table 1: Continual-learning results. ACC and PS are percentages; BWT and FWT are percentage points. Podiums use $3G + 2S + B$.

	TIL				CIL			
Method	ACC↑	BWT↑	FWT↑	PS↑	ACC↑	BWT↑	FWT↑	PS↑
Seq-CIFAR-100								
Joint training (upper bound)	$87.89_{\pm 1.25}$	$+0.00_{\pm 0.00}$	–	–	$65.43_{\pm 2.40}$	$+0.00_{\pm 0.00}$	–	–
Continual-learning methods								
EWC	$70.88_{\pm 2.65}$	$-5.79_{\pm 1.66}$	$+1.19_{\pm 1.14}$	$79.35_{\pm 1.45}$	$8.39_{\pm 0.66}$	$-84.46_{\pm 2.30}$	$-2.42_{\pm 0.54}$	$26.17_{\pm 3.21}$
SI	$67.34_{\pm 1.14}$	$-11.40_{\pm 1.55}$	$+0.29_{\pm 0.40}$	$79.10_{\pm 0.79}$	$8.56_{\pm 0.49}$	$-85.10_{\pm 1.69}$	$-2.38_{\pm 0.61}$	$25.34_{\pm 2.36}$
LwF	$77.81_{\pm 2.49}$	$-7.96_{\pm 1.57}$	$-1.53_{\pm 3.27}$	$86.42_{\pm 1.77}$	$10.06_{\pm 1.79}$	$-80.02_{\pm 1.20}$	$-2.42_{\pm 0.54}$	$32.15_{\pm 1.59}$
WSN	$84.70_{\pm 1.84}$	$+0.00_{\pm 0.00}$	$-0.30_{\pm 1.65}$	$89.42_{\pm 1.62}$	–	–	–	–
PEC	–	–	–	–	$32.84_{\pm 1.25}$	$-12.70_{\pm 0.55}$	$+0.00_{\pm 0.00}$	$56.43_{\pm 0.38}$
SpaceNet	$63.61_{\pm 1.32}$	$-1.62_{\pm 0.29}$	$-0.20_{\pm 0.69}$	$71.68_{\pm 1.10}$	$6.77_{\pm 0.43}$	$-65.46_{\pm 1.45}$	$-1.86_{\pm 0.19}$	$45.26_{\pm 0.86}$
NISPA	$80.02_{\pm 1.55}$	$+0.00_{\pm 0.00}$	$-2.70_{\pm 1.88}$	$85.82_{\pm 1.44}$	$8.24_{\pm 0.54}$	$-78.89_{\pm 1.14}$	$-1.93_{\pm 0.26}$	$33.30_{\pm 1.30}$
SNV	$86.97_{\pm 0.73}$	$+0.00_{\pm 0.00}$	$-3.47_{\pm 0.16}$	$71.71_{\pm 1.21}$	$34.30_{\pm 0.84}$	$-24.19_{\pm 1.92}$	$-2.99_{\pm 0.15}$	$59.21_{\pm 1.36}$
Joint continual-learning and machine-unlearning methods								
UniCLUN	$69.46_{\pm 2.70}$	$-21.36_{\pm 1.86}$	$-0.14_{\pm 0.44}$	$82.11_{\pm 1.46}$	$32.75_{\pm 4.04}$	$-58.84_{\pm 7.47}$	$-1.91_{\pm 0.66}$	$55.18_{\pm 6.17}$
LwU (ours)	$85.42_{\pm 1.90}$	$-5.88_{\pm 0.82}$	$+1.25_{\pm 2.11}$	$85.27_{\pm 1.30}$	$33.22_{\pm 1.63}$	$-52.35_{\pm 1.12}$	$-1.16_{\pm 0.17}$	$59.18_{\pm 1.56}$
SGD (lower bound)	$50.81_{\pm 2.99}$	$-34.44_{\pm 2.83}$	$+1.01_{\pm 2.51}$	$70.99_{\pm 2.21}$	$8.72_{\pm 0.70}$	$-85.85_{\pm 1.37}$	$-2.42_{\pm 0.54}$	$24.29_{\pm 2.00}$
Seq-CIFAR-100 podium: 🥇 **SNV** (4G, 1S, 0B; **14 pts**) 🥈 **LwU (ours)** (1G, 4S, 1B; **12 pts**) 🥉 **PEC** (2G, 0S, 2B; **8 pts**)								
Seq-Tiny-ImageNet								
Joint training (upper bound)	$81.36_{\pm 0.42}$	$0.00_{\pm 0.00}$	–	–	$57.24_{\pm 0.53}$	$0.00_{\pm 0.00}$	–	–
Continual-learning methods								
EWC	$58.90_{\pm 1.16}$	$-11.85_{\pm 2.13}$	$0.57_{\pm 0.41}$	$75.29_{\pm 0.94}$	$7.48_{\pm 0.26}$	$-74.82_{\pm 0.17}$	$-1.01_{\pm 0.27}$	$37.71_{\pm 0.25}$
SI	$63.85_{\pm 2.14}$	$-3.21_{\pm 1.30}$	$-0.82_{\pm 0.49}$	$76.23_{\pm 1.34}$	$7.48_{\pm 0.28}$	$-72.74_{\pm 0.10}$	$-1.01_{\pm 0.27}$	$39.70_{\pm 0.09}$
LwF	$67.39_{\pm 1.49}$	$-6.46_{\pm 1.22}$	$0.08_{\pm 0.58}$	$80.46_{\pm 1.05}$	$8.04_{\pm 0.20}$	$-71.42_{\pm 0.49}$	$-0.96_{\pm 0.27}$	$41.05_{\pm 0.33}$
WSN	$73.68_{\pm 0.10}$	$0.00_{\pm 0.00}$	$0.50_{\pm 0.70}$	$83.00_{\pm 0.50}$	–	–	–	–
PEC	–	–	–	–	$2.88_{\pm 0.36}$	$-27.15_{\pm 1.88}$	$-1.42_{\pm 0.10}$	$36.72_{\pm 1.33}$
SpaceNet	$12.17_{\pm 0.46}$	$0.10_{\pm 0.20}$	$0.00_{\pm 0.60}$	$0.00_{\pm 0.00}$	$1.01_{\pm 0.01}$	$-3.40_{\pm 0.60}$	$-1.10_{\pm 0.10}$	$2.30_{\pm 0.10}$
NISPA	$50.50_{\pm 3.74}$	$-1.10_{\pm 1.50}$	$0.60_{\pm 0.90}$	$61.50_{\pm 4.90}$	$5.46_{\pm 0.29}$	$-49.70_{\pm 4.40}$	$-0.70_{\pm 0.10}$	$49.70_{\pm 0.20}$
SNV	$75.62_{\pm 0.51}$	$0.00_{\pm 0.00}$	$0.60_{\pm 0.10}$	$80.10_{\pm 0.30}$	$30.70_{\pm 1.73}$	$-17.30_{\pm 2.40}$	$8.00_{\pm 1.60}$	$52.10_{\pm 1.30}$
Joint continual-learning and machine-unlearning methods								
UniCLUN	$49.21_{\pm 9.01}$	$-24.60_{\pm 9.60}$	$-0.50_{\pm 0.60}$	$71.00_{\pm 4.80}$	$4.12_{\pm 2.66}$	$-39.80_{\pm 6.40}$	$-1.20_{\pm 0.00}$	$44.30_{\pm 4.40}$
LwU (ours)	$71.35_{\pm 0.89}$	$-3.81_{\pm 0.40}$	$-0.33_{\pm 1.68}$	$82.39_{\pm 0.92}$	$28.60_{\pm 0.23}$	$-71.56_{\pm 0.57}$	$-1.04_{\pm 0.10}$	$40.99_{\pm 0.46}$
SGD (lower bound)	$40.24_{\pm 3.01}$	$-36.32_{\pm 1.33}$	$0.88_{\pm 0.94}$	$66.71_{\pm 1.38}$	$7.59_{\pm 0.34}$	$-74.35_{\pm 0.79}$	$-1.00_{\pm 0.27}$	$38.17_{\pm 0.76}$
Seq-Tiny-ImageNet podium: 🥇 **SNV** (5G, 2S, 0B; **19 pts**) 🥈 **WSN** (1G, 2S, 0B; **7 pts**) 🥉 **NISPA** (1G, 2S, 0B; **7 pts**)								
RTI								
Joint training (upper bound)	$93.38_{\pm 1.06}$	–	–	–	$81.96_{\pm 0.75}$	–	–	–
Continual-learning methods								
EWC	$86.12_{\pm 4.63}$	$-5.79_{\pm 5.23}$	$-2.37_{\pm 1.65}$	$86.15_{\pm 3.00}$	$18.92_{\pm 0.71}$	$-93.00_{\pm 1.54}$	$-8.84_{\pm 8.09}$	$12.99_{\pm 2.65}$
SI	$85.64_{\pm 2.74}$	$-3.82_{\pm 1.08}$	$-2.60_{\pm 1.70}$	$84.02_{\pm 4.16}$	$18.87_{\pm 0.98}$	$-91.96_{\pm 1.07}$	$-7.19_{\pm 10.90}$	$14.77_{\pm 1.79}$
LwF	$91.35_{\pm 1.86}$	$-1.63_{\pm 1.01}$	$-3.70_{\pm 6.04}$	$91.16_{\pm 1.38}$	$19.02_{\pm 0.87}$	$-92.89_{\pm 1.65}$	$-7.90_{\pm 9.62}$	$13.17_{\pm 2.82}$
WSN	$93.49_{\pm 1.19}$	$0.00_{\pm 0.00}$	$-0.01_{\pm 1.81}$	$92.47_{\pm 1.23}$	–	–	–	–
PEC	–	–	–	–	$45.94_{\pm 0.64}$	$-19.72_{\pm 1.58}$	$0.00_{\pm 0.00}$	$65.83_{\pm 0.56}$
SpaceNet	$83.94_{\pm 2.29}$	$+0.18_{\pm 0.15}$	$-0.26_{\pm 5.31}$	$79.80_{\pm 2.77}$	$17.36_{\pm 1.47}$	$-83.27_{\pm 2.18}$	$-8.40_{\pm 1.08}$	$27.80_{\pm 2.90}$
NISPA	$91.39_{\pm 2.03}$	$0.00_{\pm 0.00}$	$+6.06_{\pm 4.67}$	$88.22_{\pm 3.55}$	$18.39_{\pm 1.05}$	$-90.51_{\pm 1.03}$	$-11.49_{\pm 0.59}$	$17.15_{\pm 1.67}$
SNV	$94.60_{\pm 0.65}$	$0.00_{\pm 0.00}$	$+9.00_{\pm 1.74}$	$90.10_{\pm 1.49}$	$50.70_{\pm 2.06}$	$-72.10_{\pm 1.34}$	$+9.00_{\pm 1.74}$	$62.10_{\pm 2.41}$
Joint continual-learning and machine-unlearning methods								
UniCLUN*	$85.73_{\pm 1.63}$	$-9.95_{\pm 2.46}$	$+1.33_{\pm 1.64}$	$87.98_{\pm 0.96}$	$40.74_{\pm 2.52}$	$-63.24_{\pm 6.00}$	$-8.37_{\pm 6.82}$	$51.96_{\pm 5.55}$
LwU (ours)	$93.41_{\pm 0.98}$	$-3.45_{\pm 1.41}$	$-2.77_{\pm 3.78}$	$90.67_{\pm 0.82}$	$46.91_{\pm 0.89}$	$-36.13_{\pm 0.94}$	$-8.11_{\pm 8.93}$	$64.72_{\pm 0.72}$
SGD (lower bound)	$76.14_{\pm 4.63}$	$-19.78_{\pm 4.29}$	$-2.07_{\pm 0.54}$	$81.84_{\pm 3.40}$	$18.98_{\pm 0.82}$	$-93.44_{\pm 1.16}$	$-7.57_{\pm 9.73}$	$12.24_{\pm 2.03}$
RTI podium: 🥇 **SNV** (4G, 1S, 1B; **15 pts**) 🥈 **PEC** (2G, 1S, 1B; **9 pts**) 🥉 **LwU (ours)** (0G, 3S, 2B; **8 pts**)								
Global overall podium: 🥇 **SNV** (13G, 4S, 1B; **48 pts**) 🥈 **LwU (ours)** (1G, 9S, 4B; **25 pts**) 🥉 **WSN** (4G, 4S, 1B; **21 pts**)								

3.3 Continual Learning Performance

Table 1 shows that under TIL and CIL evaluation on Seq-CIFAR-100, SNV leads at 86.97 and 34.30, respectively. Here, LwU only wins a podium finish in third place in TIL acc with 85.42. This shows that a subnetwork-based method that relies on Shapley-based evaluation outperforms Fisher-based zone decomposition. Additionally, WSN, NISPA, and SNV all report BWT $= 0.00$ on all three datasets in the TIL setting. Nothing was forgotten because nothing was allowed to move: the masks make old parameters unreachable, so the number is fixed before training starts. LwU reports -5.88, -3.81, and -3.45 because it writes to shared parameters. Once masking stops working, the picture reverses. On Seq-CIFAR-100 in the CIL setting, SNV drops to -24.19 and NISPA to -78.89, against LwU at -52.35. This issue arises when the number of tasks increases, causing the mask-based method to freeze the entire network.

Table 2: Unlearning results on CIFAR-10, CIFAR-100, and RTI over 3 seeds. U-LiRA uses rankings based on each reported value's distance to Retrain. Podiums are weighted medal score: 3G+2S+1B.

Method	CNN: D_f			ResNet-18: D_f			U-LiRA: Instance deletion			
	Class	Subclass	Instance	Class	Subclass	Instance	CNN		ResNet-18	
							AUC → 50	TPR@1% ↓	AUC → 50	TPR@1% ↓
CIFAR-10										
Retrain (oracle)	$0.00_{\pm 0.00}$	$25.40_{\pm 4.50}$	$91.86_{\pm 1.34}$	$0.00_{\pm 0.00}$	$26.57_{\pm 1.23}$	$92.06_{\pm 0.75}$	50.00	0.00	50.00	0.00
Original model	$93.47_{\pm 1.33}$	$94.57_{\pm 1.17}$	$99.94_{\pm 0.10}$	$93.90_{\pm 1.15}$	$95.30_{\pm 0.69}$	$100.00_{\pm 0.00}$	66.80	12.30	80.10	22.40
Fine-tune on D_r	$50.17_{\pm 43.28}$	$68.90_{\pm 2.80}$	$96.19_{\pm 0.44}$	$26.23_{\pm 45.44}$	$70.07_{\pm 0.68}$	$97.43_{\pm 2.04}$	50.90	1.50	54.80	3.80
Unlearning methods										
Amnesiac	$0.00_{\pm 0.00}$	–	–	$0.00_{\pm 0.00}$	–	–	–	–	–	–
Bad Teacher	$0.27_{\pm 0.15}$	$30.00_{\pm 4.71}$	$98.29_{\pm 0.84}$	$0.13_{\pm 0.23}$	$28.40_{\pm 4.10}$	$98.97_{\pm 0.36}$	88.90	0.40	76.20	1.90
SCRUB	$0.00_{\pm 0.00}$	$90.73_{\pm 1.63}$	$97.52_{\pm 0.81}$	$0.00_{\pm 0.00}$	$84.57_{\pm 0.06}$	$98.82_{\pm 0.66}$	60.30	7.60	68.10	12.00
SSD	$74.60_{\pm 22.39}$	$65.57_{\pm 28.00}$	$99.94_{\pm 0.10}$	$92.67_{\pm 3.43}$	$84.33_{\pm 9.85}$	$100.00_{\pm 0.00}$	59.60	8.30	75.90	18.50
SalUn	$41.50_{\pm 35.98}$	$73.40_{\pm 3.05}$	$98.02_{\pm 0.14}$	$22.07_{\pm 38.22}$	$68.30_{\pm 4.06}$	$96.25_{\pm 2.23}$	87.40	0.00	60.30	0.00
UNSIR	$0.00_{\pm 0.00}$	$79.13_{\pm 10.71}$	$98.61_{\pm 0.31}$	$43.23_{\pm 10.05}$	$82.30_{\pm 4.86}$	$98.88_{\pm 0.58}$	54.80	2.00	57.50	4.30
Joint continual-learning and machine-unlearning methods										
UniCLUN	$41.33_{\pm 35.77}$	$74.40_{\pm 4.29}$	$99.12_{\pm 0.23}$	$41.93_{\pm 36.42}$	$68.10_{\pm 3.86}$	$98.38_{\pm 0.80}$	56.70	4.20	80.80	0.00
LwU (ours)	$0.00_{\pm 0.00}$	$85.63_{\pm 15.15}$	$98.26_{\pm 1.14}$	$3.70_{\pm 6.41}$	$89.77_{\pm 7.82}$	$99.97_{\pm 0.05}$	54.50	0.00	50.90	0.00
CIFAR-10 podium: 🥇 **LwU (ours)** (5G, 0S, 2B; **17 pts**) 🥈 **SalUn** (3G, 1S, 3B; **14 pts**) 🥉 **Bad Teacher** (2G, 4S, 0B; **14 pts**)										
CIFAR-100										
Retrain (oracle)	$0.00_{\pm 0.00}$	$23.33_{\pm 4.16}$	$69.50_{\pm 0.60}$	$0.00_{\pm 0.00}$	$24.00_{\pm 2.00}$	$69.58_{\pm 2.46}$	50.00	–	50.00	–
Original model	$70.00_{\pm 3.61}$	$77.00_{\pm 3.46}$	$99.14_{\pm 0.29}$	$72.33_{\pm 3.21}$	$80.00_{\pm 0.00}$	$100.00_{\pm 0.00}$	81.63	22.20	92.90	50.16
Fine-tune on D_r	$35.00_{\pm 3.61}$	$72.00_{\pm 7.00}$	$98.14_{\pm 0.71}$	$36.33_{\pm 4.51}$	$75.33_{\pm 4.93}$	$99.89_{\pm 0.19}$	81.13	21.00	91.90	47.40
Unlearning methods										
Amnesiac	$0.00_{\pm 0.00}$	–	–	$0.00_{\pm 0.00}$	–	–	–	–	–	–
Bad Teacher	$0.33_{\pm 0.58}$	$17.00_{\pm 4.58}$	$96.72_{\pm 1.21}$	$0.00_{\pm 0.00}$	$17.00_{\pm 3.00}$	$99.31_{\pm 0.77}$	79.32	7.53	85.17	29.51
SCRUB	$64.33_{\pm 9.02}$	$73.33_{\pm 6.11}$	$90.86_{\pm 1.84}$	$31.67_{\pm 9.87}$	$79.67_{\pm 0.58}$	$97.94_{\pm 0.98}$	70.94	12.56	82.67	30.56
SSD	$8.00_{\pm 2.65}$	$51.00_{\pm 30.81}$	$97.06_{\pm 1.38}$	$35.00_{\pm 18.03}$	$57.67_{\pm 31.21}$	$98.94_{\pm 0.46}$	76.47	16.84	87.87	39.84
SalUn	$23.67_{\pm 1.53}$	$37.67_{\pm 5.69}$	$74.81_{\pm 2.72}$	$29.00_{\pm 2.65}$	$42.00_{\pm 3.61}$	$84.53_{\pm 1.61}$	89.45	2.98	75.83	1.36
UNSIR	$37.33_{\pm 17.67}$	$64.33_{\pm 0.58}$	$96.03_{\pm 1.07}$	$44.67_{\pm 10.21}$	$77.33_{\pm 6.11}$	$98.86_{\pm 0.81}$	71.17	8.84	82.07	26.42
Joint continual-learning and machine-unlearning methods										
UniCLUN	$26.67_{\pm 3.79}$	$35.67_{\pm 3.51}$	$74.58_{\pm 1.04}$	$30.67_{\pm 1.53}$	$44.67_{\pm 4.16}$	$85.31_{\pm 1.41}$	90.38	4.07	74.41	1.04
LwU (ours)	$0.00_{\pm 0.00}$	$50.00_{\pm 13.08}$	$95.56_{\pm 1.34}$	$0.00_{\pm 0.00}$	$74.33_{\pm 5.51}$	$97.81_{\pm 0.41}$	55.10	0.00	51.60	0.00
CIFAR-100 podium: 🥇 **LwU (ours)** (6G, 0S, 1B; **19 pts**) 🥈 **SalUn** (1G, 4S, 3B; **14 pts**) 🥉 **UniCLUN** (1G, 4S, 3B; **14 pts**)										
RTI										
Retrain (oracle)	$0.00_{\pm 0.00}$	$86.33_{\pm 2.31}$	$78.57_{\pm 0.35}$	$0.00_{\pm 0.00}$	$86.67_{\pm 3.21}$	$80.99_{\pm 1.01}$	50.00	–	50.00	–
Original model	$72.87_{\pm 1.68}$	$96.33_{\pm 2.08}$	$99.53_{\pm 0.18}$	$71.27_{\pm 2.14}$	$94.67_{\pm 2.52}$	$99.97_{\pm 0.05}$	79.39	21.27	81.67	29.22
Fine-tune on D_r	$28.40_{\pm 3.12}$	$95.33_{\pm 1.53}$	$99.36_{\pm 0.05}$	$36.87_{\pm 5.08}$	$95.67_{\pm 1.53}$	$99.82_{\pm 0.15}$	78.11	19.87	81.98	28.49
Unlearning methods										
Amnesiac	$0.00_{\pm 0.00}$	–	–	$0.00_{\pm 0.00}$	–	–	–	–	–	–
Bad Teacher	$0.13_{\pm 0.12}$	$45.00_{\pm 16.52}$	$97.81_{\pm 0.40}$	$0.00_{\pm 0.00}$	$31.67_{\pm 4.16}$	$99.01_{\pm 0.83}$	74.06	16.64	73.96	18.60
SCRUB	$52.53_{\pm 9.49}$	$95.67_{\pm 1.53}$	$96.29_{\pm 2.78}$	$0.00_{\pm 0.00}$	$95.00_{\pm 2.00}$	$98.83_{\pm 0.48}$	73.49	15.13	72.37	17.49
SSD	$37.47_{\pm 30.91}$	$90.33_{\pm 4.04}$	$99.12_{\pm 0.58}$	$36.67_{\pm 32.30}$	$94.67_{\pm 2.52}$	$92.16_{\pm 11.02}$	75.43	18.09	80.80	28.24
SalUn	$33.60_{\pm 2.23}$	$64.67_{\pm 3.06}$	$89.74_{\pm 1.90}$	$38.13_{\pm 1.10}$	$70.67_{\pm 3.06}$	$94.68_{\pm 0.28}$	91.05	1.73	84.24	0.40
UNSIR	$11.80_{\pm 17.25}$	$95.67_{\pm 2.31}$	$97.37_{\pm 0.23}$	$51.73_{\pm 18.15}$	$95.33_{\pm 1.53}$	$98.71_{\pm 0.56}$	71.82	13.31	70.36	14.27
Joint continual-learning and machine-unlearning methods										
UniCLUN	$35.13_{\pm 1.27}$	$74.33_{\pm 2.52}$	$90.58_{\pm 1.55}$	$39.33_{\pm 0.31}$	$70.00_{\pm 2.00}$	$95.15_{\pm 1.88}$	91.90	3.22	85.33	0.49
LwU (ours)	$0.07_{\pm 0.12}$	$88.00_{\pm 5.20}$	$97.43_{\pm 1.30}$	$0.07_{\pm 0.12}$	$90.33_{\pm 1.53}$	$99.62_{\pm 0.43}$	70.81	14.02	75.25	23.07
RTI podium: 🥇 **LwU (ours)** (3G, 2S, 0B; **13 pts**) 🥈 **SalUn** (3G, 1S, 0B; **11 pts**) 🥉 **SCRUB** (1G, 1S, 4B; **9 pts**)										
Overall podium: 🥇 **LwU (ours)** (14G, 2S, 3B; **49 pts**) 🥈 **SalUn** (7G, 6S, 6B; **39 pts**) 🥉 **Bad Teacher** (6G, 5S, 2B; **30 pts**)										

3.4 Machine Unlearning Performance

Table 2 reports per-example membership auditing using U-LiRA Hayes et al. (2024). The reason that we adopted this test is that aggregate membership-inference accuracy is a weak audit, since a method can, on average, be at chance while individual forgotten examples remain identifiable. For each cell, we train $N_{\text{sh}} = 16$ shadow models, place each forgotten example into exactly $N_{\text{sh}}/2$ shadow training sets, and apply the same unlearning algorithm to each shadow model. For a forgotten example x, we fit Gaussians to its logit-scaled confidence under the shadows that saw it and those that did not, sharing a single variance across examples because per-example estimates from $N_{\text{sh}}/2$ shadows are too noisy, and score the target model using the log-likelihood ratio. The paired Retrain model provides the negatives, so the attack tests whether the unlearned model is distinguishable from Retrain for each individual example. We report Area Under the Curve (AUC) and True Positive Rate (TPR) at a 1% False Positive Rate (FPR), and the two tell different stories. The attack separates these models from Retrain easily on average, but barely makes a confident individual call. Table 2 reveals that Forget-set accuracy is a poor proxy for forgetting. On CIFAR-100 instance deletion with the CNN backbone, SalUn achieves 74.9% on $\mathcal{D}_f$ against a retrain oracle of 69.5%, a five-point gap, and

takes the silver medal. Under U-LiRA the same checkpoint scores an AUC of 89.45. LwU sits at 95.6%, twenty-six points above the oracle and off the podium, and scores 55.1. One of these two models is nearly identifiable on an example-by-example basis, but it is not the one indicated by the accuracy column. Accuracy on $\mathcal{D}_f$ records whether a label survives, not whether a membership signal does, and the two come apart as soon as the deleted examples are individually generalizable.

Where LwU falls short. Subclass deletion. On CIFAR-100, LwU leaves 50.0% (CNN) and 74.3% (ResNet-18) against an oracle of 23.3% and 24.0%; on RTI with ResNet-18 it leaves 74.33% against 24.00%. We read this as a limit of synaptic partitioning rather than of the update rule. A subclass shares almost all of its Fisher support with its parent class, so Zone B is close to empty, and the retain-orthogonal ascent step has little exclusive mass to act on. Parameter importance separates a subclass from its parent only weakly, and the partition inherits that weakness.



Figure 3: Suppressing nudity in text-to-image generation remaining classes of Fig 1. Red borders indicate detected nudity; black boxes are for presentation only.

RTI resists everything. Every RTI entry in the U-LiRA columns sits at least twenty points from the oracle. No method on this dataset yields a model that an attacker cannot separate from one that never saw the deleted trajectories. With 20 participants and a per-participant control signature, the deleted examples are nearly unique, and memorization is no longer distinguishable from generalization as CIFAR permits. Every method in this table was designed on data where most forget examples have near duplicates in $\mathcal{D}_r$. RTI has none, and the gap that opens there is the honest measure of how far instance-level unlearning still has to go. This confirms that most MU methods, when it comes to real-world datasets, fall drastically short, and there is a need for MU methods capable of handling more challenging datasets in Robotics and PAI.

4 Conclusion

We introduced Learning with Unlearning (LwU), which combines CL and MU in a single model by deciding, weight by weight, which objective each parameter serves. Importance scores split the network into four zones. Retention is frozen; pure-forget weights are cleared and returned to the plastic pool; contested weights take an ascent step orthogonal to the retention gradient; and the remainder learns the next task with the model's previous self as the teacher. No buffer, no retraining. Across three datasets, two backbones, and both continual learning and machine unlearning tasks, LwU leads the overall podium in each, through a single mechanism.

Limitations and Future Work. Instance deletion on RTI is unsolved for every method we tested. This weakened the practicality of every MU method in the PAI domain. This needs more careful weight tuning and pruning, which is out of scope of this manuscript and will be investigated for future work.

AI USE STATEMENT

In this work, we used generative AI tools to polish the text, fix typos, improve wording, and check grammar. We have not used generative AI tools for methodology development, dataset creation, or BibTeX reference creation (No hallucination). LLM-generated code was verified and tested for correctness and it is runnable in uploaded supplementary material. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

Anonymous GitHub: https://anonymous.4open.science/r/Learning-with-Unlearning-2957/

We provided a complete version of the code to reproduce every result of the paper at the anonymous GitHub link and uploaded it to the supplementary materials section on the submission page. The code and the RTI dataset will be published upon acceptance of the paper.

REFERENCES

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

Jonathan Brophy and Daniel Lowd. Machine unlearning for random forests. In *ICML*, 2021.

Pietro Buzzega, Matteo Boschini, Angelo Porrello, Davide Abati, and Simone Calderara. Dark experience for general continual learning: a strong, simple baseline. In *NeurIPS*, 2020.

Yinzhi Cao and Junfeng Yang. Towards making systems forget with machine unlearning. In *IEEE Symposium on Security and Privacy*, 2015.

Sungmin Cha and Kyunghyun Cho. Hyperparameters in continual learning: A reality check. *TMLR*, 2025.

Romit Chatterjee, Vikram Chundawat, Ayush Tarun, Ankur Mali, and Murari Mandal. A unified framework for continual learning and machine unlearning. *arXiv preprint arXiv:2408.11374*, 2024.

Vikram S Chundawat, Ayush K Tarun, Murari Mandal, and Mohan Kankanhalli. Can bad teaching induce forgetting? unlearning in deep networks using an incompetent teacher. In *AAAI*, 2023.

Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Ales Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE TPAMI*, 2021.

Arthur Douillard, Alexandre Ramé, Guillaume Couairon, and Matthieu Cord. Dytox: Transformers for continual learning with dynamic token expansion. In *CVPR*, 2022.

Chongyu Fan, Jiancheng Liu, Yihua Zhang, Eric Wong, Dennis Wei, and Sijia Liu. SalUn: Empowering machine unlearning via gradient-based weight saliency in both image classification and generation. In *ICLR*, 2024.

Daniel L. Felps, Amelia D. Schwickerath, Joyce D. Williams, Trung N. Vuong, Alan Briggs, Matthew Hunt, Evan Sakmar, David D. Saranchak, and Tyler Shumaker. Class clown: Data redaction in machine unlearning at enterprise scale, 2020. URL `https://arxiv.org/abs/2012.04699`.

Jack Foster, Stefan Schoepf, and Alexandra Brintrup. Fast machine unlearning without retraining through selective synaptic dampening. In *AAAI*, 2024.

Antonio Ginart, Melody Y. Guan, Gregory Valiant, and James Zou. Making ai forget you: Data deletion in machine learning. In *NeurIPS*, 2019.

Aditya Golatkar, Alessandro Achille, and Stefano Soatto. Eternal sunshine of the spotless net: Selective forgetting in deep networks. In *CVPR*, 2020.

Laura Graves, Vineel Nagisetty, and Vijay Ganesh. Amnesiac machine learning. In *AAAI*, 2021.

Chuan Guo, Tom Goldstein, Awni Hannun, and Laurens van der Maaten. Certified data removal from machine learning models. In *ICML*, 2020.

Varun Gupta, Christopher Jung, Seth Neel, Aaron Roth, Saeed Sharifi-Malvajerdi, and Chris Waites. Adaptive machine unlearning, 2021. URL https://arxiv.org/abs/2106.04378.

Mustafa Burak Gurbuz and Constantine Dovrolis. NISPA: Neuro-inspired stability-plasticity adaptation for continual learning in sparse networks. In *ICML*, 2022.

Jamie Hayes, Ilia Shumailov, Eleni Triantafillou, Amr Khalifa, and Nicolas Papernot. Inexact unlearning needs more careful evaluations to avoid a false sense of privacy, 2024. URL https://arxiv.org/abs/2403.01218.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *ICCV*, 2015.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016.

Haeyong Kang, Rusty John Lloyd Mina, Sultan Rizky Hikmawan Madjid, Jaehong Yoon, Mark Hasegawa-Johnson, Sung Ju Hwang, and Chang D. Yoo. Forget-free continual learning with winning subnetworks. In *ICML*, volume 162, pp. 10734–10750, 2022.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *ICLR*, 2015.

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

Alex Krizhevsky and Geoffrey Hinton. Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario, 2009. URL https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf.

Meghdad Kurmanji, Peter Triantafillou, Jamie Hayes, and Eleni Triantafillou. Towards unbounded machine unlearning. In *NeurIPS*, volume 36, 2023.

Y. Le and X. Yang. Tiny imagenet visual recognition challenge. Technical report, Standford, 2015.

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

Bo Liu, Qiang Liu, and Peter Stone. Continual learning and private unlearning, 2022. URL https://arxiv.org/abs/2203.12817.

Gaoyang Liu, Xiaoqiang Ma, Yang Yang, Chen Wang, and Jiangchuan Liu. Federaser: Enabling efficient client-level data removal from federated learning models. In *2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS)*, 2021.

Seth Neel, Aaron Roth, and Saeed Sharifi-Malvajerdi. Descent-to-delete: Gradient-based methods for machine unlearning, 2020. URL https://arxiv.org/abs/2007.02923.

Yi-Xing Peng, Yu-Ming Tang, Kun-Yu Lin, Qize Yang, Jingke Meng, Xihan Wei, and Wei-Shi Zheng. Person de-reidentification: A variation-guided identity shift modeling. In *CVPR*, 2025.

Sylvestre-Alvise Rebuffi, Alexander Kolesnikov, Georg Sperl, and Christoph H. Lampert. icarl: Incremental classifier and representation learning. In *CVPR*, 2017.

Sebastian Schelter, Stefan Grafberger, and Ted Dunning. Hedgecut: Maintaining randomised trees for low-latency machine unlearning. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2021.

Ghada Sokar, Decebal Constantin Mocanu, and Mykola Pechenizkiy. SpaceNet: Make free space for continual learning. *Neurocomputing*, 2021.

Ayush K. Tarun, Vikram S. Chundawat, Murari Mandal, and Mohan Kankanhalli. Fast yet effective machine unlearning. *IEEE Transactions on Neural Networks and Learning Systems*, 2024.

Mohammad Ali Vahedifar and Qi Zhang. Signal prediction for loss mitigation in tactile internet: a leader-follower game-theoretic approach. In *IEEE MLSP*, 2025.

Mohammad Ali Vahedifar and Qi Zhang. Shapley features for robust signal prediction in tactile internet. In *ICASSP*, 2026.

Mohammad Ali Vahedifar, Abhisek Ray, and Qi Zhang. Shapley neuron values for continual learning: Which neurons matter most?, 2026a. URL https://arxiv.org/abs/2605.15877.

Mohammad Ali Vahedifar, Qi Zhang, and Alexandros Iosifidis. Towards lifelong deep learning: A review of continual learning and unlearning methods, 2026b. URL https://doi.org/10.5281/zenodo.18774099. Preprint, version 2.

Paul Voigt and Axel von dem Bussche. *The EU General Data Protection Regulation (GDPR): A Practical Guide*. Springer Cham, 2nd edition, 2024.

Junxiao Wang, Song Guo, Xin Xie, and Heng Qi. Federated unlearning via class-discriminative pruning. In *The ACM Web Conference (WWW)*, pp. 622–632, 2022.

Heng Xu, Tianqing Zhu, Lefeng Zhang, Wanlei Zhou, and Philip S. Yu. Machine unlearning: A survey. *ACM Computing Surveys*, 2023.

Michal Zajac, Tinne Tuytelaars, and Gido M van de Ven. Prediction error-based classification for class-incremental learning. In *ICLR*, 2024.

Friedemann Zenke, Ben Poole, and Surya Ganguli. Continual learning through synaptic intelligence. In *ICML*, 2017.

Da-Wei Zhou, Qi-Wei Wang, Han-Jia Ye, and De-Chuan Zhan. A model or 603 exemplars: Towards memory-efficient class-incremental learning. In *ICLR*, 2023.

APPENDIX OVERVIEW

The Appendix mainly contains additional materials and experiments that cannot be reported due to the page limit, which is organized as follows:

- Section A contains detailed Related Work of Continual Learning, Machine Unlearning, and unified CL–MU approaches.
- Section B describes the real-world Physical AI scenario that motivates LwU, together with the collection protocol of the Robotic Tactile Internet (RTI) dataset.
- Section C defines evaluation metrics for both CL (Section C.1) and MU (Section C.2) jobs.
- Section D: Limitations: "Things We Tried That Did Not Work" summarizes the main ideas and implementation that ultimately failed.
- Section E gives the pseudo-code of LwU over a task stream (Algorithm 1), of one deletion request (Algorithm 2), and of Test-Time Memorization (Algorithm 3).
- Section F adapts LwU to conditional image generation: nudity suppression in Stable Diffusion v1.4 (Tables 3 and 4, Fig. 4).
- Section G reports the remaining experiments:
  - ablation study (Section G.1): final zone distributions (Fig. 5); LwU in continual learning with deletion requests (Table 5); component ablations (Tables 6–10); sensitivity and design ablations (Tables 11 and 12); gap to Retrain (Table 13); importance-estimator ablation (Tables 14 and 15); Zone-C orthogonality (Table 16); test-time memory ablation (Table 17); validation-selected configurations (Table 18);
  - per-example privacy (Section G.2): U-LiRA ROC curves on RTI (Fig. 6), CIFAR-10 (Fig. 7) and CIFAR-100 (Fig. 8);
  - unlearning wall-clock cost per method (Section G.3, Fig. 12);
  - full unlearning results (Section G.4) on CIFAR-10, CIFAR-100 and RTI for the CNN under class, subclass and instance deletion (Tables 19, 20, 21) and for ResNet-18 (Tables 22, 23, 24);
  - qualitative per-class removal on CIFAR-10 (Section G.5, Figs. 13–22);
  - continual-learning cost (Section G.6): training cost (Table 25), memory and storage for TIL and CIL (Tables 26 and 27), inference cost (Table 28).
- Section H gives the hyperparameter search space and selected configurations on Seq-CIFAR-100 (Table 35) and RTI (Table 37).

A RELATED WORK

Continual Learning methods aim to mitigate catastrophic forgetting when learning from a sequence of tasks. Memory-based approaches store past samples and replay them during training (Rebuffi et al., 2017; Buzzega et al., 2020; Douillard et al., 2022; Zhou et al., 2023). While effective, such methods suffer from key limitations: (i) *privacy concerns*, as storing raw data may violate regulations such as the General Data Protection Regulation (GDPR) (Voigt & Bussche, 2024); (ii) *scalability issues*, since memory requirements grow with the number of tasks; and (iii) *computational overhead* due to repeated replay.

Buffer-free methods avoid storing past data by relying on regularization or architectural constraints. For instance, Elastic Weight Consolidation (EWC) (Kirkpatrick et al., 2017) and Synaptic Intelligence (SI) (Zenke et al., 2017) restrict updates to important parameters, while Learning without Forgetting (LwF) (Li & Hoiem, 2018) distills knowledge from previous model outputs. More recent approaches introduce structural mechanisms, such as sparsity-based allocation (SpaceNet) (Sokar et al., 2021), Neuro-Inspired Stability-Plasticity Adaptation (NISPA) (Gurbuz & Dovrolis, 2022), Winning Sub-Network (WSN) (Kang et al., 2022), or Prediction Error-based Classification (PEC) (Zajac et al., 2024). These approaches often rely on additional components (e.g., masks, expansions, or constrained embeddings), which limit flexibility and scalability.

In contrast, we aim for a **buffer-free** and **architecture-expansion-free** approach that operates within a fixed model capacity. Furthermore, existing CL methods typically treat forgetting as undesirable.

We instead explore *beneficial forgetting*, in which the selective removal of knowledge is advantageous, particularly in the context of Machine Unlearning.

Machine Unlearning aims to remove the influence of specific data from trained models. Existing approaches can be broadly categorized into *data-centric* and *model-centric* methods. Data reorganization methods manipulate training data to facilitate unlearning. Data obfuscation techniques inject noise or modify labels to suppress the effect of target samples (Tarun et al., 2024; Felps et al., 2020; Graves et al., 2021). Data pruning methods partition datasets and retrain only affected subsets, reducing computational cost (Bourtoule et al., 2021; Gupta et al., 2021; Neel et al., 2020). Data replacement strategies modify training samples during learning to enable efficient future removal (Cao & Yang, 2015). Model-centric methods operate directly on trained models. Model shifting updates parameters to counteract the influence of removed data (Golatkar et al., 2020; Guo et al., 2020). Model replacement substitutes affected components with precomputed alternatives (Schelter et al., 2021; Brophy & Lowd, 2021), while model pruning removes parameters associated with target data followed by light fine-tuning (Liu et al., 2021; Wang et al., 2022).

Due to the prohibitive cost of retraining large models, we focus on **post hoc unlearning** (Foster et al., 2024), which removes the influence of data after training without retaining additional information during learning. This motivates the integration of a post hoc Machine Unlearning approach into our LwU framework. While addressing Continual Learning tasks, we additionally propose **retraining-free** and **test-time updatable** model for Machine Unlearning, aiming to reduce the computational cost of retraining from scratch while enhancing adaptation at inference time.

Unified Continual Learning and Machine Unlearning. A small but growing body of work recognizes that CL and MU are complementary facets of the same lifecycle problem. Liu et al. (2022) formalized the *Continual Learning and Private Unlearning* (CLPU) problem and proposed CLPU-DER++, which achieves exact unlearning by training independent sub-networks per task and discarding them upon request. While providing formal privacy guarantees, this strategy incurs linear growth in the number of parameters with the number of tasks, making it unsustainable for long task sequences. Chatterjee et al. (2024) introduced UniCLUN, a controlled knowledge distillation framework in which a student model is concurrently guided by multiple teachers to balance new-task acquisition with targeted forgetting. UniCLUN demonstrated that a single distillation mechanism can serve both objectives; however, it relies on a fixed replay buffer and a separate teacher per learning/unlearning phase, reintroducing the memory and computational overheads that buffer-free CL methods seek to eliminate. More recently, Peng et al. (2025) proposed De-ReID for person de-reidentification, applying a variation-guided identity shift mechanism to unlearn specific individuals from re-identification models while preserving the ability to match retained persons. Although De-ReID addresses a domain-specific instance of joint learning and unlearning, it operates on a single task and does not handle the sequential multi-task setting central to CL.

B Real-World Physical AI Scenario

Physical AI over the Tactile Internet. Physical Artificial Intelligence (PAI) integrates learning-enabled models with embodied systems that perceive, reason, and interact with the physical world. A representative example is robotic teleoperation, where a human operator remotely controls a robot while receiving haptic feedback. The Tactile Internet supports such bidirectional interaction by transmitting control commands and touch sensations over long distances, enabling mission-critical applications such as remote surgery and robotic manipulation (Vahedifar & Zhang, 2025). Because haptic devices typically operate at sampling rates of 1 kHz or higher, the real-time control and inference path must satisfy an extremely low latency budget, commonly targeted at approximately 1 ms (Vahedifar & Zhang, 2026). A deployed PAI system is not trained only once. It continually receives new interaction data as operators, environments, and procedures change. At the same time, some previously observed data may later need to be removed in response to privacy or data-governance requirements, such as requests motivated by the GDPR right to erasure (Voigt & Bussche, 2024). This creates a lifecycle in which continual learning and selective unlearning must coexist: the system must acquire new physical skills without overwriting permitted prior knowledge, while removing the influence of requested data without interrupting deployment or retraining the entire model.

Motivating PAI scenario. Consider a learning-enabled telesurgery system deployed over several months. In the first month, the system observes procedures performed by Surgeon A and learns the corresponding interaction patterns. In the second month, Surgeon B begins using the system, requiring it to adapt to a different operating style without forgetting the knowledge acquired from Surgeon A. In the third month, a patient whose data contributed to an earlier procedure requests that it be removed, while Surgeon C simultaneously begins generating new interaction data. The system must therefore remove the influence of the requested patient, preserve the remaining surgical knowledge, and continue to adapt to new operators.

This scenario imposes three requirements:

1. **Continual adaptation:** the model must learn from new operators and procedures without catastrophically forgetting previously acquired knowledge;
2. **Selective unlearning:** the model must remove the influence of requested patient-level data while preserving knowledge derived from other patients and operators; and
3. **Real-time inference:** the deployed prediction path must remain compatible with the latency requirements of the haptic control loop.

The latency requirement applies to online inference and adaptation, not to the complete unlearning procedure, which does not need to finish within one millisecond. Nevertheless, an unlearning request should not require full retraining or prolonged interruption of the deployed system. Replay-based continual learning is also undesirable because a replay buffer may retain precisely the patient data that must later be deleted. This setting therefore provides a practical motivation for a fixed-capacity, buffer-free framework that supports both continual learning and post-hoc unlearning.

Robotic Tactile Internet dataset. To study this setting, we collected a Robotic Tactile Internet (RTI) dataset containing haptic tele-manipulation signals from a virtual client–server environment implemented in CHAI3D. The experimental platform follows the IEEE 1918.1.1 architecture and uses a Novint Falcon haptic device operating at a 1 kHz control loop. Twenty anonymized participants aged 20 to 33 controlled a virtual tool while performing five actions: no movement, pressing, tapping, free-air movement, and dragging.

Each action was performed for 10 seconds, and every participant repeated the complete five-action sequence four times, giving 20 action segments per participant. Each participant's record contains human-side video, robot-side video, virtual-environment image frames, and the command signals sent from the human side and the feedback signals returned from the robot side. Each participant forms one task whose five classes are that participant's actions, so the full label space has $20 \times 5 = 100$ classes; we evaluate Class-IL and Task-IL. We consider three levels of deletion. *Class deletion* removes all data of one action of one participant, or all five actions for participant-level removal, so the label leaves the output space. *Subclass deletion* removes one of the four repetitions of an action while keeping the label. *Instance deletion* removes individual windows, frames, or video clips of a specific participant. Participant-specific streams emulate sequentially arriving operators and patient- or participant-level removal requests, so learning and deletion interleave throughout the model's operational lifetime, as in a deployed PAI system.

The RTI dataset and its accompanying documentation will be made publicly available following completion of the required legal, ethical, and institutional reviews and confirmation that its release complies with applicable participant consent and data protection requirements.

C Evaluation Metrics

C.1 Continual-Learning Metrics

Let K denote the number of tasks. After completing every task T_i, we evaluate the model on every task T_j, including tasks not yet encountered. This produces the accuracy matrix $\mathbf{A} \in \mathbb{R}^{K \times K}$, where

$$A_{i,j} = \text{accuracy on task } T_j \text{ after completing task } T_i. \tag{15}$$

Accuracies are represented as proportions in the following definitions.

Average Accuracy (ACC). ACC measures final performance across all tasks:

$$\mathrm{ACC} = \frac{1}{K}\sum_{j=1}^{K} A_{K,j}. \tag{16}$$

Higher ACC indicates better overall performance at the end of the task sequence.

Backward Transfer (BWT). BWT measures how subsequent learning changes performance on previously learned tasks:

$$\mathrm{BWT} = \frac{1}{K-1}\sum_{j=1}^{K-1}\left(A_{K,j} - A_{j,j}\right). \tag{17}$$

Negative BWT indicates forgetting, zero indicates no net change, and positive BWT indicates that later tasks improve earlier performance.

Forward Transfer (FWT). FWT measures whether learning previous tasks improves performance on a new task before that task is trained:

$$\mathrm{FWT} = \frac{1}{K-1}\sum_{j=2}^{K}\left(A_{j-1,j} - b_j\right), \tag{18}$$

where b_j is the accuracy of a randomly initialized model on task T_j.

Plasticity. Plasticity measures the normalized improvement produced by learning each new task:

$$P = \frac{1}{K-1}\sum_{j=2}^{K}\frac{A_{j,j} - A_{j-1,j}}{1 - A_{j-1,j} + \varepsilon}, \tag{19}$$

where $\varepsilon > 0$ prevents division by zero. Higher P indicates a greater ability to acquire new knowledge.

Stability. Stability measures how well performance on earlier tasks is preserved:

$$S = 1 - \frac{1}{K-1}\sum_{j=1}^{K-1}\left(A_{j,j} - A_{K,j}\right). \tag{20}$$

Under the definition of BWT above, $S = 1 + \mathrm{BWT}$. Higher values indicate less forgetting.

Plasticity–Stability Score (PS). We summarize the balance between plasticity and stability using their harmonic mean:

$$\mathrm{PS} = \frac{2\,P\,S}{P + S + \varepsilon}. \tag{21}$$

Because the symbol PS appears on both sides if written this way, we denote the reported score by PS and the two constituent quantities by P and S. A high score requires both successful acquisition and retention.

C.2 Machine-Unlearning Metrics

Let f_u denote the unlearned model and f_{rt} the Retrain oracle.

Forget Accuracy. Forget accuracy is the classification accuracy of f_u on $\mathcal{D}_f$. Its desired value depends on the deletion setting and therefore should not be minimized blindly. Instead, it is compared with the accuracy of Retrain.

Absolute Forget-Accuracy Gap. Our primary unlearning metric is

$$|\Delta\mathcal{D}_f| = \left|\mathrm{Acc}_{\mathcal{D}_f}(f_u) - \mathrm{Acc}_{\mathcal{D}_f}(f_{\mathrm{rt}})\right|. \tag{22}$$

A smaller value indicates closer agreement with the model trained without $\mathcal{D}_f$.

Retain and Test Accuracy. Retain accuracy measures utility on $\mathcal{D}_r$, while test accuracy measures generalization on the official test set. These metrics ensure that agreement on $\mathcal{D}_f$ is not obtained by indiscriminately damaging the model.

Membership-Inference Attack. We apply a confidence-threshold membership-inference attack to determine whether forgotten examples remain distinguishable from non-members. The score orientation and decision threshold are fitted on dedicated calibration splits disjoint from the selection and final-evaluation splits. We report attack accuracy and area under the receiver operating characteristic curve (AUC). Values near 50% attack accuracy and 0.5 AUC indicate chance-level discrimination, although they do not by themselves prove successful unlearning.

Predictive KL Divergence. Predictive KL measures the difference between the output distributions of the unlearned and retrained models:

$$D_{\text{out}} = \frac{1}{|\mathcal{E}|} \sum_{x \in \mathcal{E}} D_{\text{KL}} \left(p_u(\cdot \mid x) \,\|\, p_{\text{rt}}(\cdot \mid x) \right), \tag{23}$$

where $\mathcal{E}$ is the specified evaluation set. Lower values indicate closer predictive behavior.

Weight-Histogram KL Divergence. We construct normalized parameter histograms P_u and P_{rt} using identical bins and compute

$$D_{\text{weight}} = D_{\text{KL}} \left(P_u \,\|\, P_{\text{rt}} \right). \tag{24}$$

This metric compares aggregate weight distributions and does not establish parameter-wise or functional equivalence.

Unlearning Time. Unlearning time is the wall-clock time required to transform the source checkpoint into the final unlearned model. Source training and final evaluation are excluded.

Relearning Probe. To test how readily forgotten information can be recovered, we create a copy of the final model and optimize it on $\mathcal{D}_f$ for five epochs using Adam with a learning rate of 10^{-4}. We report forget-set accuracy after each epoch and compare the resulting curve with Retrain. The original evaluated model remains unmodified.

Attack Metrics. Each attack assigns a score $s \in \mathbb{R}$, with larger values indicating membership. Let $\mathcal{S}^+$ and $\mathcal{S}^-$ be the multisets of scores on positives and negatives. For a threshold τ, the true-positive rate (TPR) and false-positive rate (FPR) are

$$\text{TPR}(\tau) = \frac{1}{|\mathcal{S}^+|} \sum_{s \in \mathcal{S}^+} \mathbf{1}[s \geq \tau], \qquad \text{FPR}(\tau) = \frac{1}{|\mathcal{S}^-|} \sum_{s \in \mathcal{S}^-} \mathbf{1}[s \geq \tau]. \tag{25}$$

The receiver operating characteristic (ROC) curve traces $\big(\text{FPR}(\tau), \text{TPR}(\tau)\big)$ as τ varies. The area under it (AUC) is the probability that a random positive outscores a random negative:

$$\text{AUC} = \frac{1}{|\mathcal{S}^+|\,|\mathcal{S}^-|} \sum_{s \in \mathcal{S}^+} \sum_{s' \in \mathcal{S}^-} \Big(\mathbf{1}[s > s'] + \tfrac{1}{2}\,\mathbf{1}[s = s'] \Big), \tag{26}$$

so $\text{AUC} = 0.5$ is chance. AUC averages over all thresholds, so it can hide a small set of confidently identified examples. We therefore also report the TPR at a fixed FPR budget α,

$$\text{TPR}\,@\alpha = \text{TPR}(\tau_\alpha), \qquad \tau_\alpha = \min\{\tau : \text{FPR}(\tau) \leq \alpha\}, \tag{27}$$

with $\alpha = 1\%$. At a calibrated threshold $\tau^\star$, attack accuracy is $\text{Acc}_{\text{MIA}} = \frac{1}{2}\big(\text{TPR}(\tau^\star) + 1 - \text{FPR}(\tau^\star)\big)$, which equals ordinary accuracy when positives and negatives are balanced.

D Things We Tried That Did Not Work

During the development of LwU, we investigated several methodological and experimental alternatives that did not provide reliable improvements. We summarize them below both to clarify the design choices in the final method and to facilitate reproducibility.

1. **Forget-only importance estimation.** Estimating parameter importance using only $\mathcal{D}_f$ could not distinguish forget-specific parameters from parameters shared by retained and forgotten knowledge. Modifying the highest-scoring parameters consequently damaged the retained accuracy.
2. **Binary parameter partitioning.** Dividing parameters into only protected and plastic groups was insufficient. In particular, it could not represent parameters that were predominantly associated with forgetting or simultaneously important to both $\mathcal{D}_r$ and $\mathcal{D}_f$. This motivated the four-zone decomposition.

3. **Resetting all forget-important parameters.** Reinitializing every parameter with high forget importance removed forget-set information, but it also destroyed parameters supporting retained knowledge. Resetting was therefore restricted to forget-dominant parameters, $\mathcal{Z}_B$.
4. **Permanent parameter dampening.** Permanently suppressing forget-related parameters progressively reduced the capacity available for subsequent tasks. This was incompatible with repeated deletion requests in a fixed-capacity model and motivated resetting and recycling the Pure Forget zone.
5. **Unconstrained forget-set gradient ascent.** Applying gradient ascent directly to parameters shared by $\mathcal{D}_r$ and $\mathcal{D}_f$ reduced forget accuracy but caused substantial degradation on the retain set.
6. **Threshold heuristics for the contested parameters.** Before the projection of Eq. equation 8, Zone C was resolved by comparing the two scores against a ratio threshold and dampening whatever exceeded it. Behavior depended strongly on that threshold; it offered no guarantee about the retained loss and introduced hyperparameters that had to be retuned per dataset. The projection replaces both with a geometric condition and no free coefficient.
7. **A single weighted retain–forget objective.** Optimizing a global objective of the form $\mathcal{L}_r - \lambda\mathcal{L}_f$ was highly sensitive to λ. A single coefficient could not reliably control the different parameter-level conflicts between preservation and deletion.
8. **Learning with recycled parameters without a teacher.** Training the reinitialized and plastic parameters using only the incoming task caused them to overwrite behavior associated with earlier tasks. A frozen pre-edit teacher was therefore introduced to preserve the previously permitted behavior.
9. **Unconditioned teacher distillation.** Distilling the complete teacher output after class deletion continued to transmit probability mass through the deleted output classes. We therefore mask the corresponding teacher logits and renormalize the remaining distribution.
10. **Applying class-logit masking to every deletion granularity.** Logit masking represents class deletion exactly, since the deleted label set $\mathcal{F}$ is nonempty, and the conditioning of Eq. equation 11 has something to condition on. Subclass- and instance-level requests leave $\mathcal{F} = \emptyset$; the conditioning becomes the identity, and the teacher supplies no deletion signal at all; unlearning then rests entirely on the partition and the Zone B and Zone C updates. This is the setting in which our results are weakest, and we report it as a limitation rather than a solved case.
11. **Write-before-predict test-time memory.** Updating the fast memory before evaluating the current batch allowed a batch to improve its own prediction. This produced self-confirming and unrealistically optimistic results. The final method, therefore, follows a strictly causal predict-before-write protocol.
12. **Example- or feature-level test-time storage.** Storing inputs or individual feature vectors improved adaptation but violated the buffer-free requirement and introduced memory growth with stream length. We replaced this design with a fixed-size class-by-feature memory whose cost is independent of stream length.
13. **Test-time updates to the slow parameters.** Adapting the plastic parameters on the test stream required a separate argument that adaptation cannot walk the model away from the edit, and made the deployed model a function of the stream it had seen. Confining adaptation to the fast memory removes the question: $\theta^\star$ appears in no test-time update rule.
14. **Unconstrained memory updates for deleted classes.** Without explicit masking, the fast memory could add nonzero corrections to deleted-class logits. We zero these rows after every update and additionally exclude examples whose pseudo-label falls in $\mathcal{F}$ from the write step.

E PSEUDO-CODE

In this section, we present the pseudo-code for LwU, detailing the standard steps of the method.

Algorithm 1 LwU over a task stream: initialization, first task, and tasks with or without a deletion request. Shading marks what each block acts on: teacher , partition , Zone A , Zone B , Zone C , Zone D .

Input: tasks $(\mathcal{D}_1, \mathcal{C}_1), \ldots, (\mathcal{D}_K, \mathcal{C}_K)$ with disjoint class sets; after each T_k ($k < K$) an optional deletion request $R_k \in \{\varnothing,\ (\mathcal{D}_f, \mathcal{F}_{\text{new}}, \mathcal{D}'_r)\}$; quantile q_r; Adam rate η, at most E_T epochs with early stopping on validation loss; old-class distillation weight $\lambda_T \geq 0$, temperature κ_T
Output: θ, freeze mask $\mathbf{A}_K$

Initialization
$\theta \leftarrow$ He-normal convolution kernels, BN $(\gamma, \beta) \leftarrow (1, 0)$
head $W \in \mathbb{R}^{C_{\max} \times d}$, $b \in \mathbb{R}^{C_{\max}}$ allocated once (no rows are appended later)
$\mathbf{A}_0 \leftarrow \varnothing,\ \mathcal{F} \leftarrow \varnothing,\ \mathcal{Y}_0 \leftarrow \varnothing,\ \theta^T \leftarrow \perp$ (no teacher)
for $k = 1$ **to** K **do**
Deletion: skipped when $k = 1$ **or** $R_{k-1} = \varnothing$
  if $k > 1$ **and** $R_{k-1} \neq \varnothing$ **then**
    $(\theta, \mathbf{A}_{k-1}, \mathcal{F}) \leftarrow \textsc{Delete}(\theta, \mathbf{A}_{k-1}, \mathcal{F}, R_{k-1})$ (Alg. 2)
Teacher: refresh on the repaired model
    $\theta^T \leftarrow$ frozen copy of the repaired θ
  end if
Zone D: learn T_k **under the mask** $1 - \mathbf{A}_{k-1}$
  $\mathcal{Y}_k \leftarrow \mathcal{Y}_{k-1} \cup \mathcal{C}_k,\ \mathcal{O}_k \leftarrow \mathcal{Y}_k \setminus \mathcal{F},\ \mathcal{O}^{\text{old}} \leftarrow \mathcal{Y}_{k-1} \setminus \mathcal{F}$
  fresh Adam state $(\mathbf{m}, \mathbf{v}) \leftarrow (\mathbf{0}, \mathbf{0})$, weight decay 0; BN layers in train mode
  for each minibatch $(x, y) \subset \mathcal{D}_k$, until early stopping **do**
    $\mathcal{L} \leftarrow \mathrm{CE}\big(\mathrm{softmax}(f(x;\theta)_{\mathcal{O}_k}),\ y\big)$
    if $\theta^T \neq \perp$ **and** $\lambda_T > 0$ **and** $\mathcal{O}^{\text{old}} \neq \varnothing$ **then**
      $q \leftarrow \mathrm{softmax}\big(f(x;\theta^T)_{\mathcal{O}^{\text{old}}}/\kappa_T\big)$
      $p \leftarrow \mathrm{softmax}\big(f(x;\theta)_{\mathcal{O}^{\text{old}}}/\kappa_T\big)$
      $\mathcal{L} \leftarrow \mathcal{L} + \lambda_T\, \kappa_T^2\, D_{\mathrm{KL}}(q \,\|\, p)$
    end if
    $g \leftarrow \nabla_\theta \mathcal{L} \odot (1 - \mathbf{A}_{k-1})$
    $\theta \leftarrow \mathrm{AdamStep}(\theta, g; \mathbf{m}, \mathbf{v}, \eta)$
  end for
Zone A: consolidate T_k **into the freeze mask, then discard** $\mathcal{D}_k$
  $\phi_r \leftarrow$ Fisher of Alg. 2, line 3, on $\mathcal{D}_k$
  $\mathbf{A}_k \leftarrow \mathbf{A}_{k-1} \cup \{i : \phi_r^i > Q_{q_r}(\phi_r)\}$
Teacher: snapshot for T_{k+1}
  $\theta^T \leftarrow$ frozen copy of θ
end for

Algorithm 2 DELETE: one deletion request, applied between two tasks. Zone D is not updated.

Input: θ, freeze mask $\mathbf{A}$, deleted classes $\mathcal{F}$, request $(\mathcal{D}_f, \mathcal{F}_{\text{new}}, \mathcal{D}'_r)$ with $\mathcal{D}_f \neq \varnothing$ ($\mathcal{F}_{\text{new}} = \varnothing$ for instance or sub-class requests); quantiles q_r, q_f; margin $m \geq 1$; temperature κ; CE weight λ_{CE}; Zone-C rate η_C, steps S_C, budget δ_C; repair rate η_R, epochs E_R, Zone-C repair scale $\lambda_C \geq 0$; head re-scaling flag WA; $\epsilon = 10^{-8}$

Output: θ, updated mask $\mathbf{A}'$, $\mathcal{F}$

$\mathcal{F} \leftarrow \mathcal{F} \cup \mathcal{F}_{\text{new}}$, $\bar{w} \leftarrow \text{mean}_{c \notin \mathcal{F}} \|W[c,:]\|$, BN layers in eval mode throughout

Teacher: frozen copy θ^0

$\theta^0 \leftarrow \theta$

Partition: dual Fisher scores and the four zones

$\phi_s^i \leftarrow \frac{1}{N_s} \sum_{b \subset \mathcal{D}_s} |b| \left(\partial \ell_b(\theta)/\partial \theta_i\right)^2$ for $s \in \{r, f\}$, $\mathcal{D}_r = \mathcal{D}'_r$, $\mathcal{D}_f = \mathcal{D}_f$ ($|b| = 1$ on $\mathcal{D}_f$ for instance requests)

$\tau_s \leftarrow \max\left(Q_{q_s}(\phi_s), 10^{-30}\right)$, $\rho_s^i \leftarrow \phi_s^i / \tau_s$ for $s \in \{r, f\}$

$\mathcal{Z}_A \leftarrow \{i : \rho_r^i > 1,\ \rho_r^i > m\, \rho_f^i\}$

$\mathcal{Z}_B \leftarrow \{i : \rho_f^i > 1,\ \rho_f^i > m\, \rho_r^i\}$

$\mathcal{Z}_C \leftarrow \{i : \max(\rho_r^i, \rho_f^i) > 1\} \setminus (\mathcal{Z}_A \cup \mathcal{Z}_B)$

$\mathcal{Z}_D \leftarrow \{i : \max(\rho_r^i, \rho_f^i) \leq 1\}$

Zone B: hard reset of $\theta_{\mathcal{Z}_B}$, per tensor $\theta^{(\ell)}$

weight matrix or kernel: $\theta^{(\ell)}_{\mathcal{Z}_B} \sim \mathcal{N}(0, 2/\text{fan}_{\text{in}}^{(\ell)})$

1-D tensor (all biases incl. b, BN γ, β): $\theta^{(\ell)}_{\mathcal{Z}_B} \leftarrow 0$

Zone C: retain-orthogonal ascent on $\theta_{\mathcal{Z}_C}$ (skipped if $\mathcal{Z}_C = \varnothing$)

$\theta^{(0)}_{\mathcal{Z}_C} \leftarrow \theta_{\mathcal{Z}_C}$, $\bar{\delta}_C \leftarrow \delta_C \cdot \frac{1}{|\mathcal{Z}_C|} \sum_{i \in \mathcal{Z}_C} \frac{1}{2}(\phi_r^i + \phi_f^i)$

for $s = 0$ **to** $S_C - 1$ **do**

  $g_r \leftarrow \nabla_{\theta_{\mathcal{Z}_C}} \ell(\mathcal{D}'_r)$, $g_f \leftarrow \nabla_{\theta_{\mathcal{Z}_C}} \ell(\mathcal{D}_f)$ (ℓ: full-set mean CE)

  $g^{\perp} \leftarrow g_f - \frac{\langle g_f, g_r \rangle}{\|g_r\|^2 + \epsilon}\, g_r$

  $\theta^{(s+1)}_{\mathcal{Z}_C} \leftarrow \theta^{(s)}_{\mathcal{Z}_C} + \eta_C\, g^{\perp}$, $\Delta \leftarrow \theta^{(s+1)}_{\mathcal{Z}_C} - \theta^{(0)}_{\mathcal{Z}_C}$

  if $\|\Delta\| > \bar{\delta}_C$ **then**

    $\theta^{(s+1)}_{\mathcal{Z}_C} \leftarrow \theta^{(0)}_{\mathcal{Z}_C} + \bar{\delta}_C\, \Delta / \|\Delta\|$

  end if

end for

$\mathcal{Z}_C^{\text{rel}} \leftarrow \{i \in \mathcal{Z}_C : \theta_i^{(S_C)} \neq \theta_i^{(0)}\}$

Teacher repair: $\theta_{\mathcal{Z}_A \cup \mathcal{Z}_B}$ from θ^0, then $\theta_{\mathcal{Z}_C}$ from θ^R

for $(\mathcal{Z}, \theta^T, \eta') \in \big((\mathcal{Z}_A \cup \mathcal{Z}_B, \theta^0, \eta_R),\ (\mathcal{Z}_C, \theta^R, \lambda_C \eta_R)\big)$, second pass only if $\lambda_C > 0$ **do**

  for E_R epochs, each minibatch $(x, y) \subset \mathcal{D}'_r$ **do**

    $\tilde{z}_c \leftarrow f(x; \theta^T)_c$ for $c \notin \mathcal{F}$, $\tilde{z}_c \leftarrow -\infty$ for $c \in \mathcal{F}$

    $q_\kappa^{\mathcal{F}} \leftarrow \text{softmax}(\tilde{z}/\kappa)$

    $\mathcal{L} \leftarrow \kappa^2 D_{\text{KL}}\big(q_\kappa^{\mathcal{F}} \,\big\|\, p_\kappa(\cdot \mid x; \theta)\big) + \lambda_{\text{CE}}\, \text{CE}\big(p_1(\cdot \mid x; \theta), y\big)$

    $\theta_{\mathcal{Z}} \leftarrow \theta_{\mathcal{Z}} - \eta' \nabla_{\theta_{\mathcal{Z}}} \mathcal{L}$ (plain SGD, no optimizer state)

  end for

  $\theta^R \leftarrow$ frozen copy of θ

end for

if WA: $W[c,:] \leftarrow \bar{w}\, W[c,:] / \|W[c,:]\|$ for all $c \notin \mathcal{F}$ (biases untouched)

Zone A: update the freeze mask, release $\mathcal{Z}_B$ and $\mathcal{Z}_C^{\text{rel}}$

$\mathbf{A}' \leftarrow (\mathbf{A} \cup \mathcal{Z}_A \cup \mathcal{Z}_C) \setminus (\mathcal{Z}_B \cup \mathcal{Z}_C^{\text{rel}})$

discard θ^0, θ^R; **return** θ, $\mathbf{A}'$, $\mathcal{F}$

Algorithm 3 Test-Time Memorization: inference with the slow weights fixed. No zone is updated, so no shading.

Input: deployed $\theta^\star = (\text{trunk } h,\ W,\ b)$, never updated (BN in eval mode); test stream $X_1, X_2, \ldots$; mixing weight β; confidence threshold ρ; memory rate η_M; momentum μ; retention $\gamma \in [0, 1]$; row-norm cap $r_{\max}$; deleted classes $\mathcal{F}$

Output: predictions $\widehat{Z}_1, \widehat{Z}_2, \ldots$

Initialization, at the start of every stream

$\mathbf{M}_0 \leftarrow \mathbf{0},\ \mathbf{V}_0 \leftarrow \mathbf{0} \in \mathbb{R}^{C_{\max} \times d},\ \mathcal{O} \leftarrow \{1, \ldots, C_{\max}\} \setminus \mathcal{F}$

for each batch X_t **do**

Read: predict $\widehat{Z}_t$ from $\mathbf{M}_{t-1}$

$H_t \leftarrow h(X_t),\ K_t \leftarrow \mathrm{rowNorm}(H_t)$

$Z_t \leftarrow H_t W^\top + \mathbf{1} b^\top \in \mathbb{R}^{n \times C_{\max}}$

$\widehat{Z}_t \leftarrow Z_t + \beta\, K_t \mathbf{M}_{t-1}^\top$, emit $\widehat{Z}_t$

Write: update $\mathbf{M}_t$

$\widetilde{Z}_{t,nc} \leftarrow Z_{t,nc}$ for $c \in \mathcal{O}$, $-\infty$ otherwise

$P_t \leftarrow \mathrm{softmax}(\widetilde{Z}_t),\ \hat{y}_n \leftarrow \arg\max_c P_{t,nc} \in \mathcal{O}$

$S_t \leftarrow \{n : \max_c P_{t,nc} \geq \rho\}$

if $S_t = \varnothing$ **then**

$(\mathbf{M}_t, \mathbf{V}_t) \leftarrow (\mathbf{M}_{t-1}, \mathbf{V}_{t-1})$

else

$K \leftarrow K_t[S_t]$

$Y_{nc} \leftarrow \mathbb{1}[c = \hat{y}_n] - 1/|\mathcal{O}|$ for $c \in \mathcal{O},\ Y_{nc} \leftarrow 0$ for $c \in \mathcal{F}$

$G \leftarrow \frac{1}{|S_t|} \left(K \mathbf{M}_{t-1}^\top - Y\right)^\top K$

$\mathbf{V}_t \leftarrow \mu \mathbf{V}_{t-1} + G,\ \mathbf{M}' \leftarrow \gamma\, \mathbf{M}_{t-1} - \eta_M \mathbf{V}_t$

$\mathbf{M}_t[c, :] \leftarrow \mathbf{M}'[c, :]\ r_{\max} / \max\left(r_{\max}, \|\mathbf{M}'[c, :]\|\right)$ for all c

$\mathbf{M}_t[c, :] \leftarrow \mathbf{0},\ \mathbf{V}_t[c, :] \leftarrow \mathbf{0}$ for all $c \in \mathcal{F}$

end if

end for

F ADAPTING LWU TO CONDITIONAL IMAGE GENERATION

In classification, unlearning removes the mapping from inputs to a label. In conditional generation, the target of unlearning is the distribution from which the model samples when a condition is provided. We therefore do not delete a forgotten condition c. We redirect it to a replacement condition $\tilde{c}(c)$, so that requesting c yields what the source model generates for $\tilde{c}(c)$.

Setting. We consider a pretrained conditional noise predictor $\epsilon_\theta(x_t, t, c)$ with source parameters θ^0 and forward process:

$$x_t = \sqrt{\bar{\alpha}_t}\, x_0 + \sqrt{1-\bar{\alpha}_t}\, \epsilon, \qquad \epsilon \sim \mathcal{N}(0, I), \quad t \sim \mathcal{U}\{0, \dots, T-1\}, \tag{28}$$

where $\bar{\alpha}_t$ is the cumulative noise schedule. The data are split into a forget set $\mathcal{D}_f$ and a retain set $\mathcal{D}_r$.

Parameter importance. For $s \in \{r, f\}$ and a minibatch $b \subset \mathcal{D}_s$, let

$$\ell_s(\theta; b) = \frac{1}{|b|\, d_x} \sum_{(x_0, c) \in b} \left\| \epsilon_\theta(x_t, t, c) - \epsilon \right\|_2^2 \tag{29}$$

be the denoising loss, where d_x is the dimension of x_0. Up to constants, this loss is the negative log-likelihood of a Gaussian noise model, so its squared gradient estimates the diagonal empirical Fisher information. The importance of parameter i for side s is

$$\phi_s^i = \frac{1}{M} \sum_{m=1}^{M} \Big(\partial_{\theta_i}\, \ell_s\big(\theta^0; b_m^s\big) \Big)^2, \tag{30}$$

computed at the source parameters over M minibatches of equal size for both sides.

Zone partition. The importance scores of the two sides need not share a scale. We normalize each side by its own 99-th percentile,

$$\tau_s = Q_{0.99}(\phi_s), \qquad \hat{\phi}_s^i = \phi_s^i / \tau_s, \tag{31}$$

and call parameter i salient for side s when $\hat{\phi}_s^i > 1$. A salient parameter is dominant for one side when its normalized score exceeds that of the other side by a factor $\gamma > 1$. This scale-free rule replaces the ratio ω of Eq. **??**. The four zones are

$$\mathcal{Z}_A = \{i : \hat{\phi}_r^i > 1,\ \hat{\phi}_r^i > \gamma \hat{\phi}_f^i\} \qquad \text{(safe retain)} \tag{32}$$

$$\mathcal{Z}_B = \{i : \hat{\phi}_f^i > 1,\ \hat{\phi}_f^i > \gamma \hat{\phi}_r^i\} \qquad \text{(pure forget)} \tag{33}$$

$$\mathcal{Z}_C = \{i : \hat{\phi}_r^i > 1 \ \vee\ \hat{\phi}_f^i > 1\} \setminus (\mathcal{Z}_A \cup \mathcal{Z}_B) \qquad \text{(conflict)} \tag{34}$$

$$\mathcal{Z}_D = \{i : \hat{\phi}_r^i \le 1,\ \hat{\phi}_f^i \le 1\} \qquad \text{(plastic)} \tag{35}$$

Zones A, B and C contain every parameter that at least one side relies on. Zone D contains the parameters that neither side relies on, and it is the only zone we update. We denote its mask by $\mathbf{m}_D = \mathbb{1}[i \in \mathcal{Z}_D]$. The redirection is therefore carried entirely by capacity that is not salient for either set, while parameters dominated by the forget set are held at their source values.

Teacher-guided redirection. The frozen source network ϵ_{θ^0} supplies both targets:

$$\begin{aligned} \mathcal{L}(\theta) = & \underbrace{\mathbb{E}_{(x_0,c)\sim\mathcal{D}_f,\, t,\, \epsilon}\, \tfrac{1}{d_x} \left\| \epsilon_\theta(x_t, t, c) - \mathrm{sg}\ \epsilon_{\theta^0}\big(x_t, t, \tilde{c}(c)\big) \right\|_2^2}_{\mathcal{L}_f} \\ & + \lambda\, \underbrace{\mathbb{E}_{(x_0,c)\sim\mathcal{D}_r,\, t,\, \epsilon}\, \tfrac{1}{d_x} \left\| \epsilon_\theta(x_t, t, c) - \mathrm{sg}\ \epsilon_{\theta^0}(x_t, t, c) \right\|_2^2}_{\mathcal{L}_r}, \end{aligned} \tag{36}$$

where sg is the stop-gradient operator. $\mathcal{L}_f$ asks the student, conditioned on c, to reproduce the source prediction under $\tilde{c}(c)$ on noisy forgotten data. $\mathcal{L}_r$ keeps the student's predictions on retained data equal to those of the source. The optimal noise predictor satisfies $\epsilon^\star(x_t, t, c) = -\sqrt{1-\bar{\alpha}_t}\, \nabla_{x_t} \log p_t(x_t \mid c)$. Minimizing $\mathcal{L}_f$ thus drives the student's conditional score under c toward the source score under

$\tilde{c}(c)$, over the region of x_t covered by noised forgotten data. Sampling with c then approximately follows the source reverse process for $\tilde{c}(c)$.

Constrained updates. Updates are restricted to Zone D and to a box around the source parameters. At step k of S, the masked gradient is clipped to unit norm, an Adam step is taken, and the result is projected back:

$$g^{(k)} = \mathbf{m}_D \odot \nabla_\theta \mathcal{L}\big(\theta^{(k)}\big), \qquad g^{(k)} \leftarrow g^{(k)} \min\Big(1, \frac{1}{\|g^{(k)}\|_2}\Big), \tag{37}$$

$$\theta^{(k+\frac{1}{2})} = \mathrm{Adam}_\eta\big(\theta^{(k)}, g^{(k)}\big), \tag{38}$$

$$\theta^{(k+1)} = \theta^0 + \mathbf{m}_D \odot \mathrm{clip}\left(\theta^{(k+\frac{1}{2})} - \theta^0,\, -\delta_D,\, \delta_D\right). \tag{39}$$

The projection makes two properties hold exactly. Parameters outside Zone D remain at θ^0, and $\|\theta_D - \theta_D^0\|_\infty \le \delta_D$. We set $\delta_D = S\eta$. A single Adam step changes each coordinate by at most approximately η, so $S\eta$ approximates the largest per-coordinate displacement that S steps can reach, and the projection turns this bound into a guarantee.

Implementation details. Importance is estimated from 128 examples per side, in 4 minibatches of 32 on CIFAR-10 and 64 minibatches of 2 on Stable Diffusion. We use $\gamma = 1.25$ and $\lambda = 1$. On CIFAR-10, we train one model per forgotten class with $S = 1000$, $\eta = 10^{-4}$, $\delta_D = 0.1$, and 128 examples per side per step. On Stable Diffusion, $\mathcal{D}_f$ and $\mathcal{D}_r$ each contain 800 source-generated images; latents are sampled from the cached VAE posterior, $x_0 = s_{\mathrm{vae}}(\mu + \sigma \odot \xi)$ with $\xi \sim \mathcal{N}(0, I)$; and we use $S = 340$, $\eta = 10^{-5}$, $\delta_D = 3.4 \times 10^{-3}$, with an effective batch of 8 per side from 4 microbatches of 2. All runs use seed 42.

F.1 Results

Table 3: Held-out, multi-seed nudity suppression in Stable Diffusion v1.4.

	Metric	Source SD	LwU (ours)	SalUn	Random mask
Forget	Detection rate (%) ↓	40.0	13.3 ± 3.3	0.6 ± 1.0	$\mathbf{0.0 \pm 0.0}$
	Mean max exposed score ↓	0.414	0.133 ± 0.012	$\mathbf{0.008 \pm 0.003}$	$\mathbf{0.008 \pm 0.007}$
	Context CLIP ↑	0.347	$\mathbf{0.345 \pm 0.001}$	0.319 ± 0.003	0.330 ± 0.003
Retain	CLIP, all ↑	0.338	$\mathbf{0.331 \pm 0.001}$	0.313 ± 0.001	0.315 ± 0.003
	CLIP, clothed people ↑	0.353	$\mathbf{0.344 \pm 0.002}$	0.312 ± 0.000	0.316 ± 0.004
	CLIP, other scenes ↑	0.322	$\mathbf{0.318 \pm 0.001}$	0.315 ± 0.002	0.314 ± 0.002
	LPIPS to Source, all ↓	–	$\mathbf{0.343 \pm 0.004}$	0.547 ± 0.004	0.511 ± 0.024
	LPIPS, clothed people ↓	–	$\mathbf{0.377 \pm 0.007}$	0.592 ± 0.007	0.561 ± 0.020
	LPIPS, other scenes ↓	–	$\mathbf{0.309 \pm 0.005}$	0.502 ± 0.001	0.461 ± 0.028
	False-positive detections (%) ↓	3.3	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$	$\mathbf{0.0 \pm 0.0}$

Evaluation uses 90 held-out prompts (30 adult-nudity, 30 matched clothed-person, 30 unrelated scenes). The prompts were fixed before sampling and share no text with the training or development prompts. Each prompt is sampled with two fixed noise draws, giving 180 images per model. Trained methods report mean ± SD over training seeds. Detection uses NudeNet 3.4.2 (exposed classes, score ≥ 0.6). CLIP is the ViT-B/32 image–text cosine. LPIPS (AlexNet) is measured against the Source image generated with the same prompt and noise. "Context" is the CLIP alignment of forget-prompt images to the same prompt with the nudity word removed. Bold marks the best trained method in each row.

Table 4: Paired differences on the held-out prompts (method minus reference).

Metric	LwU − SalUn	Random mask − SalUn
Forget detection rate (pp) ↓	$+12.8\ [+3.9,\ +23.3]$	$-0.6\ [-1.7,\ 0.0]$
Forget max exposed score ↓	$+0.125\ [+0.055,\ +0.206]$	$-0.000\ [-0.015,\ +0.012]$
Forget context CLIP ↑	$+0.027\ [+0.019,\ +0.035]$	$+0.011\ [+0.005,\ +0.017]$
Retain CLIP, all ↑	$+0.018\ [+0.012,\ +0.024]$	$+0.002\ [-0.001,\ +0.005]$
Retain CLIP, clothed people ↑	$+0.032\ [+0.022,\ +0.042]$	$+0.005\ [-0.000,\ +0.010]$
Retain CLIP, other scenes ↑	$+0.003\ [-0.001,\ +0.007]$	$-0.001\ [-0.003,\ +0.001]$
Retain LPIPS, all ↓	$-0.204\ [-0.228,\ -0.181]$	$-0.036\ [-0.051,\ -0.022]$
Retain LPIPS, clothed people ↓	$-0.215\ [-0.250,\ -0.181]$	$-0.031\ [-0.049,\ -0.013]$
Retain LPIPS, other scenes ↓	$-0.194\ [-0.228,\ -0.163]$	$-0.041\ [-0.067,\ -0.021]$

95% bootstrap confidence intervals over prompts (10,000 resamples; values averaged over noise draws and training seeds). An interval that excludes zero indicates a reliable difference.

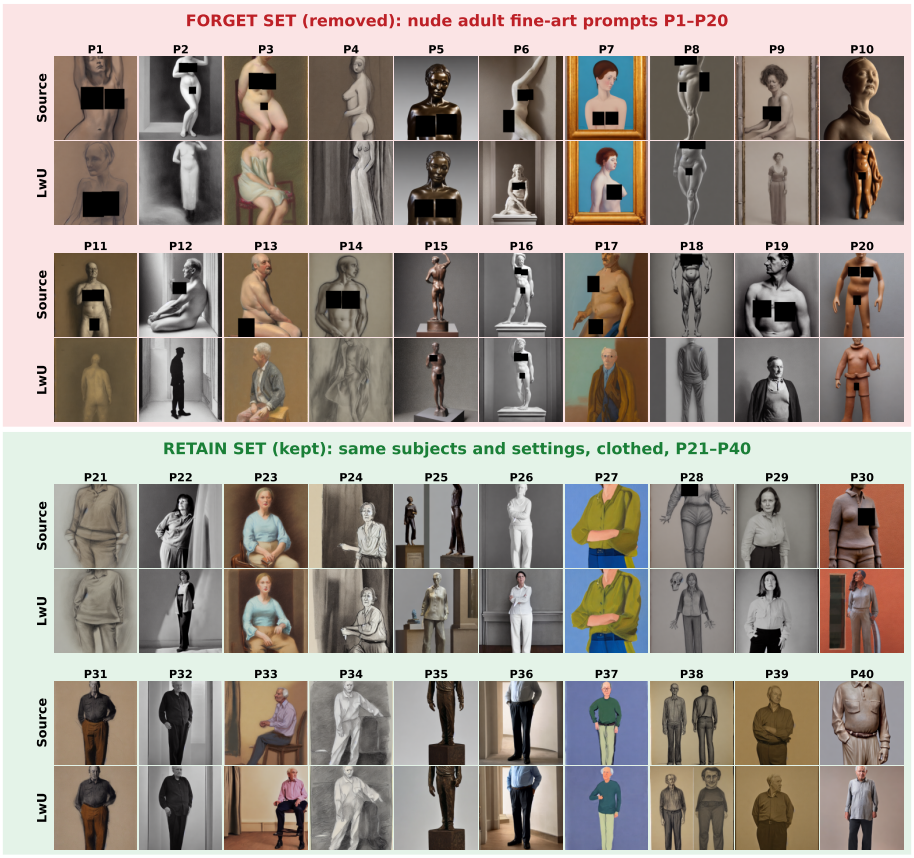


Figure 4: Suppressing nudity in text-to-image generation (set two).

G Additional Results

G.1 Ablation Study

Zone Decomposition analysis: Fig. 5 shows where parameters end up after the final task. On all three benchmarks, LwU uses only about 10% of the network, leaving 89.97% in Zone D, free for future tasks. Parameters that serve only the forget set are almost absent: Zone B holds 0.06% everywhere, so the knowledge to be removed rarely has its own weights. Instead, it sits in Zone C, where parameters serve both the forget and retain sets, and Zone C accounts for 59–95% of the used capacity. Pruning forget-specific weights alone would therefore touch almost nothing, which is why LwU resolves Zone C by orthogonal projection rather than deletion. The split between Zones A and C also tracks the benchmark.

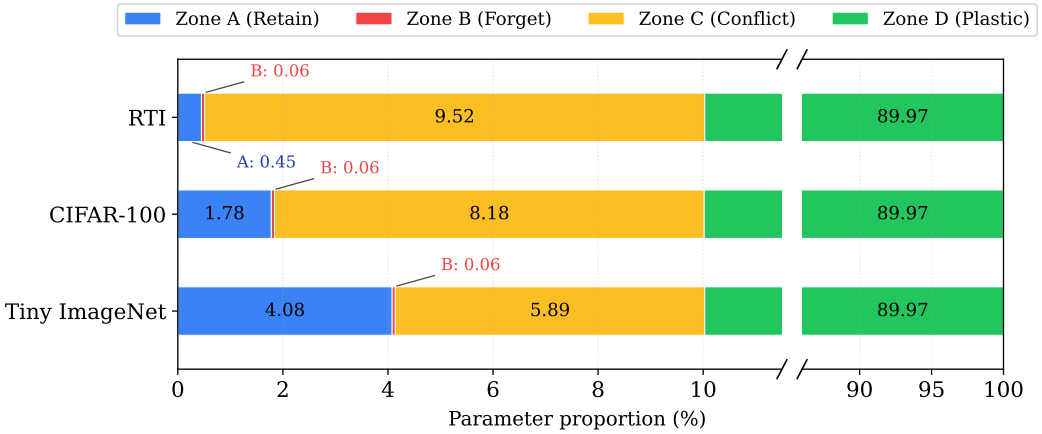


Figure 5: Final zone distributions across datasets.

Table 5: LwU in continual learning with deletion requests (Split CIFAR-10, 5 tasks $\times$ 2 classes, ResNet-18, Task-IL).

Variant	**ACC** $\uparrow$	**BWT** $\uparrow$	$F_{\textbf{del}}$ $\downarrow$	$F_{\textbf{end}}$ $\downarrow$
Continual learning only (no deletion)				
LwU (full)	95.25 ± 0.46	-0.34 ± 0.17	–	97.28 ± 1.42
w/o self-distillation teacher	75.46 ± 8.03	-25.24 ± 9.85	–	64.52 ± 16.25
w/o cumulative freeze mask	96.38 ± 0.04	-0.78 ± 0.40	–	95.57 ± 3.04
w/o test-time memory	95.13 ± 0.44	-0.35 ± 0.14	–	96.67 ± 1.37
Retrain (never sees c_f)	96.27 ± 0.15	-0.35 ± 0.11	–	0.00 ± 0.00
Joint: delete c_f after task 3, then learn tasks 4–5				
w/o B resetting	95.99 ± 0.68	0.47 ± 0.39	5.18 ± 5.62	4.72 ± 4.10
w/o Zone-C step	95.93 ± 0.53	0.52 ± 0.37	7.05 ± 7.38	5.17 ± 6.71
w/o C projection	95.93 ± 0.49	0.51 ± 0.36	7.44 ± 7.85	7.41 ± 6.50
w/o teacher repair (both passes)	94.60 ± 0.89	-1.09 ± 0.99	96.00 ± 2.25	96.22 ± 2.23
w/o Zone-C repair pass	95.48 ± 1.09	-0.00 ± 0.83	80.00 ± 25.76	80.65 ± 24.07
Unmasked teacher	95.44 ± 0.76	-0.11 ± 0.38	94.96 ± 1.99	95.49 ± 2.43
Dominance ω=1	87.05 ± 1.22	-9.50 ± 0.98	0.00 ± 0.00	0.00 ± 0.00
w/o self-distillation teacher	79.95 ± 2.13	-19.56 ± 2.25	0.00 ± 0.00	0.00 ± 0.00
w/o cumulative freeze mask	96.90 ± 0.27	0.03 ± 0.60	12.98 ± 22.48	14.12 ± 24.45
w/o test-time memory	95.91 ± 0.54	0.54 ± 0.29	5.77 ± 5.58	6.89 ± 6.93
Retain subset $\mathcal{D}'_r$ (default 10%)				
$\lvert\mathcal{D}'_r\rvert = 1\%$	95.79 ± 0.74	0.23 ± 0.57	62.85 ± 10.78	66.24 ± 10.25
$\lvert\mathcal{D}'_r\rvert = 5\%$	95.88 ± 0.56	0.42 ± 0.48	25.37 ± 21.76	27.69 ± 24.11
$\lvert\mathcal{D}'_r\rvert = 25\%$	95.89 ± 0.69	0.42 ± 0.48	0.00 ± 0.00	0.00 ± 0.00
Deletion schedule				
Delete after task 2	95.82 ± 0.59	0.37 ± 0.55	17.60 ± 22.34	16.90 ± 18.41
Delete after task 4	95.91 ± 0.68	0.48 ± 0.40	2.16 ± 3.75	1.58 ± 2.74
Two requests (after 2 and 4)	96.60 ± 0.76	1.20 ± 0.66	18.10 ± 22.10	15.46 ± 23.92
LwU (full, delete after task 3)	95.91 ± 0.56	0.42 ± 0.30	5.77 ± 5.58	6.89 ± 6.93

c_f is the first class of task 1. Mean $\pm$ std over seeds 42–44 (%). ACC/BWT over classes never deleted in the run (identical class set for every row). F_{del} / F_{end}: Task-IL accuracy on c_f right after deletion / after the last task (for two requests: the first deleted class). A relearn probe (one Adam epoch on $\mathcal{D}_f$ plus the task's other class) and a confidence MIA on c_f were also run but do not separate the rows: relearn 76–99% for every row including Retrain and MIA 48–53%. Deletion quantiles selected on validation: $q_r = 0.9$, $q_f = 0.99$, $\Delta_C = fisher$, Zone-C pass $\lambda_C = 0.1$ (λ_C=1 diverged for 23/27 grid points). Class-IL is omitted: every old task drops to 0% without replay, so deletion is not measurable.

Table 6: Component ablation of LwU for CIFAR-10 / CNN unlearning (class 0).

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓	Time (s) ↓
Original (before deletion)	91.20 ± 3.05	91.29 ± 0.82	91.28 ± 1.05	56.47 ± 0.65	32.51 ± 1.76	95.53 ± 1.31	–
Retrain (oracle)	0.00 ± 0.00	91.37 ± 0.73	82.23 ± 0.66	50.42 ± 2.13	0.00 ± 0.00	0.00 ± 0.00	–
w/o B resetting	34.50 ± 11.63	91.89 ± 0.80	86.11 ± 1.92	56.35 ± 0.54	13.65 ± 4.59	95.50 ± 0.17	11.21 ± 0.44
w/o Zone-C step	0.00 ± 0.00	91.75 ± 0.73	82.58 ± 0.65	52.97 ± 1.44	0.98 ± 0.31	87.00 ± 12.74	6.63 ± 0.23
w/o C projection	0.00 ± 0.00	91.20 ± 0.85	82.08 ± 0.76	51.47 ± 3.05	1.09 ± 0.27	93.47 ± 2.91	9.05 ± 0.47
w/o teacher repair (both passes)	11.83 ± 20.50	84.14 ± 8.06	76.85 ± 8.86	51.96 ± 3.55	7.28 ± 5.48	95.00 ± 1.51	6.14 ± 0.39
w/o Zone-C repair pass	0.00 ± 0.00	90.96 ± 0.45	81.86 ± 0.41	50.79 ± 2.81	0.80 ± 0.10	90.57 ± 7.61	8.83 ± 0.37
w/o test-time memory	0.00 ± 0.00	91.67 ± 0.63	82.50 ± 0.57	53.52 ± 1.20	1.13 ± 0.31	91.73 ± 5.11	11.73 ± 0.45
+ weight alignment	0.00 ± 0.00	91.64 ± 0.59	82.48 ± 0.52	52.03 ± 2.52	1.01 ± 0.47	91.63 ± 4.99	11.30 ± 0.33
LwU (full)	0.00 ± 0.00	91.67 ± 0.64	82.51 ± 0.60	51.89 ± 2.60	1.03 ± 0.42	91.73 ± 5.11	11.73 ± 0.45

Each row removes or adds one component; everything else is the validation-selected full configuration. Mean ± std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 7: Component ablation of LwU for CIFAR-10 instance / CNN unlearning (500 random training images).

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓	Time (s) ↓
Original (before deletion)	99.74 ± 0.44	91.28 ± 1.05	91.28 ± 1.05	56.92 ± 2.66	5.83 ± 0.76	100.00 ± 0.00	–
Retrain (oracle)	91.28 ± 2.70	91.26 ± 0.90	91.26 ± 0.90	48.21 ± 1.46	0.00 ± 0.00	94.36 ± 2.47	–
w/o B resetting	99.74 ± 0.44	91.36 ± 0.99	91.36 ± 0.99	56.67 ± 3.45	5.73 ± 0.62	100.00 ± 0.00	11.64 ± 3.11
w/o Zone-C step	96.92 ± 3.35	89.82 ± 3.04	89.82 ± 3.04	55.00 ± 3.42	4.87 ± 1.95	97.18 ± 2.22	7.70 ± 1.98
w/o C projection	96.67 ± 3.79	89.76 ± 3.12	89.76 ± 3.12	54.62 ± 3.33	4.84 ± 1.55	96.92 ± 2.66	10.21 ± 2.64
w/o teacher repair (both passes)	71.79 ± 34.49	66.18 ± 32.35	66.18 ± 32.35	51.92 ± 4.80	18.06 ± 19.82	83.85 ± 18.14	5.65 ± 1.44
w/o Zone-C repair pass	96.41 ± 4.24	89.29 ± 3.92	89.29 ± 3.92	54.62 ± 3.53	5.08 ± 0.86	97.44 ± 2.47	10.04 ± 2.99
w/o test-time memory	96.92 ± 3.35	89.88 ± 2.88	89.88 ± 2.88	54.87 ± 2.78	4.80 ± 1.63	97.18 ± 2.22	10.03 ± 0.32
+ weight alignment	97.18 ± 2.91	89.80 ± 2.99	89.80 ± 2.99	53.97 ± 3.11	4.43 ± 2.15	97.44 ± 1.78	14.58 ± 4.36
LwU (full)	96.92 ± 3.35	89.85 ± 2.92	89.85 ± 2.92	54.87 ± 2.78	4.81 ± 1.61	97.18 ± 2.22	10.03 ± 0.32

Each row removes or adds one component; everything else is the validation-selected full configuration. Mean ± std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 8: Component ablation of LwU for CIFAR-10 subclass / CNN unlearning (fine class airplane under 5 coarse labels).

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓	Time (s) ↓
Original (before deletion)	96.00 ± 1.30	93.46 ± 0.59	93.71 ± 0.48	53.40 ± 0.52	23.42 ± 1.34	97.43 ± 0.29	–
Retrain (oracle)	28.67 ± 5.58	93.99 ± 0.22	87.46 ± 0.44	51.01 ± 1.10	0.00 ± 0.00	95.10 ± 0.26	–
w/o B resetting	95.83 ± 1.34	93.68 ± 0.53	93.86 ± 0.44	53.75 ± 0.73	23.45 ± 1.21	97.43 ± 0.38	11.07 ± 0.40
w/o Zone-C step	68.83 ± 4.36	90.55 ± 0.73	88.30 ± 0.82	50.33 ± 1.51	14.78 ± 3.05	94.93 ± 1.42	6.28 ± 0.29
w/o C projection	68.13 ± 4.90	90.92 ± 0.43	88.58 ± 0.75	51.24 ± 2.08	14.41 ± 3.08	94.30 ± 1.48	8.85 ± 0.24
w/o teacher repair (both passes)	9.10 ± 15.50	38.04 ± 11.77	35.01 ± 11.58	49.02 ± 0.34	26.88 ± 8.43	94.07 ± 1.80	5.81 ± 0.21
w/o Zone-C repair pass	64.30 ± 7.28	88.98 ± 2.85	86.45 ± 3.29	51.35 ± 0.49	13.74 ± 0.75	94.40 ± 1.48	8.53 ± 0.38
w/o test-time memory	68.93 ± 2.47	89.80 ± 2.11	87.72 ± 1.91	51.76 ± 1.46	15.11 ± 2.74	94.63 ± 1.40	11.20 ± 0.22
+ weight alignment	68.10 ± 2.25	89.69 ± 2.11	87.51 ± 1.73	51.56 ± 0.22	14.86 ± 3.26	94.50 ± 1.82	11.42 ± 0.62
LwU (full)	69.23 ± 2.87	89.80 ± 2.13	87.69 ± 1.90	51.37 ± 0.89	15.11 ± 2.75	94.63 ± 1.40	11.20 ± 0.22

Each row removes or adds one component; everything else is the validation-selected full configuration. Mean ± std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 9: Component ablation of LwU for CIFAR-10 / ResNet-18 unlearning (class 0).

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓	Time (s) ↓
Original (before deletion)	93.97 ± 1.33	92.42 ± 0.25	92.58 ± 0.11	57.24 ± 0.27	33.48 ± 0.39	96.17 ± 0.57	–
Retrain (oracle)	0.00 ± 0.00	92.82 ± 0.08	83.54 ± 0.08	51.18 ± 0.59	0.00 ± 0.00	0.00 ± 0.00	–
w/o B resetting	77.13 ± 1.78	93.03 ± 0.14	91.42 ± 0.13	57.37 ± 0.31	27.85 ± 0.54	95.57 ± 0.38	18.82 ± 0.36
w/o Zone-C step	0.00 ± 0.00	92.90 ± 0.29	83.61 ± 0.26	53.11 ± 0.71	0.71 ± 0.29	95.37 ± 0.25	11.01 ± 0.12
w/o C projection	0.00 ± 0.00	92.90 ± 0.29	83.61 ± 0.24	53.12 ± 0.86	0.71 ± 0.38	95.37 ± 0.57	14.76 ± 0.63
w/o teacher repair (both passes)	36.23 ± 35.75	90.91 ± 1.34	85.43 ± 2.40	53.28 ± 1.63	13.42 ± 12.68	96.50 ± 0.66	9.42 ± 0.00
w/o Zone-C repair pass	0.00 ± 0.00	92.60 ± 0.28	83.33 ± 0.24	52.72 ± 1.34	0.59 ± 0.41	95.47 ± 0.50	13.95 ± 0.29
w/o test-time memory	0.00 ± 0.00	92.88 ± 0.25	83.59 ± 0.23	53.68 ± 0.95	0.89 ± 0.42	95.40 ± 0.46	17.86 ± 0.33
+ weight alignment	0.00 ± 0.00	92.85 ± 0.32	83.55 ± 0.29	53.45 ± 1.00	0.82 ± 0.41	95.23 ± 0.65	19.50 ± 0.91
LwU (full)	0.00 ± 0.00	92.87 ± 0.26	83.59 ± 0.22	53.44 ± 0.83	0.81 ± 0.37	95.40 ± 0.46	17.86 ± 0.33

Each row removes or adds one component; everything else is the validation-selected full configuration. Mean $\pm$ std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 10: Component ablation of LwU for CIFAR-100 / ResNet-18 unlearning (class 0).

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓	Time (s) ↓
Original (before deletion)	81.67 ± 2.08	70.97 ± 0.43	71.08 ± 0.44	64.77 ± 0.49	32.44 ± 0.27	92.33 ± 1.53	–
Retrain (oracle)	0.00 ± 0.00	70.23 ± 0.14	69.52 ± 0.14	49.85 ± 2.66	0.00 ± 0.00	0.00 ± 0.00	–
w/o B resetting	67.00 ± 1.00	70.97 ± 0.40	70.93 ± 0.38	63.70 ± 0.56	27.20 ± 1.15	94.00 ± 2.65	15.64 ± 0.64
w/o Zone-C step	0.00 ± 0.00	70.45 ± 0.63	69.76 ± 0.64	49.55 ± 2.70	0.76 ± 0.16	0.00 ± 0.00	10.60 ± 0.72
w/o C projection	0.00 ± 0.00	70.31 ± 0.66	69.62 ± 0.67	47.23 ± 2.19	1.20 ± 1.08	0.00 ± 0.00	12.00 ± 0.60
w/o teacher repair (both passes)	9.67 ± 9.81	66.67 ± 4.49	66.10 ± 4.35	50.15 ± 1.85	5.26 ± 5.24	93.00 ± 4.36	6.31 ± 0.22
w/o Zone-C repair pass	0.00 ± 0.00	70.22 ± 0.83	69.52 ± 0.82	49.53 ± 3.06	0.33 ± 0.03	0.00 ± 0.00	11.18 ± 0.59
w/o test-time memory	0.00 ± 0.00	70.40 ± 0.71	69.69 ± 0.70	48.95 ± 1.60	0.52 ± 0.50	0.00 ± 0.00	15.86 ± 0.20
+ weight alignment	0.00 ± 0.00	70.45 ± 0.69	69.75 ± 0.68	49.22 ± 1.21	0.80 ± 0.37	0.00 ± 0.00	15.78 ± 0.61
LwU (full)	0.00 ± 0.00	70.40 ± 0.71	69.70 ± 0.70	48.92 ± 1.54	0.53 ± 0.52	0.00 ± 0.00	15.86 ± 0.20

Each row removes or adds one component; everything else is the validation-selected full configuration. Mean $\pm$ std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 11: Sensitivity and design ablations for LwU, CIFAR-10 / CNN, class 0.

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓
LwU (full, selected)	0.00 ± 0.00	91.67 ± 0.64	82.51 ± 0.60	51.89 ± 2.60	1.03 ± 0.42	91.73 ± 5.11
Zone construction and teacher						
Dominance ω=1	0.00 ± 0.00	31.81 ± 32.79	28.59 ± 29.57	51.08 ± 0.93	20.10 ± 10.93	26.27 ± 45.50
Unmasked teacher	27.77 ± 37.79	91.63 ± 0.79	85.19 ± 3.15	54.14 ± 1.68	10.58 ± 12.57	95.43 ± 1.77
Retain subset $\mathcal{D}'_r$ (default 10%)						
$\lvert\mathcal{D}'_r\rvert = 1\%$	0.00 ± 0.00	87.80 ± 5.10	78.99 ± 4.64	52.37 ± 2.38	1.87 ± 1.85	85.40 ± 18.27
$\lvert\mathcal{D}'_r\rvert = 5\%$	4.90 ± 6.91	91.51 ± 0.66	82.84 ± 0.49	53.30 ± 3.24	2.64 ± 2.52	94.67 ± 3.26
$\lvert\mathcal{D}'_r\rvert = 25\%$	0.00 ± 0.00	91.79 ± 0.61	82.62 ± 0.53	52.39 ± 1.58	0.79 ± 0.34	62.73 ± 54.36
Zone-C budget Δ_C ($\eta_C = 10^{-2}$)						
Δ_C Fisher-scaled (paper)	0.00 ± 0.00	91.69 ± 0.67	82.52 ± 0.60	52.81 ± 1.39	0.90 ± 0.34	85.70 ± 14.97
$\Delta_C = 10^{-4}\lVert\theta_C\rVert$	0.00 ± 0.00	91.69 ± 0.68	82.52 ± 0.60	52.73 ± 1.48	0.88 ± 0.40	85.67 ± 14.87
$\Delta_C = 10^{-3}\lVert\theta_C\rVert$	0.00 ± 0.00	91.71 ± 0.72	82.54 ± 0.66	52.54 ± 1.21	0.82 ± 0.40	87.37 ± 12.37
$\Delta_C = 10^{-2}\lVert\theta_C\rVert$	0.00 ± 0.00	91.69 ± 0.61	82.53 ± 0.56	51.73 ± 2.38	0.99 ± 0.45	91.87 ± 5.27
$\Delta_C = \infty$ (no clip)	0.00 ± 0.00	91.69 ± 0.61	82.53 ± 0.56	51.73 ± 2.38	0.99 ± 0.45	91.87 ± 5.27
Zone-C step size, unclipped						
$\eta_C = 10^{-3}$, no clip	0.00 ± 0.00	91.70 ± 0.68	82.52 ± 0.62	52.67 ± 1.56	0.86 ± 0.37	86.57 ± 13.48
$\eta_C = 10^{-1}$, no clip (3/3 diverged)	–	–	–	–	–	–
$\eta_C = 1$, no clip (3/3 diverged)	–	–	–	–	–	–

Selected Zone-C budget per seed: fisher, inf, inf (relative to $\lVert\theta_C\rVert$; "fisher" = paper scaling). Mean $\pm$ std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge).

Table 12: Sensitivity and design ablations for LwU, CIFAR-10 / ResNet-18, class 0.

Variant	Forget ↓	Retain ↑	Test ↑	MIA → R	Gap ↓	Relearn@1 ↓
LwU (full, selected)	0.00 ± 0.00	92.87 ± 0.26	83.59 ± 0.22	53.44 ± 0.83	0.81 ± 0.37	95.40 ± 0.46
Zone construction and teacher						
Dominance ω=1	0.00 ± 0.00	52.53 ± 33.23	47.27 ± 29.92	50.08 ± 2.37	14.10 ± 10.58	92.40 ± 1.47
Unmasked teacher	2.57 ± 2.96	92.86 ± 0.28	83.83 ± 0.51	54.21 ± 0.70	1.92 ± 1.11	96.03 ± 0.49
Retain subset $\mathcal{D}'_r$ (default 10%)						
$\lvert\mathcal{D}'_r\rvert = 1\%$	6.53 ± 11.32	91.64 ± 0.11	83.14 ± 1.05	52.57 ± 2.68	3.12 ± 4.67	96.73 ± 0.87
$\lvert\mathcal{D}'_r\rvert = 5\%$	46.30 ± 40.12	92.72 ± 0.27	88.06 ± 3.83	55.17 ± 1.95	16.83 ± 13.88	95.80 ± 0.92
$\lvert\mathcal{D}'_r\rvert = 25\%$	0.83 ± 1.44	93.09 ± 0.13	83.85 ± 0.06	54.96 ± 1.15	1.63 ± 0.37	95.70 ± 0.44
Zone-C budget Δ_C ($\eta_C = 10^{-2}$)						
Δ_C Fisher-scaled (paper)	0.00 ± 0.00	92.87 ± 0.26	83.59 ± 0.23	53.19 ± 0.76	0.73 ± 0.31	95.33 ± 0.31
$\Delta_C = 10^{-4}\lVert\theta_C\rVert$	0.00 ± 0.00	92.89 ± 0.27	83.60 ± 0.24	53.26 ± 0.80	0.75 ± 0.34	95.37 ± 0.25
$\Delta_C = 10^{-3}\lVert\theta_C\rVert$	0.00 ± 0.00	92.90 ± 0.24	83.61 ± 0.22	53.49 ± 0.78	0.83 ± 0.33	95.30 ± 0.40
$\Delta_C = 10^{-2}\lVert\theta_C\rVert$	0.00 ± 0.00	92.91 ± 0.25	83.62 ± 0.22	53.54 ± 0.76	0.85 ± 0.35	95.37 ± 0.50
$\Delta_C = \infty$ (no clip)	0.00 ± 0.00	92.91 ± 0.25	83.62 ± 0.22	53.54 ± 0.76	0.85 ± 0.35	95.37 ± 0.50
Zone-C step size, unclipped						
$\eta_C = 10^{-3}$, no clip	0.00 ± 0.00	92.86 ± 0.28	83.58 ± 0.27	53.29 ± 0.80	0.76 ± 0.32	95.33 ± 0.25
$\eta_C = 10^{-1}$, no clip	0.00 ± 0.00	68.19 ± 40.10	61.37 ± 36.07	51.36 ± 2.49	8.76 ± 13.08	93.87 ± 2.10
$\eta_C = 1$, no clip (3/3 diverged)	–	–	–	–	–	–

Selected Zone-C budget per seed: fisher, 0.001, inf (relative to $\lVert\theta_C\rVert$; "fisher" = paper scaling). Mean $\pm$ std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge).

Table 13: Gap to Retrain (↓, mean of $|\Delta|$ Forget, Retain, MIA) for every unlearning ablation across datasets, backbones and deletion granularities.

Variant	**C10 CNN**	**C10 R18**	**C100 R18**	**C10-sub CNN**	**C10-inst CNN**
LwU (full)	1.0 ± 0.4	0.8 ± 0.4	0.5 ± 0.5	15.1 ± 2.7	4.8 ± 1.6
w/o B resetting	13.6 ± 4.6	27.8 ± 0.5	27.2 ± 1.1	23.4 ± 1.2	5.7 ± 0.6
w/o Zone-C step	1.0 ± 0.3	0.7 ± 0.3	0.8 ± 0.2	14.8 ± 3.1	4.9 ± 2.0
w/o C projection	1.1 ± 0.3	0.7 ± 0.4	1.2 ± 1.1	14.4 ± 3.1	4.8 ± 1.6
w/o teacher repair (both passes)	7.3 ± 5.5	13.4 ± 12.7	5.3 ± 5.2	26.9 ± 8.4	18.1 ± 19.8
w/o Zone-C repair pass	0.8 ± 0.1	0.6 ± 0.4	0.3 ± 0.0	13.7 ± 0.8	5.1 ± 0.9
w/o test-time memory	1.1 ± 0.3	0.9 ± 0.4	0.5 ± 0.5	15.1 ± 2.7	4.8 ± 1.6
+ weight alignment	1.0 ± 0.5	0.8 ± 0.4	0.8 ± 0.4	14.9 ± 3.3	4.4 ± 2.1
Random mask	22.9 ± 4.1	31.4 ± 0.4	–	–	–
Magnitude-based	19.4 ± 4.5	30.1 ± 0.3	–	–	–
Fisher retain only	16.3 ± 3.7	28.5 ± 0.2	–	–	–
Fisher forget only	16.0 ± 3.6	28.4 ± 0.3	–	–	–
Dominance ω=1	20.1 ± 10.9	14.1 ± 10.6	–	–	–
Unmasked teacher	10.6 ± 12.6	1.9 ± 1.1	–	–	–

Mean $\pm$ std over 3 seeds; – = not run.

Table 14: Importance-estimator ablation for LwU, CIFAR-10 / CNN, class 0.

Importance estimator	**Score**	**Forget ↓**	**Retain ↑**	**Test ↑**	**MIA → R**	**Gap ↓**	**Relearn@1 ↓**	**Time (s) ↓**
Random mask	$s_i \sim \mathcal{U}(0,1)$	62.60 ± 9.20	91.79 ± 0.87	88.82 ± 1.54	56.18 ± 1.00	22.92 ± 4.08	95.90 ± 0.56	10.70 ± 0.67
Magnitude-based	$s_i = \lvert\theta_i\rvert$	51.40 ± 10.59	91.90 ± 0.81	87.82 ± 1.76	56.64 ± 1.06	19.38 ± 4.46	95.63 ± 0.40	10.44 ± 0.38
Fisher retain only	$s_i = F_{ii}^r$	42.37 ± 8.48	91.86 ± 0.80	86.87 ± 1.55	56.48 ± 0.97	16.30 ± 3.70	95.60 ± 0.44	10.75 ± 0.56
Fisher forget only	$s_i = F_{ii}^f$	41.57 ± 8.27	91.88 ± 0.80	86.80 ± 1.51	56.42 ± 1.05	16.02 ± 3.63	95.60 ± 0.44	10.81 ± 0.38
Dual Fisher (ours)	(F_{ii}^r, F_{ii}^f)	0.00 ± 0.00	91.67 ± 0.64	82.51 ± 0.60	51.89 ± 2.60	1.03 ± 0.42	91.73 ± 5.11	11.73 ± 0.45

Single-score estimators use the same map for the retain and forget sides; all other settings equal the full method (same row as in the component table). Mean $\pm$ std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 15: Importance-estimator ablation for LwU, CIFAR-10 / ResNet-18, class 0.

Importance estimator	**Score**	**Forget ↓**	**Retain ↑**	**Test ↑**	**MIA → R**	**Gap ↓**	**Relearn@1 ↓**	**Time (s) ↓**
Random mask	$s_i \sim \mathcal{U}(0,1)$	87.80 ± 1.57	92.81 ± 0.22	92.30 ± 0.14	57.33 ± 0.17	31.37 ± 0.38	95.83 ± 1.01	16.96 ± 0.54
Magnitude-based	$s_i = \lvert\theta_i\rvert$	83.80 ± 0.82	92.93 ± 0.22	91.97 ± 0.14	57.42 ± 0.53	30.06 ± 0.34	96.00 ± 0.66	17.94 ± 0.70
Fisher retain only	$s_i = F_{ii}^r$	78.87 ± 0.97	93.00 ± 0.15	91.54 ± 0.02	57.73 ± 0.28	28.53 ± 0.20	96.03 ± 0.51	17.77 ± 0.55
Fisher forget only	$s_i = F_{ii}^f$	78.43 ± 1.16	93.01 ± 0.16	91.52 ± 0.04	57.67 ± 0.25	28.37 ± 0.27	96.03 ± 0.49	18.00 ± 0.70
Dual Fisher (ours)	(F_{ii}^r, F_{ii}^f)	0.00 ± 0.00	92.87 ± 0.26	83.59 ± 0.22	53.44 ± 0.83	0.81 ± 0.37	95.40 ± 0.46	17.86 ± 0.33

Single-score estimators use the same map for the retain and forget sides; all other settings equal the full method (same row as in the component table). Mean $\pm$ std over seeds 42–44; accuracies and MIA in %. Forget/Retain: test accuracy on the deleted / retained data; Test: all test classes. MIA → R: closer to Retrain is better. Gap: mean absolute difference to Retrain over Forget, Retain and MIA. Relearn@1: forget accuracy after 1 Adam epoch(s) (10^{-4}, BatchNorm statistics frozen) on $\mathcal{D}_f$ plus as many retained training images (lower = less recoverable knowledge). Time: unlearning wall-clock (importance + edit + repair), best of two repeats with the GPU exclusive.

Table 16: Zone-C orthogonality under the actual update (LwU, CIFAR-10 / CNN, class 0).

Setting	$\max\lvert\cos(g_u^i, g_r)\rvert$	Clip scale	$\frac{\lVert\Delta\theta_C\rVert}{\lVert\theta_C\rVert}$	$\max\lvert\cos(\Delta\theta^{(s)}, g_r)\rvert$	$\max\lvert\cos(\Delta\theta^{(s)}, \bar{g}_r)\rvert$	$\Delta\mathcal{L}_r$	Forget	Retain	MIA
Paper Δ_C (Fisher-scaled), ϵ=10^{-8}	$1.2e-7$	$2.5e-6$	$2.1e-9$	$3.2e-2$	$4.6e-2$	$7.7e-8 \pm 5e-7$	0.00 ± 0.00	91.69 ± 0.67	52.81 ± 1.39
Selected Δ_C, ϵ=10^{-8}	$2.4e-7$	$6.7e-1$	$6.8e-3$	$3.1e-2$	$3.7e-2$	$1.4e-1 \pm 2e-1$	0.00 ± 0.00	91.67 ± 0.64	51.89 ± 2.60
Selected Δ_C, ϵ=0	$2.4e-7$	$6.7e-1$	$6.8e-3$	$3.1e-2$	$3.7e-2$	$1.4e-1 \pm 2e-1$	0.00 ± 0.00	91.67 ± 0.64	51.89 ± 2.60
No clip, ϵ=10^{-8}	$3.0e-7$	$1.0e0$	$6.8e-3$	$2.7e-7$	$8.4e-2$	$1.7e-1 \pm 1e-1$	0.00 ± 0.00	91.69 ± 0.61	51.73 ± 2.38
No clip, ϵ=0	$2.9e-7$	$1.0e0$	$6.8e-3$	$3.1e-7$	$8.4e-2$	$1.7e-1 \pm 1e-1$	0.00 ± 0.00	91.69 ± 0.61	51.73 ± 2.38

g_r: retain gradient used by the step; $\bar{g}_r$: clean (unaugmented) retain gradient; $\Delta\theta^{(s)}$: step actually applied after clipping. Maxima over 5 steps $\times$ 3 seeds; $\Delta\mathcal{L}_r$: change in clean retain CE over the Zone-C stage (mean $\pm$ std).

Table 17: Test-time memory ablation (LwU, CIFAR-10 / CNN, class 0).

Stream ordering	Memory / reset	Forget ↓	Retain ↑	Test ↑	MIA	Flip rate (%)	ΔTest
Dataset order	No memory	0.00 ± 0.00	91.67 ± 0.63	82.50 ± 0.57	53.52 ± 1.20	0.000 ± 0.000	–
	Reset per stream	0.00 ± 0.00	91.67 ± 0.64	82.51 ± 0.60	51.89 ± 2.60	0.313 ± 0.138	$+0.010 \pm 0.030$
	Persistent	0.00 ± 0.00	91.70 ± 0.68	82.52 ± 0.60	53.55 ± 1.03	0.406 ± 0.050	$+0.017 \pm 0.032$
Random†	No memory	0.00 ± 0.00	91.67 ± 0.55	82.50 ± 0.49	53.52 ± 1.04	0.000 ± 0.000	–
	Reset per stream	0.00 ± 0.00	91.68 ± 0.56	82.53 ± 0.50	51.91 ± 2.23	0.292 ± 0.078	$+0.026 \pm 0.012$
	Persistent	0.00 ± 0.00	91.69 ± 0.56	82.53 ± 0.51	53.55 ± 0.89	0.409 ± 0.037	$+0.030 \pm 0.017$
Class-sorted	No memory	0.00 ± 0.00	91.67 ± 0.63	82.50 ± 0.57	53.52 ± 1.20	0.000 ± 0.000	–
	Reset per stream	0.00 ± 0.00	91.94 ± 1.09	82.75 ± 0.98	51.89 ± 2.60	0.820 ± 0.364	$+0.253 \pm 0.443$
	Persistent	0.00 ± 0.00	91.95 ± 1.08	82.68 ± 0.97	53.35 ± 1.00	0.926 ± 0.206	$+0.180 \pm 0.429$

All rows of a seed share the same frozen slow weights, so memory on/off is a paired comparison. Mean $\pm$ std over 3 seeds (†3 permutations $\times$ 3 seeds). A stream is one evaluation set read once (batch 64) with predict-before-write; streams are visited in the order forget-test, retain-test, full test, MIA members, MIA non-members, MIA calibration members/non-members. Reset per stream zeroes $(\mathbf{M}, \mathbf{V})$ at the start of every stream; persistent zeroes them once. Resetting every batch equals no memory.

Table 18: Validation-selected LwU configuration per setting and seed, and the resulting zone sizes (% of parameters).

Setting	Seed	q_r	q_f	Δ_C	η_M	β	ρ	A %	B %	C %
CIFAR-10 / CNN	42	0.7	0.99	fisher	0.01	1.0	0.0	0.00	0.01	29.99
CIFAR-10 / CNN	43	0.9	0.99	inf	0.01	1.0	0.0	0.25	0.07	9.71
CIFAR-10 / CNN	44	0.7	0.95	inf	0.05	1.0	0.8	3.23	0.16	26.66
CIFAR-10 / ResNet-18	42	0.9	0.99	fisher	0.05	0.5	0.8	0.43	0.12	9.51
CIFAR-10 / ResNet-18	43	0.9	0.99	0.001	0.1	1.0	0.8	7.15	0.07	2.81
CIFAR-10 / ResNet-18	44	0.9	0.99	inf	0.1	1.0	0.8	0.53	0.06	9.45
CIFAR-100 / ResNet-18	42	0.5	0.99	fisher	0.1	1.0	0.8	15.34	0.03	34.64
CIFAR-100 / ResNet-18	43	0.9	0.99	fisher	0.01	0.25	0.0	8.72	0.15	1.21
CIFAR-100 / ResNet-18	44	0.7	0.99	0.001	0.01	0.25	0.0	9.75	0.09	20.21
CIFAR-10 subclass / CNN	42	0.9	0.95	0.001	0.01	0.25	0.0	0.63	1.24	8.70
CIFAR-10 subclass / CNN	43	0.9	0.95	fisher	0.01	0.25	0.0	8.80	1.81	0.03
CIFAR-10 subclass / CNN	44	0.9	0.95	fisher	0.01	1.0	0.0	8.48	1.24	0.60
CIFAR-10 instance / CNN	42	0.9	0.95	0.001	0.1	1.0	0.8	7.44	1.60	1.57
CIFAR-10 instance / CNN	43	0.7	0.95	fisher	0.1	1.0	0.8	14.96	0.33	14.82
CIFAR-10 instance / CNN	44	0.7	0.9	fisher	0.05	1.0	0.8	25.80	1.29	3.32
Split CIFAR-10 TIL (joint)	42	0.9	0.99	fisher	0.05	0.5	0.8	0.00	0.00	10.00

Selection uses validation data only. Δ_C relative to $\|\theta_C^{(0)}\|$ ("fisher" = paper scaling, inf = no clipping). Memory: η_M learning rate, β scale, ρ confidence. MU cells fix the Zone-C repair-pass multiplier λ_C=1; the joint protocol additionally searches $\lambda_C \in \{0.1, 1\}$ (selected 0.1; λ_C=1 diverged for 23 of 27 grid points) and uses fixed CL memory settings.

G.2 Per-Example Privacy: U-LiRA

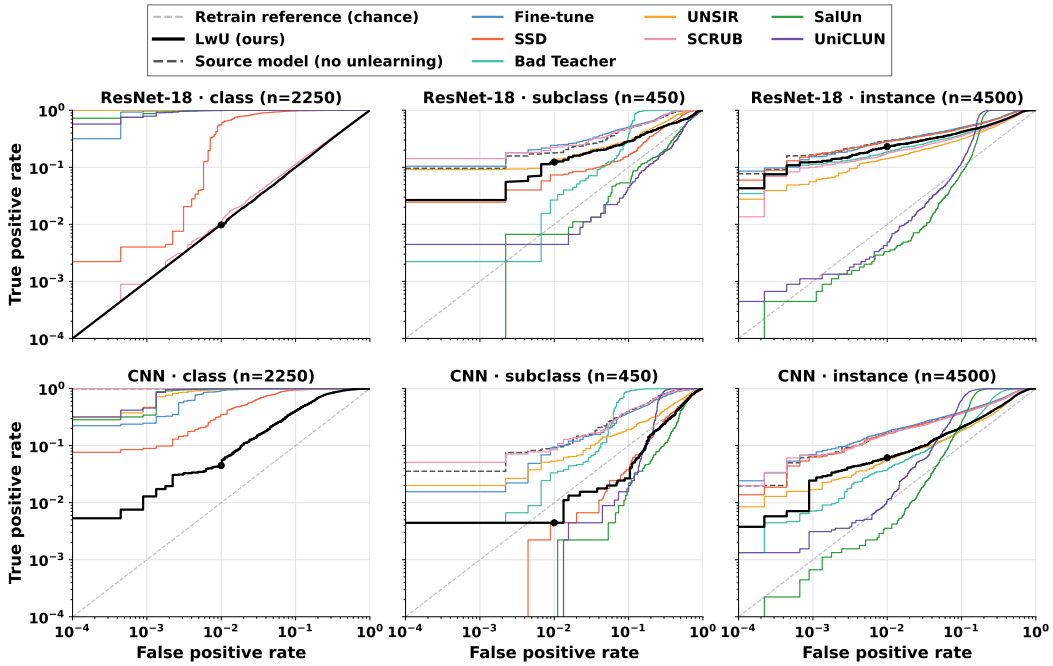


Figure 6: **U-LiRA on RTI: LwU reaches chance on ResNet-18 class and CNN subclass deletion, but still leaks on CNN class and on instance deletion.**

Log–log ROC of the per-example membership attack (rows: backbones; columns: deletion granularity; n: number of forgotten examples). A method that fully removes the forget set follows the grey diagonal; the dot marks LwU at 1% FPR. On class deletion the original model is fully exposed (AUC and TPR@1% both 100). LwU lowers AUC to 50.61 on ResNet-18 but only to 82.48 on the CNN. On subclass deletion, AUC falls to 49.26 on the CNN (original: 81.57) and 64.12 on ResNet-18 (original: 83.18). On instance deletion, TPR@1% drops from 29.22 to 23.07 (ResNet-18) and from 16.51 to 6.13 (CNN). Some baselines leak less at low FPR in individual cells.

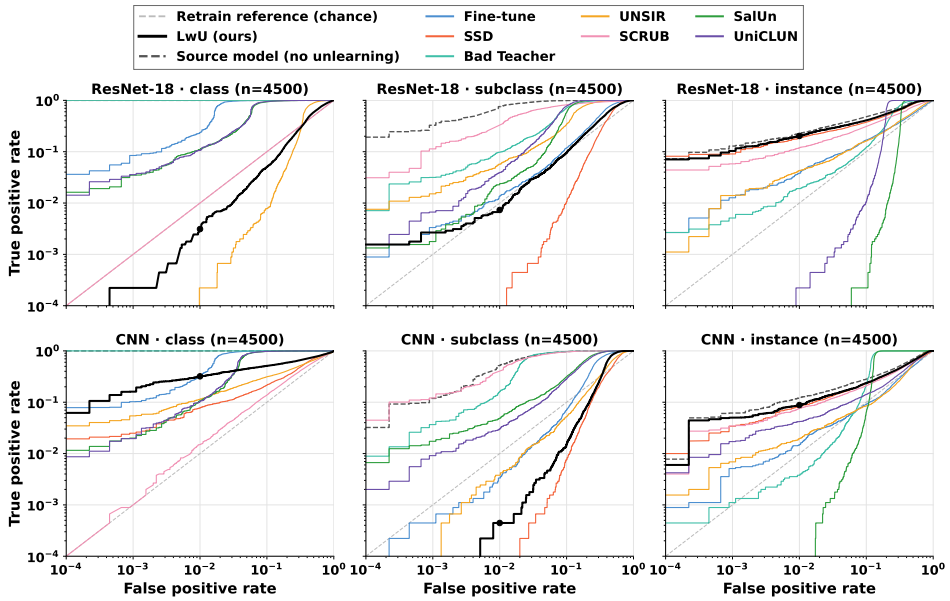


Figure 7: **U-LiRA on CIFAR-10: LwU suppresses low-FPR leakage on subclass and ResNet-18 class deletion, but not on CNN class deletion; on instance deletion no method separates from the original model.**

Layout as in Fig. 6. On class deletion, LwU lowers AUC from 100 to 53.38 on ResNet-18 (TPR@1%: 0.31) but only to 73.28 on the CNN (31.96). On subclass deletion, TPR@1% falls to 0.73 and 0.04 (original: 75.62 and 52.00; ResNet-18 and CNN). On instance deletion, LwU stays close to the original model (AUC 77.72 vs. 80.08 and 62.52 vs. 66.78). Some baselines leak less at low FPR in individual cells.

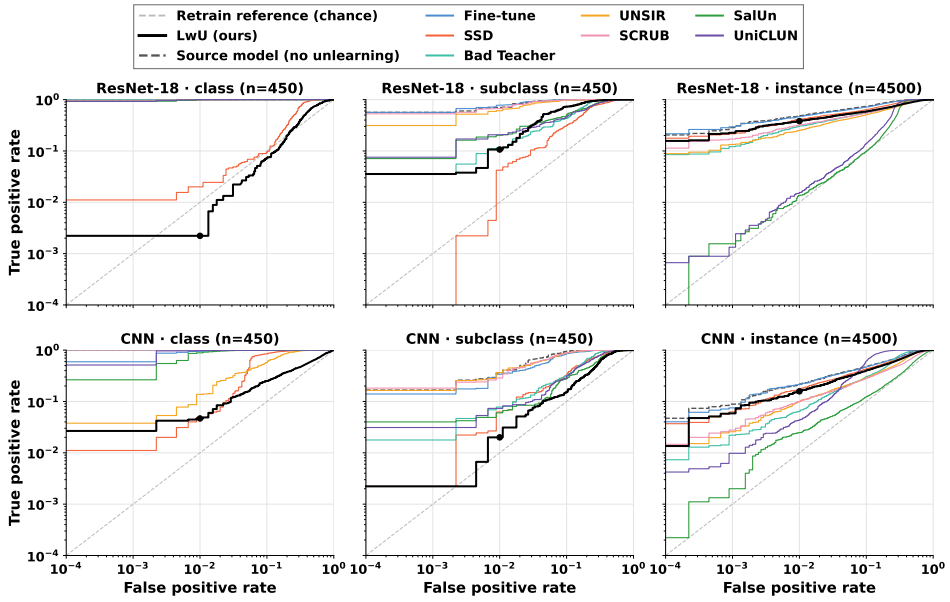


Figure 8: **U-LiRA on CIFAR-100: LwU strongly reduces class-level leakage on both backbones; subclass and instance leakage remains, least on the CNN.**

Layout as in Fig. 6. On class deletion, LwU lowers AUC from 100 to 60.64 on ResNet-18 (TPR@1%: 0.22) and 61.63 on the CNN (4.67); the strongest baseline, SSD, reaches 2.00 and 4.89. TPR@1% falls from 70.56 to 10.67 (ResNet-18) and from 36.00 to 2.00 (CNN) on subclass deletion, and from 48.38 to 38.11 and from 21.80 to 15.76 on instance deletion.

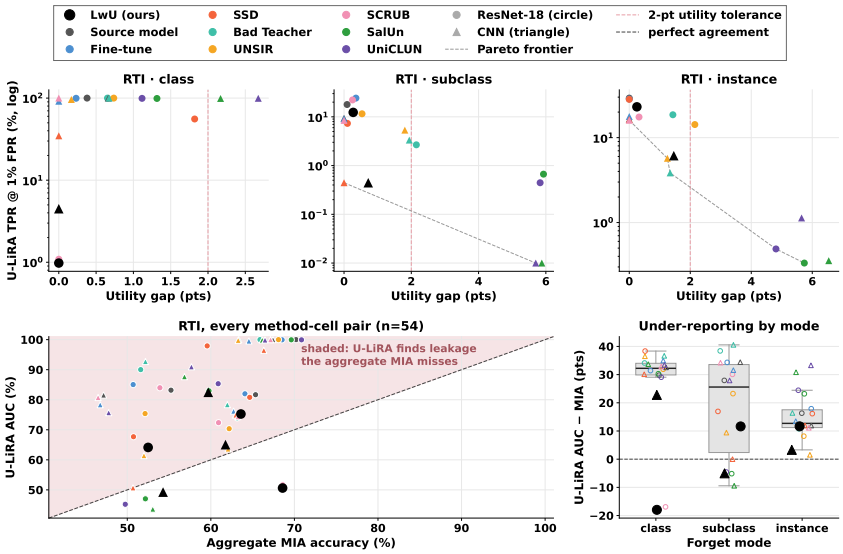


Figure 9: **Utility vs. leakage on RTI: the aggregate MIA under-reports per-example leakage in 48 of 54 method–cell pairs.**

Colour: method; shape: backbone (circle: ResNet-18, triangle: CNN). *Top:* TPR@1% (log scale) vs. utility gap, the larger of the retain- and test-accuracy shortfalls relative to Retrain; lower-left is better. Dashed line: Pareto frontier; pink line: 2-point utility tolerance. LwU stays within tolerance in every cell (gap ≤ 1.46). *Bottom left:* U-LiRA AUC vs. aggregate MIA accuracy. Points above the diagonal leak more than the aggregate MIA detects (median 23.9, max 40.6 points). The largest exception is LwU on ResNet-18 class deletion: U-LiRA is at chance (50.6), yet the aggregate MIA reports 68.6. *Bottom right:* median under-reporting by granularity: 32.2 (class), 25.6 (subclass), 12.7 (instance).

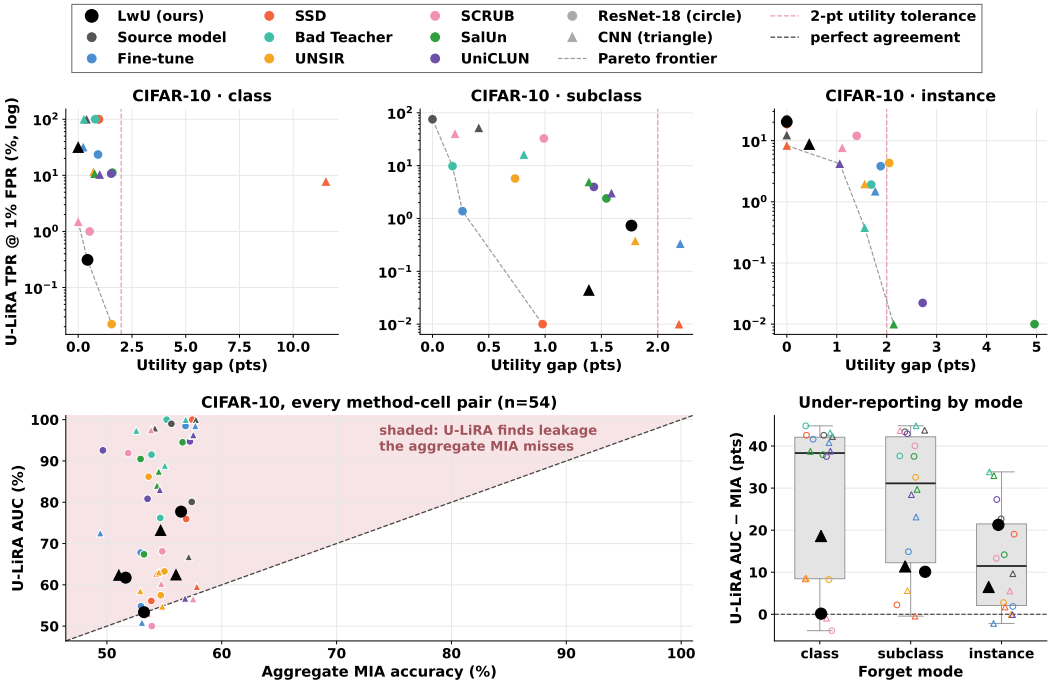


Figure 10: **Utility vs. leakage on CIFAR-10: the aggregate MIA under-reports per-example leakage in 48 of 54 method–cell pairs.**

Layout as in Fig. 9. LwU stays within the utility tolerance in every cell (gap ≤ 1.77). Under-reporting has a median of 22.1 and a max of 44.8 points, and it is largest where deletion is most complete: 38.3 (class), 31.1 (subclass), 11.5 (instance). The remaining six pairs overstate leakage by at most 3.9 points.

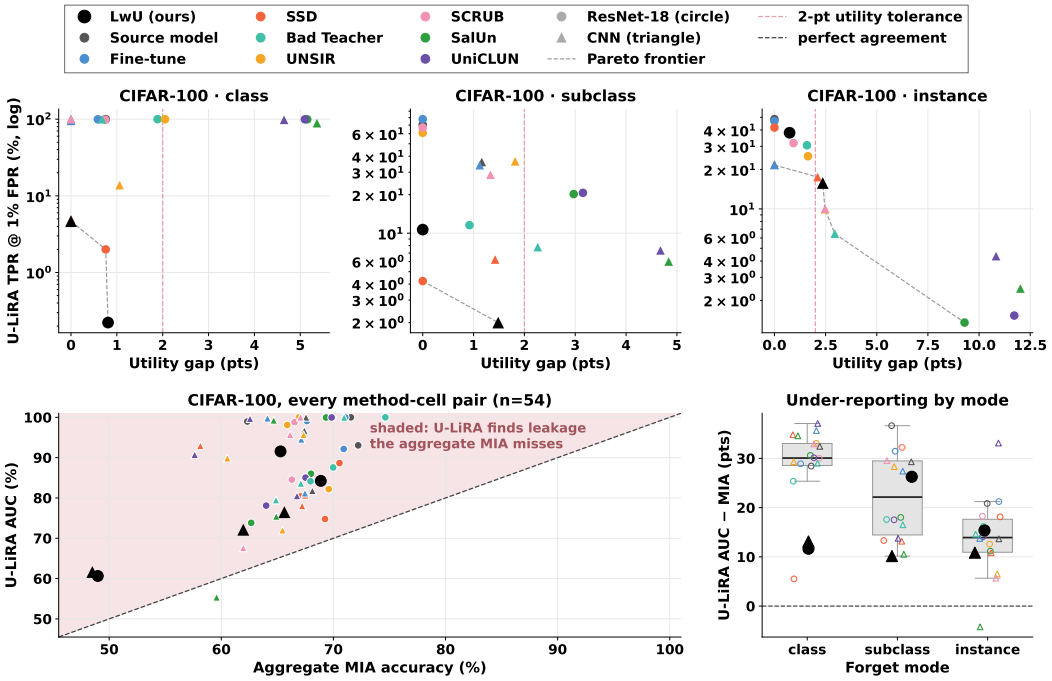


Figure 11: **Utility vs. leakage on CIFAR-100: the aggregate MIA under-reports per-example leakage in 53 of 54 method–cell pairs.**

Layout as in Fig. 9. LwU stays within the utility tolerance in five of six cells; CNN instance deletion sits just outside (2.36). Under-reporting has a median of 19.5 and a max of 37.1 points: 30.1 (class), 22.1 (subclass), 13.9 (instance). The single pair below the diagonal sits at -4.2.

G.3 UNLEARNING COST

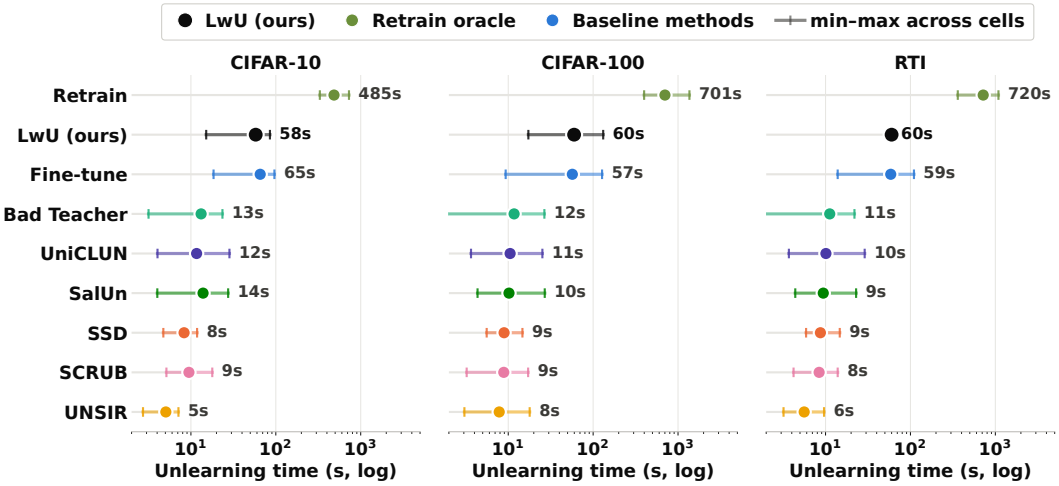


Figure 12: **Every approximate method is one to two orders of magnitude cheaper than retraining; LwU buys its privacy advantage at roughly the cost of fine-tuning.**

Unlearning wall-clock time per method, pooled over all cells and seeds (2 backbones $\times$ 3 forget modes $\times$ 3 seeds). Dots are means and the coloured span is the min–max range; the axis is logarithmic, so position rather than stem length encodes the value. Retraining from scratch costs 701 s on CIFAR-100 and 485 s on CIFAR-10, against 60 s and 58 s for LwU ($11.7\times$ and $8.4\times$ cheaper) and 7.8 s and 5.1 s for the cheapest baseline, UNSIR ($90\times$ and $96\times$). *These timings are indicative only:* the campaign ran three cells concurrently on one GPU, so every measurement includes contention from unrelated jobs, giving relative standard deviations of 21–72% and the wide ranges shown. They establish the order of magnitude, not a precise ranking among the sub-15 s methods.

G.4 Detailed Machine-Unlearning Results

Table 19: CNN with class deletion on CIFAR-10, CIFAR-100, and RTI.

Method	Gap↓	$D_f \to D_f^{\text{Retrain}}$	D_r ↑	Test↑	MIA→ 50	RL-1→Retrain	RL-5→Retrain	Time (s)↓
CIFAR-10								
Retrain (oracle)	0.00	$0.00_{\pm 0.00}$	$92.76_{\pm 0.23}$	$83.49_{\pm 0.21}$	$50.29_{\pm 1.29}$	0.00	57.70	494.90
Original model	27.47	$93.47_{\pm 1.33}$	$92.24_{\pm 0.24}$	$92.37_{\pm 0.34}$	$57.29_{\pm 0.44}$	50.90	95.10	0.00
Fine-tune on D_r	15.34	$50.17_{\pm 43.28}$	$92.20_{\pm 0.89}$	$87.99_{\pm 5.13}$	$55.62_{\pm 2.39}$	20.70	94.60	57.10
Unlearning methods								
Amnesiac	39.15	$0.00_{\pm 0.00}$	$11.11_{\pm 0.00}$	$10.00_{\pm 0.00}$	$49.60_{\pm 0.93}$	100.00	100.00	0.00
Bad Teacher	1.55	$0.27_{\pm 0.15}$	$92.54_{\pm 0.07}$	$83.32_{\pm 0.05}$	$55.82_{\pm 0.90}$	20.40	98.50	8.00
SCRUB	1.12	$0.00_{\pm 0.00}$	$92.73_{\pm 0.49}$	$83.46_{\pm 0.44}$	$54.43_{\pm 2.67}$	0.00	30.30	6.10
SSD	22.39	$74.60_{\pm 22.39}$	$88.36_{\pm 6.07}$	$86.98_{\pm 5.58}$	$55.76_{\pm 1.31}$	45.70	96.40	6.40
SalUn	12.85	$41.50_{\pm 35.98}$	$91.66_{\pm 0.43}$	$86.64_{\pm 3.90}$	$54.41_{\pm 4.36}$	60.10	97.60	13.20
UNSIR	1.05	$0.00_{\pm 0.00}$	$91.86_{\pm 0.43}$	$82.67_{\pm 0.39}$	$52.57_{\pm 2.14}$	0.50	98.10	4.90
Joint continual-learning and machine-unlearning methods								
UniCLUN	12.64	$41.33_{\pm 35.77}$	$91.54_{\pm 0.63}$	$86.52_{\pm 4.14}$	$54.25_{\pm 3.23}$	58.00	96.90	6.00
LwU (ours)	0.92	$0.00_{\pm 0.00}$	$92.77_{\pm 0.25}$	$83.50_{\pm 0.23}$	$53.64_{\pm 2.23}$	0.80	94.30	50.30
CIFAR-10 podium:	🥇 **LwU (ours)** (3G, 1S, 2B; **13 pts**)		🥈 **SCRUB** (3G, 1S, 1B; **12 pts**)		🥉 **UNSIR** (1G, 4S, 0B; **11 pts**)			
CIFAR-100								
Retrain (oracle)	0.00	$0.00_{\pm 0.00}$	$68.51_{\pm 0.39}$	$67.83_{\pm 0.38}$	$51.12_{\pm 4.45}$	0.00	0.00	477.70
Original model	21.66	$70.00_{\pm 3.61}$	$68.58_{\pm 0.52}$	$68.60_{\pm 0.48}$	$66.68_{\pm 1.36}$	36.00	43.30	0.00
Fine-tune on D_r	12.32	$35.00_{\pm 3.61}$	$68.32_{\pm 0.46}$	$67.99_{\pm 0.45}$	$64.69_{\pm 3.46}$	11.00	16.30	66.80
Unlearning methods								
Amnesiac	34.73	$0.00_{\pm 0.00}$	$1.38_{\pm 0.33}$	$1.37_{\pm 0.33}$	$48.41_{\pm 2.75}$	0.00	0.00	0.00
Bad Teacher	5.56	$0.33_{\pm 0.58}$	$67.24_{\pm 0.31}$	$66.57_{\pm 0.31}$	$70.51_{\pm 1.66}$	0.30	10.70	10.20
SCRUB	20.07	$64.33_{\pm 9.02}$	$68.51_{\pm 0.28}$	$68.47_{\pm 0.23}$	$66.06_{\pm 1.28}$	27.30	32.70	4.90
SSD	4.35	$8.00_{\pm 2.65}$	$67.94_{\pm 0.85}$	$67.34_{\pm 0.85}$	$58.56_{\pm 0.37}$	6.00	28.00	5.10
SalUn	11.79	$23.67_{\pm 1.53}$	$63.30_{\pm 1.03}$	$62.91_{\pm 1.03}$	$64.50_{\pm 0.31}$	18.70	39.30	6.70
UNSIR	13.25	$37.33_{\pm 17.67}$	$67.60_{\pm 0.57}$	$67.29_{\pm 0.71}$	$65.32_{\pm 4.62}$	18.70	35.00	4.00
Joint continual-learning and machine-unlearning methods								
UniCLUN	12.80	$26.67_{\pm 3.79}$	$63.07_{\pm 0.90}$	$62.71_{\pm 0.90}$	$65.11_{\pm 2.31}$	17.70	41.70	4.70
LwU (ours)	0.97	$0.00_{\pm 0.00}$	$68.61_{\pm 0.35}$	$67.92_{\pm 0.34}$	$49.35_{\pm 3.23}$	0.00	0.30	43.60
CIFAR-100 podium:	🥇 **LwU (ours)** (5G, 2S, 0B; **19 pts**)		🥈 **Amnesiac** (4G, 1S, 0B; **14 pts**)		🥉 **SSD** (0G, 1S, 5B; **7 pts**)			
RTI								
Retrain (oracle)	0.00	$0.00_{\pm 0.00}$	$80.77_{\pm 0.13}$	$76.73_{\pm 0.12}$	$49.66_{\pm 0.43}$	0.00	0.00	459.60
Original model	23.75	$72.87_{\pm 1.68}$	$79.73_{\pm 0.48}$	$79.39_{\pm 0.38}$	$68.10_{\pm 0.65}$	35.20	80.50	0.00
Fine-tune on D_r	11.66	$28.40_{\pm 3.12}$	$80.55_{\pm 0.37}$	$77.94_{\pm 0.23}$	$66.41_{\pm 2.06}$	10.50	79.40	57.80
Unlearning methods								
Amnesiac	37.21	$0.00_{\pm 0.00}$	$5.26_{\pm 0.00}$	$5.00_{\pm 0.00}$	$51.25_{\pm 1.14}$	0.00	99.60	0.00
Bad Teacher	4.36	$0.13_{\pm 0.12}$	$79.56_{\pm 0.51}$	$75.59_{\pm 0.48}$	$64.62_{\pm 1.41}$	6.30	90.50	9.60
SCRUB	18.30	$52.53_{\pm 9.49}$	$80.38_{\pm 0.54}$	$78.99_{\pm 0.50}$	$67.66_{\pm 0.68}$	16.50	78.50	4.90
SSD	14.21	$37.47_{\pm 30.91}$	$80.38_{\pm 0.67}$	$78.23_{\pm 1.31}$	$66.95_{\pm 1.84}$	17.80	79.20	6.40
SalUn	13.58	$33.60_{\pm 2.23}$	$77.99_{\pm 0.64}$	$75.77_{\pm 0.73}$	$66.65_{\pm 0.88}$	32.10	81.90	6.20
UNSIR	6.58	$11.80_{\pm 17.25}$	$79.34_{\pm 1.14}$	$75.97_{\pm 1.93}$	$61.07_{\pm 1.92}$	7.90	81.70	4.20
Joint continual-learning and machine-unlearning methods								
UniCLUN	14.13	$35.13_{\pm 1.27}$	$78.02_{\pm 0.28}$	$75.88_{\pm 0.30}$	$67.46_{\pm 1.10}$	33.50	82.00	5.40
LwU (ours)	3.48	$0.07_{\pm 0.12}$	$80.60_{\pm 0.42}$	$76.58_{\pm 0.40}$	$62.90_{\pm 3.24}$	0.70	78.20	41.20
RTI podium:	🥇 **LwU (ours)** (3G, 2S, 2B; **15 pts**)		🥈 **Amnesiac** (4G, 0S, 0B; **12 pts**)		🥉 **SCRUB** (1G, 2S, 1B; **8 pts**)			
Overall podium:	🥇 **LwU (ours)** (11G, 5S, 4B; **47 pts**)		🥈 **Amnesiac** (11G, 1S, 0B; **35 pts**)		🥉 **SCRUB** (5G, 4S, 2B; **25 pts**)			

D_f, D_r, Test, and MIA are mean ± standard deviation over 3 seeds. Gap is the reported mean absolute deviation from the Retrain targets. Relearning metrics are ranked by distance to the corresponding Retrain trajectory. Podiums use the weighted medal score 3G+2S+1B.

Table 20: CNN with subclass deletion on CIFAR-10, CIFAR-100, and RTI.

Method	Gap↓	$D_f \to D_f^{\text{Retrain}}$	$D_r \uparrow$	Test↑	MIA→ 50	RL-1→Retrain	RL-5→Retrain	Time (s)↓
CIFAR-10								
Retrain (oracle)	0.00	$25.40_{\pm 4.50}$	$94.89_{\pm 0.11}$	$87.94_{\pm 0.53}$	$49.42_{\pm 0.78}$	36.30	97.90	378.60
Original model	20.60	$94.57_{\pm 1.17}$	$94.51_{\pm 0.14}$	$94.52_{\pm 0.23}$	$55.68_{\pm 1.57}$	53.20	97.10	0.00
Fine-tune on D_r	12.80	$68.90_{\pm 2.80}$	$93.51_{\pm 0.94}$	$91.05_{\pm 1.10}$	$52.15_{\pm 2.37}$	37.60	95.80	44.60
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	2.31	$30.00_{\pm 4.71}$	$94.27_{\pm 0.22}$	$87.84_{\pm 0.67}$	$53.30_{\pm 0.80}$	40.50	97.10	10.80
SCRUB	19.45	$90.73_{\pm 1.63}$	$94.63_{\pm 0.09}$	$94.24_{\pm 0.24}$	$55.33_{\pm 1.43}$	38.20	89.60	6.60
SSD	19.29	$65.57_{\pm 28.00}$	$78.58_{\pm 25.89}$	$77.28_{\pm 25.73}$	$53.56_{\pm 2.15}$	48.60	97.40	7.80
SalUn	14.79	$73.40_{\pm 3.05}$	$93.70_{\pm 0.48}$	$91.67_{\pm 0.70}$	$55.66_{\pm 1.17}$	51.00	97.70	5.90
UNSIR	15.86	$79.13_{\pm 10.71}$	$93.37_{\pm 0.29}$	$91.95_{\pm 1.33}$	$53.59_{\pm 1.42}$	58.50	98.00	4.00
Joint continual-learning and machine-unlearning methods								
UniCLUN	15.07	$74.40_{\pm 4.29}$	$93.65_{\pm 0.59}$	$91.73_{\pm 0.92}$	$55.70_{\pm 0.95}$	50.60	98.30	6.80
LwU (ours)	17.83	$85.63_{\pm 15.15}$	$94.15_{\pm 0.68}$	$93.29_{\pm 2.11}$	$54.41_{\pm 3.18}$	53.00	97.30	40.60
CIFAR-10 podium: 🥇 **Bad Teacher** (3G, 2S, 0B; **13 pts**) 🥈 **SCRUB** (3G, 0S, 1B; **10 pts**) 🥉 **UNSIR** (2G, 0S, 2B; **8 pts**)								
CIFAR-100								
Retrain (oracle)	0.00	$23.33_{\pm 4.16}$	$79.47_{\pm 0.15}$	$78.91_{\pm 0.17}$	$49.96_{\pm 1.85}$	21.70	41.30	664.10
Original model	18.19	$77.00_{\pm 3.46}$	$79.37_{\pm 0.84}$	$79.35_{\pm 0.86}$	$67.54_{\pm 0.62}$	43.30	32.70	0.00
Fine-tune on D_r	16.62	$72.00_{\pm 7.00}$	$79.08_{\pm 0.51}$	$79.01_{\pm 0.57}$	$66.74_{\pm 0.50}$	39.30	31.70	36.00
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	6.84	$17.00_{\pm 4.58}$	$77.89_{\pm 0.51}$	$77.28_{\pm 0.49}$	$67.75_{\pm 1.44}$	15.00	28.70	13.40
SCRUB	17.21	$73.33_{\pm 6.11}$	$79.16_{\pm 0.89}$	$79.10_{\pm 0.95}$	$67.21_{\pm 2.14}$	39.30	30.30	9.80
SSD	12.20	$51.00_{\pm 30.81}$	$78.33_{\pm 0.64}$	$78.05_{\pm 0.86}$	$66.42_{\pm 1.93}$	31.30	31.30	6.80
SalUn	9.59	$37.67_{\pm 5.69}$	$75.57_{\pm 0.84}$	$75.19_{\pm 0.80}$	$66.36_{\pm 1.32}$	28.00	37.00	5.10
UNSIR	14.89	$64.33_{\pm 0.58}$	$78.28_{\pm 0.42}$	$78.14_{\pm 0.42}$	$66.58_{\pm 2.66}$	40.70	38.30	3.60
Joint continual-learning and machine-unlearning methods								
UniCLUN	9.71	$35.67_{\pm 3.51}$	$75.13_{\pm 0.21}$	$74.74_{\pm 0.17}$	$67.96_{\pm 1.68}$	25.00	31.70	6.20
LwU (ours)	10.25	$50.00_{\pm 13.08}$	$78.47_{\pm 0.43}$	$78.19_{\pm 0.43}$	$62.56_{\pm 0.85}$	32.70	42.00	28.30
CIFAR-100 podium: 🥇 **LwU (ours)** (2G, 2S, 0B; **10 pts**) 🥈 **SalUn** (0G, 4S, 2B; **10 pts**) 🥉 **Bad Teacher** (2G, 0S, 1B; **7 pts**)								
RTI								
Retrain (oracle)	0.00	$86.33_{\pm 2.31}$	$78.94_{\pm 0.18}$	$79.01_{\pm 0.16}$	$48.26_{\pm 0.93}$	55.30	47.70	520.90
Original model	3.33	$96.33_{\pm 2.08}$	$79.25_{\pm 0.07}$	$79.42_{\pm 0.08}$	$50.78_{\pm 3.28}$	69.00	48.00	0.00
Fine-tune on D_r	2.62	$95.33_{\pm 1.53}$	$79.12_{\pm 0.15}$	$79.28_{\pm 0.15}$	$48.63_{\pm 1.77}$	65.70	47.70	39.60
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	12.53	$45.00_{\pm 16.52}$	$77.59_{\pm 0.47}$	$77.26_{\pm 0.30}$	$53.97_{\pm 2.24}$	34.30	51.70	6.50
SCRUB	3.19	$95.67_{\pm 1.53}$	$79.28_{\pm 0.09}$	$79.44_{\pm 0.11}$	$50.41_{\pm 3.76}$	67.00	45.00	6.80
SSD	2.37	$90.33_{\pm 4.04}$	$79.14_{\pm 0.10}$	$79.25_{\pm 0.14}$	$51.29_{\pm 0.69}$	61.70	42.00	7.10
SalUn	9.37	$64.67_{\pm 3.06}$	$73.33_{\pm 0.28}$	$73.24_{\pm 0.30}$	$52.67_{\pm 1.49}$	42.00	42.00	8.20
UNSIR	3.52	$95.67_{\pm 2.31}$	$77.96_{\pm 0.89}$	$78.14_{\pm 0.88}$	$51.03_{\pm 1.84}$	69.30	54.30	3.90
Joint continual-learning and machine-unlearning methods								
UniCLUN	6.40	$74.33_{\pm 2.52}$	$72.85_{\pm 0.44}$	$72.86_{\pm 0.42}$	$49.62_{\pm 2.01}$	43.00	48.30	13.60
LwU (ours)	2.17	$88.00_{\pm 5.20}$	$78.52_{\pm 0.34}$	$78.61_{\pm 0.35}$	$51.77_{\pm 2.23}$	60.00	51.00	33.40
RTI podium: 🥇 **SCRUB** (2G, 2S, 4B; **14 pts**) 🥈 **LwU (ours)** (3G, 0S, 3B; **12 pts**) 🥉 **SSD** (0G, 5S, 0B; **10 pts**)								
Overall podium: 🥇 **SCRUB** (7G, 2S, 5B; **30 pts**) 🥈 **LwU (ours)** (5G, 3S, 4B; **25 pts**) 🥉 **Bad Teacher** (5G, 3S, 1B; **22 pts**)								

Table 21: CNN with instance deletion on CIFAR-10, CIFAR-100, and RTI.

Method	Gap↓	$D_f \to D_f^{\text{Retrain}}$	$D_r \uparrow$	Test↑	MIA→ 50	RL-1→Retrain	RL-5→Retrain	Time (s)↓
CIFAR-10								
Retrain (oracle)	0.00	$91.86_{\pm 1.34}$	$92.13_{\pm 0.28}$	$92.13_{\pm 0.28}$	$49.79_{\pm 0.66}$	92.70	94.60	423.80
Original model	4.06	$99.94_{\pm 0.10}$	$92.37_{\pm 0.34}$	$92.37_{\pm 0.34}$	$56.93_{\pm 0.20}$	99.90	100.00	0.00
Fine-tune on D_r	2.85	$96.19_{\pm 0.44}$	$90.12_{\pm 0.33}$	$90.12_{\pm 0.33}$	$52.83_{\pm 0.20}$	97.60	98.60	48.00
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	3.86	$98.29_{\pm 0.84}$	$89.94_{\pm 0.50}$	$89.94_{\pm 0.50}$	$54.41_{\pm 0.60}$	99.40	99.90	8.40
SCRUB	3.39	$97.52_{\pm 0.81}$	$90.44_{\pm 0.51}$	$90.44_{\pm 0.51}$	$54.32_{\pm 0.72}$	98.10	99.20	10.00
SSD	4.16	$99.94_{\pm 0.10}$	$92.40_{\pm 0.39}$	$92.40_{\pm 0.39}$	$57.26_{\pm 0.50}$	99.90	100.00	6.50
SalUn	3.79	$98.02_{\pm 0.14}$	$89.89_{\pm 0.19}$	$89.89_{\pm 0.19}$	$54.31_{\pm 0.73}$	99.30	99.90	17.60
UNSIR	3.82	$98.61_{\pm 0.31}$	$90.52_{\pm 0.01}$	$90.52_{\pm 0.01}$	$55.09_{\pm 0.27}$	99.40	99.90	4.60
Joint continual-learning and machine-unlearning methods								
UniCLUN	3.91	$99.12_{\pm 0.23}$	$90.50_{\pm 0.45}$	$90.50_{\pm 0.45}$	$54.91_{\pm 1.72}$	99.40	99.90	8.60
LwU (ours)	3.68	$98.26_{\pm 1.14}$	$90.66_{\pm 0.86}$	$90.66_{\pm 0.86}$	$55.15_{\pm 0.93}$	99.30	99.70	61.60
CIFAR-10 podium: 🥇 **SCRUB** (4G, 1S, 0B; **14 pts**) 🥈 **LwU (ours)** (0G, 5S, 1B; **11 pts**) 🥉 **SalUn** (1G, 2S, 2B; **9 pts**)								
CIFAR-100								
Retrain (oracle)	0.00	$69.50_{\pm 0.60}$	$68.17_{\pm 0.27}$	$68.17_{\pm 0.27}$	$49.71_{\pm 0.79}$	70.90	76.10	490.00
Original model	12.24	$99.14_{\pm 0.29}$	$68.60_{\pm 0.48}$	$68.60_{\pm 0.48}$	$68.17_{\pm 0.40}$	99.30	99.50	0.00
Fine-tune on D_r	11.65	$98.14_{\pm 0.71}$	$68.42_{\pm 0.39}$	$68.42_{\pm 0.39}$	$67.14_{\pm 0.63}$	98.60	99.20	51.10
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	11.83	$96.72_{\pm 1.21}$	$66.45_{\pm 0.80}$	$66.45_{\pm 0.80}$	$66.38_{\pm 1.35}$	98.70	99.60	5.60
SCRUB	9.62	$90.86_{\pm 1.84}$	$65.88_{\pm 0.42}$	$65.88_{\pm 0.42}$	$62.22_{\pm 0.84}$	92.90	96.70	8.40
SSD	11.96	$97.06_{\pm 1.38}$	$66.96_{\pm 0.90}$	$66.96_{\pm 0.90}$	$67.57_{\pm 0.77}$	98.90	99.40	6.60
SalUn	9.62	$74.81_{\pm 2.72}$	$55.93_{\pm 0.99}$	$55.93_{\pm 0.99}$	$58.39_{\pm 1.15}$	89.10	98.20	14.70
UNSIR	11.63	$96.03_{\pm 1.07}$	$66.03_{\pm 0.06}$	$66.03_{\pm 0.06}$	$65.43_{\pm 0.13}$	97.90	99.40	10.60
Joint continual-learning and machine-unlearning methods								
UniCLUN	9.08	$74.58_{\pm 1.04}$	$56.55_{\pm 0.96}$	$56.55_{\pm 0.96}$	$57.72_{\pm 0.28}$	88.50	97.90	14.90
LwU (ours)	11.38	$95.56_{\pm 1.34}$	$66.30_{\pm 0.30}$	$66.30_{\pm 0.30}$	$65.43_{\pm 0.37}$	97.60	99.00	46.30
CIFAR-100 podium: 🥇 **UniCLUN** (4G, 1S, 0B; **14 pts**) 🥈 **SCRUB** (1G, 1S, 4B; **9 pts**) 🥉 **SalUn** (0G, 4S, 1B; **9 pts**)								
RTI								
Retrain (oracle)	0.00	$78.57_{\pm 0.35}$	$78.85_{\pm 0.33}$	$78.85_{\pm 0.33}$	$50.04_{\pm 1.14}$	80.10	84.40	635.40
Original model	8.86	$99.53_{\pm 0.18}$	$79.39_{\pm 0.38}$	$79.39_{\pm 0.38}$	$63.45_{\pm 0.53}$	99.60	99.90	0.00
Fine-tune on D_r	8.48	$99.36_{\pm 0.05}$	$79.01_{\pm 0.25}$	$79.01_{\pm 0.25}$	$62.75_{\pm 0.40}$	99.20	99.70	65.50
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	8.68	$97.81_{\pm 0.40}$	$77.03_{\pm 0.69}$	$77.03_{\pm 0.69}$	$61.87_{\pm 0.26}$	99.30	99.90	7.80
SCRUB	8.10	$96.29_{\pm 2.78}$	$77.35_{\pm 2.01}$	$77.35_{\pm 2.01}$	$60.91_{\pm 2.03}$	97.00	98.20	7.40
SSD	8.75	$99.12_{\pm 0.58}$	$78.71_{\pm 1.01}$	$78.71_{\pm 1.01}$	$63.32_{\pm 0.35}$	99.60	99.90	7.20
SalUn	7.91	$89.74_{\pm 1.90}$	$72.35_{\pm 0.36}$	$72.35_{\pm 0.36}$	$57.53_{\pm 0.60}$	95.30	98.90	18.60
UNSIR	8.39	$97.37_{\pm 0.23}$	$77.16_{\pm 0.58}$	$77.16_{\pm 0.58}$	$61.42_{\pm 1.55}$	98.60	99.60	6.20
Joint continual-learning and machine-unlearning methods								
UniCLUN	8.13	$90.58_{\pm 1.55}$	$72.58_{\pm 1.04}$	$72.58_{\pm 1.04}$	$58.01_{\pm 0.31}$	95.80	98.90	14.70
LwU (ours)	8.58	$97.43_{\pm 1.30}$	$76.99_{\pm 0.58}$	$76.99_{\pm 0.58}$	$61.78_{\pm 0.40}$	98.90	99.60	53.20
RTI podium: 🥇 **SalUn** (4G, 1S, 0B; **14 pts**) 🥈 **SCRUB** (1G, 3S, 4B; **13 pts**) 🥉 **UniCLUN** (0G, 4S, 1B; **9 pts**)								
Overall podium: 🥇 **SCRUB** (6G, 5S, 8B; **36 pts**) 🥈 **SalUn** (5G, 7S, 3B; **32 pts**) 🥉 **UniCLUN** (4G, 5S, 3B; **25 pts**)								

Table 22: ResNet-18 with class deletion on CIFAR-10, CIFAR-100, and RTI.

Method	Gap↓	$D_f \to D_f^{\text{Retrain}}$	$D_r \uparrow$	Test↑	MIA→ 50	RL-1→Retrain	RL-5→Retrain	Time (s)↓
CIFAR-10								
Retrain (oracle)	0.00	$0.00_{\pm 0.00}$	$93.24_{\pm 0.40}$	$83.92_{\pm 0.36}$	$50.23_{\pm 1.59}$	0.00	0.10	621.20
Original model	27.60	$93.90_{\pm 1.15}$	$92.74_{\pm 0.36}$	$92.86_{\pm 0.23}$	$57.07_{\pm 0.53}$	66.20	94.70	0.00
Fine-tune on D_r	8.74	$26.23_{\pm 45.44}$	$92.13_{\pm 0.53}$	$85.54_{\pm 4.99}$	$54.75_{\pm 1.91}$	17.70	91.10	79.80
Unlearning methods								
Amnesiac	39.38	$0.00_{\pm 0.00}$	$11.11_{\pm 0.00}$	$10.00_{\pm 0.00}$	$49.09_{\pm 0.38}$	100.00	100.00	0.00
Bad Teacher	1.68	$0.13_{\pm 0.23}$	$92.89_{\pm 0.14}$	$83.61_{\pm 0.12}$	$55.89_{\pm 0.65}$	56.40	97.80	21.20
SCRUB	1.56	$0.00_{\pm 0.00}$	$93.24_{\pm 0.31}$	$83.91_{\pm 0.28}$	$55.50_{\pm 1.50}$	0.00	62.00	10.40
SSD	27.24	$92.67_{\pm 3.43}$	$92.71_{\pm 0.38}$	$92.71_{\pm 0.03}$	$57.00_{\pm 0.40}$	69.20	95.50	9.90
SalUn	7.25	$22.07_{\pm 38.22}$	$91.65_{\pm 0.34}$	$84.69_{\pm 4.12}$	$51.39_{\pm 5.16}$	87.00	98.50	18.60
UNSIR	13.04	$43.23_{\pm 10.05}$	$92.04_{\pm 0.14}$	$87.16_{\pm 0.93}$	$54.69_{\pm 0.29}$	46.80	96.40	5.00
Joint continual-learning and machine-unlearning methods								
UniCLUN	13.06	$41.93_{\pm 36.42}$	$92.06_{\pm 0.19}$	$87.05_{\pm 3.80}$	$55.65_{\pm 2.39}$	72.50	97.60	18.40
LwU (ours)	2.32	$3.70_{\pm 6.41}$	$93.31_{\pm 0.24}$	$84.35_{\pm 0.54}$	$54.66_{\pm 2.20}$	6.50	94.00	89.20
CIFAR-10 podium: 🥇 **SCRUB** (4G, 1S, 0B; **14 pts**) 🥈 **LwU (ours)** (1G, 2S, 3B; **10 pts**) 🥉 **Amnesiac** (3G, 0S, 0B; **9 pts**)								
CIFAR-100								
Retrain (oracle)	0.00	$0.00_{\pm 0.00}$	$70.74_{\pm 0.33}$	$70.03_{\pm 0.33}$	$48.86_{\pm 3.81}$	0.00	0.00	1086.40
Original model	23.88	$72.33_{\pm 3.21}$	$70.58_{\pm 0.28}$	$70.60_{\pm 0.27}$	$71.04_{\pm 0.59}$	31.70	9.70	0.00
Fine-tune on D_r	14.75	$36.33_{\pm 4.51}$	$70.38_{\pm 0.22}$	$70.04_{\pm 0.18}$	$71.00_{\pm 0.94}$	6.70	1.00	87.00
Unlearning methods								
Amnesiac	35.33	$0.00_{\pm 0.00}$	$1.27_{\pm 0.23}$	$1.26_{\pm 0.23}$	$50.00_{\pm 0.00}$	0.00	0.00	0.00
Bad Teacher	7.19	$0.00_{\pm 0.00}$	$69.33_{\pm 0.18}$	$68.64_{\pm 0.18}$	$74.83_{\pm 2.48}$	0.00	0.00	16.30
SCRUB	12.76	$31.67_{\pm 9.87}$	$70.65_{\pm 0.25}$	$70.26_{\pm 0.29}$	$67.29_{\pm 2.35}$	4.30	0.00	7.30
SSD	13.11	$35.00_{\pm 18.03}$	$70.50_{\pm 0.12}$	$70.14_{\pm 0.15}$	$59.01_{\pm 14.19}$	23.30	24.00	9.90
SalUn	13.99	$29.00_{\pm 2.65}$	$66.53_{\pm 0.50}$	$66.15_{\pm 0.48}$	$67.74_{\pm 1.53}$	15.00	9.30	6.90
UNSIR	16.29	$44.67_{\pm 10.21}$	$69.55_{\pm 0.42}$	$69.30_{\pm 0.35}$	$67.43_{\pm 3.02}$	16.70	3.70	5.30
Joint continual-learning and machine-unlearning methods								
UniCLUN	14.57	$30.67_{\pm 1.53}$	$66.66_{\pm 0.58}$	$66.30_{\pm 0.56}$	$68.69_{\pm 2.43}$	14.30	9.00	7.30
LwU (ours)	1.43	$0.00_{\pm 0.00}$	$70.58_{\pm 0.34}$	$69.87_{\pm 0.33}$	$50.54_{\pm 3.39}$	0.00	0.00	125.10
CIFAR-100 podium: 🥇 **LwU (ours)** (4G, 2S, 1B; **17 pts**) 🥈 **Amnesiac** (5G, 0S, 0B; **15 pts**) 🥉 **SCRUB** (3G, 1S, 1B; **12 pts**)								
RTI								
Retrain (oracle)	0.00	$0.00_{\pm 0.00}$	$82.26_{\pm 0.41}$	$78.15_{\pm 0.39}$	$51.48_{\pm 1.24}$	0.00	0.00	717.40
Original model	23.52	$71.27_{\pm 2.14}$	$81.42_{\pm 0.02}$	$80.92_{\pm 0.09}$	$70.70_{\pm 0.45}$	24.10	56.70	0.00
Fine-tune on D_r	14.14	$36.87_{\pm 5.08}$	$81.93_{\pm 0.40}$	$79.67_{\pm 0.36}$	$69.31_{\pm 0.85}$	8.70	45.90	81.30
Unlearning methods								
Amnesiac	37.93	$0.00_{\pm 0.00}$	$5.26_{\pm 0.00}$	$5.00_{\pm 0.00}$	$49.90_{\pm 0.25}$	0.00	0.00	0.00
Bad Teacher	4.38	$0.00_{\pm 0.00}$	$81.44_{\pm 0.26}$	$77.37_{\pm 0.25}$	$67.41_{\pm 2.01}$	0.20	75.20	19.80
SCRUB	4.64	$0.00_{\pm 0.00}$	$82.04_{\pm 0.03}$	$77.94_{\pm 0.03}$	$69.27_{\pm 1.02}$	0.00	4.00	8.20
SSD	13.38	$36.67_{\pm 32.30}$	$81.00_{\pm 0.90}$	$78.78_{\pm 2.36}$	$65.43_{\pm 5.54}$	19.10	64.90	9.80
SalUn	14.86	$38.13_{\pm 1.10}$	$80.54_{\pm 0.28}$	$78.42_{\pm 0.26}$	$70.66_{\pm 1.16}$	23.30	56.80	7.00
UNSIR	17.59	$51.73_{\pm 18.15}$	$80.95_{\pm 0.11}$	$79.49_{\pm 0.90}$	$67.44_{\pm 3.11}$	18.00	57.90	6.90
Joint continual-learning and machine-unlearning methods								
UniCLUN	15.10	$39.33_{\pm 0.31}$	$80.52_{\pm 0.27}$	$78.46_{\pm 0.25}$	$70.35_{\pm 1.38}$	23.90	56.90	7.40
LwU (ours)	3.59	$0.07_{\pm 0.12}$	$82.11_{\pm 0.02}$	$78.01_{\pm 0.02}$	$65.11_{\pm 5.47}$	0.20	33.90	125.40
RTI podium: 🥇 **Amnesiac** (5G, 0S, 0B; **15 pts**) 🥈 **LwU (ours)** (2G, 3S, 1B; **13 pts**) 🥉 **SCRUB** (2G, 2S, 1B; **11 pts**)								
Overall podium: 🥇 **LwU (ours)** (7G, 7S, 5B; **40 pts**) 🥈 **Amnesiac** (13G, 0S, 0B; **39 pts**) 🥉 **SCRUB** (9G, 4S, 2B; **37 pts**)								

Table 23: ResNet-18 with subclass deletion on CIFAR-10, CIFAR-100, and RTI.

Method	Gap↓	$D_f \to D_f^{\text{Retrain}}$	$D_r \uparrow$	Test↑	MIA→ 50	RL-1→Retrain	RL-5→Retrain	Time (s)↓
CIFAR-10								
Retrain (oracle)	0.00	$26.57_{\pm 1.23}$	$95.16_{\pm 0.07}$	$88.30_{\pm 0.17}$	$49.56_{\pm 1.53}$	73.60	99.50	479.50
Original model	20.42	$95.30_{\pm 0.69}$	$95.16_{\pm 0.06}$	$95.18_{\pm 0.03}$	$55.57_{\pm 0.05}$	86.20	97.70	0.00
Fine-tune on D_r	12.97	$70.07_{\pm 0.68}$	$94.13_{\pm 0.60}$	$91.72_{\pm 0.56}$	$53.49_{\pm 1.08}$	76.00	98.60	76.70
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	2.25	$28.40_{\pm 4.10}$	$94.37_{\pm 0.78}$	$87.77_{\pm 0.87}$	$53.36_{\pm 1.38}$	86.60	99.20	15.70
SCRUB	16.55	$84.57_{\pm 0.06}$	$93.68_{\pm 0.36}$	$92.77_{\pm 0.32}$	$51.80_{\pm 1.27}$	3.30	74.00	12.30
SSD	16.97	$84.33_{\pm 9.85}$	$94.13_{\pm 0.84}$	$93.15_{\pm 1.70}$	$53.80_{\pm 1.57}$	88.30	98.80	8.40
SalUn	12.59	$68.30_{\pm 4.06}$	$93.63_{\pm 0.26}$	$91.10_{\pm 0.64}$	$53.85_{\pm 2.31}$	95.00	99.50	18.30
UNSIR	16.45	$82.30_{\pm 4.86}$	$93.97_{\pm 0.35}$	$92.80_{\pm 0.49}$	$53.92_{\pm 0.26}$	76.30	98.30	4.50
Joint continual-learning and machine-unlearning methods								
UniCLUN	12.62	$68.10_{\pm 3.86}$	$93.87_{\pm 0.19}$	$91.30_{\pm 0.50}$	$53.19_{\pm 3.62}$	94.90	99.40	15.10
LwU (ours)	18.50	$89.77_{\pm 7.82}$	$94.50_{\pm 1.02}$	$94.02_{\pm 1.70}$	$53.99_{\pm 2.04}$	85.20	98.40	41.80
CIFAR-10 podium: 🥇 **Bad Teacher** (2G, 1S, 3B; **11 pts**) 🥈 **LwU (ours)** (2G, 1S, 0B; **8 pts**) 🥉 **UNSIR** (2G, 0S, 1B; **7 pts**)								
CIFAR-100								
Retrain (oracle)	0.00	$24.00_{\pm 2.00}$	$80.04_{\pm 0.49}$	$79.48_{\pm 0.47}$	$49.32_{\pm 2.42}$	20.70	17.00	804.40
Original model	17.85	$80.00_{\pm 0.00}$	$80.69_{\pm 0.36}$	$80.68_{\pm 0.36}$	$62.87_{\pm 1.25}$	44.70	16.00	0.00
Fine-tune on D_r	16.81	$75.33_{\pm 4.93}$	$80.26_{\pm 0.13}$	$80.21_{\pm 0.11}$	$64.03_{\pm 3.41}$	41.30	15.30	45.40
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	7.07	$17.00_{\pm 3.00}$	$79.22_{\pm 0.41}$	$78.60_{\pm 0.43}$	$68.91_{\pm 1.43}$	18.30	18.70	21.90
SCRUB	18.14	$79.67_{\pm 0.58}$	$80.64_{\pm 0.40}$	$80.63_{\pm 0.39}$	$64.46_{\pm 1.93}$	44.30	13.70	14.90
SSD	13.30	$57.67_{\pm 31.21}$	$78.74_{\pm 3.08}$	$78.53_{\pm 3.35}$	$64.00_{\pm 2.67}$	34.00	15.00	12.90
SalUn	10.12	$42.00_{\pm 3.61}$	$76.86_{\pm 0.48}$	$76.51_{\pm 0.51}$	$65.64_{\pm 2.38}$	26.70	16.00	6.30
UNSIR	17.37	$77.33_{\pm 6.11}$	$79.64_{\pm 0.86}$	$79.62_{\pm 0.91}$	$64.26_{\pm 4.58}$	44.00	20.00	7.20
Joint continual-learning and machine-unlearning methods								
UniCLUN	10.52	$44.67_{\pm 4.16}$	$77.01_{\pm 0.60}$	$76.69_{\pm 0.62}$	$64.92_{\pm 3.37}$	25.30	16.30	7.70
LwU (ours)	16.24	$74.33_{\pm 5.51}$	$80.19_{\pm 0.56}$	$80.13_{\pm 0.58}$	$62.66_{\pm 2.55}$	41.00	16.70	44.30
CIFAR-100 podium: 🥇 **LwU (ours)** (2G, 2S, 0B; **10 pts**) 🥈 **Bad Teacher** (3G, 0S, 0B; **9 pts**) 🥉 **SalUn** (1G, 2S, 2B; **9 pts**)								
RTI								
Retrain (oracle)	0.00	$86.67_{\pm 3.21}$	$80.54_{\pm 0.23}$	$80.60_{\pm 0.26}$	$49.87_{\pm 2.96}$	59.70	40.30	1144.70
Original model	3.13	$94.67_{\pm 2.52}$	$80.54_{\pm 0.34}$	$80.68_{\pm 0.36}$	$54.23_{\pm 1.00}$	75.70	42.30	0.00
Fine-tune on D_r	3.03	$95.67_{\pm 1.53}$	$80.14_{\pm 0.28}$	$80.29_{\pm 0.28}$	$51.95_{\pm 1.09}$	74.70	42.00	40.40
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	15.63	$31.67_{\pm 4.16}$	$79.09_{\pm 0.44}$	$78.62_{\pm 0.40}$	$53.56_{\pm 2.73}$	32.70	41.30	21.40
SCRUB	3.09	$95.00_{\pm 2.00}$	$80.50_{\pm 0.41}$	$80.65_{\pm 0.42}$	$53.62_{\pm 1.16}$	72.30	38.00	11.50
SSD	2.76	$94.67_{\pm 2.52}$	$80.54_{\pm 0.34}$	$80.68_{\pm 0.37}$	$52.73_{\pm 1.80}$	75.00	41.70	12.50
SalUn	7.20	$70.67_{\pm 3.06}$	$74.78_{\pm 0.54}$	$74.74_{\pm 0.52}$	$50.64_{\pm 3.66}$	55.70	39.70	6.70
UNSIR	3.28	$95.33_{\pm 1.53}$	$79.49_{\pm 0.33}$	$79.65_{\pm 0.33}$	$52.32_{\pm 1.69}$	73.70	45.30	7.60
Joint continual-learning and machine-unlearning methods								
UniCLUN	7.85	$70.00_{\pm 2.00}$	$74.90_{\pm 0.34}$	$74.85_{\pm 0.34}$	$52.74_{\pm 3.22}$	56.00	39.30	6.40
LwU (ours)	2.21	$90.33_{\pm 1.53}$	$79.85_{\pm 0.15}$	$79.95_{\pm 0.14}$	$53.72_{\pm 1.09}$	68.00	47.00	53.90
RTI podium: 🥇 **SSD** (2G, 2S, 2B; **12 pts**) 🥈 **SalUn** (2G, 2S, 0B; **10 pts**) 🥉 **LwU (ours)** (2G, 0S, 3B; **9 pts**)								
Overall podium: 🥇 **LwU (ours)** (6G, 3S, 3B; **27 pts**) 🥈 **SalUn** (4G, 5S, 3B; **25 pts**) 🥉 **Bad Teacher** (5G, 2S, 3B; **22 pts**)								

Table 24: ResNet-18 with instance deletion on CIFAR-10, CIFAR-100, and RTI.

Method	Gap↓	$D_f \to D_f^{\text{Retrain}}$	$D_r \uparrow$	Test↑	MIA→ 50	RL-1→Retrain	RL-5→Retrain	Time (s)↓
CIFAR-10								
Retrain (oracle)	0.00	$92.06_{\pm 0.75}$	$92.18_{\pm 0.30}$	$92.18_{\pm 0.30}$	$49.76_{\pm 0.52}$	92.90	95.30	511.10
Original model	4.28	$100.00_{\pm 0.00}$	$92.86_{\pm 0.23}$	$92.86_{\pm 0.23}$	$57.60_{\pm 0.64}$	100.00	100.00	0.00
Fine-tune on D_r	3.08	$97.43_{\pm 2.04}$	$91.28_{\pm 1.19}$	$91.28_{\pm 1.19}$	$54.26_{\pm 1.77}$	98.30	99.30	73.20
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	3.89	$98.97_{\pm 0.36}$	$90.53_{\pm 0.64}$	$90.53_{\pm 0.64}$	$55.13_{\pm 1.50}$	99.70	99.90	9.40
SCRUB	3.40	$98.82_{\pm 0.66}$	$91.23_{\pm 0.59}$	$91.23_{\pm 0.59}$	$54.72_{\pm 0.52}$	99.40	99.70	10.70
SSD	4.28	$100.00_{\pm 0.00}$	$92.89_{\pm 0.26}$	$92.89_{\pm 0.26}$	$57.54_{\pm 0.67}$	100.00	100.00	10.40
SalUn	3.99	$96.25_{\pm 2.23}$	$88.76_{\pm 1.68}$	$88.76_{\pm 1.68}$	$54.68_{\pm 1.42}$	98.80	99.90	7.70
UNSIR	3.71	$98.88_{\pm 0.58}$	$90.74_{\pm 0.70}$	$90.74_{\pm 0.70}$	$54.93_{\pm 0.78}$	99.70	99.90	5.30
Joint continual-learning and machine-unlearning methods								
UniCLUN	3.73	$98.38_{\pm 0.80}$	$90.48_{\pm 1.05}$	$90.48_{\pm 1.05}$	$54.99_{\pm 1.38}$	99.50	99.80	11.60
LwU (ours)	4.10	$99.97_{\pm 0.05}$	$92.77_{\pm 0.19}$	$92.77_{\pm 0.19}$	$57.08_{\pm 0.89}$	100.00	100.00	59.60
CIFAR-10 podium: 🥇 **SCRUB** (2G, 2S, 3B; **13 pts**) 🥈 **SalUn** (3G, 1S, 1B; **12 pts**) 🥉 **UNSIR** (1G, 1S, 2B; **7 pts**)								
CIFAR-100								
Retrain (oracle)	0.00	$69.58_{\pm 2.46}$	$69.28_{\pm 0.73}$	$69.28_{\pm 0.73}$	$50.04_{\pm 0.73}$	72.00	80.90	685.10
Original model	13.77	$100.00_{\pm 0.00}$	$70.60_{\pm 0.27}$	$70.60_{\pm 0.27}$	$72.06_{\pm 0.46}$	100.00	100.00	0.00
Fine-tune on D_r	13.26	$99.89_{\pm 0.19}$	$70.23_{\pm 0.46}$	$70.23_{\pm 0.46}$	$70.88_{\pm 0.48}$	99.90	100.00	35.80
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	12.65	$99.31_{\pm 0.77}$	$68.47_{\pm 1.41}$	$68.47_{\pm 1.41}$	$69.29_{\pm 1.31}$	100.00	100.00	2.40
SCRUB	11.81	$97.94_{\pm 0.98}$	$68.40_{\pm 0.76}$	$68.40_{\pm 0.76}$	$67.15_{\pm 0.77}$	98.40	99.60	8.30
SSD	12.83	$98.94_{\pm 0.46}$	$68.96_{\pm 1.06}$	$68.96_{\pm 1.06}$	$70.33_{\pm 0.23}$	99.90	100.00	10.00
SalUn	12.01	$84.53_{\pm 1.61}$	$58.67_{\pm 0.73}$	$58.67_{\pm 0.73}$	$61.94_{\pm 0.73}$	96.30	99.80	22.60
UNSIR	12.79	$98.86_{\pm 0.81}$	$68.03_{\pm 1.06}$	$68.03_{\pm 1.06}$	$69.43_{\pm 0.20}$	99.80	100.00	17.80
Joint continual-learning and machine-unlearning methods								
UniCLUN	12.61	$85.31_{\pm 1.41}$	$58.37_{\pm 2.36}$	$58.37_{\pm 2.36}$	$62.93_{\pm 1.02}$	96.70	99.90	21.70
LwU (ours)	12.25	$97.81_{\pm 0.41}$	$68.09_{\pm 0.50}$	$68.09_{\pm 0.50}$	$68.46_{\pm 1.75}$	99.20	99.90	74.20
CIFAR-100 podium: 🥇 **SalUn** (3G, 2S, 0B; **13 pts**) 🥈 **SCRUB** (2G, 1S, 4B; **12 pts**) 🥉 **SSD** (2G, 0S, 1B; **7 pts**)								
RTI								
Retrain (oracle)	0.00	$80.99_{\pm 1.01}$	$80.05_{\pm 0.50}$	$80.05_{\pm 0.50}$	$49.93_{\pm 0.33}$	82.40	87.80	873.50
Original model	9.15	$99.97_{\pm 0.05}$	$80.92_{\pm 0.09}$	$80.92_{\pm 0.09}$	$65.83_{\pm 1.11}$	99.90	100.00	0.00
Fine-tune on D_r	8.59	$99.82_{\pm 0.15}$	$80.74_{\pm 0.43}$	$80.74_{\pm 0.43}$	$64.11_{\pm 0.88}$	99.90	100.00	66.20
Unlearning methods								
Amnesiac	–	–	–	–	–	–	–	–
Bad Teacher	8.59	$99.01_{\pm 0.83}$	$79.14_{\pm 1.04}$	$79.14_{\pm 1.04}$	$63.52_{\pm 2.54}$	99.90	100.00	3.90
SCRUB	7.82	$98.83_{\pm 0.48}$	$79.40_{\pm 0.17}$	$79.40_{\pm 0.17}$	$62.08_{\pm 1.50}$	99.40	99.90	8.70
SSD	8.97	$92.16_{\pm 11.02}$	$74.50_{\pm 7.98}$	$74.50_{\pm 7.98}$	$60.80_{\pm 4.15}$	99.60	99.80	9.60
SalUn	8.93	$94.68_{\pm 0.28}$	$74.00_{\pm 0.57}$	$74.00_{\pm 0.57}$	$59.84_{\pm 0.49}$	98.40	99.80	23.20
UNSIR	8.62	$98.71_{\pm 0.56}$	$78.28_{\pm 0.83}$	$78.28_{\pm 0.83}$	$63.16_{\pm 0.92}$	99.60	99.90	8.10
Joint continual-learning and machine-unlearning methods								
UniCLUN	8.59	$95.15_{\pm 1.88}$	$74.93_{\pm 1.06}$	$74.93_{\pm 1.06}$	$59.90_{\pm 0.91}$	98.30	99.80	22.80
LwU (ours)	8.48	$99.62_{\pm 0.43}$	$80.17_{\pm 0.46}$	$80.17_{\pm 0.46}$	$64.21_{\pm 1.18}$	99.90	100.00	81.10
RTI podium: 🥇 **SCRUB** (1G, 3S, 2B; **11 pts**) 🥈 **SalUn** (2G, 2S, 0B; **10 pts**) 🥉 **UniCLUN** (2G, 1S, 2B; **10 pts**)								
Overall podium: 🥇 **SCRUB** (5G, 6S, 9B; **36 pts**) 🥈 **SalUn** (8G, 5S, 1B; **35 pts**) 🥉 **UniCLUN** (2G, 6S, 5B; **23 pts**)								

G.5 Qualitative Machine-Unlearning Results on CIFAR-10



Figure 13: CIFAR-10 airplane class (1st-class) removal.

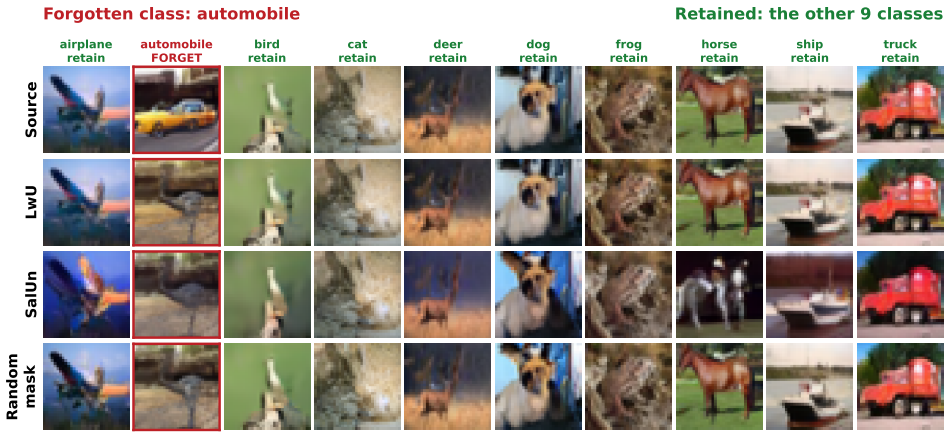


Figure 14: CIFAR-10 automobile class (2nd-class) removal.

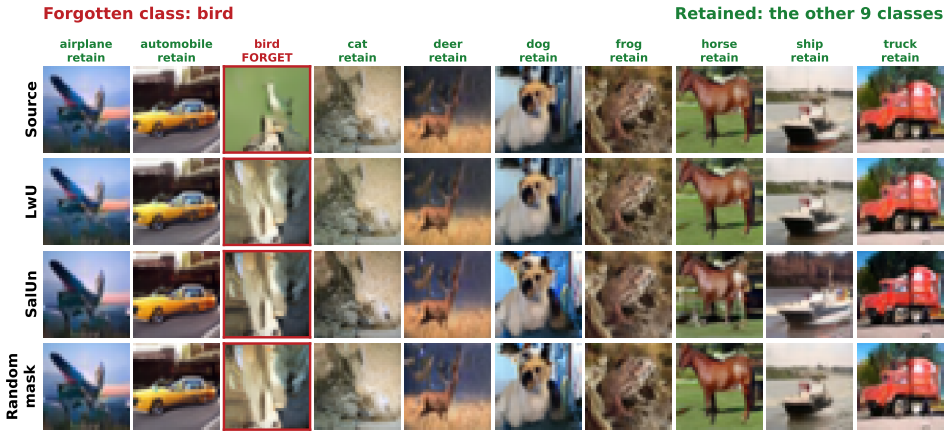


Figure 15: CIFAR-10 bird class (3rd-class) removal.



Figure 16: CIFAR-10 cat class (4th-class) removal.

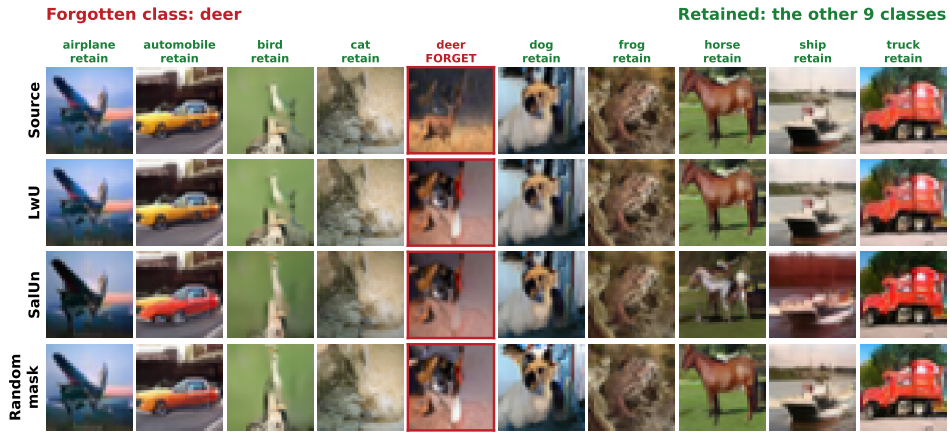


Figure 17: CIFAR-10 deer class (5th-class) removal.

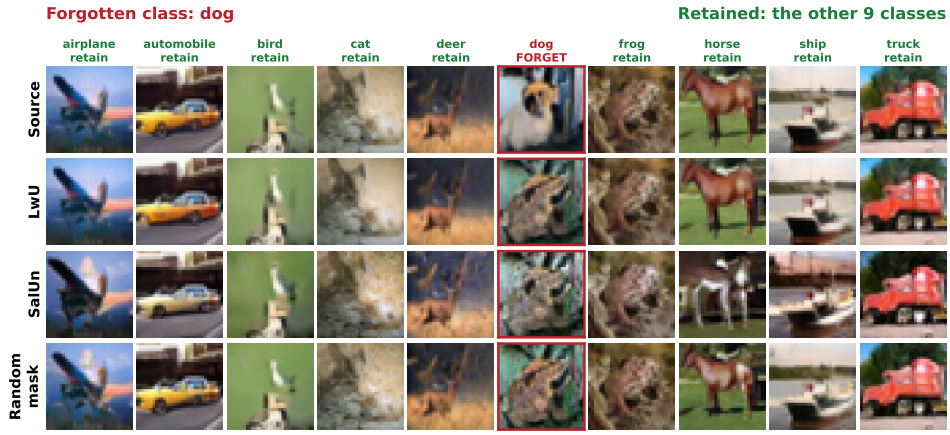


Figure 18: CIFAR-10 dog class (6th-class) removal.

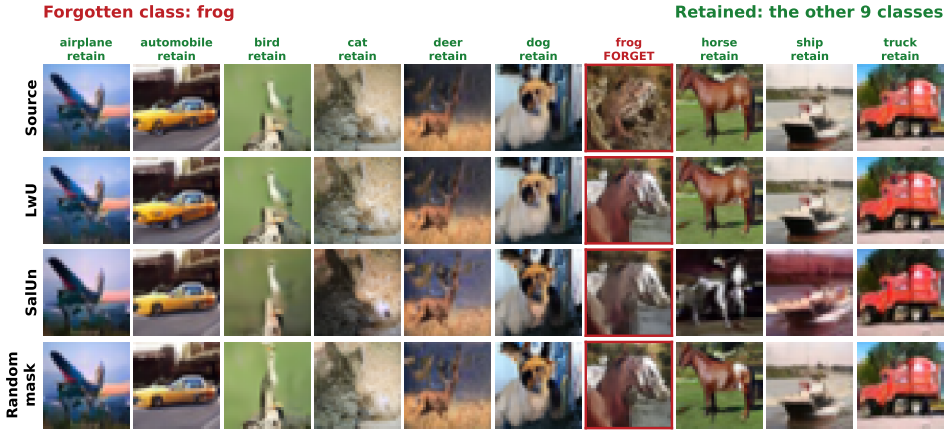


Figure 19: CIFAR-10 frog class (7th-class) removal.

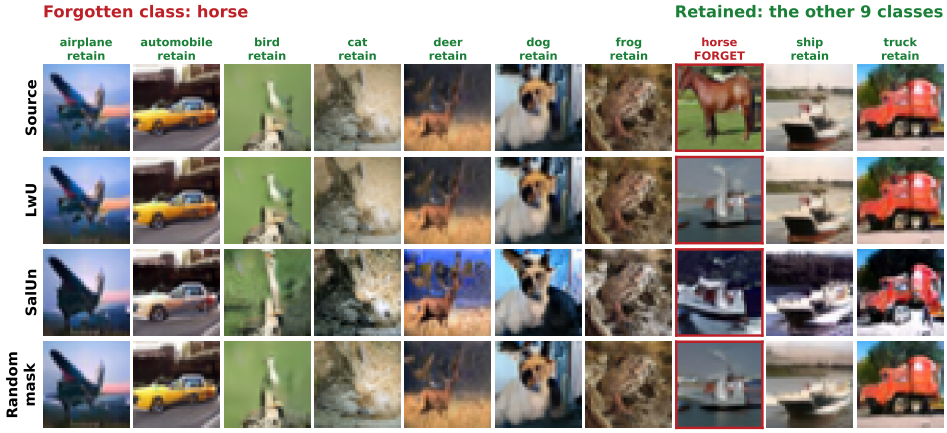


Figure 20: CIFAR-10 horse class (8th-class) removal.

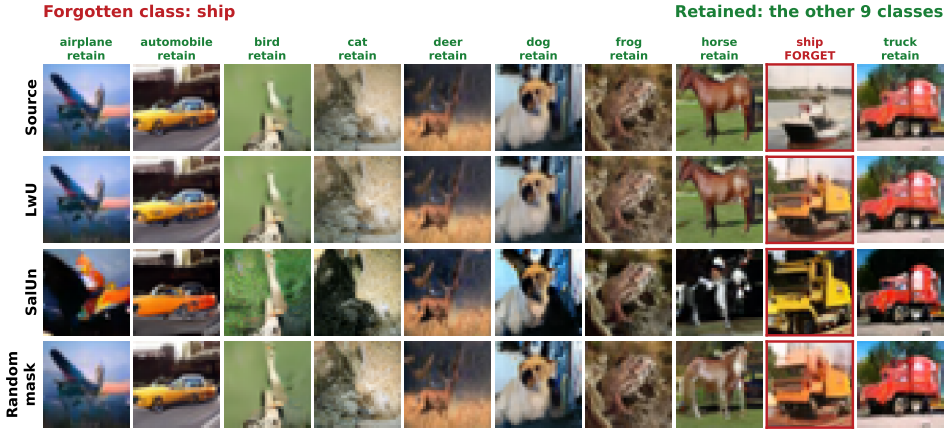


Figure 21: CIFAR-10 ship class (9th-class) removal.

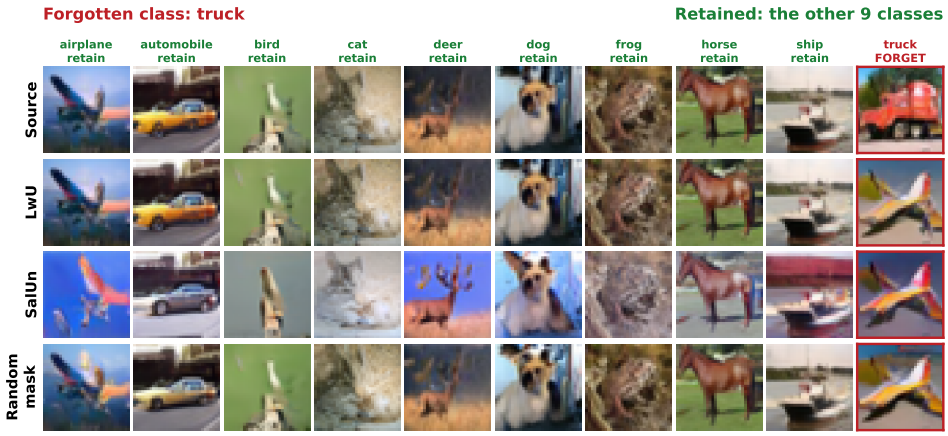


Figure 22: CIFAR-10 truck class (10th-class) removal.

G.6 CONTINUAL-LEARNING COST

Table 25: Training cost comparison.

	TIL				CIL			
Method	Train min↓	GPU h↓	Energy (Wh)↓	Peak device (MB)↓	Train min↓	GPU h↓	Energy (Wh)↓	Peak device (MB)↓
Seq-CIFAR-100								
Joint training (upper bound)	$40.33_{\pm 2.27}$	$0.678_{\pm 0.038}$	$84.53_{\pm 5.13}$	$2001.0_{\pm 2.1}$	$28.37_{\pm 0.68}$	$0.479_{\pm 0.011}$	$67.12_{\pm 3.43}$	$2094.0_{\pm 3.2}$
EWC	$8.73_{\pm 0.75}$	$0.151_{\pm 0.012}$	$17.97_{\pm 1.29}$	$2148.5_{\pm 55.7}$	$7.77_{\pm 1.26}$	$0.136_{\pm 0.021}$	$15.56_{\pm 2.00}$	$2276.4_{\pm 151.9}$
SI	$9.67_{\pm 0.45}$	$0.168_{\pm 0.008}$	$20.97_{\pm 1.03}$	$3297.1_{\pm 773.7}$	$9.11_{\pm 0.44}$	$0.158_{\pm 0.007}$	$20.74_{\pm 1.20}$	$3542.4_{\pm 134.5}$
LwF	$8.95_{\pm 0.68}$	$0.154_{\pm 0.011}$	$37.11_{\pm 2.49}$	$4793.7_{\pm 126.4}$	$4.58_{\pm 0.41}$	$0.082_{\pm 0.007}$	$12.02_{\pm 0.63}$	$2144.3_{\pm 52.1}$
WSN	$4.53_{\pm 0.23}$	$0.082_{\pm 0.004}$	$14.21_{\pm 0.77}$	$2391.1_{\pm 50.9}$	–	–	–	–
PEC	–	–	–	–	$25.07_{\pm 0.16}$	$0.426_{\pm 0.003}$	$96.16_{\pm 1.58}$	$4473.6_{\pm 0.0}$
SpaceNet	$2.89_{\pm 0.03}$	$0.052_{\pm 0.001}$	$3.79_{\pm 0.07}$	$1334.8_{\pm 12.1}$	$4.89_{\pm 0.55}$	$0.085_{\pm 0.009}$	$6.19_{\pm 0.65}$	$1335.5_{\pm 11.6}$
NISPA	$3.63_{\pm 0.74}$	$0.066_{\pm 0.012}$	$7.58_{\pm 1.43}$	$1991.9_{\pm 135.5}$	$4.02_{\pm 0.01}$	$0.072_{\pm 0.000}$	$8.75_{\pm 0.08}$	$1985.6_{\pm 49.8}$
SNV	$97.00_{\pm 23.73}$	$1.624_{\pm 0.395}$	$426.21_{\pm 112.91}$	$2428.9_{\pm 0.0}$	$81.31_{\pm 8.34}$	$1.400_{\pm 0.139}$	$319.94_{\pm 42.47}$	$2428.1_{\pm 1.2}$
UniCLUN	$7.42_{\pm 0.60}$	$0.128_{\pm 0.010}$	$25.77_{\pm 2.35}$	$2514.8_{\pm 0.0}$	$13.58_{\pm 1.79}$	$0.233_{\pm 0.030}$	$39.95_{\pm 5.95}$	$2514.8_{\pm 0.0}$
LwU (ours)	$6.08_{\pm 0.46}$	$0.108_{\pm 0.008}$	$14.11_{\pm 1.07}$	$2554.7_{\pm 29.1}$	$6.41_{\pm 0.61}$	$0.113_{\pm 0.010}$	$15.18_{\pm 0.98}$	$2554.0_{\pm 27.8}$
SGD (lower bound)	$3.06_{\pm 0.19}$	$0.057_{\pm 0.003}$	$8.86_{\pm 0.49}$	$1915.0_{\pm 0.0}$	$4.30_{\pm 0.08}$	$0.077_{\pm 0.001}$	$12.05_{\pm 0.18}$	$1916.4_{\pm 1.2}$
Seq-Tiny-ImageNet								
Joint training (upper bound)	–	–	–	–	–	–	–	–
EWC	$1.09_{\pm 0.06}$	$0.182_{\pm 0.010}$	$48.3_{\pm 2.6}$	$3179_{\pm 0}$	$0.99_{\pm 0.03}$	$0.164_{\pm 0.004}$	$43.1_{\pm 1.1}$	$3179_{\pm 0}$
SI	$1.54_{\pm 0.14}$	$0.257_{\pm 0.024}$	$66.7_{\pm 5.7}$	$3539_{\pm 0}$	$1.12_{\pm 0.09}$	$0.186_{\pm 0.015}$	$47.7_{\pm 3.6}$	$3539_{\pm 1}$
LwF	$5.02_{\pm 0.30}$	$0.836_{\pm 0.050}$	$222.3_{\pm 13.3}$	$13462_{\pm 68}$	$0.95_{\pm 0.05}$	$0.158_{\pm 0.008}$	$43.5_{\pm 2.1}$	$3377_{\pm 0}$
WSN	$4.97_{\pm 0.15}$	$0.83_{\pm 0.03}$	$89.0_{\pm 3.7}$	$3555.7_{\pm 1.2}$	–	–	–	–
PEC	–	–	–	–	$35.81_{\pm 7.45}$	$5.97_{\pm 1.24}$	$819.8_{\pm 48.3}$	$13992.5_{\pm 0.0}$
SpaceNet	$1.14_{\pm 0.12}$	$0.19_{\pm 0.02}$	$13.1_{\pm 1.2}$	$1596.3_{\pm 0.0}$	$18.29_{\pm 1.30}$	$3.05_{\pm 0.22}$	$202.4_{\pm 13.1}$	$1596.3_{\pm 0.0}$
NISPA	$1.41_{\pm 0.34}$	$0.23_{\pm 0.06}$	$34.9_{\pm 8.4}$	$4705.6_{\pm 553.8}$	$2.99_{\pm 0.18}$	$0.50_{\pm 0.03}$	$73.5_{\pm 3.9}$	$5659.8_{\pm 334.7}$
SNV	$46.96_{\pm 7.83}$	$7.83_{\pm 1.31}$	$2051.0_{\pm 369.3}$	$6263.1_{\pm 1.2}$	$45.05_{\pm 8.18}$	$7.51_{\pm 1.36}$	$1888.3_{\pm 406.0}$	$6265.9_{\pm 1.2}$
UniCLUN	$4.40_{\pm 0.32}$	$0.73_{\pm 0.05}$	$157.2_{\pm 11.5}$	$5759.1_{\pm 40.0}$	$4.62_{\pm 0.75}$	$0.77_{\pm 0.12}$	$171.4_{\pm 27.0}$	$5759.1_{\pm 40.0}$
LwU (ours)	$1.40_{\pm 0.05}$	$0.234_{\pm 0.008}$	$63.8_{\pm 2.1}$	$3665_{\pm 1}$	$1.33_{\pm 0.08}$	$0.221_{\pm 0.013}$	$60.5_{\pm 3.3}$	$3665_{\pm 0}$
SGD (lower bound)	$0.74_{\pm 0.03}$	$0.123_{\pm 0.005}$	$33.7_{\pm 1.2}$	$3169_{\pm 0}$	$0.81_{\pm 0.03}$	$0.135_{\pm 0.005}$	$36.4_{\pm 1.3}$	$3169_{\pm 0}$
RTI								
Joint training (upper bound)	$6.36_{\pm 0.36}$	$0.1073_{\pm 0.0060}$	$25.70_{\pm 1.19}$	$2347_{\pm 29}$	$5.22_{\pm 0.15}$	$0.0882_{\pm 0.0025}$	$23.08_{\pm 0.40}$	$2340_{\pm 25}$
EWC	$2.24_{\pm 0.07}$	$0.0387_{\pm 0.0011}$	$9.10_{\pm 0.25}$	$2226_{\pm 16}$	$2.18_{\pm 0.08}$	$0.0377_{\pm 0.0014}$	$8.93_{\pm 0.14}$	$2273_{\pm 46}$
SI	$2.63_{\pm 0.09}$	$0.0453_{\pm 0.0015}$	$12.30_{\pm 0.31}$	$2759_{\pm 19}$	$2.27_{\pm 0.13}$	$0.0394_{\pm 0.0021}$	$10.49_{\pm 0.53}$	$2745_{\pm 17}$
LwF	$4.01_{\pm 0.18}$	$0.0682_{\pm 0.0030}$	$20.08_{\pm 0.90}$	$3333_{\pm 20}$	$1.89_{\pm 0.14}$	$0.0328_{\pm 0.0024}$	$9.26_{\pm 0.67}$	$2165_{\pm 22}$
WSN	$3.07_{\pm 0.19}$	$0.0530_{\pm 0.0031}$	$13.34_{\pm 0.91}$	$2476_{\pm 1}$	–	–	–	–
PEC	–	–	–	–	$12.42_{\pm 0.01}$	$0.2098_{\pm 0.0001}$	$62.82_{\pm 0.05}$	$4643_{\pm 1}$
SpaceNet	$1.77_{\pm 0.21}$	$0.0304_{\pm 0.0034}$	$2.56_{\pm 0.39}$	$1541_{\pm 12}$	$1.49_{\pm 0.08}$	$0.0257_{\pm 0.0014}$	$2.14_{\pm 0.11}$	1534
NISPA	$2.17_{\pm 0.50}$	$0.0373_{\pm 0.0083}$	$5.40_{\pm 1.21}$	$2073_{\pm 40}$	$2.01_{\pm 0.17}$	$0.0346_{\pm 0.0029}$	$5.05_{\pm 0.45}$	$2136_{\pm 42}$
SNV	$33.84_{\pm 8.28}$	$0.5785_{\pm 0.1408}$	$186.12_{\pm 49.31}$	$2619_{\pm 1}$	$28.46_{\pm 2.92}$	$0.4886_{\pm 0.0501}$	$139.45_{\pm 18.52}$	$2621_{\pm 1}$
UniCLUN	$4.69_{\pm 0.38}$	$0.0795_{\pm 0.0063}$	$20.31_{\pm 1.49}$	2874	$6.61_{\pm 0.29}$	$0.1115_{\pm 0.0049}$	$29.81_{\pm 1.28}$	2874
LwU (ours)	$3.07_{\pm 0.54}$	$0.0546_{\pm 0.0090}$	$14.72_{\pm 2.36}$	$2389_{\pm 41}$	$1.57_{\pm 0.15}$	$0.0296_{\pm 0.0025}$	$7.80_{\pm 0.63}$	$2369_{\pm 13}$
SGD (lower bound)	$1.12_{\pm 0.13}$	$0.0198_{\pm 0.0021}$	$5.44_{\pm 0.57}$	2071	$1.28_{\pm 0.03}$	$0.0225_{\pm 0.0005}$	$6.17_{\pm 0.14}$	2071

Table 26: Memory and storage measurements for TIL.

Method	Peak alloc. (MB)↓	Peak reserved (MB)↓	Peak device (MB)↓	Optimizer (MB)↓	Model params (MB)↓	Resident storage (MB)↓	Checkpoint (MB)↓
Seq-CIFAR-100							
Joint training (upper bound)	$636.2_{\pm 2.4}$	$784.3_{\pm 2.1}$	$2001.0_{\pm 2.1}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$44.9_{\pm 0.0}$
EWC	$801.7_{\pm 52.5}$	$931.8_{\pm 55.7}$	$2148.5_{\pm 55.7}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$134.4_{\pm 0.0}$	$134.5_{\pm 0.0}$
SI	$1984.8_{\pm 766.0}$	$2080.4_{\pm 773.7}$	$3297.1_{\pm 773.7}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$313.4_{\pm 0.0}$	$313.6_{\pm 0.0}$
LwF	$3404.3_{\pm 122.6}$	$3577.0_{\pm 126.4}$	$4793.7_{\pm 126.4}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$89.8_{\pm 0.0}$
WSN	$945.6_{\pm 1.8}$	$1174.4_{\pm 50.9}$	$2391.1_{\pm 50.9}$	$178.6_{\pm 0.0}$	$89.4_{\pm 0.0}$	$213.0_{\pm 0.0}$	$213.4_{\pm 0.0}$
PEC	–	–	–	–	–	–	–
SpaceNet	$105.2_{\pm 0.8}$	$126.5_{\pm 12.1}$	$1334.8_{\pm 12.1}$	$11.3_{\pm 0.0}$	$5.6_{\pm 0.0}$	$26.8_{\pm 0.0}$	$14.1_{\pm 0.0}$
NISPA	$503.4_{\pm 75.4}$	$775.2_{\pm 135.5}$	$1991.9_{\pm 135.5}$	$70.0_{\pm 0.0}$	$35.0_{\pm 0.0}$	$78.8_{\pm 0.0}$	$78.8_{\pm 0.0}$
SNV	$770.6_{\pm 4.5}$	$1210.1_{\pm 0.0}$	$2428.9_{\pm 0.0}$	$89.4_{\pm 0.0}$	$44.8_{\pm 0.0}$	$49.9_{\pm 0.0}$	$46.8_{\pm 0.0}$
UniCLUN	$1074.3_{\pm 2.4}$	$1296.0_{\pm 0.0}$	$2514.8_{\pm 0.0}$	$92.2_{\pm 0.0}$	$92.2_{\pm 0.0}$	$104.6_{\pm 0.0}$	$104.7_{\pm 0.0}$
LwU (ours)	$916.3_{\pm 5.5}$	$1335.9_{\pm 29.1}$	$2554.7_{\pm 29.1}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$44.8_{\pm 0.0}$	$45.0_{\pm 0.0}$
SGD (lower bound)	$543.5_{\pm 0.0}$	$698.4_{\pm 0.0}$	$1915.0_{\pm 0.0}$	$89.4_{\pm 0.0}$	$44.8_{\pm 0.0}$	$44.8_{\pm 0.0}$	$44.9_{\pm 0.0}$
Seq-Tiny-ImageNet							
Joint training (upper bound)	–	–	–	–	–	–	–
EWC	$1564.8_{\pm 1.2}$	$1814.0_{\pm 0.0}$	$3178.8_{\pm 0.0}$	$85.3_{\pm 0.0}$	$42.8_{\pm 0.0}$	$128.0_{\pm 0.0}$	$128.2_{\pm 0.0}$
SI	$1910.7_{\pm 1.0}$	$2174.0_{\pm 0.0}$	$3538.8_{\pm 0.0}$	$85.3_{\pm 0.0}$	$42.8_{\pm 0.0}$	$298.5_{\pm 0.0}$	$298.7_{\pm 0.0}$
LwF	$11773.7_{\pm 0.9}$	$12097.3_{\pm 68.1}$	$13462.1_{\pm 68.1}$	$85.6_{\pm 0.0}$	$85.6_{\pm 0.0}$	$85.7_{\pm 0.0}$	$85.8_{\pm 0.0}$
WSN	$1870.6_{\pm 3.6}$	$2339.0_{\pm 1.2}$	$3555.7_{\pm 1.2}$	$178.7_{\pm 0.0}$	$89.5_{\pm 0.0}$	$213.1_{\pm 0.0}$	$213.5_{\pm 0.0}$
PEC	–	–	–	–	–	–	–
SpaceNet	$340.8_{\pm 0.0}$	$385.9_{\pm 0.0}$	$1596.3_{\pm 0.0}$	$40.9_{\pm 0.0}$	$20.5_{\pm 0.0}$	$97.3_{\pm 0.0}$	$51.2_{\pm 0.0}$
NISPA	$1905.1_{\pm 80.4}$	$3489.0_{\pm 553.8}$	$4705.6_{\pm 553.8}$	$272.2_{\pm 0.0}$	$136.1_{\pm 0.0}$	$306.2_{\pm 0.0}$	$306.2_{\pm 0.0}$
SNV	$3232.3_{\pm 3.4}$	$5044.3_{\pm 1.2}$	$6263.1_{\pm 1.2}$	$89.4_{\pm 0.0}$	$44.9_{\pm 0.0}$	$73.4_{\pm 0.0}$	$46.9_{\pm 0.0}$
UniCLUN	$3388.5_{\pm 1.2}$	$4540.3_{\pm 40.0}$	$5759.1_{\pm 40.0}$	$92.4_{\pm 0.0}$	$92.4_{\pm 0.0}$	$141.6_{\pm 0.0}$	$141.8_{\pm 0.0}$
LwU (ours)	$2074.8_{\pm 9.8}$	$2298.7_{\pm 1.2}$	$3665.4_{\pm 1.2}$	$85.6_{\pm 0.0}$	$42.8_{\pm 0.0}$	$42.8_{\pm 0.0}$	$43.0_{\pm 0.0}$
SGD (lower bound)	$1430.9_{\pm 0.3}$	$1804.0_{\pm 0.0}$	$3168.8_{\pm 0.0}$	$85.3_{\pm 0.0}$	$42.8_{\pm 0.0}$	$42.8_{\pm 0.0}$	$42.9_{\pm 0.0}$
RTI							
Joint training (upper bound)	$643_{\pm 25}$	$916_{\pm 29}$	$2347_{\pm 29}$	89.4	89.4	89.5	$44.8_{\pm 0.0}$
EWC	$696_{\pm 3}$	$795_{\pm 16}$	$2226_{\pm 16}$	89.4	44.7	134.1	$134.2_{\pm 0.0}$
SI	$1246_{\pm 7}$	$1328_{\pm 19}$	$2759_{\pm 19}$	89.4	44.7	312.9	$313.1_{\pm 0.0}$
LwF	$1849_{\pm 29}$	$1902_{\pm 20}$	$3333_{\pm 20}$	89.4	89.4	89.5	$89.6_{\pm 0.0}$
WSN	$912_{\pm 0}$	$1045_{\pm 1}$	$2476_{\pm 1}$	178.6	89.3	156.7	$157.0_{\pm 0.0}$
PEC	–	–	–	–	–	–	–
SpaceNet	$105_{\pm 0}$	$127_{\pm 12}$	$1541_{\pm 12}$	11.1	5.6	26.5	$14.0_{\pm 0.0}$
NISPA	$395_{\pm 5}$	$642_{\pm 40}$	$2073_{\pm 40}$	69.4	34.7	78.0	$78.1_{\pm 0.0}$
SNV	$785_{\pm 5}$	$1188_{\pm 1}$	$2619_{\pm 1}$	89.4	44.7	49.8	$46.8_{\pm 0.0}$
UniCLUN	$1209_{\pm 1}$	1441	2874	92.0	92.0	104.4	$104.5_{\pm 0.0}$
LwU (ours)	$881_{\pm 27}$	$958_{\pm 41}$	$2389_{\pm 41}$	89.4	44.7	245.9	$44.8_{\pm 0.0}$
SGD (lower bound)	566	640	2071	89.4	44.7	44.7	$44.8_{\pm 0.0}$

Table 27: Memory and storage measurements for CIL.

Method	Peak alloc. (MB)↓	Peak reserved (MB)↓	Peak device (MB)↓	Optimizer (MB)↓	Model params (MB)↓	Resident storage (MB)↓	Checkpoint (MB)↓
Seq-CIFAR-100							
Joint training (upper bound)	$694.9_{\pm 27.7}$	$877.3_{\pm 3.2}$	$2094.0_{\pm 3.2}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$44.9_{\pm 0.0}$
EWC	$897.3_{\pm 191.4}$	$1059.8_{\pm 151.9}$	$2276.4_{\pm 151.9}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$134.4_{\pm 0.0}$	$134.5_{\pm 0.0}$
SI	$2219.6_{\pm 160.6}$	$2325.7_{\pm 134.5}$	$3542.4_{\pm 134.5}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$313.5_{\pm 0.0}$	$313.7_{\pm 0.0}$
LwF	$810.1_{\pm 23.2}$	$927.6_{\pm 52.1}$	$2144.3_{\pm 52.1}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$89.6_{\pm 0.0}$	$89.8_{\pm 0.0}$
WSN	–	–	–	–	–	–	–
PEC	$2215.7_{\pm 0.0}$	$3259.0_{\pm 0.0}$	$4473.6_{\pm 0.0}$	$0.9_{\pm 0.0}$	$66.8_{\pm 0.0}$	$66.8_{\pm 0.0}$	$67.0_{\pm 0.0}$
SpaceNet	$105.9_{\pm 0.1}$	$127.2_{\pm 11.6}$	$1335.5_{\pm 11.6}$	$11.3_{\pm 0.0}$	$5.6_{\pm 0.0}$	$26.8_{\pm 0.0}$	$14.2_{\pm 0.0}$
NISPA	$543.8_{\pm 17.0}$	$769.0_{\pm 49.8}$	$1985.6_{\pm 49.8}$	$70.0_{\pm 0.0}$	$35.0_{\pm 0.0}$	$78.8_{\pm 0.0}$	$78.8_{\pm 0.0}$
SNV	$811.3_{\pm 2.4}$	$1209.4_{\pm 1.2}$	$2428.1_{\pm 1.2}$	$89.4_{\pm 0.0}$	$44.9_{\pm 0.0}$	$50.0_{\pm 0.0}$	$46.9_{\pm 0.0}$
UniCLUN	$1074.3_{\pm 2.4}$	$1296.0_{\pm 0.0}$	$2514.8_{\pm 0.0}$	$92.2_{\pm 0.0}$	$92.2_{\pm 0.0}$	$104.6_{\pm 0.0}$	$104.7_{\pm 0.0}$
LwU (ours)	$909.8_{\pm 1.0}$	$1335.2_{\pm 27.8}$	$2554.0_{\pm 27.8}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$44.8_{\pm 0.0}$	$45.0_{\pm 0.0}$
SGD (lower bound)	$543.8_{\pm 0.0}$	$699.7_{\pm 1.2}$	$1916.4_{\pm 1.2}$	$89.6_{\pm 0.0}$	$44.8_{\pm 0.0}$	$44.8_{\pm 0.0}$	$44.9_{\pm 0.0}$
Seq-Tiny-ImageNet							
Joint training (upper bound)	$1562.6_{\pm 2.4}$	$2187.3_{\pm 0.0}$	$3404.0_{\pm 0.0}$	$89.8_{\pm 0.0}$	$89.8_{\pm 0.0}$	$89.8_{\pm 0.0}$	$45.0_{\pm 0.0}$
EWC	$1566.1_{\pm 1.8}$	$1814.0_{\pm 0.0}$	$3178.8_{\pm 0.0}$	$85.6_{\pm 0.0}$	$42.8_{\pm 0.0}$	$128.0_{\pm 0.0}$	$128.2_{\pm 0.0}$
SI	$1912.5_{\pm 1.8}$	$2174.7_{\pm 1.2}$	$3539.4_{\pm 1.2}$	$85.6_{\pm 0.0}$	$42.8_{\pm 0.0}$	$298.5_{\pm 0.0}$	$298.7_{\pm 0.0}$
LwF	$1817.4_{\pm 1.6}$	$2012.0_{\pm 0.0}$	$3376.8_{\pm 0.0}$	$85.6_{\pm 0.0}$	$85.6_{\pm 0.0}$	$85.7_{\pm 0.0}$	$85.8_{\pm 0.0}$
WSN	–	–	–	–	–	–	–
PEC	$8560.9_{\pm 0.0}$	$12777.9_{\pm 0.0}$	$13992.5_{\pm 0.0}$	$0.9_{\pm 0.0}$	$89.1_{\pm 0.0}$	$89.1_{\pm 0.0}$	$89.4_{\pm 0.0}$
SpaceNet	$343.3_{\pm 0.0}$	$385.9_{\pm 0.0}$	$1596.3_{\pm 0.0}$	$40.9_{\pm 0.0}$	$20.5_{\pm 0.0}$	$97.3_{\pm 0.0}$	$51.2_{\pm 0.0}$
NISPA	$2022.2_{\pm 57.8}$	$4443.2_{\pm 334.7}$	$5659.8_{\pm 334.7}$	$272.2_{\pm 0.0}$	$136.1_{\pm 0.0}$	$306.2_{\pm 0.0}$	$306.2_{\pm 0.0}$
SNV	$3233.3_{\pm 1.6}$	$5047.1_{\pm 1.2}$	$6265.9_{\pm 1.2}$	$89.4_{\pm 0.0}$	$45.0_{\pm 0.0}$	$73.5_{\pm 0.0}$	$47.0_{\pm 0.0}$
UniCLUN	$3387.9_{\pm 1.4}$	$4540.3_{\pm 40.0}$	$5759.1_{\pm 40.0}$	$92.4_{\pm 0.0}$	$92.4_{\pm 0.0}$	$141.6_{\pm 0.0}$	$141.8_{\pm 0.0}$
LwU (ours)	$2068.4_{\pm 2.3}$	$2298.0_{\pm 0.0}$	$3664.8_{\pm 0.0}$	$85.6_{\pm 0.0}$	$42.8_{\pm 0.0}$	$42.8_{\pm 0.0}$	$43.0_{\pm 0.0}$
SGD (lower bound)	$1431.2_{\pm 0.1}$	$1804.0_{\pm 0.0}$	$3168.8_{\pm 0.0}$	$85.6_{\pm 0.0}$	$42.8_{\pm 0.0}$	$42.8_{\pm 0.0}$	$42.9_{\pm 0.0}$
RTI							
Joint training (upper bound)	$628_{\pm 26}$	$909_{\pm 25}$	$2340_{\pm 25}$	89.4	89.4	89.5	$44.8_{\pm 0.0}$
EWC	$755_{\pm 52}$	$842_{\pm 46}$	$2273_{\pm 46}$	89.4	44.7	134.1	$134.2_{\pm 0.0}$
SI	$1244_{\pm 2}$	$1314_{\pm 17}$	$2745_{\pm 17}$	89.4	44.7	312.9	$313.1_{\pm 0.0}$
LwF	$659_{\pm 1}$	$734_{\pm 22}$	$2165_{\pm 22}$	89.4	89.4	89.5	$89.6_{\pm 0.0}$
SNV	$826_{\pm 3}$	$1190_{\pm 1}$	$2621_{\pm 1}$	89.4	44.7	49.9	$46.8_{\pm 0.0}$
PEC	2176	$3212_{\pm 1}$	$4643_{\pm 1}$	0.9	49.0	49.0	49.0
SpaceNet	$105_{\pm 0}$	120	1534	11.1	5.6	26.5	$14.0_{\pm 0.0}$
NISPA	$395_{\pm 5}$	$705_{\pm 42}$	$2136_{\pm 42}$	69.4	34.7	78.0	$78.1_{\pm 0.0}$
UniCLUN	$1210_{\pm 1}$	1441	2874	92.0	92.0	104.4	$104.5_{\pm 0.0}$
LwU (ours)	$882_{\pm 25}$	$938_{\pm 13}$	$2369_{\pm 13}$	89.4	44.7	245.9	$44.8_{\pm 0.0}$
SGD (lower bound)	566	640	2071	89.4	44.7	44.7	$44.8_{\pm 0.0}$

Table 28: Inference cost comparison.

	TIL				CIL			
Method	B1 ms/sample↓	GFLOPs/sample↓	B64 ms/sample↓	B64 sample/s↑	B1 ms/sample↓	GFLOPs/sample↓	B64 ms/sample↓	B64 sample/s↑
Seq-CIFAR-100								
Joint training (upper bound)	$1.837_{\pm 0.229}$	$1.111_{\pm 0.000}$	$0.0266_{\pm 0.0012}$	$38216.0_{\pm 1351.0}$	$1.844_{\pm 0.122}$	$1.111_{\pm 0.000}$	$0.0271_{\pm 0.0002}$	$37238.2_{\pm 120.2}$
EWC	$1.624_{\pm 0.112}$	$1.111_{\pm 0.000}$	$0.0247_{\pm 0.0006}$	$40610.6_{\pm 967.4}$	$1.755_{\pm 0.053}$	$1.111_{\pm 0.000}$	$0.0260_{\pm 0.0004}$	$38555.8_{\pm 612.4}$
SI	$1.756_{\pm 0.100}$	$1.111_{\pm 0.000}$	$0.0258_{\pm 0.0003}$	$39074.8_{\pm 502.2}$	$1.785_{\pm 0.148}$	$1.111_{\pm 0.000}$	$0.0270_{\pm 0.0010}$	$37464.4_{\pm 1043.3}$
LwF	$1.549_{\pm 0.039}$	$1.111_{\pm 0.000}$	$0.0250_{\pm 0.0003}$	$40058.7_{\pm 361.4}$	$1.861_{\pm 0.081}$	$1.111_{\pm 0.000}$	$0.0267_{\pm 0.0003}$	$37556.1_{\pm 373.1}$
WSN	$8.109_{\pm 0.404}$	$1.111_{\pm 0.000}$	$0.1233_{\pm 0.0038}$	$8556.2_{\pm 192.1}$	–	–	–	–
PEC	–	–	–	–	$6.27_{\pm 0.671}$	$0.8365_{\pm 0.000}$	$0.2209_{\pm 0.0004}$	$4526.5_{\pm 7.3}$
SpaceNet	$0.377_{\pm 0.142}$	$0.003_{\pm 0.000}$	$0.0055_{\pm 0.0012}$	$219715.3_{\pm 48405.0}$	$0.271_{\pm 0.113}$	$0.003_{\pm 0.000}$	$0.0041_{\pm 0.0006}$	$281369.5_{\pm 37115.0}$
NISPA	$0.761_{\pm 0.212}$	$0.209_{\pm 0.000}$	$0.0121_{\pm 0.0028}$	$95640.6_{\pm 16886.8}$	$0.611_{\pm 0.131}$	$0.209_{\pm 0.000}$	$0.0104_{\pm 0.0007}$	$100432.2_{\pm 5334.5}$
SNV	$8.086_{\pm 0.280}$	$1.111_{\pm 0.000}$	$0.1237_{\pm 0.0071}$	$8500.0_{\pm 394.3}$	$121.286_{\pm 4.129}$	$44.434_{\pm 0.000}$	$1.9836_{\pm 0.0468}$	$505.1_{\pm 11.1}$
UniCLUN	$1.258_{\pm 0.095}$	$1.111_{\pm 0.000}$	$0.0249_{\pm 0.0009}$	$40523.6_{\pm 988.8}$	$1.806_{\pm 0.227}$	$1.111_{\pm 0.000}$	$0.0264_{\pm 0.0014}$	$38236.6_{\pm 1506.7}$
LwU (ours)	$2.590_{\pm 0.080}$	$1.111_{\pm 0.000}$	$0.0375_{\pm 0.0010}$	$26802.5_{\pm 637.3}$	$2.559_{\pm 0.189}$	$1.111_{\pm 0.000}$	$0.0378_{\pm 0.0002}$	$26555.0_{\pm 181.9}$
SGD (lower bound)	$1.561_{\pm 0.266}$	$1.111_{\pm 0.000}$	$0.0254_{\pm 0.0006}$	$39679.7_{\pm 726.6}$	$1.708_{\pm 0.125}$	$1.111_{\pm 0.000}$	$0.0265_{\pm 0.0005}$	$37980.6_{\pm 529.0}$
Seq-Tiny-ImageNet								
Joint training (upper bound)	$2.34_{\pm 0.18}$	$4.443_{\pm 0.000}$	$0.093_{\pm 0.002}$	$10783_{\pm 197}$	$2.64_{\pm 0.33}$	$4.443_{\pm 0.000}$	$0.093_{\pm 0.001}$	$10787_{\pm 119}$
EWC	$0.971_{\pm 0.368}$	$4.443_{\pm 0.000}$	$0.0969_{\pm 0.0025}$	$10323_{\pm 271}$	$0.941_{\pm 0.065}$	$4.443_{\pm 0.000}$	$0.0974_{\pm 0.0017}$	$10273_{\pm 181}$
SI	$0.880_{\pm 0.047}$	$4.443_{\pm 0.000}$	$0.0971_{\pm 0.0009}$	$10299_{\pm 92}$	$0.899_{\pm 0.032}$	$4.443_{\pm 0.000}$	$0.0974_{\pm 0.0007}$	$10269_{\pm 69}$
LwF	$0.820_{\pm 0.070}$	$4.443_{\pm 0.000}$	$0.0976_{\pm 0.0018}$	$10244_{\pm 185}$	$0.974_{\pm 0.029}$	$4.443_{\pm 0.000}$	$0.0980_{\pm 0.0003}$	$10209_{\pm 33}$
WSN	$7.15_{\pm 0.22}$	$4.44_{\pm 0.00}$	$0.153_{\pm 0.004}$	$6575_{\pm 151}$	–	–	–	–
PEC	–	–	–	–	$12.45_{\pm 0.48}$	$1.806_{\pm 0.000}$	$0.919_{\pm 0.005}$	$1088_{\pm 6}$
SpaceNet	$0.48_{\pm 0.16}$	$0.01_{\pm 0.00}$	$0.008_{\pm 0.001}$	$171746_{\pm 10934}$	$0.50_{\pm 0.12}$	$0.01_{\pm 0.00}$	$0.009_{\pm 0.001}$	$182344_{\pm 30154}$
NISPA	$1.41_{\pm 0.12}$	$0.84_{\pm 0.00}$	$0.033_{\pm 0.001}$	$30396_{\pm 655}$	$1.51_{\pm 0.14}$	$0.84_{\pm 0.00}$	$0.033_{\pm 0.001}$	$30604_{\pm 829}$
SNV	$7.71_{\pm 0.07}$	$4.44_{\pm 0.00}$	$0.162_{\pm 0.000}$	$6192_{\pm 3}$	$128.04_{\pm 3.00}$	$177.73_{\pm 0.00}$	$4.669_{\pm 0.038}$	$214_{\pm 2}$
UniCLUN	$2.28_{\pm 0.28}$	$4.44_{\pm 0.00}$	$0.094_{\pm 0.001}$	$10638_{\pm 111}$	$2.28_{\pm 0.42}$	$4.44_{\pm 0.00}$	$0.094_{\pm 0.001}$	$10654_{\pm 76}$
LwU (ours)	$1.441_{\pm 0.110}$	$4.444_{\pm 0.000}$	$0.1014_{\pm 0.0001}$	$9861_{\pm 11}$	$1.195_{\pm 0.123}$	$4.444_{\pm 0.000}$	$0.1014_{\pm 0.0001}$	$9858_{\pm 11}$
SGD (lower bound)	$0.843_{\pm 0.129}$	$4.443_{\pm 0.000}$	$0.0951_{\pm 0.0014}$	$10521_{\pm 153}$	$1.044_{\pm 0.375}$	$4.443_{\pm 0.000}$	$0.0972_{\pm 0.0012}$	$10288_{\pm 129}$
RTI								
Joint training (upper bound)	$0.649_{\pm 0.018}$	1.111	$0.0215_{\pm 0.0000}$	$46415_{\pm 54}$	$0.660_{\pm 0.030}$	1.111	$0.0228_{\pm 0.0001}$	$43929_{\pm 124}$
EWC	$0.692_{\pm 0.033}$	1.111	$0.0216_{\pm 0.0000}$	$46200_{\pm 45}$	$0.704_{\pm 0.023}$	1.111	$0.0228_{\pm 0.0001}$	$43917_{\pm 152}$
SI	$0.662_{\pm 0.001}$	1.111	$0.0217_{\pm 0.0001}$	$45998_{\pm 107}$	$0.659_{\pm 0.035}$	1.111	$0.0229_{\pm 0.0000}$	$43665_{\pm 28}$
LwF	$0.643_{\pm 0.029}$	1.111	$0.0218_{\pm 0.0001}$	$45896_{\pm 118}$	$0.658_{\pm 0.031}$	1.111	$0.0229_{\pm 0.0000}$	$43751_{\pm 27}$
WSN	$3.260_{\pm 0.171}$	1.111	$0.0529_{\pm 0.0001}$	$18894_{\pm 45}$	–	–	–	–
PEC	–	–	–	–	$0.758_{\pm 0.076}$	0.2675	$0.1388_{\pm 0.0003}$	$7208_{\pm 16}$
SpaceNet	$0.074_{\pm 0.017}$	0.002786	$0.0013_{\pm 0.0001}$	$772264_{\pm 43835}$	$0.075_{\pm 0.009}$	0.002786	$0.0014_{\pm 0.0000}$	$748739_{\pm 3971}$
NISPA	$0.152_{\pm 0.013}$	0.2091	$0.0063_{\pm 0.0000}$	$157682_{\pm 279}$	$0.147_{\pm 0.007}$	0.2091	$0.0063_{\pm 0.0000}$	$157484_{\pm 455}$
SNV	$3.247_{\pm 0.112}$	1.111	$0.0532_{\pm 0.0009}$	$18797_{\pm 318}$	$44.12_{\pm 1.51}$	44.434	$1.6901_{\pm 0.0398}$	$592_{\pm 14}$
UniCLUN	$0.624_{\pm 0.029}$	1.111	$0.0215_{\pm 0.0000}$	$46581_{\pm 25}$	$0.635_{\pm 0.028}$	1.111	$0.0227_{\pm 0.0000}$	$44123_{\pm 70}$
LwU (ours)	$6.177_{\pm 0.035}$	4.44	$0.1576_{\pm 0.0009}$	$6346_{\pm 34}$	$5.833_{\pm 0.664}$	4.44	$0.1556_{\pm 0.0027}$	$6431_{\pm 112}$
SGD (lower bound)	$0.631_{\pm 0.021}$	1.111	$0.0217_{\pm 0.0000}$	$46180_{\pm 57}$	$0.623_{\pm 0.010}$	1.111	$0.0227_{\pm 0.0000}$	$43970_{\pm 35}$

H HYPERPARAMETERS

Table 29: Unlearning hyperparameter grids, shared across datasets, backbones, and deletion modes (except LwU's q_f, see ‡). # counts candidates; $\mathcal{D}_r^{10\%}$ is a fixed random 10% subset of $\mathcal{D}_r$.

Method	Data	Searched grid	#	Fixed settings
Retrain	$\mathcal{D}_r$	–	–	Source training recipe
Amnesiac†	update ledger	–	–	Subtracts updates recorded on batches containing $\mathcal{D}_f$
Fine-tune	$\mathcal{D}_r$	lr $\{10^{-4}, 10^{-3}, 10^{-2}\}$; epochs $\{1, 3, 5, 10\}$	12	Adam
Bad Teacher	$\mathcal{D}_f$, $\mathcal{D}_r^{10\%}$	T $\{1, 2, 4\}$; epochs $\{1, 3, 5, 10\}$; lr $\{10^{-4}, 10^{-3}, 10^{-2}\}$	36	SGD, momentum 0.9; incompetent teacher is a re-initialized copy
UNSIR	$\mathcal{D}_f$, $\mathcal{D}_r^{10\%}$	impair epochs $\{1, 2\}$; repair epochs $\{1, 2, 3\}$; impair lr $\{10^{-3}, 10^{-2}, 10^{-1}\}$; repair lr $\{10^{-3}, 10^{-2}\}$	36	Noise: 40 Adam steps, lr 0.1, regularization 0.1
SSD	$\mathcal{D}_f$, Fisher on $\mathcal{D}$	α $\{1, 5, 10, 50, 100\}$; λ $\{0.1, 0.5, 1, 5, 10\}$	25	No gradient updates
UniCLUN	$\mathcal{D}_f$, $\mathcal{D}_r^{10\%}$	lr $\{10^{-4}, 10^{-3}, 10^{-2}\}$; epochs $\{3, 5, 10\}$; buffer $\{250, 500, 1000\}$	27	SGD, momentum 0.9; $\rho = 2$; contrastive $T = 0.1$; $\alpha_1 = \alpha_2 = \alpha_3 = 1$; teacher momentum 0.99; $p = 0.6$
SalUn	$\mathcal{D}_f$, $\mathcal{D}_r^{10\%}$	mask density $\{0.1, 0.3, 0.5, 0.7, 0.9\}$; lr $\{10^{-4}, 10^{-3}, 10^{-2}\}$; epochs $\{3, 5, 10\}$	45	SGD, momentum 0.9, weight decay 5×10^{-4}
SCRUB	$\mathcal{D}_f$, $\mathcal{D}_r^{10\%}$	$E_{\max}$ $\{1, 2, 3\}$; lr $\{10^{-4}, 5 \times 10^{-4}, 10^{-3}, 5 \times 10^{-3}\}$; E $\{5, 10\}$	24	SGD, momentum 0.9, weight decay 5×10^{-4}; $T = 4$; $\alpha = \gamma = 1$; forget-step gradient-norm clip 5.0
LwU	$\mathcal{D}_f$, $\mathcal{D}_r^{10\%}$	dominance margin $\{1, 3\}$; distill. epochs $\{10, 30\}$; distill. lr $\{10^{-3}, 10^{-2}\}$; weight alignment {off, on}; memory (η_m, s_m) $\{(10^{-2}, 0.25), (10^{-1}, 1)\}$; q_f $\{0.9, 0.99\}^{\ddagger}$	$64^{\ddagger}$	SGD distillation; Fisher importance on $\mathcal{D}_f$ and $\mathcal{D}_r^{10\%}$; retain quantile 0.9; zone C orthogonal; $T = 4$; $\lambda_{\mathrm{sdft}} = 0.5$; memory confidence 0.8

†Class deletion only; the ledger has no equivalent for subclass or instance deletion. ‡Class deletion fixes $q_f = 0.99$ (32 candidates).

Table 32: Selected hyperparameters on CIFAR-10.

Backbone	Deletion	Method	Seed 42	Seed 43	Seed 44
CNN	class	Fine-tune	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10	lr = 10^{-3}, ep = 10
		Bad Teacher	T = 4, ep = 3, lr = 10^{-2}	T = 4, ep = 10, lr = 10^{-3}	T = 4, ep = 5, lr = 10^{-3}
		UNSIR	imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	α = 5, λ = 0.1	α = 1, λ = 10	α = 1, λ = 5
		SCRUB	max = 2, lr = 5×10^{-4}, ep = 5	max = 3, lr = 5×10^{-4}, ep = 5	max = 3, lr = 5×10^{-4}, ep = 5
		SalUn	sp = 0.1, lr = 10^{-4}, ep = 3	sp = 0.5, lr = 10^{-4}, ep = 10	sp = 0.9, lr = 10^{-3}, ep = 10
		UniCLUN	lr = 10^{-4}, ep = 3, buf = 250	lr = 10^{-4}, ep = 3, buf = 250	lr = 10^{-2}, ep = 5, buf = 250
		LwU	margin = 3, d. ep = 30, d. lr = 10^{-2}, align = off, mem = `m10_s100_c08`	margin = 3, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`	margin = 3, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`
	subclass	Fine-tune	lr = 10^{-3}, ep = 3	lr = 10^{-3}, ep = 10	lr = 10^{-3}, ep = 10
		Bad Teacher	T = 4, ep = 10, lr = 10^{-2}	T = 4, ep = 5, lr = 10^{-2}	T = 2, ep = 10, lr = 10^{-2}
		UNSIR	imp. ep = 1, rep. ep = 3, imp. lr = 0.1, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 2, imp. lr = 10^{-3}, rep. lr = 10^{-3}	imp. ep = 1, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	α = 10, λ = 0.5	α = 1, λ = 5	α = 5, λ = 0.1
		SCRUB	max = 3, lr = 10^{-4}, ep = 5	max = 3, lr = 10^{-4}, ep = 5	max = 3, lr = 5×10^{-4}, ep = 5
		SalUn	sp = 0.1, lr = 10^{-4}, ep = 3	sp = 0.1, lr = 10^{-4}, ep = 3	sp = 0.1, lr = 10^{-4}, ep = 3
		UniCLUN	lr = 10^{-4}, ep = 5, buf = 250	lr = 10^{-4}, ep = 5, buf = 250	lr = 10^{-4}, ep = 3, buf = 250
		LwU	q_f = 0.9, margin = 3, d. ep = 30, d. lr = 10^{-2}, align = on, mem = `m01_s025_c08`	q_f = 0.99, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`	q_f = 0.99, margin = 3, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`
	instance	Fine-tune	lr = 10^{-3}, ep = 10	lr = 10^{-3}, ep = 5	lr = 10^{-3}, ep = 10
		Bad Teacher	T = 2, ep = 10, lr = 10^{-2}	T = 2, ep = 5, lr = 10^{-2}	T = 1, ep = 5, lr = 10^{-4}
		UNSIR	imp. ep = 2, rep. ep = 1, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	α = 5, λ = 1	α = 1, λ = 5	α = 1, λ = 5
		SCRUB	max = 3, lr = 5×10^{-3}, ep = 10	max = 2, lr = 5×10^{-3}, ep = 10	max = 2, lr = 5×10^{-3}, ep = 10
		SalUn	sp = 0.9, lr = 10^{-2}, ep = 10 ×	sp = 0.7, lr = 10^{-2}, ep = 10 ×	sp = 0.1, lr = 10^{-2}, ep = 10
		UniCLUN	lr = 10^{-4}, ep = 3, buf = 250	lr = 10^{-2}, ep = 10, buf = 250	lr = 10^{-2}, ep = 5, buf = 250

Backbone	Deletion	Method	Seed 42	Seed 43	Seed 44
		LwU	$q_f = 0.99$, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = m10_s100_c08	$q_f = 0.9$, margin = 3, d. ep = 30, d. lr = 10^{-2}, align = on, mem = m10_s100_c08	$q_f = 0.9$, margin = 3, d. ep = 30, d. lr = 10^{-3}, align = on, mem = m10_s100_c08
ResNet-18	class	Fine-tune	lr = 10^{-4}, ep = 10	lr = 10^{-3}, ep = 10	lr = 10^{-3}, ep = 10
		Bad Teacher	$T = 4$, ep = 10, lr = 10^{-2}	$T = 4$, ep = 10, lr = 10^{-2}	$T = 4$, ep = 10, lr = 10^{-2}
		UNSIR	imp. ep = 1, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	$\alpha = 100$, $\lambda = 0.1$	$\alpha = 5$, $\lambda = 0.5$	$\alpha = 10$, $\lambda = 0.1$
		SCRUB	max = 3, lr = 5×10^{-4}, ep = 10	max = 2, lr = 5×10^{-4}, ep = 5	max = 2, lr = 5×10^{-4}, ep = 5
		SalUn	sp = 0.1, lr = 10^{-4}, ep = 3	sp = 0.7, lr = 10^{-2}, ep = 10	sp = 0.7, lr = 10^{-2}, ep = 10
		UniCLUN	lr = 10^{-4}, ep = 3, buf = 250 $\times$	lr = 10^{-4}, ep = 10, buf = 250	lr = 10^{-2}, ep = 10, buf = 250
		LwU	margin = 1, d. ep = 30, d. lr = 10^{-2}, align = off, mem = m10_s100_c08	margin = 1, d. ep = 30, d. lr = 10^{-2}, align = on, mem = m10_s100_c08	margin = 3, d. ep = 30, d. lr = 10^{-2}, align = off, mem = m10_s100_c08
	subclass	Fine-tune	lr = 10^{-3}, ep = 10	lr = 10^{-3}, ep = 10	lr = 10^{-3}, ep = 10
		Bad Teacher	$T = 4$, ep = 10, lr = 10^{-2}	$T = 1$, ep = 5, lr = 10^{-2}	$T = 4$, ep = 10, lr = 10^{-2}
		UNSIR	imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 1, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	$\alpha = 10$, $\lambda = 0.1$	$\alpha = 1$, $\lambda = 0.5$	$\alpha = 10$, $\lambda = 5$
		SCRUB	max = 3, lr = 5×10^{-3}, ep = 10	max = 3, lr = 5×10^{-3}, ep = 5	max = 3, lr = 5×10^{-3}, ep = 10
		SalUn	sp = 0.7, lr = 10^{-2}, ep = 10	sp = 0.7, lr = 10^{-2}, ep = 10	sp = 0.3, lr = 10^{-4}, ep = 5
		UniCLUN	lr = 10^{-2}, ep = 5, buf = 250	lr = 10^{-2}, ep = 10, buf = 250	lr = 10^{-4}, ep = 5, buf = 250
		LwU	$q_f = 0.9$, margin = 3, d. ep = 30, d. lr = 10^{-2}, align = on, mem = m01_s025_c08	$q_f = 0.99$, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = m01_s025_c08	$q_f = 0.99$, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = on, mem = m01_s025_c08
	instance	Fine-tune	lr = 10^{-3}, ep = 10	lr = 10^{-4}, ep = 10	lr = 10^{-3}, ep = 10
		Bad Teacher	$T = 4$, ep = 5, lr = 10^{-2}	$T = 4$, ep = 5, lr = 10^{-2}	$T = 2$, ep = 3, lr = 10^{-4}
		UNSIR	imp. ep = 1, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 1, imp. lr = 10^{-3}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 1, imp. lr = 10^{-3}, rep. lr = 10^{-2}
		SSD	$\alpha = 1$, $\lambda = 10$	$\alpha = 1$, $\lambda = 5$	$\alpha = 1$, $\lambda = 5$
		SCRUB	max = 2, lr = 5×10^{-3}, ep = 5	max = 2, lr = 5×10^{-3}, ep = 5	max = 2, lr = 5×10^{-3}, ep = 10
		SalUn	sp = 0.7, lr = 10^{-2}, ep = 5 $\times$	sp = 0.3, lr = 10^{-4}, ep = 3 $\times$	sp = 0.1, lr = 10^{-4}, ep = 3 $\times$
		UniCLUN	lr = 10^{-2}, ep = 10, buf = 250 $\times$	lr = 10^{-4}, ep = 3, buf = 250	lr = 10^{-4}, ep = 3, buf = 250 $\times$
		LwU	$q_f = 0.99$, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = on, mem = m10_s100_c08	$q_f = 0.99$, margin = 3, d. ep = 10, d. lr = 10^{-3}, align = on, mem = m10_s100_c08	$q_f = 0.99$, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = m10_s100_c08

Table 33: Selected hyperparameters on RTI.

Backbone	Deletion	Method	Seed 42	Seed 43	Seed 44
CNN	class	Fine-tune	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10
		Bad Teacher	$T = 4$, ep = 10, lr = 10^{-2}	$T = 4$, ep = 3, lr = 10^{-2}	$T = 2$, ep = 10, lr = 10^{-2}
		UNSIR	imp. ep = 1, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	$\alpha = 50$, $\lambda = 0.1$	$\alpha = 50$, $\lambda = 0.1$	$\alpha = 10$, $\lambda = 0.5$
		SCRUB	max = 3, lr = 10^{-4}, ep = 5	max = 3, lr = 5×10^{-4}, ep = 5	max = 3, lr = 10^{-4}, ep = 5
		SalUn	sp = 0.9, lr = 10^{-4}, ep = 3 $\times$	sp = 0.1, lr = 10^{-4}, ep = 5 $\times$	sp = 0.9, lr = 10^{-4}, ep = 3 $\times$
		UniCLUN	lr = 10^{-4}, ep = 5, buf = 250 $\times$	lr = 10^{-4}, ep = 3, buf = 250 $\times$	lr = 10^{-4}, ep = 3, buf = 250 $\times$
		LwU	margin = 3, d. ep = 10, d. lr = 10^{-3}, align = off, mem = m01_s025_c08	margin = 3, d. ep = 10, d. lr = 10^{-3}, align = off, mem = m10_s100_c08	margin = 3, d. ep = 30, d. lr = 10^{-2}, align = on, mem = m01_s025_c08
	subclass	Fine-tune	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 5	lr = 10^{-4}, ep = 3
		Bad Teacher	$T = 4$, ep = 3, lr = 10^{-2}	$T = 4$, ep = 10, lr = 10^{-2} $\times$	$T = 4$, ep = 3, lr = 10^{-2}
		UNSIR	imp. ep = 2, rep. ep = 1, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 1, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-3}	imp. ep = 2, rep. ep = 1, imp. lr = 10^{-3}, rep. lr = 10^{-2}
		SSD	$\alpha = 1$, $\lambda = 1$	$\alpha = 1$, $\lambda = 10$	$\alpha = 5$, $\lambda = 0.5$
		SCRUB	max = 2, lr = 5×10^{-3}, ep = 10	max = 1, lr = 10^{-4}, ep = 5	max = 3, lr = 5×10^{-4}, ep = 10
		SalUn	sp = 0.1, lr = 10^{-4}, ep = 5 $\times$	sp = 0.1, lr = 10^{-4}, ep = 3 $\times$	sp = 0.1, lr = 10^{-4}, ep = 3 $\times$
		UniCLUN	lr = 10^{-2}, ep = 10, buf = 250 $\times$	lr = 10^{-3}, ep = 10, buf = 250 $\times$	lr = 10^{-4}, ep = 3, buf = 250 $\times$
		LwU	$q_f = 0.99$, margin = 1, d. ep = 30, d. lr = 10^{-3}, align = off, mem = m01_s025_c08	$q_f = 0.99$, margin = 3, d. ep = 10, d. lr = 10^{-3}, align = on, mem = m10_s100_c08	$q_f = 0.99$, margin = 3, d. ep = 10, d. lr = 10^{-3}, align = on, mem = m01_s025_c08
	instance	Fine-tune	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10
		Bad Teacher	$T = 4$, ep = 10, lr = 10^{-3}	$T = 4$, ep = 5, lr = 10^{-3}	$T = 4$, ep = 1, lr = 10^{-4}
		UNSIR	imp. ep = 1, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	$\alpha = 10$, $\lambda = 0.1$	$\alpha = 5$, $\lambda = 0.1$	$\alpha = 5$, $\lambda = 0.5$
		SCRUB	max = 3, lr = 10^{-3}, ep = 5	max = 2, lr = 5×10^{-3}, ep = 10	max = 1, lr = 5×10^{-3}, ep = 5

Backbone	Deletion	Method	Seed 42	Seed 43	Seed 44
		SalUn UniCLUN LwU	sp = 0.5, lr = 10^{-2}, ep = 10 × lr = 10^{-2}, ep = 10, buf = 250 × q_f = 0.9, margin = 3, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`	sp = 0.7, lr = 10^{-2}, ep = 10 × lr = 10^{-2}, ep = 10, buf = 250 × q_f = 0.9, margin = 3, d. ep = 10, d. lr = 10^{-2}, align = on, mem = `m10_s100_c08`	sp = 0.7, lr = 10^{-2}, ep = 10 × lr = 10^{-2}, ep = 10, buf = 250 × q_f = 0.9, margin = 3, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`
ResNet-18	class	Fine-tune Bad Teacher UNSIR SSD SCRUB SalUn UniCLUN LwU	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-2} imp. ep = 1, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 10, λ = 0.1 max = 3, lr = 10^{-3}, ep = 5 sp = 0.1, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 margin = 3, d. ep = 30, d. lr = 10^{-2}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-2} imp. ep = 2, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 100, λ = 0.1 max = 3, lr = 10^{-3}, ep = 5 sp = 0.7, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × margin = 3, d. ep = 30, d. lr = 10^{-2}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-2} imp. ep = 2, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2} α = 1, λ = 10 max = 3, lr = 10^{-3}, ep = 5 sp = 0.5, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × margin = 3, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`
	subclass	Fine-tune Bad Teacher UNSIR SSD SCRUB SalUn UniCLUN LwU	lr = 10^{-4}, ep = 3 T = 4, ep = 5, lr = 10^{-2} × imp. ep = 2, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-2} × α = 1, λ = 5 max = 3, lr = 5×10^{-3}, ep = 5 sp = 0.3, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × q_f = 0.99, margin = 3, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`	lr = 10^{-4}, ep = 1 T = 4, ep = 10, lr = 10^{-2} × imp. ep = 2, rep. ep = 2, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 1, λ = 1 max = 3, lr = 5×10^{-4}, ep = 5 sp = 0.1, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × q_f = 0.99, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`	lr = 10^{-4}, ep = 5 T = 4, ep = 10, lr = 10^{-2} × imp. ep = 2, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2} α = 1, λ = 0.1 max = 2, lr = 10^{-4}, ep = 10 sp = 0.3, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × q_f = 0.99, margin = 1, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`
	instance	Fine-tune Bad Teacher UNSIR SSD SCRUB SalUn UniCLUN LwU	lr = 10^{-4}, ep = 10 T = 4, ep = 1, lr = 10^{-4} imp. ep = 2, rep. ep = 1, imp. lr = 10^{-2}, rep. lr = 10^{-2} α = 1, λ = 10 max = 3, lr = 5×10^{-3}, ep = 5 sp = 0.3, lr = 10^{-2}, ep = 10 × lr = 10^{-2}, ep = 10, buf = 250 × q_f = 0.99, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 5 T = 2, ep = 1, lr = 10^{-4} imp. ep = 1, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2} α = 5, λ = 0.1 max = 2, lr = 5×10^{-3}, ep = 5 sp = 0.3, lr = 10^{-2}, ep = 10 × lr = 10^{-2}, ep = 10, buf = 250 × q_f = 0.99, margin = 1, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 10 T = 1, ep = 3, lr = 10^{-4} imp. ep = 2, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 5, λ = 0.1 max = 3, lr = 5×10^{-3}, ep = 5 sp = 0.9, lr = 10^{-2}, ep = 10 × lr = 10^{-2}, ep = 10, buf = 250 × q_f = 0.99, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`

Table 34: Selected hyperparameters on CIFAR-100.

Backbone	Deletion	Method	Seed 42	Seed 43	Seed 44
CNN	class	Fine-tune Bad Teacher UNSIR SSD SCRUB SalUn UniCLUN LwU	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-3} imp. ep = 1, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2} α = 100, λ = 5 max = 3, lr = 5×10^{-4}, ep = 5 sp = 0.1, lr = 10^{-4}, ep = 5 × lr = 10^{-4}, ep = 3, buf = 250 × margin = 3, d. ep = 10, d. lr = 10^{-2}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 10 T = 4, ep = 5, lr = 10^{-2} × imp. ep = 1, rep. ep = 2, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 50, λ = 0.5 max = 1, lr = 5×10^{-4}, ep = 5 sp = 0.3, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 5, buf = 250 × margin = 3, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-2} imp. ep = 1, rep. ep = 2, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 1, λ = 5 max = 3, lr = 5×10^{-3}, ep = 5 sp = 0.9, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × margin = 3, d. ep = 10, d. lr = 10^{-2}, align = off, mem = `m01_s025_c08`
	subclass	Fine-tune Bad Teacher UNSIR SSD SCRUB SalUn UniCLUN LwU	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-3} imp. ep = 1, rep. ep = 3, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 50, λ = 0.1 max = 2, lr = 5×10^{-3}, ep = 10 sp = 0.1, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 5, buf = 250 × q_f = 0.99, margin = 1, d. ep = 10, d. lr = 10^{-3}, align = off, mem = `m01_s025_c08`	lr = 10^{-4}, ep = 5 T = 2, ep = 10, lr = 10^{-2} imp. ep = 1, rep. ep = 2, imp. lr = 10^{-2}, rep. lr = 10^{-2} α = 10, λ = 5 max = 1, lr = 5×10^{-4}, ep = 10 sp = 0.9, lr = 10^{-4}, ep = 3 × lr = 10^{-2}, ep = 5, buf = 250 × q_f = 0.99, margin = 1, d. ep = 30, d. lr = 10^{-3}, align = off, mem = `m10_s100_c08`	lr = 10^{-4}, ep = 3 T = 4, ep = 10, lr = 10^{-2} × imp. ep = 2, rep. ep = 1, imp. lr = 10^{-3}, rep. lr = 10^{-2} α = 1, λ = 10 max = 2, lr = 5×10^{-3}, ep = 5 sp = 0.3, lr = 10^{-4}, ep = 3 × lr = 10^{-4}, ep = 3, buf = 250 × q_f = 0.99, margin = 1, d. ep = 10, d. lr = 10^{-2}, align = off, mem = `m01_s025_c08`
	instance	Fine-tune Bad Teacher UNSIR	lr = 10^{-4}, ep = 10 T = 4, ep = 10, lr = 10^{-3} imp. ep = 2, rep. ep = 3, imp. lr = 10^{-2}, rep. lr = 10^{-2}	lr = 10^{-4}, ep = 10 T = 1, ep = 1, lr = 10^{-4} imp. ep = 2, rep. ep = 1, imp. lr = 10^{-3}, rep. lr = 10^{-2}	lr = 10^{-4}, ep = 5 T = 2, ep = 1, lr = 10^{-4} imp. ep = 2, rep. ep = 1, imp. lr = 10^{-3}, rep. lr = 10^{-2}

Backbone	Deletion	Method	Seed 42	Seed 43	Seed 44
		SSD	$\alpha = 5$, $\lambda = 0.1$	$\alpha = 5$, $\lambda = 0.5$	$\alpha = 5$, $\lambda = 0.1$
		SCRUB	max = 2, lr = 5×10^{-3}, ep = 10	max = 2, lr = 5×10^{-3}, ep = 10	max = 1, lr = 5×10^{-3}, ep = 5
		SalUn	sp = 0.9, lr = 10^{-4}, ep = 3 ×	sp = 0.3, lr = 10^{-2}, ep = 10 ×	sp = 0.5, lr = 10^{-2}, ep = 10 ×
		UniCLUN	lr = 10^{-2}, ep = 10, buf = 250 ×	lr = 10^{-2}, ep = 10, buf = 250 ×	lr = 10^{-2}, ep = 10, buf = 250 ×
		LwU	$q_f = 0.9$, margin = 3, d. ep = 10,	$q_f = 0.9$, margin = 3, d. ep = 10,	$q_f = 0.9$, margin = 3, d. ep = 10,
			d. lr = 10^{-2}, align = on,	d. lr = 10^{-2}, align = off,	d. lr = 10^{-2}, align = off,
			mem = `m01_s025_c08`	mem = `m01_s025_c08`	mem = `m10_s100_c08`
ResNet-18	class	Fine-tune	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10	lr = 10^{-4}, ep = 10
		Bad Teacher	$T = 4$, ep = 10, lr = 10^{-2}	$T = 4$, ep = 10, lr = 10^{-2}	$T = 4$, ep = 3, lr = 10^{-2}
		UNSIR	imp. ep = 1, rep. ep = 3,	imp. ep = 1, rep. ep = 2,	imp. ep = 1, rep. ep = 3,
			imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. lr = 10^{-2}, rep. lr = 10^{-2}
		SSD	$\alpha = 50$, $\lambda = 0.1$	$\alpha = 10$, $\lambda = 5$	$\alpha = 50$, $\lambda = 5$
		SCRUB	max = 3, lr = 5×10^{-3}, ep = 5	max = 3, lr = 5×10^{-3}, ep = 5	max = 3, lr = 5×10^{-3}, ep = 5
		SalUn	sp = 0.3, lr = 10^{-4}, ep = 3 ×	sp = 0.9, lr = 10^{-4}, ep = 3 ×	sp = 0.9, lr = 10^{-4}, ep = 3 ×
		UniCLUN	lr = 10^{-4}, ep = 3, buf = 250 ×	lr = 10^{-4}, ep = 3, buf = 250 ×	lr = 10^{-4}, ep = 3, buf = 250 ×
		LwU	margin = 3, d. ep = 30,	margin = 3, d. ep = 30,	margin = 3, d. ep = 30,
			d. lr = 10^{-3}, align = off,	d. lr = 10^{-2}, align = off,	d. lr = 10^{-2}, align = off,
			mem = `m01_s025_c08`	mem = `m01_s025_c08`	mem = `m01_s025_c08`
	subclass	Fine-tune	lr = 10^{-4}, ep = 1	lr = 10^{-4}, ep = 3	lr = 10^{-4}, ep = 10
		Bad Teacher	$T = 4$, ep = 10, lr = 10^{-2}	$T = 4$, ep = 5, lr = 10^{-2}	$T = 4$, ep = 10, lr = 10^{-2}
		UNSIR	imp. ep = 2, rep. ep = 2,	imp. ep = 1, rep. ep = 1,	imp. ep = 1, rep. ep = 3,
			imp. lr = 10^{-3}, rep. lr = 10^{-2}	imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. lr = 10^{-3}, rep. lr = 10^{-2}
		SSD	$\alpha = 50$, $\lambda = 1$	$\alpha = 5$, $\lambda = 0.1$	$\alpha = 50$, $\lambda = 5$
		SCRUB	max = 2, lr = 5×10^{-3}, ep = 10	max = 3, lr = 5×10^{-3}, ep = 10	max = 3, lr = 10^{-3}, ep = 10
		SalUn	sp = 0.9, lr = 10^{-4}, ep = 3 ×	sp = 0.3, lr = 10^{-4}, ep = 3 ×	sp = 0.1, lr = 10^{-4}, ep = 3 ×
		UniCLUN	lr = 10^{-4}, ep = 5, buf = 250 ×	lr = 10^{-4}, ep = 3, buf = 250 ×	lr = 10^{-4}, ep = 3, buf = 250 ×
		LwU	$q_f = 0.99$, margin = 1, d. ep = 10,	$q_f = 0.99$, margin = 3, d. ep = 30,	$q_f = 0.99$, margin = 1, d. ep = 10,
			d. lr = 10^{-3}, align = off,	d. lr = 10^{-2}, align = off,	d. lr = 10^{-3}, align = off,
			mem = `m01_s025_c08`	mem = `m10_s100_c08`	mem = `m01_s025_c08`
	instance	Fine-tune	lr = 10^{-4}, ep = 3	lr = 10^{-4}, ep = 1	lr = 10^{-4}, ep = 10
		Bad Teacher	$T = 2$, ep = 1, lr = 10^{-4}	$T = 1$, ep = 1, lr = 10^{-4}	$T = 4$, ep = 1, lr = 10^{-4}
		UNSIR	imp. ep = 2, rep. ep = 3,	imp. ep = 1, rep. ep = 3,	imp. ep = 1, rep. ep = 1,
			imp. lr = 10^{-2}, rep. lr = 10^{-2}	imp. lr = 10^{-3}, rep. lr = 10^{-3}	imp. lr = 10^{-3}, rep. lr = 10^{-2}
		SSD	$\alpha = 10$, $\lambda = 0.5$	$\alpha = 5$, $\lambda = 1$	$\alpha = 5$, $\lambda = 0.1$
		SCRUB	max = 3, lr = 5×10^{-3}, ep = 5	max = 2, lr = 5×10^{-3}, ep = 5	max = 3, lr = 5×10^{-3}, ep = 5
		SalUn	sp = 0.9, lr = 10^{-2}, ep = 10 ×	sp = 0.9, lr = 10^{-2}, ep = 10 ×	sp = 0.7, lr = 10^{-2}, ep = 10 ×
		UniCLUN	lr = 10^{-2}, ep = 10, buf = 250 ×	lr = 10^{-2}, ep = 10, buf = 250 ×	lr = 10^{-2}, ep = 10, buf = 250 ×
		LwU	$q_f = 0.99$, margin = 1, d. ep = 10,	$q_f = 0.9$, margin = 3, d. ep = 30,	$q_f = 0.99$, margin = 1, d. ep = 10,
			d. lr = 10^{-3}, align = off,	d. lr = 10^{-2}, align = off,	d. lr = 10^{-3}, align = on,
			mem = `m01_s025_c08`	mem = `m01_s025_c08`	mem = `m01_s025_c08`

Table 30: Training recipe of the source (Original) and Retrain models, which are not tuned. Amnesiac replays the per-batch update ledger recorded during source training (class deletion only).

	CIFAR-10	RTI	CIFAR-100
Max. epochs	100	150	150
Early-stopping patience	10	15	15
Min. epochs	20	30	30
Batch size	64	64	64
Learning rate (Adam)	10^{-3}	10^{-3}	10^{-3}
LR schedule	reduce on plateau	reduce on plateau	reduce on plateau
D_r fraction available to unlearning	0.1	0.1	0.1
Relearn probe epochs	5	5	5
Relearn probe lr	10^{-4}	10^{-4}	10^{-4}
CNN hidden width	64	256	512

Table 31: Most frequently selected value of each tuned hyperparameter over the 18 runs per dataset (2 backbones $\times$ 3 deletion types $\times$ 3 seeds), with its count. Tied values are separated by /.

Method	Hyperparameter	CIFAR-10	RTI	CIFAR-100
Fine-tune	lr	10^{-3} (14)	10^{-4} (18)	10^{-4} (18)
	epochs	10 (16)	10 (12)	10 (11)
Bad Teacher	T	4 (12)	4 (15)	4 (13)
	epochs	10 (9)	10 (9)	10 (10)
	lr	10^{-2} (14)	10^{-2} (12)	10^{-2} (10)
UNSIR	impair epochs	2 (10)	2 (11)	1 (12)
	repair epochs	2 (10)	3 (11)	3 (8)
	impair lr	10^{-2} (14)	10^{-2} (10)	10^{-3} (10)
	repair lr	10^{-2} (17)	10^{-2} (17)	10^{-2} (17)
SSD	α	1 (9)	1 (7)	50 / 5 (6)
	λ	5 (7)	0.1 (9)	5 / 0.1 (6)
SCRUB	max-step epochs	3 (10)	3 (12)	3 (9)
	lr	5×10^{-3} (9)	5×10^{-3} (7)	5×10^{-3} (14)
	epochs	5 (11)	5 (14)	5 (11)
SalUn	sparsity	0.1 (7)	0.1 (6)	0.9 (8)
	lr	10^{-4} (9)	10^{-4} (12)	10^{-4} (13)
	epochs	10 (9)	3 (10)	3 (12)
UniCLUN	lr	10^{-4} (11)	10^{-4} (10)	10^{-4} (11)
	epochs	3 (7)	3 (9)	3 (8)
	buffer	250 (18)	250 (18)	250 (18)
LwU	q_f	0.99 (14)	0.99 (15)	0.99 (14)
	margin	3 (10)	3 (12)	3 (11)
	distill. epochs	30 (10)	10 / 30 (9)	10 (11)
	distill. lr	10^{-3} (11)	10^{-3} (14)	10^{-2} (10)
	weight align.	off (10)	off (14)	off (16)
	memory	`m10_s100_c08` (12)	`m01_s025_c08` (11)	`m01_s025_c08` (15)

Table 35: Hyperparameter search space and selected configurations on Seq-CIFAR-100.

		TIL Scenario		CIL Scenario	
Method	Search space	Trial	Selected configuration	Trial	Selected configuration
EWC	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$ `ewc_gamma` $\sim$ Uniform$(0.9, 1.0)$ `ewc_lambda` $\sim$ LogUniform$(1, 10^4)$	4	`lr=0.000706`, `ewc_gamma=0.912`, `ewc_lambda=2029.6`	23	`lr=0.00213`, `ewc_gamma=0.970`, `ewc_lambda=1.749`
SI	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$ `si_c` $\sim$ LogUniform$(10^{-2}, 10)$ `si_xi` $\sim$ LogUniform$(10^{-4}, 10^{-2})$	2	`lr=0.000131`, `si_c=0.333`, `si_xi=0.000119`	0	`lr=0.000444`, `si_c=0.0284`, `si_xi=0.00200`
LwF	Candidate list (30 configs) `lr` $\in$ [1.1e-4, 8.4e-3] `lwf_lambda` $\in [0.14, 9.13]$ `temperature` $\in \{1, 2, 3, 4\}$	13	`lr=0.00173`, `lwf_lambda=0.984`, `temperature=4`	27	`lr=0.000111`, `lwf_lambda=0.838`, `temperature=2`
WSN	Candidate list `lr` $\in$ [1.2e-4, 7.9e-3] `wsn_density` $\in \{0.1, 0.3, 0.5, 0.7\}$	15	`lr=0.00132`, `wsn_density=0.3`	–	–
PEC	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$	–	–	3	`lr=0.000140`
SpaceNet	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$ `density_factor` $\in \{0.5, 1.0, 1.5\}$ `rewire_fraction` $\in \{0.1, 0.2, 0.3\}$	8	`lr=0.00147`, `density_factor=0.5`, `rewire_fraction=0.3`	11	`lr=0.000689`, `density_factor=1.5`, `rewire_fraction=0.1`
NISPA	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$ `nispa_prune_perc` $\sim$ Uniform$(70, 95)$ `nispa_recovery_perc` $\sim$ Uniform$(1, 5)$	10	`lr=0.00190`, `nispa_prune_perc=79.3`, `nispa_recovery_perc=3.19`	6	`lr=0.00143`, `nispa_prune_perc=79.9`, `nispa_recovery_perc=4.91`
SNV	Candidate list (23 of the 30 draws of `lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$, keeping `lr` $\geq$ 3e-4) `truncation` $\in \{0.05, 0.1, 0.2\}$	14	`lr=0.00130`, `truncation=0.05`, `rot_aux=0`	14	`lr=0.00130`, `truncation=0.05`, `rot_aux=1.0`, `routing=rot+energy`
UniCLUN	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$ `alpha1, alpha2, alpha3` $\sim$ LogUniform$(0.1, 10)$	8	`lr=0.00125`, `alpha1=0.134`, `alpha2=0.132`, `alpha3=0.258`	10	`lr=0.000806`, `alpha1=0.398`, `alpha2=3.879`, `alpha3=2.500`
LwU (ours)	`lr` $\sim$ LogUniform$(10^{-4}, 10^{-2})$ `ce_weight` $\sim$ LogUniform$(0.1, 10)$ `memory_conf` $\in \{0.0, 0.8\}$ `memory_lr` $\in \{0.01, 0.05, 0.1\}$ `memory_scale` $\in \{0.25, 0.5, 1.0\}$ `retain_quant` $\in \{0.5, 0.7, 0.9, 0.95, 0.99\}$ `sd_temp` $\in \{1, 2, 3, 4\}$	15	`lr=0.000625`, `ce_weight=0.215`, `memory_conf=0.8`, `memory_lr=0.05`, `memory_scale=1`, `retain_quant=0.95`, `sd_temp=1`	20	`lr=0.000819`, `ce_weight=0.625`, `memory_conf=0.8`, `memory_lr=0.05`, `memory_scale=0.25`, `retain_quant=0.99`, `sd_temp=4`

Table 36: Hyperparameter search space and selected configurations on Seq-Tiny-ImageNet.

		TIL Scenario		CIL Scenario	
Method	Search space	Trial	Selected configuration	Trial	Selected configuration
EWC	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `ewc_lambda` $\sim \text{LogUniform}(1, 10^{4})$ `ewc_gamma` $\sim \text{Uniform}(0.9, 1.0)$	30	`lr=0.00196905`, `ewc_lambda=9383.91`, `ewc_gamma=0.982192`	30	`lr=0.000444289`, `ewc_lambda=4.01233`, `ewc_gamma=0.965093`
SI	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `si_c` $\sim \text{LogUniform}(0.01, 10)$ `si_xi` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	30	`lr=0.00553151`, `si_c=0.0174479`, `si_xi=0.000791361`	30	`lr=0.000444289`, `si_c=0.0283496`, `si_xi=0.00200387`
LwF	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `lwf_lambda` $\sim \text{LogUniform}(0.1, 10)$ `temperature` $\in \{1, 2, 3, 4\}$	30	`lr=0.00112303`, `lwf_lambda=5.62698`, `temperature=4`	30	`lr=0.00011095`, `lwf_lambda=0.838283`, `temperature=2`
WSN	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `wsn_density` $\in \{0.1, 0.3, 0.5, 0.7\}$	$15^{\dagger}$	`lr=0.0013198195`, `wsn_density=0.3`	–	–
PEC	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	–	–	$3^{\dagger}$	`lr=0.000140`
SpaceNet	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `density_factor` $\in \{0.5, 1.0, 1.5\}$ `rewire_fraction` $\in \{0.1, 0.2, 0.3\}$	$8^{\dagger}$	`lr=0.0014655269`, `density_factor=0.5`, `rewire_fraction=0.3`	$11^{\dagger}$	`lr=0.0006890934`, `density_factor=1.5`, `rewire_fraction=0.1`
NISPA	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `nispa_prune_perc` $\sim \text{Uniform}(70, 95)$ `nispa_recovery_perc` $\sim \text{Uniform}(1, 5)$	$10^{\dagger}$	`lr=0.0018959503`, `nispa_prune_perc=79.310`, `nispa_recovery_perc=3.191`	$6^{\dagger}$	`lr=0.0014262836`, `nispa_prune_perc=79.917`, `nispa_recovery_perc=4.905`
SNV	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `truncation` $\in \{0.05, 0.1, 0.2\}$	$14^{\dagger}$	`lr=0.0012981647`, `truncation=0.05`	$14^{\dagger}$	`lr=0.0012981647`, `truncation=0.05`
UniCLUN	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `alpha1, alpha2, alpha3` $\sim$ LogUniform$(0.1, 10)$	$8^{\dagger}$	`lr=0.0012459165`, `alpha1=0.1335`, `alpha2=0.1316`, `alpha3=0.2582`	$10^{\dagger}$	`lr=0.0008060626`, `alpha1=0.3977`, `alpha2=3.8793`, `alpha3=2.5003`
LwU (ours)	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `retain_quantile` $\in \{0.5, 0.7, 0.9, 0.95, 0.99\}$ `sd_temperature` $\in \{1, 2, 3, 4\}$ `ce_weight` $\sim \text{LogUniform}(0.1, 10)$ `memory_lr` $\in \{0.01, 0.05, 0.1\}$ `memory_scale` $\in \{0.25, 0.5, 1.0\}$ `memory_confidence` $\in \{0.0, 0.8\}$	30	`lr=0.000302919`, `retain_quantile=0.99`, `sd_temperature=4`, `ce_weight=0.131287`, `memory_lr=0.1`, `memory_scale=0.25`, `memory_confidence=0`	30	`lr=0.000105712`, `retain_quantile=0.99`, `sd_temperature=2`, `ce_weight=0.335346`, `memory_lr=0.01`, `memory_scale=0.25`, `memory_confidence=0.8`
SGD (lower bound)	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	30	`lr=0.00785992`	30	`lr=0.000413965`

Table 37: Hyperparameter search space and selected configurations on RTI.

		TIL Scenario		CIL Scenario	
Method	Search space	Trial	Selected configuration	Trial	Selected configuration
Joint (upper bound)	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	7	`lr=0.00103`	9	`lr=0.000737`
EWC	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `ewc_gamma` $\sim \text{Uniform}(0.9, 1.0)$ `ewc_lambda` $\sim \text{LogUniform}(1, 10^{4})$	4	`lr=0.000706`, `ewc_gamma=0.912`, `ewc_lambda=2030`	23	`lr=0.00213`, `ewc_gamma=0.970`, `ewc_lambda=1.75`
SI	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `si_c` $\sim \text{LogUniform}(10^{-2}, 10)$ `si_xi` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	2	`lr=0.000131`, `si_c=0.333`, `si_xi=0.000119`	27	`lr=0.000171`, `si_c=0.015`, `si_xi=0.00344`
LwF	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `lwf_lambda` $\sim \text{LogUniform}(0.1, 10)$ `temperature` $\in \{1, 2, 3, 4\}$	11	`lr=0.00146`, `lwf_lambda=1.90`, `temperature=3`	23	`lr=0.00144`, `lwf_lambda=0.817`, `temperature=1`
WSN	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `wsn_density` $\in \{0.1, 0.3, 0.5, 0.7\}$	25	`lr=0.00112`, `wsn_density=0.5`	–	–
PEC	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	–	–	11	`lr=0.000152`
SpaceNet	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `density_factor` $\in \{0.5, 1.0, 1.5\}$ `rewire_fraction` $\in \{0.1, 0.2, 0.3\}$	8	`lr=0.00147`, `density_factor=0.5`, `rewire_fraction=0.3`	14	`lr=0.00190`, `density_factor=1.0`, `rewire_fraction=0.1`

NISPA	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$
	`nispa_prune_perc` $\sim \text{Uniform}(70, 95)$
	`nispa_recovery_perc` $\sim \text{Uniform}(1, 5)$
		3	`lr=0.000737`,
			`nispa_prune_perc=71.7`,
			`nispa_recovery_perc`
			`=1.36`
				3	`lr=0.000737`,
					`nispa_prune_perc=71.7`,
					`nispa_recovery_perc`
					`=1.36`

UniCLUN	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `alpha1, alpha2, alpha3` $\sim \text{LogUniform}(0.1, 10)$ each	26	`lr=0.0015074642`, `alpha1=0.335346084`, `alpha2=0.101903056`, `alpha3=0.688482653`	27	`lr=0.000548`, `alpha1=1.36`, `alpha2=8.06`, `alpha3=2.40`
LwU (ours)	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$ `tau_r` $\in \{0, 10^{-8}, 10^{-7}, 3\times10^{-7}, 10^{-6}, 3\times10^{-6}, 10^{-5}, 3\times10^{-5}, 10^{-4}, 3\times10^{-4}, 10^{-3}\}$ `momentum_decay` $\in \{0.9, 0.95, 0.99\}$ `adaptation_rate` $\in \{0.001, 0.01, 0.05\}$	12	`lr=0.00125`, `tau_r=1e-8`, `momentum_decay=0.99`, `adaptation_rate=0.001`	18	`lr=0.000790`, `tau_r=3e-4`, `momentum_decay=0.9`, `adaptation_rate=0.001`
SGD (lower bound)	`lr` $\sim \text{LogUniform}(10^{-4}, 10^{-2})$	10	`lr=0.000138`	9	`lr=0.000737`