ONE-WRITE ASSOCIATIVE SKETCHING FOR PARTITION-INVARIANT LIFELONG CROSS-MODAL HASHING

Anonymous authors
Paper under double-blind review

ABSTRACT

Lifelong cross-modal hashing updates compact image-text indexes as paired data arrive. Task-boundary updates can nevertheless change the items returned for the same query when an ordered stream is segmented differently, even at similar retrieval accuracy. The challenge is to preserve both the current query head and codes written by earlier heads. We propose Partition-Invariant Lifelong Cross-Modal Hashing (PILCH), which couples prefix learning with one-write indexing. With frozen features, permanent Dice-derived semantic targets and additive moments recover the full-prefix code-and-relation objective. We prove that its unique ridge solution and shared sample-count clock yield identical query heads, historical codes and rankings across taskifications. Perturbation bounds connect approximate moments to retrieval changes. We evaluate cross-partition top-k agreement using the top-k retrieval reproducibility rate, the mean Jaccard overlap of returned sets, alongside relevance. Across four datasets and three paired seeds, the analytic core returns identical sets and ordered lists at $k = 10, 20, 50$. Clocked discrete extensions retain this agreement, and fixed-gallery query calibration achieves the highest mean average precision on all four datasets, reaching 92.68% on MIRFlickr-25K at 64 bits when averaged over both retrieval directions.

1 INTRODUCTION

Cross-modal hashing retrieves images from text and vice versa through compact binary codes (Jiang & Li, 2017; Cao et al., 2018; Hu et al., 2023). Growing collections require learning from new observations while keeping stored codes usable. Online and lifelong hashing address this need through incremental updates (Xie et al., 2016; Wang et al., 2020b; Zhan et al., 2022; Xu et al., 2024).

Task boundaries often trigger model updates and database writes. Consider one ordered stream $z_1 \to z_2 \to \cdots \to z_N$ of N paired observations, with z_i the ith arrival, processed under two segmentations. A boundary can change the head that encodes an arriving item. In Figure 1, final heads agree, but different historical codes lead the same query to different items.

Zhuang et al. (2022) recover joint analytic predictors from accumulated statistics. Backward-compatible learning aligns new representations with existing indexes (Shen et al., 2020; Jang & Lim, 2025). These approaches control the current predictor and cross-version compatibility, respectively. A one-write gallery also retains outputs from earlier heads: a newly recovered predictor does not regenerate those historical codes.

We address this temporal dependence with PILCH. Each arrival contributes feature and semantic-target moments in a fixed coordinate system. Adding these moments recovers the objective for all observations so far, including relations between segments. Ridge regularization gives a unique head; updates and writes occur at shared sample counts. Changing external boundaries thus preserves the head that encodes each arriving item.

Our contributions are threefold. We derive a mergeable formulation of code-and-relation learning for one-write indexing. We establish partition invariance of heads, historical codes and rankings, and bound retrieval changes under approximate computation. We transfer the write schedule to discrete

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

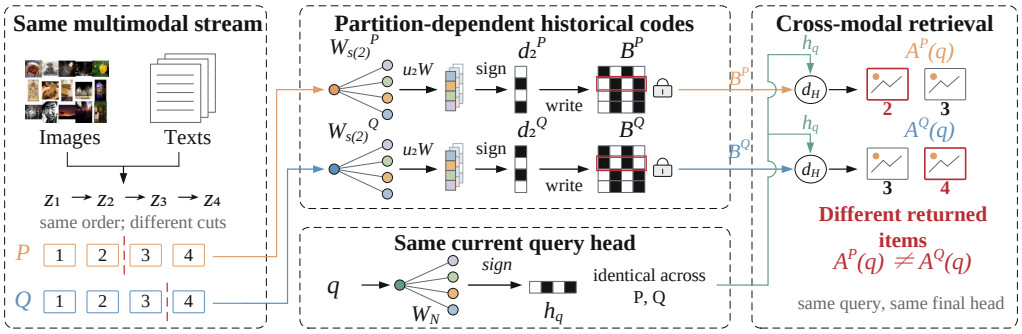


Figure 1: Boundary-induced retrieval drift. Under partitions P, Q, the heads $W_{s(2)}$ used when item 2 is written produce different codes d_2, galleries B and top-two sets $A(q)$. Final heads W_N and query codes h_q agree. Red: item 2; white/black: $+1/-1$. Hamming ties use item ID.

learners. Across four datasets and three seeds, fixed-gallery calibration of this extension achieves the highest direction-averaged mean average precision (mAP) among the evaluated methods at 64 bits.

2 Related Work

Semantic hash learning. Semantic relations guide cross-modal learning in binary space. DCMH jointly fits features and codes to labels (Jiang & Li, 2017); CMHH concentrates relevant codes within small Hamming balls through a focal loss (Cao et al., 2018). Without labels, UCCH uses paired contrastive learning (Hu et al., 2023), while UCMFH exploits CLIP features and multimodal fusion (Xia et al., 2023). To enrich pairwise relations, HINT builds hierarchical trees and mixes community-derived proxies (Xiao et al., 2026); DGHDGH propagates batch-wide relations to generate informative negatives (Cheng et al., 2026). These structures strengthen semantic neighborhoods, whose coordinates can change during subsequent learning.

Online cross-modal hashing. Online updates carry learned information into each incoming chunk. OCMH retains latent codes while adapting a transform (Xie et al., 2016); OCMFH maintains a collective factorization (Wang et al., 2020a), and LEMON adds semantic label embeddings (Wang et al., 2020b). Discrete optimization addresses quantization error: DOCH fits new-old similarities directly in Hamming space (Zhan et al., 2022); DOCMH also models continuous labels and training-retrieval modality deviation (Kang et al., 2024). For partially labeled streams, OMGH uses anchor-guided manifold structure (Liu et al., 2023). Old-new relations inform these updates, whose timing and data grouping remain tied to chunk boundaries.

Lifelong learning and index compatibility. Lifelong hashing maintains indexed semantics as categories and models change. CT-CMH penalizes code conflicts between old and new classes (Zheng et al., 2023); DLCH fixes historical codes (Xu et al., 2024), and MCDLCH adds multiscale consistency and modality alignment (Xu et al., 2025). OH-ELS supports label expansion through code-anchor replay and classifier consistency (Fan et al., 2026); SILCH balances gradients across unequal tasks (Yang et al., 2026). Model upgrades pose a related problem: backward-compatible training aligns new and old representations (Shen et al., 2020), extended to vision-language retrieval by XBT (Jang & Lim, 2025). Compatibility constrains comparisons across model versions; the timing of historical code generation remains unspecified.

Boundary-free and analytic learning. Streaming learning separates task identities from observation accumulation. Task-free methods dispense with supplied task IDs (Aljundi et al., 2019), whereas temporal taskification examines data grouping (Filat et al., 2026). Mergeable summaries and matrix sketches support blockwise accumulation (Batagelj, 2023; Ghashami et al., 2016; Shi & Phillips, 2021). ACIL recovers joint classifiers through recursive updates (Zhuang et al., 2022); RanPAC combines pretrained features, random projections and decorrelated prototypes (McDonnell et al., 2023). HESTIA couples order-invariant linearized adaptation with density-guided adapter

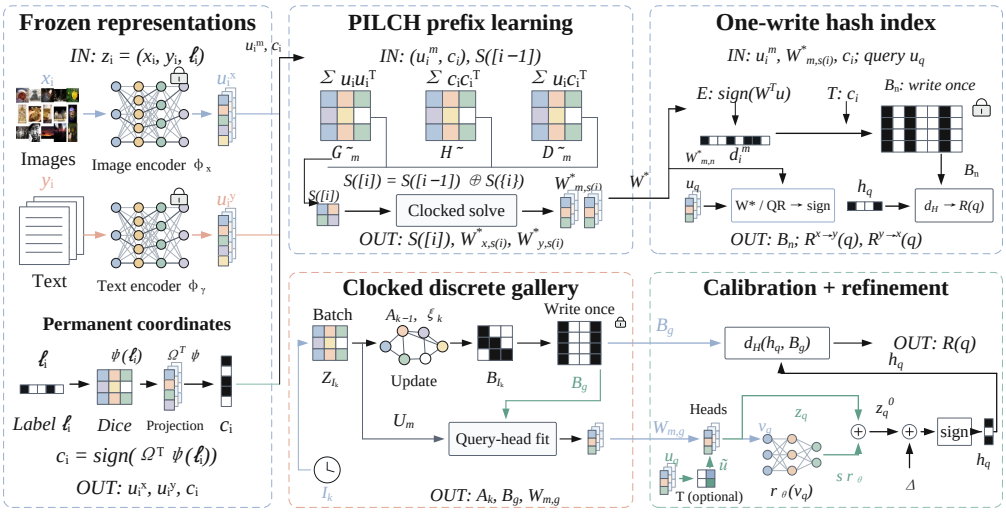


Figure 2: PILCH and its fixed-clock extension. Permanent targets and additive moments produce analytic heads for encoding new items and queries. The lower branch updates a discrete learner at fixed sample counts; query refinement leaves the corresponding gallery unchanged.

routing (Le et al., 2026). Accumulated predictors summarize the current prefix, while an immutable gallery retains outputs from intermediate predictors throughout learning.

3 Methodology

3.1 Problem setup

PILCH maintains modality-specific heads and a one-write gallery. For a positive integer a, write $[a] = \{1, \dots, a\}$; N is the stream size. Arrival $i \in [N]$ is $z_i = (x_i, y_i, \ell_i)$: an image $x_i \in \mathcal{X}$, text $y_i \in \mathcal{Y}$ and a revealed nonzero multi-hot label $\ell_i \in \{0,1\}^{d_\ell} \setminus \{0\}$ over d_ℓ categories. Here $\mathcal{X}, \mathcal{Y}$ are the image and text domains. The prefix $\mathcal{Z}_{1:n} = (z_1, \dots, z_n)$ contains the first $n \in [N]$ observations. A bootstrap of $1 \le n_0 < N$ items is indexed after the initial head solve.

A taskification divides the remaining stream into $T \in \mathbb{N}_{>0}$ contiguous segments with sample-count endpoints $\mathcal{P} = (n_0 = \tau_0 < \tau_1 < \cdots < \tau_T = N)$. Here τ_t is the last sample count of task $t \in [T]$. *Taskification-equivalent* runs differ only in boundaries, sharing ordered examples, label-disclosure times, preprocessing, hyperparameters, seeds, code length, gallery item IDs, queries and sample-count checkpoints $\mathcal{C} \subseteq \{n_0, \dots, N\}$. The solve clock $\mathcal{C} \cup \{n_0\} \subseteq \mathcal{C}_s \subseteq \{n_0, \dots, N\}$ specifies head-update counts.

With code length $b \in \mathbb{N}_{>0}$, a query q at prefix n in direction $g \in \{x \to y, y \to x\}$ returns the ranked item-ID list $R^g_{\mathcal{P},n}(q)$, ordered by Hamming distance d_H, the number of unequal bits, with item-ID ties.

In Figure 2, the analytic core fits heads from additive moments; the discrete extension clocks an existing binary-code learner. Both write each item once and allow query refinement against a fixed gallery.

3.2 Permanent semantic teacher coordinates

Both modalities fit label-based targets fixed throughout the stream. For $i, j \in [N]$, define the Dice kernel with Euclidean norm $\|\cdot\|_2$:

$$k_{\mathrm{D}}(\ell_i, \ell_j) = \frac{2\ell_i^\top \ell_j}{\|\ell_i\|_2^2 + \|\ell_j\|_2^2}.$$

Factorization over the d_ℓ possible nonzero label counts gives a fixed map $\psi(\ell) \in \mathbb{R}^{d_\ell^2}$ satisfying $\psi(\ell_i)^\top \psi(\ell_j) = k_{\mathrm{D}}(\ell_i, \ell_j)$ and $\|\psi(\ell_i)\|_2 = 1$. One stream-wide draw $\Omega = [\omega_1, \ldots, \omega_b] \in \mathbb{R}^{d_\ell^2 \times b}$ with independent standard Gaussian columns ω_r, $r \in [b]$, gives the permanent target code

$$c_i = \operatorname{sign}(\Omega^\top \psi(\ell_i)) \in \{-1, +1\}^b, \qquad \operatorname{sign}(0) = +1. \tag{1}$$

Here sign acts elementwise and c_{ir} is bit r of c_i. Random hyperplane hashing gives, over Ω, $\Pr[c_{ir} \neq c_{jr}] = \arccos(\psi(\ell_i)^\top \psi(\ell_j))/\pi$ (Charikar, 2002). Thus $\mathbb{E}[c_i^\top c_j / b] = 1 - \frac{2}{\pi} \arccos(k_{\mathrm{D}}(\ell_i, \ell_j))$, an angular monotone transform of Dice similarity.

3.3 Associative prefix learning

We fit each modality to the permanent codes and their relations. For modality $m \in \{x, y\}$, let $z_i^x = x_i$ and $z_i^y = y_i$. The frozen map ϕ_m gives r_m-dimensional column features $u_i^m = \phi_m(z_i^m)$ with norm bounded by R_m. A fixed projection can reduce dimension; r_m includes a constant bias feature. An optional adapter is bootstrap-trained and then frozen within ϕ_m.

For $E \subseteq [N]$, the state contains its count $n_E = |E|$ and moments:

$$\mathcal{S}(E) = \big(n_E, \widetilde{G}_x(E), \widetilde{G}_y(E), \widetilde{H}(E), \widetilde{D}_x(E), \widetilde{D}_y(E)\big), \tag{2}$$
$$\widetilde{G}_m(E) = \sum_{i \in E} u_i^m u_i^{m\top}, \quad \widetilde{H}(E) = \sum_{i \in E} c_i c_i^\top, \quad \widetilde{D}_m(E) = \sum_{i \in E} u_i^m c_i^\top.$$

For disjoint index sets $A, B \subseteq [N]$, componentwise addition satisfies

$$\mathcal{S}(A \cup B) = \mathcal{S}(A) \oplus \mathcal{S}(B). \tag{3}$$

$\widetilde{G}_m \in \mathbb{R}^{r_m \times r_m}$ stores feature geometry, $\widetilde{H} \in \mathbb{R}^{b \times b}$ target-bit correlations, and $\widetilde{D}_m \in \mathbb{R}^{r_m \times b}$ feature-target associations. At prefix $E = [n]$, normalize by count: $G_m = \widetilde{G}_m/n, H = \widetilde{H}/n, D_m = \widetilde{D}_m/n$.

Stack feature and target transposes in $U_m \in \mathbb{R}^{n \times r_m}$ and $C \in \{-1, +1\}^{n \times b}$. With Frobenius norm $\|\cdot\|_F$, head $W_m \in \mathbb{R}^{r_m \times b}$ fits codes and their pairwise relations through

$$\mathcal{L}_{\mathrm{rel}}^m(W_m) = \frac{1}{n^2} \left\| \frac{1}{b} U_m W_m C^\top - \frac{1}{b} C C^\top \right\|_F^2, \tag{4}$$

$$\mathcal{L}_{\mathrm{code}}^m(W_m) = \frac{1}{nb} \|U_m W_m - C\|_F^2, \tag{5}$$

$$\mathcal{F}_m(W_m) = \beta \mathcal{L}_{\mathrm{rel}}^m + \alpha \mathcal{L}_{\mathrm{code}}^m + \frac{\lambda}{2} \|W_m\|_F^2, \quad \alpha, \beta \geq 0,\ \lambda > 0. \tag{6}$$

Weights α, β control code and relation fitting; λ controls ridge regularization. The target $CC^\top/b$ has the angular expectation above. Trace expansion recovers both losses from G_m, H, D_m, including cross-segment pairs, without an $n \times n$ similarity matrix.

To solve the strongly convex objective, diagonalize $G_m = P_m \operatorname{diag}(g) P_m^\top$ and $H = Q \operatorname{diag}(h) Q^\top$, and set $J_m = 2\beta D_m H/b^2 + 2\alpha D_m/b$. Here diag forms a diagonal matrix, P_m, Q are orthogonal, and $g \in \mathbb{R}^{r_m}, h \in \mathbb{R}^b$ hold nonnegative eigenvalues. Rotation into these bases decouples the head entries. For $a \in [r_m]$ and $r \in [b]$,

$$W_m^\star = P_m \overline{W}_m Q^\top, \qquad (\overline{W}_m)_{ar} = \frac{(P_m^\top J_m Q)_{ar}}{2\beta g_a h_r/b^2 + 2\alpha g_a/b + \lambda}. \tag{7}$$

Positive denominators give the unique rotated head $\overline{W}_m$ and count-n solution $W_{m,n}^\star$; preceding equations suppress n. At $n \in \mathcal{C}_s$, query q has feature $u_q = \phi_m(q)$ and code $h_{m,n}(q) = \operatorname{sign}(W_{m,n}^{\star\top} u_q)$, compared with the stored codes of the opposite modality.

For $\alpha > 0$ and $r \in [b]$, define $\kappa_r = \alpha/b + \beta h_r/b^2$. With I the identity of order r_m, the eigenbasis of H gives

$$(W_m^\star Q)_{:r} = \left(G_m + \frac{\lambda}{2\kappa_r} I \right)^{-1} (D_m Q)_{:r}. \tag{8}$$

Here $(:\, r)$ selects column r. Relations set direction-specific ridge strengths. For positive-definite G_m, $\lim_{\lambda \to 0^+} W_m^\star = G_m^{-1} D_m$, independently of β.

3.4 STREAM UPDATE

Accumulate $\mathcal{S}([n_0])$ and solve Equation 7 before indexing the bootstrap. Each later arrival i adds $\mathcal{S}(\{i\})$. At $i \in \mathcal{C}_s$, normalize and solve before writing; otherwise retain the heads. Let $s(i) = n_0$ for bootstrap items and $s(i) = \max\{s \in \mathcal{C}_s : s \le i\}$ thereafter. The written code d_i^m is

$$d_i^m = \begin{cases} \operatorname{sign}(W_{m,s(i)}^{\star\top} u_i^m), & \text{E: encoder-write,} \\ c_i, & \text{T: target-write.} \end{cases} \tag{9}$$

Mode E is the default. Gallery B_n^m stacks $\{d_i^m\}_{i \le n}$ as immutable rows; both modes encode query features with the current head.

Analytic prefix moments require $O(r_x^2 + r_y^2 + b^2 + (r_x + r_y)b)$ storage and per-item work. Heads, the semantic map and the $O(Nb)$-bit gallery are additional.

3.5 PREFIX RECOVERY AND PARTITION INVARIANCE

Assume (A1) the same ordered stream and label-disclosure times; (A2) fixed nonzero labels, maps, features, randomness and indexing policy. Computation uses (A3) a shared bootstrap, sample-clock solves and one contribution per item; (A4) $\lambda > 0$ and exact prefix minimization; (A5) exact real arithmetic; and (A6) deterministic tie-breaking. Partition subscripts identify each run.

Theorem 1 (Exact task-partition invariance)**.** *For any taskification-equivalent partitions $\mathcal{P}, \mathcal{Q}$ and any shared checkpoint n,*

$$\mathcal{S}_{\mathcal{P},n} = \mathcal{S}_{\mathcal{Q},n}, \quad W_{m,\mathcal{P},n}^{\star} = W_{m,\mathcal{Q},n}^{\star}, \quad \{d_{i,\mathcal{P}}^m\}_{i \le n} = \{d_{i,\mathcal{Q}}^m\}_{i \le n}.$$

Both retrieval directions therefore have identical ranked lists:

$$R_{\mathcal{P},n}^g(q) = R_{\mathcal{Q},n}^g(q) \quad \textit{for every query } q \textit{ and direction } g. \tag{10}$$

Additivity recovers identical states; ridge gives a unique head. Shared checkpoints preserve Mode-E codes; the fixed map preserves Mode-T codes (Appendix A.3).

Approximation can flip bits with near-zero activations. Suppressing m, a computed head $\widehat{W}$ has gradient-error bounds $\varepsilon_{\text{grad}}$ from approximate moments and ε_{opt} from solving. Strong convexity gives $\|\widehat{W} - W^{\star}\|_F \le (\varepsilon_{\text{grad}} + \varepsilon_{\text{opt}})/\lambda$. Activation margins then bound bit flips, ranking inversions, average precision (AP) changes and replacements at retrieval depth k.

3.6 CLOCKED EXTENSIONS AND QUERY REFINEMENT

Fixed intervals retain the code updates of discrete learners such as LEMON (Wang et al., 2020b) while controlling their write times. For block length $c \in \mathbb{N}_{>0}$, sort $\{n_0, N\} \cup \mathcal{C} \cup \{n_0 + jc : j \in \mathbb{N}_{>0},\ n_0 + jc < N\}$ as $n_0 = a_0 < a_1 < \cdots < a_K = N$. The K intervals are, for $k \in [K]$, $I_k = \{a_{k-1} + 1, \ldots, a_k\}$. Let $\mathcal{Z}_{I_k}$ contain the ordered image-text-label tuples and U_{m,I_k} their row-stacked features. Let $\mathcal{U}$ be the learner's update, A_0 its bootstrap state, A_k its state after interval k, and ξ_k random draws keyed by seed and interval endpoints. The update returns binary codes $B_{I_k} \in \{-1, +1\}^{|I_k| \times b}$:

$$(B_{I_k}, A_k) = \mathcal{U}(\mathcal{Z}_{I_k}, A_{k-1}; \xi_k), \qquad \widetilde{D}_m \leftarrow \widetilde{D}_m + U_{m,I_k}^{\top} B_{I_k}. \tag{11}$$

Each B_{I_k} is written once; query heads use ridge regression on cumulative feature-code pairs.

Shared bootstrap, intervals and deterministic updates give identical states, writes and rankings by induction, including finite-step nonconvex updates.

For the analytic branch, query refinement (QR) refits the final-prefix head using training features and stored codes. Its quadratic objective averages input noise and aligns predictions with the frozen gallery.

For the discrete branch, calibration fixes gallery B_g and heads W_g (g denotes gallery). Training-label regressions supply semantic predictions; alternate-clock and pairwise-compatible heads are aligned to this gallery's bit coordinates before their outputs are combined.

Base logits $z_q \in \mathbb{R}^b$ sum the normalized current output and a development-weighted pairwise output. Input $v_q \in \mathbb{R}^{5b+1}$ concatenates the normalized current, aligned alternate and pairwise outputs, current-alternate elementwise products, absolute differences and a constant 1. A zero-initialized b-dimensional residual r_θ with parameters θ gives query code h_q at scale s:

$$h_q = \operatorname{sign}(z_q + s\, r_\theta(v_q)), \qquad s \in [0, 1]. \tag{12}$$

We fit θ on training-stream pairs using binary and Dice-graded listwise anchor ranking of relaxed codes (He et al., 2018; Çakir et al., 2019), stability, quantization and residual-norm penalties. Development data select the objective, architecture and scale.

At inference, covariance-based denoising $\widetilde{u} = T(u_q)$ (Chen et al., 2012) optionally precedes selected query heads; its protected variant preserves native and semantic outputs. For MIRFlickr text, calibrated logits z_q^0 then receive correction Δ before binarization: a training-fitted, gradient-balanced ranking residual plus improved-binary-target regression.

4 Experiments

4.1 Experimental setup

NUS-WIDE, COCO-2014, MIRFlickr-25K and IAPR-TC12 use frozen 512-D CLIP features and fixed 80/10/10 stream/development/test splits (Chua et al., 2009; Lin et al., 2014; Huiskes et al., 2010; Escalante et al., 2010). Wiki supplies 2,173 training pairs and 231 evaluation queries with non-CLIP features (Rasiwasia et al., 2010). After bootstrap, P_t groups 128-item microbatches into t nearly equal tasks; P_{micro} uses one task per microbatch. We measure cross-partition agreement with the top-k retrieval reproducibility rate (TRR): the Jaccard overlap of top-k item sets returned for the same query, averaged over queries and pairs of taskifications. The primary comparison covers 18 configurations at 64 bits: mAP uses P_{10}, and TRR averages all pairs of $P_1, P_{10}, P_{\mathrm{micro}}$. Additional studies use six taskifications and examine retrieval quality at 16-128 bits. A separate evaluation uses the 32,000/4,000/4,000 train/validation/test split of XMediaNet (Peng et al., 2018a;b).

Paired taskifications share stream order, labels, bootstrap, seeds, settings, query/gallery IDs and evaluation counts. Comparisons retain native update schedules, writers and backends; interval-shift controls match update counts. Relevance means a shared label; Hamming ties use sample ID. On common features and streams, LEMON (Wang et al., 2020b), DOCH (Zhan et al., 2022), OCMFH (Wang et al., 2020a), OH-ELS (Fan et al., 2026), OCMH (Xie et al., 2016), OMGH (Liu et al., 2023), OHPKL (Shu et al., 2024) and ROHLSE (Li et al., 2024) retain their published update structures; OMGH's manifold weight uses development data. CT-CMH, DLCH, SILCH and MCDLCH form the linear-head mechanism-control block in Table 1. Fixed CLIP-RP uses random hyperplanes; Joint-full solves each complete prefix. LEMON and OH-ELS use 128-item intervals in the clock intervention. Calibration selects LEMON (2,048-item intervals) or DOCH (512) galleries on development data. The core/adapter write encoder codes; discrete methods write optimized codes; Mode T writes permanent targets. Queries use modality features.

Cross-partition agreement compares returned items (Fagin et al., 2003). Let $A^k_{\mathcal{P},n,g}(q)$ contain the first k IDs of $R^q_{\mathcal{P},n}(q)$, and let $\mathcal{E}_\Pi$ be the unordered pairs of distinct taskifications in the evaluated family Π. For the eligible query set $\mathcal{Q}_{n,g}$, the primary metric is TRR:

$$\mathrm{TRR}^{n,g}_k = \frac{1}{|\mathcal{Q}_{n,g}|\,|\mathcal{E}_\Pi|} \sum_{q \in \mathcal{Q}_{n,g}} \sum_{\{\mathcal{P},\mathcal{Q}\} \in \mathcal{E}_\Pi} \frac{|A^k_{\mathcal{P},n,g}(q) \cap A^k_{\mathcal{Q},n,g}(q)|}{|A^k_{\mathcal{P},n,g}(q) \cup A^k_{\mathcal{Q},n,g}(q)|}. \tag{13}$$

TRR equals one for identical sets. We use $k \in \{10, 20, 50\}$, $k \le M_{n,g}$; ordered-list agreement also requires identical order. mAP measures binary relevance; normalized discounted cumulative gain (nDCG) at 100 uses Dice-graded relevance. Normalized Kendall distance ($\mathrm{PG}_{\mathrm{rank}}$) and absolute mAP gaps ($\mathrm{PG}_{\mathrm{MAP}}$) measure full-ranking changes. Scores are percentages, with means and sample SDs over three joint stream/train/code seeds paired across methods; TRR compares partitions within each seed. Core effects use 10,000-draw hierarchical bootstraps and Holm correction; discrete effects use paired seed-level t intervals. Integer scoring covers all eligible queries; query-refinement comparisons hold galleries fixed.

Table 1: Final-prefix retrieval at 64 bits (%; two directions, three seeds). Upper blocks: TRR@20 / mAP; lower block: mAP. Shading ranks mAP: **Best** 2nd 3rd (ties share rank).

Method	NUS-WIDE	COCO	MIRFlickr	IAPR-TC12
LEMON	82.61 / 81.85	66.78 / 82.02	68.67 / 89.78	49.26 / 67.59
DOCH	2.15 / 83.20	1.97 / 70.97	7.74 / 92.20	10.12 / 66.80
OCMFH	67.24 / 57.89	22.26 / 51.47	31.17 / 67.92	10.15 / 37.25
OH-ELS	3.17 / 75.45	3.81 / 57.36	2.83 / 89.38	3.63 / 51.76
OCMH	20.25 / 43.32	16.90 / 60.23	9.54 / 61.77	14.46 / 39.44
OMGH	1.90 / 67.14	0.23 / 56.50	2.26 / 81.63	1.25 / 44.50
OHPKL	21.32 / 75.28	22.61 / 75.72	14.11 / 85.38	14.67 / 64.29
ROHLSE	14.64 / 72.86	15.91 / 66.28	6.47 / 84.38	4.98 / 55.72
Linear-head mechanism controls[a]				
CT-CMH	56.28 / 71.62	57.49 / 74.68	60.73 / 77.41	50.85 / 44.57
DLCH	56.24 / 71.62	57.55 / 74.69	60.72 / 77.41	50.76 / 44.57
SILCH	31.38 / 71.67	29.60 / 74.90	30.08 / 78.27	25.41 / 45.33
MCDLCH	44.62 / 71.50	52.68 / 74.65	67.82 / 76.86	58.49 / 44.10
mAP (TRR@20 = 100%)				
Fixed CLIP-RP	33.16	38.47	57.03	32.72
Joint-full	71.79	75.07	81.26	51.87
PILCH	71.75	74.93	80.44	50.09
PILCH + adapter	72.27	75.53	81.02	50.50
LEMON + fixed clock	81.84	81.79	89.79	67.57
Clocked + calibration	**83.94**	**84.46**	**92.68**	**69.84**

[a] These controls share linear heads, teacher codes, initialization and fitting loss; their retained mechanisms include replay, alignment, gradient balancing and optimizer-state components.

4.2 Retrieval across datasets and code lengths

The calibrated clocked extension attains the highest mean mAP on all four datasets in Table 1, which compares online hashing methods and fixed-feature references. For the analytic encoder-write branch, the bootstrap adapter improves mAP by +0.527/+0.605/ +0.586/+0.409 points on NUS-WIDE/COCO/MIRFlickr/IAPR-TC12, preserving zero partition gap and exceeding the six linear-head controls in this encoder-write study. The analytic and clocked-discrete studies cover 16-128 bits; the final calibrator uses 64 bits.

Calibration improves ranking with the 64-bit sample-clock galleries held fixed. It exceeds the development-selected comparator's mean mAP by 0.04-2.66 points in all 10 dataset-direction cells. Mean nDCG@100 also improves in every direction. COCO image-to-text retrieval gains 2.66 mAP and 3.57 nDCG@100 points, showing that a query-side update improves both binary and graded ranking. Positive mAP effects occur in 26 of 30 seed-level comparisons; four are negative, and short-list precision varies by dataset. Table 1 reports this calibrated reference before input denoising, with gains summarized at the direction level.

On XMediaNet's 32,000-item stream and 4,000 test queries, the permanent-target special case gains 12.150/2.943 mAP points over the strongest comparator under this protocol, with both paired 95% intervals above zero. This evaluates label-visible writes with shared bootstrap-prototype features and validation-selected ridge settings, separately from the encoder-write studies.

Figure 3 shows a flower query with its original tags. Its first 30 results include 20 flower matches and 10 additional plant-only matches, distinguishing the specific subject from its broader semantic category. The upper panels show local neighborhoods before and after hashing for each modality. Their independently fitted coordinates have no shared orientation or scale.

4.3 Boundary and input robustness

PILCH preserves returned items across taskifications at all three retrieval depths. Across the four datasets, LEMON attains 49.26-82.61% TRR@20, DOCH 1.97-10.12%, and OCMFH 10.15-67.24%. LEMON's mean mAP span is at most 0.233 point, yet on IAPR, 55.50% of queries have

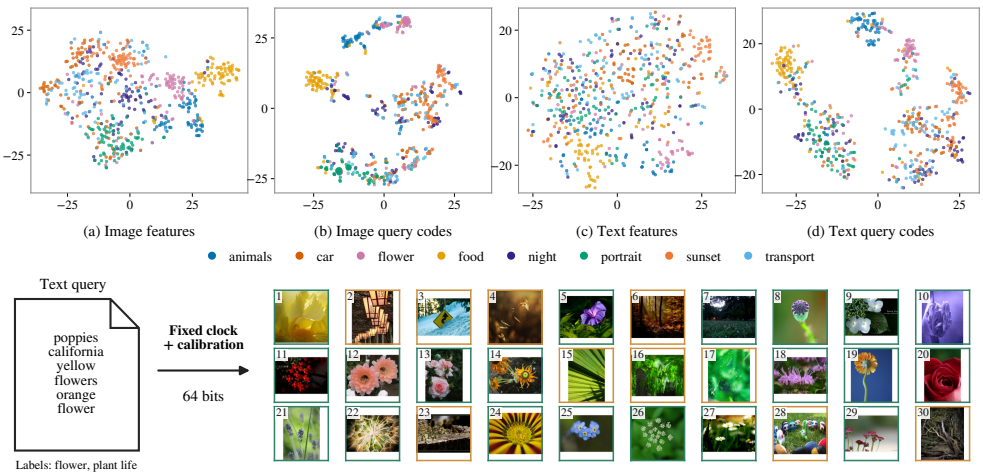


Figure 3: MIRFlickr at 64 bits (seed 0). Top: separate t-SNE projections of image/text features and current query codes for the same 600 items; colors denote semantic strata. Bottom: top-30 text-to-image results for a flower query, selected by item-ID hash. Green borders mark flower matches; amber marks plant-only matches, and red marks neither label.

mean LEMON TRR@20 below 0.5. The core's ordered lists also agree at $k = 10, 20, 50$. Aggregate relevance scores thus conceal substantial changes in returned identities.

Among methods with 100% TRR (Table 1), clocked LEMON exceeds fixed CLIP-RP (32.72-57.03% mAP) on every dataset. Identical returned-item agreement can therefore coexist with different relevance.

The 128-sample LEMON clock returns identical sets in the 30 reference-partition comparisons without database rewrites. Across five datasets, its 64-bit mean mAP differs from LEMON by at most 0.233 point; four paired intervals include zero. OHELS provides a second optimizer-family comparison. Clocking LEMON thus recovers result-set agreement with little change in mean relevance.

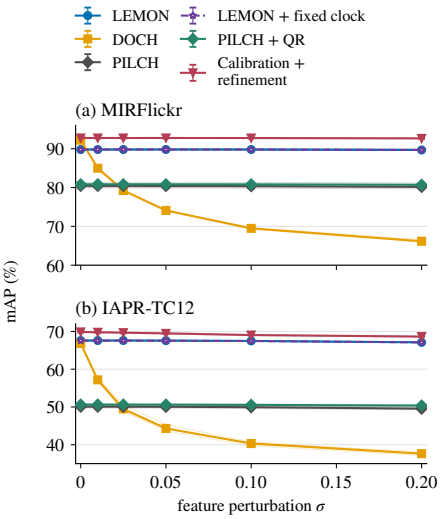


Figure 4: Query-noise mAP (64 bits). Lines: direction-averaged seeds; markers/bars: mean/SD.

Figure 4 compares six pipelines after adding noise of norm σ and renormalizing features. The refined discrete extension leads mean mAP on both datasets; QR improves the analytic branch. At the largest perturbation ($\sigma = 0.2$), the refined extension retains 92.66/68.63% mAP on MIRFlickr/IAPR, versus 89.68/67.11% for the strongest unrefined comparator in each dataset. These curves measure input sensitivity; TRR measures boundary dependence. At $\sigma = 0.1$, QR raises the core's MIRFlickr/IAPR overlap from 60.05/48.36% to 65.61/60.37%, while improving mAP and nDCG@100. For the calibrated extension, query refinement raises MIRFlickr overlap from 7.68% to 70.01%. Its text-query correction improves graded retrieval and set stability over denoising alone.

At matched denoising strength $\rho = 0.4$ (noise covariance $\rho^2 I/d$ in d feature dimensions), MIRFlickr text-query mAP at $\sigma = 0.1$ is 90.04%, versus DOCH's 89.56% and LEMON's 86.83%. LEMON retains higher nDCG@100 (59.65% versus 54.00%) and top-20 overlap (75.30% versus 66.42%): the metrics favor different pipelines.

4.4 Ablation and mechanism analysis

On IAPR, the Mode-T core leads the three ablations in mean TRR, partition-averaged mAP and nDCG@100 (Table 2). Task-local learning attains 4.30-44.35% TRR@20 across datasets, whereas the accumulated core returns identical items. One-write indexing alone leaves sensitivity in targets, retained state and optimization.

All 35,880 core ordered-list comparisons match; their top-k sets also agree. The fixed item-ID tie rule is important here: 93.34-94.14% of queries have equal Hamming distances at positions 20 and 21. Taskwise targets also change semantic quality. Under P_1, PCA-ITQ's nDCG@100 exceeds the core by 2.28 points; re-estimating those targets in smaller tasks reduces its mean. The benefit of permanent coordinates emerges when codes from separately processed segments share the same gallery.

Table 2: IAPR Mode-T ablation at 64 bits (%, mean±sample SD over three seeds). TRR averages directions and partition pairs; mAP and nDCG average directions and $P_1, P_{10}, P_{\text{micro}}$. TRR of 100% means identical returned sets. Column ranks: **Best** 2nd 3rd.

Method	TRR@20	mAP	nDCG@100
PILCH (T)	**100.00**±0.00	**57.65**±0.60	**49.53**±0.49
Taskwise PCA-ITQ (T)	0.33±0.03	40.85±0.28	28.76±0.23
Within-task state (T)	4.30±0.86	49.05±0.93	36.51±0.99
Boundary SGD (T)	33.91±0.25	56.98±0.67	47.32±0.71

Taskwise PCA-ITQ targets (A1) cost 20.127-33.958 mAP points under P_{10}: new tasks change target axes while historical codes retain the old axes. Permanent random targets (A9) avoid this drift but cost 27.266-49.268 points because their pairwise agreement omits label similarity. Dice targets preserve both coordinates and semantic neighborhoods. In addition, cross-task accumulation gains 0.432/0.684/ 3.189/7.720 points with Mode-T writes, and 0.922-14.056 with encoder writes. Accumulation retains feature-target associations in D_m and feature geometry in G_m; a task-local solve discards both. By contrast, removing the relation term changes Mode-T mAP by at most 0.005 point. Equation 8 explains this difference: β changes directional regularization, whereas resetting the state changes the observations defining the regression. The small effect is consistent with the weak-ridge limit at the chosen $\lambda = 10^{-5}$. With encoder writes, head errors affect both permanent item codes and later queries. Mode T instead fixes gallery codes from labels, isolating the query-head effect of discarding history.

Across four cohorts and three seeds, merging 13 unequal shards in three orders gives maximum relative moment/head errors of 3.40×10^{-16}/1.54×10^{-12}; all query codes agree. Raw moments give each observation one contribution, regardless of shard size. Reduction-order errors do not change any query-bit signs in these runs, preserving rankings despite finite-precision summation.

Vocabulary expansion costs 7.377-13.888 mAP points by changing target relations across arrivals; order controls with fixed moments differ by less than 0.001. With equal update counts, interval shifts change mean mAP by at most 0.177 point, with paired intervals including zero. Coordinates and observation weights determine the analytic head; the write clock aligns historical encoder codes.

On cached features, the analytic core is faster than DOCH by 50.2× on NUS-WIDE, 35.1× on COCO and 19.2× on MIRFlickr.

5 Conclusion

PILCH couples recovery of the full-prefix code-and-relation objective with sample-clock indexing. Additive statistics preserve cross-segment relations; common prefix heads and write times preserve historical codes. Theorem 1 establishes identical rankings across taskifications, and the approximation analysis connects moment and solver errors to bit flips, ranking inversions and retrieval changes. The write schedule also supports discrete learners. Fixed-gallery query calibration improves relevance in both directions on the four main datasets.

AI USE STATEMENT

In this work, we did not use generative AI tools for any task requiring disclosure. We have not used generative AI tools for research ideation, method design, code generation, experiment execution, data analysis, result interpretation, or artifact creation, and the remaining required disclosure tasks are not applicable to this work. Additionally, we used generative AI for language polishing and grammar checking. We have reviewed all AI assisted work. Specifically, all AI suggested language edits were manually checked by the authors against the original technical content, and no AI generated research ideas, code, experimental results, or scientific claims were used. We take responsibility for the final content of this work, including text, claims, or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

We have made the following efforts to ensure reproducibility. Section 3 specifies the objective, prefix statistics, solver, and indexing procedure. Appendix A provides complete derivations of the semantic map, the unique minimizer, the partition-invariance proof, and the approximation bounds. Appendix B records dataset construction, preprocessing steps, implementation details of all comparison methods, hyperparameter settings, and the query-refinement configurations. Section 4.1 and Appendix B.7 define the evaluation populations, three-seed aggregation, deterministic tie-breaking, and all metric definitions. Appendix C reports additional results, matched controls, code-length studies, and numerical-precision measurements. Anonymous source code, data-processing scripts, and evaluation code are included in the supplementary material.

REFERENCES

Amirali Aghazadeh, Andrew S. Lan, Anshumali Shrivastava, and Richard G. Baraniuk. Rhash: Robust hashing via l_infinity-norm distortion. In Carles Sierra (ed.), *Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence, IJCAI 2017, Melbourne, Australia, August 19-25, 2017*, pp. 1386–1394. ijcai.org, 2017. doi: 10.24963/IJCAI.2017/192. URL https://doi.org/10.24963/ijcai.2017/192.

Rahaf Aljundi, Klaas Kelchtermans, and Tinne Tuytelaars. Task-free continual learning. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 11254–11263. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.01151. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Aljundi_Task-Free_Continual_Learning_CVPR_2019_paper.html.

Vladimir Batagelj. Exactly mergeable summaries. *CoRR*, abs/2303.15465, 2023. doi: 10.48550/ARXIV.2303.15465. URL https://doi.org/10.48550/arXiv.2303.15465.

Christopher M. Bishop. Training with noise is equivalent to tikhonov regularization. *Neural Comput.*, 7(1):108–116, 1995. doi: 10.1162/NECO.1995.7.1.108. URL https://doi.org/10.1162/neco.1995.7.1.108.

Fatih Çakir, Kun He, Xide Xia, Brian Kulis, and Stan Sclaroff. Deep metric learning to rank. In *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2019, Long Beach, CA, USA, June 16-20, 2019*, pp. 1861–1870. Computer Vision Foundation / IEEE, 2019. doi: 10.1109/CVPR.2019.00196. URL http://openaccess.thecvf.com/content_CVPR_2019/html/Cakir_Deep_Metric_Learning_to_Rank_CVPR_2019_paper.html.

Yue Cao, Bin Liu, Mingsheng Long, and Jianmin Wang. Cross-modal hamming hashing. In Vittorio Ferrari, Martial Hebert, Cristian Sminchisescu, and Yair Weiss (eds.), *Computer Vision - ECCV 2018 - 15th European Conference, Munich, Germany, September 8-14, 2018, Proceedings, Part I*, volume 11205 of *Lecture Notes in Computer Science*, pp. 207–223. Springer, 2018. doi: 10.1007/978-3-030-01246-5_13. URL https://doi.org/10.1007/978-3-030-01246-5_13.

Moses Charikar. Similarity estimation techniques from rounding algorithms. In John H. Reif (ed.), *Proceedings on 34th Annual ACM Symposium on Theory of Computing, May 19-21, 2002, Montréal, Québec, Canada*, pp. 380–388. ACM, 2002. doi: 10.1145/509907.509965. URL https://doi.org/10.1145/509907.509965.

Minmin Chen, Zhixiang Eddie Xu, Kilian Q. Weinberger, and Fei Sha. Marginalized denoising autoencoders for domain adaptation. In *Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012*. icml.cc / Omnipress, 2012. URL http://icml.cc/2012/papers/416.pdf.

Zhao Chen, Vijay Badrinarayanan, Chen-Yu Lee, and Andrew Rabinovich. Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks. In Jennifer G. Dy and Andreas Krause (eds.), *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pp. 793–802. PMLR, 2018. URL http://proceedings.mlr.press/v80/chen18a.html.

Kun Cheng, Qibing Qin, Wenfeng Zhang, Lei Huang, and Jie Nie. Deep global-sense hard-negative discriminative generation hashing for cross-modal retrieval. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 113564–113585, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/b87c1fadf9218a831d53a353db605e1b-Paper-Conference.pdf.

Tat-Seng Chua, Jinhui Tang, Richang Hong, Haojie Li, Zhiping Luo, and Yantao Zheng. NUS-WIDE: a real-world web image database from national university of singapore. In Stéphane Marchand-Maillet and Yiannis Kompatsiaris (eds.), *Proceedings of the 8th ACM International Conference on Image and Video Retrieval, CIVR 2009, Santorini Island, Greece, July 8-10, 2009*. ACM, 2009. doi: 10.1145/1646396.1646452. URL https://doi.org/10.1145/1646396.1646452.

Hugo Jair Escalante, Carlos A. Hernández, Jesús A. González, Aurelio López-López, Manuel Montes-y-Gómez, Eduardo F. Morales, Luis Enrique Sucar, Luis Villaseñor-Pineda, and Michael Grubinger. The segmented and annotated IAPR TC-12 benchmark. *Comput. Vis. Image Underst.*, 114(4):419–428, 2010. doi: 10.1016/J.CVIU.2009.03.008. URL https://doi.org/10.1016/j.cviu.2009.03.008.

Ronald Fagin, Ravi Kumar, and D. Sivakumar. Comparing top k lists. *SIAM J. Discret. Math.*, 17 (1):134–160, 2003. doi: 10.1137/S0895480102412856. URL https://doi.org/10.1137/S0895480102412856.

Wentao Fan, Chao Zhang, Chunlin Chen, and Huaxiong Li. Online cross-modal hashing with expanding label space. In Sven Koenig, Chad Jenkins, and Matthew E. Taylor (eds.), *Fortieth AAAI Conference on Artificial Intelligence, Thirty-Eighth Conference on Innovative Applications of Artificial Intelligence, Sixteenth Symposium on Educational Advances in Artificial Intelligence, AAAI 2026, Singapore, January 20-27, 2026*, pp. 21002–21010. AAAI Press, 2026. doi: 10.1609/AAAI.V40I25.39242. URL https://doi.org/10.1609/aaai.v40i25.39242.

Nicolae Filat, Ahmed Hussain, Konstantinos Kalogiannis, and Elena Burceanu. Temporal taskification in streaming continual learning: A source of evaluation instability. *CoRR*, abs/2604.21930, 2026. doi: 10.48550/ARXIV.2604.21930. URL https://doi.org/10.48550/arXiv.2604.21930.

Mina Ghashami, Edo Liberty, Jeff M. Phillips, and David P. Woodruff. Frequent directions: Simple and deterministic matrix sketching. *SIAM J. Comput.*, 45(5):1762–1792, 2016. doi: 10.1137/15M1009718. URL https://doi.org/10.1137/15M1009718.

Kun He, Fatih Çakir, Sarah Adel Bargal, and Stan Sclaroff. Hashing as tie-aware learning to rank. In *2018 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2018, Salt Lake City, UT, USA, June 18-22, 2018*, pp. 4023–4032. Computer Vision Foundation / IEEE Computer Society, 2018. doi: 10.1109/CVPR.2018.00423. URL http://openaccess.thecvf.com/content_cvpr_2018/html/He_Hashing_as_Tie-Aware_CVPR_2018_paper.html.

Peng Hu, Hongyuan Zhu, Jie Lin, Dezhong Peng, Yin-Ping Zhao, and Xi Peng. Unsupervised contrastive cross-modal hashing. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(3):3877–3889, 2023. doi: 10.1109/TPAMI.2022.3177356. URL https://doi.org/10.1109/TPAMI.2022.3177356.

Mark J. Huiskes, Bart Thomee, and Michael S. Lew. New trends and ideas in visual concept detection: the MIR flickr retrieval evaluation initiative. In James Ze Wang, Nozha Boujemaa, Nuria Oliver Ramirez, and Apostol Natsev (eds.), *Proceedings of the 11th ACM SIGMM International Conference on Multimedia Information Retrieval, MIR 2010, Philadelphia, Pennsylvania, USA, March 29-31, 2010*, pp. 527–536. ACM, 2010. doi: 10.1145/1743384.1743475. URL https://doi.org/10.1145/1743384.1743475.

Young Kyun Jang and Ser-Nam Lim. Towards cross-modal backward-compatible representation learning for vision-language models. In *IEEE/CVF International Conference on Computer Vision, ICCV 2025, Honolulu, HI, USA, October 19-25, 2025*, pp. 1783–1792. IEEE, 2025. doi: 10.1109/ICCV51701.2025.00174. URL https://doi.org/10.1109/ICCV51701.2025.00174.

Qing-Yuan Jiang and Wu-Jun Li. Deep cross-modal hashing. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, pp. 3270–3278. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.348. URL https://doi.org/10.1109/CVPR.2017.348.

Qing-Yuan Jiang and Wu-Jun Li. Asymmetric deep supervised hashing. In Sheila A. McIlraith and Kilian Q. Weinberger (eds.), *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, pp. 3342–3349. AAAI Press, 2018. doi: 10.1609/AAAI.V32I1.11814. URL https://doi.org/10.1609/aaai.v32i1.11814.

Xiao Kang, Xingbo Liu, Wen Xue, Xuening Zhang, Xiushan Nie, and Yilong Yin. Discrete online cross-modal hashing with consistency preservation. *Pattern Recognit.*, 155:110688, 2024. doi: 10.1016/J.PATCOG.2024.110688. URL https://doi.org/10.1016/j.patcog.2024.110688.

Hang Thi-Thuy Le, Nam-Quan Nguyen, Lam-Huy Nguyen, Dien Dinh, Minh Hoang, and Trong Nghia Hoang. Task-free continual learning via order-invariant linearized adaptation and density-guided adapter routing. In Emilija Perković and Daniel Malinsky (eds.), *Proceedings of the 42nd Conference on Uncertainty in Artificial Intelligence*, volume 337 of *Proceedings of Machine Learning Research*, pp. 3274–3297. PMLR, 17–21 Aug 2026. URL https://proceedings.mlr.press/v337/le26a.html.

Kimin Lee, Kibok Lee, Honglak Lee, and Jinwoo Shin. A simple unified framework for detecting out-of-distribution samples and adversarial attacks. In Samy Bengio, Hanna M. Wallach, Hugo Larochelle, Kristen Grauman, Nicolò Cesa-Bianchi, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 31: Annual Conference on Neural Information Processing Systems 2018, NeurIPS 2018, December 3-8, 2018, Montréal, Canada*, pp. 7167–7177, 2018. URL https://proceedings.neurips.cc/paper/2018/hash/abdeb6f575ac5c6676b747bca8d09cc2-Abstract.html.

Li Li, Zhenqiu Shu, Zhengtao Yu, and Xiaojun Wu. Robust online hashing with label semantic enhancement for cross-modal retrieval. *Pattern Recognit.*, 145:109972, 2024. doi: 10.1016/J.PATCOG.2023.109972. URL https://doi.org/10.1016/j.patcog.2023.109972.

Tsung-Yi Lin, Michael Maire, Serge J. Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. In David J. Fleet, Tomás Pajdla, Bernt Schiele, and Tinne Tuytelaars (eds.), *Computer Vision - ECCV 2014 - 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V*, volume 8693 of *Lecture Notes in Computer Science*, pp. 740–755. Springer, 2014. doi: 10.1007/978-3-319-10602-1_48. URL https://doi.org/10.1007/978-3-319-10602-1_48.

Xin Liu, Jinhan Yi, Yiu-ming Cheung, Xing Xu, and Zhen Cui. OMGH: online manifold-guided hashing for flexible cross-modal retrieval. *IEEE Trans. Multim.*, 25:3811–3824, 2023. doi: 10.1109/TMM.2022.3166668. URL https://doi.org/10.1109/TMM.2022.3166668.

Mark D. McDonnell, Dong Gong, Amin Parvaneh, Ehsan Abbasnejad, and Anton van den Hengel. Ranpac: Random projections and pre-trained models for continual learning. In Alice Oh, Tristan

Naumann, Amir Globerson, Kate Saenko, Moritz Hardt, and Sergey Levine (eds.), *Advances in Neural Information Processing Systems 36: Annual Conference on Neural Information Processing Systems 2023, NeurIPS 2023, New Orleans, LA, USA, December 10 - 16, 2023*, 2023. URL http://papers.nips.cc/paper_files/paper/2023/hash/2793dc35e14003dd367684d93d236847-Abstract-Conference.html.

Yuxin Peng, Xin Huang, and Yunzhen Zhao. An overview of cross-media retrieval: Concepts, methodologies, benchmarks, and challenges. *IEEE Trans. Circuits Syst. Video Technol.*, 28(9): 2372–2385, 2018a. doi: 10.1109/TCSVT.2017.2705068. URL https://doi.org/10.1109/TCSVT.2017.2705068.

Yuxin Peng, Jinwei Qi, and Yuxin Yuan. Modality-specific cross-modal similarity measurement with recurrent attention network. *IEEE Trans. Image Process.*, 27(11):5585–5599, 2018b. doi: 10.1109/TIP.2018.2852503. URL https://doi.org/10.1109/TIP.2018.2852503.

Nikhil Rasiwasia, José Costa Pereira, Emanuele Coviello, Gabriel Doyle, Gert R. G. Lanckriet, Roger Levy, and Nuno Vasconcelos. A new approach to cross-modal multimedia retrieval. In Alberto Del Bimbo, Shih-Fu Chang, and Arnold W. M. Smeulders (eds.), *Proceedings of the 18th International Conference on Multimedia 2010, Firenze, Italy, October 25-29, 2010*, pp. 251–260. ACM, 2010. doi: 10.1145/1873951.1873987. URL https://doi.org/10.1145/1873951.1873987.

Yantao Shen, Yuanjun Xiong, Wei Xia, and Stefano Soatto. Towards backward-compatible representation learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, pp. 6367–6376. Computer Vision Foundation / IEEE, 2020. doi: 10.1109/CVPR42600.2020.00640. URL https://openaccess.thecvf.com/content_CVPR_2020/html/Shen_Towards_Backward-Compatible_Representation_Learning_CVPR_2020_paper.html.

Benwei Shi and Jeff M. Phillips. A deterministic streaming sketch for ridge regression. In Arindam Banerjee and Kenji Fukumizu (eds.), *The 24th International Conference on Artificial Intelligence and Statistics, AISTATS 2021, April 13-15, 2021, Virtual Event*, volume 130 of *Proceedings of Machine Learning Research*, pp. 586–594. PMLR, 2021. URL http://proceedings.mlr.press/v130/shi21b.html.

Zhenqiu Shu, Li Li, and Zhengtao Yu. Online hashing with partially known labels for cross-modal retrieval. *Eng. Appl. Artif. Intell.*, 138:109367, 2024. doi: 10.1016/J.ENGAPPAI.2024.109367. URL https://doi.org/10.1016/j.engappai.2024.109367.

Di Wang, Quan Wang, Yaqiang An, Xinbo Gao, and Yumin Tian. Online collective matrix factorization hashing for large-scale cross-media retrieval. In Jimmy X. Huang, Yi Chang, Xueqi Cheng, Jaap Kamps, Vanessa Murdock, Ji-Rong Wen, and Yiqun Liu (eds.), *Proceedings of the 43rd International ACM SIGIR conference on research and development in Information Retrieval, SIGIR 2020, Virtual Event, China, July 25-30, 2020*, pp. 1409–1418. ACM, 2020a. doi: 10.1145/3397271.3401132. URL https://doi.org/10.1145/3397271.3401132.

Xuanhui Wang, Cheng Li, Nadav Golbandi, Michael Bendersky, and Marc Najork. The lambdaloss framework for ranking metric optimization. In Alfredo Cuzzocrea, James Allan, Norman W. Paton, Divesh Srivastava, Rakesh Agrawal, Andrei Z. Broder, Mohammed J. Zaki, K. Selçuk Candan, Alexandros Labrinidis, Assaf Schuster, and Haixun Wang (eds.), *Proceedings of the 27th ACM International Conference on Information and Knowledge Management, CIKM 2018, Torino, Italy, October 22-26, 2018*, pp. 1313–1322. ACM, 2018. doi: 10.1145/3269206.3271784. URL https://doi.org/10.1145/3269206.3271784.

Yongxin Wang, Xin Luo, and Xin-Shun Xu. Label embedding online hashing for cross-modal retrieval. In Chang Wen Chen, Rita Cucchiara, Xian-Sheng Hua, Guo-Jun Qi, Elisa Ricci, Zhengyou Zhang, and Roger Zimmermann (eds.), *MM '20: The 28th ACM International Conference on Multimedia, Virtual Event / Seattle, WA, USA, October 12-16, 2020*, pp. 871–879. ACM, 2020b. doi: 10.1145/3394171.3413971. URL https://doi.org/10.1145/3394171.3413971.

Xinyu Xia, Guohua Dong, Fengling Li, Lei Zhu, and Xiaomin Ying. When CLIP meets cross-modal hashing retrieval: A new strong baseline. *Inf. Fusion*, 100:101968, 2023. doi: 10.1016/J.INFFUS.2023.101968. URL https://doi.org/10.1016/j.inffus.2023.101968.

Zhiping Xiao, Junyu Luo, Hang Zhou, Yusheng Zhao, Xiao Luo, Pengyun Wang, Wei Ju, Siyu Heng, and Ming Zhang. Hierarchical encoding tree with modality mixup for cross-modal hashing. In C. Vondrick, B. Hariharan, C. Raffel, L. Pinto, D. Yang, and A. Faust (eds.), *International Conference on Learning Representations*, volume 2026, pp. 126617–126641, 2026. URL https://proceedings.iclr.cc/paper_files/paper/2026/file/cd8bc4942f4d080335c6c158d0778cb7-Paper-Conference.pdf.

Liang Xie, Jialie Shen, and Lei Zhu. Online cross-modal hashing for web image retrieval. In Dale Schuurmans and Michael P. Wellman (eds.), *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, February 12-17, 2016, Phoenix, Arizona, USA*, pp. 294–300. AAAI Press, 2016. doi: 10.1609/AAAI.V30I1.9982. URL https://doi.org/10.1609/aaai.v30i1.9982.

Liming Xu, Hanqi Li, Bochuan Zheng, Weisheng Li, and Jiancheng Lv. Deep lifelong cross-modal hashing. *IEEE Trans. Circuits Syst. Video Technol.*, 34(12):13478–13493, 2024. doi: 10.1109/TCSVT.2024.3450490. URL https://doi.org/10.1109/TCSVT.2024.3450490.

Liming Xu, Hanqi Li, Jie Shao, Xianhua Zeng, and Weisheng Li. Multi-scale consistency deep lifelong cross-modal hashing. *ACM Trans. Multim. Comput. Commun. Appl.*, 21(10):283:1–283:23, 2025. doi: 10.1145/3704636. URL https://doi.org/10.1145/3704636.

Haoran Yang, Ying An, Zhan Yang, and Jun Long. Scale-invariant lifelong cross-modal hashing via gradient equilibrium. In Won Kim, Jae-Gil Lee, Kyuseok Shim, James Caverlee, David F. Gleich, Xiangliang Zhang, Qi He, Faisal Farooq, Rómer Rosales, Sayan Ranu, Haixun Wang, Yan Liu, Carl Yang, Leman Akoglu, Nitesh V. Chawla, and Vagelis Papalexakis (eds.), *Proceedings of the 32nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining V.2, KDD 2026, Jeju Island, Korea, August 9-13, 2026*, pp. 5986–5997. ACM, 2026. doi: 10.1145/3770855.3817901. URL https://doi.org/10.1145/3770855.3817901.

Yu-Wei Zhan, Yongxin Wang, Yu Sun, Xiao-Ming Wu, Xin Luo, and Xin-Shun Xu. Discrete online cross-modal hashing. *Pattern Recognit.*, 122:108262, 2022. doi: 10.1016/J.PATCOG.2021.108262. URL https://doi.org/10.1016/j.patcog.2021.108262.

Hao Zheng, Jinbao Wang, Xiantong Zhen, Jingkuan Song, Feng Zheng, Ke Lu, and Guo-Jun Qi. Continuous cross-modal hashing. *Pattern Recognit.*, 142:109662, 2023. doi: 10.1016/J.PATCOG.2023.109662. URL https://doi.org/10.1016/j.patcog.2023.109662.

Huiping Zhuang, Zhenyu Weng, Hongxin Wei, Renchunzi Xie, Kar-Ann Toh, and Zhiping Lin. ACIL: analytic class-incremental learning with absolute memorization and privacy protection. In Sanmi Koyejo, S. Mohamed, A. Agarwal, Danielle Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems 35: Annual Conference on Neural Information Processing Systems 2022, NeurIPS 2022, New Orleans, LA, USA, November 28 - December 9, 2022*, 2022. URL http://papers.nips.cc/paper_files/paper/2022/hash/4b74a42fc81fc7ee252f6bcb6e26c8be-Abstract-Conference.html.

A Theoretical Analysis

We use the assumptions in Section 3.5. For matrices, $\|\cdot\|_2$ is the spectral norm and tr is the trace; I_d is the identity of order d. Angle brackets denote the Frobenius inner product. The relation $A \succeq B$ means that $A - B$ is positive semidefinite. In single-modality derivations, the subscript m is omitted.

A.1 Permanent semantic coordinates

For nonzero ℓ_i, ℓ_j, put $a = \|\ell_i\|_2^2 + \|\ell_j\|_2^2 > 0$. Since $\int_0^\infty e^{-at}\,dt = 1/a$, with integration variable $t \geq 0$, the map $\psi_t(\ell) = \sqrt{2}e^{-t\|\ell\|_2^2}\ell$ satisfies

$$k_{\mathrm{D}}(\ell_i, \ell_j) = \int_0^\infty \psi_t(\ell_i)^\top \psi_t(\ell_j)\,dt. \tag{A.1}$$

For any real coefficients $v_1, \dots, v_n$,

$$\sum_{i,j=1}^{n} v_i v_j k_{\mathrm{D}}(\ell_i, \ell_j) = \int_0^\infty \left\| \sum_{i=1}^{n} v_i \psi_t(\ell_i) \right\|_2^2 dt \geq 0.$$

Thus k_{D} is positive semidefinite and $k_{\mathrm{D}}(\ell_i, \ell_i) = 1$. For the exact finite construction, let $\Gamma \in \mathbb{R}^{d_\ell \times d_\ell}$ have $\Gamma_{st} = 2/(s+t)$ for $s, t \in \{1, \dots, d_\ell\}$. The same integral proves $\Gamma \succeq 0$. Fix a matrix factor V such that $\Gamma = VV^\top$, with rows $v_s^\top$, and put $\nu(\ell) = \|\ell\|_0$, the number of nonzero labels, with $s_i = \nu(\ell_i)$. Using standard basis vectors $e_j \in \mathbb{R}^{d_\ell}$ and the Kronecker product $\otimes$, define

$$\psi(\ell) = \sum_{j=1}^{d_\ell} \ell_j (e_j \otimes v_{\nu(\ell)}) \in \mathbb{R}^{d_\ell^2}.$$

Then

$$\psi(\ell_i)^\top \psi(\ell_j) = \sum_{a=1}^{d_\ell} \sum_{r=1}^{d_\ell} \ell_{ia} \ell_{jr} (e_a^\top e_r)(v_{s_i}^\top v_{s_j}) = (\ell_i^\top \ell_j) \Gamma_{s_i s_j} = k_{\mathrm{D}}(\ell_i, \ell_j).$$

Its squared norm is $s_i \Gamma_{s_i s_i} = s_i / s_i = 1$.

A.2 Prefix objective and its unique minimizer

Expanding the squared losses gives

$$\mathcal{L}_{\mathrm{rel}}^m = \frac{1}{b^2} \left[\mathrm{tr}(W_m^\top G_m W_m H) - 2\,\mathrm{tr}(W_m^\top D_m H) + \mathrm{tr}(H^2) \right], \tag{A.2}$$

$$\mathcal{L}_{\mathrm{code}}^m = \frac{1}{b} \left[\mathrm{tr}(W_m^\top G_m W_m) - 2\,\mathrm{tr}(W_m^\top D_m) + \mathrm{tr}(H) \right]. \tag{A.3}$$

Cyclic permutation of the trace replaces $U^\top U/n$, $U^\top C/n$ and $C^\top C/n$ by G, D, H, respectively. Thus the relation expression retains all AA, AB, BA, BB pairs for disjoint row groups A, B. The identity uses uniform squared loss and the fixed finite-rank target.

Stacking the columns of W into a vector gives the Hessian

$$\nabla^2 \mathcal{F}_m = \frac{2\beta}{b^2}(H \otimes G_m) + \frac{2\alpha}{b}(I_b \otimes G_m) + \lambda I \succeq \lambda I. \tag{A.4}$$

Since $G, H \succeq 0$, the first two terms are positive semidefinite, and therefore $\nabla^2 \mathcal{F} \succeq \lambda I$. The objective is λ-strongly convex and has one minimizer. Setting its gradient to zero yields

$$\frac{2\beta}{b^2} G_m W_m H + \frac{2\alpha}{b} G_m W_m + \lambda W_m = \frac{2\beta}{b^2} D_m H + \frac{2\alpha}{b} D_m. \tag{A.5}$$

Substitute eigendecompositions of G and H, left-multiply by $P^\top$, and right-multiply by Q. Every coordinate satisfies

$$\left(\frac{2\beta g_a h_r}{b^2} + \frac{2\alpha g_a}{b} + \lambda \right) \overline{W}_{ar} = (P^\top J Q)_{ar}, \qquad a \in [r_m],\ r \in [b].$$

The parenthesized coefficient is at least $\lambda > 0$, proving Equation 7 and uniqueness.

Proposition 1. *Let $\alpha > 0$, $\beta \geq 0$, and $H = Q\,\mathrm{diag}(h)Q^\top$. For $r \in [b]$, put $\kappa_r = \alpha/b + \beta h_r / b^2$ and $v_r = (WQ)_{:r}$. The minimizer of Equation 6 satisfies*

$$v_r = \left(G + \frac{\lambda}{2\kappa_r} I \right)^{-1} (DQ)_{:r}.$$

If G is positive definite, then $W \to G^{-1} D$ as $\lambda \to 0^+$, independently of β.

Proof. The matrix $K = \alpha I/b + \beta H/b^2$ is positive definite. The stationary equation is $2(GW - D)K + \lambda W = 0$. Right multiplication by Q decouples its columns, giving $2\kappa_r(Gv_r - (DQ)_{:r}) + \lambda v_r = 0$. The displayed expression follows; invertibility of G gives the limit. □

The relation loss changes effective regularization along eigenvectors of the target Gram matrix H while preserving the regression target. When regularization is weak relative to informative eigenvalues of G, both terms approach the same limiting estimator. This identifies the role of β as directional regularization, rather than additional target information. Removing cross-task observations changes G, D and therefore a different part of the estimator.

A.3 PARTITION INVARIANCE

Include the optimizer variables, random-number state, stored codes and any learned feature parameters in A_k. Write the transition on fixed interval I_k as $A_k = F_k(A_{k-1}, \mathcal{Z}_{I_k}; \xi_k)$. The common bootstrap gives $A_{0,\mathcal{P}} = A_{0,\mathcal{Q}}$. If equality holds at $k-1$, both runs evaluate the same deterministic F_k on identical arguments, so $A_{k,\mathcal{P}} = A_{k,\mathcal{Q}}$. Induction proves equality of all writes and query encodings. Identical candidates and deterministic ties give identical rankings, establishing the clock-coupled observation in Section 3.6 for finite-iteration, nonconvex updates. The exact core also reconstructs the complete prefix objective from mergeable moments.

Equal update counts alone give a different condition. For the scalar stream $(-1, -1, -1, 3)$, consider the uniquely minimized objective $\sum_{k=1}^{2}(w - \bar{z}_{I_k})^2$ for scalar parameter w, where $\bar{z}_I = |I|^{-1}\sum_{i\in I} z_i$ is the mean of chunk I. The partitions $(3, 1)$ and $(1, 3)$ both perform two chunk updates, yet their minimizers are 1 and $-1/3$. A scalar query mapped by $\mathrm{sign}(w)$ therefore reverses the ranking of two database codes $+1$ and -1. In contrast, raw sums and counts recover the same full-stream mean for both partitions. Equal counts do not fix the weighting of observations.

Every legal taskification groups the same prefix $\mathcal{Z}_{1:n}$ into disjoint contiguous sets. Repeated application of Equation 3 and associativity and commutativity of addition in real arithmetic give identical

$$n,\ \widetilde{G}_x,\ \widetilde{G}_y,\ \widetilde{H},\ \widetilde{D}_x,\ \widetilde{D}_y.$$

Fixed semantic features and hyperplanes also give the same c_i. Hence both runs have the same G_m, H, D_m, objective and unique $W_m^\star$ at every sample-clock solve.

For Mode T, identical c_i directly imply identical one-write database codes. For Mode E, the fixed bootstrap prefix yields the same initial solved head and the solve schedule is shared. Bootstrap codes written at n_0 are therefore identical. Inductively, at every later item i, the most recent sample-clock head is identical by the previous paragraph; fixed features and sign convention generate identical d_i^m. Later solves do not rewrite it. Thus all database codes agree at every prefix. Current query codes, Hamming distances, and deterministic tie-breaking agree, which proves Equation 10.

A.4 APPROXIMATION-TO-RETRIEVAL CHAIN

Let $\widehat{G}, \widehat{H}, \widehat{D}$ approximate G, H, D, with $\|\widehat{G} - G\|_2 \le \varepsilon_G$, $\|\widehat{H} - H\|_2 \le \varepsilon_H$ and $\|\widehat{D} - D\|_F \le \varepsilon_D$. Write $\widehat{\mathcal{F}}$ for the objective with these approximate moments. Let $\widehat{W}$ have optimization residual $\|\nabla\widehat{\mathcal{F}}(\widehat{W})\|_F \le \varepsilon_{\mathrm{opt}}$. For nonnegative bounds $\|G\|_2 \le g$, $\|H\|_2 \le h$, $\|D\|_F \le d$, and $\|\widehat{W}\|_F \le R_W$, define

$$\begin{aligned}\varepsilon_{\mathrm{grad}} = \frac{2}{b^2}\beta\left[(\varepsilon_G R_W + \varepsilon_D)(h + \varepsilon_H) + (gR_W + d)\varepsilon_H\right] \\ + \frac{2\alpha}{b}(\varepsilon_G R_W + \varepsilon_D).\end{aligned} \tag{A.6}$$

Strong convexity gives

$$\boxed{\|\widehat{W} - W^\star\|_F \le \rho := \frac{\varepsilon_{\mathrm{grad}} + \varepsilon_{\mathrm{opt}}}{\lambda}.} \tag{A.7}$$

For continuous score $s_W(u, d) = u^\top W d/b$, $\|u\|_2 \le R_m$, and binary candidates $d_j, d_k \in \{-1, +1\}^b$, $j, k \in [n]$, define the score difference $\Delta s_W(j, k) = s_W(u, d_j) - s_W(u, d_k)$. Then

$$\left|\Delta s_{\widehat{W}}(j, k) - \Delta s_{W^\star}(j, k)\right| \le \frac{2R_m\rho}{\sqrt{b}}. \tag{A.8}$$

Thus only reference pairs within this margin can reverse. For query feature $u_q = \phi_m(q)$, let $w_r^\star, \widehat{w}_r$ denote column r of the exact and approximate heads. The number of potentially changed query bits is

$$e(q) = \left|\{r \in [b] : |w_r^{\star\top} u_q| \le R_m\|\widehat{w}_r - w_r^\star\|_2\}\right|. \tag{A.9}$$

Only these query bits can flip. With identical database codes, a candidate-distance difference changes by at most $2e(q)$. For approximate encoder-written database items, let $e_d(j)$ bound their bit changes using the corresponding historical write head. The general pairwise bound is

$$|\Delta\widehat{d}_H(q; j, k) - \Delta d_H(q; j, k)| \le 2e(q) + e_d(j) + e_d(k). \tag{A.10}$$

Here $h(q)$ abbreviates the reference query code $h_{m,n}(q)$, $\Delta d_H(q; j, k) = d_H(h(q), d_j) - d_H(h(q), d_k)$, and a hat denotes distances from approximate codes. Mode T with unchanged targets has $e_d(j) = 0$. Finally, if two full rankings differ by I_q relevant-irrelevant inversions and query q has $R_q \geq 1$ relevant items,

$$|\operatorname{AP}_{\mathcal{P}}(q) - \operatorname{AP}_{\mathcal{Q}}(q)| \leq \min\left(1, \frac{I_q}{R_q}\right). \tag{A.11}$$

For AP truncated at rank K, items crossing that cutoff are tracked in addition to full-ranking inversions.

Let $\Delta G = \widehat{G} - G$, $\Delta H = \widehat{H} - H$, and $\Delta D = \widehat{D} - D$. The difference in relation-gradient factors is

$$\begin{aligned}&(\widehat{G}\widehat{W} - \widehat{D})\widehat{H} - (G\widehat{W} - D)H \\ &= (\Delta G\widehat{W} - \Delta D)(H + \Delta H) + (G\widehat{W} - D)\Delta H.\end{aligned}$$

Submultiplicativity, followed by multiplication by $2\beta/b^2$, yields the relation part of $\varepsilon_{\text{grad}}$; the code-regression part is bounded by $2\alpha(\varepsilon_G R_W + \varepsilon_D)/b$. Strong convexity implies strong gradient monotonicity,

$$\langle \nabla\mathcal{F}(W_1) - \nabla\mathcal{F}(W_2), W_1 - W_2\rangle \geq \lambda\|W_1 - W_2\|_F^2.$$

Set $W_1 = \widehat{W}$, $W_2 = W^\star$, apply Cauchy-Schwarz, and split

$$\nabla\mathcal{F}(\widehat{W}) = [\nabla\mathcal{F}(\widehat{W}) - \nabla\widehat{\mathcal{F}}(\widehat{W})] + \nabla\widehat{\mathcal{F}}(\widehat{W}).$$

This proves Equation A.7. For two independently approximated partitions, the triangle inequality gives a parameter gap no larger than the sum of their two right-hand sides.

For any fixed binary candidates d_j, d_k,

$$\begin{aligned}&|\Delta s_{\widehat{W}}(j,k) - \Delta s_{W^\star}(j,k)| \\ &= b^{-1}|u^\top(\widehat{W} - W^\star)(d_j - d_k)| \\ &\leq b^{-1}\|u\|_2\|\widehat{W} - W^\star\|_F\|d_j - d_k\|_2 \leq 2R_m\rho/\sqrt{b},\end{aligned}$$

which proves Equation A.8. For bit r, Cauchy-Schwarz shows that the activation perturbation is at most $R_m\|\widehat{w}_r - w_r^\star\|_2$; a sign change therefore requires the event in Equation A.9. Flipping $e(q)$ query bits changes any Hamming distance by at most $e(q)$, and a pairwise distance difference by at most $2e(q)$.

For each historical write, the same activation argument bounds $e_d(j)$. The Hamming triangle inequality gives $|\widehat{d}_H(q,j) - d_H(q,j)| \leq e(q) + e_d(j)$; summing the bounds for j, k proves Equation A.10. Candidate pairs with a reference gap strictly larger than this bound retain their order. Ties and pairs at equality are included among potentially reversed pairs.

Any two rankings can be connected by adjacent swaps. Swapping two relevant or two irrelevant items leaves AP unchanged. An adjacent relevant-irrelevant swap changes one relevant item's precision contribution by at most $1/R_q$. Summing over the I_q cross-label inversions proves Equation A.11; clipping at one uses the range of AP.

B Experimental Details

Table 3 summarizes the six datasets.

B.1 Datasets and preprocessing

Stable item identifiers determine an 80/10/10 stream/development/test split for the first four datasets. Exact decoded-image duplicates are grouped before splitting, retaining the first item in source order. Query labels are used only by the scorer. NUS-WIDE follows the official image list reconciled with All_Tags Flickr identifiers; COCO pools the 2014 train and validation images while grouping all captions of an image. MIRFlickr excludes zero-label items, and IAPR aligns English captions with RefCLEF/SAIAPR regions.

Table 3: Dataset construction. Counts are post-filter stream/test sizes; every item pairs an image, the indicated text and a nonzero multi-hot label.

Dataset	Stream	Eval.	Labels	Source	Text and label construction
NUS-WIDE	156,288	19,770	21	269,648	Ordered user tags; top-21 nonzero ground-truth concepts (Chua et al., 2009).
COCO-2014	97,433	12,432	80	123,287	Official captions ordered by annotation ID; union of instance categories (Lin et al., 2014).
MIRFlickr-25K	18,004	2,278	24	25,000	Ordered Flickr tags; 24 potential-label files, excluding relevance refinements (Huiskes et al., 2010).
IAPR-TC12	15,807	1,995	255	19,997	Official English fields; union of manually annotated SAIAPR region concepts (Escalante et al., 2010).
Wiki	2,173	231	10	2,866	Supplied image/text features and category labels; source validation defines the query cohort (Rasiwasia et al., 2010).
XMediaNet	32,000	4,000	200	40,000	Published 4,096-D image/300-D text features and fixed train/validation/test split (Peng et al., 2018a;b).

Wiki evaluates the fixed-clock extension on 2,173 source-train pairs and 231 source-validation queries; the reported cohort does not use the 462 source-test queries. Both implementations include the cumulative code-latent cross moment H_5 in the Procrustes update.

For XMediaNet, each stream permutation uses its first 3,200 pairs to average one normalized prototype per class and modality; cosine similarities to those prototypes form a fixed 200-D coordinate map shared by all methods. The proposed Mode-T database writes the permanent 64-bit Dice target once. Its query heads use only the additive feature Gram and feature-target cross moment. The ridge grid $\{10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 100\}$ selects 10^{-4} for image queries and 10^{-2} for text queries on all 4,000 validation pairs aggregated over three seeds. The resulting heads score the 4,000 test queries against all 32,000 training items.

Eligibility is fixed at the first evaluation prefix. IAPR retains 1,994/1,994/1,992 queries for the three seeds from its 1,995-item test pool; the other CLIP datasets retain all listed test queries. Every comparison uses the same cached backbone features.

B.2 Implementation

For the first four datasets, images and text are encoded once by the frozen backbone, normalized, and augmented with a bias coordinate before moment accumulation. Wiki uses its supplied 4,096-D image and 300-D text features. A signed feature hash with seed 4040 maps each image coordinate to one of 512 signed buckets; text is unchanged. This transform is label- and split-independent, after which both views are normalized and biased. Table 4 fixes the core constants. The residual maps in Equation B.1 use $512 \to 128 \to 512$ projections, residual strength $\gamma = 1$, and a zero-initialized up projection. They are trained for 20 bootstrap epochs with AdamW and retained after the temporary predictors are discarded. Hyperparameters selected on NUS-WIDE development data are transferred unchanged to IAPR-TC12.

Let v_i^m be the backbone feature without the constant coordinate, and let $A_m^{\downarrow}, A_m^{\uparrow}, p_m$ denote learned affine maps. Here $\mathrm{norm}(v) = v/\|v\|_2$ and GELU is the Gaussian error linear unit. For $m \in \{x, y\}$, the residual feature and bootstrap loss are

$$a_m(v) = \mathrm{norm}\big[\mathrm{norm}(v) + \gamma A_m^{\uparrow}\big(\mathrm{GELU}\big(A_m^{\downarrow}(\mathrm{norm}(v))\big)\big)\big], \tag{B.1}$$

$$\mathcal{L}_{\mathrm{boot}} = \frac{1}{n_0 b} \sum_{i \le n_0, m} \|p_m(a_m(v_i^m)) - c_i\|_2^2 + \frac{0.1}{n_0 b} \sum_{i \le n_0} \|p_x(a_x(v_i^x)) - p_y(a_y(v_i^y))\|_2^2.$$

The temporary predictors p_m are discarded after fitting; all subsequent moments use $[a_m(v)^\top, 1]^\top$ in place of u. Optimization evaluates these averages on each minibatch with AdamW weight decay 10^{-4}.

Table 4: Shared implementation constants.

Component	Setting
Backbone	OpenAI CLIP ViT-B/32 QuickGELU; 512-D; ℓ_2 normalized
Feature/statistic precision	FP32 cache / FP64 moments and solve
Code lengths	$b \in \{16, 32, 64, 128\}$; primary $b = 64$
Objective	$\alpha = 10$, $\beta = 10$, $\lambda = 10^{-5}$
Bootstrap / microbatch	$0.1N$ / 128 items
Solve clock	$10, 25, 50, 75, 100\%$ of the stream
Evaluation clock	$25, 50, 75, 100\%$ of the stream
Taskifications	$P_1, P_2, P_5, P_{10}, P_{20}, P_{\mathrm{micro}}$
Joint seeds	$(1103, 2203, 3301)$, $(1109, 2207, 3307)$, $(1117, 2213, 3313)$
Adapter	bottleneck 128; 20 epochs; batch 512; AdamW; lr 3×10^{-3}
Adapter loss	target MSE $+0.1\times$ cross-modal MSE; weight decay 10^{-4}

B.3 Fixed-clock discrete extension

We implement the LEMON update (Wang et al., 2020b) with its cumulative Procrustes term. For a chunk of h items, let $L \in \mathbb{R}^{h \times d_\ell}$ contain labels and $\bar{L}$ their row-normalized values. Let $B \in \{-1, +1\}^{h \times b}$ be binary codes, $V \in \mathbb{R}^{h \times b}$ balanced latent codes, $R \in \mathbb{R}^{b \times b}$ an orthogonal rotation, and $P \in \mathbb{R}^{b \times d_\ell}$ the label predictor. With an all-ones column $\mathbf{1} \in \mathbb{R}^h$ and zero-initialized cumulative moments $(H_1, \ldots, H_7)$, each iteration forms a latent-update matrix $Z \in \mathbb{R}^{h \times b}$ using relation weight α_L, label weight β_L and predictor ridge γ:

$$\begin{aligned}
&P = (H_1 + V^\top V + \gamma I)^{-1}(H_2 + V^\top L),\\
&Z = BR^\top + \beta_L L P^\top + 2\alpha_L b \bar{L}(H_3 + \bar{L}^\top B) - \alpha_L b \mathbf{1}(H_4 + \mathbf{1}^\top B),\\
&V \leftarrow \mathrm{BalancedSVD}(Z), \qquad R \leftarrow \mathrm{Procrustes}(H_5 + B^\top V),\\
&B \leftarrow \mathrm{sign}\big(2\alpha_L b \bar{L}(H_6 + \bar{L}^\top V) - \alpha_L b \mathbf{1}(H_7 + \mathbf{1}^\top V) + VR\big).
\end{aligned}$$

Here BalancedSVD uses the centered polar factor of Z with $\sqrt{h}$ scaling and seeded completion below squared singular value 10^{-6}. For a singular value decomposition $M = U_M \Sigma_M V_M^\top$, $\mathrm{Procrustes}(M) = V_M U_M^\top$. After the chunk, add

$$(V^\top V, V^\top L, \bar{L}^\top B, \mathbf{1}^\top B, B^\top V, \bar{L}^\top V, \mathbf{1}^\top V)$$

to $(H_1, \ldots, H_7)$. Feature statistics add $(U_{m,I_k}^\top U_{m,I_k}, U_{m,I_k}^\top B)$ to $(\widetilde{G}_m, \widetilde{D}_m)$ and solve $W_m = (\widetilde{G}_m + \lambda_L I)^{-1}\widetilde{D}_m$ at the evaluation checkpoint. Here λ_L is the query-head ridge. The constants are $\alpha_L = \beta_L = 10^4$, $\gamma = 0.1$, $\lambda_L = 1$ and five iterations. Fixed-clock chunks close every 128 samples and at shared evaluation counts. Both methods use FP64 CUDA updates on the four CLIP datasets and FP64 NumPy updates on Wiki; paired runs always share a backend. Random draws are keyed by global seed and interval endpoints. When a short interval has fewer rows than the

requested orthogonal-completion columns, the implementation uses seeded unit-norm columns and records that event. Otherwise, completion uses the left singular vectors of a seeded Gaussian matrix. LEMON uses the same numerical rule and constants with evaluator-triggered chunks.

The two LEMON schedules use this recurrence on four datasets and three seeds: 64 bits at $P_1/P_{10}/P_{\text{micro}}$ and 16/32/128 bits at P_{10}. Each of these 144 cells is evaluated at four checkpoints in both directions.

B.4 Comparison methods

Table 5 distinguishes controlled adaptations from linear-head mechanism controls and records the retained update mechanism.

Table 5: Comparison methods and implementation scope. Adaptations retain the source update structure and database rewrite policy. Linear-head controls share the semantic teacher, bootstrap initialization and primary loss.

Method	Status	Boundary-sensitive mechanism retained
LEMON (Wang et al., 2020b)	controlled adaptation	source chunk update with cumulative Procrustes state; five inner iterations
DOCH (Zhan et al., 2022)	controlled adaptation	bootstrap/current anchors and 500-code old-anchor budget
OCMFH (Wang et al., 2020a)	controlled adaptation	collective factorization update and task-time code reconstruction
OH-ELS (Fan et al., 2026)	controlled adaptation	asymmetric binary updates, historical anchors and expanding-class classifier consistency
OCMH (Xie et al., 2016)	source implementation	shared latent codes, two modality transforms and six accumulated sufficient statistics
OMGH (Liu et al., 2023)	source implementation	anchor-guided manifold regularization and supervised tri-factorization
OHPKL (Shu et al., 2024)	controlled adaptation	pairwise and chunk similarity with cumulative feature-code regression
ROHLSE (Li et al., 2024)	controlled adaptation	label semantic enhancement and within-/cross-chunk similarity
Local update	mechanism control	local code loss, cross-modal consistency, optimizer reset
Mixed replay	mechanism control	512 mixed anchors and distillation
CT-CMH (Zheng et al., 2023)	mechanism control	512 anchors and unit distillation
DLCH (Xu et al., 2024)	mechanism control	1,024 anchors, distillation, cross-modal consistency
SILCH (Yang et al., 2026)	mechanism control	mixed 1,024-anchor replay with gradient-norm equilibration
MCDLCH (Xu et al., 2025)	mechanism control	mixed replay, gradient equilibration, persistent optimizer
Fixed CLIP-RP	feature control	fixed random hyperplanes; no stream learning
Joint-full	oracle-style control	one full-prefix ridge solve at each evaluation clock

Development-selected learning rates are 0.003 for the first four linear-head controls, 0.01 for balanced replay, and 0.001 for persistent replay; each sees one optimization epoch per global micro-batch.

OCMH uses three alternating sweeps per arriving chunk, with $\lambda = \alpha = \beta = 10^{-6}$ and $\theta = 0.3$. OMGH uses two sweeps, 500 anchors and ranks 100/50. Its manifold weight is selected from $\{10^{-8}, 10^{-6}, 10^{-4}, 10^{-2}, 1, 100\}$ on development queries, then held fixed across seeds and taskifications. The selected values are $10^{-2}, 10^{-4}, 10^{-2}, 10^{-6}$ on NUS-WIDE, COCO, MIRFlickr and

IAPR, respectively. These two implementations use Python 3.12 and PyTorch 2.8.0; their reported comparison uses the final prefix at 64 bits.

OHPKL uses five alternating sweeps per chunk with all stream labels observed. ROHLSE uses ten sweeps on MIRFlickr, COCO and IAPR, and fifteen on NUS-WIDE; its remaining settings follow the released dataset configurations, with the MIRFlickr configuration used for COCO. Both methods operate directly on the shared 512-D features and center query logits over the unlabeled query set, following their released query encoding. For the final microchunks smaller than 64 samples, the auxiliary update uses the compact rectangular polar factor. Implementations use float64 NumPy/SciPy operations; full settings and source identifiers accompany the code.

The main training runs use Python 3.10.8 and PyTorch 2.1.2 with CUDA 12.1 on RTX 4090/4090D hosts, with CPU or CUDA kernels selected by method. OH-ELS uses NumPy/Numba CPU kernels. The common retrieval scorer operates on the saved binary codes. Hardware and optimizer kernels vary by method; each paired clock intervention uses a common backend within its dataset.

B.5 OH-ELS IMPLEMENTATION

OH-ELS (Fan et al., 2026) supplies a recent discrete online comparator. We implement its asymmetric sample-level objective, replay of historical binary anchors, old-class classifier consistency and cumulative query-head regression. At 64 bits, similarity and consistency weights are 10 and 100. Each update performs 15 complete code/classifier sweeps with 50 current and 500 old anchors, sampled without replacement and clipped to the available population. Bits update sequentially with fixed anchors; zero maps to -1 during binary optimization and $+1$ in query encoding. Historical codes are immutable, and optimized B, rather than latent V, enters cumulative feature-code regression. Singular systems use the Moore-Penrose solution with relative cutoff 10^{-12}. The benchmark uses shared features and fixed label identities; the source uses an expanding-class GIST/BoW/SIFT benchmark.

Table 6 compares P_{10} with the fixed 128-sample clock under the same objective and per-update sweeps. The main table includes the P_{10} comparator. The raw results retain both directions and all intermediate checkpoints.

Table 6: OH-ELS retrieval at the final prefix and 64 bits on the common feature cache. Values are bidirectional mAP mean±sample SD over three seeds; Δ is fixed clock minus P_{10} with a paired 95% t interval. Both schedules write optimized B codes.

Dataset	OH-ELS	Fixed clock	Δ [95% CI]
NUS-WIDE	75.453±0.627	75.257±0.677	-0.196 [-0.769, +0.377]
COCO	57.360±0.511	56.793±0.099	-0.567 [-1.591, +0.457]
MIRFlickr	89.384±0.152	89.306±0.157	-0.078 [-0.167, +0.011]
IAPR-TC12	51.758±0.258	52.196±0.262	+0.438 [-0.049, +0.925]
Wiki	65.701±1.107	63.516±0.513	-2.185 [-3.804, -0.567]

B.6 MATCHED WRITERS AND EQUAL-UPDATE-COUNT INTERVENTIONS

Table 7 isolates replay and update mechanisms with the same linear heads, teacher targets and fitting objective.

Table 7: Linear-head mechanism controls at 64 bits (TRR@20 / mAP, %). Means use both directions and three seeds; mAP uses P_{10}.

Method	NUS-WIDE	COCO	MIRFlickr	IAPR-TC12
Local update	56.36 / 71.61	57.30 / 74.69	60.06 / 77.41	49.84 / 44.57
Mixed replay	56.44 / 71.61	57.47 / 74.69	60.39 / 77.41	50.33 / 44.56
CT-CMH	56.28 / 71.62	57.49 / 74.68	60.73 / 77.41	50.85 / 44.57
DLCH	56.24 / 71.62	57.55 / 74.69	60.72 / 77.41	50.76 / 44.57
SILCH	31.38 / 71.67	29.60 / 74.90	30.08 / 78.27	25.41 / 45.33
MCDLCH	44.62 / 71.50	52.68 / 74.65	67.82 / 76.86	58.49 / 44.10

Table 8 compares the core, its zero-relation variant and task-local statistics under the same encoder-write policy. All five datasets share features, bootstrap, solve counts, code seeds and queries; only task-local state resets at P_{10} boundaries. Effects average four checkpoints and both directions within each of three seeds.

Table 8: Matched encoder-write effects at 64 bits, averaged over four checkpoints and both directions. Core mAP is mean±sample SD; effects are full minus ablation points with paired 95% t intervals over three seeds.

Dataset	Core mAP	Accumulated state	Relation term
NUS-WIDE	71.736±0.732	+0.922 [+0.691, +1.153]	-0.001 [-0.008, +0.006]
COCO	74.923±0.417	+1.721 [+1.543, +1.900]	-0.005 [-0.011, +0.000]
MIRFlickr	80.161±0.589	+7.656 [+6.580, +8.731]	-0.013 [-0.026, +0.001]
IAPR-TC12	49.189±0.199	+9.585 [+9.310, +9.861]	-0.019 [-0.023, -0.014]
Wiki	35.385±0.158	+14.056 [+12.849, +15.263]	-0.207 [-0.359, -0.054]

The equal-count intervention starts from the sorted union of bootstrap, 128-sample boundaries and evaluation endpoints. Each internal nonevaluation boundary moves left or right by one quarter of its smaller adjacent interval, floored to an integer; an available one-sample move handles intervals shorter than four samples. A one-sample interval is left unchanged. Bootstrap and evaluation endpoints remain fixed, boundaries stay ordered, and every checkpoint retains the same number of optimizer updates. Both perturbed schedules retain LEMON's five inner iterations and source-corrected objective. Equal counts control the number of updates, not their data-dependent FLOPs or wall time. Table 9 gives the paired intervals.

Table 9: Equal-update-count LEMON interventions. The left/right schedules shift internal interval boundaries while preserving the number of updates at every evaluation checkpoint. Entries are shifted minus fixed-clock mAP points averaged over checkpoints and directions, with paired 95% t intervals over three seeds.

Dataset	Left shift	Right shift
NUS-WIDE	-0.000 [-0.005, +0.005]	-0.000 [-0.002, +0.001]
COCO	-0.005 [-0.015, +0.005]	-0.001 [-0.010, +0.007]
MIRFlickr	-0.003 [-0.012, +0.006]	+0.001 [-0.004, +0.005]
IAPR-TC12	-0.005 [-0.050, +0.041]	-0.003 [-0.027, +0.020]
Wiki	-0.059 [-0.313, +0.195]	-0.177 [-0.566, +0.211]

B.7 Metrics and uncertainty

Equation 13 uses the same query and candidate IDs across every partition pair. Its primary family is $\Pi = \{P_1, P_{10}, P_{\text{micro}}\}$, giving three unordered pairs. The six-partition extension uses $P_1, P_2, P_5, P_{10}, P_{20}, P_{\text{micro}}$ and all 15 pairs; it is a separate estimand. We evaluate $k = 10, 20, 50$ without selecting a cutoff from the results. All candidates exceed 50 items. For an individual pair

with intersection size h, the recorded score is $h/(2k-h)$. We average these ratios over queries and pairs, rather than take a ratio of aggregated intersections and unions. Writing R^k for an ordered top-k prefix with retrieval direction suppressed, ordered agreement averages the indicator $\mathbf{1}\{R^k_{\mathcal{P},n}(q) = R^k_{\mathcal{Q},n}(q)\}$ over the same population. An additional set-agreement rate averages $\mathbf{1}\{h = k\}$; it distinguishes partial overlap from a completely reproduced result set.

Final-prefix results average the two directions at the last checkpoint. Stream-average results give equal weight to the four checkpoints and both directions. Each joint seed is aggregated separately before reporting the mean and sample SD over three seeds. Query counts, exclusions and every pairwise value accompany the aggregate results. Final-prefix P_{10} mAP is paired with three-pair TRR@20 for each dataset; query-input sensitivity uses fixed galleries and compares each noisy query with its clean counterpart. This top-k analysis re-scores previously observed evaluation trajectories, with the analysis rule fixed before computing TRR and with no model retuning.

For direction $g \in \{x \to y, y \to x\}$, let $I_{q,\mathrm{all}}(\mathcal{P}, \mathcal{Q})$ count pair inversions among the $M_{n,g}$ candidates, and let $\mathrm{MAP}^g_{\mathcal{P},n}$ be mean AP under partition $\mathcal{P}$ at count n in direction g. The full-ranking and quality gaps are

$$D_{\mathrm{K}}(R^g_{\mathcal{P},n}, R^g_{\mathcal{Q},n}) = \frac{1}{|\mathcal{Q}_{n,g}|} \sum_{q \in \mathcal{Q}_{n,g}} \frac{I_{q,\mathrm{all}}(\mathcal{P}, \mathcal{Q})}{\binom{M_{n,g}}{2}}, \tag{B.2}$$

$$\mathrm{PG}_{\mathrm{rank}} = \max_{\mathcal{P},\mathcal{Q}} \frac{1}{2|\mathcal{C}|} \sum_g \sum_{n \in \mathcal{C}} D_{\mathrm{K}}(R^g_{\mathcal{P},n}, R^g_{\mathcal{Q},n}), \tag{B.3}$$

$$\mathrm{PG}_{\mathrm{MAP}} = \max_{\mathcal{P},\mathcal{Q}} \frac{1}{2|\mathcal{C}|} \sum_g \sum_{n \in \mathcal{C}} \left| \mathrm{MAP}^g_{\mathcal{P},n} - \mathrm{MAP}^g_{\mathcal{Q},n} \right|. \tag{B.4}$$

The pair term is zero when $M_{n,g} < 2$. AP is evaluated on the shared eligible query set with at least one relevant candidate at the first checkpoint. Both maxima range over distinct partitions in the evaluated family Π.

For one checkpoint and direction, write $M = M_{n,g}$ and $\mathcal{Q} = \mathcal{Q}_{n,g}$. For query q, let $r_{qk} \in \{0, 1\}$ indicate any shared label at rank $k \in [M]$ and $R_q = \sum_{k=1}^M r_{qk}$. The reported full-list AP is

$$\mathrm{AP}(q) = \frac{1}{R_q} \sum_{k=1}^{M} r_{qk} \frac{\sum_{j \le k} r_{qj}}{k}, \qquad \mathrm{MAP} = |\mathcal{Q}|^{-1} \sum_{q \in \mathcal{Q}} \mathrm{AP}(q).$$

nDCG@100 uses relevance $s_{qk} = k_{\mathrm{D}}(\ell_q, \ell_{(k)})$, where ℓ_q labels the query and $\ell_{(k)}$ labels the rank-k item, and gain $2^{s_{qk}} - 1$, discounted by $\log_2(k+1)$ and normalized by the ideal top-100 gain. All eligible test queries are retained at every fixed checkpoint; Hamming ties use item ID. Means and sample SDs use the three paired joint seeds. Core intervention intervals use 10,000 hierarchical bootstrap draws over seeds and eligible queries, with Holm correction within each predefined family. Fixed-clock utility intervals use paired seed differences and a Student-t interval with two degrees of freedom.

B.8 Relations between retrieval-consistency metrics

For top-k sets A, B with r replacements, let J_k be their Jaccard overlap. Their intersection has $k - r$ items and union $k + r$, so

$$J_k = \frac{k-r}{k+r}, \qquad 1 - J_k = \frac{2r}{k+r}, \qquad \frac{|A \cap B|}{k} = \frac{2J_k}{1+J_k}. \tag{B.5}$$

For a single pair, $J_k = 1/2$ means two thirds of the items are shared; this nonlinear conversion does not commute with query averaging. Reversing the first $k > 1$ results leaves $J_k = 1$ but changes the ordered list. Exact full-ranking equality implies both set and ordered top-k equality, whereas equality at a fixed cutoff leaves the remainder of the ranking unrestricted. Because every Jaccard term is at most one and receives positive weight, $\mathrm{TRR}^{n,g}_k = 1$ if and only if all its evaluated sets agree. This is a finite-family statement; Theorem 1 supplies the guarantee over all legal taskifications under its assumptions.

Let I be the number of inversions between two full rankings of the same $M \geq 2$ uniquely identified candidates. Every item leaving the top-k set reverses its order relative to every item entering it. These r^2 distinct pairs give $I \geq r^2$. Consequently, with $u = \min\{k, M-k, \lfloor\sqrt{I}\rfloor\}$,

$$J_k \geq \frac{k-u}{k+u}, \qquad u = \min\{k, M-k, \lfloor\sqrt{D_{\mathrm{K}}\binom{M}{2}}\rfloor\}. \tag{B.6}$$

Here D_{K} is the single-query normalized Kendall distance, before query aggregation. For a large database, even a small normalized distance can permit $u = k$ and therefore a zero lower bound. Applying the bound query by query and then averaging preserves Equation 13's denominator. For numerical perturbations, Equation A.10 yields

$$I \leq \sum_{1 \leq i < j \leq M} \mathbf{1}\{|\Delta d_H(q; i, j)| \leq 2e(q) + e_d(i) + e_d(j)\}. \tag{B.7}$$

Pairs outside this margin retain their order under deterministic tie-breaking. Inserting this upper bound for I into Equation B.6 connects bit stability to top-k overlap. Here I counts all inversions; the relevant-irrelevant count I_q in the AP bound is insufficient because exchanging two relevant items can also change a top-k set.

Consider the two rankings $(1, \ldots, 1000)$ and $(11, \ldots, 20, 1, \ldots, 10, 21, \ldots, 1000)$. If items 1-20 are relevant, both have AP equal to one. Swapping the two ten-item blocks creates exactly 100 inversions, so $D_{\mathrm{K}} = 100/499500 \simeq 0.0002002$, yet their top-10 sets are disjoint and $J_{10} = 0$. This constructed example concerns metric properties, independently of the experimental datasets.

A deterministic but irrelevant ranking can attain TRR equal to one. mAP and nDCG describe its relevance, while TRR describes how consistently the same items are returned. The quality-consistency plane retains both outcomes without assigning an arbitrary exchange rate between them.

B.9 Discrete-gallery calibration

Use row-vector features in this subsection. For paired training matrices X, Y and ridge $\eta > 0$, define regression by $\mathcal{R}_\eta(X, Y) = (X^\top X + \eta I)^{-1} X^\top Y$. For gallery codes B_g, frozen gallery query head $W_{m,g}$, full-prefix labels L, modality features U_m and query u, define

$$\begin{aligned} A_m &= (U_m^\top U_m + \epsilon I)^{-1} U_m^\top L, & C_g &= (L^\top L + \epsilon I)^{-1} L^\top B_g, \\ h_{m,g,w}(u) &= \operatorname{sign}(\operatorname{norm}(uW_{m,g}) + w \operatorname{norm}(S_m(u) C_g)), \\ S_m(u) &= \sigma\big(\tau_{\mathrm{sem}}(uA_m - \tfrac{1}{2})\big), \end{aligned} \tag{B.8}$$

where the ridge is $\epsilon = 10^{-3}$, semantic sharpness is $\tau_{\mathrm{sem}} = 8$, and transfer weight $w \in \{0, 0.25\}$; σ is logistic, and norm is rowwise ℓ_2 normalization. Development mAP over three seeds selects w per gallery and direction, followed by one gallery selection. Both ridge maps use training-stream labels. The fixed-prefix maps and clocked gallery preserve pathwise equality.

For a gallery head W_g, the normalized direct logits are $a(u) = \operatorname{norm}(uW_g)$. The semantic logits are $e(u) = \operatorname{norm}(S_m(u) C_g)$ from Equation B.8, using training-stream labels to fit S_m, C_g. Define $f(u) = [a(u), e(u), 1]$. The compatible current head is either $a(u) + we(u)$, $w \in \{0, 0.25\}$, or $f(u)\mathcal{R}_\eta(F, B_g)$, where F stacks the training rows $f(u)$ and $\eta \in \{0.001, 0.1, 1, 10\}$; development data choose between these forms.

The alternate clocked learner supplies normalized logits A_o, aligned to the selected gallery by $\mathcal{R}_{0.1}(A_o, B_g)$. For keyed training and anchor index sets $Q, A \subseteq [n]$, write $S_{QA} = 2k_{\mathrm{D}}(L_Q, L_A) - 1$ entrywise. The pairwise-compatible head regresses F_Q, which stacks $f(u)$ for rows Q, onto targets formed from anchor codes $B_A = (B_g)_{A,:}$:

$$T_Q = (S_{QA} B_A / b)(B_A^\top B_A / b^2 + 0.1I)^{-1}, \qquad W_p = \mathcal{R}_{0.1}(F_Q, T_Q).$$

Let $c_q, o_q^{\mathrm{alt}}, p_q$ be the row-normalized outputs of the current, aligned alternate and pairwise heads. The residual input is the 321-D vector $v_q = [c_q, o_q^{\mathrm{alt}}, p_q, c_q \odot o_q^{\mathrm{alt}}, |c_q - o_q^{\mathrm{alt}}|, 1]$ at 64 bits, where $\odot$ and absolute value act elementwise. Its base logits are $z_q = c_q + \omega p_q$, with $\omega \in \{0, 0.25, 0.5, 1\}$ selected before residual training. The linear residual has no additional bias; the MLP uses a 256-unit

GELU hidden layer and a zero-initialized output layer. AdamW uses learning rate 0.002 and zero weight decay.

For binarization sharpness κ and row-vector anchor score $a_{qj} = \tanh(\kappa(z_q + r_\theta(v_q)))B_{g,j}^\top/b$, let P_q contain anchors sharing a label with query q and N_q the remaining anchors. Define $\pi_{qj} = k_{\mathrm{D}}(\ell_q, \ell_j)/\sum_a k_{\mathrm{D}}(\ell_q, \ell_a)$, with the sum over anchors. At ranking temperature τ_{rank}, training uses

$$\mathcal{L}_{\mathrm{bin}} = \tau_{\mathrm{rank}} \operatorname{softplus}\left(\mathrm{LME}_{j\in N_q} \frac{a_{qj}}{\tau_{\mathrm{rank}}} + \mathrm{LME}_{j\in P_q} \frac{-a_{qj}}{\tau_{\mathrm{rank}}}\right), \tag{B.9}$$

$$\mathcal{L}_{\mathrm{grade}} = -\sum_j \pi_{qj} \log \operatorname{softmax}_j(a_{qj}/\tau_{\mathrm{rank}}), \tag{B.10}$$

where $\mathrm{LME}_{j\in E}\, t_j = \log(|E|^{-1}\sum_{j\in E}\exp(t_j))$, $\operatorname{softplus}(t) = \log(1+e^t)$, and softmax normalizes exponentiated anchor scores to sum to one. With $o_q = \tanh(\kappa(z_q + r_\theta(v_q)))$ and $o_q^0 = \tanh(\kappa z_q)$, the optimized loss is

$$\lambda_{\mathrm{rank}}(\mathcal{L}_{\mathrm{bin}} + 0.2\,\mathbf{1}_{\mathrm{grade}}\mathcal{L}_{\mathrm{grade}}) + 0.05\|o_q - o_q^0\|_2^2/b + 0.01\||o_q| - \mathbf{1}\|_2^2/b + 0.001\|r_\theta(v_q)\|_2^2/b, \tag{B.11}$$

where $\mathbf{1}_{\mathrm{grade}}$ enables the graded loss, $\mathbf{1}$ is the all-ones vector, $\lambda_{\mathrm{rank}} \in \{0.05, 0.2\}$ weights ranking, and κ increases linearly from 1 to 5 over the 15 training epochs. Ranking losses average training examples having at least one positive and one negative anchor. Stability, quantization and residual penalties average all sampled examples and bits. The calibration trains on at most 4,096 keyed stream examples against 2,048 keyed gallery anchors for 15 epochs (batch 256, $\tau_{\mathrm{rank}} = 0.08$). The candidate set contains linear/MLP residuals, binary/graded ranking objectives, rank weights $\{0.05, 0.2\}$ and scales $\{0.1, 0.25, 0.5, 0.75, 1\}$. A 256-query development screen chooses feasible candidates; the disjoint remainder must repeat a 0.10-point mAP gain with at least two positive seeds and at most a 0.10-point nDCG@100 loss, otherwise the frozen base head is retained. The standard test scoring follows the frozen selector. Because these test IDs were used in earlier diagnostic analyses, this matrix is reported as retrospective and the candidate set remains fixed after scoring. Integer rescoring covers 15 seed-dataset cells under $P_1, P_{10}, P_{\mathrm{micro}}$; the 90 direction-partition-pair comparisons yield identical lists at $k = 10, 20, 50$. The development rule retains the base head for NUS-WIDE T→I and Wiki I→T, and selects a residual in the other eight directions.

B.10 Noise-aware gallery alignment

Training with zero-mean input noise adds a quadratic regularizer (Bishop, 1995). Let $\Sigma = \rho^2 \operatorname{diag}(I_d, 0)/d$, where ρ is training-noise strength and $d = 512$ excludes the bias. Rows of Ξ are independent, zero-mean noise with covariance Σ. For the opposite-modality historical codes $B \in \{-1, +1\}^{n\times b}$ and semantic affinities $S_{ij} = 2k_{\mathrm{D}}(\ell_i, \ell_j) - 1$ for $i, j \in [n]$, use gallery-alignment weight $\gamma \geq 0$ to refine the objective:

$$\mathbb{E}_\Xi\left[\mathcal{F}_m(W; U_m + \Xi) + \frac{\gamma}{n^2 b^2}\|(U_m + \Xi)WB^\top - bS\|_F^2\right]. \tag{B.12}$$

The asymmetric term aligns query predictions to the actual gallery rather than regenerating its codes (Jiang & Li, 2018). Write $H_B = B^\top B/n$ and $D_S = U_m^\top SB/n^2$. Independence and zero mean give $\bar{G}_m = G_m + \Sigma$ and leave cross-moments unchanged, yielding

$$\begin{aligned} \bar{G}_m WA + \tfrac{\lambda}{2}W &= R, \\ A &= \beta H/b^2 + \alpha I/b + \gamma H_B/b^2, \\ R &= \beta D_m H/b^2 + \alpha D_m/b + \gamma D_S/b. \end{aligned} \tag{B.13}$$

Diagonalizing $\bar{G}_m$ and A gives the unique solution as in Equation 7. At $\gamma = 0$ this recovers noise-only regularization. The expectation is exact for additive training noise; the evaluation perturbation additionally normalizes each noisy feature.

For label cardinalities $a_i = \|\ell_i\|_0$, the product SB is computed without storing pairwise similarities:

$$(SB)_i = 4\sum_{r=1}^{d_\ell} \frac{\ell_i^\top \sum_{j:a_j=r} \ell_j B_j}{a_i + r} - \sum_j B_j.$$

The additional training-prefix pass fits the query head and preserves gallery codes.

The development search uses all eligible queries, both directions and three seeds, with $\rho \in \{0, 0.025, 0.05, 0.1, 0.2, 0.4\}$ for noise alone and $\gamma \in \{0.1, 1, 10, 100\}$ at $\rho \in \{0.2, 0.4\}$ for gallery alignment, followed by the logarithmic midpoint $\gamma = \sqrt{10}$ at $\rho = 0.4$. Selection requires improved clean/noisy mAP and nDCG means in every direction, two positive stability seeds and a two-point overlap gain. It chooses $(\rho, \gamma) = (0.4, \sqrt{10})$ with $\alpha = \beta = 10$, $\lambda = 10^{-5}$.

B.11 Denoising the calibrated query stages

For row features $u \in \mathbb{R}^d$ ($d = 512$), training-feature covariance C and mean μ, denoising uses

$$T_\rho = (C + \rho^2 I/d)^{-1} C, \qquad \widetilde{u} = \mathrm{normalize}(\mu + (u - \mu)T_\rho)\,.$$

For zero-mean isotropic noise with covariance $\rho^2 I/d$, the linear map minimizes expected squared reconstruction error (Chen et al., 2012; Bishop, 1995). Here normalize divides by the Euclidean norm; $\rho = 0$ returns the original input.

Development tests $\rho \in \{0, 0.005, 0.01, 0.025, 0.05, 0.1, 0.2\}$ over all eligible queries and three seeds. Selection preserves clean mAP/nDCG and noisy nDCG means, requires two positive stability seeds and a two-point overlap gain, then maximizes noisy mAP. The required noisy-mAP gain is 2 points for MIRFlickr T→I and 0.1 elsewhere, fixed before each search. The retained MIRFlickr image-query denoiser uses $\rho = 0.2$; text queries use $\rho = 0.2$ during correction training and 0.4 at inference, as specified below. IAPR image queries denoise the alternate-clock expert at $\rho = 0.1$; text queries use the prediction-preserving map below.

Let P project onto the column space of the native and semantic heads, excluding their bias rows. Minimizing expected linear reconstruction error subject to $TP = P$ gives

$$T_{\rho,P} = (C + \lambda I)^{-1}(C + \lambda P), \qquad \lambda = \rho^2/d.$$

Indeed, $(C + \lambda I)T_{\rho,P} - C = \lambda P$, which is orthogonal to every feasible matrix perturbation E with $EP = 0$ in the Frobenius inner product. Strict convexity for $\lambda > 0$ gives uniqueness. The affine transformation $\mu + (u - \mu)T_{\rho,P}$ preserves the protected head outputs; it omits unit renormalization, which would break that identity. Development compares no, native-only and joint protection over the six positive strengths, selecting joint protection at $\rho = 0.01$ for typical IAPR text queries. Motivated by feature-space detection (Lee et al., 2018), a residual-energy detector selects an unprotected map for atypical inputs. A deterministic 80/20 training split fits the lowest 128 covariance-eigenvalue directions and calibrates their mean squared projection score. Development searches quantiles $\{0.90, 0.95, 0.99\}$ and unprotected strengths $\{0.01, 0.05, 0.1\}$, selecting 0.90 and $\rho = 0.1$ under the same gate. Fitting adds training passes and $O(d^2)$ state.

B.12 Composed query correction

Binary relevance treats any shared label as a match, whereas graded nDCG distinguishes degrees of semantic overlap. We therefore refine the MIRFlickr text-query head using both metric-aware gradients and binary training targets, leaving the 18,004-item gallery fixed. Let z_q^0 denote the complete frozen calibrated logits and $v_q \in \mathbb{R}^{321}$ its expert features, evaluated after denoising. For anchors sorted by immutable ID, define $g_{qj} = 2^{k_{\mathrm{D}}(\ell_q, \ell_j)} - 1$ and $d(r) = \mathbf{1}[r \le 100]/\log_2(r+1)$, the rank-$r$ discount. Let r_i be anchor i's rank and $\mathrm{IDCG}_{q,100}$ the ideal top-100 discounted gain. At the current binary ranking, the absolute nDCG change from swapping anchors i, j is

$$\delta_{qij} = \frac{|g_{qi} - g_{qj}|\,|d(r_i) - d(r_j)|}{\mathrm{IDCG}_{q,100}}.$$

We use these detached weights in a sampled pairwise logistic loss $\mathcal{L}_\Delta$, following metric-weighted ranking (Wang et al., 2018). Each of the first 100 anchors pairs with four gain-ranked and four ID-ranked partners, cycled deterministically by optimization step. Self-pairs and equal-gain pairs have zero weight; loss weights normalize within each query. For residual weights $W_\Delta \in \mathbb{R}^{321\times 64}$, relaxed scores are $\tanh(\kappa(z_q^0 + v_q W_\Delta))B_j^\top/64$, where B_j is anchor j's code, with temperature 0.08 and κ increasing from 1 to 5.

Training-gradient measurements identify an imbalance between the binary and graded objectives. Gradient magnitudes provide a way to calibrate their scales (Chen et al., 2018). Here the factor is static: $a = 6.7831$ is the median ratio of their weighted gradient norms on 24 training-only diagnostic batches across three seeds and two sharpnesses. Writing $o_q = \tanh(\kappa(z_q^0 + v_q W_\Delta))$ and $o_q^0 = \tanh(\kappa z_q^0)$, the selected objective is

$$0.2\mathcal{L}_{\text{bin}} + 0.04a\mathcal{L}_\Delta + 0.05\|o_q - o_q^0\|_2^2/64 + 0.001\|v_q W_\Delta\|_2^2/64,$$

averaged over training queries. The separate quantization penalty is omitted after the fixed-factor, penalty-on/off development comparison. Zero-initialized W_Δ trains on all 18,004 stream items for 15 epochs, batch size 128, with AdamW learning rate 0.002, zero weight decay and 2,048 ID-keyed anchors: 2,115 updates per seed.

The second residual learns from exact binary improvements on those training anchors. Starting from $\text{sign}(z_q^0)$, two full coordinate sweeps accept a bit flip only if anchor AP does not decrease and nDCG@100 strictly increases, excluding the query's own ID. Let b_q^* be the resulting target and define $D_{qr} = 0$ for unchanged bits $r \in [b]$, otherwise $D_{qr} = b_{qr}^* \max(|z_{qr}^0|, 10^{-6}) - z_{qr}^0$. Stack training rows v_q and target increments D_q into V and D, respectively. With n training rows, ridge regression fits

$$W_D = (V^\top V/n + \eta I)^{-1}(V^\top D/n), \qquad h(u) = \text{sign}\big(z^0(\widetilde{u}) + v(\widetilde{u})\,[\tfrac{1}{2}W_\Delta + \tfrac{1}{2}W_D]\big)\,.$$

The fitted components use training inputs denoised at $\rho = 0.2$. Development selects $\eta = 0.01$ from $\{10^{-4}, 10^{-3}, 10^{-2}\}$ and the displayed residual scales from $\{0.1, 0.25, 0.5, 1\}$. They remain fixed while the inference denoiser tests $\rho \in \{0.2, 0.225, 0.25, 0.3, 0.4\}$. The complete six-severity development grid selects $\rho = 0.4$: candidates must improve clean mean nDCG by at least 0.25 point with three positive seeds, preserve AP, nDCG and top-20 overlap means at every severity, and retain at least 95% of clean bit entropy and distinct code count in every seed. Among passing candidates, the minimum mean overlap gain over nonzero severities determines selection. The gain-normalized cross-entropy control uses the same training budget; the correction must also preserve its clean AP and improve its clean nDCG by at least 0.1 point. Inference uses the query features and these fitted maps. Final-prefix calibration costs are separate from the analytic core's streaming cost.

C Additional Experimental Results

C.1 Top-k agreement across taskifications

Figure 5 reports set agreement at three retrieval depths. The analytic core and clocked LEMON reproduce both sets and ordered lists at all three cutoffs.

Trajectories requiring per-query reruns use Python 3.10.12, NumPy 2.2.6, PyTorch 2.7.1 and Numba 0.61.2 with fixed model settings. Relevance and TRR come from the same trajectories. OCMFH retains checkpoint-specific rewrites; Joint-full re-encodes each prefix. The primary and six-taskification cohorts contain 927 distinct trajectories in 309 paired groups.

Over all 15 pairs of six taskifications, the Mode-T core again has identical sets and ordered lists. The six structural controls span 29.84-73.15% mean TRR@20 across the four datasets. Table 10 separates the 72 mechanism-study cells by taskification. PCA-ITQ leads nDCG@100 under P_1 but loses this advantage when smaller tasks re-estimate its targets. Of 35,880 ordered-list comparisons per method, the core matches every list, PCA-ITQ matches none, within-task state matches 368 and boundary SGD 1,616.

C.2 Query-feature perturbation stress

Input perturbations test a different source of retrieval sensitivity from task boundaries (Aghazadeh et al., 2017). We use all eligible MIRFlickr and IAPR test queries against their complete final-prefix galleries, with the existing P_{10} heads of LEMON, DOCH, the analytic core and clocked LEMON, plus the final calibrated discrete pipeline. The three original joint seeds and both directions are retained. For unit feature u, the intervention is

$$u_\sigma = \frac{u + \sigma v}{\|u + \sigma v\|_2}, \qquad v = \frac{z}{\|z\|_2}, \quad z \sim \mathcal{N}(0, I), \qquad \sigma \in \{0, 0.01, 0.025, 0.05, 0.1, 0.2\}.$$

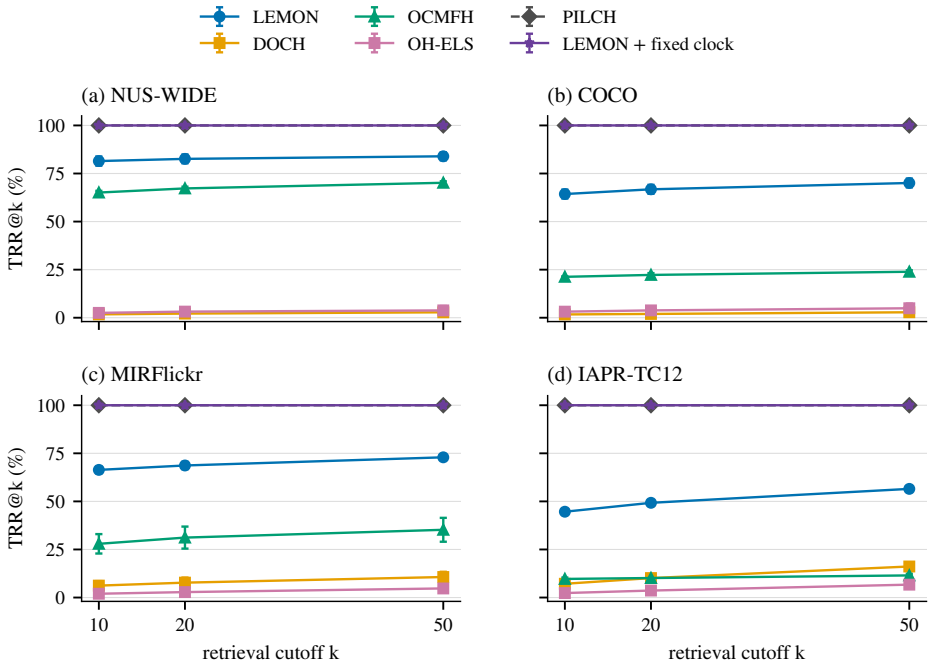


Figure 5: Final-prefix TRR versus retrieval cutoff at 64 bits. Each point averages all three taskification pairs and both directions; error bars show sample SD across three seeds. The core and clocked LEMON coincide at 100% for all cutoffs.

Table 10: Partition-specific relevance for the IAPR Mode-T study (64 bits). Each column pair gives mAP / nDCG@100 (%), averaged over directions and three seeds. Column ranks: **Best** 2nd 3rd; ties share rank at the shown precision.

Method	P_1		P_{10}		P_{micro}	
PILCH (T)	**57.65**	49.53	**57.65**	**49.53**	**57.65**	**49.53**
Taskwise PCA-ITQ (T)	55.46	**51.81**	35.54	20.28	31.56	14.18
Within-task state (T)	57.55	49.43	49.93	38.77	39.68	21.31
Boundary SGD (T)	57.39	47.93	56.79	47.19	56.77	46.83

Noise is keyed by dataset, query ID, direction and seed, shared across methods and severities. The intercept, gallery and tie rule stay fixed; $\sigma = 0$ recovers the saved clean codes. Each positive level re-encodes every query with its frozen head. Table 11 reports the final refined pipeline at all six severities under the retrospective protocol.

Figure 6 reports overlap and paired relevance gains. QR reduces changed query bits from 1.90/3.75% to 1.54/2.44% on MIRFlickr/IAPR at $\sigma = 0.1$, accompanying its relevance gains. For MIRFlickr text queries, the final composed correction increases mean mAP from 89.93% to 90.04%, nDCG@100 from 53.64% to 54.00%, and top-20 overlap from 63.47% to 66.42%. Every nonzero seed-severity overlap also increases. The stricter ordered-list statistic has two small seed-level decreases, while its six severity means increase or remain equal. These paired results show improved graded retrieval and set stability on the evaluated feature perturbations.

The same frontend is also applied to the frozen LEMON and DOCH heads, using $\rho \in \{0.2, 0.4\}$ and all six severities. The 72 cells retain all 2,278 queries and 18,004 gallery items. Table 12 compares the pipelines at $\sigma = 0.1$. Raising ρ improves DOCH's overlap from 60.31% to 64.20%, almost closing its gap to the corrected pipeline's 66.42%. LEMON reaches 75.30% overlap and 59.65% nDCG, above the corrected pipeline's 54.00% nDCG; the latter has higher mAP, 90.04% versus 86.83%. Thus preprocessing explains part of the stability gain, while the remaining ranking differences depend on the query head and gallery.

Table 11: Current refined queries across all six feature-perturbation levels (%, mean ± sample SD; three seeds). Overlap compares each noisy ranking with its own clean ranking; at $\sigma = 0$ it is the identity comparison. Every cell uses the complete eligible query set and final gallery.

Dataset	Dir.	σ	mAP	nDCG@100	Top-20 overlap
MIRFlickr	I→T	0	95.46 ± 0.10	66.19 ± 0.44	100.00 ± 0.00
		0.01	95.46 ± 0.10	66.20 ± 0.44	96.49 ± 0.16
		0.025	95.46 ± 0.10	66.19 ± 0.43	91.91 ± 0.35
		0.05	95.47 ± 0.08	66.16 ± 0.44	84.53 ± 0.42
		0.1	95.46 ± 0.11	66.14 ± 0.50	73.59 ± 1.25
		0.2	95.42 ± 0.11	65.99 ± 0.58	57.69 ± 0.73
MIRFlickr	T→I	0	90.08 ± 0.11	54.09 ± 0.25	100.00 ± 0.00
		0.01	90.08 ± 0.11	54.09 ± 0.24	95.48 ± 0.53
		0.025	90.07 ± 0.12	54.07 ± 0.20	88.99 ± 0.53
		0.05	90.07 ± 0.12	54.07 ± 0.17	80.08 ± 0.67
		0.1	90.04 ± 0.12	54.00 ± 0.19	66.42 ± 0.69
		0.2	89.91 ± 0.13	53.84 ± 0.22	47.94 ± 1.47
IAPR-TC12	I→T	0	71.01 ± 0.17	53.16 ± 0.34	100.00 ± 0.00
		0.01	70.98 ± 0.17	53.14 ± 0.35	93.37 ± 0.41
		0.025	70.95 ± 0.19	53.08 ± 0.41	84.74 ± 0.36
		0.05	70.87 ± 0.21	53.00 ± 0.37	73.41 ± 0.89
		0.1	70.59 ± 0.19	52.57 ± 0.30	58.21 ± 0.76
		0.2	69.73 ± 0.27	51.51 ± 0.34	39.72 ± 0.60
IAPR-TC12	T→I	0	68.73 ± 0.14	49.78 ± 0.28	100.00 ± 0.00
		0.01	68.58 ± 0.14	49.66 ± 0.37	70.66 ± 1.74
		0.025	68.39 ± 0.13	49.51 ± 0.44	59.11 ± 1.12
		0.05	68.09 ± 0.13	49.19 ± 0.31	50.53 ± 0.81
		0.1	67.49 ± 0.10	48.43 ± 0.37	41.97 ± 0.82
		0.2	67.53 ± 0.13	48.14 ± 0.54	44.93 ± 0.53

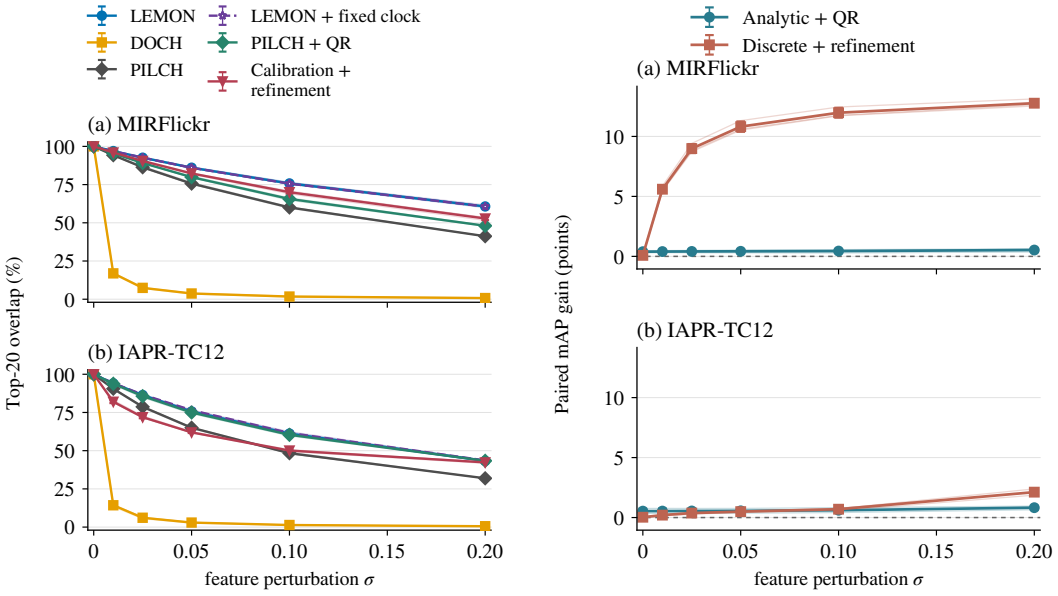


Figure 6: Left: cross-method clean/noisy top-20 Jaccard on the full query populations. Right: paired mAP gains against each branch's pre-refinement head with its gallery fixed. All datasets, directions, severities and seeds are retained. Thin curves average directions within each seed; markers and bars give three-seed means and sample SDs.

Table 12: Matched denoising controls on MIRFlickr T→I at $\sigma = 0.1$ (%, mean ± sample SD; three seeds). All methods use the same training-covariance frontend and inference strength ρ within each strength group. Heads and galleries retain their respective training procedures, so this control matches preprocessing. **Best** 2nd 3rd among the three pipelines inside each strength group, at the shown precision.

Query pipeline	ρ	mAP	nDCG@100	Top-20 overlap
LEMON	0.2	86.91 ± 0.10	**59.87** ± 0.42	**73.32** ± 0.68
DOCH	0.2	89.56 ± 0.10	53.44 ± 0.23	60.31 ± 0.19
Calibration + denoising	0.2	**89.93** ± 0.12	53.64 ± 0.29	63.47 ± 0.86
LEMON	0.4	86.83 ± 0.12	**59.65** ± 0.51	**75.30** ± 1.79
DOCH	0.4	89.56 ± 0.05	53.43 ± 0.20	64.20 ± 0.19
Calibration + correction	0.4	**90.04** ± 0.12	54.00 ± 0.19	66.42 ± 0.69

C.3 Current directional retrieval

Table 13 combines the unchanged NUS-WIDE, COCO and Wiki heads with the current MIRFlickr and IAPR query refinements. It reports the complete three-seed directional matrix. NUS-WIDE T→I retains losses of 0.071/0.095 Precision@50/100 points relative to its comparator, although mean mAP improves. Four of the 30 seed-direction mAP effects are negative: seeds 1 and 2 for NUS-WIDE T→I and Wiki I→T. The refinements on the other datasets do not change these outcomes.

Table 13: Current final-prefix retrieval at 64 bits (%, mean ± sample SD; three seeds). MIRFlickr and IAPR include the selected query refinements. NUS-WIDE, COCO and Wiki retain the frozen calibration. Historical gallery codes are unchanged.

Dataset	Dir.	mAP	nDCG@100
NUS-WIDE	I→T	87.00 ± 0.24	66.89 ± 0.57
	T→I	80.88 ± 0.04	59.80 ± 0.27
COCO	I→T	83.63 ± 0.20	65.97 ± 0.24
	T→I	85.28 ± 0.17	70.90 ± 0.44
MIRFlickr	I→T	95.46 ± 0.10	66.19 ± 0.44
	T→I	90.08 ± 0.11	54.09 ± 0.25
IAPR-TC12	I→T	71.01 ± 0.17	53.16 ± 0.34
	T→I	68.73 ± 0.14	49.78 ± 0.28
Wiki	I→T	57.45 ± 1.60	50.01 ± 1.77
	T→I	87.26 ± 0.86	83.51 ± 1.46

The retrieval visualization uses current MIRFlickr seed-0 refined query codes and the 18,004-item clocked-DOCH gallery. Each label stratum contains 75 stable-hash-selected items; the rarest positive test label assigns multi-label items, with label-index ties. Shared item IDs enter independent PCA-50 and t-SNE fits (perplexity 30, 1,000 iterations, seed 5603). The query is the minimum stable-hash item among flower test queries whose labels are confined to flower and plant life, with three to eight tags including flower or flowers. Selection uses annotations and identifiers, without AP or the retrieved images. Top-30 images follow Hamming distance and item-ID ties without reranking or cropping. Integer scoring covers both 2,278-query directions.

The external XMediaNet study tests the permanent-target core, separately from fixed-gallery query calibration. The mAP gains are 12.150/2.943 points, with paired intervals $[10.701, 13.599]$/$[1.247, 4.639]$; nDCG@100 gains are 13.689/2.926 points.

C.4 CODE LENGTH

Table 14 gives the paired LEMON clock intervention at 16, 32, 64 and 128 bits, using three seeds and the stated database writer. These results concern the uncalibrated learners; the final calibrated-discrete matrix uses 64 bits.

Table 14: Complete source-corrected code-length comparison at P_{10}. Fixed clock and LEMON columns report bidirectional mAP mean±sample SD; directional columns give paired mAP differences (fixed clock minus LEMON), also mean±sample SD, over three seeds. All values are percentage points.

Dataset	Bits	Fixed clock	LEMON	Δ I2T	Δ T2I
NUS-WIDE	16	79.816±0.146	79.830±0.159	-0.021±0.021	-0.007±0.005
	32	80.896±0.525	80.899±0.498	-0.008±0.026	0.001±0.031
	64	81.844±0.328	81.847±0.326	-0.011±0.009	0.005±0.009
	128	82.132±0.388	82.141±0.382	-0.012±0.006	-0.006±0.006
COCO	16	74.908±0.247	74.958±0.254	-0.047±0.025	-0.054±0.025
	32	78.388±0.099	78.505±0.084	-0.099±0.030	-0.136±0.034
	64	81.788±0.191	82.021±0.186	-0.207±0.014	-0.259±0.010
	128	83.538±0.197	83.840±0.153	-0.277±0.049	-0.327±0.047
MIRFlickr	16	88.309±0.247	88.318±0.236	-0.007±0.009	-0.011±0.034
	32	89.253±0.222	89.211±0.200	0.027±0.036	0.057±0.027
	64	89.789±0.083	89.784±0.091	-0.015±0.029	0.026±0.005
	128	89.918±0.107	89.880±0.097	0.039±0.015	0.038±0.014
IAPR-TC12	16	61.003±0.201	61.017±0.253	-0.040±0.086	0.013±0.053
	32	64.923±0.177	64.908±0.191	-0.001±0.039	0.030±0.044
	64	67.570±0.129	67.588±0.062	-0.003±0.113	-0.032±0.048
	128	69.050±0.148	69.112±0.202	-0.062±0.019	-0.062±0.092

C.5 ANALYTIC-CORE INTERVENTIONS

Table 15: Cross-dataset mechanism effects at 64 bits and P_{10}. Entries are full minus ablation mAP points with paired hierarchical 95% bootstrap intervals; positive favors the full method.

NUS-WIDE		
ID	Intervention	Effect [95% CI]
A1	Taskwise PCA-ITQ targets	+21.093 [+20.535, +21.960]
A2	Within-task statistics	+0.432 [+0.317, +0.523]
A3	Boundary-local SGD	+0.365 [-0.418, +1.064]
A5	No relation term	+0.004 [-0.001, +0.008]
A9	Random permanent target	+49.268 [+48.255, +51.025]

Table 15: Cross-dataset mechanism effects (continued).

COCO		
ID	Intervention	Effect [95% CI]
A1	Taskwise PCA-ITQ targets	+33.958 [+33.360, +34.428]
A2	Within-task statistics	+0.684 [+0.611, +0.753]
A3	Boundary-local SGD	+0.499 [+0.462, +0.537]
A5	No relation term	-0.003 [-0.006, +0.000]
A9	Random permanent target	+44.312 [+43.865, +44.717]

MIRFlickr		
ID	Intervention	Effect [95% CI]
A1	Taskwise PCA-ITQ targets	+20.127 [+18.913, +20.889]
A2	Within-task statistics	+3.189 [+2.958, +3.460]
A3	Boundary-local SGD	+0.235 [+0.154, +0.321]
A5	No relation term	-0.005 [-0.012, +0.001]
A9	Random permanent target	+33.219 [+32.802, +33.678]

IAPR-TC12		
ID	Intervention	Effect [95% CI]
A1	Taskwise PCA-ITQ targets	+22.108 [+21.334, +22.732]
A2	Within-task statistics	+7.720 [+7.428, +7.991]
A3	Boundary-local SGD	+0.859 [+0.711, +1.009]
A5	No relation term	-0.005 [-0.014, +0.005]
A9	Random permanent target	+27.266 [+26.628, +27.810]

Table 15 contains the matched Mode-T effects. A1 re-centers fixed Dice scores within each task, fits task-specific PCA coordinates and learns a separate ITQ rotation before thresholding. This changes the targets as well as their coordinates. A2 resets prefix state, A3 moves optimization to task boundaries, A5 removes the relation term and A9 replaces semantic targets by random permanent codes. A3's interval includes zero on NUS-WIDE; A5's intervals include zero on all four datasets. A2 is the task-local learner used in the state-accumulation comparison, not an independent second intervention.

C.6 Prefix recovery and numerical precision

At 25/50/75/100% of each stream, 13 nonempty ID-keyed shards are reduced left-to-right, in reverse and by a balanced tree, then compared with single-batch accumulation. Shared settings are 64 bits, FP64, biased CLIP features, permanent Dice targets, $\alpha = \beta = 10$ and $\lambda = 10^{-5}$. Four datasets and three seeds yield 288 modality-level comparisons of the five moment matrices, heads and all query codes. Reported maxima retain every comparison. The shared Mode-T gallery connects query-code equality to ranking equality; the main stream study tests historical Mode-E writes.

Against FP64, FP32 query-bit flip rates are 1.84×10^{-6} on NUS-WIDE and 6.28×10^{-7} on COCO, with top-100 set overlap 0.999985 and 0.999997. Scaled FP16 raises flip rates to 0.415 and 0.319 and lowers overlap to 0.185 and 0.169. The main protocol therefore uses FP64 statistics and solves.

C.7 CACHED-FEATURE COST

Table 16: Cached-feature update and evaluation wall seconds: median above [Q1, Q3] over five isolated repeats at seed s0. State counts the implementation's retained arrays; the analytic core counts prefix moments. Backbone extraction, fixed maps, query heads and binary database storage sit outside that core state count. **Best** 2nd 3rd in each column with lower is better, at the shown precision.

Method	NUS-WIDE	COCO	MIRFlickr	State (MiB)
DOCH	3378.1 [3310.4, 3762.1]	1829.9 [1827.4, 1845.4]	138.2 [138.0, 138.5]	14.10
OCMFH	102.5 [101.8, 103.3]	61.3 [61.2, 61.4]	10.3 [10.1, 10.4]	82.36
Local update	68.4 [68.4, 68.5]	53.3 [53.2, 53.4]	7.4 [7.4, 7.4]	5.55
Mixed replay	75.3 [75.3, 75.7]	57.7 [57.6, 57.7]	8.2 [8.1, 8.2]	6.05
CT-CMH	72.4 [72.3, 72.6]	55.7 [55.6, 55.8]	7.8 [7.8, 8.0]	6.05
DLCH	74.3 [74.3, 74.4]	57.0 [57.0, 57.3]	8.1 [8.0, 8.2]	6.56
SILCH	77.4 [77.3, 77.5]	59.4 [59.3, 59.4]	8.6 [8.4, 8.6]	6.56
MCDLCH	77.8 [77.4, 78.0]	59.5 [59.3, 59.9]	8.7 [8.5, 8.8]	6.56
PILCH	**67.3 [67.3, 67.5]**	**52.1 [51.9, 52.2]**	**7.2 [7.2, 7.2]**	**4.55**

With cached features, the analytic core takes 67.3/52.1/7.2 seconds on NUS-WIDE/COCO/MIRFlickr and reports 4.55 MiB of prefix moments, as counted by the implementation's summary-byte function. The measurement covers prefix moments; heads and the semantic map add storage. Timings use five isolated repeats. The binary database grows by eight bytes per stored 64-bit code. The timings exclude backbone extraction and query calibration; the storage count excludes the feature cache. Table 16 reports the same measurement for every compared method, so the speed ratios quoted in Section 4.4 can be read off directly.

Manifests retain run- and query-level metrics, analysis populations and content hashes. Small-prefix tests compare the sufficient-statistic solution with an explicit quadratic oracle. The final calibrated-discrete records contain 330 seed-direction-metric rows across five datasets; raw trajectory records retain the checkpoint-level and taskification-pair measurements.