CROSS-DOMAIN CROSS-MODAL HASHING VIA SEMANTIC RELATION TRANSFER AND INVERSE-FREE BROAD LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Cross-domain cross-modal hashing aims to leverage semantically rich knowledge from an auxiliary domain to improve cross-modal retrieval in an unlabeled or weakly labeled target domain. Existing methods face two major challenges: the lack of reliable semantic supervision in the target domain, and the increased complexity of cross-domain transfer caused by modality discrepancies and disjoint label spaces, which further impedes the exploitation of semantic information. To address these challenges, we propose a hashing framework based on semantic relation transfer and QR-based incremental learning. For cross-domain semantic transfer, we learn transferable cross-modal representations from the semantically rich auxiliary domain and bridge the auxiliary and target domains through parameter transfer and cross-domain semantically aligned samples. Meanwhile, we construct an auxiliary-domain semantic relation graph and a weak semantic relation graph for the target domain using CLIP-generated pseudo-labels. By jointly modeling semantic relations across the two domains, our method provides effective semantic guidance for the supervision-deficient target domain, thereby facilitating knowledge transfer and target-domain adaptation. For representation learning, we introduce an incrementally extensible broad learning-based hashing layer that dynamically expands the representation space by appending feature nodes and enhancement nodes. We further develop an incremental QR decomposition-based update of the R factor and employ triangular solves to analytically compute the expanded-layer weights, eliminating the need for the explicit pseudoinverse computation required by conventional broad learning methods. By integrating auxiliary-to-target knowledge transfer with incremental target-domain learning in a lightweight hashing framework, together with dual semantic relation graph modeling, the proposed method simultaneously enables cross-domain knowledge transfer and target-domain semantic enhancement. Extensive experiments on two benchmark datasets demonstrate significant improvements over existing methods, validating the effectiveness and superiority of the proposed framework.

1 INTRODUCTION

Cross-modal hashing maps heterogeneous modalities into a shared compact Hamming space, and has become an important paradigm for large-scale image-text retrieval, cross-modal semantic matching, and multimedia information retrieval (Yang et al., 2020; Zeng et al., 2022; Li et al., 2024b; Shen et al., 2024; Yang et al., 2025). Broad Learning Systems (BLS) exploit horizontal structural expansion and efficient incremental learning, offering advantages in rapid representation learning, incremental model updating, and lightweight computation (Chen & Liu, 2018).

However, real-world applications inevitably involve substantial distributional variations across domains. (Wang & Deng, 2022; Hu et al., 2019; Josifovski et al., 2025) Differences in acquisition conditions, content distributions, semantic categories, and modality-specific expressions make cross-modal representations learned within a single domain difficult to generalize to new retrieval environments. To bridge this gap, cross-domain cross-modal hashing seeks to leverage semantically rich auxiliary domains to guide learning in supervision-scarce target domains, enabling heterogeneous modalities across domains to preserve consistent semantic structures in a shared Hamming

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.

space (Li et al., 2024a). Existing approaches mainly follow two directions: one transfers knowledge from the auxiliary domain through parameter transfer, feature alignment, or relation transfer; the other exploits target-domain semantics through sample similarities, pseudo-labels, or pretrained vision-language models to construct weak supervisory signals. The former provides external semantic priors, whereas the latter captures the intrinsic structure of the target domain.

Despite continued progress, existing methods remain constrained by the difficulty of simultaneously achieving reliable semantic transfer and efficient target-domain adaptation. First, unlabeled target domains lack direct semantic supervision, and relying solely on sample similarity or individual pseudo-labels is insufficient to capture complex cross-modal relationships (Yu et al., 2022). Second, modality heterogeneity, distribution shifts, and mismatched label spaces between auxiliary and target domains make direct transfer strategies difficult to establish reliable semantic correspondences. Meanwhile, although Broad Learning Systems (BLS) provide efficient structural adaptation through horizontal node expansion, conventional incremental solvers rely on Moore–Penrose pseudoinverse operations, resulting in additional computational costs and potential numerical instability during continuous expansion (Chen & Liu, 2018; Chen et al., 2026).

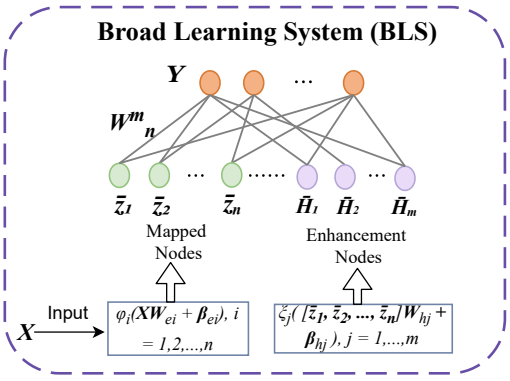


Figure 1: The structure of broad learning system

Based on these observations, this work proposes a cross-domain cross-modal hashing framework named **SRIF-BLS** (Semantic Relation transfer and Inverse-Free Broad Learning System), which integrates semantic relation transfer with inverse-free incremental broad learning. SRIF-BLS establishes a unified learning paradigm where semantic relation transfer provides reliable supervision signals and inverse-free broad learning enables efficient target-domain adaptation under limited supervision. At the semantic transfer level, SRIF-BLS learns transferable cross-modal representations from the auxiliary domain and combines cross-domain semantic consistency with CLIP-based weak semantic relations in the target domain, enabling knowledge transfer through relational structures rather than direct class correspondence. At the representation learning level, SRIF-BLS introduces a QR-based inverse-free broad learning mechanism. By maintaining the R-factor and solving triangular systems through forward and backward substitution, SRIF-BLS obtains analytical output weights without explicit pseudoinverse computation, improving the efficiency and numerical stability of incremental adaptation.

The main contributions of this work are summarized as follows:

1. The proposed SRIF-BLS framework establishes a unified paradigm for cross-domain cross-modal hashing by integrating semantic knowledge transfer and efficient target-domain adaptation, enhancing semantic representation learning under limited supervision.

2. The proposed method develops a cross-domain semantic relation transfer mechanism that jointly exploits cross-domain semantic consistency and CLIP-based weak semantic information, providing reliable supervision for unlabeled target domains under modality heterogeneity and mismatched label spaces.

3. The proposed method introduces an inverse-free incremental broad learning mechanism that efficiently adapts hashing models under transferred semantic supervision through QR-based R-factor updates and triangular solving, avoiding explicit pseudoinverse computation while improving the efficiency and numerical stability of model expansion.

2 RELATED WORK

2.1 CROSS-MODAL HASHING

Cross-modal hashing aims to map heterogeneous data, such as images and texts, into a shared Hamming space, enabling efficient cross-modal retrieval with compact binary codes. Existing approaches

can be broadly categorized into supervised methods (Jiang & Li, 2017; Deng et al., 2018; Yang et al., 2022; Teng et al., 2024; Fang & Ren, 2020; Li et al., 2024c; Luo et al., 2018; Shi et al., 2019; Yi et al., 2023) and unsupervised methods (Wang et al., 2020; Zhang et al., 2018; Xu et al., 2024; Hoang et al., 2022). The latter typically exploit intrinsic data structures, self-supervised signals, or pseudo-semantic relations (Yu et al., 2021; Shi et al., 2022; Tu et al., 2023) to uncover latent cross-modal correspondences. However, the resulting semantic supervision is often weak, making it difficult to learn stable and discriminative hash representations in the presence of complex semantic relationships and distributional variations.

2.2 Cross-Domain Cross-Modal Transfer

Cross-domain cross-modal transfer seeks to alleviate the lack of supervision in the target domain by leveraging semantically rich knowledge from an auxiliary domain, while learning transferable shared representations under domain and modality discrepancies. Existing methods primarily achieve knowledge adaptation through domain alignment and semantic structure transfer. CDTH (Li et al., 2024a) transfers semantic structures from an auxiliary domain to an unlabeled target domain; ATH (Lu et al., 2023) combines asymmetric hashing with adaptive bipartite graph learning to mitigate cross-domain distribution discrepancies; subsequent similarity transfer methods (Liang et al., 2022) further enhance cross-modal alignment by transferring semantic relations in the feature space. Despite encouraging progress, substantial discrepancies in feature distributions and label spaces across domains still make reliable semantic correspondences difficult to establish, limiting the effectiveness of cross-domain knowledge transfer.

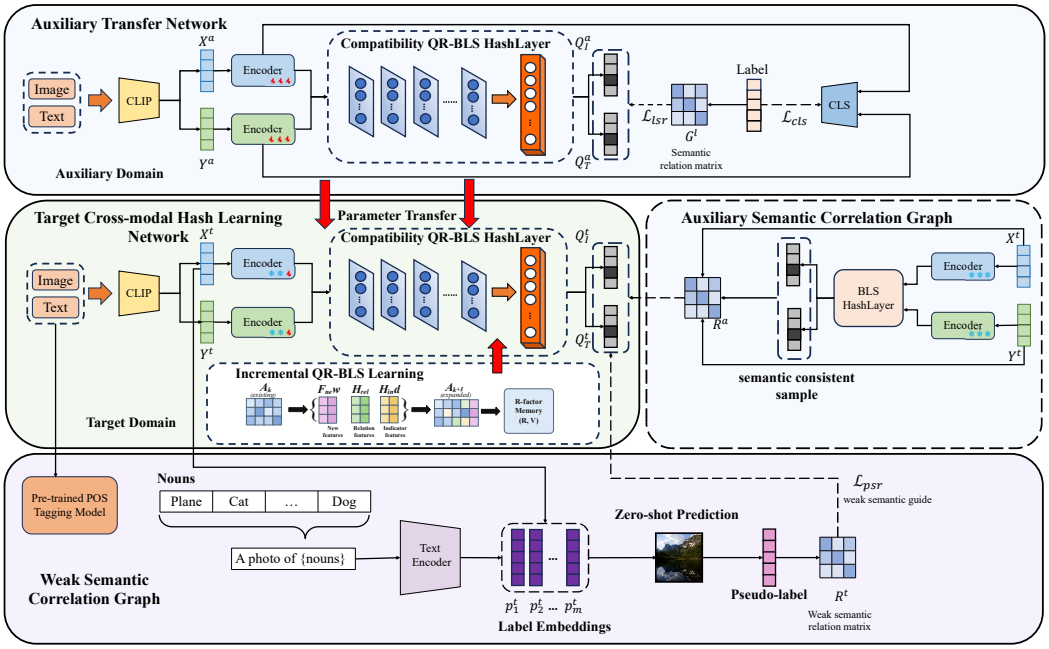


Figure 2: Overview of SRIF-BLS. The framework first transfers semantic relations from the auxiliary domain to construct cross-domain transferable representations, then performs target-domain hash learning through an inverse-free broad learning architecture with QR-based incremental updates.

2.3 Broad Learning and Inverse-Free Incremental Learning

The Broad Learning System (BLS) adopts a flat, horizontally expandable architecture and enables efficient incremental learning by introducing additional feature and enhancement nodes (Chen & Liu, 2018; Wang et al., 2021; Hu et al., 2023; Zhou et al., 2021). However, its analytical solution and incremental updates typically rely on the Moore–Penrose pseudoinverse, leading to considerable computational overhead and numerical stability issues as the model continues to expand. InvF-BLS

reformulates the BLS least-squares problem into semi-normal equations, maintains an (R)-factor through incremental QR decomposition, and obtains analytical weights via forward and backward triangular solves, thereby eliminating the need for explicit matrix inversion (Chen et al., 2026). In this work, we further incorporate this inverse-free incremental mechanism into the target-domain broad learning process of cross-domain cross-modal hashing, enabling more efficient and numerically stable representation expansion.

3 METHODS

3.1 PROBLEM SETUP AND REPRESENTATION TRANSFER

We learn image and text hash functions from a labeled auxiliary domain $\mathcal{D}_a = \{(x_i^a, y_i^a, l_i^a)\}_{i=1}^{n_a}$ and an unlabeled target domain $\mathcal{D}_t = \{(x_i^t, y_i^t)\}_{i=1}^{n_t}$. The paired inputs are precomputed image–text features; their domains may have disjoint label spaces. Superscripts a, t denote domains and subscripts I, T modalities. For matrices U, V, let $\mathcal{C}(U, V) = \bar{U}\bar{V}^\top$ be row-wise cosine similarity, and let $\phi(z) = 2\sigma(z) - 1$. We use $\mathcal{P} = \{(I, I), (T, T), (I, T)\}$ for the three relations evaluated on a mini-batch of size b.

Each modality has an independent three-block encoder $U_m = g_{m,3} \circ g_{m,2} \circ g_{m,1}(X_m)$ with widths $512 \to 4096 \to 2048 \to 512$. Its initial broad representation (Chen & Liu, 2018) comprises feature windows $F_{m,j}$ and an enhancement block H_m:

$$F_{m,j} = \tanh(U_m W_{m,j}^f + b_{m,j}^f), \qquad H_m = \tanh([F_{m,j}]_{j=1}^J W_m^h + b_m^h). \tag{1}$$

Concatenating these outputs gives the design $A_m = [F_{m,1}, \ldots, F_{m,J}, H_m]$ and continuous code

$$Q_m = \tanh(A_m W_m^o). \tag{2}$$

The initial layer uses eight windows of 64 nodes and 400 enhancement nodes, so A_m has 912 columns; the code length r is independent of this width. The source encoders, mappings, output projections, and shared classifier are jointly trained. Their label-relation, classification, and shared-code quantization losses combine as

$$\mathcal{L}_a = \mathcal{L}_{\text{lsr}} + \mathcal{L}_{\text{cls}} + \mathcal{L}_{\text{quant}}^a. \tag{3}$$

The label relation is $G^a = \phi(L^a(L^a)^\top)$; its preservation across $\mathcal{P}$ establishes a semantic structure without requiring target labels. The exact three loss terms are given in Appendix A.1. The first two source encoder blocks and the complete initial hashing layer, including W_m^o, initialize the target branches. The target third blocks are newly initialized; only these blocks and W_m^o receive target gradient updates. The transferred feature and enhancement mappings remain fixed, but pass gradients to the third block. The frozen early blocks may update their batch-normalization running statistics during target training.

3.2 DUAL SEMANTIC RELATION GRAPHS

Weak semantic graph. Let p_c be the embedding of candidate semantic term c (Radford et al., 2021). Temperature-scaled image–term similarity gives a soft assignment:

$$a_{ic} = \frac{\exp(\tau \cos(x_i^t, p_c))}{\sum_h \exp(\tau \cos(x_i^t, p_h))}. \tag{4}$$

We ℓ_2-normalize a_i and replace only its largest component by one; the others remain graded. Denote this reinforced vector by $\hat{a}_i$. The fixed weak graph is

$$R_{ij}^{wk} = \begin{cases} \phi(\hat{a}_i \hat{a}_j^\top), & i \neq j, \\ 2, & i = j. \end{cases} \tag{5}$$

Its off-diagonal entries express weak semantic affinity; the larger diagonal emphasizes paired image–text agreement. Details of the assignment and diagonal treatment appear in Appendix A.2.

Auxiliary relation graph. Apply the frozen source model to target pairs. Write $T_a = \mathcal{C}(Q_T^{a\to t}, Q_T^{a\to t})$ for transferred text-code similarity, and $I_x = \mathcal{C}(X_I^t, X_I^t)$, $S_x^t = \mathcal{C}(X_T^t, X_T^t)$ for target feature similarities. For each i, select k neighbors by I_x and compare the mean and population variance of T_a and S_x^t on those same neighbors. We retain the instances whose local discrepancies in both statistics are smaller than their global discrepancies; call this set $\mathcal{G}$. After excluding the least-similar target-text entries in each row, the resulting row mask M_T defines the supported relations:

$$R_{ij}^{tr} = \mathbf{1}[i \in \mathcal{G}]\, \mathbf{1}[j \in \mathcal{G}]\, (M_T)_{ij}. \tag{6}$$

The row mask can be asymmetric. Appendix A.3 gives the precise selection rule; the target categories never have to match source categories.

The supported pairs receive both direct and neighborhood-level feature evidence. Set $S_p = (1 - \alpha_1)I_x + \alpha_1 S_x^t$. Let P be its row-normalized neighborhood profile after removing low-similarity entries and self-relations. Their fusion is

$$B_g = (1 - \alpha_2)\phi(S_p) + \alpha_2 \phi(\gamma P P^\top). \tag{7}$$

Here γ controls the profile-similarity scale and α_1, α_2 the two fusions. Unlike a probability transition matrix, P may contain negative entries. Its exact normalization is specified in Appendix A.3.

At the end of each epoch, target image codes are recomputed on an aligned evaluation-mode snapshot. Define $D^a = \phi(\mathcal{C}(Q_I^{a\to t}, Q_I^{a\to t}))$ and $D_e^t = \phi(\mathcal{C}(Q_{I,e}^t, Q_{I,e}^t))$. For numerical stability, let Δ_e be their entrywise difference with a symmetric, bounded exponent; the bound is an implementation safeguard, not a semantic threshold. The dynamically weighted graph is

$$R_e^a = R^{tr} \odot B_g \odot \exp\big(\mathrm{sgn}(B_g) \odot \Delta_e\big). \tag{8}$$

Initially the exponential factor is one. Subsequently, the support and feature graph remain fixed while the source–target image-relation discrepancy adjusts the weight of each supported pair. Appendix A.3 states the bounded update exactly.

Target objective. For $(u, v) \in \mathcal{P}$, define $S_{uv}^t = \mathcal{C}(Q_u^t, Q_v^t)$ and take the batch-indexed submatrices of R^{wk}, R_e^a. The weak graph is a similarity target, yielding

$$\mathcal{L}_{\mathrm{psr}} = \frac{1}{b^2} \sum_{(u,v)\in\mathcal{P}} \|S_{uv}^t - R^{wk}\|_F^2. \tag{9}$$

In contrast, the transfer graph weights attraction toward unit similarity:

$$\mathcal{L}_{\mathrm{tf}} = \frac{1}{b^2} \sum_{(u,v)\in\mathcal{P}} \|R_e^a \odot (S_{uv}^t - \mathbf{1})\|_F^2. \tag{10}$$

The weights enter quadratically, so negative entries do not repel pairs. With $B^t = \mathrm{sign}_+((Q_I^t + Q_T^t)/2)$, the shared binary target minimizes the sum of the two modalities' quantization errors for fixed continuous codes. We optimize

$$\mathcal{L}_t = \mathcal{L}_{\mathrm{psr}} + \lambda_1 \mathcal{L}_{\mathrm{tf}} + \lambda_2 \mathcal{L}_{\mathrm{quant}}^t. \tag{11}$$

Here $\mathcal{L}_{\mathrm{quant}}^t = (br)^{-1} \sum_m \|Q_m^t - B^t\|_F^2$. The three terms preserve target semantic structure, selectively transfer reliable relations, and align discrete codes.

The constructed semantic relations provide informative supervision for the unlabeled target domain. However, effectively exploiting these transferred signals requires an efficient adaptation mechanism, since conventional hashing models usually rely on repeated optimization when the target distribution changes. To address this issue, SRIF-BLS introduces an inverse-free broad learning mechanism that performs lightweight target-domain adaptation through QR-based incremental updates.

3.3 Incremental QR-Based Width Expansion

At a scheduled expansion, each branch appends one feature window, enhancement nodes driven by that window, and enhancement nodes driven by all feature windows available at creation. Calling their concatenated outputs D_s, the design grows by columns:

$$A_s = [A_{s-1}, D_s]. \tag{12}$$

Existing enhancement blocks keep their original input ranges. Thus the old columns remain unchanged on the same input snapshot, and zero-padding old output weights preserves the previous predictor before refitting. The two scheduled increments add 64 feature, 75 new-window enhancement, and 125 all-window enhancement columns each, giving $912 \to 1176 \to 1440$. Appendix A.4 specifies the new-node mappings.

The trainable third block changes the design between epochs. Immediately before each increment we therefore collect both modalities' aligned full-training snapshot in evaluation mode, construct the detached shared binary target $Y = B^t$, and synchronize the current factor state. For either branch, the output projection is refitted through the ridge surrogate

$$W^\star = \arg\min_W \left\{ \|AW - Y\|_F^2 + \rho \|W\|_F^2 \right\}, \quad \rho > 0. \tag{13}$$

Factoring the augmented old design $[A^\top, \sqrt{\rho} I_p]^\top = Q_R R$ gives the upper-triangular state (Chen et al., 2026)

$$R^\top R = A^\top A + \rho I_p, \qquad V = A^\top Y. \tag{14}$$

The QR state is recomputed before each growth step because either A or Y may have changed. This is a linear fit to the pre-$\tanh$ response, not an exact minimization of the semantic loss.

Let $D \in \mathbb{R}^{n_t \times q}$ be the new columns of this same snapshot. With $L = R^\top$, a triangular solve yields the cross-block coupling E:

$$LE^\top = A^\top D. \tag{15}$$

The residual block has a Cholesky factor G defined by

$$GG^\top = D^\top D + \rho I_q - EE^\top. \tag{16}$$

The factors and right-hand side are then expanded, without refactorizing the entire wider design:

$$R^+ = \begin{bmatrix} R & E^\top \\ 0 & G^\top \end{bmatrix}, \qquad V^+ = \begin{bmatrix} V \\ D^\top Y \end{bmatrix}. \tag{17}$$

Finally, forward and backward triangular substitution replace the whole output matrix:

$$(R^+)^\top K^+ = V^+, \qquad R^+ W^+ = K^+. \tag{18}$$

For $\rho > 0$, the residual block in Equation (16) is positive definite, even with dependent new columns, and Equation (18) returns the unique expanded ridge solution. The proof and the precise fixed-snapshot scope of this claim are in Appendix A.4.1.

3.4 Optimization and Retrieval

Auxiliary optimization precedes parameter transfer and construction of the target graphs. On each target mini-batch, Equation (11) updates the trainable third blocks and output projections while treating the shared binary target and graph submatrices as fixed. The source relation graph, when active, is refreshed after each epoch. At the two scheduled expansion stages, the QR refit of Equations (14)–(18) initializes the enlarged output projections; gradient-based semantic optimization then resumes. If the transfer graph is active, it is refreshed again after expansion. At inference, each modality independently produces $\operatorname{sign}(Q_m^t)$; image-to-text and text-to-image retrieval rank opposite-modality codes by Hamming distance.

4 Experiments

4.1 Experimental Setup

Datasets and evaluation. We conduct experiments on MIRFlickr-25K (Huiskes & Lew, 2008) and NUS-WIDE (Chua et al., 2009), following the data protocol used in the cross-domain hashing setting (Su et al., 2019). MIRFlickr-25K contains 25,000 image–text pairs over 24 semantic concepts; after removing unlabeled samples, 20,015 valid pairs remain. We randomly select 2,000 pairs as queries, use the remaining 18,015 pairs as the retrieval set, and sample 5,000 retrieval pairs for training. For NUS-WIDE, we retain the 181,365 valid pairs from the ten most frequent concepts, with 2,000 query pairs, 179,365 retrieval pairs, and 5,000 training pairs. Unless otherwise stated,

MIRFlickr-25K is used as the auxiliary domain and NUS-WIDE as the unlabeled target domain. Retrieval is evaluated by mean average precision (mAP) for image-to-text (I→T) and text-to-image (T→I) retrieval with 16, 32, and 64-bit hash codes.

Implementation details. All experiments are conducted on a single NVIDIA RTX 5070 Ti GPU with a random seed of 1234. For both transfer directions, the auxiliary-domain model is trained for 100 epochs and the target-domain adaptation stage is performed for 40 epochs, with a batch size of 256 for both stages. The auxiliary-domain optimizer uses a learning rate of 1×10^{-4} and a weight decay of 5×10^{-4}.

The default configuration corresponds to **MIRFlickr-25K→NUS-WIDE**, where MIRFlickr-25K serves as the auxiliary domain and NUS-WIDE as the target domain. For this transfer direction, the target-domain learning rate and weight decay are set to 5×10^{-5} and 2.5×10^{-4}, respectively. The semantic graph parameters are $a_1 = 0.01$, $a_2 = 0.7$, $\mathrm{TopK} = 30$, $K_{\mathrm{num}} = 3500$, and $T = 80$. The weights of the relation-transfer and quantization losses are set to $\lambda_1 = 0.65$ and $\lambda_2 = 1$, respectively.

For the reverse direction, **NUS-WIDE→MIRFlickr-25K**, where NUS-WIDE serves as the auxiliary domain and MIRFlickr-25K as the target domain, the target-domain learning rate and weight decay are set to 1×10^{-4} and 5×10^{-4}, respectively. The corresponding semantic graph parameters are $a_1 = 0.4$, $a_2 = 0.9$, $\mathrm{TopK} = 45$, $K_{\mathrm{num}} = 3000$, and $T = 80$, while the loss weights are set to $\lambda_1 = 0.5$ and $\lambda_2 = 1$.

For the inverse-free broad learning stage, the initial structure contains 912 nodes and performs two append-only expansions at an interval of 10 target-training epochs, resulting in widths of $912 \rightarrow 1176 \rightarrow 1440$. At each expansion stage, one feature window, 75 related enhancement nodes, and 125 additional enhancement nodes are incrementally introduced. The QR-based solver uses binary targets and a shared ridge coefficient of $\alpha = 1.0$. Retrieval performance is evaluated after every target-domain epoch, the mean of I→T and T→I mAP is used for model selection, and checkpoints are saved every 10 epochs.

Table 1: Cross-domain cross-modal retrieval performance (mAP%) on NUS-WIDE.

Method	**I→T**			**T→I**		
	16	**32**	**64**	**16**	**32**	**64**
CVH	42.80	41.40	40.10	43.27	41.79	40.39
IMH	45.31	44.78	43.56	46.88	45.99	44.52
CMFH	54.56	56.82	57.87	55.59	57.95	59.11
LSSH	59.42	54.60	54.45	49.61	48.08	48.91
DLFH	60.46	64.36	65.82	64.10	68.33	**70.25**
SCRATCH	63.81	65.45	66.47	64.58	65.58	66.82
JDSH	53.65	54.40	53.04	53.56	54.38	56.09
CDTH	64.00	66.70	68.39	64.17	67.28	68.69
VTM-UCH	61.60	63.30	63.90	63.50	64.90	65.10
Ours	**67.72**	**68.58**	**68.53**	**68.72**	**69.23**	69.26

Table 2: Cross-domain cross-modal retrieval performance (mAP%) on MIRFlickr-25K.

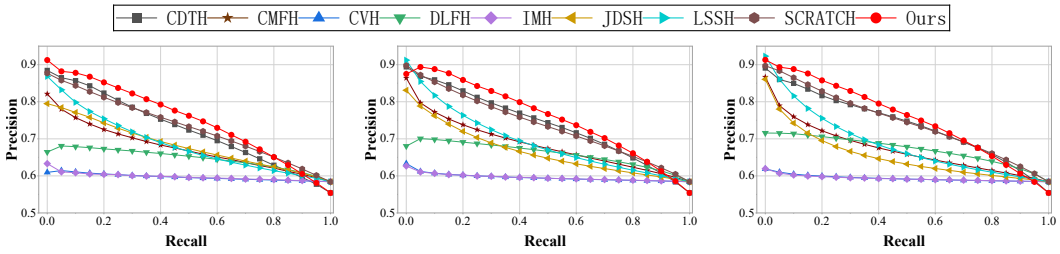
Method	**I→T**			**T→I**		
	16	**32**	**64**	**16**	**32**	**64**
CVH	59.46	59.27	59.11	59.47	59.52	59.36
IMH	59.04	59.27	59.44	59.55	59.88	60.11
CMFH	67.13	67.90	66.72	66.83	67.54	66.37
LSSH	68.07	68.70	68.09	65.09	65.28	64.56
DLFH	64.52	65.76	66.88	69.62	71.12	72.45
SCRATCH	72.84	73.19	74.11	73.79	73.76	74.92
JDSH	68.16	66.41	65.07	67.94	66.46	64.93
CDTH	72.34	73.71	73.73	72.76	74.63	74.57
VTM-UCH	74.30	75.40	75.10	75.10	75.80	**76.00**
Ours	**74.90**	**75.68**	**75.24**	**75.24**	**75.90**	75.28

Compared methods. We include eight representative baselines spanning classical, discrete, deep unsupervised, and cross-domain hashing. CVH (Kumar & Udupa, 2011) learns view-specific hash functions from multi-view data so that cross-view similarities are preserved in the learned binary codes. IMH (Song et al., 2013) models correlations among heterogeneous media and maps them into a common Hamming space, with linear regression used to generate codes for unseen samples. CMFH (Ding et al., 2014) employs collective matrix factorization to learn a unified latent representation and shared hash codes from multiple modalities, whereas LSSH (Zhou et al., 2014) combines sparse coding for visual structure with matrix factorization for textual latent concepts before mapping both modalities into a joint abstraction space. DLFH (Jiang & Li, 2019) directly learns binary codes through a discrete latent-factor formulation, while SCRATCH (Chen et al., 2019) uses scalable discrete matrix factorization to optimize cross-modal binary codes without relaxing the binary constraint. JDSH (Liu et al., 2020) constructs a joint-modal similarity matrix together with distribution-based similarity decision and weighting for unsupervised deep hashing. CDTH (Li et al., 2024a) transfers semantic structure from a semantically rich auxiliary domain to the target domain for efficient cross-domain cross-modal retrieval.VTM-UCH (Fan & Cao, 2025) mines text semantics by matching words with image or object representations through CLIP, fuses fine-grained image and text similarities, and applies community detection to construct pseudo-centers for refining the joint similarity matrix in unsupervised cross-modal hashing.

4.2 Comparison with Existing Methods

We compare SRIF-BLS with representative cross-domain cross-modal hashing methods, including conventional projection-based approaches (e.g., CVH, IMH, CMFH, and LSSH), deep hashing methods (e.g., DLFH and JDSH), and recent domain adaptation hashing methods (e.g., SCRATCH and CDTH). All methods are evaluated under the same retrieval protocol with identical hash lengths. Tables 1 and 2 report the I→T and T→I retrieval results, where SRIF-BLS is compared against existing methods from both retrieval directions.

Figure 3 presents representative precision–recall curves in the main text. The remaining precision–recall curves for other hash-code lengths, retrieval directions, and the NUS-WIDE dataset are provided in Appendix A.5 (Figure 4).



16 bits (I→T) on MIRFlickr-25K 32 bits (I→T) on MIRFlickr-25K 64 bits (I→T) on MIRFlickr-25K

Figure 3: Precision–recall curves on MIRFlickr-25K under different hash-code lengths and retrieval directions.

4.3 Ablation Study

To investigate the contribution of each component in SRIF-BLS, we conduct ablation studies by progressively removing the key modules from the full model while keeping the experimental settings unchanged. The evaluated variants analyze the effects of semantic relation transfer, target-domain semantic guidance, quantization constraint, transfer initialization, and the proposed inverse-free broad learning mechanism.

The variants are defined as follows:

w/o WSR: Removes the target-domain weak semantic relation modeling module and replaces the semantic relation graph with a uniform similarity structure.
w/o RSG: Replaces the pseudo-label-based semantic guidance with random semantic relations to

Table 3: Ablation results (mAP%) on MIRFlickr-25K.

Variants	I→T			T→I		
	16	32	64	16	32	64
w/o WSR	72.38	73.67	74.16	72.03	73.42	74.11
w/o RSG	72.99	73.15	73.95	72.82	73.27	73.98
w/o TR	73.27	74.45	74.80	73.54	74.63	74.87
w/o QL	65.97	66.42	65.63	65.90	66.40	65.30
w/o TI	73.53	74.67	74.68	73.61	74.79	74.53
w/o IF-BLS	73.17	74.61	74.50	73.15	74.64	74.64
SRIF-BLS	**74.90**	**75.68**	**75.24**	**75.24**	**75.90**	**75.28**

Table 4: Ablation results (mAP%) on NUS-WIDE.

Variants	I→T			T→I		
	16	32	64	16	32	64
w/o WSR	64.33	64.57	65.58	64.93	64.90	66.34
w/o RSG	64.42	65.24	63.89	65.15	66.01	64.58
w/o TR	66.74	66.85	67.89	67.75	67.82	68.56
w/o QL	54.49	55.42	59.93	54.89	55.60	59.44
w/o TI	66.19	67.10	66.68	67.09	67.85	67.68
w/o IF-BLS	65.96	66.13	67.00	67.88	68.15	69.10
SRIF-BLS	**67.72**	**68.58**	**68.53**	**68.72**	**69.23**	**69.26**

verify the effectiveness of reliable semantic information.
w/o TR: Removes the auxiliary-domain semantic relation transfer module and eliminates transferable semantic knowledge.
w/o QL: Removes the quantization loss to evaluate its influence on learning compact binary hash representations.
w/o TI: Removes the transfer initialization strategy and optimizes the target-domain model without transferred parameter initialization.
w/o IF-BLS: Replaces the proposed QR-based inverse-free broad learning mechanism with conventional optimization.
SRIF-BLS: The complete model with all proposed components.

As shown in Tables 3 and 4, removing any individual component leads to performance degradation, demonstrating the effectiveness of each module in SRIF-BLS. Removing WSR or RSG decreases retrieval accuracy, indicating that reliable target-domain semantic structures provide complementary supervision under limited labels. The degradation caused by removing TR verifies the importance of auxiliary-domain semantic knowledge transfer. Furthermore, removing QL affects the compact binary representation learning process, while replacing IF-BLS with conventional optimization reduces retrieval performance, confirming the effectiveness of the proposed QR-based incremental adaptation mechanism.

5 Conclusion

This paper presents SRIF-BLS, a cross-domain cross-modal hashing framework that integrates semantic relation transfer with inverse-free broad learning. By exploiting transferable semantic relations from the auxiliary domain and efficient QR-based incremental adaptation, SRIF-BLS alleviates the challenges of limited target-domain supervision and costly model updating. Extensive experiments on benchmark datasets demonstrate the effectiveness of SRIF-BLS over existing approaches, validating its capability for efficient and scalable cross-domain retrieval.

AI use statement

Generative AI tools were used only for language refinement and improving manuscript clarity. The authors independently designed the methodology, implemented SRIF-BLS, conducted experiments, analyzed results, and verified all technical content. All AI-assisted suggestions were reviewed and revised by the authors, who take full responsibility for the final manuscript.

Reproducibility statement

To ensure reproducibility, we provide detailed descriptions of our experimental settings, including data preprocessing, model architecture, hyperparameters, and training procedures. Our code is available at: `https://gitee.com/A1emon/SRIF-BLS`.

References

C. L. Philip Chen and Zhulin Liu. Broad learning system: An effective and efficient incremental learning system without the need for deep architecture. *IEEE Transactions on Neural Networks and Learning Systems*, 29(1):10–24, 2018.

Guang-Ze Chen, Chunyu Lei, Zhulin Liu, C. L. Philip Chen, and Hai-Wei Sun. Efficient incremental learning for inverse matrix-free broad learning system. *Information Fusion*, 127:103842, 2026.

Zhen-Duo Chen, Chuan-Xiang Li, Xin Luo, Liqiang Nie, Wei Zhang, and Xin-Shun Xu. Scratch: A scalable discrete matrix factorization hashing framework for cross-modal retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(7):2262–2275, 2019.

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 *Proceedings of the ACM international conference on image and video retrieval*, pp. 1–9, 2009.

Cheng Deng, Zhaojia Chen, Xianglong Liu, Xinbo Gao, and Dacheng Tao. Triplet-based deep hashing network for cross-modal retrieval. *IEEE Transactions on Image Processing*, 27(8):3893–3903, 2018.

Guiguang Ding, Yuchen Guo, and Jile Zhou. Collective matrix factorization hashing for multimodal data. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2075–2082, 2014.

Haozhi Fan and Yuan Cao. Vision-guided text mining for unsupervised cross-modal hashing with community similarity quantization. In *Proceedings of the AAAI conference on artificial intelligence*, volume 39, pp. 2843–2851, 2025.

Yixian Fang and Yuwei Ren. Supervised discrete cross-modal hashing based on kernel discriminant analysis. *Pattern Recognition*, 98:107062, 2020.

Tuan Hoang, Thanh-Toan Do, Tam V Nguyen, and Ngai-Man Cheung. Multimodal mutual information maximization: A novel approach for unsupervised deep cross-modal hashing. *IEEE Transactions on Neural Networks and Learning Systems*, 34(9):6289–6302, 2022.

Dongmei Hu, Xinying Xu, Jiwei Xing, Gang Xie, Jijun Tang, and Rong Li. Broad learning system based on fusion features. In *International Conference on Cognitive Systems and Signal Processing*, pp. 3–19. Springer, 2023.

Guangneng Hu, Yu Zhang, and Qiang Yang. Transfer meets hybrid: A synthetic approach for cross-domain collaborative filtering with text. In *The world wide web conference*, pp. 2822–2829, 2019.

Mark J Huiskes and Michael S Lew. The mir flickr retrieval evaluation. In *Proceedings of the 1st ACM international conference on Multimedia information retrieval*, pp. 39–43, 2008.

Qing-Yuan Jiang and Wu-Jun Li. Deep cross-modal hashing. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3232–3240, 2017.

Qing-Yuan Jiang and Wu-Jun Li. Discrete latent factor model for cross-modal hashing. *IEEE Transactions on Image Processing*, 28(7):3490–3501, 2019.

Josip Josifovski, Shangding Gu, Mohammadhossein Malmir, Haoliang Huang, Sayantan Auddy, Nicolás Navarro-Guerrero, Costas Spanos, and Alois Knoll. Safe continual domain adaptation after sim2real transfer of reinforcement learning policies in robotics. *arXiv preprint arXiv:2503.10949*, 2025.

Shaishav Kumar and Raghavendra Udupa. Learning hash functions for cross-view similarity search. In *IJCAI proceedings-international joint conference on artificial intelligence*, volume 22, pp. 1360, 2011.

Fengling Li, Bowen Wang, Lei Zhu, Jingjing Li, Zheng Zhang, and Xiaojun Chang. Cross-domain transfer hashing for efficient cross-modal retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(10):9664–9677, 2024a.

Li Li, Zhenqiu Shu, Zhengtao Yu, and Xiao-Jun Wu. Robust online hashing with label semantic enhancement for cross-modal retrieval. *Pattern Recognition*, 145:109972, 2024b.

Ze Li, Tao Yao, Lili Wang, Ying Li, and Gang Wang. Supervised contrastive discrete hashing for cross-modal retrieval. *Knowledge-Based Systems*, 295:111837, 2024c.

Huan Liang, Hairong Wang, and Dong Wang. Cross-modal hashing network based on self-attention similarity transfer. *Journal of Beijing University of Aeronautics and Astronautics*, 50(2):615–622, 2022.

Song Liu, Shengsheng Qian, Yang Guan, Jiawei Zhan, and Long Ying. Joint-modal distribution-based similarity hashing for large-scale unsupervised deep cross-modal retrieval. In *Proceedings of the 43rd international ACM SIGIR conference on research and development in information retrieval*, pp. 1379–1388, 2020.

Jianglin Lu, Jie Zhou, Yudong Chen, Witold Pedrycz, and Kwok-Wai Hung. Asymmetric transfer hashing with adaptive bipartite graph learning. *IEEE Transactions on Cybernetics*, 54(1):533–545, 2023.

Xin Luo, Xiao-Ya Yin, Liqiang Nie, Xuemeng Song, Yongxin Wang, Xin-Shun Xu, et al. Sdmch: Supervised discrete manifold-embedded cross-modal hashing. In *IJCAI*, pp. 2518–2524, 2018.

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

Xiaobo Shen, Qianxin Huang, Long Lan, and Yuhui Zheng. Contrastive transformer cross-modal hashing for video-text retrieval. In *IJCAI*, pp. 1227–1235, 2024.

Yufeng Shi, Xinge You, Feng Zheng, Shuo Wang, and Qinmu Peng. Equally-guided discriminative hashing for cross-modal retrieval. In *Ijcai*, pp. 4767–4773, 2019.

Yufeng Shi, Yue Zhao, Xin Liu, Feng Zheng, Weihua Ou, Xinge You, and Qinmu Peng. Deep adaptively-enhanced hashing with discriminative similarity guidance for unsupervised cross-modal retrieval. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):7255–7268, 2022.

Jingkuan Song, Yang Yang, Yi Yang, Zi Huang, and Heng Tao Shen. Inter-media hashing for large-scale retrieval from heterogeneous data sources. In *Proceedings of the 2013 ACM SIGMOD international conference on management of data*, pp. 785–796, 2013.

Shupeng Su, Zhisheng Zhong, and Chao Zhang. Deep joint-semantics reconstructing hashing for large-scale unsupervised cross-modal retrieval. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3027–3035, 2019.

Shaohua Teng, Shengjie Lin, Luyao Teng, Naiqi Wu, Zefeng Zheng, Lunke Fei, and Wei Zhang. Joint specifics and dual-semantic hashing learning for cross-modal retrieval. *Neurocomputing*, 565:126993, 2024.

Rong-Cheng Tu, Jie Jiang, Qinghong Lin, Chengfei Cai, Shangxuan Tian, Hongfa Wang, and Wei Liu. Unsupervised cross-modal hashing with modality-interaction. *IEEE Transactions on Circuits and Systems for Video Technology*, 33(9):5296–5308, 2023.

M Wang and W Deng. Deep visual domain adaptation: A survey. neurocomputing 312, 135–153 (2018). *Scientific Reports*, 12:1040, 2022.

Tong Wang, Lei Zhu, Zhiyong Cheng, Jingjing Li, and Zan Gao. Unsupervised deep cross-modal hashing with virtual label regression. *Neurocomputing*, 386:84–96, 2020.

Zhaoxin Wang, Huaping Liu, Xinying Xu, and Fuchun Sun. Multi-modal broad learning for material recognition. *Cognitive Computation and Systems*, 3(2):123–130, 2021.

Mengying Xu, Linyin Luo, Hanjiang Lai, and Jian Yin. Category-level contrastive learning for unsupervised hashing in cross-modal retrieval. *Data Science and Engineering*, 9(3):251–263, 2024.

Fan Yang, Zheng Wang, Jing Xiao, and Shin'ichi Satoh. Mining on heterogeneous manifolds for zero-shot cross-modal image retrieval. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pp. 12589–12596, 2020.

Fan Yang, Xiaojian Ding, Yufeng Liu, Fumin Ma, and Jie Cao. Scalable semantic-enhanced supervised hashing for cross-modal retrieval. *Knowledge-based systems*, 251:109176, 2022.

Fan Yang, Xinqi Liu, Fumin Ma, Xiaojian Ding, and Kaixiang Wang. Online asymmetric supervised discrete cross-modal hashing for streaming multimedia data. *Pattern Recognition*, 165:111604, 2025.

Zeqian Yi, Xinghui Zhu, Runbing Wu, Zhuoyang Zou, Yi Liu, and Lei Zhu. Multi-label weighted contrastive cross-modal hashing. *Applied Sciences*, 14(1):93, 2023.

Jun Yu, Hao Zhou, Yibing Zhan, and Dacheng Tao. Deep graph-neighbor coherence preserving network for unsupervised cross-modal hashing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pp. 4626–4634, 2021.

Jun Yu, Xiao-Jun Wu, and Donglin Zhang. Unsupervised multi-modal hashing for cross-modal retrieval. *Cognitive Computation*, 14(3):1159–1171, 2022.

Zelong Zeng, Zheng Wang, Fan Yang, and Shin'ichi Satoh. Geo-localization via ground-to-satellite cross-view image retrieval. *IEEE Transactions on Multimedia*, 25:2176–2188, 2022.

Jian Zhang, Yuxin Peng, and Mingkuan Yuan. Unsupervised generative adversarial cross-modal hashing. In *Proceedings of the AAAI conference on artificial intelligence*, volume 32, 2018.

Jile Zhou, Guiguang Ding, and Yuchen Guo. Latent semantic sparse hashing for cross-modal similarity search. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pp. 415–424, 2014.

Yanhong Zhou, Jie Du, Kai Guan, and Tianfu Wang. Multi-modal broad learning system for medical image and text-based classification. In *2021 43rd annual international conference of the IEEE engineering in medicine & biology society (EMBC)*, pp. 3439–3442. IEEE, 2021.

A Appendix

A.1 Auxiliary Objective and Target Quantization

For an auxiliary batch, write $S^a_{uv} = \mathcal{C}(Q^a_u, Q^a_v)$ and $G^a = \phi(L^a(L^a)^\top)$. The three terms in Equation (3) are

$$\mathcal{L}_{\text{lsr}} = b^{-2} \sum_{(u,v)\in\mathcal{P}} \|S^a_{uv} - G^a\|^2_F. \tag{19}$$

The shared linear classifier h acts on the third encoder output, and its softmax cross-entropy uses the floating multi-label vectors directly, as in the implementation:

$$\mathcal{L}_{\text{cls}} = -b^{-1} \sum_{m,i,c} l^a_{ic} \log[\text{softmax}(h(U^a_{m,i}))]_c. \tag{20}$$

Writing $B^a = \text{sign}((Q^a_I + Q^a_T)/2)$, the auxiliary quantization term is

$$\mathcal{L}^a_{\text{quant}} = (br)^{-1} \sum_m \|Q^a_m - B^a\|^2_F. \tag{21}$$

The target uses sign_+, assigning a tie to $+1$, whereas the auxiliary quantizer applies the ordinary sign. Indeed, expanding the two squared errors shows that $B^t = \text{sign}_+((Q^t_I + Q^t_T)/2)$ minimizes $\sum_m \|Q^t_m - B\|^2_F$ over binary B for fixed continuous codes. The codes are detached when constructing the target.

A.2 Semantic Graph Construction

The candidate terms are embedded in the same input space as target images. In this formulation, x^t_i denotes the i-th target-domain sample, $a_i \in \mathbb{R}^C$ denotes its CLIP-based semantic assignment vector over C semantic concepts, and a_{ic} is the confidence assigned to concept c. The variable $\tau > 0$ in Equation (4) controls assignment concentration. With $\tilde{a}_i = a_i/\|a_i\|_2$ and $c^*_i = \arg\max_c a_{ic}$, reinforcement sets $\hat{a}_{ic^*_i} = 1$ and retains $\hat{a}_{ic} = \tilde{a}_{ic}$ elsewhere. Thus $\hat{a}_i$ is not a one-hot distribution. The graph's diagonal is set to two after pairwise construction. Within-modal self-cosine is one for nonzero codes, so the corresponding squared diagonal residual is constant; the cross-modal diagonal still supplies a paired-instance gradient.

A.3 Transfer Support and Dynamic Weighting

Let $\mathcal{N}_k(i)$ be the nearest other target instances under I_x. For $z \in \{a, x\}$, the text-relation statistics are

$$\mu_{z,i} = k^{-1} \sum_{j\in\mathcal{N}_k(i)} (T_z)_{ij}, \qquad v_{z,i} = k^{-1} \sum_{j\in\mathcal{N}_k(i)} ((T_z)_{ij} - \mu_{z,i})^2. \tag{22}$$

Let $\bar{\mu}_z = n_t^{-1} \sum_i \mu_{z,i}$ and $\bar{v}_z = n_t^{-1} \sum_i v_{z,i}$. The default consistency criterion is

$$\mathcal{G} = \{i : \ |\mu_{a,i} - \mu_{x,i}| < |\bar{\mu}_a - \bar{\mu}_x|, \ |v_{a,i} - v_{x,i}| < |\bar{v}_a - \bar{v}_x|\}. \tag{23}$$

The two inequalities test agreement of local mean and dispersion against their respective global differences; they do not assert label correctness. The row mask M_T is one except at the κ smallest entries of each row of S^t_x, where it is zero. Here, S^t_x denotes the target-domain semantic relation matrix obtained from the auxiliary-domain transferred representation, and M_T selects reliable transferable relations while suppressing low-confidence neighbors. Thus Equation (6) may be asymmetric.

To construct the profile P in Equation (7), S_p denotes the semantic similarity matrix transferred from the auxiliary domain, while P is the normalized semantic profile matrix derived from S_p. Specifically, zero the κ smallest entries in every row of S_p and then its diagonal, obtaining Z. Write $s_i = \sum_j Z_{ij}$ and $d_i = s_i$ if $|s_i| \geq \epsilon$, otherwise $d_i = 1$, where ϵ is a small numerical tolerance. Set $P_{ij} = Z_{ij}/d_i$. Since negative feature similarities can survive pruning, these rows need not represent probabilities.

At initialization $R^a_0 = R^{tr} \odot B_g$. Here, R^{tr} represents the transferable auxiliary-domain relation support, while B_g denotes the fused semantic guidance matrix combining transferred and target

weak semantic relations. For subsequent epochs, let $\delta_e = D^a - D_e^t$ denote the semantic discrepancy between the auxiliary and target domains at epoch e, and define $\Delta_e = \mathrm{clip}(\delta_e, -c, c)$ entrywise, with $c > 0$ a fixed numerical bound. Thus Equation (8) is equivalently a positive exponential factor for nonnegative B_g and a negative exponent for negative B_g, multiplied in either case by $R^{tr} \odot B_g$. Clipping prevents exponential overflow but does not change the support. The three branch losses in Equation (10) use the same batch-indexed graph; because $\|R_e^a \odot (S - \mathbf{1})\|_F^2 = \sum_{ij} (R_{e,ij}^a)^2 (S_{ij} - 1)^2$, the sign of a weight cannot create repulsion.

A.4 Append-Only Blocks and QR Derivation

At stage s, A_{s-1} denotes the existing broad feature matrix before expansion, and D_s denotes the newly appended feature and enhancement nodes. The new feature nodes are $F_s = \tanh(UW_s^f + b_s^f)$. The new-window enhancement $H_s^r = \tanh(F_s W_s^r + b_s^r)$ uses only this window. The other block $H_s^g = \tanh([F^{\mathrm{old}}, F_s] W_s^g + b_s^g)$ uses all windows available at this stage, and $D_s = [F_s, H_s^r, H_s^g]$. Every enhancement block permanently stores its creation-time input range. For the same encoder snapshot, $A_s[:, 1:p] = A_{s-1}$, so $A_s[W^\top, 0]^\top = A_{s-1} W$. The refitted weights can nevertheless change on both old and new columns.

A.4.1 Exactness of the Block Update

For $\rho > 0$, augmented QR yields $R^\top R = A^\top A + \rho I_p$ and $V = A^\top Y$. The old-width solution satisfies $R^\top K = V$ and $RW = K$; it is synchronized without writing W to the network before expansion. Set $L = R^\top$, where R is the upper triangular factor maintained by QR decomposition, and $S_D = D^\top D + \rho I_q - EE^\top$ is the Schur complement associated with the newly added nodes. Because $LE^\top = A^\top D$, Schur elimination gives, for every $z \neq 0$,

$$z^\top S_D z = min_x \{ \|Ax + Dz\|_2^2 + \rho \|x\|_2^2 + \rho \|z\|_2^2 \}. \tag{24}$$

The right-hand side is at least $\rho \|z\|_2^2 > 0$; hence the Cholesky factor G exists even for dependent columns. Multiplication of Equation (17) gives

$$(R^+)^\top R^+ = [A, D]^\top [A, D] + \rho I_{p+q}. \tag{25}$$

With $V^+ = [A, D]^\top Y$, the triangular solves in Equation (18) therefore satisfy the expanded ridge normal equations and give its unique minimizer. Zero-padded old weights are feasible for the expanded problem, so its optimal ridge objective on this same snapshot cannot exceed the old objective. The statement does not imply monotonic decrease of the nonlinear semantic objective or improvement in retrieval accuracy. The implementation uses double precision for the factorization and may add a small diagonal perturbation to the new Schur block if finite-precision Cholesky fails; the exact equivalence above assumes no perturbation. The fitted output matrix is subsequently refined by mini-batch semantic training.

A.5 Additional Precision–Recall Curves

Figure 4 reports the remaining precision–recall curves for MIRFlickr-25K and NUS-WIDE across additional hash-code lengths and retrieval directions. The same method legend as in the main-text PR figure is used for all panels.

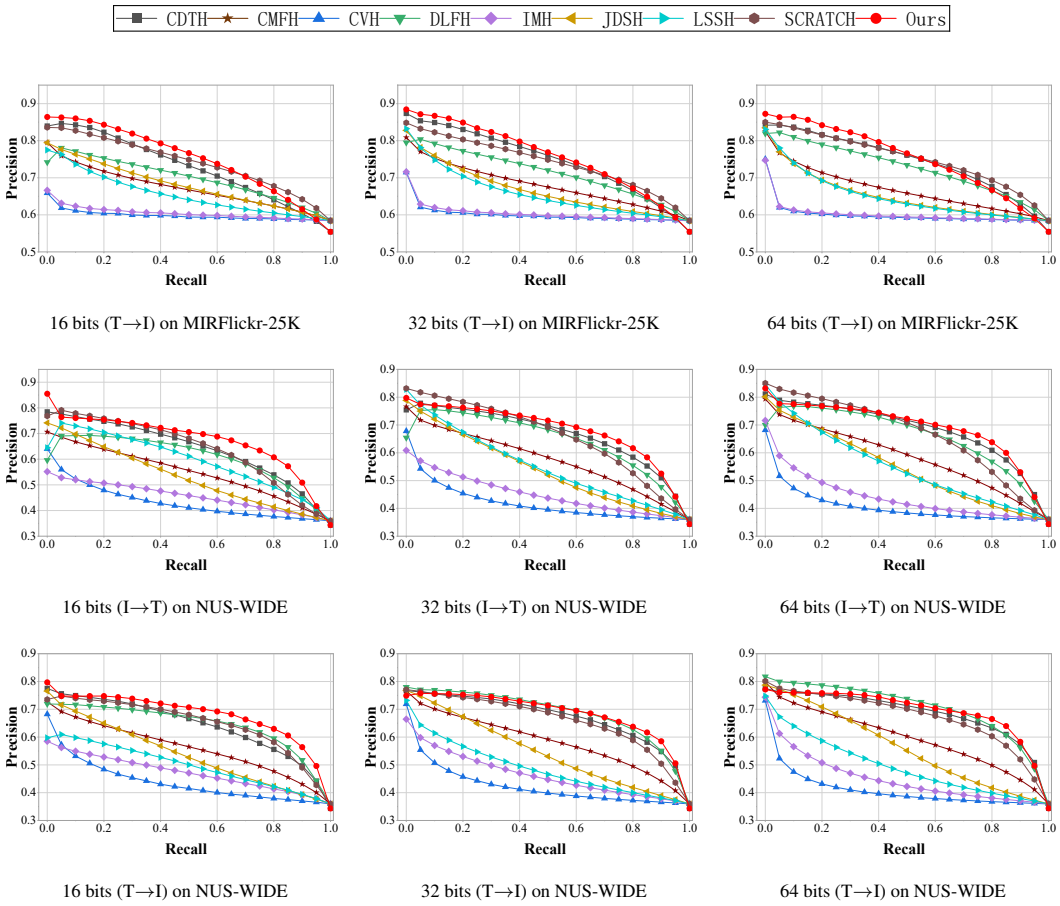


Figure 4: Additional precision–recall curves on MIRFlickr-25K and NUS-WIDE for the remaining hash-code lengths and retrieval directions.