Online Cross-Modal Hashing with Expanding Label Space

Wentao Fan, Chao Zhang, Chunlin Chen, Huaxiong Li *

Nanjing University, China
{fanwentao0955,chzhang}@smail.nju.edu.cn, {clchen, huaxiongli}@nju.edu.cn

Abstract

Due to the continuous increase of multimedia data on the internet, online hashing has garnered considerable attention for handling multi-modal data streams. However, most existing online hashing approaches focus solely on data growth of samples, overlooking the dynamics of classes. In this paper, we simultaneously address the challenges of both sample-level and class-level growth, and propose a novel Online Hashing method with Expanding Label Space (OH-ELS) for cross-modal retrieval. In OH-ELS, multi-modal data arrives continuously, and incoming data may introduce new classes. To avoid catastrophic forgetting, we transfer the historical knowledge at both the sample and class levels. At the sample-level, a small subset of anchor codes from old data are replayed to preserve the similarities between new data and old data. At the class-level, a consistency regularizer is applied to new classifiers to leverage the priors of historical classes. To ensure both efficiency and accuracy, a discrete optimization algorithm is proposed to solve the binary-constrained optimization problem without relaxation. Experimental results illustrate the effectiveness and superiority of OH-ELS in class-incremental cross-modal retrieval compared with the state-of-the-art methods.

Introduction

The explosion of multimedia data on the internet has significantly increased the demand for effective cross-modal retrieval techniques, where users can retrieve semantically relevant items across heterogeneous modalities using a query from another. The high complexities of traditional search methods make them unsuitable for practical use, particularly in large-scale or resource-constrained environments. As an Approximate Nearest Neighbor (ANN) method, hashing has attracted growing interest in recent years owing to its advantages in rapid query speed and low storage requirements (Wu et al. 2022; Jin et al. 2024; Sun et al. 2024; Li et al. 2025b). Cross-modal hashing aims to obtain a low-dimensional binary representation of high-dimensional multi-modal features while preserving the original associations, which enables efficient similarity search through XOR operation (Kang et al. 2025; Li, Long, and Yang 2025).

*Corresponding Author

Although existing hashing methods have shown promising results, most of them adopt an offline learning mechanism, which presumes complete access to the dataset before training (Meng et al. 2020; Li et al. 2021; Lei et al. 2024). However, in many practical applications, multimedia data commonly arrives in the form of dynamic streams. This raises two primary challenges. The first is computational efficiency: with the arrival of new data, offline hashing methods have to recompute binary codes for the entire dataset and retrain the hashing model, incurring substantial computational overhead. The second is storage scalability: the continuous influx of new data leads to ever-growing memory demands, which limits offline methods' scalability to large-scale datasets. To accommodate dynamic data streams, an online hash learning scheme is introduced, which incrementally updates the hash model when new data arrives (Xie, Shen, and Zhu 2016; Yao et al. 2019; Yi et al. 2021; Liu et al. 2022; Zhang, Wu, and Chen 2023; Jiang et al. 2023b,a; Han et al. 2024; Li et al. 2025a). Furthermore, after each training iteration, the data from the current round is commonly discarded from memory. Therefore, both the computational complexity and memory requirements can be well reduced.

Online hashing offers an efficient solution for processing continuous data streams in real time. However, several critical challenges remain to be addressed: 1) Existing online hashing approaches struggle to handle incoming data that involves an expanding label space, as they typically assume that all classes are known and fixed during training. In real-world scenarios, new classes often emerge over time, which requires a novel learning scheme that can incrementally incorporate knowledge of newly introduced classes while maintaining effectiveness for previously learned ones (Peng et al. 2024); 2) The commonly-used relaxation strategy in online hash learning may cause information loss during the quantization of continuous representations into discrete hash codes (Liu et al. 2025).

To accommodate both sample growth and class expansion, this study develops a novel online cross-modal hashing approach, termed Online Hashing with Expanding Label Space (OH-ELS). At the sample level, a subset of binary codes from previous rounds is selected as anchor codes. By replaying the anchor codes, OH-ELS not only measures the neighboring relationships among data points but also captures semantic correlations between previous and new data.

At the class level, OH-ELS performs hash learning within a classification framework, where a classifier is learned at each round to leverage class information as the semantic guidance for hash codes generation. A consistency regularizer is proposed to preserve and transfer the knowledge of historical classes, serving as prior information for training new hash model. To prevent information loss caused by continuous relaxation, we propose a discrete learning algorithm that directly learns hash codes while preserving binary constraints. This study makes the following contributions:

- We propose a novel online hashing method to address the simultaneous growth of samples and classes — a critical yet under-explored problem in online cross-modal retrieval.
- To mitigate catastrophic forgetting, we employ anchor code replay and consistency regularization to facilitate effective knowledge transfer at both the sample and class levels.
- A discrete optimization algorithm is devised to directly generate hash codes with binary constraints. Comprehensive experimental evaluations verify the efficacy of the proposed approach in class-incremental retrieval tasks.

Related Work

In this section, an overview of related research on online cross-modal hashing and class-incremental learning is presented.

Online Cross-Modal Hashing

Recent years have witnessed the emergence of multiple online cross-modal hashing methods. Compared with offline hashing, online hashing can update hash model incrementally to achieve more flexible cross-modal retrieval. Depending on the availability of label supervision, existing online cross-modal hashing approaches can be broadly divided into unsupervised and supervised paradigms. Unsupervised approaches, such as OCMFH (Wang et al. 2020), generate hash representations by modeling the joint distribution of features across multiple modalities. Leveraging label supervision, supervised online hashing can achieve better retrieval performance. LEMON (Wang, Luo, and Xu 2020) refines the original asymmetric learning strategy to capture the semantic relationships across historical and incoming data. DOCH (Zhan et al. 2022) formulates hash code generation as a log-likelihood maximization task. Several studies report that $\{0, 1\}$ logical annotations lack the capacity to encode complex semantic structures, potentially leading to semantic drift in streaming scenarios. To address this issue, ODCH (Kang et al. 2023) and DOCHM (Kang et al. 2024) introduce label reconstruction strategies aimed at embedding more fine-grained semantic details. ROH (Jiang et al. 2023a) introduces a novel approach to efficiently maintain pairwise similarity relationships throughout online learning. ROHLSE (Li et al. 2024) and OHPKL (Shu, Li, and Yu 2024) are developed to handle partially labeled data in online settings. Most existing online hashing methods assume a fixed label space, while few studies consider expanding label spaces.

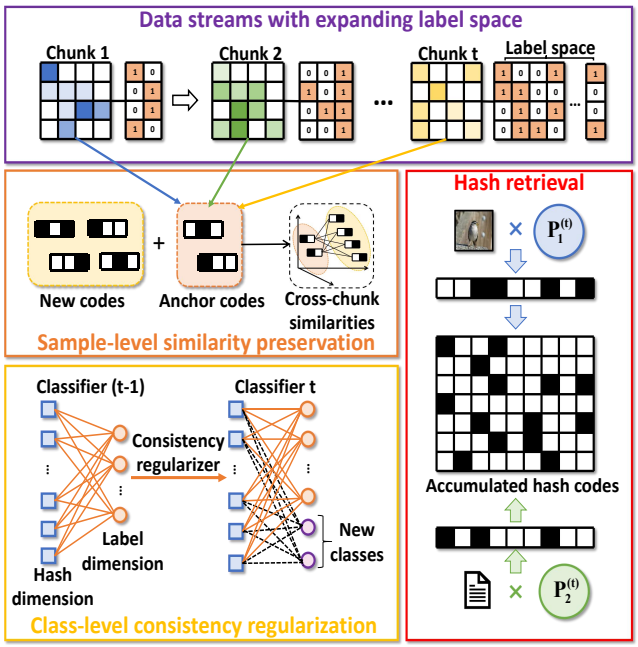


Figure 1: Overview of the OH-ELS approach.

Class-Incremental Learning

Class-incremental learning studies the problem of continuously incorporating new classes while retaining knowledge of previously learned ones (Park, Kang, and Han 2021; Zhou et al. 2024). The main challenge lies in mitigating catastrophic forgetting. To address this issue, several methods employ data replay, which revisits a small set of representative samples from previous rounds by saving an extra exemplar set (Rebuffi et al. 2017). Rather than replaying previous data, an alternative strategy proposes to regularize the model based on former data and control the optimization direction, which guarantees that training on new classes does not compromise knowledge of former ones (Tang et al. 2021). Distillation-based methods aim to convey the knowledge learned by the previous model to the new model by building distillation relationships (Hou et al. 2019), including logit distillation (Zhou, Ye, and Zhan 2021), feature distillation (Kang, Park, and Han 2022), and relational distillation (Gao et al. 2022). For template-based methods, a prototype is learned separately for each class to serve as a representative semantic embedding (De Lange and Tuytelaars 2021). Some methods further consider the scenario where feature and label dimensions increase simultaneously (Hou et al. 2023). Incorporating insights from class-incremental learning will enhance the capability of hash models in handling data with continuously growing label spaces.

Model Formulation

In this section, we first list the notations and give the problem definition. Subsequently, we elaborate the formulation of the proposed method, which covers sample-level similarity preservation and class-level consistency regularization. Figure 1 illustrates the overview of OH-ELS.

Notations and Problem Definition

Capitalized symbols (e.g., $\mathbf{A}$) are adopted to indicate matrices in this work. $\mathbf{A}_{i*}$ indicates the i-th row of $\mathbf{A}$ and $\mathbf{A}_{*i}$ corresponds to the i-th column of $\mathbf{A}$. $\mathbf{A}^T$ and $\mathbf{A}^{-1}$ refer to the transpose and inverse of $\mathbf{A}$, respectively. $\mathbf{I}$ denotes the identity matrix. $\mathrm{sign}(.)$ is a sign function.

Assume that the training samples consist of total M modalities, and at round t, a new data chunk $\{\vec{\mathbf{X}}_m^{(t)}\}_{m=1}^M$ is added to the database, where $\vec{\mathbf{X}}_m^{(t)} \in \mathbb{R}^{d_m \times n_t}$ is the feature matrix of modality m, and d_m, n_t denote the feature dimensions and data size, respectively. The corresponding label matrix is defined as $\vec{\mathbf{L}}^{(t)} \in \{0,1\}^{(C_{t-1}+c_t)\times n_t}$, where c_t is the number of incremental classes at round t, and $C_{t-1} = \sum_{i=1}^{t-1} c_i$ refers to the total count of classes observed up to round $(t-1)$.

This paper addresses the scenario where data continuously arrives with an expanding label space ($c_t > 0$). At each round, the objective is to learn semantically consistent hash codes, i.e., $\{-1,1\}^{r\times n_t}$, for new data without relearning the codes of previously seen data, in which r represents the length of hash codes. In addition, multiple hash functions are learned for out-of-sample extension.

Sample-Level Similarity Preservation

Preserving pairwise similarities among training samples is essential in cross-modal hashing, as it ensures that semantically similar instances have proximate representations in the Hamming space. Let $\vec{\mathbf{V}}^{(t)} \in \{-1,1\}^{r\times n_t}$ denote the binary common representation learned from $\{\vec{\mathbf{X}}_m^{(t)}\}_{m=1}^M$. An asymmetric strategy, i.e., $f(\vec{\mathbf{V}}^{(t)^T}\vec{\mathbf{B}}^{(t)};\mathbf{S}^{(t)})$, is adopted to capture the pairwise relationships between training data. $\mathbf{S}^{(t)}$ records the similarities between data that arrives at round t. To avoid the complex optimization problem caused by the discrete symmetric inner product $\vec{\mathbf{V}}^{(t)^T}\vec{\mathbf{V}}^{(t)}$, $\vec{\mathbf{B}}^{(t)} \in \{-1,1\}^{r\times n_t}$ is learned at each round as the approximation of $\vec{\mathbf{V}}^{(t)}$. By minimizing the difference between matrix $\mathbf{S}^{(t)}$ and binary code inner product $\vec{\mathbf{V}}^{(t)^T}\vec{\mathbf{B}}^{(t)}$, the pairwise similarities can be well preserved. We formulate the pairwise similarity preservation as the following maximum log-likelihood problem:

$$\begin{aligned}\max_{\vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)}} \log P(\mathbf{S}^{(t)}|\vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)}) \\ = \sum_{i=1}^{n_t}\sum_{j=1}^{n_t}(\mathbf{S}_{ij}^{(t)}\mathbf{\Theta}_{ij}^{(t)} - \log(1+\exp(\mathbf{\Theta}_{ij}^{(t)}))) \quad (1)\\ s.t.\ \ \vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)} \in \{-1,1\}^{r\times n_t},\end{aligned}$$

where $\mathbf{\Theta}_{ij}^{(t)} = \frac{\lambda}{r}\vec{\mathbf{V}}_{*i}^{(t)^T}\vec{\mathbf{B}}_{*j}^{(t)}$, and λ is a hyper-parameter. The semantic affinity across data is measured by the inner product of their labels:

$$\mathbf{S}_{ij}^{(t)} = \begin{cases} 1, & \langle \vec{\mathbf{L}}_{*i}^{(t)^T}\vec{\mathbf{L}}_{*j}^{(t)}\rangle > 0, \\ 0, & \langle \vec{\mathbf{L}}_{*i}^{(t)^T}\vec{\mathbf{L}}_{*j}^{(t)}\rangle \le 0. \end{cases} \quad (2)$$

In online scenarios, the correlations between existing and incoming data should be considered to mitigate catastrophic forgetting. To this end, an additional loss function $f(\vec{\mathbf{V}}^{(t)^T}\widetilde{\mathbf{B}}^{(t)};\hat{\mathbf{S}}^{(t)})$ is introduced, in which $\widetilde{\mathbf{B}}^{(t)} = [\vec{\mathbf{B}}^{(1)},\vec{\mathbf{B}}^{(2)},\cdots,\vec{\mathbf{B}}^{(t-1)}]$. $\hat{\mathbf{S}}^{(t)}$ captures the cross-chunk sample correlations, following the formulation defined in Eq. (2). The label space of the previous data is zero-padded to ensure computational efficiency.

Nevertheless, incorporating all historical binary codes into training is impractical. To reduce computational expenses, we select N_q samples from $\widetilde{\mathbf{B}}^{(t)}$ as anchor codes $\widetilde{\mathbf{B}}_{\mathbf{q}}^{(t)}$. By replaying anchor codes $\widetilde{\mathbf{B}}_{\mathbf{q}}^{(t)}$, our proposed method not only preserves the sample correlations of new data embedded in itself ($f(\vec{\mathbf{V}}^{(t)^T}\vec{\mathbf{B}}^{(t)};\mathbf{S}^{(t)})$), but also captures the similarities between previous and new data ($f(\vec{\mathbf{V}}^{(t)^T}\widetilde{\mathbf{B}}_{\mathbf{q}}^{(t)};\hat{\mathbf{S}}^{(t)})$). Then, Eq. (1) evolves into:

$$\begin{aligned}\max_{\vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)}} \sum_{i=1}^{n_t}\sum_{j=1}^{n_t}(\mathbf{S}_{ij}^{(t)}\mathbf{\Theta}_{ij}^{(t)} - \log(1+\exp(\mathbf{\Theta}_{ij}^{(t)}))) \\ + \sum_{i=1}^{n_t}\sum_{j=1}^{N_q}(\hat{\mathbf{S}}_{ij}^{(t)}\hat{\mathbf{\Theta}}_{ij}^{(t)} - \log(1+\exp(\hat{\mathbf{\Theta}}_{ij}^{(t)}))) \quad (3)\\ s.t.\ \ \vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)} \in \{-1,1\}^{r\times n_t},\end{aligned}$$

where $\hat{\mathbf{\Theta}}_{ij}^{(t)} = \frac{\lambda}{r}\vec{\mathbf{V}}_{*i}^{(t)^T}\widetilde{\mathbf{B}}_{\mathbf{q}*j}^{(t)}$. Eq. (3) enables a comprehensive exploitation of streaming data. Moreover, instead of adopting a relaxation-based strategy, hash codes are directly learned under binary constraints, thereby mitigating semantic loss caused by quantizing continuous variables.

Class-Level Consistency Regularization

Unlike existing online approaches that focus exclusively on sample growth, we additionally take class expansion into consideration. We conduct hash learning within a classification framework to exploit class information as the semantic guidance for hash codes generation:

$$\min_{\vec{\mathbf{V}}^{(t)},\mathbf{W}^{(t)}} \|\mathbf{W}^{(t)}\vec{\mathbf{V}}^{(t)} - \vec{\mathbf{L}}^{(t)}\|_F^2 \ \ s.t.\ \ \vec{\mathbf{V}}^{(t)} \in \{-1,1\}^{r\times n_t}, \quad (4)$$

where $\mathbf{W}^{(t)} \in \mathbb{R}^{(C_{t-1}+c_t)\times r}$ is the linear classifier at round t. Eq. (4) aims to obtain discriminative hash codes that can be easily classified using the linear classifier. The main challenge in handling class-incremental data is to continuously learn new classes while preserving knowledge of previously learned ones. Eq. (4) shows that the classifier is retrained on the new label matrix at each round, which results in the loss of knowledge acquired before. To retain knowledge of historical classes and ensure that it contributes to the learning of new classifiers, a consistency regularizer is introduced. Specifically, suppose in round $(t-1)$, we have a well-trained hash model with classifier defined as $\mathbf{W}^{(t-1)} \in \mathbb{R}^{C_{t-1}\times r}$. Since $\mathbf{W}^{(t-1)}$ and $\mathbf{W}^{(t)}$ play the same role in classification of data belonging to the first C_{t-1} classes, the knowledge from $\mathbf{W}^{(t-1)}$ can be transferred to guide the learning of $\mathbf{W}^{(t)}$ for efficient hash learning. Inspired by knowledge distillation, a consistency regularizer is defined by constraining the classifier parameter as:

$$\mathcal{R}(\mathbf{W}^{(t)}) = \|\bar{\mathbf{W}}^{(t)} - \mathbf{W}^{(t-1)}\|_F^2, \quad (5)$$

where $\bar{\mathbf{W}}^{(t)}$ denotes the first C_{t-1} rows of $\mathbf{W}^{(t)}$. Following the teacher-student distillation paradigm, the new model can directly inherit the previous knowledge by minimizing the Euclidean distance, which is simple and effective.

Overall Objective Function

The optimization problem for OH-ELS can be derived by integrating Eq. (3 - 5):

$$\begin{aligned}\min_{\vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)},\mathbf{W}^{(t)}} &\sum_{i=1}^{n_t}\sum_{j=1}^{n_t}(\log(1+\exp(\mathbf{\Theta}_{ij}^{(t)}))-\mathbf{S}_{ij}^{(t)}\mathbf{\Theta}_{ij}^{(t)})\\ &+\sum_{i=1}^{n_t}\sum_{j=1}^{N_q}(\log(1+\exp(\hat{\mathbf{\Theta}}_{ij}^{(t)}))-\hat{\mathbf{S}}_{ij}^{(t)}\hat{\mathbf{\Theta}}_{ij}^{(t)})\\ &+\|\mathbf{W}^{(t)}\vec{\mathbf{V}}^{(t)}-\vec{\mathbf{L}}^{(t)}\|_F^2+\alpha\mathcal{R}(\mathbf{W}^{(t)})\\ s.t.\ \ &\vec{\mathbf{V}}^{(t)},\vec{\mathbf{B}}^{(t)}\in\{-1,1\}^{r\times n_t},\end{aligned} \tag{6}$$

where α is a hyper-parameter. By solving the above problem, OH-ELS transfers historical knowledge at both the sample and class levels to support retrieval in evolving label spaces.

Online Optimization

An efficient alternating optimization strategy is devised to solve Eq. (6), where variables are updated sequentially with others held fixed.

$\vec{\mathbf{V}}^{(t)}$-Step: To solve $\vec{\mathbf{V}}^{(t)}$, we first remove the irrelevant part, i.e., $\alpha\mathcal{R}(\mathbf{W}^{(t)})$ in Eq. (6), and obtain the sub-equation defined as $\mathcal{L}(\vec{\mathbf{V}}^{(t)})$. To alleviate computational complexity, we select n_q samples from $\vec{\mathbf{B}}^{(t)}$ as anchor codes $\vec{\mathbf{B}}_{\mathbf{q}}^{(t)}$. $\mathbf{\Theta}^{(t)}$ is reformulated as $\frac{\lambda}{r}\vec{\mathbf{V}}^{(t)^T}\vec{\mathbf{B}}_{\mathbf{q}}^{(t)}$. A surrogate strategy is adopted to update $\vec{\mathbf{V}}^{(t)}$ bit by bit through maximizing the lower bound of $\mathcal{L}(\vec{\mathbf{V}}_{i*}^{(t)})$. In specific, we first compute the gradient and Hessian matrix of $\mathcal{L}(\vec{\mathbf{V}}^{(t)})$ over $\vec{\mathbf{V}}_{i*}^{(t)}$:

$$\begin{aligned}\frac{\partial\mathcal{L}}{\partial\vec{\mathbf{V}}_{i*}^{(t)}}=&\frac{\lambda}{r}\sum_{j}^{n_q}(\mathbf{S}_{*j}^{(t)^T}-\mathbf{F}_{*j}^{(t)^T})\vec{\mathbf{B}}_{\mathbf{q}_{ij}}^{(t)}\\ &+\frac{\lambda}{r}\sum_{j}^{N_q}(\hat{\mathbf{S}}_{*j}^{(t)^T}-\hat{\mathbf{F}}_{*j}^{(t)^T})\widetilde{\mathbf{B}}_{\mathbf{q}_{ij}}^{(t)}-2\mathbf{E}_{i*}^{(t)},\end{aligned} \tag{7}$$

and

$$\mathbf{H}_{\mathcal{L}}=-(\frac{\lambda^2}{4r^2}(n_q+N_q)+2)\mathbf{I}, \tag{8}$$

where $\mathbf{E}^{(t)}=\mathbf{W}^{(t)^T}\mathbf{W}^{(t)}\vec{\mathbf{V}}^{(t)}-\mathbf{W}^{(t)^T}\vec{\mathbf{L}}^{(t)}$, $\mathbf{F}_{ij}^{(t)}=1/(1+\exp(-\mathbf{\Theta}_{ij}^{(t)}))$, and $\hat{\mathbf{F}}_{ij}^{(t)}=1/(1+\exp(-\hat{\mathbf{\Theta}}_{ij}^{(t)}))$. Given by the theorem in (Jiang and Li 2019), we can derive the lower bound $\hat{\mathcal{L}}(\vec{\mathbf{V}}_{i*}^{(t)})$:

$$\hat{\mathcal{L}}(\vec{\mathbf{V}}_{i*}^{(t)})=\vec{\mathbf{V}}_{i*}^{(t)}(\frac{\partial\mathcal{L}}{\partial\vec{\mathbf{V}}_{i*}^{(t)}}(g)-\mathbf{H}_{\mathcal{L}}\vec{\mathbf{V}}_{i*}^{(t)}(g))^T+\text{const}, \tag{9}$$

where $\vec{\mathbf{V}}_{i*}^{(t)}(g)$ and $\frac{\partial\mathcal{L}}{\partial\vec{\mathbf{V}}_{i*}^{(t)}}(g)$ represents the value of $\vec{\mathbf{V}}_{i*}^{(t)}$ and the gradient w.r.t. $\vec{\mathbf{V}}_{i*}^{(t)}$ at the g-th iteration, respectively. Because of the binary constraint imposed on $\vec{\mathbf{V}}_{i*}^{(t)}$, to maximize $\hat{\mathcal{L}}(\vec{\mathbf{V}}_{i*}^{(t)})$, $\forall e\in n_t$, $\vec{\mathbf{V}}_{ie}^{(t)}$ should be set to 1 when the e-th element of $(\frac{\partial\mathcal{L}}{\partial\vec{\mathbf{V}}_{i*}^{(t)}}(g)-\mathbf{H}\vec{\mathbf{V}}_{i*}^{(t)}(g))$ is greater than 0; otherwise $\vec{\mathbf{V}}_{ie}^{(t)}=-1$. Consequently, the optimal solution for $\vec{\mathbf{V}}_{i*}^{(t)}$ is derived:

$$\vec{\mathbf{V}}_{i*}^{(t)}(g+1)=\text{sign}(\frac{\partial\mathcal{L}}{\partial\vec{\mathbf{V}}_{i*}^{(t)}}(g)-\mathbf{H}_{\mathcal{L}}\vec{\mathbf{V}}_{i*}^{(t)}(g)). \tag{10}$$

$\vec{\mathbf{B}}^{(t)}$-Step: When updating $\vec{\mathbf{B}}^{(t)}$, we also select samples from $\vec{\mathbf{V}}^{(t)}$ as anchor codes $\vec{\mathbf{V}}_{\mathbf{q}}^{(t)}$ to avert computational costs, and $\mathbf{\Theta}^{(t)}$ is reformulated as $\frac{\lambda}{r}\vec{\mathbf{B}}^{(t)^T}\vec{\mathbf{V}}_{\mathbf{q}}^{(t)}$. $\mathcal{J}(\vec{\mathbf{B}}^{(t)})$ is defined as the objective function that removes the irrelevant parts. Similar to updating $\vec{\mathbf{V}}^{(t)}$, we optimize $\vec{\mathbf{B}}^{(t)}$ bit by bit:

$$\vec{\mathbf{B}}_{i*}^{(t)}(g+1)=\text{sign}(\frac{\partial\mathcal{J}}{\partial\vec{\mathbf{B}}_{i*}^{(t)}}(g)-\mathbf{H}_{\mathcal{J}}\vec{\mathbf{B}}_{i*}^{(t)}(g)), \tag{11}$$

where

$$\frac{\partial\mathcal{J}}{\partial\vec{\mathbf{B}}_{i*}^{(t)}}=\frac{\lambda}{r}\sum_{j}^{n_q}(\mathbf{S}_{*j}^{(t)^T}-\mathbf{F}_{*j}^{(t)^T})\vec{\mathbf{V}}_{\mathbf{q}_{ij}}^{(t)}, \tag{12}$$

and

$$\mathbf{H}_{\mathcal{J}}=-\frac{\lambda^2 n_q}{4r^2}\mathbf{I}. \tag{13}$$

$\mathbf{W}^{(t)}$-Step: Since the regularization is imposed on parts of the classifier, we divide $\mathbf{W}^{(t)}$ as:

$$\min_{\mathbf{W}^{(t)}}\|\begin{bmatrix}\bar{\mathbf{W}}^{(t)}\\ \check{\mathbf{W}}^{(t)}\end{bmatrix}\vec{\mathbf{V}}^{(t)}-\vec{\mathbf{L}}^{(t)}\|_F^2+\alpha\|\bar{\mathbf{W}}^{(t)}-\mathbf{W}^{(t-1)}\|_F^2, \tag{14}$$

where $\check{\mathbf{W}}^{(t)}$ denotes the last c_t rows of $\mathbf{W}^{(t)}$. By equating the derivative of Eq. (14) to 0, we obtain the solutions for $\bar{\mathbf{W}}^{(t)}$ and $\check{\mathbf{W}}^{(t)}$:

$$\bar{\mathbf{W}}^{(t)}=(\vec{\mathbf{L}}_1^{(t)}\vec{\mathbf{V}}^{(t)^T}+\alpha\mathbf{W}^{(t-1)})(\vec{\mathbf{V}}^{(t)}\vec{\mathbf{V}}^{(t)^T}+\alpha\mathbf{I})^{-1}, \tag{15}$$

and

$$\check{\mathbf{W}}^{(t)}=(\vec{\mathbf{L}}_2^{(t)}\vec{\mathbf{V}}^{(t)^T})(\vec{\mathbf{V}}^{(t)}\vec{\mathbf{V}}^{(t)^T})^{-1}, \tag{16}$$

where $\vec{\mathbf{L}}_1^{(t)}$ and $\vec{\mathbf{L}}_2^{(t)}$ represent the first C_{t-1} rows and the last c_t rows of $\vec{\mathbf{L}}^{(t)}$, respectively.

Efficient Online Hash Function Learning

Based on the obtained binary codes, corresponding hash mappings are subsequently learned for out-of-sample extension:

$$\min_{\mathbf{P}_m^{(t)}}\|\mathbf{P}_m^{(t)}\vec{\mathbf{X}}_m^{(t)}-\vec{\mathbf{B}}^{(t)}\|_F^2+\|\mathbf{P}_m^{(t)}\widetilde{\mathbf{X}}_m^{(t)}-\widetilde{\mathbf{B}}^{(t)}\|_F^2, \tag{17}$$

where $\mathbf{P}_m^{(t)}$ denotes the hash function for modality m and $\widetilde{\mathbf{X}}_m^{(t)}=[\vec{\mathbf{X}}_m^{(1)},...,\vec{\mathbf{X}}_m^{(t-1)},\vec{\mathbf{X}}_m^{(t)}]$. It should be noted that we do not explicitly utilize the previous data but instead obtain the solution for $\mathbf{P}_m^{(t)}$ through the following procedure:

$$\mathbf{P}_m^{(t)}=\mathbf{K}_m^{(t)}(\mathbf{G}_m^{(t)})^{-1}, \tag{18}$$

where $\mathbf{K}_m^{(t)}=\mathbf{K}_m^{(t-1)}+\vec{\mathbf{B}}^{(t)}\vec{\mathbf{X}}_m^{(t)^T}$ and $\mathbf{G}_m^{(t)}=\mathbf{G}_m^{(t-1)}+\vec{\mathbf{X}}_m^{(t)}\vec{\mathbf{X}}_m^{(t)^T}$. By updating $\mathbf{K}_m^{(t)}\in\mathbb{R}^{r\times d_m}$ and $\mathbf{G}_m^{(t)}\in\mathbb{R}^{d_m\times d_m}$ at each round, the previously acquired knowledge is effectively preserved, and the memory consumption remains independent of the size of data.

Methods	Image to Text			Text to Image		
	16 bits	32 bits	64 bits	16 bits	32 bits	64 bits
OCMFH	0.5995	0.5997	0.5995	0.5998	0.6043	0.6102
LEMON	0.7115	0.7078	0.7228	0.7760	0.7746	0.7896
DOCH	0.7098	0.7301	0.7477	0.7415	0.7650	0.7915
ROH	0.7191	0.7253	0.7272	0.7743	0.7852	0.7928
ODCH	0.6742	0.6579	0.6678	0.7898	0.7958	0.7994
ROHLSE	0.6958	0.6979	0.7040	0.7138	0.7200	0.7223
OHPKL	0.6832	0.6948	0.6988	0.7047	0.7177	0.7204
OH-ELS	**0.8114**	**0.8361**	**0.8425**	**0.8533**	**0.8613**	**0.8688**

Methods	Image to Text			Text to Image		
	16 bits	32 bits	64 bits	16 bits	32 bits	64 bits
OCMFH	0.4601	0.4615	0.4637	0.4593	0.4633	0.4676
LEMON	0.6547	0.6615	0.6569	0.7326	0.7417	0.7372
DOCH	0.6743	0.6851	0.6987	0.7429	0.7655	0.7735
ROH	0.5404	0.5712	0.5691	0.7120	0.7426	0.7445
ODCH	0.6309	0.6360	0.6331	0.7353	0.7332	0.7377
ROHLSE	0.5988	0.6093	0.6020	0.6439	0.6417	0.6367
OHPKL	0.6015	0.6033	0.6118	0.6236	0.6435	0.6491
OH-ELS	**0.7161**	**0.7729**	**0.7856**	**0.8190**	**0.8445**	**0.8550**

Table 1: Average mAP scores on MIRFlickr-25K (left) and NUS-WIDE (right) across all rounds.

Methods	Image to Text			Text to Image		
	16 bits	32 bits	64 bits	16 bits	32 bits	64 bits
OCMFH	0.5631	0.5696	0.5703	0.5620	0.5708	0.5789
LEMON	0.7090	0.7105	0.7158	0.7500	0.7542	0.7583
DOCH	0.7517	0.7796	0.7867	0.7716	0.8048	0.8237
ROH	0.7035	0.7255	0.7285	0.7475	0.7727	0.7749
ODCH	0.5468	0.5253	0.5283	0.7566	0.7698	0.7707
ROHLSE	0.6489	0.6610	0.6646	0.6523	0.6668	0.6655
OHPKL	0.6432	0.6531	0.6582	0.6437	0.6613	0.6671
OH-ELS	**0.8206**	**0.8223**	**0.8497**	**0.8503**	**0.8538**	**0.8850**

Methods	Image to Text			Text to Image		
	16 bits	32 bits	64 bits	16 bits	32 bits	64 bits
OCMFH	0.3646	0.3655	0.3664	0.3638	0.3637	0.3653
LEMON	0.5478	0.5613	0.5523	0.6356	0.6756	0.6614
DOCH	0.5948	0.6067	0.6087	0.6690	0.7099	0.7380
ROH	0.4932	0.5335	0.5218	0.6891	0.6919	0.7013
ODCH	0.6016	0.6205	0.6263	0.6574	0.6638	0.6971
ROHLSE	0.5036	0.5002	0.5198	0.5683	0.5623	0.5745
OHPKL	0.5039	0.5095	0.5122	0.5506	0.5865	0.5718
OH-ELS	**0.6070**	**0.6370**	**0.6435**	**0.7482**	**0.7826**	**0.8041**

Table 2: mAP scores on MIRFlickr-25K (left) and NUS-WIDE (right) at the last round.

Computational Complexity Analysis

At round t, the computational complexity contains $O(((C_{t-1}+c_t)r^2+r^2n_t+(C_{t-1}+c_t)rn_t+rN_qn_t+rn_qn_t)k)$ for updating $\vec{\mathbf{V}}^{(t)}$, $O((rn_qn_t)k)$ for updating $\vec{\mathbf{B}}^{(t)}$ and $O(((C_{t-1}+c_t)rn_t+r^3)k)$ for updating $\mathbf{W}^{(t)}$, where k is the maximum iteration number. For hash function learning, solving $\mathbf{P}_m^{(t)}$ requires $O((rd_mn_t+d_m^2n_t+d_m^3)k)$. For the anchor codes, $n_q \ll n_t$ and $N_q \ll n_t$. Accordingly, the overall computational cost of OH-ELS at each round grows linearly with the dataset scale n_t.

Experiments

Datasets

The proposed method is evaluated on two standard multi-modal benchmarks, MIRFlickr-25K and NUS-WIDE. MIRFlickr-25K contains 25,000 images with corresponding textual annotations. Following (Zhang et al. 2022; Sun et al. 2023), we select 20,015 data for the experiment, and each one is labeled by at least one of the 24 classes. The visual data and textual data are depicted by 512-dimensional GIST features and 1386-dimensional BoW features, respectively. NUS-WIDE contains 269,468 image-text pairs of 81 classes. Each image is portrayed through a 500-dimensional SIFT feature vector, and each text is conveyed via a 1000-dimensional binary tagging vector. Following (Chen et al. 2020; Qin et al. 2022), 186,577 image-text pairs with the 10 most frequent classes are selected for the experiment.

Experimental Settings

We compare OH-ELS with the state-of-the-art online cross-modal hashing methods, i.e., OCMFH (Wang et al. 2020), LEMON (Wang, Luo, and Xu 2020), DOCH (Zhan et al. 2022), ROH (Jiang et al. 2023a), ODCH (Kang et al. 2023), ROHLSE (Li et al. 2024) and OHPKL (Shu, Li, and Yu 2024), among which OCMFH is unsupervised approach and the others are supervised ones. Parameter α is set to $1e2$ and parameter λ is set to [4, 8, 10] for [16-bit, 32-bit, 64-bit].

To simulate class-incremental scenarios, we divide MIRFlickr-25K and NUS-WIDE into multiple chunks and ensure that in each chunk, the data contains both old classes from previous rounds and novel classes not seen before ($c_t > 0$). At each round, 10% of incoming data is selected as the query set to dynamically assess retrieval performance. We carry out two fundamental cross-modal retrieval tasks, i.e., Image-to-Text (I2T) and Text-to-Image (T2I). The mean average precision (mAP) is leveraged to evaluate the retrieval performance. To make the baseline adaptable, we zero-padded the labels to make the label dimension constant during the training phase. The experiments were conducted using an AMD AI 9-HX370 CPU running at 5.1 GHz and 32 GB of RAM.

Retrieval Performance Evaluation

The average mAP scores across all rounds on MIRFlickr-25K and NUS-WIDE are listed in Table 1, while mAP scores for the last round are recorded in Table 2. The hash code length ranges from 16 to 64 bits, with the optimal outcomes indicated in bold. OH-ELS outperforms all baselines by achieving the highest mAP scores, demonstrating its effectiveness for cross-modal retrieval under class-incremental settings. In terms of the last round mAP scores, OH-ELS outperforms the baselines with an average improvement of 7.60% in I2T and 7.91% in T2I on MIRFlickr-25K. On NUS-WIDE, it achieves improvements of 2.10% and 9.26% in I2T and T2I, respectively. Referring to average mAP scores, OH-ELS surpasses the second-best method by 13.34% in I2T and 8.34% in T2I on MIRFlickr-25K. On NUS-WIDE, OH-ELS takes the lead by 10.48% in I2T and 10.37% in T2I.

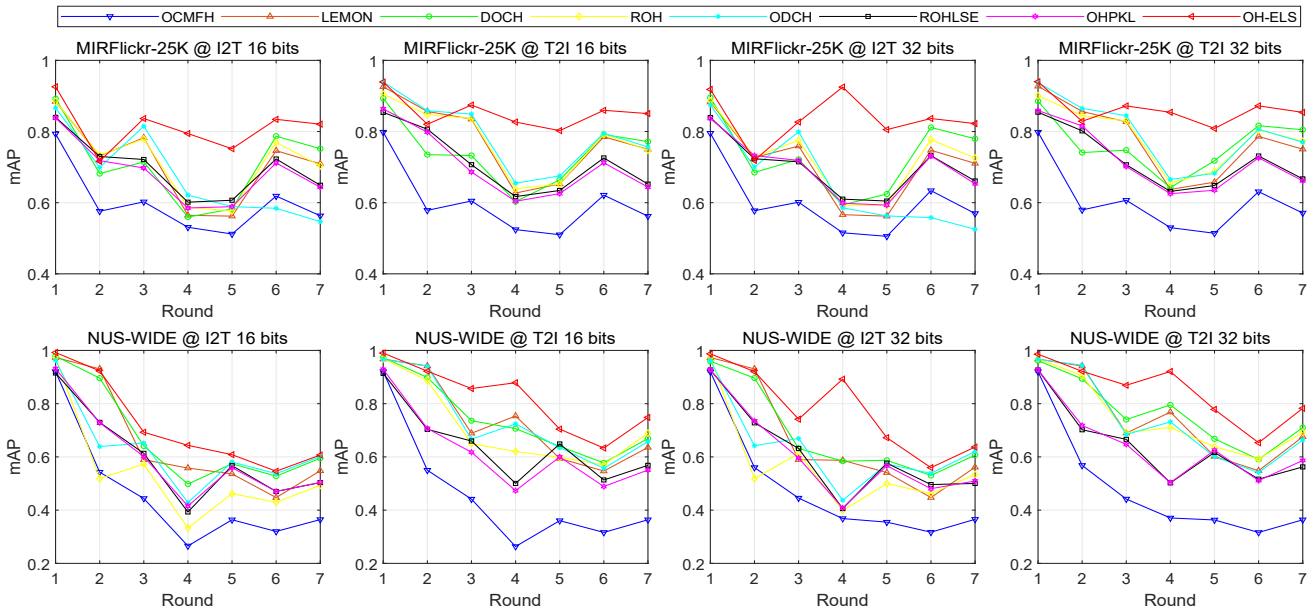
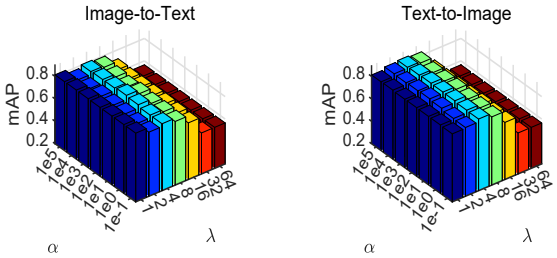
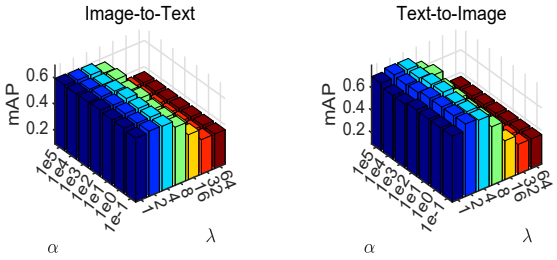


Figure 2: mAP-round curves on MIRFlickr-25K and NUS-WIDE.



(a) MIRFlickr-25K



(b) NUS-WIDE

Figure 3: mAP values versus α and λ.

Figure 2 displays the mAP-round curves for 16-bit and 32-bit codes, which indicates that OH-ELS consistently outperforms the baseline methods in almost all training rounds. Lacking label supervision, OCMFH performs less effectively than the other approaches. LEMON captures both the internal associations within the new data and the cross-chunk correlations between existing and incoming data. However, the relaxation strategy it adopts introduces semantic loss during the quantization of continuous representations into discrete hash codes, which in turn compromises retrieval accuracy. While DOCH leverages discrete optimization to enhance hash code quality, it fails to preserve the intrinsic neighborhood relationships among data points. More importantly, the above methods focus solely on sample-level increments while neglecting class-level expansion. By jointly employing anchor codes replay and consistency regularization, OH-ELS effectively preserves acquired knowledge at both the sample and class levels, and leverages it to guide the learning of new hash codes, thereby achieving superior retrieval performance in class-incremental scenarios.

Bits	Method	R1	R2	R3	R4	R5	R6	R7
16	OCMFH	5.81	6.76	6.93	8.08	6.21	7.28	6.89
	LEMON	0.07	0.10	0.06	0.08	0.06	0.08	0.08
	DOCH	0.73	1.22	1.31	2.56	1.71	2.82	3.09
	ROH	3.88	4.39	3.91	5.14	4.21	4.88	4.67
	ODCH	0.87	1.11	0.81	1.18	0.78	1.02	0.97
	ROHLSE	1.63	1.07	0.94	1.11	0.90	0.99	0.94
	OHPKL	2.12	2.28	2.32	2.42	2.24	2.32	2.33
	OH-ELS	0.36	1.20	0.89	2.24	1.04	2.11	1.97
32	OCMFH	6.96	7.84	6.82	8.33	6.48	7.32	6.93
	LEMON	0.08	1.01	0.07	1.00	0.06	0.08	0.08
	DOCH	0.96	2.01	2.17	4.40	2.91	5.14	5.88
	ROH	4.10	4.55	4.24	4.82	4.16	4.53	4.32
	ODCH	0.94	1.19	0.94	1.36	0.88	1.15	1.08
	ROHLSE	1.61	1.09	1.00	1.17	0.93	1.05	1.03
	OHPKL	2.22	2.32	2.37	2.39	2.33	2.44	2.37
	OH-ELS	0.61	2.24	1.53	3.91	1.96	4.06	3.73

Table 3: Training time evaluation for MIRFlickr-25K (in seconds).

Parameter Sensitivity Analysis

In this section, we examine the sensitivity of the parameters α and λ on MIRFlickr-25K and NUS-WIDE. For each parameter, a candidate set is constructed. The parameter of interest is systematically varied within this set while keeping the other parameter fixed, and the resulting mAP values are collected to select the best-performing configuration, as

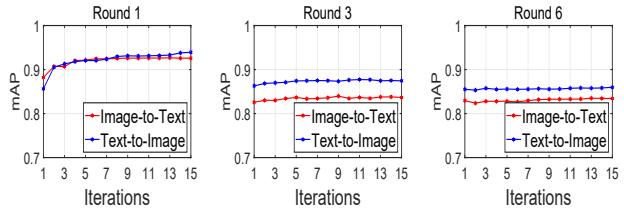


Figure 4: Convergence analysis on MIRFlickr-25K.

shown in Figure 3. α controls the consistency regularization term and λ influences similarities preservation. Experimental results reveal that the proposed method exhibits a relatively high sensitivity to λ. Specifically, as λ increases, the mAP values drop on both MIRFlickr-25K and NUS-WIDE. In contrast, the method demonstrates low sensitivity to variations in α; the mAP remains relatively stable under different α values.

Training Time Evaluation

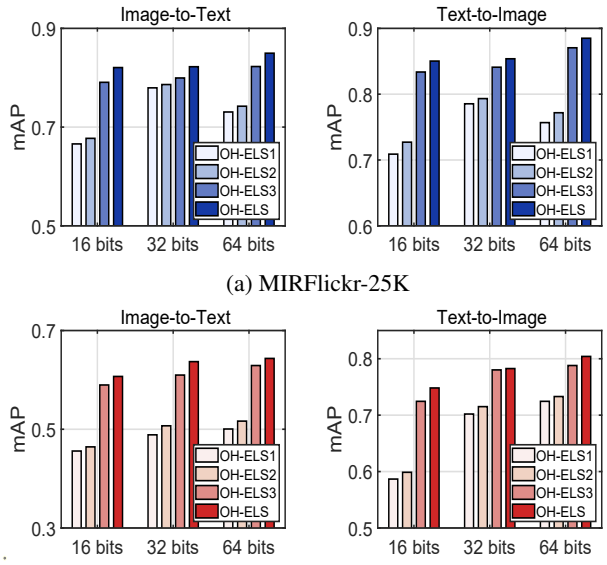
A comparative analysis of the training time between OH-ELS and the baseline methods on MIRFlickr-25K is provided in Table 3. OCMFH exhibits the lowest efficiency because it keeps revising the binary codes of existing data as the model evolves. LEMON achieves the highest efficiency, owing to its efficient online optimization algorithm. ROH suffers from reduced efficiency as it optimizes over a large number of variables. As both DOCH and OH-ELS update hash codes bit by bit, their training efficiencies are comparable, with OH-ELS showing a slight advantage. In general, OH-ELS achieves optimal accuracy while maintaining an acceptable training efficiency.

Convergence Analysis

The proposed method is updated using an alternating optimization scheme. We investigate its convergence property by recording the mAP values at different iterations across several representative rounds. Specifically, we plot the mAP values for the first, third, and sixth rounds with respect to iteration steps on MIRFlickr-25K, as shown in Figure 4. The results suggest the model needs multiple iterations for convergence in the initial round, whereas for $t \geq 2$, the mAP curves of both I2T and T2I show a steady trend across iteration steps, demonstrating that the value of the objective function monotonously decreases and eventually stabilizes at the optimal value. These findings empirically validate the efficiency of the proposed discrete optimization scheme.

Ablation Studies

By transferring historical knowledge at both sample and class levels, our proposed OH-ELS method achieves strong retrieval performance under class-incremental settings. To evaluate the impact of each component of OH-ELS, three ablated variants are developed: OH-ELS1, OH-ELS2 and OH-ELS3. OH-ELS2 retains only the pairwise relationships within newly arrived data, which ignores the sample correlations between previous and new data (by removing the second term from Eq. (6)). OH-ELS3 drops the consistency regularization term, i.e., $\|\bar{\mathbf{W}}^{(t)} - \mathbf{W}^{(t-1)}\|_F^2$. OH-ELS1 serves as a base model that excludes both terms. As illustrated in Figure 5, OH-ELS demonstrates the highest performance among all examined variants in the ablation study, while OH-ELS1, which does not consider knowledge transfer at all, obtains the lowest result, confirming the effectiveness of each component. The results of OH-ELS2 show that retrieval performance significantly declines on both MIRFlickr-25K and NUS-WIDE when the cross-chunk similarities preservation term is removed, which validates the importance of preserving sample correlations between historical and new data. The results of OH-ELS3 also prove the effectiveness of the consistency regularizer applied to classifiers.



(a) MIRFlickr-25K

(b) NUS-WIDE

Figure 5: Ablation results on MIRFlickr-25K and NUS-WIDE.

Conclusion

This paper investigates a critical but under-explored challenge: retrieval over multimodal data with an expanding label space. Most existing online hashing methods assume a fixed and known label space, which reduces their effectiveness in practical scenarios where novel classes appear progressively. To overcome this limitation, we introduce OH-ELS, which manages sample- and class-level dynamics through knowledge transfer across both levels. At the sample level, a subset of binary codes from previous data is revisited to capture both local and cross-chunk sample correlations. At the class level, knowledge from historical classes is leveraged to guide the training of new classifiers via a consistency regularization mechanism. Extensive experiments demonstrate that OH-ELS outperforms state-of-the-art online cross-modal hashing approaches under class-incremental settings.

Acknowledgments

This work was partially supported by the National Natural Science Foundation of China under Grants Nos. 62576161, 62176116, and 62276136.

References

Chen, Y.; Zhang, H.; Tian, Z.; Wang, J.; Zhang, D.; and Li, X. 2020. Enhanced discrete multi-modal hashing: More constraints yet less time to learn. *IEEE TKDE*, 34(3): 1177–1190.

De Lange, M.; and Tuytelaars, T. 2021. Continual prototype evolution: Learning online from non-stationary data streams. In *ICCV*, 8250–8259.

Gao, Q.; Zhao, C.; Ghanem, B.; and Zhang, J. 2022. R-dfcil: Relation-guided representation learning for data-free class incremental learning. In *ECCV*, 423–439. Springer.

Han, K.; Liu, Y.; Wei, R.; Zhou, K.; Xu, J.; and Long, K. 2024. Supervised Hierarchical Online Hashing for Cross-modal Retrieval. *ACM TOMM*, 20(4): 1–23.

Hou, C.; Gu, S.; Xu, C.; and Qian, Y. 2023. Incremental learning for simultaneous augmentation of feature and class. *IEEE TPAMI*, 45(12): 14789–14806.

Hou, S.; Pan, X.; Loy, C. C.; Wang, Z.; and Lin, D. 2019. Learning a unified classifier incrementally via rebalancing. In *CVPR*, 831–839.

Jiang, K.; Wong, W. K.; Fang, X.; Li, J.; Qin, J.; and Xie, S. 2023a. Random online hashing for cross-modal retrieval. *IEEE TNNLS*.

Jiang, Q.-Y.; and Li, W.-J. 2019. Discrete latent factor model for cross-modal hashing. *IEEE TIP*, 28(7): 3490–3501.

Jiang, X.; Liu, X.; Cheung, Y.-M.; Xu, X.; Zheng, S.; and Li, T. 2023b. Label-Semantic-Enhanced Online Hashing for Efficient Cross-modal Retrieval. In *ICME*, 984–989. IEEE.

Jin, H.; Zhang, Y.; Shi, L.; Zhang, S.; Kou, F.; Yang, J.; Zhu, C.; and Luo, J. 2024. An end-to-end graph attention network hashing for cross-modal retrieval. *NeurIPS*, 37: 2106–2126.

Kang, M.; Park, J.; and Han, B. 2022. Class-incremental learning by knowledge distillation with adaptive feature consolidation. In *CVPR*, 16071–16080.

Kang, X.; Liu, X.; Xue, W.; Zhang, X.; Nie, X.; and Yin, Y. 2024. Discrete online cross-modal hashing with consistency preservation. *Pattern Recognition*, 110688.

Kang, X.; Liu, X.; Zhang, X.; Nie, X.; and Yin, Y. 2023. Online discriminative cross-modal hashing. *IEEE TCSVT*, 34(7): 5242–5254.

Kang, X.; Liu, X.; Zhang, X.; Xue, W.; Nie, X.; and Yin, Y. 2025. Semi-Supervised Online Cross-Modal Hashing. In *AAAI*, volume 39, 17770–17778.

Lei, F.; Zhang, C.; Li, H.; Gao, Y.; and Chen, C. 2024. Label Distribution Guided Hashing for Cross-Modal Retrieval. *ACM TKDD*, 19(1): 1–23.

Li, H.; Zhang, C.; Jia, X.; Gao, Y.; and Chen, C. 2021. Adaptive label correlation based asymmetric discrete hashing for cross-modal retrieval. *IEEE TKDE*, 35(2): 1185–1199.

Li, J.; Jiang, L.; Ma, Z.; Jiang, K.; Fang, X.; and Wen, J. 2025a. Lightweight Contrastive Distilled Hashing for Online Cross-modal Retrieval. In *AAAI*, volume 39, 4779–4787.

Li, L.; Shu, Z.; Yu, Z.; and Wu, X.-J. 2024. Robust online hashing with label semantic enhancement for cross-modal retrieval. *Pattern Recognition*, 145: 109972.

Li, Y.; Long, J.; and Yang, Z. 2025. Asymmetric Cross-Modal Hashing Based on Formal Concept Analysis. In *AAAI*, volume 39, 1392–1401.

Li, Y.; Zhen, L.; Sun, Y.; Peng, D.; Peng, X.; and Hu, P. 2025b. Deep Evidential Hashing for Trustworthy Cross-Modal Retrieval. In *AAAI*, volume 39, 18566–18574.

Liu, X.; Yi, J.; Cheung, Y.-m.; Xu, X.; and Cui, Z. 2022. OMGH: Online manifold-guided hashing for flexible cross-modal retrieval. *IEEE TMM*, 25: 3811–3824.

Liu, Y.; Zhang, Y.; Fu, H.; and Gu, G. 2025. Adaptive Asymmetric Online Hashing for Cross-Modal Retrieval. In *ICMR*, 908–916.

Meng, M.; Wang, H.; Yu, J.; Chen, H.; and Wu, J. 2020. Asymmetric supervised consistent and specific hashing for cross-modal retrieval. *IEEE TIP*, 30: 986–1000.

Park, J.; Kang, M.; and Han, B. 2021. Class-incremental learning for action recognition in videos. In *ICCV*, 13698–13707.

Peng, S.-J.; Yi, J.; Liu, X.; Cheung, Y.-m.; Cui, Z.; and Li, T. 2024. OLCH: Online Label Consistent Hashing for streaming cross-modal retrieval. *Pattern Recognition*, 150: 110335.

Qin, J.; Fei, L.; Zhang, Z.; Wen, J.; Xu, Y.; and Zhang, D. 2022. Joint specifics and consistency hash learning for large-scale cross-modal retrieval. *IEEE TIP*, 31: 5343–5358.

Rebuffi, S.-A.; Kolesnikov, A.; Sperl, G.; and Lampert, C. H. 2017. icarl: Incremental classifier and representation learning. In *CVPR*, 2001–2010.

Shu, Z.; Li, L.; and Yu, Z. 2024. Online hashing with partially known labels for cross-modal retrieval. *Engineering Applications of Artificial Intelligence*, 138: 109367.

Sun, Y.; Dai, J.; Ren, Z.; Chen, Y.; Peng, D.; and Hu, P. 2024. Dual Self-Paced Cross-Modal Hashing. In *AAAI*, volume 38, 15184–15192.

Sun, Y.; Ren, Z.; Hu, P.; Peng, D.; and Wang, X. 2023. Hierarchical consensus hashing for cross-modal retrieval. *IEEE TMM*, 26: 824–836.

Tang, S.; Chen, D.; Zhu, J.; Yu, S.; and Ouyang, W. 2021. Layerwise optimization by gradient decomposition for continual learning. In *CVPR*, 9634–9643.

Wang, D.; Wang, Q.; An, Y.; Gao, X.; and Tian, Y. 2020. Online collective matrix factorization hashing for large-scale cross-media retrieval. In *SIGIR*, 1409–1418.

Wang, Y.; Luo, X.; and Xu, X.-S. 2020. Label embedding online hashing for cross-modal retrieval. In *MM*, 871–879.

Wu, X.-M.; Luo, X.; Zhan, Y.-W.; Ding, C.-L.; Chen, Z.-D.; and Xu, X.-S. 2022. Online enhanced semantic hashing: Towards effective and efficient retrieval for streaming multi-modal data. In *AAAI*, volume 36, 4263–4271.

Xie, L.; Shen, J.; and Zhu, L. 2016. Online cross-modal hashing for web image retrieval. In *AAAI*, volume 30.

Yao, T.; Wang, G.; Yan, L.; Kong, X.; Su, Q.; Zhang, C.; and Tian, Q. 2019. Online latent semantic hashing for cross-media retrieval. *Pattern Recognition*, 89: 1–11.

Yi, J.; Liu, X.; Cheung, Y.-m.; Xu, X.; Fan, W.; and He, Y. 2021. Efficient online label consistent hashing for large-scale cross-modal retrieval. In *ICME*, 1–6. IEEE.

Zhan, Y.-W.; Wang, Y.; Sun, Y.; Wu, X.-M.; Luo, X.; and Xu, X.-S. 2022. Discrete online cross-modal hashing. *Pattern Recognition*, 122: 108262.

Zhang, C.; Li, H.; Gao, Y.; and Chen, C. 2022. Weakly-supervised enhanced semantic-aware hashing for cross-modal retrieval. *IEEE TKDE*, 35(6): 6475–6488.

Zhang, D.; Wu, X.-J.; and Chen, G. 2023. ONION: Online Semantic Autoencoder Hashing for Cross-Modal Retrieval. *ACM TIST*, 14(2): 1–18.

Zhou, D.-W.; Wang, Q.-W.; Qi, Z.-H.; Ye, H.-J.; Zhan, D.-C.; and Liu, Z. 2024. Class-incremental learning: A survey. *IEEE TPAMI*, 46(12): 9851–9873.

Zhou, D.-W.; Ye, H.-J.; and Zhan, D.-C. 2021. Co-transport for class-incremental learning. In *MM*, 1645–1654.