Asymmetric Cross-Modal Hashing Based on Formal Concept Analysis

Yinan Li, Jun Long, Zhan Yang*

Big Data Institute, Central South University, Changsha, China
{liyinan, junlong, zyang22}@csu.edu.cn

Abstract

Hashing has been widely applied in large-scale multimodal retrieval by mapping heterogeneous modalities data into binary codes. However, most cross-modal hashing methods cannot make the most of semantic information to construct the association relations of sample pairs, resulting in unsatisfactory retrieval accuracy. Concept lattice is a powerful tool for data mining and information retrieval, and for all we know, this is the first time to combine formal concept analysis and hash learning to improve cross-modal hashing retrieval performance. In this paper, we propose a novel framework for **A**symmetric **C**ross-modal **H**ashing based on **F**ormal **C**oncept **A**nalysis, denoted as **ACHFCA**. Initially, a flash-projection three-layer semantic enhancement descriptor is designed to extract latent representations from heterogeneous modalities. Subsequently, an asymmetric hash learning framework is established to enhance the semantics of different layers based on the fine-grained similarity values reconstructed from concept lattice to reinforce the discriminative competence of the model. Finally, an effective discrete optimization algorithm is proposed, which can directly learn compact hash codes. Comprehensive experiments on MIRFlickr, NUS-WIDE and IAPR-TC12 datasets demonstrate the superior performance of **ACHFCA** to state-of-the-art hashing approaches.

Introduction

With the explosive increase in multimodal data, conventional information retrieval methodologies are encountering significant obstacles (Bin et al. 2023; Song, Chen, and Jiang 2023; Zhao et al. 2023). To improve retrieval efficiency, hashing-based encoding mechanisms have become popular, particularly for dealing with data from diverse modalities (Liu et al. 2023; Tu et al. 2023; Li et al. 2023b). Hashing intends to project high-dimensional data into binary formats within a Hamming space, by which sample similarity is calculated via Hamming distances. Taking advantage of the above benefits, researchers have designed plenty of cross-modal hashing approaches from various perspectives.

Existing cross-modal hashing can be separated into supervised (Luo et al. 2023; Yang et al. 2024; Sun et al. 2024b) and unsupervised (Liu et al. 2020; Li, Zheng, and Sun 2022; Zhong et al. 2023) methods depending on whether semantic labels are used or not. In the absence of semantic information, unsupervised methods rely solely on original data for training the model, resulting in unsatisfactory accuracy. FSH (Liu et al. 2017) integrates fusion similarity based on graph structure within the Hamming space to optimize hash code learning. JIMFH (Wang et al. 2020) derives the semantic representation relations from diverse modalities through matrix factorization. CUH (Wang et al. 2021a) maps multimodal data into a latent space through multi-view clustering, enabling the learning of linear hash function, and thus generating compact hash codes. However, the above methods are unable to capture the actual semantic relevance among heterogeneous modalities.

In contrast, supervised methods employ semantic annotation to improve the discriminative capacity of hash codes, leading to superior retrieval performance. SMFH (Liu et al. 2016) establishes the architecture between graph regularization and matrix factorization preserving sample similarity. SCRATCH (Chen et al. 2020) gets the utmost out of semantic annotation to derive the correlation from diverse modalities, and utilizes the two-step learning strategy to decrease the training cost of model. With the aid of label information, recent supervised methods can accurately represent the relevance among multimodal data, and perform superior in semantic alignment. SRLCH (Shen et al. 2021) leverages label information to adjust the relations within the Hamming subspace and utilizes a symmetric framework to minimize the gap between binary representation and corresponding correlations, ensuring similar data that come from diverse modalities are positioned nearer to one another within the Hamming space. The label is regarded as an additional modality in conjunction with image and text in BATCH (Wang et al. 2021b), facilitating the embedding of similar information into binary codes through the minimization of distance discrepancy. ASCSH (Meng et al. 2021) factorizes the projection matrix into common and specific components to investigate the inherent correlations among diverse types of data through the asymmetric framework. DSCMH (Sun et al. 2024a) mimics human beings in constructing hashing algorithms from "easy" to "hard" and thus can reduce the negative impact of the learning process.

Although the above methods have achieved considerable results, humanized similarity reconstruction schemes

*Corresponding author

for the semantic gap among multi-modalities are the focus of cross-modal research. Formal Concept Analysis (FCA) (Wille 1982) has attracted increasing popularity across various domains (Akram, Nawaz, and Deveci 2023; Hu et al. 2023), such as potential relation mining and information retrieval (Zhi and Li 2023). Furthermore, three-way concept analysis (Qi, Qian, and Wei 2016) was proposed to simultaneously exploit both positive and negative attributes. The initial information of FCA is formal context, *i.e.*, a Boolean matrix, which is comprised of objects and attributes among binary relations. The data structure derived from formal context is concept lattice, consisting of a certain amount of specific pairs, *i.e.*, formal concepts, describing the target set through common attributes. When the objects are fixed, the selection of attributes determines the structure and size of the concept lattice. The generalization and specialization between concepts can reflect the hierarchical relations via the concept lattice.

The loci of maximum transmission will indicate the regions in which the primary wave had the same phase as the modified wave (GABOR 1948), in general, a flash source can be used to illuminate the object to magnify which features in a more detailed view of the particular perspective of the projected surface. Motivated by the above theory, we treat the label information as the flash source, the multimodal data as the object to be irradiated, and the hash codes as the detailed semantic projection of heterogeneous modal data, *i.e.*, the flash-projection model. Besides, although the conventional similarity reconstruction strategy covers a wide range of labels, it has problems of data noise and semantic error due to only considering the corresponding relations between instances and labels, ignoring the concept information that acts as a bridge. In order to improve the accuracy of sample similarity calculation, it is necessary to add a concept layer between the instances and labels.

To maintain the information integrality among heterogeneous modalities, we propose a novel framework for **A**symmetric **C**ross-modal **H**ashing based on FCA (**ACHFCA**), which intends to derive latent representations from heterogeneous modalities via a three-layer semantic enhancement descriptor. Subsequently, an asymmetric hash learning framework based on FCA among different layers of the semantic enhancement descriptor is proposed, which can learn more discriminative hash codes during optimization. And for all we know, this is the first time to combine FCA and hashing to improve hashing retrieval performance. The contributions lie in three folds.

- A flash-projection three-layer semantic enhancement descriptor is designed to extract latent representations from heterogeneous modalities.
- A novel double asymmetric hash learning framework based on FCA is established, which can leverage the abundant semantic information extracted from different layers of the semantic enhancement descriptor to directly learn compact hash codes.
- The proposed **ACHFCA** is easily extensible and more powerful than many state-of-the-art supervised cross-modal hashing methods.

Proposed Method

Notations

Assume a set of training data with M modalities, $\mathbf{X} = \{\mathbf{X}^1, ..., \mathbf{X}^{(m)}\}$, where $\mathbf{X}^{(m)} = \{\mathbf{x}_1, ..., \mathbf{x}_n\} \in \mathbb{R}^{d_m \times n}$, d_m is the dimension of modality $\mathbf{X}^{(m)}$, n is the number of instances. In addition, $\mathbf{B} \in \{-1, 1\}^{k \times n}$ and $\mathbf{L} \in \mathbb{R}^{c \times n}$ represent the hash codes and semantic labels, where k is the length of hash codes, c is the number of tags. $\mathbf{B}^\top$ and $tr(\mathbf{B})$ represent the transpose and the trace of $\mathbf{B}$, respectively. For simplicity, in this paper, matrices and vectors are denoted by bold uppercase and bold lowercase letters, and scalars are represented by regular lowercase letters.

Basic Definitions in FCA

Let $\mathbf{N} = (U, V, I)$ be a formal context. Concretely, U is a nonempty set of objects, V is a nonempty set of attributes, and I denotes the relation between U and V. Besides, we use $I(y, z) = 1$ (resp. $I(y, z) = 0$) to express that the object y contains (resp. does not contain) the attribute z.

Definition 1. (Wille 1982) For $Y \in 2^U$ and $Z \in 2^V$, a pair of operators $* : 2^U \to 2^V$ and $* : 2^V \to 2^U$ are defined as:

$$\begin{aligned} Y^* &= \{z \in V \mid \forall y \in Y, I(y, z) = 1\}, \\ Z^* &= \{y \in U \mid \forall z \in Z, I(y, z) = 1\}. \end{aligned} \tag{1}$$

If $Y^* = Z$ and $Z^* = Y$, then (Y, Z) is called a formal concept, Y and Z are called the extent and intent of the concept, respectively. All concepts contained in $\mathbf{N}$ form a complete lattice, which is denoted by $\mathcal{L}(\mathbf{N})$. Besides, (y^{**}, y^*) is called an object concept.

An implicit concept layer between objects and attributes in the hierarchical space is presented. Different objects share concepts of various granularity. When the granularity of concept is general, the corresponding objects are more and the attributes are less, and vice versa. The semantic relation between objects can be measured from the concept viewpoint. Particularly, when two objects co-appear in multiple concepts, it indicates that the two objects process a strong semantic relation.

Proposition 1. The similarity among concepts can be calculated via structural position in the concept lattice, *i.e.*, searching for the nearest common parent node and shortest path of two concepts. Specifically, for any concept T_i and T_j, the similarity is calculated as,

$$\text{pos}(T_i, T_j) = \frac{2 \times \text{depth(NCP)}}{2 \times \text{depth(NCP)} + \text{distance}(T_i, T_j)}, \tag{2}$$

where depth(NCP) represents the degree of nearest common parent node, and distance(T_i, T_j) represents the shortest path length between two concepts.

Example 1. In Table 1, formal context $\mathbf{N} = (U, V, I)$ demonstrates four instances with five characteristics from MIRFlickr (Huiskes and Lew 2008) dataset. Specifically, U is an object set consisting of four instances and V is an attribute set consisting of five characteristics. For example, l_1, l_2, l_3, l_4 and l_5 denote *sky*, *water*, *night*, *sunset* and *clouds*, respectively.

Fig. 1 depicts six implicit concepts in the lattice $\mathcal{L}(\mathbf{N})$ and corresponding hierarchical relations, by which common

U	l_1	l_2	l_3	l_4	l_5
1	1	1	0	1	1
2	1	1	1	0	0
3	0	0	0	1	0
4	1	1	1	0	0

Table 1: The formal context $\mathbf{N}$ of Example 1.

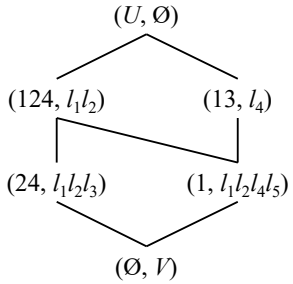

Figure 1: The concept lattice $\mathcal{L}(\mathbf{N})$ of Example 1. For convenience, as an example, we use $(124, l_1l_2)$ instead of $(\{1, 2, 4\}, (\{l_1, l_2\})$.

characteristics of instances can be analyzed. For instance, concept $(\{2, 4\}, (\{l_1, l_2, l_3\})$ manifests that both instances 2 and 4 possess *sky*, *water* and *night*. Specifically, each lower concept inherits all of the attributes in the upper concept, the further down the concept lattice, the more specific the concept, and the less the corresponding extent.

ACHFCA Framework

The proposed **ACHFCA** depicts a flash-projection three-layer semantic enhancement descriptor from label information to hash codes in Fig. 2. In addition, a double asymmetric hashing framework via extracted heterogeneous modalities representations is constructed, which can directly learn hash codes. The details of **ACHFCA** framework are as follows.

Semantic hash mapping layer-$\mathbf{W}_1^{(m)}$ First of all, the semantic hash mapping layer is designed, *i.e.*, the 1st layer, considers the semantic annotation into primary binary codes as follows,

$$\mathbf{G}_1^{(m)} = \mathbf{W}_1^{(m)}\mathbf{L}, \tag{3}$$

where $\mathbf{W}_1^{(m)} \in \mathbb{R}^{k\times c}$ is the semantic hash mapping matrix of m-th modality, $\mathbf{G}_1^{(m)} \in \mathbb{R}^{k\times n}$ is the output of the 1st layer for m-th modality. Considering the regularization terms, Eq. (3) can be reformulated as,

$$L_1 = \sum_{m=1}^{M}(||\mathbf{W}_1^{(m)}\mathbf{L} - \mathbf{G}_1^{(m)}||^2 + Re(\mathbf{W}_1^{(m)})), \tag{4}$$

where $Re(\cdot)$ is the regularization term.

Feature alignment layer-$\mathbf{W}_2^{(m)}$ The initial features of heterogeneous modalities data comprise imprecise information. In order to eliminate the negative impacts, a feature alignment layer is proposed, *i.e.*, the 2nd layer, to align the latent representations as follows,

$$\mathbf{G}_2^{(m)} = \mathbf{W}_2^{(m)}\mathbf{G}_1^{(m)}, \tag{5}$$

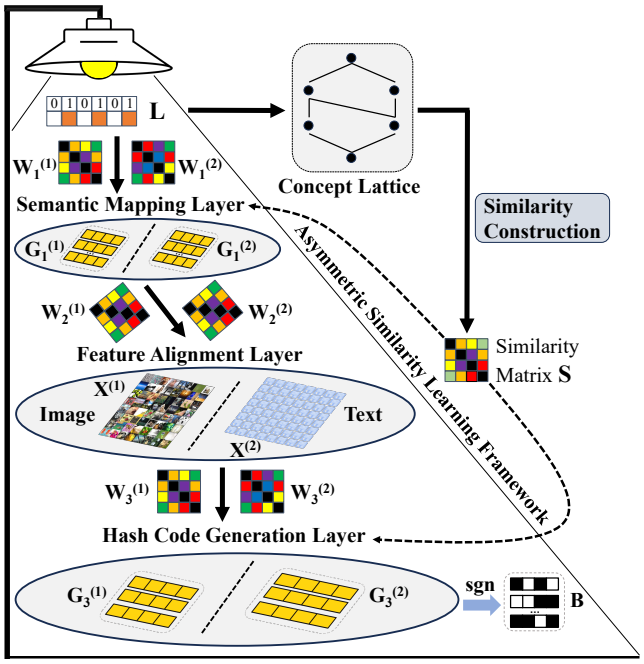

Figure 2: The pipeline of the proposed **A**symmetric **C**ross-modal **H**ashing based on **F**ormal **C**oncept **A**nalysis (**ACHFCA**).

where $\mathbf{W}_2^{(m)} \in \mathbb{R}^{d_m\times k}$ is the alignment mapping matrix of m-th modality representation, $\mathbf{G}_2^{(m)}$ is the output of the 2nd layer for m-th modality. Note that the output of the 2nd layer is related to initial features, *i.e.*, $\mathbf{X}^{(m)} = \mathbf{G}_2^{(m)}$, to improve the discriminative competency of hash codes. Considering the regularization terms, Eq. (5) can be reformulated as,

$$L_2 = \sum_{m=1}^{M}(||\mathbf{W}_2^{(m)}\mathbf{W}_1^{(m)}\mathbf{L} - \mathbf{X}^{(m)}||^2 + Re(\mathbf{W}_2^{(m)})). \tag{6}$$

Hash code generation layer-$\mathbf{W}_3^{(m)}$ In order to learn compact hash codes, a code generation layer is designed, *i.e.*, the 3rd layer, which is devoted to integrating the variables of previous layers and generat hash codes as follows,

$$\mathbf{G}_3^{(m)} = \mathbf{W}_3^{(m)}\mathbf{G}_2^{(m)}, \tag{7}$$

where $\mathbf{W}_3^{(m)} \in \mathbb{R}^{k\times d_m}$ is the hash mapping matrix of m-th modality, $\mathbf{G}_3^{(m)} \in \mathbb{R}^{k\times n}$ is the output of the 3rd layer for m-th modality. Besides, $\mathbf{G}_3^{(m)}$ is designed to generate the hash codes, which can be formulated as $\mathbf{B}^{(m)} = \text{sgn}(\mathbf{G}_3^{(m)})$, where $\text{sgn}(\cdot)$ denotes the element-wise indicator. Considering the regularization terms and minimizing the quantization gap between real-valued and binary embedding, Eq. (7) can be reformulated as,

$$L_3 = \sum_{m=1}^{M}(||\mathbf{W}_3^{(m)}\mathbf{W}_2^{(m)}\mathbf{W}_1^{(m)}\mathbf{L} - \mathbf{B}^{(m)}||^2 + Re(\mathbf{W}_3^{(m)})). \tag{8}$$

Asymmetric semantic hashing based on FCA Asymmetric semantic hash learning (Liu et al. 2012) preserves the linkage of pairs with two identical binary representations, which can be formulated as,

$$\min_{\mathbf{B}} ||\mathbf{B}^\top \mathbf{B} - k\mathbf{S}||^2 \; s.t. \; \mathbf{B} \in \{-1, 1\}^{k \times n}, \tag{9}$$

where $\mathbf{S}$ is the **pairwise similarity matrix**.

For label matrix $\mathbf{L}$, we treat the instances as objects, and semantic annotations as attributes. Therefore, $\mathbf{L}$ can be regarded as a formal context, for any object y_i and y_j, the similarity can be calculated as,

$$\text{sim}(y_i, y_j) = \rho |{y_i}^* \cap {y_j}^*| + (1 - \rho) \frac{\text{pos}(T_i, T_j)}{|{y_i}^* \cup {y_j}^*|}, \tag{10}$$

where ρ is the trade-off parameter of generalization and specialization between concepts, and $T_i = (y_i^{**}, y_i^*)$, $T_j = (y_j^{**}, y_j^*)$.

Therefore, the humanized fine-grained similarity matrix $\mathbf{S}$ among all objects can be calculated via Eq. (10). Some works (Da et al. 2017; Meng et al. 2021; Yang et al. 2024) demonstrate that the asymmetric structure can solve the accuracy. Thus, we construct the asymmetric learning framework as follows,

$$\min_{\mathbf{B}_i, \mathbf{B}_j} ||{\mathbf{B}_i}^\top \mathbf{B}_j - k\mathbf{S}||^2 \; s.t. \; \mathbf{B}_i, \mathbf{B}_j \in \{-1, 1\}^{k \times n}. \tag{11}$$

Note that $\mathbf{B}_i$ and $\mathbf{B}_j$ are hash codes generated from different layers of the semantic enhancement descriptor. Let $\mathbf{B} = \mathbf{B}_i = \mathbf{B}_j$ instead of minimizing $||\mathbf{B}_i - \mathbf{B}_j||$. Thus, an asymmetric learning framework is constructed by using real-value $\mathbf{G}$ instead of the binary-value $\mathbf{B}$ to get closer to the optimal approximate solution. Combining the discrete constraints, Eq. (11) can be rewritten as,

$$\begin{aligned} \min_{\mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}, \mathbf{B}} & ||{\mathbf{G}_1^{(1)}}^\top \mathbf{G}_1^{(2)} - k\mathbf{S}||^2 + \eta ||{\mathbf{G}_3^{(1)}}^\top \mathbf{G}_3^{(2)} - k\mathbf{S}||^2 \\ & + \lambda ||\mathbf{B} - \mathbf{G}_3^{(1)}||^2 + \omega ||\mathbf{B} - \mathbf{G}_3^{(2)}||^2 \\ & s.t. \; \mathbf{B} \in \{-1, 1\}^{k \times n}, \end{aligned} \tag{12}$$

where η, λ and ω are balance parameters.

Objective function Following deliberation on the three layers of semantic enhancement descriptor, the overall objective function can be formulated as,

$$\begin{aligned} \min_{\substack{\mathbf{W}_i^{(m)}|_{i=1,2,3}^{m=1,2}, \\ \mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}, \mathbf{B}}} & \sum_{m=1}^{M} (||\mathbf{W}_3^{(m)} \mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{G}_3^{(m)}||^2 \\ & + ||\mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{X}^{(m)}||^2 + ||\mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{G}_1^{(m)}||^2 \\ & + \gamma Re(\mathbf{W}_1^{(m)}, \mathbf{W}_2^{(m)}, \mathbf{W}_3^{(m)})) \\ & + ||{\mathbf{G}_1^{(1)}}^\top \mathbf{G}_1^{(2)} - k\mathbf{S}||^2 + \eta ||{\mathbf{G}_3^{(1)}}^\top \mathbf{G}_3^{(2)} - k\mathbf{S}||^2 \\ & + \lambda ||\mathbf{B} - \mathbf{G}_3^{(1)}||^2 + \omega ||\mathbf{B} - \mathbf{G}_3^{(2)}||^2 \\ & s.t. \; \mathbf{B} \in \{-1, 1\}^{k \times n}, \end{aligned} \tag{13}$$

where γ is the balance parameter.

Optimization

The optimization problem in Eq. (13) is not jointly convex when all variables are considered. Thus, we adopt the alternating direction minimization algorithm (Lin, Liu, and Su 2011) to optimize each variable with other variables fixed.

Update $\mathbf{W}_1^{(m)}$ When $\mathbf{W}_i^{(m)}|_{i=2,3}^{m=1,2}$, $\mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}$, and $\mathbf{B}$ are fixed, the optimization for $\mathbf{W}_1^{(m)}$ can be transformed into solving the following issue,

$$\begin{aligned} \min_{\mathbf{W}_1^{(m)}} \sum_{m=1}^{M} (& ||\mathbf{W}_3^{(m)} \mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{G}_3^{(m)}||^2 \\ & + ||\mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{X}^{(m)}||^2 \\ & + ||\mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{G}_1^{(m)}||^2 + \gamma Re(\mathbf{W}_1^{(m)})). \end{aligned} \tag{14}$$

Setting the derivative with respect to $\mathbf{W}_1^{(m)}$ to zero, the optimization for $\mathbf{W}_1^{(m)}$ equals the following,

$$\begin{aligned} & \mathbf{W}_1^{(m)} = \\ & ({\mathbf{W}_2^{(m)}}^\top {\mathbf{W}_3^{(m)}}^\top \mathbf{W}_3^{(m)} \mathbf{W}_2^{(m)} + {\mathbf{W}_2^{(m)}}^\top \mathbf{W}_2^{(m)} + (1 + \gamma) \mathbf{I}_k)^{-1} \\ & \times ({\mathbf{W}_2^{(m)}}^\top {\mathbf{W}_3^{(m)}}^\top \mathbf{G}_3^{(m)} \mathbf{L}^\top + {\mathbf{W}_2^{(m)}}^\top \mathbf{F}^{(m)} + \mathbf{G}_1^{(m)} \mathbf{L}^\top) \\ & \times (3\mathbf{E} + \gamma \mathbf{I}_c)^{-1}, \end{aligned} \tag{15}$$

where $\mathbf{E} = \mathbf{L}\mathbf{L}^\top$ and $\mathbf{F}^{(m)} = \mathbf{X}^{(m)} \mathbf{L}^\top$, which can be regarded as fixed variables calculated once outside of iterations, reducing the computation complexity from $\mathcal{O}(n)$ to $\mathcal{O}(1)$.

Update $\mathbf{W}_2^{(m)}$ When $\mathbf{W}_i^{(m)}|_{i=1,3}^{m=1,2}$, $\mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}$, and $\mathbf{B}$ are fixed, the optimization for $\mathbf{W}_2^{(m)}$ can be transformed into the following,

$$\begin{aligned} \min_{\mathbf{W}_2^{(m)}} \sum_{m=1}^{M} (& ||\mathbf{W}_3^{(m)} \mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{G}_3^{(m)}||^2 \\ & + ||\mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{X}^{(m)}||^2 + \gamma Re(\mathbf{W}_2^{(m)})). \end{aligned} \tag{16}$$

The optimization for $\mathbf{W}_2^{(m)}$ equals the following,

$$\begin{aligned} \mathbf{W}_2^{(m)} = & ({\mathbf{W}_3^{(m)}}^\top \mathbf{W}_3^{(m)} + (1 + \gamma) \mathbf{I}_{d_m})^{-1} \\ & \times ({\mathbf{W}_3^{(m)}}^\top \mathbf{G}_3^{(m)} \mathbf{L}^\top {\mathbf{W}_1^{(m)}}^\top + \mathbf{F}^{(m)} {\mathbf{W}_1^{(m)}}^\top) \\ & \times (2\mathbf{W}_1^{(m)} \mathbf{E} {\mathbf{W}_1^{(m)}}^\top + \gamma \mathbf{I}_k)^{-1}. \end{aligned} \tag{17}$$

Update $\mathbf{W}_3^{(m)}$ When $\mathbf{W}_i^{(m)}|_{i=1,2}^{m=1,2}$, $\mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}$, and $\mathbf{B}$ are fixed, the optimization for $\mathbf{W}_3^{(m)}$ can be transformed into the following,

$$\min_{\mathbf{W}_3^{(m)}} \sum_{m=1}^{M} (||\mathbf{W}_3^{(m)} \mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{L} - \mathbf{G}_3^{(m)}||^2 + \gamma Re(\mathbf{W}_3^{(m)})). \tag{18}$$

The optimization for $\mathbf{W}_3^{(m)}$ equals the following,

$$\begin{aligned} \mathbf{W}_3^{(m)} = & (\mathbf{G}_3^{(m)} \mathbf{L}^\top {\mathbf{W}_1^{(m)}}^\top {\mathbf{W}_2^{(m)}}^\top) \\ & \times (\mathbf{W}_2^{(m)} \mathbf{W}_1^{(m)} \mathbf{E} {\mathbf{W}_1^{(m)}}^\top {\mathbf{W}_2^{(m)}}^\top + \gamma \mathbf{I}_{d_m})^{-1}. \end{aligned} \tag{19}$$

Update $\mathbf{G}_1^{(1)}$ Fixing corresponding variables, the optimization for $\mathbf{G}_1^{(1)}$ can be transformed into the following,

$$\min_{\mathbf{G}_1^{(1)}} ||\mathbf{W}_1^{(1)} \mathbf{L} - \mathbf{G}_1^{(1)}||^2 + ||{\mathbf{G}_1^{(1)}}^\top \mathbf{G}_1^{(2)} - k\mathbf{S}||^2. \tag{20}$$

The optimization for $\mathbf{G}_1^{(1)}$ equals the following,

$$\mathbf{G}_1^{(1)} = (\mathbf{G}_1^{(2)} {\mathbf{G}_1^{(2)}}^\top + \mathbf{I}_k)^{-1} (\mathbf{W}_1^{(1)} \mathbf{L} + k\mathbf{G}_1^{(2)} \mathbf{S}^\top). \tag{21}$$

Update $\mathbf{G}_1^{(2)}$ Fixing corresponding variables, the optimization for $\mathbf{G}_1^{(2)}$ can be transformed into the following,

$$\min_{\mathbf{G}_1^{(2)}} ||\mathbf{W}_1^{(2)}\mathbf{L} - \mathbf{G}_1^{(2)}||^2 + ||{\mathbf{G}_1^{(1)}}^\top \mathbf{G}_1^{(2)} - k\mathbf{S}||^2. \tag{22}$$

The optimization for $\mathbf{G}_1^{(2)}$ equals the following,

$$\mathbf{G}_1^{(2)} = (\mathbf{G}_1^{(1)}{\mathbf{G}_1^{(1)}}^\top + \mathbf{I}_k)^{-1}(\mathbf{W}_1^{(2)}\mathbf{L} + k\mathbf{G}_1^{(1)}\mathbf{S}). \tag{23}$$

Update $\mathbf{G}_3^{(1)}$ Correspondingly, the optimization for $\mathbf{G}_3^{(1)}$ can be transformed into the following,

$$\begin{aligned}\min_{\mathbf{G}_3^{(1)}} &||\mathbf{W}_3^{(1)}\mathbf{W}_2^{(1)}\mathbf{W}_1^{(1)}\mathbf{L} - \mathbf{G}_3^{(1)}||^2 + \eta||{\mathbf{G}_3^{(1)}}^\top \mathbf{G}_3^{(2)} - k\mathbf{S}||^2 \\ &+ \lambda||\mathbf{B} - \mathbf{G}_3^{(1)}||^2.\end{aligned} \tag{24}$$

The optimization for $\mathbf{G}_3^{(1)}$ equals the following,

$$\begin{aligned}\mathbf{G}_3^{(1)} =&(\eta\mathbf{G}_3^{(2)}{\mathbf{G}_3^{(2)}}^\top + (1+\lambda)\mathbf{I}_k)^{-1} \\ &\times(\mathbf{W}_3^{(1)}\mathbf{W}_2^{(1)}\mathbf{W}_1^{(1)}\mathbf{L} + k\eta\mathbf{G}_3^{(2)}\mathbf{S}^\top + \lambda\mathbf{B}).\end{aligned} \tag{25}$$

Update $\mathbf{G}_3^{(2)}$ Correspondingly, the optimization for $\mathbf{G}_3^{(2)}$ can be transformed into the following,

$$\begin{aligned}\min_{\mathbf{G}_3^{(2)}} &||\mathbf{W}_3^{(2)}\mathbf{W}_2^{(2)}\mathbf{W}_1^{(2)}\mathbf{L} - \mathbf{G}_3^{(2)}||^2 + \eta||{\mathbf{G}_3^{(1)}}^\top \mathbf{G}_3^{(2)} - k\mathbf{S}||^2 \\ &+ \omega||\mathbf{B} - \mathbf{G}_3^{(2)}||^2.\end{aligned} \tag{26}$$

The optimization for $\mathbf{G}_3^{(2)}$ equals the following,

$$\begin{aligned}\mathbf{G}_3^{(2)} =&(\eta\mathbf{G}_3^{(1)}{\mathbf{G}_3^{(1)}}^\top + (1+\omega)\mathbf{I}_k)^{-1} \\ &\times(\mathbf{W}_3^{(2)}\mathbf{W}_2^{(2)}\mathbf{W}_1^{(2)}\mathbf{L} + k\eta\mathbf{G}_3^{(1)}\mathbf{S} + \omega\mathbf{B}).\end{aligned} \tag{27}$$

Update $\mathbf{B}$ When $\mathbf{W}_i^{(m)}|_{i=1,2,3}^{m=1,2}$ and $\mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}$ are fixed, the optimization for $\mathbf{B}$ can be transformed into the following,

$$\min_{\mathbf{B}} \lambda||\mathbf{B} - \mathbf{G}_3^{(1)}||^2 + \omega||\mathbf{B} - \mathbf{G}_3^{(2)}||^2 \; s.t. \; \mathbf{B} \in \{-1, 1\}^{k\times n}. \tag{28}$$

The optimization for $\mathbf{B}$ equals the following,

$$\mathbf{B} = \text{sgn}(\lambda\mathbf{G}_3^{(1)} + \omega\mathbf{G}_3^{(2)}). \tag{29}$$

According to Eq. (29), hash codes $\mathbf{B}$ can be generated discretely during optimization, and all of which bits are optimized concurrently.

Out-of-sample Extension

For a new query $\mathbf{x}_q^{(m)}$, **ACHFCA** generates the hash code $\mathbf{b}_q$ as follows,

$$\mathbf{b}_q = \text{sgn}(\mathbf{W}_3^{(m)}\mathbf{x}_q^{(m)}). \tag{30}$$

We summarize the complete procedure of optimization in **Algorithm 1**.

Algorithm 1: The optimization of **ACHFCA**

Input: Training instances $\mathbf{X}^{(m)}$, label matrix $\mathbf{L}$, balance parameters γ, η, λ, ω, maximum iteration number ξ.
Output: Binary codes $\mathbf{B}$.

1: Construct concept lattice $\mathcal{L}(\mathbf{L})$;
2: Calculate $\mathbf{S}$ via Eq. (10);
3: Initialize: $\mathbf{W}_i^{(m)}|_{i=1,2,3}^{m=1,2}$, $\mathbf{G}_i^{(m)}|_{i=1,3}^{m=1,2}$, $\mathbf{B}$ with standard normal distribution;
4: $\mathbf{E} = \mathbf{L}\mathbf{L}^\top$, $\mathbf{F}^{(m)} = \mathbf{X}^{(m)}\mathbf{L}^\top$.
5: **Repeat**
  Update $\mathbf{W}_1^{(m)}$, $\mathbf{W}_2^{(m)}$, $\mathbf{W}_3^{(m)}$ via Eq. (15), (17), (19), respectively;
  Update $\mathbf{G}_1^{(1)}$, $\mathbf{G}_1^{(2)}$ via Eq. (21), (23), respectively;
  Update $\mathbf{G}_3^{(1)}$, $\mathbf{G}_3^{(2)}$ via Eq. (25), (27), respectively;
  Update $\mathbf{B}$ via Eq. (29);
6: **Until** up to ξ.
7: **Return** Hash function $\text{sgn}(\mathbf{W}_3^{(m)}\mathbf{X}^{(m)})$.

Dataset	MIRFlickr	NUS-WIDE	IAPR-TC12
Total	20,015	186,577	20,000
Tags	24	10	255
Training/Test	18,015/2,000	184,710/1,867	18,000/2,000
Image Feature	512-D GIST	500-D SIFT	512-D GIST
Text Feature	1386-D BoW	1000-D BoW	2912-D BoW

Table 2: Statistics information of the three datasets.

Experiments

Datasets

In this paper, we use MIRFlickr, NUS-WIDE (Chua et al. 2009) and IAPR-TC12 (Escalante et al. 2010) datasets for evaluation. For MIRFlickr dataset, we randomly choose 18,015 image-text pairs as the training set while the residual as the test set. For NUS-WIDE dataset, we select 10 typical tags and randomly choose 1,867 image-text pairs as the query set. For IAPR-TC12 dataset, we randomly divide the image-text pairs into 18,000/2,000 training/test sets. The statistics information of the three datasets are listed in Table 2.

Compared Baselines and Evaluation Metrics

In the experiments, supervised methods BATCH (Wang et al. 2021b), SRLCH (Shen et al. 2021), ASCSH (Meng et al. 2021), ALECH (Li et al. 2023a), HCCH (Sun et al. 2024c), online method ROHLSE (Li et al. 2024) are compared with **ACHFCA** in retrieving textual data by visual

Dataset	γ	η	λ	ω
MIRFlickr	10^0	10^{-5}	10^{-5}	10^{-4}
NUS-WIDE	10^2	10^1	10^2	10^2
IAPR-TC12	10^2	10^1	10^{-1}	10^4

Table 3: Best parameter configurations on the three datasets.

Task	Method	MIRFlickr				NUS-WIDE				IAPR-TC12			
		8 bits	16 bits	32 bits	64 bits	8 bits	16 bits	32 bits	64 bits	8 bits	16 bits	32 bits	64 bits
I → T	BATCH	0.7245	0.7326	0.7467	0.7478	0.6089	0.6283	0.6514	0.6578	0.4510	0.4738	0.5079	0.5256
	SRLCH	0.6051	0.6368	0.6506	0.6841	0.5777	0.5916	0.6445	0.6471	0.3352	0.3660	0.3769	0.4038
	ASCSH	0.7356	0.7508	0.7627	0.7711	0.6101	0.6413	0.6656	0.6703	0.4684	0.5036	0.5196	0.5333
	ALECH	0.7287	0.7418	0.7477	0.7463	0.6103	0.6386	0.6535	0.6682	0.4531	0.4781	0.5097	0.5200
	ROHLSE	0.7033	0.7214	0.7347	0.7365	0.5988	0.6197	0.6342	0.6383	0.4464	0.4689	0.4853	0.5090
	HCCH	0.7337	0.7461	0.7551	0.7650	0.6204	0.6409	0.6567	0.6701	0.4848	0.5110	0.5223	0.5345
	ACHFCA	**0.7497**	**0.7605**	**0.7774**	**0.7831**	**0.6250**	**0.6492**	**0.6753**	**0.6832**	**0.4932**	**0.5176**	**0.5336**	**0.5458**
T → I	BATCH	0.7909	0.8106	0.8204	0.8289	0.7219	0.7500	0.7723	0.7738	0.5157	0.5574	0.5945	0.6111
	SRLCH	0.6342	0.6726	0.6933	0.7522	0.6811	0.7042	0.7580	0.7605	0.3567	0.3987	0.4232	0.4794
	ASCSH	0.7802	0.8133	0.8288	0.8323	0.7201	0.7565	0.7850	0.7929	0.4819	0.5367	0.5906	0.6342
	ALECH	0.7859	0.8087	0.8172	0.8206	0.7242	0.7531	0.7797	0.7816	0.5156	0.5618	0.6074	0.6313
	ROHLSE	0.7764	0.7914	0.8143	0.8152	0.7106	0.7209	0.7321	0.7424	0.4969	0.5418	0.5898	0.6154
	HCCH	0.7836	0.8069	0.8210	0.8259	0.7220	0.7536	0.7796	0.7819	0.5221	0.5637	0.6159	0.6345
	ACHFCA	**0.7991**	**0.8225**	**0.8395**	**0.8458**	**0.7393**	**0.7618**	**0.7951**	**0.7986**	**0.5343**	**0.5715**	**0.6289**	**0.6457**

Table 4: Performance comparison (mAP) of **ACHFCA** and baselines on the three datasets with various code lengths.

query (I2T) and retrieving visual data by text query (T2I).

In order to assess the retrieval performance of **ACHFCA**, *mean average precision* (**mAP**) and *precision-recall* (**PR**) curves act as evaluation metrics.

Implementation Details

For the proposed **ACHFCA**, parameters $\gamma, \eta, \lambda, \omega$ are selected by utilizing grid search (from 10^{-5} to 10^{4}, 10 times per step). We provide the best performance of parameter configurations in Table 3. In addition, ρ is set to 0.5. For the baselines, we implement them by ourselves or their open source codes. All experiments are trialed on a server with Intel Xeon Silver 4210 Processor @2.20 GHz, 128G RAM.

Results

Table 4 manifests the mAP scores of each method on the three datasets, and hash code lengths are set to $\{8, 16, 32, 64\}$. The PR curves are demonstrated in Fig. 3. From Table 4 and Fig. 3, we can obtain the following key findings.

- The proposed **ACHFCA** surpassed all the baselines. Particularly, for the I2T task, contrasted with the best baselines, the average mAP scores of **ACHFCA** are beyond by 1.68%, 1.77% and 1.83% on MIRFlickr, NUS-WIDE and IAPR-TC12 datasets, respectively. For the T2I task, the average mAP scores of **ACHFCA** are higher by 1.62%, 1.35% and 1.89%, respectively.
- The mAP scores of mainstream approaches increase with the increase in code lengths. The PR curves depict a similar trend as mAP scores with the increase in retrieved instances.
- Although ASCSH and HCCH are efficient discrete methods, **ACHFCA** achieves higher mAP scores. The improvement in mAP scores owing to employing the double asymmetric hash learning framework based on FCA, which adequately leverages the abundant semantic information into binary codes.

Ablation Experiments

Impacts of kernelization To verify the impacts of kernelization, we utilize the RBF kernel to project heterogeneous modalities data into a nonlinear space (Yao et al. 2020), termed **ACHFCA**-$\mathcal{K}$. For each instance of m-th modality $\mathbf{x}_i^{(m)}$, the kernelized feature $\phi(\mathbf{x}_i^{(m)})$ can be formulated as,

$$\left[\exp\left(-\frac{||\mathbf{x}_i^{(m)} - \mathbf{a}_1^{(m)}||^2}{2\sigma_m^2}\right), ..., \exp\left(-\frac{||\mathbf{x}_i^{(m)} - \mathbf{a}_q^{(m)}||^2}{2\sigma_m^2}\right)\right], \tag{31}$$

where $\sigma_m = \frac{1}{qn}\sum_{i=1}^{n}\sum_{j=1}^{q}||\mathbf{x}_i^{(m)} - \mathbf{a}_j^{(m)}||$ is the kernel width and $\{\mathbf{a}_1^{(m)}, \mathbf{a}_2^{(m)}, ..., \mathbf{a}_q^{(m)}\}$ are q anchor instances.

Specifically, q is set to 2000 in the experiment. From Table 5, it can be found that the kernelized feature hash codes result in a great loss in performance, especially in T2I tasks. The reason is that kernelization causes individual modality-specific representation loss and overfitting, thereby some characteristic noises come to the feature alignment layer of the semantic enhancement descriptor.

Impacts of classical similarity reconstruction strategy To testify the impacts of classical similarity reconstruction strategy, we present a variant of **ACHFCA**, termed **ACHFCA**-$\mathcal{C}$. Particularly, after normalizing each row of label matrix $\mathbf{L}$ with 2-norm and obtaining $\tilde{\mathbf{L}}$, $\mathbf{S}$ can be generated as,

$$\mathbf{S} = 2\tilde{\mathbf{L}}^{\top}\tilde{\mathbf{L}} - \mathbf{1}\mathbf{1}^{\top}. \tag{32}$$

From Table 5, it can be found that the variant cannot obtain superior performance. The reason is that the classical similarity reconstruction strategy failed to make the most of latent semantic information into binary codes.

Impacts of deep image features Deep learning-based hash models have made exciting achievements in the field of information retrieval. In order to verify the generalization of our proposed **ACHFCA**, we compare **ACHFCA** with DCHMT (Tu et al. 2022) and DSPH (Huo et al. 2024). To ensure the fairness of experimental results between pseudo-deep and end-to-end deep methods, we utilize the 4096-D

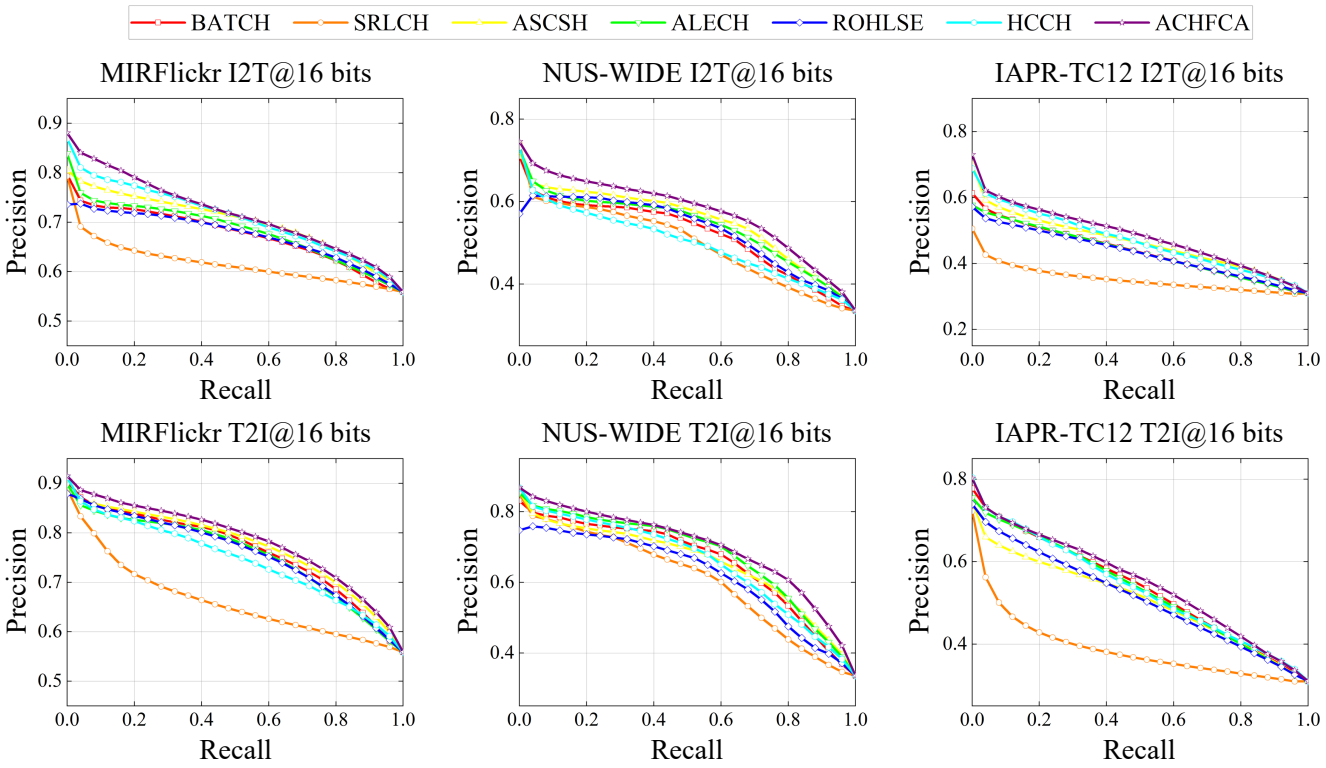

Figure 3: PR curves of **ACHFCA** and baselines on the three datasets.

Task	Method	MIRFlickr		NUS-WIDE	
		16 bits	32 bits	16 bits	32 bits
I → T	**ACHFCA**-$\mathcal{K}$	0.7278	0.7329	0.6290	0.6544
	ACHFCA-$\mathcal{C}$	0.7418	0.7531	0.6297	0.6506
	ACHFCA	0.7605	0.7774	0.6492	0.6753
T → I	**ACHFCA**-$\mathcal{K}$	0.7324	0.7424	0.6647	0.6961
	ACHFCA-$\mathcal{C}$	0.7937	0.8027	0.7388	0.7612
	ACHFCA	0.8225	0.8395	0.7618	0.7951

Table 5: The mAP results of **ACHFCA** and its variants on MIRFlickr and NUS-WIDE datasets.

Task	Method	MIRFlickr		NUS-WIDE	
		16 bits	32 bits	16 bits	32 bits
I → T	DCHMT	0.8201	0.8253	0.6596	0.6706
	DSPH	0.8129	0.8482	0.6830	0.6979
	ACHFCA-$\mathcal{D}$	0.7761	0.7840	0.6562	0.6827
T → I	DCHMT	0.7983	0.8048	0.6761	0.6837
	DSPH	0.8000	0.8238	0.6997	0.7153
	ACHFCA-$\mathcal{D}$	0.8481	0.8579	0.7822	0.8097

Table 6: The mAP results of **ACHFCA**-$\mathcal{D}$ and baselines on MIRFlickr and NUS-WIDE datasets.

image features extracted from a 19-layer VGG network (Simonyan and Zisserman 2015) on ImageNet dataset (Deng et al. 2009) with BoW textual features, termed **ACHFCA**-$\mathcal{D}$. For training, we randomly select a subset of 2,000 image-text pairs, designating the remainder as the test set.

The mAP results are presented in Table 6, it is observed that the mAP scores realized by **ACHFCA**-$\mathcal{D}$ are lower than the compared end-to-end deep baselines in the I2T tasks, indicating that the increased dimensionality of deep image representations leads to increased difficulty of the projection from labels to semantic space, which affects the margin of the performance growth in the I2T tasks. Conversely, the performance of the T2I tasks is higher than that of the compared baselines, manifesting that our proposed similarity matrix construction and discrete optimization strategies are important for hash code generation.

Parameter Sensitivity Analysis

To ascertain the influence of diverse parameters on the three datasets, we conducted the grid search for γ, η, λ, ω preserving the remaining parameters constant, with the hash code length at 16 bits. From Fig. 4, it can be observed that the performance is stable when γ is set to $\{10^0, 10^4\}$, $\{10^1, 10^4\}$, $\{10^1, 10^4\}$, η is in the range of $\{10^{-5}, 10^0\}$, $\{10^{-2}, 10^2\}$, $\{10^{-5}, 10^{-2}\}$, λ is set to $\{10^{-5}, 10^0\}$, $\{10^{-5}, 10^2\}$, $\{10^{-1}, 10^4\}$, ω is in the range of $\{10^{-5}, 10^1\}$, $\{10^{-3}, 10^2\}$, $\{10^{-1}, 10^4\}$ on MIRFlickr, NUS-WIDE and IAPR-TC12 datasets, respectively.

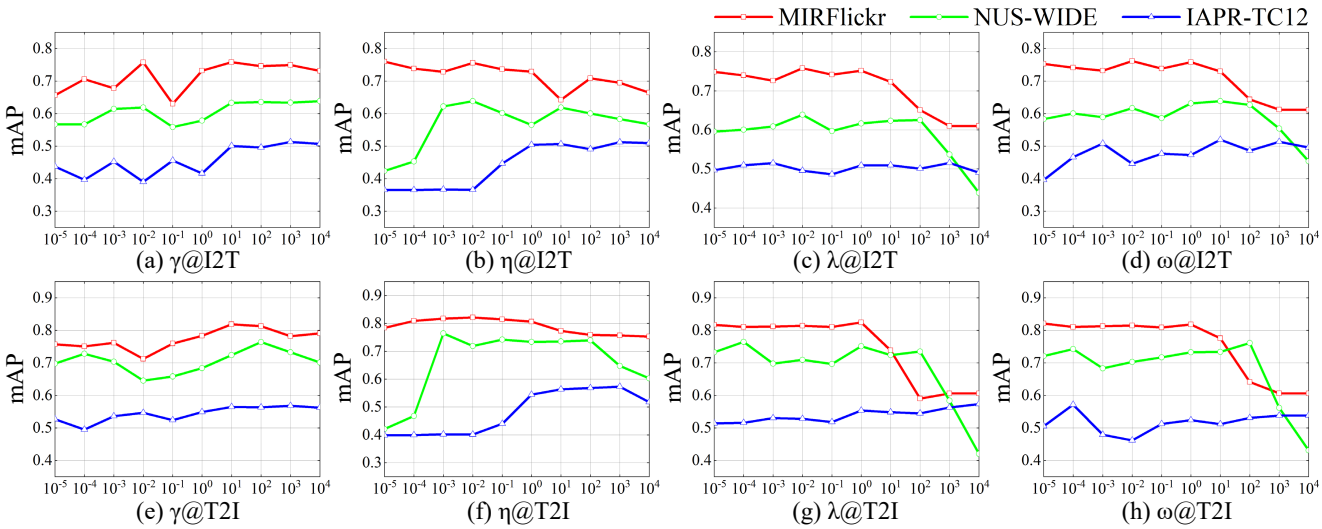

Figure 4: Parameter sensitivity for hyper-parameters.

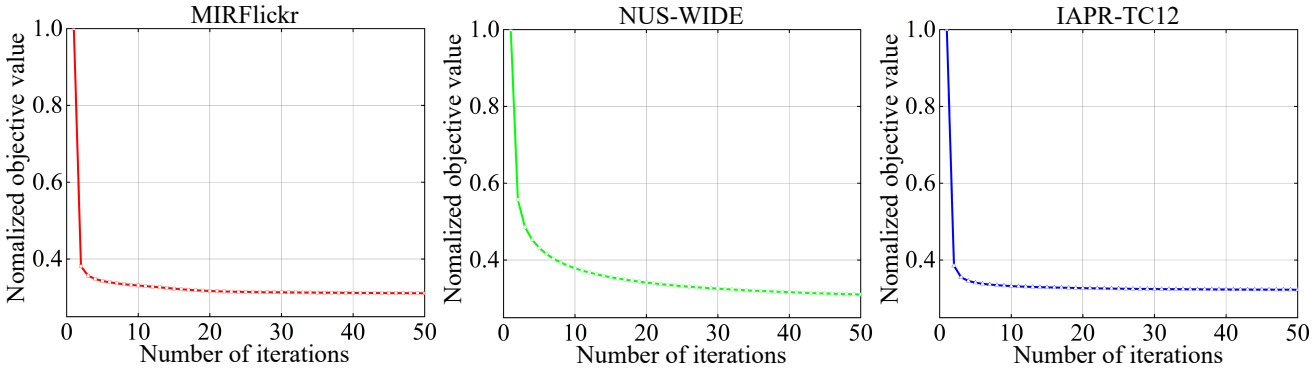

Figure 5: Convergence curves of **ACHFCA** on the three datasets.

For the most part, from Fig. 4, it can be found that **ACHFCA** obtains superior retrieval accuracy with parameters in the middle (*i.e.*, around 10^0). Consequently, the parameters exhibit insensitivity to cross-modal tasks across a wide value range, thereby highlighting the robustness of **ACHFCA**.

Convergence Analysis

The convergence analysis experiments were conducted on the three datasets with 16 bits. The convergence values of Eq. (13) are plotted in Fig. 5, it can be observed that the objective function can converge quickly within 20, 50 and 10 iterations on MIRFlickr, NUS-WIDE, and IAPR-TC12 datasets, respectively, which verifies the efficiency of the presented optimization algorithm.

Conclusion

In this paper, we propose a novel asymmetric supervised approach based on FCA for cross-modal hashing, which adopts a flash-projection three-layer semantic enhancement descriptor to extract latent representations from heterogeneous modalities. To improve the discriminative capacity of the proposed model and generate compact hash codes, we construct a humanized fine-grained similarity matrix based on FCA to enhance the semantics of the enhancement descriptor in the asymmetric hash learning framework. Comprehensive experiments on the three datasets demonstrate the superior accuracy of **ACHFCA** to state-of-the-art hashing approaches. In future work, since FCA plays a meaningful role in fine-grained relation reconstruction, it would be interesting to introduce FCA and three-way concept analysis into self-supervised cross-modal hashing.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant No. 62202501, in part by the Science and Technology Plan of Hunan Province under Grant No. 2023GK2013, and in part by the National Key R&D Program of China under Grant No. 2021YFB3900902.

References

Akram, M.; Nawaz, H. S.; and Deveci, M. 2023. Attribute reduction and information granulation in Pythagorean fuzzy formal contexts. *Expert Syst. Appl.*, 222: 119794.

Bin, Y.; Li, H.; Xu, Y.; Xu, X.; Yang, Y.; and Shen, H. T. 2023. Unifying Two-Stream Encoders with Transformers for Cross-Modal Retrieval. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3041–3050. ACM.

Chen, Z.; Li, C.; Luo, X.; Nie, L.; Zhang, W.; and Xu, X. 2020. SCRATCH: A Scalable Discrete Matrix Factorization Hashing Framework for Cross-Modal Retrieval. *IEEE Trans. Circuits Syst. Video Technol.*, 30(7): 2262–2275.

Chua, T.; Tang, J.; Hong, R.; Li, H.; Luo, Z.; and Zheng, Y. 2009. NUS-WIDE: a real-world web image database from National University of Singapore. In *Proceedings of the 8th ACM International Conference on Image and Video Retrieval*. ACM.

Da, C.; Xu, S.; Ding, K.; Meng, G.; Xiang, S.; and Pan, C. 2017. AMVH: Asymmetric Multi-Valued hashing. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 898–906. IEEE Computer Society.

Deng, J.; Dong, W.; Socher, R.; Li, L.; Li, K.; and Fei-Fei, L. 2009. ImageNet: A large-scale hierarchical image database. In *2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2009)*, 248–255. IEEE Computer Society.

Escalante, H. J.; Hernández, C. A.; González, J. A.; López-López, A.; Montes-y-Gómez, M.; Morales, E. F.; Sucar, L. E.; Pineda, L. V.; and Grubinger, M. 2010. The segmented and annotated IAPR TC-12 benchmark. *Comput. Vis. Image Underst.*, 114(4): 419–428.

GABOR, D. 1948. A New Microscopic Principle. *Nature*, 161: 777–778.

Hu, Q.; Yuan, Z.; Qin, K.; and Zhang, J. 2023. A novel outlier detection approach based on formal concept analysis. *Knowl. Based Syst.*, 268: 110486.

Huiskes, M. J.; and Lew, M. S. 2008. The MIR flickr retrieval evaluation. In *Proceedings of the 1st ACM SIGMM International Conference on Multimedia Information Retrieval*, 39–43. ACM.

Huo, Y.; Qin, Q.; Dai, J.; Wang, L.; Zhang, W.; Huang, L.; and Wang, C. 2024. Deep Semantic-Aware Proxy Hashing for Multi-Label Cross-Modal Retrieval. *IEEE Trans. Circuits Syst. Video Technol.*, 34(1): 576–589.

Li, H.; Zhang, C.; Jia, X.; Gao, Y.; and Chen, C. 2023a. Adaptive Label Correlation Based Asymmetric Discrete Hashing for Cross-Modal Retrieval. *IEEE Trans. Knowl. Data Eng.*, 35(2): 1185–1199.

Li, J.; Li, F.; Zhu, L.; Cui, H.; and Li, J. 2023b. Prototype-guided Knowledge Transfer for Federated Unsupervised Cross-modal Hashing. In *Proceedings of the 31st ACM International Conference on Multimedia*, 1013–1022. ACM.

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

Li, L.; Zheng, B.; and Sun, W. 2022. Adaptive Structural Similarity Preserving for Unsupervised Cross Modal Hashing. In *The 30th ACM International Conference on Multimedia*, 3712–3721. ACM.

Lin, Z.; Liu, R.; and Su, Z. 2011. Linearized Alternating Direction Method with Adaptive Penalty for Low-Rank Representation. In *Advances in Neural Information Processing Systems 24: 25th Annual Conference on Neural Information Processing Systems 2011*, 612–620.

Liu, H.; Ji, R.; Wu, Y.; and Hua, G. 2016. Supervised Matrix Factorization for Cross-Modality Hashing. In Kambhampati, S., ed., *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence*, 1767–1773. IJCAI/AAAI Press.

Liu, H.; Ji, R.; Wu, Y.; Huang, F.; and Zhang, B. 2017. Cross-Modality Binary Code Learning via Fusion Similarity Hashing. In *2017 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017, Honolulu, HI, USA, July 21-26, 2017*, 6345–6353. IEEE Computer Society.

Liu, S.; Qian, S.; Guan, Y.; Zhan, J.; and Ying, L. 2020. 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*, 1379–1388. ACM.

Liu, W.; Wang, J.; Ji, R.; Jiang, Y.; and Chang, S. 2012. Supervised hashing with kernels. In *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2074–2081. IEEE Computer Society.

Liu, Y.; Wu, Q.; Zhang, Z.; Zhang, J.; and Lu, G. 2023. Multi-Granularity Interactive Transformer Hashing for Cross-modal Retrieval. In *Proceedings of the 31st ACM International Conference on Multimedia*, 893–902. ACM.

Luo, K.; Zhang, C.; Li, H.; Jia, X.; and Chen, C. 2023. Adaptive Marginalized Semantic Hashing for Unpaired Cross-Modal Retrieval. *IEEE Trans. Multim.*, 25: 9082–9095.

Meng, M.; Wang, H.; Yu, J.; Chen, H.; and Wu, J. 2021. Asymmetric Supervised Consistent and Specific Hashing for Cross-Modal Retrieval. *IEEE Trans. Image Process.*, 30: 986–1000.

Qi, J.; Qian, T.; and Wei, L. 2016. The connections between three-way and classical concept lattices. *Knowledge-Based Systems*, 91: 143–151. Three-way Decisions and Granular Computing.

Shen, H. T.; Liu, L.; Yang, Y.; Xu, X.; Huang, Z.; Shen, F.; and Hong, R. 2021. Exploiting Subspace Relation in Semantic Labels for Cross-Modal Hashing. *IEEE Trans. Knowl. Data Eng.*, 33(10): 3351–3365.

Simonyan, K.; and Zisserman, A. 2015. Very Deep Convolutional Networks for Large-Scale Image Recognition. In *3rd International Conference on Learning Representations*.

Song, X.; Chen, J.; and Jiang, Y. 2023. Relation Triplet Construction for Cross-modal Text-to-Video Retrieval. In *Proceedings of the 31st ACM International Conference on Multimedia*, 4759–4767. ACM.

Sun, Y.; Dai, J.; Ren, Z.; Chen, Y.; Peng, D.; and Hu, P. 2024a. Dual Self-Paced Cross-Modal Hashing. In *Thirty-Eighth AAAI Conference on Artificial Intelligence*, 15184–15192. AAAI Press.

Sun, Y.; Ren, Z.; Hu, P.; Peng, D.; and Wang, X. 2024b. Hierarchical Consensus Hashing for Cross-Modal Retrieval. *IEEE Trans. Multim.*, 26: 824–836.

Sun, Y.; Ren, Z.; Hu, P.; Peng, D.; and Wang, X. 2024c. Hierarchical Consensus Hashing for Cross-Modal Retrieval. *IEEE Trans. Multim.*, 26: 824–836.

Tu, J.; Liu, X.; Lin, Z.; Hong, R.; and Wang, M. 2022. Differentiable Cross-modal Hashing via Multimodal Transformers. In Magalhães, J.; Bimbo, A. D.; Satoh, S.; Sebe, N.; Alameda-Pineda, X.; Jin, Q.; Oria, V.; and Toni, L., eds., *MM '22: The 30th ACM International Conference on Multimedia, Lisboa, Portugal, October 10 - 14, 2022*, 453–461. ACM.

Tu, R.; Mao, X.; Ji, W.; Wei, W.; and Huang, H. 2023. Data-Aware Proxy Hashing for Cross-modal Retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 686–696. ACM.

Wang, D.; Wang, Q.; He, L.; Gao, X.; and Tian, Y. 2020. Joint and individual matrix factorization hashing for large-scale cross-modal retrieval. *Pattern Recognit.*, 107: 107479.

Wang, L.; Yang, J.; Zareapoor, M.; and Zheng, Z. 2021a. Cluster-wise unsupervised hashing for cross-modal similarity search. *Pattern Recognit.*, 111: 107732.

Wang, Y.; Luo, X.; Nie, L.; Song, J.; Zhang, W.; and Xu, X. 2021b. BATCH: A Scalable Asymmetric Discrete Cross-Modal Hashing. *IEEE Trans. Knowl. Data Eng.*, 33(11): 3507–3519.

Wille, R. 1982. Restructuring Lattice Theory: An Approach Based on Hierarchies of Concepts. In *Ordered Sets*, 445–470. Dordrecht: Springer Netherlands.

Yang, Z.; Deng, X.; Guo, L.; and Long, J. 2024. Asymmetric Supervised Fusion-Oriented Hashing for Cross-Modal Retrieval. *IEEE Trans. Cybern.*, 54(2): 851–864.

Yao, T.; Yan, L.; Ma, Y.; Yu, H.; Su, Q.; Wang, G.; and Tian, Q. 2020. Fast discrete cross-modal hashing with semantic consistency. *Neural Networks*, 125: 142–152.

Zhao, S.; Xu, L.; Liu, Y.; and Du, S. 2023. Multi-grained Representation Learning for Cross-modal Retrieval. In *Proceedings of the 46th International ACM SIGIR Conference on Research and Development in Information Retrieval*, 2194–2198. ACM.

Zhi, H.; and Li, Y. 2023. Attribute granulation in fuzzy formal contexts based on L-fuzzy concepts. *Int. J. Approx. Reason.*, 159: 108947.

Zhong, F.; Chu, C.; Zhu, Z.; and Chen, Z. 2023. Hypergraph-Enhanced Hashing for Unsupervised Cross-Modal Retrieval via Robust Similarity Guidance. In *Proceedings of the 31st ACM International Conference on Multimedia, MM 2023*, 3517–3527. ACM.