Discretization is Not Always Better: Rethinking Deep Quantization for Asymmetric Image Retrieval

Xinze Liu[1, 2], **Dayan Wu**[1*], **Hengjie Zhu**[1, 2], **Chenming Wu**[3], **Pengwen Dai**[4]
[1]Institute of Information Engineering, Chinese Academy of Sciences
[2]School of Cyber Security, University of Chinese Academy of Sciences
[3]Baidu Research
[4]School of Cyber Science and Technology, Shenzhen Campus of Sun Yat-sen University
{liuxinze, wudayan, zhuhengjie}@iie.ac.cn, wuchenming@baidu.com, daipw@mail.sysu.edu.cn

Abstract

Asymmetric image retrieval (AIR), which typically employs a compact model for the query side and a large model for the database server, has garnered significant attention in resource-constrained environments. While deep hashing methods have shown great potential in large-scale image retrieval, current attempts for the asymmetric image retrieval overlook the differences in quantization capabilities between query and gallery networks. In AIR, the conventional quantization scheme forces the outputs of small query models to approximate the discrete outputs of large models, imposing overly rigid and stringent constraints that severely limit the optimization of small query models. Furthermore, existing deep hashing methods for AIR necessitate labeled datasets from large models, which also limits their practical applicability. To this end, we reconsider the necessity of strict discretization in AIR and propose a novel asymmetric hashing method, named **D**eep **C**orrelation **A**lignment **H**ashing (DCAH). Rather than explicitly quantizing continuous query features to match discrete gallery representations, we distill the correlation across both models and introduce a **C**orrelation **A**lignment based **Q**uantization (CAQ) scheme, thereby implicitly accomplishing quantization. To preserve the similarity consistency between the query and gallery models, we further employ a correlation alignment-based knowledge distillation strategy which is intrinsically compatible with the CAQ. Notably, the proposed quantization scheme can function as a plug-and-play module that seamlessly integrates with existing AIR methods. Comprehensive evaluations on three real-world benchmark datasets demonstrate the effectiveness of the proposed quantization scheme CAQ, and also show that DCAH achieves state-of-the-art performance in asymmetric image retrieval scenarios.

Code — https://github.com/Xinze919/DCAH

Introduction

Deep hashing methods have been widely used in large-scale image retrieval due to its high efficiency. Most deep hashing methods (Cao et al. 2018; Li, Wang, and Kang 2016; Hoe et al. 2021; Zhu et al. 2016; Liu et al. 2016) typically employ the same large model to symmetrically encode both query and gallery images, named as *symmetric image retrieval* (Budnik and Avrithis 2021).

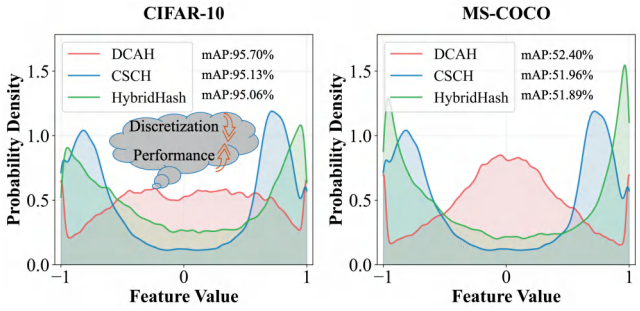

Figure 1: Feature distribution visualization on CIFAR-10 and MS-COCO with 32 bits. Despite milder discretization, our method achieves superior asymmetric image retrieval performance.

However, in resource-constrained scenarios, it is inefficient to deploy such a large model on query side. Consequently, *asymmetric image retrieval* (Wu et al. 2022b; Budnik and Avrithis 2021), which handles gallery images with a large model while employing a smaller model for the query images, is proposed to strike a balance between retrieval accuracy and efficiency. Several deep hashing methods (Jiang and Li 2018; Shen et al. 2017) have been proposed to asymmetrically learn compatible hashing models. DAPH (Shen et al. 2017) is the first deep hashing method that introduces two different models for query and gallery images. Specifically, DAPH jointly trains two models to learn pairwise similarity preserving codes in an alternative manner. However, empirical results show that such mutual training scheme will inevitably degrade the performance of large gallery model. CSCH (Xuan et al. 2024) is the first deep hashing method, which pre-trains a large gallery model by center-based deep hashing method and then transfers knowledge to a small query model.

Traditional quantization approaches explicitly enforce feature values to converge to +1 or -1 through strong binarization constraints. However, due to the inherent disparity in capacity between lightweight query models and powerful gallery models, achieving effective binarization while preserving semantic similarity remains a significant chal-

*Corresponding author.

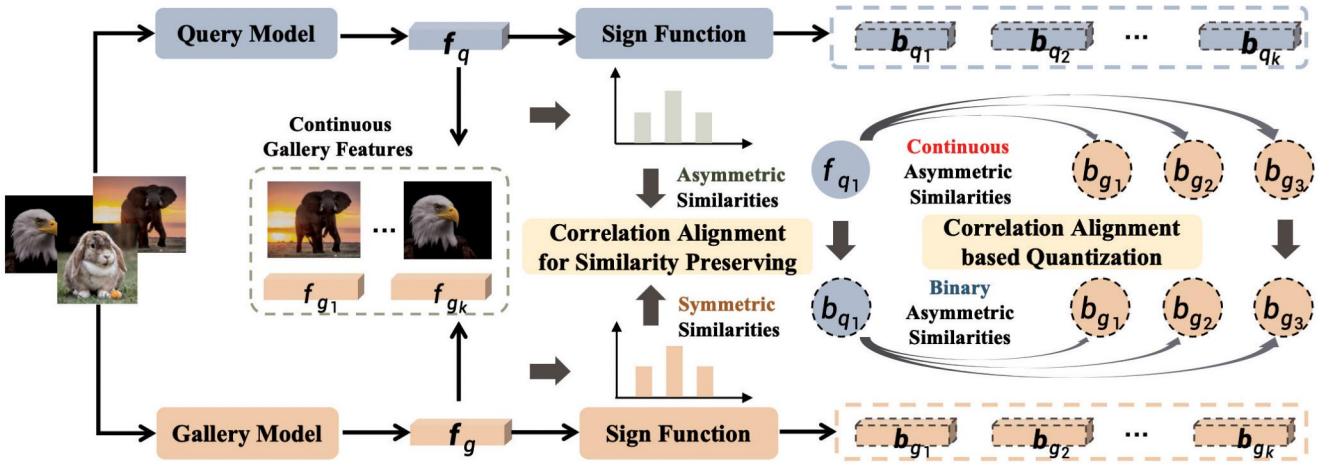

Figure 2: Framework of the proposed DCAH. The objective of correlation alignment for similarity preserving is to align asymmetric similarities with symmetric similarities. Correlation alignment based quantization explicitly minimize the gap between continuous and binary query features.

lenge for the query network. As shown in Figure 1, although our method employs a less aggressive discretization strategy, it achieves superior performance in asymmetric image retrieval, suggesting that excessive binarization may harm feature representation. This observation highlights a critical insight: in asymmetric image retrieval, we must re-examine the fundamental role of quantization and reconsider the necessity of strict binary constraint in hash code learning.

In this paper, we propose Deep Correlation Alignment Hashing (DCAH), an innovative quantization framework tailored for Asymmetric Image Retrieval (AIR). Instead of enforcing direct alignment between continuous query features and discrete gallery representations, we rethink the essence of quantization by focusing on minimizing quantization errors implicitly. Our key insight is that aligning the similarity structures—rather than the features themselves—can effectively reduce distortion. Specifically, we compute two types of asymmetric similarity metrics: (1) between the continuous small query model and the binary large gallery model, and (2) between both binarized models. By aligning these via correlation matching and a correlation distance constraint, our method achieves implicit quantization without explicit binarization. For similarity preservation, we further model two continuous similarity relationships: (1) symmetric similarities within the large model, and (2) asymmetric similarities between the pre-binarization small and large models. Enforcing consistency between them enables effective knowledge distillation. In DCAH, correlation alignment serves as the unified mechanism for both quantization and similarity preservation. The main contributions of this work can be summarized as follows:

- This paper introduces Deep Correlation Alignment Hashing (DCAH), a novel quantization framework designed specifically for AIR. DCAH is, to our knowledge, the first deep asymmetric hashing approach to challenge and relax explicit binarization, replacing rigid feature-level quantization with a correlation-driven learning strategy.
- We propose a novel Correlation Alignment-based Quantization (CAQ) loss to optimize the query model for implicit quantization. The CAQ loss is both label-free and plug-and-play, making it easily integrable with existing AIR methods.
- Comprehensive experiments have demonstrated that our method not only attains state-of-the-art performance but also validates the effectiveness of the proposed quantization scheme based on correlation alignment.

The structure of the remainder of this paper is as follows. Section *Related Work* delves into the works related to our research. Section *The Proposed method* outlines the problem statement and presents our DCAH. Experimental results are detailed in Section *Experiments*, followed by a conclusion of our work in Section *Conclusion*.

Related Work

Asymmetric Image Retrieval. Recently, asymmetric image retrieval, which leverages a large model to encode the gallery images and a small model for query ones, shows superior performance when the computing resources are limited. AML(Budnik and Avrithis 2021) proposes an asymmetric metric learning framework that adopts different optimization objectives to encourage small query model to align with large gallery model. CSD(Wu et al. 2022b) further introduces a contextual similarity distillation framework which constrains contextual similarity consistency to keep feature generated by query model compatible with that of gallery model with no labels. ROP(Wu et al. 2023b) proposes a rank-preserving framework using differentiable sorting and monotonic mapping to optimize retrieval order consistency without strict feature alignment. D3still(Xie et al. 2024) introduces a similarity difference matrix that decomposes into consistent/inconsistent knowledge components for distillation, achieving ranking consistency with-

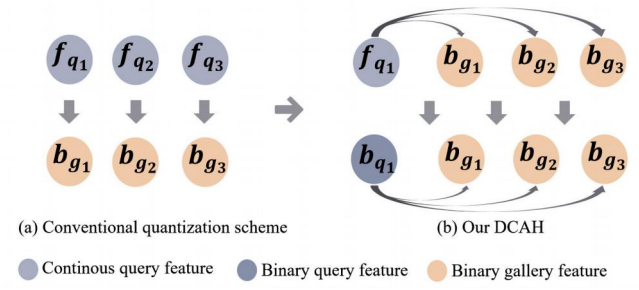

Figure 3: Comparison of quantization scheme in AIR between conventional methods and ours.

out feature alignment. GranDist(Zhang et al. 2025) presents a layered-granularity distillation with multi-scale local semantic transfer and UnamSel mechanism for fine-grained alignment without gallery retraining.

Deep Hashing Methods. Deep hashing methods (Xia et al. 2014; Li, Wang, and Kang 2016; Jiang and Li 2018; Shen et al. 2017; Yuan et al. 2020; Hoe et al. 2021; Gu et al. 2022; Wu et al. 2019; Yang et al. 2020; Sun et al. 2023; Su et al. 2024; Wu et al. 2024, 2023a; Pu et al. 2025b,a; Su, Wu, and Li 2025) have shown prominent performance improvements over non-deep hashing methods with handcrafted features (Weiss, Torralba, and Fergus 2008; Gionis et al. 1999; Gong et al. 2013). Recently, central similarity based symmetric deep hashing methods attract more attentions and present better performance (Fan et al. 2020; Yuan et al. 2020; Hoe et al. 2021; Wang et al. 2023). CSQ (Yuan et al. 2020) first generates hash centers via Hadamard matrix, and then pulls hash codes towards their corresponding centers with binary cross entropy loss. Similarly, OrthoHash (Hoe et al. 2021) proposes to maximize the cosine similarity between the continuous codes and their corresponding hash centers. Besides, many asymmetric deep hashing methods are proposed to learn hash codes for query and gallery images with different models. DAPH (Shen et al. 2017) jointly trains two different models to learn pairwise similarity preserving codes in an alternative manner. The large model is constrained to be aligned with the small one. Different from previous methods, CSCH (Xuan et al. 2024) uses central similarity to align the large and small models via end-to-end training.

Unlike existing asymmetric deep hashing methods that rely on labeled feature-space optimization, our DCAH explicitly addresses the model capability gap through a more relaxed yet highly effective correlation alignment mechanism. Furthermore, the proposed CAQ module serves as a label-free, plug-and-play component that can be seamlessly integrated with various AIR methods.

The Proposed Method

Preliminaries

Assume that we have a training set of N images $\mathbf{X} = \{\mathbf{x}_i\}_{i=1}^{N} \in \mathbb{R}^{N\times D}$ The goal of deep hashing is to learn a hashing function $\mathcal{H}$: $\mathbf{x} \mapsto \mathbf{b} \in \{-1,1\}^B$ from input space $\mathbb{R}^D$ into hamming space $\{0,1\}^B$ via deep neural networks, which $\mathbf{b} = sgn(\mathbf{f})$ is B-bit binary codes transformed from the continuous hash codes $\mathbf{f} \in \mathbb{R}^B$ through a sgn function. Under the asymmetric image retrieval setting, we use a small model $\mathcal{H}_q(\cdot)$ for query side and a large model $\mathcal{H}_g(\cdot)$ for gallery side. Given an image $\mathbf{x}$, the continuous hash codes generated from query / gallery model are denoted as $\mathbf{f}_q = \mathcal{H}_q(\mathbf{x})$ / $\mathbf{f}_g = \mathcal{H}_g(\mathbf{x})$. Therefore, the binary hash codes are denoted as $\mathbf{b}_q = sgn(\mathbf{f}_q) = \mathcal{H}_q^{'}(\mathbf{x})$ / $\mathbf{b}_g = sgn(\mathbf{f}_g) = \mathcal{H}_g^{'}(\mathbf{x})$, where $\mathcal{H}^{'}(\cdot)$ is called deep hash function (Wu et al. 2022a). The performance of a retrieval system is measured by some metrics, such as mean Average Precision (mAP), which can be denoted as $P(\mathcal{H}_q(\cdot), \mathcal{H}_g(\cdot))$. Generally, it is expected that $P(\mathcal{H}_q^{'}(\cdot), \mathcal{H}_g^{'}(\cdot)) > P(\mathcal{H}_q^{'}(\cdot), \mathcal{H}_q^{'}(\cdot))$ and $P(\mathcal{H}_q^{'}(\cdot), \mathcal{H}_g^{'}(\cdot)) \approx P(\mathcal{H}_g^{'}(\cdot), \mathcal{H}_g^{'}(\cdot))$, which allows asymmetric retrieval to strike a balance between performance and efficiency.

Deep Correlation Alignment Hashing Framework

As illustrated in Figure 2, the proposed Deep Correlation Alignment Hashing (DCAH) framework includes two components: *correlation alignment for similarity preserving* and *correlation alignment based quantization*. The *correlation alignment for similarity sreserving* part tries to achieve $P(\mathcal{H}_q(\cdot), \mathcal{H}_g(\cdot)) \approx P(\mathcal{H}_g(\cdot), \mathcal{H}_g(\cdot))$, while the *correlation alignment based quantization* part tries to achieve $P(\mathcal{H}_q(\cdot), \mathcal{H}_g^{'}(\cdot)) \approx P(\mathcal{H}_g^{'}(\cdot), \mathcal{H}_g^{'}(\cdot))$.

During the training of the lightweight query model $\mathcal{H}_q(\cdot)$, the gallery model $\mathcal{H}_g(\cdot)$ is frozen. For the t-th batch in the training set, we first extract the continuous feature $\mathbf{F}_k = [\mathbf{f}_{k_1}^t, \mathbf{f}_{k_2}^t, \mathbf{f}_{k_3}^t, \cdots, \mathbf{f}_{k_T}^t] \in \mathbb{R}^{T\times B}$ and the binary features $\mathbf{B}_k = [\mathbf{b}_{k_1}^t, \mathbf{b}_{k_2}^t, \mathbf{b}_{k_3}^t, \cdots, \mathbf{b}_{k_T}^t] \in \mathbb{R}^{T\times B}$ of images in the batch, where T denotes the batch size, $k \in \{g, q\}$ and $1 \le \mathrm{i} \le \mathrm{T}$, $1 \le \mathrm{j} \le \lceil \frac{N}{T} \rceil$.

$$\mathbf{f}_{k_i}^j = \mathcal{H}_k(\mathbf{x}_i^j) \in \mathbb{R}^B, \quad \mathbf{b}_{k_i}^j = \mathcal{H}_k^{'}(\mathbf{x}_i^j) \in \mathbb{R}^B \tag{1}$$

where $\mathbf{x}_i^j$ is the i-th image in the j-th training batch. For each training sample $\mathbf{x}_i^j$, we compute the symmetric similarities $\mathbf{C}_g$ between $\mathbf{f}_{g_i}^j$ and $\mathbf{F}_g$, and the asymmetric similarities $\mathbf{C}_q$ between $\mathbf{f}_{q_i}^j$ and $\mathbf{F}_g$, making use of correlation alignment for similarity preserving:

$$\begin{aligned} \mathbf{C}_g &= [\mathbf{f}_{g_i}^{t^T}\mathbf{f}_{g_1}^t, \mathbf{f}_{g_i}^{t^T}\mathbf{f}_{g_2}^t, \cdots, \mathbf{f}_{g_i}^{t^T}\mathbf{f}_{g_T}^t] \in \mathbb{R}^T \\ \mathbf{C}_q &= [\mathbf{f}_{q_i}^{t^T}\mathbf{f}_{g_1}^t, \mathbf{f}_{q_i}^{t^T}\mathbf{f}_{g_2}^t, \cdots, \mathbf{f}_{q_i}^{t^T}\mathbf{f}_{g_T}^t] \in \mathbb{R}^T \end{aligned} \tag{2}$$

Then, we compute the continuous asymmetric similarities $\mathbf{C}_m$ between $\mathbf{f}_{q_i}^j$ and $\mathbf{B}_g$, and the binary asymmetric similarities $\mathbf{C}_n$ between $\mathbf{b}_{q_i}^j$ and $\mathbf{B}_g$, making use of correlation alignment based quantization:

$$\begin{aligned} \mathbf{C}_m &= [\,\mathbf{f}_{q_i}^{t^T}\mathbf{b}_{g_1}^t\,,\ \mathbf{f}_{q_i}^{t^T}\mathbf{b}_{g_2}^t\,, \cdots,\ \mathbf{f}_{q_i}^{t^T}\mathbf{b}_{g_T}^t\,] \in \mathbb{R}^T \\ \mathbf{C}_n &= [\,\mathbf{b}_{q_i}^{t^T}\mathbf{b}_{g_1}^t\,,\ \mathbf{b}_{q_i}^{t^T}\mathbf{b}_{g_2}^t\,, \cdots,\ \mathbf{b}_{q_i}^{t^T}\mathbf{b}_{g_T}^t\,] \in \mathbb{R}^T \end{aligned} \tag{3}$$

Correlation Alignment for Similarity Preserving

The objective of correlation alignment for similarity preserving is to align asymmetric similarities with symmetric similarities, thereby facilitating the implicit convergence of continuous features from query models to those of gallery models. Specifically, this process first leverages a temperature-scaled softmax the asymmetric similarities between query and gallery models with their symmetric similarities, smoothing the probability distributions. Building on this, it further enforces similarity alignment through a relaxed yet precise correlation constraint, refining the consistency between the two types of similarities. Ultimately, the optimized query model maintains feature similarity with the gallery model in the continuous domain, effectively preserving the desired similarity.

KL Divergence. The KL divergence D_{KL} is employed to optimize the consistency between symmetric similarities and asymmetric similarities. To apply KL divergence, we first transform the symmetric and asymmetric similarities into probability distributions via temperature-scaled softmax operations:

$$p_j^i = \frac{\exp\left(C_j^i/\tau_j\right)}{\sum_{l=1}^{T}\exp\left(C_j^l/\tau_j\right)}, \text{ for } i = 1, 2, \cdots, T. \quad (4)$$

where τ_j denotes the similarity-preserving temperature coefficient and $j \in \{q, g, m, n\}$. Specifically, τ_g is set to be smaller than 1 to regulate the sharpness of the probability distribution, ensuring that the $\mathcal{H}_q(\cdot)$ concentrates primarily on the local neighborhood structure of training images rather than distant outliers. The defined correlation alignment for similarity preserving loss $\mathcal{L}_{DL}$ is as follows.

$$\mathcal{L}_{DL} = D_{KL}\left(p_g||p_q\right) = \sum_{l=1}^{T} p_g^l \log\frac{p_g^l}{p_q^l}. \quad (5)$$

Correlation Alignment Based Quantization

As illustrated in Figure , unlike conventional quantization schemes that explicitly minimize the gap between continuous and binary query features, our approach focuses on aligning semantic similarities rather than enforcing strict feature-level binarization. Specifically, it comprises two components: the *correlation alignment loss* and the *correlation distance loss*.

Correlation Alignment Loss. From Eq. 4, p_m and p_n are derived, which represent the transformation of continuous and binary asymmetric similarities into probability distributions, respectively. Temperature coefficient τ_m is set less than τ_n keep $\mathcal{H}_q^{'}(\cdot)$ primarily focuses on the local neighborhood correlation of training images, rather than distant outliers. This is because a lower τ_m encourages the student model to learn generalized knowledge from the teacher's smooth distribution (high τ_n) while focusing on local features for precise binary coding. The defined correlation alignment loss $\mathcal{L}_{CA}$ is as follows.

$$\mathcal{L}_{CA} = D_{KL}\left(q_m||q_n\right) = \sum_{l=1}^{T} q_m^l \log\frac{q_m^l}{q_n^l}. \quad (6)$$

Correlation Distance Loss. A naive strategy is to promote both models to have close distance for C_m and C_n. To quantify the closeness between continuous asymmetric similarities and binary asymmetric similarities, $\mathcal{L}_1$ and $\mathcal{L}_2$ distance metrics are two most widely, with which we define the correlation distance loss $\mathcal{L}_s$ as follows,

$$\mathcal{L}_s = \left(\sum_{i=1}^{T}\left|C_m^i - C_n^i\right|^{\alpha}\right)^{\frac{1}{\alpha}}, \ \alpha = 1, 2. \quad (7)$$

By merging Eq. (6) with Eq. (7), we arrive at the definitive formulation of correlation alignment based quantization (CAQ) loss, enabling the model to jointly optimize multiple objectives simultaneously.

$$\mathcal{L}_{CAQ} = \gamma\mathcal{L}_{CA} + \beta\mathcal{L}_s. \quad (8)$$

where γ and β is a hyper-parameter that makes a trade-off in $\mathcal{L}_{CA}$ and $\mathcal{L}_s$, and $\gamma \in \{0, 1\}$, β in the range of [0, 1].

In a nutshell, by merging Eq. (5) with Eq. (8), we arrive at the definitive formulation of $\mathcal{L}$ loss for training:

$$\mathcal{L} = \mathcal{L}_{DL} + \mathcal{L}_{CAQ}. \quad (9)$$

Experiments

Datasets

We conduct empirical evaluations of our proposed method on three widely used datasets: CIFAR-10, MS-COCO, and ImageNet.

CIFAR-10 consists of 60,000 32×32 images in 10 classes. Following Refs. (Su et al. 2018), we use 5000 images per class as the train set, 1000 images per class as the query set, and 5000 images per class as the gallery database.

MS-COCO contains 82,783 images in the training set and 40,504 images in the validation set with multi-label annotations across 80 categories. Following Ref. (Cao et al. 2017), we obtain 122,218 images with category information. Then, the dataset is partitioned into 5,000 queries and a gallery of 117,218 images, from which 10,000 are randomly selected for training.

ImageNet contains 1.2M training images and 50k validation images from 1000 classes. We follow Refs. (Cao et al. 2017) to randomly 100 categories, utilizing all training-set images from these categories as the gallery database and all validation-set images as queries.

Baselines and Backbone Networks

Baselines. We selected seven state-of-the-art deep hashing methods—DPSH (Li, Wang, and Kang 2016), HashNet (Cao et al. 2017), DSDH (Li et al. 2017), CSQ (Yuan et al. 2020), MDSH (Wang et al. 2023), HybridHash (He and Wei

Method	Datasets (%)											
	CIFAR-10				MS-COCO				ImageNet			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
Sign-Based Quantization	93.51	94.75	95.10	95.37	49.31	51.26	53.26	53.73	47.28	56.48	59.17	62.27
(A) *Supervised deep hashing methods for symmetric image retrieval*												
DPSH (Li, Wang, and Kang 2016)	88.03	90.30	91.01	92.89	47.45	48.39	49.78	50.34	46.97	55.49	58.33	60.23
HashNet (Cao et al. 2017)	94.10	94.98	95.29	95.80	49.77	52.32	54.11	55.89	49.31	57.92	60.66	62.17
DSDH (Li et al. 2017)	93.89	93.19	93.45	93.98	47.55	49.76	50.30	50.76	48.61	56.61	59.78	61.83
CSQ (Yuan et al. 2020)	94.09	95.37	95.66	96.01	49.56	51.72	53.66	54.32	49.23	57.29	60.90	62.96
MDSH (Wang et al. 2023)	85.03	85.55	85.40	87.49	43.45	45.71	47.12	48.77	40.34	48.46	50.09	51.33
HybridHash (He and Wei 2024)	94.19	95.06	95.51	96.12	49.60	51.89	53.96	55.07	49.51	57.89	60.78	62.13
(B) *Unsupervised deep hashing methods for asymmetric image retrieval*												
CSCH* (Xuan et al. 2024)	93.94	95.13	95.41	95.71	49.49	51.96	53.31	55.41	49.45	58.12	60.50	62.04
DCAH	**94.41**	**95.70**	**96.02**	**96.41**	**49.84**	**52.40**	**54.64**	**55.74**	**49.81**	**58.67**	**61.35**	**63.88**

Table 1: Performance comparison of deep hashing methods.

2024), and CSCH* (Xuan et al. 2024)—alongside six state-of-the-art asymmetric retrieval methods—CC (Peng et al. 2019), CSD (Wu et al. 2022b), ROP (Wu et al. 2023b), RAML (Suma and Tolias 2023), RKD (Xiao and Yamasaki 2024), and GranDist (Zhang et al. 2025)—as baselines. Notably, for CSCH*, we employed correlation alignment for similarity preserving to faithfully reproduce the original methodology, aligning the continuous features of the query model with the binary features of the gallery model. In our experimental setup, the query model is trained for the comparative evaluation, whereas the gallery model is kept fixed as the pre-trained hashing network. We employ deep hashing baselines to quantify the query model, comparing its performance with the CAQ method, and correlation alignment for similarity preserving component is implemented using asymmetric retrieval baselines.

Backbone networks. We adopt RN50 (ResNet50) (He et al. 2016) and RN101 (ResNet101) (He et al. 2016) as the large gallery models, while using MNetv3-S (MobileNetV3-small) (Howard et al. 2019) and MNetv3-L (MobileNetV3-large) (Howard et al. 2019) as the small query models. Specifically, we conduct experiments using two model combinations: MNetv3-S with RN50, and MNetv3-L with RN101. All backbone networks are pre-trained on ImageNet.

Implementation Details and Metric

Implementation Details. The proposed DCAH framework is implemented using PyTorch (Paszke et al. 2019). We train all models for 120 epochs on the three datasets with a batch size of 96. The Adam optimizer (Kingma and Ba 2015) is used for optimization. For MNetv3-S, the learning rate and weight decay are both set to 0.0001 across CIFAR-10, MS-COCO, and ImageNet. The hyper-parameter β is set to 0.01 on CIFAR-10 and ImageNet, and to 0 on MS-COCO. The query temperature τ_n is fixed at 0.5, while the gallery temperature τ_m is set to $[0.05, 0.01, 0.005, 0.0025]$ for 16-bit to 64-bit hash codes on CIFAR-10, and $[0.1, 0.05, 0.05, 0.005]$ on MS-COCO and ImageNet, respectively. For the quantization of the sign function, we employ the estimated straight-through estimator. All experiments are conducted on a workstation equipped with an Intel Xeon Silver 4214 CPU @ 2.20 GHz, 125 GB RAM, and an NVIDIA Tesla A800 GPU.

Evaluation Metric. We evaluate asymmetric image retrieval performance using Mean Average Precision at 1,000 retrieved items (mAP@1k), Precision-Recall curves (PR curves), and Precision at top-k (P@top-k). In contrast to symmetric hashing methods (Li, Wang, and Kang 2016; Cao et al. 2017; Li et al. 2017; Yuan et al. 2020; Wang et al. 2023; He and Wei 2024), our evaluation follows an asymmetric protocol: query images are encoded using a lightweight model to generate hash codes, while database (gallery) images are encoded using a more powerful and complex model. This asymmetric setup is consistently applied across all performance assessments, including mAP computation and the generation of PR and P@top-k curves, ensuring a fair and realistic evaluation of efficiency and accuracy in practical retrieval scenarios.

Accuracy Comparison

mAP Comparisons with Representative Hashing Methods. We conduct benchmarking experiments on six representative supervised hashing methods and one unsupervised hashing method using MNetV3-S as the query model and RN50 as the gallery model. Sign-based quantization refers to the process of directly converting floating-point features into binary features during inference. Remarkably, our method outperforms most supervised learning approaches.

Method	Datasets (%)											
	CIFAR-10				MS-COCO				ImageNet			
	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit	16bit	32bit	48bit	64bit
CC (Peng et al. 2019)	90.79	92.34	92.89	93.48	46.58	48.32	50.89	51.45	49.59	57.86	60.13	61.07
CC+CAQ	**91.66**	**92.93**	**93.36**	**94.10**	**47.78**	**49.93**	**52.56**	**52.34**	**50.97**	**59.59**	**61.19**	**62.51**
CSD (Wu et al. 2022b)	93.51	94.75	95.10	95.37	49.31	51.26	53.26	53.73	47.28	56.48	59.17	62.27
CSD+CAQ	**94.41**	**95.70**	**96.02**	**96.41**	**49.84**	**52.40**	**54.64**	**55.74**	**49.81**	**58.67**	**61.35**	**63.88**
ROP (Wu et al. 2023b)	93.58	94.67	94.89	95.18	49.58	51.22	52.94	53.67	48.59	55.17	57.82	58.48
ROP+CAQ	**94.39**	**95.18**	**95.66**	**95.80**	**50.23**	**52.18**	**54.26**	**54.98**	**49.83**	**56.94**	**58.23**	**60.21**
RAML (Suma and Tolias 2023)	93.01	95.12	95.34	95.67	47.86	50.19	52.64	53.23	46.26	54.60	58.57	60.03
RAML+CAQ	**94.19**	**95.76**	**95.98**	**96.23**	**48.53**	**51.56**	**53.58**	**54.43**	**48.37**	**56.54**	**60.42**	**61.57**
RKD (Xiao and Yamasaki 2024)	93.20	95.01	95.28	95.45	47.87	50.95	52.65	53.46	49.09	57.90	60.81	61.81
RKD+CAQ	**94.38**	**95.40**	**95.67**	**95.99**	**50.09**	**51.88**	**53.37**	**54.89**	**50.68**	**58.81**	**61.31**	**63.12**
GranDist (Zhang et al. 2025)	92.30	93.78	94.42	94.79	48.97	50.42	53.43	53.95	44.90	53.71	57.34	59.96
GranDist+CAQ	**93.11**	**94.50**	**95.09**	**95.89**	**50.43**	**51.98**	**54.12**	**55.23**	**46.55**	**55.23**	**58.12**	**61.03**

Table 2: Quantitative effects of plug-and-play CAQ on existing asymmetric retrieval methods.

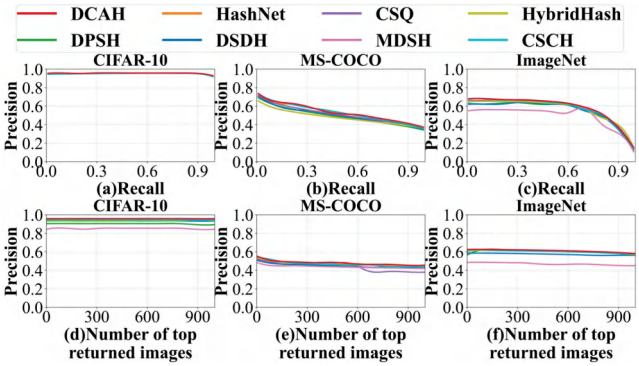

Figure 4: Precision–recall curves and Precision@top-1k on CIFAR-10, ImageNet, and MS-COCO with 32 bits.

As shown in table 1, comprehensive evaluations on CIFAR-10, MS-COCO, and ImageNet100 demonstrate DCAH's consistent superiority: it achieves 96.59% mAP on CIFAR-10, outperforming the HybridHash (96.12%) by 0.47% in 64-bit. HashNet, CSQ, and HybridHash exhibit quite good retrieval performance in benchmarks. For instance, HashNet reaches 55.89% mAP in 64-bit hashing through its stepwise quantization strategy. These detailed labels provide the necessary supervisory signals for HashNet's progressive quantization to effectively minimize information loss during discrete encoding. Our experimental results show consistent performance across 16-bit to 48-bit hashing, suggesting the method maintains reasonable generalization across different encoding lengths.

The Effects of the CAQ Plug-and-Play Hashing Method. We systematically integrate the CAQ plug-and-play hashing module into state-of-the-art asymmetric retrieval frameworks. For the AIR method, we directly use the sign symbol for quantification. Without requiring additional inputs, by optimizing the transition from discrete asymmetric similarity approximation to continuous asymmetric similarity preservation within a lightweight query model, it not only reduces quantization error in the query model but also effectively mitigates the performance gap in asymmetric hashing. Specifically, the table 2 shows consistent mAP improvements of 0.5–2% across CIFAR-10, MS-COCO, and ImageNet datasets under 16–64-bit configurations with MNetV3-S (query model) and RN50 (gallery model). These results validate the advantages of the CAQ plug-and-play hashing method, that is readily quantifiable through straightforward integration of CAQ into asymmetric image retrieval frameworks.

Ablation Study

Impact of Different Loss Terms. We evaluate different combinations of losses $\mathcal{L}_{DL}$, $\mathcal{L}_{CA}$ and $\mathcal{L}_s$. As shown in Table 3, when removing $\mathcal{L}_{CA}$ and $\mathcal{L}_s$, the performance reduces significantly on the datasets for all bits. Quantitative analysis demonstrates that the exclusion of quantization techniques invariably diminishes the training efficacy of hashing networks in asymmetric retrieval scenarios. The introduction of correlation alignment loss $\mathcal{L}_{CA}$ and correlation distance loss $\mathcal{L}_s$ yields significant performance gains on datasets, validating its critical contribution to the quantization paradigm. We observe that $\mathcal{L}_{CA}$ primarily drives the process, with $\mathcal{L}_s$ acting as an auxiliary component. In the MS-COCO dataset, the combination of $\mathcal{L}_s$ and $\mathcal{L}_{CA}$ exhibits side effects, leading to weight of $\mathcal{L}_s$ being set to zero.

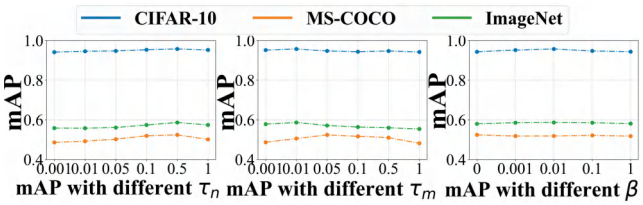

Figure 5: mAP changes with different τ_n, τ_m, β at 32 bits code length on CIFAR-10, MS-COCO and ImageNet.

$\mathcal{L}_{DL}$	$\mathcal{L}_{CA}$	$\mathcal{L}_1$	$\mathcal{L}_2$	Datasets(%)		
				CIFAR-10	MS-COCO	ImageNet
✓				94.75	51.26	56.48
✓	✓			95.26	52.40	58.02
✓		✓		94.95	51.58	57.30
✓			✓	94.90	51.49	57.21
✓	✓		✓	95.66	52.40	58.49
✓	✓	✓		**95.70**	**52.40**	**58.67**

Table 3: mAP comparisons with different combinations of losses with 32 bits.

PR and P@top-k Comparisons

To comprehensively evaluate the retrieval quality of DCAH, we present PR curves and p@k curves for the top-1,000 retrieved images using 32-bit hash codes. As shown in Figure 4, DCAH substantially outperforms all existing deep hashing methods, achieving significant margins across both evaluation metrics. Notably, DCAH demonstrates superior retrieval characteristics: it maintains higher precision at low recall levels while retrieving more relevant samples in top-ranking positions. These experimental findings confirm DCAH's efficacy in practical large-scale asymmetric retrieval scenarios.

Parameter Sensitivity

To further analyze the sensitivity of these parameters, we conduct experiments under different values of τ_n, τ_m and β on CIFAR-10, MS-COCO and ImageNet datasets. The temperature parameter τ_n and τ_m adjust the smoothness and sharpness of the distribution, enabling a more accurate calculation of the KL divergence in Eq. (4). And β is the hyper-parameter that makes a trade-off between correlation alignment loss and correlation quantization loss in Eq. (9).

The variation in mAP with different parameter settings at a code length of 32 bits is plotted in Figure 5, with τ_n, τ_m in the range of [0.0001, 1] and β in [0, 1]. We find that the trade-off value of τ_m and β is dataset-dependent, and the values of τ_m and β differ when obtaining the highest performance on different datasets. And our systematic parameter optimization revealed peak performance on CIFAR-10 at τ_n=0.5, τ_m=0.05, on MS-COCO at τ_n=0.5, τ_m=0.05, and on ImageNet at τ_n=0.5, τ_m=0.01. We observe that an appropriate β will minimize the intra-class variance and contribute to high-quality hash codes. Our CAQ hashing method achieve peak mAP at β=0.01 on ImageNet, while β=0 yields optimal MS-COCO performance and β=0.1 maximizes results on CIFAR-10.

Figure 6: Results visualization on ImageNet with 32 bits.

Results Visualization

We randomly choose one query image from the single-label dataset ImageNet to perform the similarity search. The figure 6 shows the top-5 retrieved images of DCAH and CSCH*. Our DCAH exhibits the ability to produce high-quality feature, resulting in more appropriate and useful retrieval outcomes for users. Notably, while CSCH*'s features exhibit stronger discretization with values clustered near -1 and 1, DCAH demonstrably outperforms CSCH* in retrieval results.

Efficiency Comparison

Training Efficiency. In the DCAH framework, our total training time consumption is divided into two main components: similarity preservation and quantification. Under the same experimental setup, DCAH demonstrates the fastest training time per batch (0.774s), outperforming all compared methods including DPSH (0.896s), HashNet (0.947s), DSDH (0.976s), CSQ (0.865s), MDCSQ (1.145s), HybridHash (0.935s), and CSCH* (0.803s).

Conclusions

This paper proposes Deep Correlation Alignment Hashing (DCAH), a novel label-free deep hashing method for AIR. We argue that achieving effective binarization while preserving semantic similarity is difficult and also unnecessary for the query network. Actually, discretization is not the only way to reduce quantization errors. In this paper, DCAH introduces a correlation alignment based quantization scheme to achieve implicit quantization. The proposed correlation alignment based quantization scheme integrates correlation distribution approximation and correlation distance constraints. For similarity preserving, we further adopt a correlation alignment-based knowledge distillation strategy that is inherently compatible with the quantization scheme. Extensive experiments across three benchmarks demonstrate DCAH's consistent superiority, enhancing lightweight model performance and outperforming state-of-the-art approaches.

Acknowledgments

This work was supported by the National Natural Science Foundation of China under Grant 62472420 and the Natural Science Foundation of Guangdong under Grant 2025A1515011224.

References

Budnik, M.; and Avrithis, Y. 2021. Asymmetric metric learning for knowledge transfer. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Cao, Y.; Long, M.; Liu, B.; and Wang, J. 2018. Deep cauchy hashing for hamming space retrieval. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Cao, Z.; Long, M.; Wang, J.; and Yu, P. S. 2017. Hashnet: Deep learning to hash by continuation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Fan, L.; Ng, K. W.; Ju, C.; Zhang, T.; and Chan, C. S. 2020. Deep Polarized Network for Supervised Learning of Accurate Binary Hashing Codes. In *Proc. of the Intl. Conf. on Artificial Intelligence (IJCAI)*.

Gionis, A.; Indyk, P.; Motwani, R.; et al. 1999. Similarity search in high dimensions via hashing. In *Proc. of the International Conference on Very Large Data Bases (VLDB)*.

Gong, Y.; Lazebnik, S.; Gordo, A.; and Perronnin, F. 2013. Iterative Quantization: A Procrustean Approach to Learning Binary Codes for Large-Scale Image Retrieval. *IEEE Trans. on Pattern Analalysis and Machine Intelligence (TPAMI)*.

Gu, J.; Wu, D.; Fu, P.; Li, B.; and Wang, W. 2022. Deep piecewise hashing for efficient hamming space retrieval. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*.

He, C.; and Wei, H. 2024. HybridHash: Hybrid convolutional and self-attention deep hashing for image retrieval. In *Proc. of international conference on multimedia retrieval (ICMR)*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep Residual Learning for Image Recognition. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Hoe, J. T.; Ng, K. W.; Zhang, T.; Chan, C. S.; Song, Y.; and Xiang, T. 2021. One Loss for All: Deep Hashing with a Single Cosine Similarity based Learning Objective. In *Proc. of the Conference on Neural Information Processing Systems (NeurIPS)*.

Howard, A.; Pang, R.; Adam, H.; Le, Q. V.; Sandler, M.; Chen, B.; Wang, W.; Chen, L.; Tan, M.; Chu, G.; Vasudevan, V.; and Zhu, Y. 2019. Searching for MobileNetV3. In *Proc. of the IEEE/CVF Intl. Conf. on Computer Vision (ICCV)*.

Jiang, Q.; and Li, W. 2018. Asymmetric Deep Supervised Hashing. In *Proc. of the Conference on Advancements of Artificial Intelligence (AAAI)*.

Kingma, D. P.; and Ba, J. 2015. Adam: A Method for Stochastic Optimization. In *Proc. of the Int. Conf. on Learning Representations (ICLR)*.

Li, Q.; Sun, Z.; He, R.; and Tan, T. 2017. Deep Supervised Discrete Hashing. In *Proc. of the Conference on Neural Information Processing Systems (NeurIPS)*.

Li, W.; Wang, S.; and Kang, W. 2016. Feature Learning Based Deep Supervised Hashing with Pairwise Labels. In *Proc. of the Intl. Conf. on Artificial Intelligence (IJCAI)*.

Liu, H.; Wang, R.; Shan, S.; and Chen, X. 2016. Deep Supervised Hashing for Fast Image Retrieval. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; Desmaison, A.; Köpf, A.; Yang, E. Z.; DeVito, Z.; Raison, M.; Tejani, A.; Chilamkurthy, S.; Steiner, B.; Fang, L.; Bai, J.; and Chintala, S. 2019. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In *Proc. of the Conference on Neural Information Processing Systems (NeurIPS)*.

Peng, B.; Jin, X.; Liu, J.; Li, D.; Wu, Y.; Liu, Y.; Zhou, S.; and Zhang, Z. 2019. Correlation congruence for knowledge distillation. In *Proc. of the IEEE/CVF Intl. Conf. on Computer Vision (ICCV)*.

Pu, R.; Qin, Y.; Song, X.; Peng, D.; Ren, Z.; and Sun, Y. 2025a. SHE: Streaming-media Hashing Retrieval. In *Proc. of the Int. Conf. on Machine Learning (ICML)*.

Pu, R.; Sun, Y.; Qin, Y.; Ren, Z.; Song, X.; Zheng, H.; and Peng, D. 2025b. Robust Self-Paced Hashing for Cross-Modal Retrieval with Noisy Labels. In *Proc. of the Conf. Artif. Intell. (AAAI)*.

Shen, F.; Gao, X.; Liu, L.; Yang, Y.; and Shen, H. T. 2017. Deep Asymmetric Pairwise Hashing. In *Proc. of the ACM International Conference on Multimedia (ACM MM)*.

Su, Q.; Wu, D.; and Li, B. 2025. Boundary-aware Prototype Augmentation and Dual-level Knowledge Distillation for Non-Exemplar Class-Incremental Hashing. *Knowledge-Based Systems (KBS)*.

Su, Q.; Wu, D.; Wu, C.; Li, B.; and Wang, W. 2024. From data to optimization: Data-free deep incremental hashing with data disambiguation and adaptive proxies. *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*.

Su, S.; Zhang, C.; Han, K.; and Tian, Y. 2018. Greedy hash: Towards fast optimization for accurate hash coding in cnn. *Proc. of the Advances in Neural Information Processing Systems (NIPS)*.

Suma, P.; and Tolias, G. 2023. Large-to-small image resolution asymmetry in deep metric learning. In *Proc. of the IEEE Winter Conf. on Applications of Computer Vision (WACV)*.

Sun, Y.; Ren, Z.; Hu, P.; Peng, D.; and Wang, X. 2023. Hierarchical consensus hashing for cross-modal retrieval. *IEEE Trans. on Multimedia (TMM)*.

Wang, L.; Pan, Y.; Liu, C.; Lai, H.; Yin, J.; and Liu, Y. 2023. Deep Hashing With Minimal-Distance-Separated Hash Centers. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Weiss, Y.; Torralba, A.; and Fergus, R. 2008. Spectral Hashing. In *Proc. of the Conference on Neural Information Processing Systems (NeurIPS)*.

Wu, D.; Dai, Q.; Li, B.; and Wang, W. 2023a. Deep uncoupled discrete hashing via similarity matrix decomposition. *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM)*.

Wu, D.; Dai, Q.; Liu, J.; Li, B.; and Wang, W. 2019. Deep incremental hashing network for efficient image retrieval. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Wu, D.; Su, Q.; Li, B.; and Wang, W. 2022a. Efficient Hash Code Expansion by Recycling Old Bits. In *Proc. of the ACM International Conference on Multimedia (ACM MM)*.

Wu, D.; Su, Q.; Li, B.; and Wang, W. 2024. Pairwise-label-based deep incremental hashing with simultaneous code expansion. In *Proc. of the Conf. Artif. Intell. (AAAI)*.

Wu, H.; Wang, M.; Zhou, W.; and Li, H. 2023b. A general rank preserving framework for asymmetric image retrieval. In *Proc. of the Int. Conf. on Learning Representations (ICLR)*.

Wu, H.; Wang, M.; Zhou, W.; Li, H.; and Tian, Q. 2022b. Contextual Similarity Distillation for Asymmetric Image Retrieval. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Xia, R.; Pan, Y.; Lai, H.; Liu, C.; and Yan, S. 2014. Supervised Hashing for Image Retrieval via Image Representation Learning. In *Proc. of the Conf. Artif. Intell. (AAAI)*.

Xiao, L.; and Yamasaki, T. 2024. Boosting fine-grained fashion retrieval with relational knowledge distillation. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Xie, Y.; Lin, Y.; Cai, W.; Xu, X.; Zhang, H.; Du, Y.; and He, S. 2024. D3still: Decoupled differential distillation for asymmetric image retrieval. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Xuan, Z.; Wu, D.; Zhang, W.; Su, Q.; Li, B.; and Wang, W. 2024. Central similarity consistency hashing for asymmetric image retrieval. *Computational Visual Media (CVMJ)*.

Yang, D.; Wu, D.; Zhang, W.; Zhang, H.; Li, B.; and Wang, W. 2020. Deep semantic-alignment hashing for unsupervised cross-modal retrieval. In *Proc. of international conference on multimedia retrieval (ICMR)*.

Yuan, L.; Wang, T.; Zhang, X.; Tay, F. E. H.; Jie, Z.; Liu, W.; and Feng, J. 2020. Central Similarity Quantization for Efficient Image and Video Retrieval. In *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*.

Zhang, H.; Xie, Y.; Zhang, H.; Xu, C.; Luo, X.; Chen, D.; Xu, X.; Zhang, H.; Heng, P. A.; and He, S. 2025. Unambiguous granularity distillation for asymmetric image retrieval. *Neural Networks (NN)*.

Zhu, H.; Long, M.; Wang, J.; and Cao, Y. 2016. Deep Hashing Network for Efficient Similarity Retrieval. In *Proc. of the Conf. Artif. Intell. (AAAI)*.