计 算 机 研 究 与 发 展
Journal of Computer Research and Development
DOI: 10.7544/issn1000-1239.202330433
61(11): 2923−2936, 2024

双重结构保持的在线跨模态哈希

康 潇[1] 刘兴波[2] 卢鹏宇[2] 赵志杰[2] 聂秀山[2] 王少华[2] 尹义龙[1]
[1](山东大学软件学院 济南 250101)
[2](山东建筑大学计算机科学与技术学院 济南 250101)
(sckx@mail.sdu.edu.cn)

Online Cross-Modal Hashing with Double Structure Preserving

Kang Xiao[1], Liu Xingbo[2], Lu Pengyu[2], Zhao Zhijie[2], Nie Xiushan[2], Wang Shaohua[2], and Yin Yilong[1]
[1](*School of Software*, *Shandong University*, *Jinan* 250101)
[2](*School of Computer Science and Technology*, *Shandong Jianzhu University*, *Jinan* 250101)

Abstract Online cross-modal hashing has received increasing attention due to its efficiency and effectiveness in handling cross-modal streaming data retrieval. Despite promising progress, most existing methods rely on accurate and clear supervised information. There are only a few online cross-modal hashing studies concentrating on unsupervised learning mode, and numerous challenges still need to be tackled. For example, streaming data usually suffer from the unbalanced distribution problem due to the limited volume of data in each chunk. Most existing methods neglect this problem, resulting in heightened sensitivity to outlier samples and compromised robustness. Moreover, existing models typically exploit global data distribution, while ignoring local neighborhood information that can promote hash learning. To solve these problems, we propose an unsupervised online cross-modal hashing method with double structure-preserving, called SPOCH (structure preserving online cross-modal hashing). It simultaneously explores the global structure and local structure to generate the corresponding common representation; thereafter, the learned common representation can be used to guide the hash learning process. In terms of global structure-preserving, we design the loss function based on $L_{2,1}$ norm, which can alleviate the sensitivity of outlier samples. In terms of local structure-preserving, we reconstruct sample representation based on neighbor relations that integrates the multi-modality information. In addition, to alleviate the forgetting problem, we propose joint optimization on streaming data, and design the corresponding update strategy to improve the training efficiency. We conduct experiments on two widely-used cross-modal retrieval datasets. Compared with the existing state-of-the-art unsupervised online cross-modal hashing methods, SPOCH achieves superior retrieval accuracy within a comparable or even shorter training time, validating the effectiveness of the proposed approach.

Key words cross-modal retrieval; unsupervised learning; online hashing; double structure preserving; robustness

摘 要 近年来,在线跨模态哈希因其能处理更为贴近现实的流数据场景而受到广泛关注.虽然取得了不错的进展,但现有方法大都依赖准确清晰的数据标记.目前,针对无监督学习模式下在线跨模态哈希的研究相对较少,还有很多问题有待解决.例如,新到达的数据流通常规模较小,因此常常存在分布不平衡的现象.而现有模型极少关注这个问题,导致模型对离群样本敏感,鲁棒性较差.并且,现有方法大都关注样

收稿日期: 2023-06-05; **修回日期**: 2023-12-19
基金项目: 国家自然科学基金项目 (62176141, 62176139, 62206160); 山东省自然科学基金项目 (ZR2022QF082); 山东建筑大学特聘教授基金项目
This work was supported by the National Natural Science Foundation of China (62176141, 62176139, 62206160), the Natural Science Foundation of Shandong Province (ZR2022QF082), and the Special Funds for Distinguished Professors of Shandong Jianzhu University.
通信作者: 刘兴波 (sclxb@mail.sdu.edu.cn)

本的整体结构而忽视了邻域信息对于生成公共哈希码的帮助.为了解决上述问题,提出了基于双重结构保持的无监督在线跨模态哈希方法,称为 SPOCH（structure preserving online cross-modal hashing）.该方法的基本思想是同时挖掘样本空间的全局结构信息和邻域结构信息来生成相应的公共表示,用以指导哈希码和哈希函数的学习.针对全局结构的学习,引入$L_{2,1}$范数取代L_2范数来约束损失函数,利用$L_{2,1}$范数结构化稀疏的性质缓解模型对离群样本的敏感性;针对邻域结构的学习,利用多模态融合的邻域样本进行样本重构,使得所学公共表示更好地表征多模态信息.此外,为了缓解遗忘问题,提出在新旧数据上联合优化,并设计相应的更新策略提高算法的训练效率,实现在线检索.在 2 个广泛使用的跨模态检索数据集上进行的实验结果表明,较现有最先进的无监督在线跨模态哈希方法,SPOCH 在可比较甚至更短的训练时间内取得了更优的检索精度,验证了所提方法的有效性.

关键词 跨模态检索;无监督学习;在线哈希;双重结构保持;鲁棒性

中图法分类号 TP391

随着互联网的普及和信息技术的快速发展,媒体数据呈爆发式增长.由于传播媒介的多样化,媒体数据通常包含多种模态,并且不同模态数据之间往往是相互关联的[1-2].用户在检索某个主题时期望返回尽可能丰富且与主题相关的多模态数据.因此,如何高效准确地实现跨模态检索成为多媒体领域亟待解决的问题.跨模态哈希学习作为最有代表性的跨模态检索方法之一受到研究者的广泛关注[3-6].其旨在通过对不同模态样本信息的挖掘,将多模态数据从原始特征空间映射到公共海明空间,并尽可能保持数据的原始相似关系,从而实现快速的跨模态检索.尽管现已取得了较好的检索性能,但传统的跨模态哈希方法大都采用批量学习模式,即利用当前累积的所有训练样本学习哈希函数[5-6].当数据频繁更新且数据量不断增长时,这类方法无法根据数据的变化自适应更新模型,而是需要重新进行哈希学习,从而导致巨大的时间成本和计算开销.为了应对这一挑战,在线跨模态哈希方法应运而生.

不同于传统的离线跨模态哈希,在线方法需要在不重复训练旧数据的前提下,为新到来的数据学习哈希码,同时不断更新哈希函数使之保持对全体数据的适应性[7-12].为实现这一目的,广泛采用的设计思路是尽可能地挖掘数据间的关联关系,利用样本关系实现对旧数据的记忆和重利用.典型的策略包括挖掘新旧数据之间的相似关系和多模态特征之间的关联关系.前者[10-11]期望通过对新旧数据间分布关系的保持达到对全局数据分布的保持;后者[8,12]则期望利用模态间的特异和互补信息,减小公共哈希码与多模态特征之间的误差.尽管现有的在线跨模态哈希方法取得了不错的检索性能,但其模型实现通常依赖于准确的语义信息指导,即监督训练模式.然而,在现实生活中,监督方法存在数据标注成本高、难以获取足够的标记数据、标记数据质量差等问题.

无监督在线跨模态哈希学习或许是解决这一问题的有效方案[7-9].其基本思想是在没有标记数据的情况下,仅利用数据的内在结构和相似性信息学习哈希码和哈希函数.虽然避免了标记数据难获取的弊端,但现有的无监督在线跨模态哈希方法仍存在很多问题需要解决.例如:1)现有模型对离群样本敏感,模型鲁棒性较差.现实场景中,流数据因更新频繁且数据块规模较小,往往存在类别分布不平衡问题[13].换言之,数据块中存在一些样本数量很少的类别,在无监督训练模式下,这些样本从数据分布上表现为离群点.现有的在线跨模态哈希方法大都采用均方误差来优化损失函数.这一损失通常用于衡量模型整体的拟合情况,因而会放大离群样本的偏差对于模型的影响,导致检索性能下降.2)现有模型对原始空间中样本的分布信息挖掘不充分.无监督学习模式下,由于缺乏多模态一致的高层语义信息的指导,充分挖掘多模态一致的表征信息显得尤为重要[7-8].现有模型大都通过学习线性映射将多模态特征转移到公共空间中,并期望学习到的公共表示具有模态一致性和样本可区分性.然而,受限于多模态数据异构,很难仅凭线性变换充分挖掘出模态间的一致信息.因而,公共空间的表征能力仍有待进一步提升.

为了缓解上述问题,我们提出一种新的无监督在线跨模态方法.该方法设计了双重结构保持框架.一方面,引入$L_{2,1}$范数取代L_2范数学习全局结构保持的公共表示,从而缓解模型对于离群样本敏感的问题,提升模型的鲁棒性.另一方面,通过融合的多模态特征进行邻域样本重构,学习具有邻域结构保持的公共表示.最后,将两者结合用于指导哈希码和哈希函数的学习.此外,为了保证哈希函数对于新旧数据的适应性,我们采用新旧数据联合优化的模式,并

设计相应的优化策略使模型的更新效率仅于新增数据的规模相关，从而实现在线检索.

本文的主要贡献包括 3 个方面：

1）提出了基于双重结构保持的在线跨模态哈希方法. 在不依赖语义标签的前提下，学习具有全局结构保持和邻域信息保持的公共表示用以指导哈希学习，从而提升哈希码的判别性和鲁棒性.

2）为了避免在线优化过程中的灾难性遗忘问题，我们引入新旧数据的联合优化，并设计相应的在线优化方案降低训练的时间复杂度，从而实现灵活高效的在线检索.

3）在 2 个广泛使用的多模态数据集上的对比实验表明本文方法在表现出更优的检索精度的同时，保持着不错的训练效率，证实了本文方法在在线检索中的有效性和实用性.

1 相关工作

在本节中，我们将从单模态和跨模态这 2 个方面，简单回顾在线哈希学习相关的工作.

1）单模态在线哈希方法

早期的在线哈希研究主要关注单模态检索任务，根据是否利用监督信息，可以粗略地分为监督方法和无监督方法. 监督在线哈希方法主要依靠标签信息或成对相似关系来学习哈希码，以实现高效的图像检索、语义检索等任务. 2015 年，Cakir 等人[14] 提出在线监督哈希（online supervised hashing, OSH）. 该方法采用 2 步哈希学习框架，先根据标签信息为每个类分配 1 个 ECOC（error correct output code）向量，然后通过优化哈希码和 ECOC 向量之间的损失来学习哈希函数. 在 OSH 的基础上，基于哈达玛矩阵的在线哈希（Hadamard matrix guided online hashing, HMOH）[15] 将哈达玛矩阵引入在线哈希学习框架中. 该方法将哈达玛矩阵中的列作为类标签的代理码，并采用局部灵敏度哈希（local sensitivity hashing, LSH）映射[2] 对齐代理码和待学习哈希码，从而加快在线训练速度. 相似保持链接哈希（similarity preserving linking hashing, SPLH）[16] 将语义标签和哈希码映射到链接空间，通过这个链接空间可以更准确地衡量语义相似度. 平衡相似度在线离散哈希（balance similarity online discrete hashing, BSODH）[13] 尝试构造新旧数据间的非对称成对相似图，并引入平衡因子来调节相似样本对和不相似样本对的权重，从而缓解了新旧数据规模差异大所导致的更新不平衡问题. 快速类更新在线哈希（fast class-wise online hashing, FCOH）[17] 则设计逐类更新的方法，在解决更新不平衡问题的同时提高哈希学习的效率.

不同于监督哈希，无监督在线哈希方法不需要借助标签或成对相似矩阵的指导，而是通过挖掘样本特征表示的内在结构和分布关系来学习哈希码和哈希函数[18-19]. 在线核哈希（online kernel hashing, OKH）[20] 是无监督在线哈希领域的第 1 次尝试，它将核函数引入到在线哈希学习中，利用在线凸优化算法实现哈希函数的在线优化. 由于 OKH 在训练过程中每轮只能处理 1 对数据，因此不适合大规模数据检索. 为此，研究者提出基于数据草图的方法，如 OSketchH（online sketching hashing）[21]，FROSH（faster online sketching hashing）[22]，MOSH（making online sketching hashing even faster）[23]. 这些方法从每个数据块中提取概要信息，并根据这些概要信息更新哈希模型，由此降低了计算复杂度，训练效率也得到有效提升.

2）跨模态在线哈希方法

近年来，随着流数据的爆发式增长，在线跨模态哈希引起了越来越多的关注. 现有的在线跨模态哈希方法可以大致分为 2 类：监督方法和无监督方法. 与单模态方法类似，监督在线跨模态方法延续了利用语义标签和成对相似关系来学习哈希码的技术路线. 例如，在线潜在语义哈希（online latent semantic hashing, OLSH）[10] 将离散的标签映射到连续的潜在语义空间，从而更准确地度量多模态流数据的语义相似度. 基于标签嵌入的在线哈希（label embedding online hashing, LEMON）[11] 通过建立非对称相似矩阵来探索新旧数据之间、不同模态之间的关联关系，从而缓解模型对旧数据的遗忘并弥合不同模态间的语义鸿沟. 离散在线跨模态哈希（discrete online cross-modal hashing, DOCH）方法[12] 则通过极大似然函数来约束数据块之间的非对称相似性，从而更好地保持新旧数据关系.

相对地，无监督方法主要通过分解和重构特征表示来探索多模态流数据之间的底层几何结构，进而学习哈希码和哈希函数. 由于其新颖性和挑战性，仅有少量的无监督在线跨模态哈希方法被提出. Xie 等人[7] 提出在线跨模态哈希（online cross-modal hashing, OCMH），通过优化共享潜在码和动态转移矩阵来学习哈希码. 在线跨模态矩阵分解哈希（online cross-modal matrix factorization hashing, OCMFH）[8] 通过定制的在线优化策略使模型的训练时间线性于新数据规模. 并且，OCMFH 提出在不访问旧数据特征表示的前提下更新旧数据哈希码的方案，使得哈希

函数和哈希码的学习更加协调. 在线无监督跨模态离散哈希(online unsupervised cross-modal discrete hashing, OUCDH)[9]引入了非对称锚点图策略,在每轮的数据块上学习代表性数据作为锚点,通过锚点的分布来表征累计数据分布,从而缓解遗忘问题.

2 基于双重结构保持的在线跨模态哈希方法

本节将从4个方面详细介绍所提方法,包括:符号定义、模型构建、算法优化过程及时间复杂度分析.

2.1 符号定义

假设训练样本以流数据的形式到达,每个样本具有M个模态表示. 在第t轮训练时会加入新的数据块$\boldsymbol{X}^{(t)}=\left(\boldsymbol{X}_1^{(t)},\boldsymbol{X}_2^{(t)},\cdots,\boldsymbol{X}_M^{(t)}\right)$,其中$\boldsymbol{X}_m^{(t)}\in\mathbb{R}^{d_m\times n_t}$,$d_m$表示第$m$个模态的样本维度,$n_t$表示新到的样本数. 前$t-1$轮的累计数据用$\boldsymbol{X}^{(t-1)}=\left(\boldsymbol{X}_1^{(t-1)},\boldsymbol{X}_2^{(t-1)},\cdots,\boldsymbol{X}_M^{(t-1)}\right)$表示,其中$\boldsymbol{X}_m^{(t-1)}\in\mathbb{R}^{d_m\times N_t}$,$N_t$表示累积数据的样本数. $\boldsymbol{B}=\left(\boldsymbol{B}^{(t)},\boldsymbol{B}^{(t-1)}\right)$表示所学哈希码,其中$\boldsymbol{B}^{(t)}\in\mathbb{R}^{r\times n_t}$,$\boldsymbol{B}^{(t-1)}\in\mathbb{R}^{r\times N_t}$,$r$表示哈希码长. $\|\cdot\|_2$和$\|\cdot\|_{2,1}$分别表示L_2范数和$L_{2,1}$范数,sgn(·)为符号函数,如果$x>0$,则$\mathrm{sgn}(x)=1$,否则$\mathrm{sgn}(x)=-1$.

2.2 模型构建

如图1所示,本文所提的SPOCH(structure preserving online cross-modal hashing)主要包括3个模块:鲁棒的全局结构保持、融合的邻域结构保持和基于双重结构保持的在线哈希学习. 在本节中,我们将从这3个部分入手,详细介绍模型构建过程.

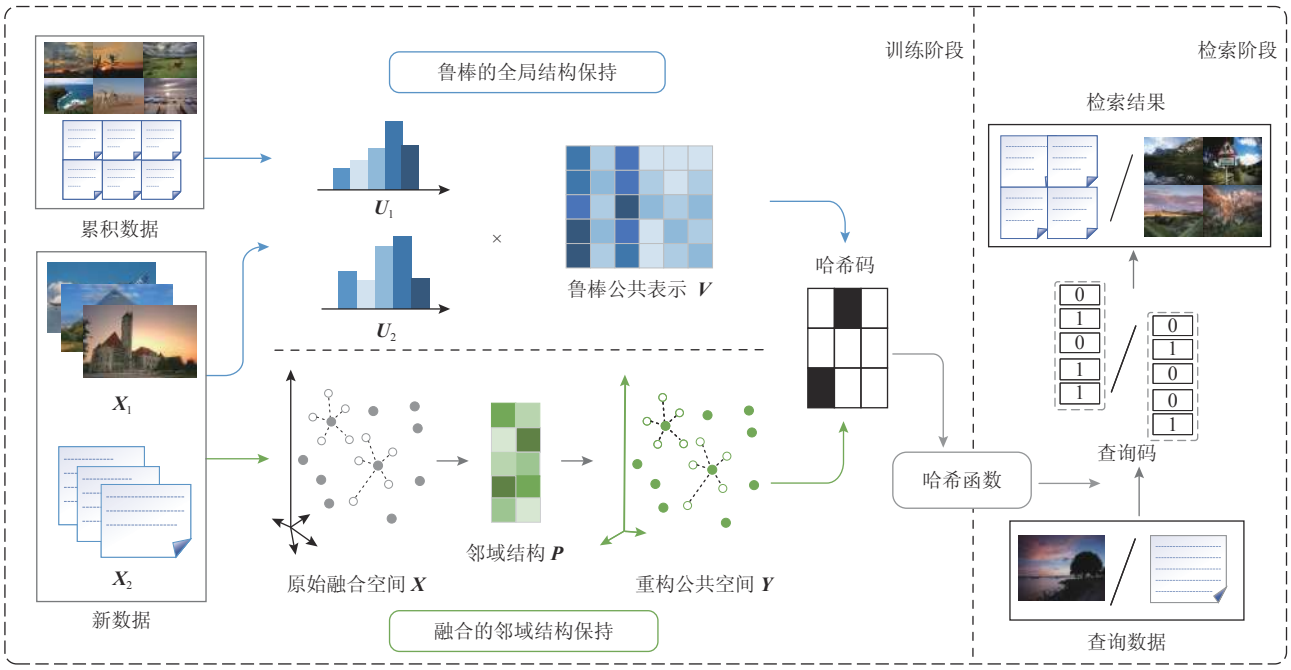


Fig. 1 Framework of SPOCH

图1 SPOCH框架

2.2.1 鲁棒的全局结构保持

由于缺乏可信的监督信息指导,无监督跨模态哈希学习通常需要充分挖掘多模态数据的潜在几何结构[7-9]. 广泛采纳的学习框架是联合矩阵分解[24],它通过为每个模态学习对应的特征转移矩阵$\boldsymbol{U}_m$,将不同模态的特征表示$\boldsymbol{X}_m$映射为连续的公共表示$\boldsymbol{V}$. 这个过程可以建模为

$$\min_{\boldsymbol{U}_m,\boldsymbol{V}}\sum_{m=1}^{M}\alpha_m\|\boldsymbol{X}_m-\boldsymbol{U}_m\boldsymbol{V}\|_2\,, \tag{1}$$

其中α_m为超参数.

尽管在传统的离线跨模态哈希方法中,式(1)所示的联合矩阵分解框架可以取得不错的检索性能,但在更为贴近现实生活的在线检索场景下,流数据样本频繁更新,每个批次的样本数据量较少,常常会出现数据类别分布不平衡的现象,从而导致每个批次可用训练样本中存在少量的离群样本[13]. 而式(1)采用均方误差来衡量优化损失,其对离群样本较为敏感,使得模型的鲁棒性较差,影响模型的检索性能.

为此,我们提出采用$L_{2,1}$范数取代L_2范数来约束损失函数. 矩阵的$L_{2,1}$范数为矩阵中所有行的L_2范数

之和，即 $\|\mathcal{A}\|_{2,1}=\sum_i\|\mathcal{A}_i\|$. 不同于 L_2 范数使矩阵中所有元素的均方误差最小化，$L_{2,1}$ 范数旨在实现损失矩阵的结构化稀疏[25]. 换言之，$L_{2,1}$ 范数期望出现更多优化损失小的样本(即更多的 0 或接近 0 的值)，而非追求全部样本的平均优化损失最小化，从而弱化模型对离群样本的关注，提升鲁棒性. 进一步地，为了适应在线学习的需求，我们在新旧数据上联合优化转移矩阵 $\boldsymbol{U}_m$，并为映射矩阵 $\boldsymbol{U}_m$ 添加 L_2 范数正则化以防止过拟合，提高模型的泛化能力[26]. 这个过程可以建模为

$$\min_{\boldsymbol{U}_m,\boldsymbol{V}^{(t)}}\sum_{m=1}^{M}\alpha_m\left(\left\|\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right\|_{2,1}+\left\|\boldsymbol{X}_m^{(t-1)}-\boldsymbol{U}_m\boldsymbol{V}^{(t-1)}\right\|_{2,1}\right)+\sum_{m=1}^{M}\gamma\|\boldsymbol{U}_m\|_2. \tag{2}$$

2.2.2 融合的邻域结构保持

多模态数据的异构性使得不同模态的特征表示之间存在巨大的语义鸿沟，这导致不同模态的数据之间很难相互检索. 在本节中，我们提出基于多模态特征融合的邻域结构保持框架，通过挖掘多模态融合后的特征 $\boldsymbol{X}^{(t)}=\left(\boldsymbol{X}_1^{(t)},\boldsymbol{X}_2^{(t)},\cdots,\boldsymbol{X}_m^{(t)}\right)$ 的邻域结构信息，为样本学习邻域结构保持的多模态公共表示. 具体地，我们首先学习基于邻域的线性组合 $\boldsymbol{P}^{(t)}$ 来对训练样本重表示，这个过程可以建模为

$$\begin{aligned}&\min_{\boldsymbol{P}^{(t)}}\sum_{i=1}^{n_t}\left\|\boldsymbol{x}_i^{(t)}-\sum_{j\in Q_i}P_{ij}^{(t)}\boldsymbol{x}_j^{(t)}\right\|_2^2,\\&\text{s.t.}\ \sum_{j\in Q_i}P_{ij}^{(t)}=1,\end{aligned} \tag{3}$$

其中 Q_i 表示第 i 个样本 $\boldsymbol{x}_i^{(t)}$ 的邻域样本集合. 训练样本的邻域结构信息将被保存在亲和矩阵 $\boldsymbol{P}^{(t)}$ 中. 利用 $\boldsymbol{P}^{(t)}$ 学习邻域结构保持的 $\boldsymbol{Y}^{(t)}=\left(\boldsymbol{y}_1^{(t)},\boldsymbol{y}_2^{(t)},\cdots,\boldsymbol{y}_{n_t}^{(t)}\right)$，并约束 $\boldsymbol{Y}^{(t)}$ 中的样本两两正交，从而提升其表征能力，减少冗余信息. 该过程表示为

$$\begin{aligned}&\min_{\boldsymbol{Y}^{(t)}}\sum_{i=1}^{n_t}\left\|\boldsymbol{y}_i^{(t)}-\sum_{j\in Q_i}P_{ij}^{(t)}\boldsymbol{y}_j^{(t)}\right\|_2^2,\\&\text{s.t.}\ \boldsymbol{y}_i^{(t)\mathrm{T}}\boldsymbol{y}_i^{(t)}=1.\end{aligned} \tag{4}$$

2.2.3 在线哈希学习

为了充分保持数据在原始空间的相似关系，我们将 2.2.1 节和 2.2.2 节所学的鲁棒全局公共表示 $\boldsymbol{V}^{(t)}$ 和邻域结构保持的公共表示 $\boldsymbol{Y}^{(t)}$ 进行拼接，并通过正交映射 $\boldsymbol{R}$ 将其映射到海明空间中，用以指导哈希码 $\boldsymbol{B}^{(t)}$ 的学习. 同时，为了实现样本外扩展，我们学习映射 $\boldsymbol{W}_m$，将原始特征 $\boldsymbol{X}_m^{(t)}$ 映射到海明空间中. 为了保证哈希函数对全局样本的适应性，我们在新旧数据上联合优化哈希函数，并为映射矩阵 $\boldsymbol{W}_m$ 添加 L_2 范数正则化以防止过拟合，提高模型的泛化能力[26]. 这个过程可以表示为

$$\begin{aligned}&\min_{\boldsymbol{B}^{(t)},\boldsymbol{R},\boldsymbol{W}_m}\left\|\boldsymbol{B}^{(t)}-\boldsymbol{R}\left[\boldsymbol{V}^{(t)};\boldsymbol{Y}^{(t)}\right]\right\|_2+\sum_{m=1}^{M}\gamma\|\boldsymbol{W}_m\|_2+\\&\sum_{m=1}^{M}\beta_m\left\|\left[\boldsymbol{B}^{(t-1)},\boldsymbol{B}^{(t)}\right]-\boldsymbol{W}_m\left[\boldsymbol{X}_m^{(t-1)},\boldsymbol{X}_m^{(t)}\right]\right\|_2,\\&\text{s.t.}\ \boldsymbol{B}^{(t)}\in\{-1,+1\}^{n_t\times r},\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}=\boldsymbol{I},\end{aligned} \tag{5}$$

其中 β_m 和 γ 为超参数.

2.3 算法优化过程

在本节中，我们将介绍所提模型的优化过程. 该模型的优化分为 3 个部分：

1)鲁棒的全局结构保持模块中通过优化式(2)来学习转移矩阵 $\boldsymbol{U}_m$ 和鲁棒公共表示 $\boldsymbol{V}^{(t)}$.

2)融合的邻域结构保持模块中通过优化式(3)学习亲和矩阵 $\boldsymbol{P}^{(t)}$，通过优化式(4)学习邻域保持公共表示 $\boldsymbol{Y}^{(t)}$.

3)在线哈希学习模块中通过优化式(5)学习哈希码 $\boldsymbol{B}^{(t)}$，映射矩阵 $\boldsymbol{R}$ 和 $\boldsymbol{W}_m$.

2.3.1 优化全局结构保持模块

由于存在 $L_{2,1}$ 范数约束，式(2)难以直接优化，为此，我们通过拆解变量的方式，将式(2)改写为

$$\begin{aligned}&\min_{\boldsymbol{U}_m,\boldsymbol{V}^{(t)}}\sum_{m=1}^{M}\left\{\alpha_m\mathrm{tr}\left(\left(\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right)\cdot\boldsymbol{D}_m^{(t)}\left(\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right)^{\mathrm{T}}\right)+\gamma\|\boldsymbol{U}_m\|_2\right\}+\\&\sum_{m=1}^{M}\alpha_m\mathrm{tr}\left(\left(\boldsymbol{X}_m^{(t-1)}-\boldsymbol{U}_m\boldsymbol{V}^{(t-1)}\right)\cdot\boldsymbol{D}_m^{(t-1)}\left(\boldsymbol{X}_m^{(t-1)}-\boldsymbol{U}_m\boldsymbol{V}^{(t-1)}\right)^{\mathrm{T}}\right),\end{aligned} \tag{6}$$

其中 $\boldsymbol{D}_m$ 是对角矩阵，其第 i 行的对角元素表示为

$$\begin{aligned}\left(\boldsymbol{D}_m^{(t)}\right)_{i,i}&=1\Big/\left\|\left(\boldsymbol{x}_m^{(t)}\right)_i-\boldsymbol{U}_m\left(\boldsymbol{v}^{(t)}\right)_i\right\|_2,\\\left(\boldsymbol{D}_m^{(t-1)}\right)_{i,i}&=1\Big/\left\|\left(\boldsymbol{x}_m^{(t-1)}\right)_i-\boldsymbol{U}_m\left(\boldsymbol{v}^{(t-1)}\right)_i\right\|_2,\end{aligned} \tag{7}$$

其中 $\left(\boldsymbol{x}_m^{(t)}\right)_i$ 表示第 m 个模态中的第 i 个样本的特征. 此时，式(6)可以通过迭代优化算法求解，包括 U-Step 和 V-Step，具体步骤如下.

U-Step：固定变量 $\boldsymbol{V}^{(t)}$，优化模态转移矩阵 $\boldsymbol{U}_m$，那么式(6)可以被改写为

$$\begin{aligned}&\min_{\boldsymbol{U}_m}\alpha_m\mathrm{tr}\left(\left(\boldsymbol{X}_m^{(t-1)}-\boldsymbol{U}_m\boldsymbol{V}^{(t-1)}\right)\boldsymbol{D}_m^{(t-1)}\left(\boldsymbol{X}_m^{(t-1)}-\boldsymbol{U}_m\boldsymbol{V}^{(t-1)}\right)^{\mathrm{T}}\right)+\\&\alpha_m\mathrm{tr}\left(\left(\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right)\boldsymbol{D}_m^{(t)}\left(\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right)^{\mathrm{T}}\right)+\gamma\|\boldsymbol{U}_m\|_2.\end{aligned} \tag{8}$$

通过将 $\boldsymbol{U}_m$ 的导数设为 $\boldsymbol{0}$，$\boldsymbol{U}_m$ 的闭式解可表示为

$$\boldsymbol{U}_m=\boldsymbol{C}_1^{(t)}\left(\boldsymbol{C}_2^{(t)}+\frac{\gamma}{\alpha_m}\boldsymbol{I}\right)^{-1}, \tag{9}$$

$$\begin{aligned}\boldsymbol{C}_1^{(t)}&=\boldsymbol{X}_m^{(t)}\boldsymbol{D}_m^{(t)}\boldsymbol{V}^{(t)\mathrm{T}}+\boldsymbol{X}_m^{(t-1)}\boldsymbol{D}_m^{(t-1)}\boldsymbol{V}^{(t-1)\mathrm{T}}=\\&\quad\boldsymbol{X}_m^{(t)}\boldsymbol{D}_m^{(t)}\boldsymbol{V}^{(t)\mathrm{T}}+\boldsymbol{C}_1^{(t-1)},\\\boldsymbol{C}_2^{(t)}&=\boldsymbol{V}^{(t)}\boldsymbol{D}_m^{(t)}\boldsymbol{V}^{(t)\mathrm{T}}+\boldsymbol{V}^{(t-1)}\boldsymbol{D}_m^{(t-1)}\boldsymbol{V}^{(t-1)\mathrm{T}}=\\&\quad\boldsymbol{V}^{(t)}\boldsymbol{D}_m^{(t)}\boldsymbol{V}^{(t)\mathrm{T}}+\boldsymbol{C}_2^{(t-1)}.\end{aligned}\tag{10}$$

V-Step: 固定变量$\boldsymbol{U}_m$, 优化公共表示$\boldsymbol{V}^{(t)}$, 此时, 该优化问题可表示为

$$\min_{\boldsymbol{V}^{(t)}}\sum_{m=1}^{M}\alpha_m\mathrm{tr}\left(\left(\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right)\boldsymbol{D}_m^{(t)}\left(\boldsymbol{X}_m^{(t)}-\boldsymbol{U}_m\boldsymbol{V}^{(t)}\right)^{\mathrm{T}}\right).\tag{11}$$

对式(11)求导, 并令导数为$\boldsymbol{0}$, 可得$\boldsymbol{V}^{(t)}$的闭式解为

$$\boldsymbol{V}^{(t)}=\sum_{m=1}^{M}\alpha_m\left(\boldsymbol{U}_m^{\mathrm{T}}\boldsymbol{U}_m\right)^{-1}\boldsymbol{U}_m^{\mathrm{T}}\boldsymbol{X}_m^{(t)}.\tag{12}$$

2.3.2 优化邻域结构保持模块

该模块中变量$\boldsymbol{P}^{(t)}$和变量$\boldsymbol{Y}^{(t)}$可单独优化.

P-Step: 亲和矩阵$\boldsymbol{P}^{(t)}$可以利用拉格朗日乘子法求解. 此时式(3)可被重写为

$$\begin{aligned}&\min_{\boldsymbol{P}^{(t)}}\sum_{i=1}^{n_i}\boldsymbol{P}_i^{(t)\mathrm{T}}\boldsymbol{Z}_i\boldsymbol{P}_i^{(t)},\\&\text{s.t. }\sum_{j\in Q_i}\boldsymbol{P}_i^{(t)\mathrm{T}}\boldsymbol{1}=1,\end{aligned}\tag{13}$$

其中$\boldsymbol{Z}_i=\sum_{j\in Q_i}\left(\boldsymbol{x}_i^{(t)}-\boldsymbol{x}_j^{(t)}\right)\left(\boldsymbol{x}_i^{(t)}-\boldsymbol{x}_j^{(t)}\right)^{\mathrm{T}}$. 具体的优化过程可参考文献[27].

Y-Step: 为了获得局域结构保持的表示$\boldsymbol{Y}^{(t)}$, 我们设置$\boldsymbol{Y}^{(t)}=\boldsymbol{A}\boldsymbol{X}^{(t)}$, 并将式(4)改写为

$$\begin{aligned}&\min_{\boldsymbol{A}}\boldsymbol{A}\boldsymbol{X}^{(t)}\left(\boldsymbol{I}-\boldsymbol{P}^{(t)}\right)^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{P}^{(t)}\right)\boldsymbol{X}^{(t)\mathrm{T}}\boldsymbol{A}^{\mathrm{T}},\\&\text{s.t. }\boldsymbol{A}\boldsymbol{X}^{(t)}\boldsymbol{X}^{(t)\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}=\boldsymbol{I}.\end{aligned}\tag{14}$$

转移矩阵$\boldsymbol{A}$可通过以下广义特征向量问题[28]求解:

$$\boldsymbol{X}^{(t)}\left(\boldsymbol{I}-\boldsymbol{P}^{(t)}\right)^{\mathrm{T}}\left(\boldsymbol{I}-\boldsymbol{P}^{(t)}\right)\boldsymbol{X}^{(t)\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}=\lambda'\boldsymbol{X}^{(t)}\boldsymbol{X}^{(t)\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}.\tag{15}$$

2.3.3 优化在线哈希学习模块

式(5)的优化为非凸非连续问题, 很难直接求解, 因此我们采用迭代优化. 具体步骤如下.

W-Step: 固定变量$\boldsymbol{B}^{(t)}$和$\boldsymbol{R}$, 优化映射矩阵$\boldsymbol{W}_m$, 式(5)可以改写为

$$\min_{\boldsymbol{W}_m}\beta_m\left(\left\|\boldsymbol{B}^{(t-1)}-\boldsymbol{W}_m\boldsymbol{X}_m^{(t-1)}\right\|_2+\left\|\boldsymbol{B}^{(t)}-\boldsymbol{W}_m\boldsymbol{X}_m^{(t)}\right\|_2\right)+\gamma\|\boldsymbol{W}_m\|_2.\tag{16}$$

令其导数为$\boldsymbol{0}$, $\boldsymbol{W}_m$的闭式解可表示为

$$\boldsymbol{W}_m=\boldsymbol{H}_m^{(t)}\left(\boldsymbol{F}_m^{(t)}+\frac{\gamma}{\beta_m}\boldsymbol{I}\right)^{-1},\tag{17}$$

$$\begin{aligned}\boldsymbol{H}_m^{(t)}&=\boldsymbol{B}^{(t)}\boldsymbol{X}_m^{(t)\mathrm{T}}+\boldsymbol{B}^{(t-1)}\boldsymbol{X}_m^{(t-1)\mathrm{T}}=\boldsymbol{B}^{(t)}\boldsymbol{X}_m^{(t)\mathrm{T}}+\boldsymbol{H}_m^{(t-1)},\\\boldsymbol{F}_m^{(t)}&=\boldsymbol{X}_m^{(t)}\boldsymbol{X}_m^{(t)\mathrm{T}}+\boldsymbol{X}_m^{(t-1)}\boldsymbol{X}_m^{(t-1)\mathrm{T}}=\boldsymbol{X}_m^{(t)}\boldsymbol{X}_m^{(t)\mathrm{T}}+\boldsymbol{F}_m^{(t-1)}.\end{aligned}\tag{18}$$

R-Step: 固定变量$\boldsymbol{B}^{(t)}$和$\boldsymbol{W}_m$, 优化映射矩阵$\boldsymbol{R}$. 经过简单的变换, 式(5)可以改写为

$$\begin{aligned}&\max_{\boldsymbol{R}}\mathrm{tr}\left(\left[\boldsymbol{V}^{(t)};\boldsymbol{Y}^{(t)}\right]\boldsymbol{B}^{(t)\mathrm{T}}\boldsymbol{R}\right),\\&\text{s.t. }\boldsymbol{R}^{\mathrm{T}}\boldsymbol{R}=\boldsymbol{I}.\end{aligned}\tag{19}$$

为了优化式(19), 我们首先执行奇异值分解, 也就是$[\boldsymbol{V}^{(t)};\boldsymbol{Y}^{(t)}]\boldsymbol{B}^{(t)\mathrm{T}}=\boldsymbol{K}\boldsymbol{\tau}\boldsymbol{O}^{\mathrm{T}}$. $\boldsymbol{R}$的解可以表示为$\boldsymbol{R}=\boldsymbol{O}\hat{\boldsymbol{K}}^{\mathrm{T}}$, 其中$\hat{\boldsymbol{K}}$表示矩阵$\boldsymbol{K}$的前$r$个特征向量.

B-Step: 固定$\boldsymbol{W}_m$和$\boldsymbol{R}$, 优化哈希码$\boldsymbol{B}^{(t)}$, 此时

$$\begin{aligned}&\min_{\boldsymbol{B}^{(t)}}\left\|\boldsymbol{B}^{(t)}-\boldsymbol{R}\left[\boldsymbol{V}^{(t)};\boldsymbol{Y}^{(t)}\right]\right\|_2+\sum_{m=1}^{M}\beta_m\left\|\boldsymbol{B}^{(t)}-\boldsymbol{W}_m\boldsymbol{X}_m^{(t)}\right\|_2,\\&\text{s.t. }\boldsymbol{B}^{(t)}\in\{-1,+1\}^{n_i\times r}.\end{aligned}\tag{20}$$

通过公式变换可得

$$\boldsymbol{B}^{(t)}=\mathrm{sgn}\left(\boldsymbol{R}\left[\boldsymbol{V}^{(t)};\boldsymbol{Y}^{(t)}\right]+\sum_{m=1}^{M}\beta_m\boldsymbol{W}_m\boldsymbol{X}_m^{(t)}\right).\tag{21}$$

具体的优化过程见算法1、算法2.

算法1. 双重结构保持在线跨模态哈希算法.

输入: 流数据特征表示$\boldsymbol{X}_m$, 参数α_m, β_m, γ, 邻居节点数k, 哈希码长r;

输出: 哈希码$\boldsymbol{B}$, 映射矩阵$\boldsymbol{W}_m$.

① if t=1 then

② 初始化$\boldsymbol{B}^{(0)}$, $\boldsymbol{C}_1^{(0)}$, $\boldsymbol{C}_2^{(0)}$, $\boldsymbol{H}_m^{(0)}$, $\boldsymbol{F}_m^{(0)}$为$\boldsymbol{0}$;

③ end if

④ for t = 1: s do

⑤ 通过算法2学习$\boldsymbol{B}^{(t)}$, $\boldsymbol{C}_1^{(t)}$, $\boldsymbol{C}_2^{(t)}$, $\boldsymbol{H}_m^{(t)}$, $\boldsymbol{F}_m^{(t)}$;

⑥ 设$\boldsymbol{B}^{(t-1)}=[\boldsymbol{B}^{(t)},\boldsymbol{B}^{(t-1)}]$;

⑦ end for

算法2. 第t轮更新算法.

输入: $\boldsymbol{X}_m^{(t)}$, $\boldsymbol{C}_1^{(t-1)}$, $\boldsymbol{C}_2^{(t-1)}$, $\boldsymbol{H}_m^{(t-1)}$, $\boldsymbol{F}_m^{(t-1)}$, 哈希码长r, 参数α_m, β_m, γ, k, 迭代次数T;

输出: $\boldsymbol{B}^{(t)}$, $\boldsymbol{C}_1^{(t)}$, $\boldsymbol{C}_2^{(t)}$, $\boldsymbol{H}_m^{(t)}$, $\boldsymbol{F}_m^{(t)}$, $\boldsymbol{W}_m$.

① 随机初始化$\boldsymbol{V}^{(t)}$为零均值矩阵;

② 随机初始化$\boldsymbol{B}^{(t)}$为二值矩阵;

③ 初始化$\boldsymbol{D}_m^{(t)}$为单位阵;

④ 根据式(13)更新$\boldsymbol{P}^{(t)}$;

⑤ 根据式(15)更新$\boldsymbol{A}$, 计算$\boldsymbol{Y}^{(t)}=\boldsymbol{A}\boldsymbol{X}^{(t)}$;

⑥ for $i\leqslant T$ do

⑦ 根据式(9)更新$\boldsymbol{U}_m$;

⑧ 根据式(12)更新$\boldsymbol{V}^{(t)}$;

⑨ 根据式(7)更新$\boldsymbol{D}_m^{(t)}$;

⑩ end for

⑪ for $i\leqslant T$ do

⑫ 根据式(17)更新变量$\boldsymbol{W}_m$;

⑬ 根据式(19)更新变量$\boldsymbol{R}$;

⑭ 根据式(21)更新变量$\boldsymbol{B}^{(t)}$;

⑮ end for

2.4 时间复杂度分析

映射矩阵$\boldsymbol{U}_m$、$\boldsymbol{W}_m$和公共表示$\boldsymbol{V}^{(t)}$可以直接求导得到其解析解，其时间复杂度分为$O(d_m r n_t + r^2 n_t)$，$O(d_m r n_t + r^2 d_m)$，$O(d_m^2 n_t + d_m^2 r)$. 哈希码$\boldsymbol{B}^{(t)}$通过符号函数优化，其时间复杂度为$O(d_m r n_t + r^2 n_t)$. 正交映射$\boldsymbol{R}$需要通过奇异值分解优化，其时间复杂度为$O(d_m r n_t + r^2 n_t)$. 优化公共表示$\boldsymbol{Y}^{(t)}$的时间复杂度为$O(d_m r n_t + k^2 n_t)$. 由于样本数n_t远大于r和k，因此，本文所提出的SPOCH方法时间复杂度可简化为$T \times O(d_m^2 n_t + d_m r n_t + d_m^2 r)$，其中$T$表示迭代次数.

3 实验及分析

为了验证所提方法的有效性，我们在2个广泛使用的数据集上进行实验. 本节将从数据集、实验设置和实验结果分析3个方面具体介绍实验内容.

3.1 数据集

1)IAPRTC-12[29]. 该数据集包含20 000个图像文本对，分属255个类别. 我们使用512维GIST特征来描述图像模态样本，使用2 912维BOW特征来描述文本模态样本. 在实验中，我们随机选取2 000个图像文本对作为查询集，其余的作为训练集. 为了适应在线检索的场景，训练集被分成9个数据块，每个数据块包含2 000个样本对.

2)MIRFLICKR[30]. 该数据集由20 015个图像文本对组成. 每个样本对带有1个或多个标签，标签总共24类. 我们使用512维GIST特征和1 386维BOW特征分别描述该数据集中图像模态和文本模态. 不失一般性地，我们随机选取2 000个样本对作为查询集，其余的作为训练集. 为了适应在线场景，训练集被分成9个数据块，其中8个数据块包含2 000个样本对，1个数据块包含2 015个样本对.

3.2 实验设置

1)对比方法. 在实验中，我们与8种当前最好的无监督跨模态方法进行对比. 其中：CMFH[24], RFDH[31], JIMFH[32], HSSPH[33]为离线方法；OCMH[7], OCMFH[8], OUCDH[9], OMGHUN[34]为在线方法. 其中OCMH, OUCDH, OMGHUN是我们根据论文复现的，其余对比方法均采用论文作者公开的源代码. 离线方法在每个数据块上重新训练累积数据，而在线方法仅使用新数据进行训练.

2)参数设置. 为了公平，对于对比方法，我们均采用原文中建议的参数设置. 对于本文所提方法，我们利用交叉验证与网格搜索结合，对参数设置为：模态平衡参数$\alpha_1 = \alpha_2 = 10^4$, $\beta_1 = \beta_2 = 1$，正则化参数$\gamma = 10$，邻居节点个数$k = 50$. 此外，根据经验，我们设置迭代次数T=5. 所有的实验结果都在Intel® Core™ i9-10900K CPU@ 3.70GHz 64GB RAM的机器上运行得到.

3)评价指标. 为了验证跨模态检索的性能，在本实验中，我们实施2种检索任务：文本检索图像和图像检索文本. 我们采用2种最常见的检索评估指标，即平均精确率均值(mAP@all)和前K精度(Precision@K)，来评估模型检索准确性. 其中K=100. 此外，为了评估模型的训练效率，我们记录了所提的SPOCH方法及所有对比方法的运行时间.

3.3 实验结果及分析

在本节中，我们将从所提方法与对比方法的性能比较、消融实验、参数敏感性实验和训练时间4个方面展示实验结果并进行分析.

3.3.1 实验结果对比与分析

表1和表2展示了我们所提的SPOCH方法和对比方法在IAPRTC-12和MIRFLICKR这2个广泛使用的跨态检索数据集上的mAP@all实验结果. 为了保证实验结果的可靠性，我们报道5次运行的均值并提供了标准差. 从表1和表2可以看出，SPOCH在几乎所有码长上明显优于4种无监督在线跨模态方法OCMH, OCMFH, OUCDH, OMGHUN，尤其是在IAPRTC-12数据集上取得显著提升. 可能的原因是，该数据集包含的类别数较多，在不同数据块之间的类别不平衡现象更为明显. 同时，我们还将SPOCH与4个离线跨模态哈希方法进行了对比. 一般来说，由于进行了重复训练，离线方法的性能优于在线方法，一定程度上是以时间为代价提升了检索精度. 然而，从实验结果中可以看出，SPOCH几乎取得了与离线方法相当的性能，尤其是在IAPRTC-12数据集上，SPOCH的精度在一些码长上甚至优于离线方法.

进一步地，图2和图3中分别展示了mAP@all和Precision@100这2种评估指标在32 b时随数据块的变化. 可以看出，SPOCH在不同数据块之间性能较为平稳，表明所提方法对新数据的适应性良好，适合用于在线检索的场景. 并且，在不同数据块上，所提方法的mAP@all和Precision@100性能都显著优于4种在线无监督跨模态哈希，且与离线方法相比也取得了相当甚至更优的检索表现，充分验证了所提方法在在线检索场景下的有效性.

Table 1 Experimental Results Comparison of mAP@all and Standard Deviation on IAPRTC-12 Dataset

表 1 IAPRTC-12 数据集上 mAP@all 和标准差的实验结果对比

检索任务	对比方法	哈希码长				
		8b	16b	32b	64b	128b
图像检索文本	CMFH	0.314 8±0.002 8	0.315 8±0.003 6	0.307 7±0.002 2	0.311 5±0.002 0	0.315 1±0.002 8
	RFDH	0.347 4±0.003 0	0.351 6±0.003 1	0.353 5±0.005 9	0.360 3±0.007 2	0.369 0±0.003 7
	JIMFH	0.322 4±0.008 7	0.331 7±0.005 5	0.336 7±0.004 9	0.344 4±0.002 8	0.348 2±0.004 3
	HSSPH	0.351 4±0.007 8	0.352 1±0.006 7	0.354 2±0.010 1	0.364 5±0.007 2	0.368 8±0.008 8
	OCMH	0.303 9±0.005 0	0.302 5±0.002 7	0.301 2±0.001 8	0.301 9±0.001 7	0.301 7±0.001 8
	OCMFH	0.300 7±0.003 9	0.300 6±0.003 2	0.298 9±0.002 6	0.298 3±0.003 3	0.298 4±0.004 0
	OUCDH	0.303 0±0.004 0	0.301 0±0.003 2	0.300 5±0.003 1	0.302 1±0.005 0	0.306 6±0.003 4
	OMGHUN	0.302 3±0.006 6	0.302 2±0.006 6	0.302 5±0.006 4	0.302 3±0.006 5	0.302 4±0.006 3
	SPOCH(本文)	**0.352 6±0.009 7**	**0.356 8±0.014 0**	**0.361 9±0.010 6**	**0.378 5±0.006 0**	**0.378 8±0.006 0**
文本检索图像	CMFH	0.376 0±0.004 5	0.388 2±0.004 3	0.402 5±0.005 3	0.411 8±0.004 1	0.417 6±0.004 3
	RFDH	0.347 7±0.002 1	0.356 0±0.002 4	0.364 6±0.003 6	0.373 1±0.004 2	0.383 8±0.003 4
	JIMFH	0.360 2±0.005 7	0.369 1±0.001 7	0.377 9±0.002 5	0.393 9±0.003 2	0.406 2±0.003 6
	HSSPH	0.357 4±0.006 8	0.356 8±0.005 8	0.361 5±0.009 9	0.360 3±0.007 2	0.369 0±0.003 7
	OCMH	0.301 0±0.010 8	0.303 1±0.006 1	0.302 9±0.003 3	0.300 7±0.001 8	0.303 7±0.003 8
	OCMFH	0.337 7±0.004 2	0.336 9±0.003 6	0.333 1±0.003 2	0.330 3±0.002 0	0.327 0±0.002 9
	OUCDH	0.302 7±0.004 0	0.302 9±0.003 4	0.300 5±0.003 2	0.302 1±0.004 7	0.302 7±0.003 9
	OMGHUN	0.302 3±0.006 7	0.302 2±0.006 0	0.302 4±0.006 2	0.302 3±0.006 5	0.302 6±0.006 0
	SPOCH(本文)	**0.360 9±0.009 9**	**0.365 0±0.015 1**	**0.376 2±0.013 0**	**0.397 8±0.007 3**	**0.402 7±0.007 4**

注：最优的在线检索结果用加粗数值表示.

Table 2 Experimental Results Comparison of mAP@all and Standard Deviation on MIRFLICKR Dataset

表 2 MIRFLICKR 数据集上 mAP@all 和标准差的实验结果对比

检索任务	对比方法	哈希码长				
		8b	16b	32b	64b	128b
图像检索文本	CMFH	0.578 8±0.003 5	0.575 1±0.006 5	0.574 2±0.006 2	0.577 7±0.003 5	0.577 3±0.005 0
	RFDH	0.559 2±0.002 0	0.580 0±0.003 3	0.582 1±0.002 5	0.584 0±0.001 6	0.584 3±0.001 9
	JIMFH	0.582 0±0.007 4	0.587 1±0.007 1	0.588 2±0.007 0	0.587 2±0.006 0	0.593 7±0.006 9
	HSSPH	0.591 2±0.006 7	0.596 1±0.008 3	0.593 1±0.006 8	0.593 8±0.009 9	0.596 9±0.003 1
	OCMH	0.555 3±0.001 7	0.556 5±0.001 5	0.555 1±0.002 4	0.555 7±0.001 6	0.555 6±0.001 7
	OCMFH	**0.564 1±0.004 5**	0.561 8±0.003 9	0.560 5±0.003 7	0.559 0±0.003 3	0.558 6±0.003 4
	OUCDH	0.552 2±0.002 1	0.553 0±0.001 7	0.551 5±0.002 2	0.551 6±0.002 6	0.552 0±0.002 5
	OMGHUN	0.553 6±0.002 6	0.553 7±0.001 5	0.552 8±0.001 1	0.553 0±0.002 2	0.553 6±0.002 3
	SPOCH(本文)	0.563 8±0.004 6	**0.571 9±0.008 4**	**0.577 9±0.009 7**	**0.587 7±0.009 4**	**0.593 1±0.002 1**
文本检索图像	CMFH	0.593 1±0.002 0	0.593 3±0.003 5	0.590 6±0.003 6	0.588 4±0.004 3	0.584 7±0.003 8
	RFDH	0.559 1±0.002 4	0.579 4±0.003 5	0.584 7±0.004 6	0.586 1±0.003 3	0.585 6±0.002 4
	JIMFH	0.600 2±0.006 6	0.602 8±0.007 4	0.602 0±0.005 5	0.601 2±0.003 4	0.606 7±0.004 1
	HSSPH	0.591 3±0.006 7	0.602 8±0.006 0	0.601 4±0.006 2	0.603 4±0.006 5	0.601 5±0.006 0
	OCMH	0.552 3±0.003 0	0.555 2±0.002 4	0.556 1±0.003 9	0.555 5±0.001 7	0.555 5±0.001 8
	OCMFH	0.560 4±0.003 4	0.559 8±0.003 0	0.559 4±0.003 5	0.559 0±0.003 5	0.559 0±0.003 2
	OUCDH	0.552 0±0.002 0	0.552 8±0.001 6	0.551 7±0.002 1	0.551 7±0.002 4	0.552 0±0.002 4
	OMGHUN	0.554 4±0.001 9	0.553 7±0.001 8	0.553 4±0.001 4	0.553 4±0.002 0	0.553 6±0.002 3
	SPOCH(本文)	**0.563 7±0.003 1**	**0.572 3±0.008 6**	**0.580 4±0.009 9**	**0.590 0±0.009 5**	**0.597 6±0.002 9**

注：最优的在线检索结果用加粗数值表示.

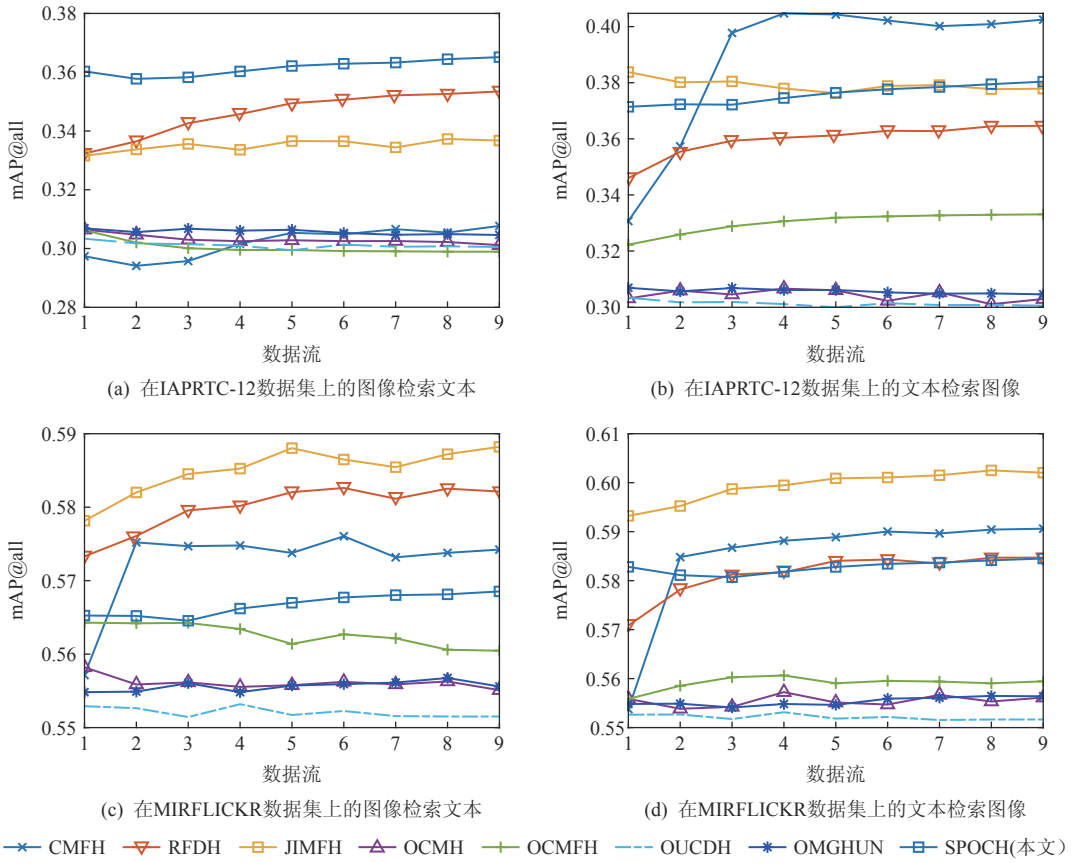


Fig. 2 mAP@all varies with chunks

图 2 mAP@all 随数据块的变化

3.3.2 消融实验

根据模型的设计方案，我们从 4 个角度验证所提方法的有效性. 表 3 和表 4 展示了消融实验的结果.

1）SPOCH-f. 为了验证结构化稀疏是否可以缓解离群样本的影响，我们采用 L_2 范数而非 $L_{2,1}$ 范数来学习公共表示 $\boldsymbol{V}$. 实验结果表明，SPOCH-f 在 2 个数据集上都较 SPOCH 有所下降，验证了结构化稀疏可以提高模型鲁棒性. 特别地，SPOCH-f 在 IAPRTC-12 数据集上下降尤为明显，可能的原因是该数据集的类别不平衡现象更为显著.

2）SPOCH-v. 为了验证全局结构信息对于哈希码学习的指导作用，我们在 SPOCH-v 的实验中去掉了鲁棒全局结构保持模块. 从表 3 和表 4 可以看出，该模块对于性能的提升十分重要，尤其是在短码的情况下，去掉公共表示的指导会导致性能显著下降.

3）SPOCH-y. 该模块用于验证邻域结构保持对于检索性能的影响. 为此，我们去掉了对邻域结构信息的学习，仅采用全局公共表示 $\boldsymbol{V}$ 来指导哈希码的生成. 从表 3 和表 4 可见，在短码的情况下，SPOCH-y 的性能较 SPOCH 有显著下降. 结合 SPOCH-v 的结果可知，仅通过单一的指导信息无法习得高质量的短码，由此验证了本文所提的双重结构保持方案的有效性.

4）SPOCH-w. 该模块用于验证新旧数据联合优化对于检索性能的影响. 在 SPOCH-w 的实验中，我们去掉了转移矩阵 $\boldsymbol{U}_m$ 和映射矩阵 $\boldsymbol{W}_m$ 在新旧数据上的联合优化. SPOCH-w 在 2 个数据集各个位上都呈现明显的性能下降，可见联合学习及相应的优化策略可以缓解流数据场景下的灾难性遗忘问题.

3.3.3 参数实验

为了进一步验证模型的稳定性，我们对模型中的平衡参数 α_m 和 β_m、邻居节点个数 k 和邻域公共表示长度 c，分别做敏感性测试. 图 4 和图 5 展示了平衡参数 α_m 和 β_m 随着参数取值的变化时对应的 mAP@all 性能的变化. 从实验结果来看，平衡参数的选择对于检索性能的影响非常微弱. 由此可见，所提模型对这 2 个参数不敏感.

图 6 展示了邻居节点个数 k 和邻域结构保持的公共表示的长度 c 在参数变化时对应的 mAP@all 性能的变化. 从实验结果来看，邻域范围的选择及公共表示的长度对于检索性能的影响并不明显，即模型

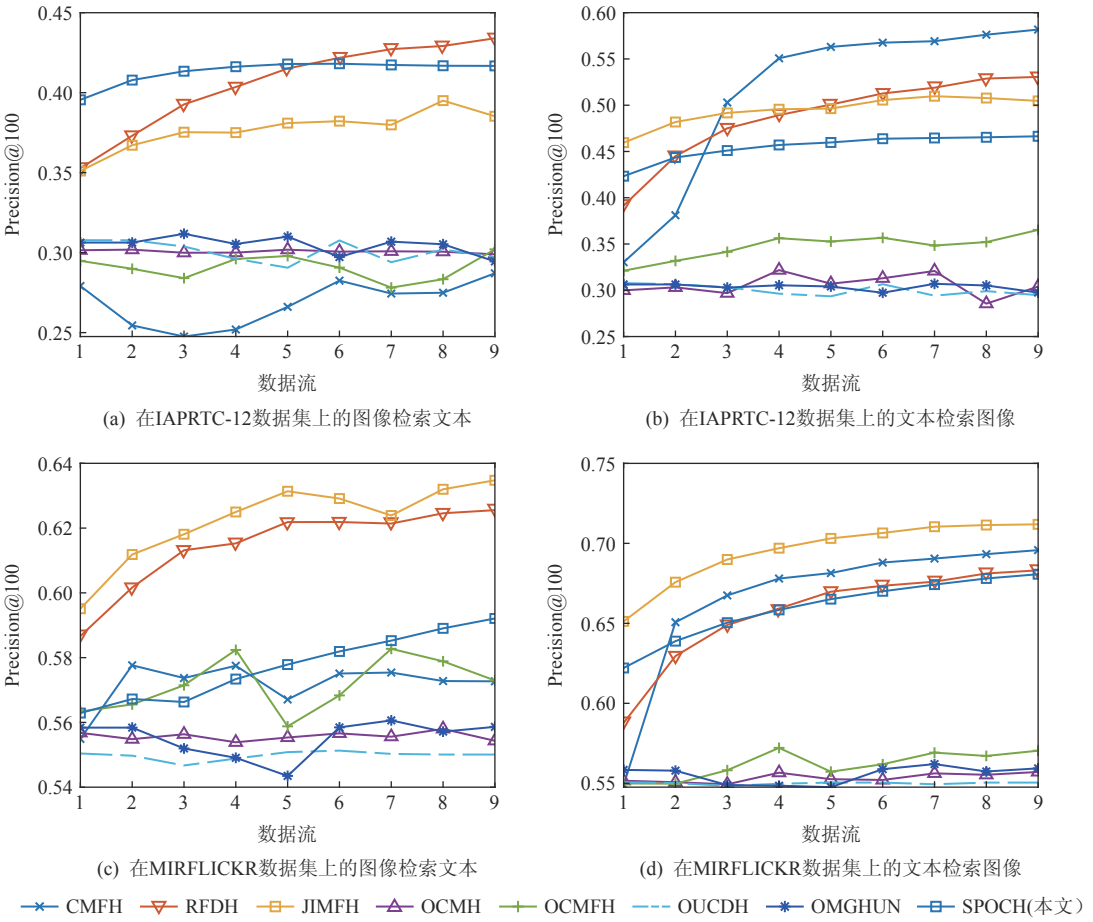


Fig. 3 Precision@100 varies with chunks

图 3 Precision@100 随数据块的变化

Table 3 Ablation Study Results (mAP@100 and Standard Deviation) on IAPRTC-12 Dataset

表 3 IAPRTC-12 数据集上模型消融实验 (mAP@100 和标准差) 结果

检索任务	对比方法	哈希码长				
		8b	16b	32b	64b	128b
图像检索文本	SPOCH-f	0.346 3±0.009 0	0.350 5±0.008 4	0.345 1±0.015 9	0.367 5±0.010 3	0.374 2±0.007 0
	SPOCH-v	0.314 5±0.021 0	0.349 4±0.014 7	0.377 2±0.003 1	0.370 2±0.003 9	0.380 5±0.004 2
	SPOCH-y	0.328 6±0.015 9	0.320 1±0.021 2	0.371 0±0.006 5	0.377 6±0.003 8	0.371 4±0.005 9
	SPOCH-w	0.307 9±0.003 9	0.308 6±0.000 7	0.305 9±0.003 9	0.311 2±0.003 4	0.310 1±0.004 9
	SPOCH(本文)	0.352 6±0.009 7	0.356 8±0.014 0	0.361 9±0.010 6	0.378 5±0.006 0	0.378 8±0.006 0
文本检索图像	SPOCH-f	0.350 3±0.011 7	0.360 2±0.010 1	0.362 7±0.015 5	0.387 0±0.010 7	0.398 4±0.008 3
	SPOCH-v	0.326 8±0.018 1	0.359 5±0.013 7	0.392 2±0.003 6	0.388 3±0.006 1	0.408 2±0.004 4
	SPOCH-y	0.337 4±0.017 3	0.326 6±0.020 8	0.384 5±0.008 9	0.395 5±0.006 0	0.394 9±0.006 2
	SPOCH-w	0.305 8±0.001 6	0.307 5±0.001 6	0.307 7±0.003 1	0.314 9±0.003 6	0.315 7±0.005 2
	SPOCH(本文)	0.360 9±0.009 9	0.365 0±0.015 1	0.376 2±0.013 0	0.397 8±0.007 3	0.402 7±0.007 4

对于这 2 个参数具有良好的稳定性.

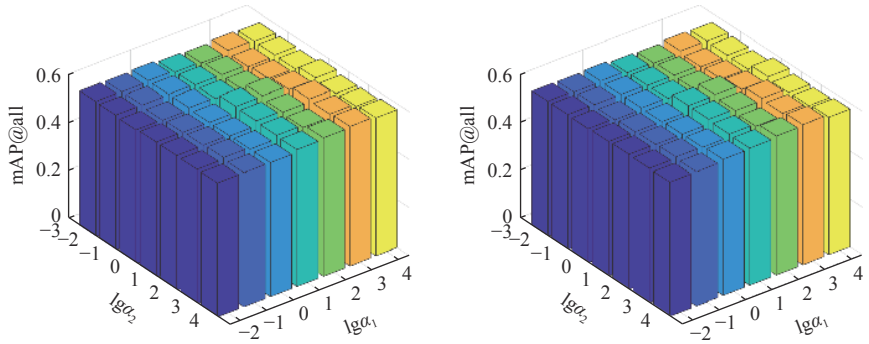
3.3.4 训练时间分析

图 7 记录了 2 个数据集上 SPOCH 和对比方法的训练时间, 哈希码长设置为 32 b. 由于离线方法需要在累计数据集上重复训练, 因此离线方法的训练时间随数据流显著增加, 此时检索效率会受到很大影响. 而在线方法的训练时间与累积数据无关, 因此更加适用于流数据的检索. 从实验结果可以看出, 随着数据量的累积, SPOCH 的效率逐渐优于离线方法. 并且, SPOCH 在检索精度显著优于现有方法的前提下

Table 4 Ablation Study Results (mAP@100 and Standard Deviation) on MIRFLICKR Dataset

表 4 MIRFLICKR 数据集上模型消融实验 (mAP@100 和标准差) 结果

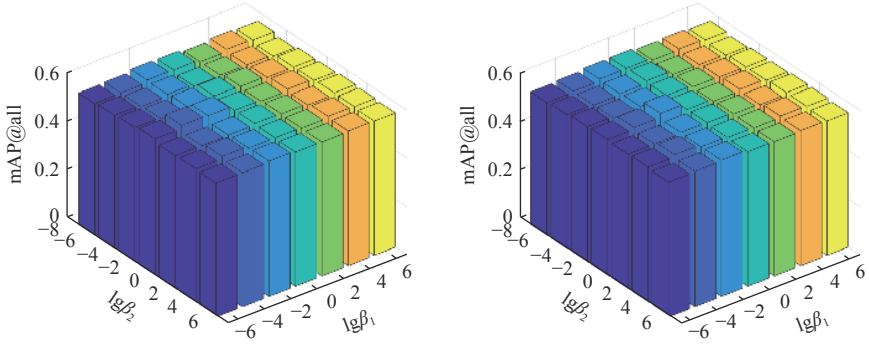
检索任务	对比方法	哈希码长				
		8b	16b	32b	64b	128b
图像检索文本	SPOCH-f	0.576 3±0.008 6	0.557 5±0.018 3	0.576 3±0.007 5	0.577 6±0.003 3	0.586 1±0.001 7
	SPOCH-v	0.558 4±0.012 9	0.567 6±0.018 8	0.578 7±0.005 6	0.588 9±0.003 0	0.587 2±0.008 4
	SPOCH-y	0.562 8±0.018 0	0.589 5±0.014 8	0.576 5±0.006 4	0.578 3±0.003 5	0.586 1±0.004 4
	SPOCH-w	0.551 6±0.003 2	0.552 1±0.003 8	0.558 1±0.002 6	0.554 7±0.002 9	0.556 0±0.002 8
	SPOCH(本文)	0.563 8±0.004 6	0.571 9±0.008 4	0.577 9±0.009 7	0.587 7±0.009 4	0.593 1±0.002 1
文本检索图像	SPOCH-f	0.577 7±0.008 0	0.557 9±0.016 8	0.577 2±0.007 9	0.582 0±0.002 8	0.592 9±0.000 4
	SPOCH-v	0.559 6±0.011 7	0.568 7±0.017 9	0.580 7±0.005 7	0.590 9±0.004 3	0.593 0±0.008 8
	SPOCH-y	0.564 4±0.016 7	0.589 7±0.013 3	0.579 9±0.006 9	0.581 5±0.003 4	0.594 5±0.004 9
	SPOCH-w	0.553 0±0.002 3	0.552 9±0.003 6	0.556 9±0.001 8	0.556 0±0.002 5	0.556 1±0.002 6
	SPOCH(本文)	0.563 7±0.003 1	0.572 3±0.008 6	0.580 4±0.009 9	0.590 0±0.009 5	0.597 6±0.002 9



(a) 图像检索文本 (b) 文本检索图像

Fig. 4 Parameter sensitivity analysis result of α_m on MIRFLICKR dataset

图 4 MIRFLICKR 数据集上α_m参数的敏感性分析结果



(a) 图像检索文本 (b) 文本检索图像

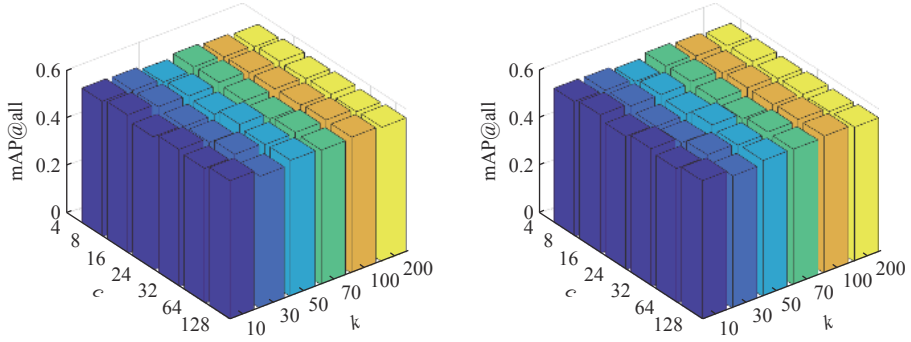
Fig. 5 Parameter sensitivity analysis result of β_m on MIRFLICKR dataset

图 5 MIRFLICKR 数据集上β_m参数的敏感性分析结果

取得了可比较甚至更优的训练效率.

4 结 论

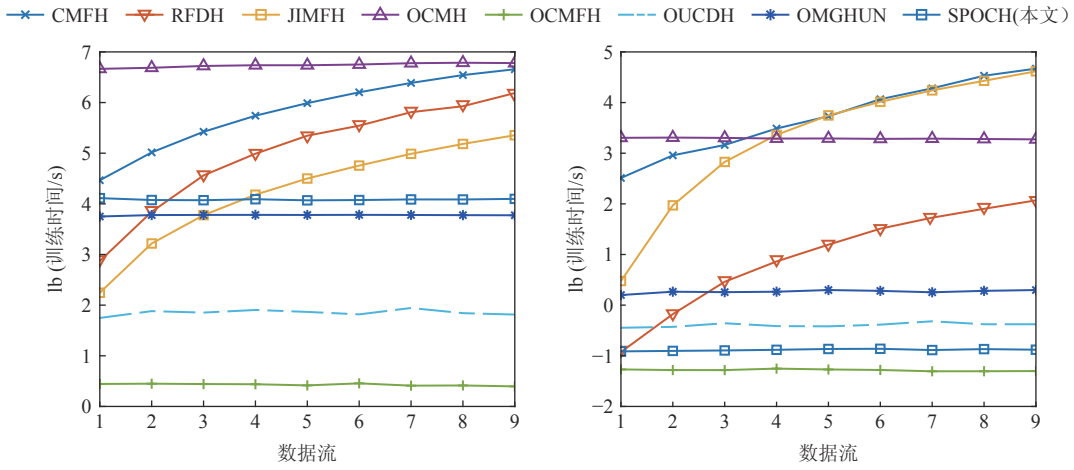
本文提出基于双重结构保持的在线跨模态哈希方法. 该模型可以在不依赖监督信息的情况下为多模态流数据学习对应的哈希函数和哈希码, 用于后续的在线跨模态检索任务. 在该模型中, 我们同时挖掘流数据中的全局结构信息和邻域结构信息, 使得所学的公共表示可以更好地表征多模态数据, 从而促进高质量哈希码的生成. 并且, 我们基于$L_{2,1}$范数设计损失函数, 利用$L_{2,1}$范数结构化稀疏的性质缓解



(a) 图像检索文本 (b) 文本检索图像

Fig. 6 Parameter sensitivity analysis result of k and c on MIRFLICKR dataset

图 6 MIRFLICKR 数据集上 k 和 c 参数的敏感性分析结果



(a) IAPRTC-12 (b) MIRFLICKR

Fig. 7 Training time varies with chunks on two datasets

图 7 在 2 个数据集上训练时间随数据块的变化

了离群样本对模型的影响, 提高了模型鲁棒性. 此外, 我们设计了相应的在线优化策略, 使得算法的时间复杂度线性于新数据的规模, 提高了检索效率.

作者贡献声明: 康潇提出算法思路, 完成实验并撰写论文; 刘兴波提出指导意见并修改论文; 卢鹏宇、赵志杰和王少华协助完成实验和撰写论文; 聂秀山和尹义龙提出指导意见和审核论文.

参 考 文 献

[1] Xu Xing, Shen Fumin, Yang Yang, et al. Learning discriminative binary codes for large-scale cross-modal retrieval[J]. IEEE Transactions on Image Processing, 2017, 26(5): 2494−2507

[2] Raginsky M, Lazebnik S. Locality-sensitive binary codes from shift-invariant kernels[C]//Proc of the 22nd Int Conf on Neural Information Processing Systems. New York: ACM, 2009, 1509−1517

[3] Liu Xingbo, Nie Xiushan, Yin Yilong. Supervised discrete cross-modal hashing method based on two-way linear regression[J]. Journal of Computer Research and Development, 2020, 57(8): 1707−1714 (in Chinese)
(刘兴波, 聂秀山, 尹义龙. 基于双向线性回归的监督离散跨模态散列方法[J]. 计算机研究与发展, 2020, 57(8): 1707−1714)

[4] Zheng Chaoqun, Zhu Lei, Zhang Zheng, et al. Efficient semi-supervised multimodal hashing with importance differentiation regression[J]. IEEE Transactions on Image Processing, 2022, 31: 5881−5892

[5] Zhuo Junbao, Su Chi, Wang Shuhui, et al. Least entropy transfer against hashing method[J]. Journal of Computer Research and Development, 2020, 57(4): 888−896 (in Chinese)
(卓君宝, 苏驰, 王树徽, 等. 最小熵迁移对抗散列方法[J]. 计算机研究与发展, 2020, 57(4): 888−896)

[6] Lin Zijia, Ding Guiguang, Hu Mingqing, et al. Semantics-preserving hashing for cross-view retrieval[C]// Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2015: 3864−3872

[7] Xie Liang, Shen Jialie, Zhu Lei, et al. Online cross-modal hashing for web image retrieval[C]// Proc of the 30th AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2016: 294−300

[8] Wang Di, Wang Quan, An Yaqiang, et al. Online collective matrix factorization hashing for large-scale cross-media retrieval[C]// Proc of the 43rd Int ACM SIGIR Conf on Research and Development in Information Retrieval. New York: ACM, 2020: 1409−1418

[9] Li Xuan, Wu Wei, Yuan Yunhao, et al. Online unsupervised cross-view discrete hashing for large-scale retrieval[J]. Applied Intelligence, 2022, 52(13): 14905−14917

[10] Yao Tao, Wang Gang, Yan Lianshan, et al. Online latent semantic hashing for cross-media retrieval[J]. Pattern Recognition, 2019, 89: 1−11

[11] Wang Yongxin, Luo Xin, Xu Xinshun, et al. Label embedding online hashing for cross-modal retrieval[C]// Proc of the 28th ACM Int Conf on Multimedia. New York: ACM, 2020: 871−879

[12] Zhan Yuwei, Wang Yongxin, Sun Yu, et al. Discrete online cross-modal hashing[J]. Pattern Recognition, 2022, 122: 108262

[13] Lin Mingbao, Ji Rongrong, Liu Hong, et al. Towards optimal discrete online hashing with balanced similarity[C]//Proc of the 33rd AAAI Conf on Artificial Intelligence. Palo Alto, CA: AAAI, 2019: 8722−8729

[14] Cakir F, Sclaroff S. Online supervised hashing[C]// Proc of the IEEE Int Conf on Image Processing. Piscataway, NJ: IEEE, 2015: 2606−2610

[15] Lin Mingbao, Ji Rongrong, Liu Hong, et al. Hadamard matrix guided online hashing[J]. International Journal of Computer Vision, 2020, 128(8): 2279−2306

[16] Lin Mingbao, Ji Rongrong, Chen Shen, et al. Similarity-preserving linkage hashing for online image retrieval[J]. IEEE Transactions on Image Processing, 2020, 29: 5289−5300

[17] Lin Mingbao, Ji Rongrong, Sun Xiaoshuai, et al. Fast class-wise updating for online hashing[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2022, 44(5): 2453−2467

[18] Fang Yuzhi, Liu Li. Scalable supervised online hashing for image retrieval[J]. Journal of Computational Design and Engineering, 2021, 8(5): 1391−1406

[19] Hao Yanbin, Mu Tingting, Goulermas J, et al. Unsupervised t-distributed video hashing and its deep hashing extension[J]. IEEE Transactions on Image Processing, 2017, 26(11): 5531−5544

[20] Huang Longkai, Yang Qiang, Zheng Weishi. Online hashing[J]. IEEE Transactions on Neural Networks and Learning Systems, 2018, 29(6): 2309−2322

[21] Leng Cong, Wu Jiaxiang, Cheng Jian, et al. Online sketching hashing[C]// Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Los Alamitos, CA: IEEE Computer Society, 2015: 2503−2511

[22] Chen Xixian, King I, Lyu M. FROSH: Faster online sketching hashing[C/OL]// Proc of the 30th Conf on Uncertainty in Artificial Intelligence. Corvallis: AUAI, 2017[2023-11-21]. http://auai.org/uai2017/proceedings/papers/12.pdf

[23] Chen Xixian, Yang Haiqin, Zhao Shenglin, et al. Making online sketching hashing even faster[J]. IEEE Transactions on Knowledge and Data Engineering, 2021, 33(3): 1089−1101

[24] Ding Guiguang, Guo Yuchen, Zhou Jile. Collective matrix factorization hashing for multimodal data[C]//Proc of the IEEE Conf on Computer Vision and Pattern Recognition. Piscataway, NJ: IEEE, 2014: 2083−2090

[25] Sun Yuan, Wang Xu , Peng Dezhong, et al. Hierarchical hashing learning for image set classification[J]. IEEE Transactions on Image Processing, 2023, 32: 1732−1744

[26] Hoerl A, Kennard R. Ridge regression: Applications to nonorthogonal problems[J]. Technometrics, 1970, 12(1): 69−82

[27] Roweis T, Saul K. Nonlinear dimensionality reduction by locally linear embedding[J]. Science, 2000, 290(5500): 2323−2326

[28] He Xiaofei, Cai Deng, Yan Shuicheng, et al. Neighborhood preserving embedding[C]// Proc of the 10th IEEE Int Conf on Computer Vision. Piscataway, NJ: IEEE, 2005: 1208−1213

[29] Escalante H, Hernandez C, Gonzalez J, et al. The segmented and annotated IAPRTC-12 benchmark[J]. Computer Vision and Image Understanding, 2010, 144(4): 419−428

[30] Huiskes M , Lew M. The MIR flickr retrieval evaluation[C]// Proc of the 1st ACM SIGMM Int Conf on Multimedia Information Retrieval. New York: ACM, 2008: 9−43

[31] Wang Di, Wang Quan, Gao Xinbo. Robust and flexible discrete hashing for cross-modal similarity search[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2018, 28(10): 2703−2715

[32] Wang Di, Wang Quan, He Lihuo, et al. Joint and individual matrix factorization hashing for large-scale cross-modal retrieval[J]. Pattern Recognition, 2020, 107: 107479

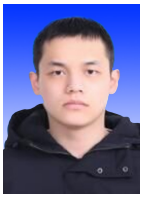
[33] Wang, Di, Caiping Zhang, Quan Wang, et al. Hierarchical semantic structure preserving hashing for cross-modal retrieval[J]. IEEE Transactions on Multimedia, 2023, 25: 1217−1229

[34] Liu Xin, Yi Jinhan, Cheung Yiuming, et al. OMGHUN: Online manifold-guided hashing for flexible cross-modal retrieval[J]. IEEE Transactions on Multimedia, 2023, 25: 3811−3824



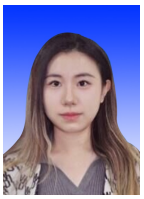
Kang Xiao, born in 1997. PhD. Her main research interests include multimedia retrieval and computer vision.

康　潇, 1997 年生. 博士. 主要研究方向为多媒体检索、计算机视觉.



Liu Xingbo, born in 1995. PhD, professor. His main research interests include multimedia retrieval and computer vision.

刘兴波, 1995 年生. 博士, 教授. 主要研究方向为多媒体检索、计算机视觉.



Lu Pengyu, born in 1998. PhD candidate. Her main research interests include machine learning and artificial neural network. (periispengyu@163.com)

卢鹏宇, 1998 年生. 博士研究生. 主要研究方向为机器学习、人工神经网络.



Zhao Zhijie, born in 1999. Master candidate. Her main research interests include multimedia retrieval and computer vision. (zhaozhijie_01@163.com)

赵志杰, 1999 年生. 硕士研究生. 主要研究方向为多媒体检索、计算机视觉.



Nie Xiushan, born in 1981. PhD, professor. His main research interests include multimedia retrieval, multimedia security, and computer vision. (niexiushan@163.com)

聂秀山, 1981 年生. 博士，教授. 主要研究方向为多媒体检索、多媒体安全、计算机视觉.



Wang Shaohua, born in 1979. PhD, senior engineer. Member of CCF. Her main research interests include multimedia retrieval and computer vision. (wsh@sdjzu.edu.cn)

王少华, 1979 年生. 博士，高级工程师. CCF 会员. 主要研究方向为多媒体检索、计算机视觉.



Yin Yilong, born in 1972. PhD, professor, PhD supervisor. His main research interests include machine learning and data mining. (ylyin@sdu.edu.cn)

尹义龙, 1972 年生. 博士，教授，博士生导师. 主要研究方向为机器学习、数据挖掘.