用于多媒体检索的深度度量多视图哈希

Jian Zhu 浙江实验室 中国浙江省杭州市
qijian.zhu@zhejianglab.com

Xiaohu Ruan vivo AI 实验室 中国浙江省杭州市
xiaohu.ruan@vivo.com

程永立 福州大学 中国福建省福州市
chengyongli@fzu.edu.cn

Zhangmin Huang 浙江实验室 中国浙江省杭州市
zmhuang@zhejianglab.com

崔宇 浙江实验室 中国浙江省杭州市
cui.yu@zhejianglab.com

Lingfang Zeng*浙江实验室 中国浙江省杭州市
zenglf@zhejianglab.com

arXiv:2304.06358v1 [cs.CV] 13 Apr 2023

摘要—学习多视图异构数据的哈希表示是多媒体检索中的一项重要任务。然而，现有方法无法有效融合多视图特征，也未能利用不相似样本提供的度量信息，导致检索精确率有限。当前方法采用加权和或拼接的方式融合多视图特征。我们认为，这些融合方法无法捕捉不同视图之间的交互。此外，这些方法忽略了不相似样本提供的信息。为解决上述问题，我们提出了一种新颖的深度度量多视图哈希（DMMVH）方法。大量实验证据表明，基于门控的融合优于典型方法。我们将深度度量学习引入多视图哈希问题，从而能够利用不相似样本的度量信息。在 MIR-Flickr25K、MS COCO 和 NUS-WIDE 数据集上，我们的方法以较大优势超越了当前最先进的方法（平均精度均值（mAP）最高提升15.28 个百分点）。

索引词—**多视图哈希,多模态哈希,深度度量学习, 多媒体检索**

I. 引言

多视图哈希用于解决多媒体检索问题。设计良好的多视图哈希算法可以显著提高多媒体检索任务的精确率。与仅以单视图方式进行搜索的单视图哈希不同，多视图哈希可以利用来自不同来源的数据（例如图像、文本、音频和视频）。多视图哈希表示学习首先从不同视图中提取异构特征，然后融合多视图特征，以捕捉不同视图的全局表示。

当前的多视图哈希算法存在检索精确率较低的问题。这主要由以下两个因素造成。
方面。首先，当前的多视图哈希算法对多视图特征的融合并不充分。为了获得全局表示，典型的多视图哈希方法（例如，深度协同多视图哈希）（DCMVH） [1]，灵活多模态哈希（FMH） [2]）利用加权和或

*曾玲芳是通讯作者。

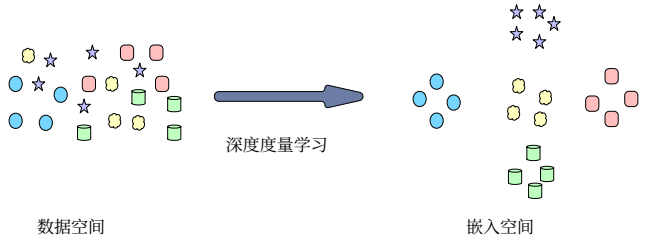

图1：深度度量学习示意图。输入在数据空间中随机分布。深度度量学习将输入投影到嵌入空间中，其中嵌入表示根据其语义含义进行分配。

通过拼接来融合多视图特征。在融合过程中忽略了文本与图像之间的关系，这导致所获得的全局表示的表达能力较弱。其次，当前方法受限于相似样本所提供的信息。衡量不相似样本之间距离的重要性被低估了。例如，灵活图卷积多模态哈希（FGCMH） [3] 是一种基于 GCN的 [4]多视图哈希方法，该方法基于相似性构建图的边，并聚合相邻节点的特征。因此，不相似样本在这一过程中不起作用。

我们提出了一种深度度量多视图哈希方法，称为DMMVH。它利用上下文门控 [5]来学习图像和文本特征之间的交互与依赖关系。与典型方法不同，DMMVH将多视图特征融合为全局表示，同时不会丧失对这些特征的依赖。此外，DMMVH中引入了深度度量学习。如图1所示，最初，样本在原始数据空间中随机分布。通过深度度量学习，语义相似的样本彼此靠近，而不相似样本则被推远。为了利用不相似样本的距离信息，我们设计了一种深度度量损失函数。此外，我们引入了一

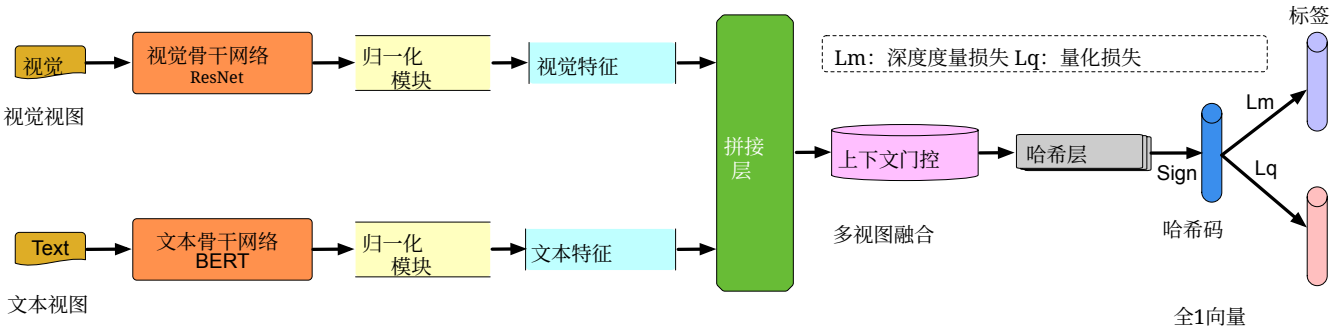

图 2：DMMVH 方法的流程图。图像和文本特征分别由 ResNet 和 BERT 提取。归一化模块对特征进行归一化，并将其连接在一起。多视图融合模块对连接后的特征执行上下文门控，在保留依赖关系的同时融合多视图特征。最后，哈希层根据融合表示生成哈希码。

用于降低所设计损失函数复杂度的超参数。通过深度度量学习获得最优嵌入空间，该学习遵循哈希表示学习的语义保持原则。

我们在多视图哈希表示学习基准上，使用 MIR-Flickr25K、MS COCO和 NUS-WIDE 数据集评估了我们的方法。所提出的方法在基准测试中带来了最高达 15。28% 的平均精度均值提升。

我们的主要贡献如下：

•我们提出了一种新颖的多视图哈希方法，在多媒体检索中取得了最先进的结果。

• 我们利用上下文门控学习不同视图更优的全局表示，以解决融合不充分的问题。

• 首次将深度度量学习引入多视图哈希。我们设计并优化了一种具有线性复杂度的深度度量损失。

II. 提出的方法

DMMVH旨在利用新设计的深度度量损失训练深度多视图哈希网络。我们首先介绍深度多视图哈希网络，该网络将多视图特征深度融合为全局表示。然后阐述新的深度度量损失。最后，引入超参数 λ 以降低复杂度。

A. 深度多视图哈希网络

深度多视图哈希网络旨在将多视图数据转换为哈希码。如图 2 所示，DMMVH 由视觉骨干网络、文本骨干网络、归一化模块、多视图融合模块和哈希层组成。下面将详细介绍这些模块。

1)视觉骨干网络： 采用 Deep ResNet [6] 来生成视觉特征。2) **文本骨干网络：** 采用 BERT-base [7] 来提取文本特征。3) **归一化模块：** 归一化模块将多视图特征（视觉特征和文本特征）投影到相同维度并进行阈值化。

4) 多视图融合模块： 我们采用 ContextGating 融合拼接后的视觉特征和文本特征。多视图融合模块将输入的多视图特征投影为新的全局表示，如下所示：

$$X_{\text{fusion}} = \sigma(w_{\text{fusion}} X_{\text{concat}} + b_{\text{fusion}}) \circ X_{\text{concat}}, \quad (1)$$

其中，$X_{\text{concat}} \in \mathbb{R}^n$ 是多视图特征向量，σ 是逐元素 sigmoid 激活，$\circ$ 是逐元素乘法。$w_{\text{fusion}} \in \mathbb{R}^{n\times n}$ 和$b_{\text{fusion}} \in \mathbb{R}^n$ 是可训练参数。权重向量 $\sigma(w_{\text{fusion}} X_{\text{concat}} + b_{\text{fusion}}) \in [0,1]$ 表示一组学习得到的门控值，用于作用于输入特征 X_{concat}的各个维度。

5) 哈希层： 采用带有 tanh 激活函数的线性层作为哈希层，可表示为$h_{\text{k-bit}} = \text{sgn}[\tanh(w_{\text{hash}} X_{\text{fusion}} + b_{\text{hash}})]$, 其中，$sgn$ 表示符号函数。$w_{\text{hash}} \in \mathbb{R}^{n\times n}$ 和 $b_{\text{hash}} \in \mathbb{R}^n$是可训练参数。输出的维度数与哈希码相同。

B. 深度度量损失

假设训练数据集 $\mathcal{X} = \left\{\{(x_i, y_i)\}_{i=1}^N\right\}$，其中 $x_i \in \mathbb{R}^D$ 表示一个多视图实例，y_i 表示 x_i的类别信息。此外，$F: x \mapsto h$ 表示深度多视图哈希网络，它将输入空间 $\mathbb{R}^D$ 映射到 K 位汉明空间 $\{-1,1\}^K$。

设 $h_i = F(x_i)$ 为 x_i 的哈希码。然后，汉明距离 $dist_H(\cdot,\cdot)$ 与内积 $\langle\cdot,\cdot\rangle$ 之间存在优雅的线性关系

$$dist_H(h_i, h_j) = \frac{1}{2}(K - \phi_{ij}), \quad (2)$$

其中 $\phi_{ij} = \langle h_i, h_j\rangle$。对于 x_i，其标签 $y_i \in \{0,1\}^C$，而 C是类别数。请注意，一个样本可能属于多个类别。给定语义标签信息，两两相似度矩阵 $S = \{s_{ij}\}$ 可定义如下：如果 x_i 和 x_j 在语义上相似，则 $s_{ij} = 1$；否则，$s_{ij} = 0$。

给定矩阵 $\Phi = (\phi_{ij})$ 和 $S = (s_{ij})$，将交叉熵损失与深度度量学习相结合，可得损失函数

$$L'_m = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[s_{ij}\log(1+e^{\phi_{ij}}) - s_{ij}\phi_{ij}]. \quad (3)$$

由于 s_{ij} 只能是 0 或 1，当 $s_{ij} = 0$ 时，损失 L'_m 消失，这意味着不相似样本在训练中不起任何作用。注意，度量损失的第一部分是 $\log(1 + e^{\phi_{ij}})$。考虑到汉明距离与内积之间优雅的线性关系，即式（2），随着内积 ϕ_{ij} 减小，汉明距离将增大。因此，这一部分是合适的度量损失。它会惩罚嵌入空间中距离较近的不相似样本，同时奖励它们之间较大的距离。基于上述分析，我们将式（3）修正为

$$L_m = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}[w_d\log(1+e^{\phi_{ij}}) - s_{ij}\phi_{ij}]. \quad (4)$$

w_d 表示不相似样本对的损失权重。经过这一修正，不相似样本也能助力训练。度量损失的推导过程见附录。

C. 超参数 λ

注意，计算矩阵 Φ 或 S 的复杂度为 $O(N^2)$。引入超参数 λ 后，计算其中任意一个的复杂度仅为 $O(\lambda^2 bN)$，其中 b 为批次大小。

我们随机选取其中一部分来计算相似度矩阵，而不是为每个样本计算全局相似度矩阵 S。假设样本已打乱。令 b 为批量大小，λ 为超参数。我们取前 λb 个和后 λb 个样本来计算损失。具体而言，令 $H_{\text{prec}} = \{h_1, h_2, \ldots, h_{\lambda b}\}$，$H_{\text{rest}} = \{h_{(1-\lambda)b+1}, h_{(1-\lambda)b+2}, \ldots, h_b\}$，$Y_{\text{prec}} = \{y_1, y_2, \ldots, y_{\lambda b}\}$和 $Y_{\text{rest}} = \{y_{(1-\lambda)b+1}, y_{(1-\lambda)b+2}, \ldots, y_b\}$。于是可得两个矩阵：$\Phi_{batch}$ 和 S_{batch}

$$\Phi_{\text{batch}} = H_{\text{prec}} \times H_{\text{rest}}^T = \left[\phi_{ij}\right] \quad (5)$$

$$S_{\text{batch}} = Y_{\text{prec}} \times Y_{\text{rest}}^T = \left[s_{ij}\right], \quad (6)$$

其中 $\times$ 表示矩阵乘法运算。通过这种设计，式（4）可简化为

$$L_m = \frac{1}{(\lambda b)^2}\sum_{i=1}^{\lambda b}\sum_{j=(1-\lambda)b+1}^{b}[w_d\log(1+e^{\phi_{ij}}) - s_{ij}\phi_{ij}]. \quad (7)$$

最后，引入量化损失来优化生成的哈希码，其表示如下：

$$L_q = \frac{1}{b}\sum_{i\in I}(\||\boldsymbol{h}_i| - \mathbf{1}\|_2), \quad (8)$$

其中 $I = \{i \mid 1 \le i \le \lambda b, i \in \mathbb{N}\} \cup \{i \mid (1-\lambda)b + 1 \le i \le b, i \in \mathbb{N}\}$。通过加权和结合度量损失与量化损失，可得到我们方法的总损失函数

$$L_{\text{Total}} = L_m + \mu L_q, \quad (9)$$

其中 μ 是在我们的工作中通过网格搜索得到的超参数。

III. 实验

我们进行了大量实验，在三个公开基准数据集上，将所提出的 DMMVH 方法与十一种最先进的多视图哈希方法进行比较。

数据集： 采用了三个基因数据集：MIR-Flickr25K [8], NUS-WIDE [9], 和 MS COCO [10]。这些数据集已被广泛用于评估多媒体检索性能。三个数据集的统计信息汇总于表 I 中。

评估指标： 我们采用平均精度均值（mAP）作为评估指标。

基线方法： 为评估检索性能，我们将所提出的方法与十一种多视图哈希方法进行比较，其中包括四种无监督方法（MFH [11],MAH [12],MVLH [13], 和 MvDH [14]）以及七种有监督方法（MFKH [15], DMVH [16],FDMH [17], FOMH[18],DCMVH [1],SAPMH [19],和 FGCMH [3]）。

实现细节： 我们的实现基于 PyTorch 平台。对于特征提取骨干网络，我们使用预训练模型，具体包括 ResNet-50 和 BERT- base。将丢弃概率设置为 0.1，以提升泛化能力。我们采用 AdamW优化器，初始学习率为 1×10^{-5}，并设置 $\beta_1 = 0.9$、$\beta_2 = 0.999$。深度度量学习损失函数的超参数 λ 为 0.5。总损失函数的组合系数 μ 设置为 0.5。不相似样本对的损失权重 w_d为 1.5。

A. 实验结果分析

平均精度均值： 结果如表 II 所示，表明 DMMVH 整体上以较大优势优于所有对比的多视图哈希方法。例如，与当前最先进的多视图哈希方法 FGCMH 相比，我们方法的平均 mAP 分数在 MIR-Flickr25K、NUS-WIDE 和 MS COCO 上分别提升了 3.51%、9.58% 和 13.85%。这表明，深度度量学习确实能够增强哈希码的判别能力。

哈希码长度： 直观而言，较长的哈希码应能保留更多语义信息，并取得更好的精确率。此外，我们研究了哈希码长度对多媒体检索平均精度均值的影响。对于不同的方法，通过设置相同的码长来学习哈希码。从表II 中可以看出，随着哈希码长度的增加，我们方法的平均精度均值也随之提高。在 MS COCO 数据集上，当哈希码长度从 16位增加到 128 位时，我们的方法取得了 5.25%的性能提升。在其他数据集上的实验也显示出相同的结论。然而，一些

表 I：总体统计信息　　三个数据集的数据集规模、类别数量和特征维度均已列出　　d.

数据集	训练集规模	检索集规模	查询集规模	类别	视觉特征	文本特征
MIR-Flickr25K	5000	17772	2243	24	ResNet（768-D）	BERT（768-D）
MS COCO	18000	82783	5981	80	ResNet(768-D)	BERT(768-D)
NUS-WIDE	21000	193749	2085	21	ResNet(768-D)	BERT(768-D)

表 II：MIR-Flickr25K、NUS-WIDE 和 MS COCO 上的平均精度均值比较结果。最佳结果以粗体标示，先前最先进的结果以下划线标示。* 表示该数据集上的结果具有统计显著性。

方法	Ref.	MIR-Flickr25K*				NUS-WIDE*				MS COCO*			
		16位	32位	64位	128位	16位	32位	64位	128位	16位	32位	64位	128位
MFH	TMM13	0.5795	0.5824	0.5831	0.5836	0.3603	0.3611	0.3625	0.3629	0.3948	0.3699	0.3960	0.3980
MAH	TIP15	0.6488	0.6649	0.6990	0.7114	0.4633	0.4945	0.5381	0.5476	0.3967	0.3943	0.3966	0.3988
MVLH	MM15	0.6541	0.6421	0.6044	0.5982	0.4182	0.4092	0.3789	0.3897	0.3993	0.4012	0.4065	0.4099
MvDH	TIST18	0.6828	0.7210	0.7344	0.7527	0.4947	0.5661	0.5789	0.6122	0.3978	0.3966	0.3977	0.3998
MFKH	MM12	0.6369	0.6128	0.5985	0.5807	0.4768	0.4359	0.4342	0.3956	0.4216	0.4211	0.4230	0.4229
DMVH	ICMR17	0.7231	0.7326	0.7495	0.7641	0.5676	0.5883	0.6902	0.6279	0.4123	0.4288	0.4355	0.4563
FOMH	MM19	0.7557	0.7632	0.7564	0.7705	0.6329	0.6456	0.6678	0.6791	0.5008	0.5148	0.5172	0.5294
FDMH	NPL20	0.7802	0.7963	0.8094	0.8181	0.6575	0.6665	0.6712	0.6823	0.5404	0.5485	0.5600	0.5674
DCMVH	TIP20	0.8097	0.8279	0.8354	0.8467	0.6509	0.6625	0.6905	0.7023	0.5387	0.5427	0.5490	0.5576
SAPMH	TMM21	0.7657	0.8098	0.8188	0.8191	0.6503	0.6703	0.6898	0.6901	0.5467	0.5502	0.5563	0.5672
FGCMH	MM21	0.8173	0.8358	0.8377	0.8606	0.6677	0.6874	0.6936	0.7011	0.5641	0.5273	0.5797	0.5862
DMMVH	提出	**0.8587**	**0.8707**	**0.8798**	**0.8827**	**0.7714**	**0.7820**	**0.7879**	**0.7916**	**0.6716**	**0.7030**	**0.7122**	**0.7244**

表 III：消融　　实验：在三个数据集上的实验。深度多视图哈希网络架构的影响。

方法	MIR-Flickr25K				NUS-WIDE				MS COCO			
	16位	32位	64位	128位	16位	32位	64位	128位	16位	32位	64位	128位
DMMVH-度量	0.8531	0.8614	0.8708	0.8738	0.7671	0.7766	0.7809	0.7872	0.6686	0.6970	0.7066	0.7078
DMMVH-量化	0.5531	0.5531	0.5531	0.5531	0.3085	0.3085	0.3085	0.3085	0.3502	0.3502	0.3502	0.3502
DMMVH-文本	0.6047	0.6107	0.6104	0.6119	0.3623	0.3585	0.3649	0.3631	0.5819	0.5886	0.5955	0.5992
DMMVH-图像	0.8292	0.8425	0.8547	0.8631	0.7530	0.7593	0.7689	0.7778	0.6598	0.6886	0.7033	0.7160
DMMVH-拼接	0.8498	0.8633	0.8742	0.8777	0.7635	0.7713	0.7827	0.7866	0.6615	0.6932	0.7056	0.7188
DMMVH	**0.8587**	**0.8707**	**0.8798**	**0.8827**	**0.7714**	**0.7820**	**0.7879**	**0.7916**	**0.6716**	**0.7030**	**0.7122**	**0.7244**

以往方法在增加哈希位数时精确率会下降，这表明这些方法难以很好地扩展到哈希码长度更长的哈希任务。相反，我们的结果表明，随着哈希码长度增加，所提出的方法在平均精度均值方面有明显提升。最后，附录中详细介绍了超参数实验。

B. 消融实验

实验设置：为评估我们方法的有效性，我们采用不同设置进行消融研究，并报告实验结果。

- DMMVH-metric：量化损失已移除。
- DMMVH-quant：移除了度量损失。
- DMMVH-图像：仅使用视觉特征。
- DMMVH-文本：仅使用文本特征。
- DMMVH-拼接：通过拼接融合图像和文本特征。

- DMMVH：我们的完整框架。

消融分析：比较结果如表 III 所示。首先来看损失函数，量化损失无法对嵌入进行任何优化。该方法会随机检索数据，导致其在所有任务上的平均精度均值都很差。相反，深度度量损失有助于该方法更好地学习嵌入。我们注意到，由于缺少二值化约束，DMMVH-度量的表现略逊于完整方法。从视图角度来看，DMMVH-文本仅略优于 DMMVH-量化。DMMVH-图像在所有任务上的表现都大幅优于 DMMVH-文本，这表明图像特征所包含的信息多于文本特征。采用拼接的多视图特征后，我们的方法已经优于最先进的方法。但上下文门控进一步提升了平均精度均值。此外，旧骨干网络的对比实验详见附录。

C. 收敛性分析

我们通过实验验证 DMMVH 的收敛性和泛化能力。我们在 MIR-Flickr25K 数据集上针对不同编码长度运行哈希基准测试。结果如图3所示。该图展示了训练损失和测试平均精度均值，以供分析。随着训练的进行，损失逐渐降低。经过 500 个训练轮次后，损失趋于稳定，这意味着已达到局部最小值。就测试性能而言，平均精度均值在训练初期迅速上升

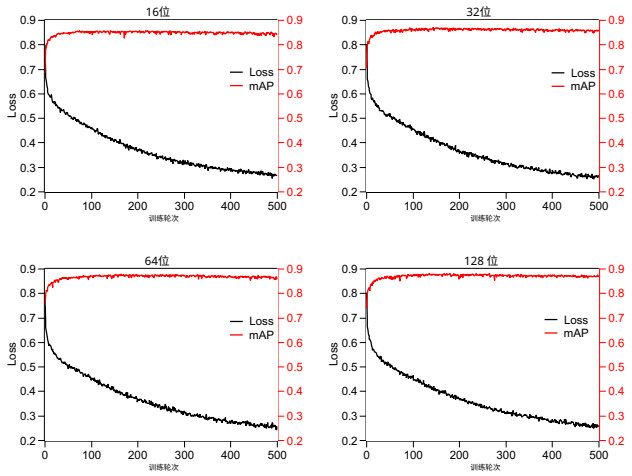

图3：上方曲线为测试平均精度均值，下方曲线为 MIR-Flickr25K 数据集上的训练损失。

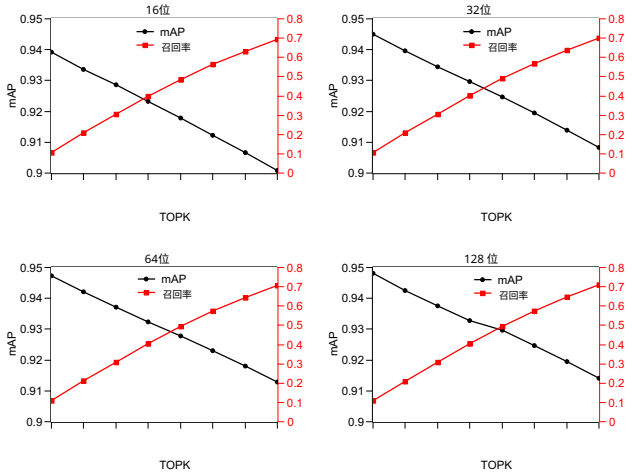

图 4：MIR-Flickr25K 数据集上的 mAP@K 和召回率@K 曲线。

。经过 100 个训练轮次后，测试平均精度均值保持稳定。继续训练后，测试平均精度均值没有出现下降，表明模型具有良好的泛化能力。在其他数据集上也观察到了类似的收敛结果。

D. mAP@K 和召回率@K

图 4 展示了在 MIR-Flickr25K 数据集上，采用不同编码长度时，随着检索结果数量增加，mAP@K 和召回率@K 曲线的变化情况。随着 K 增加，四种情况下的平均精度均值略有下降，而召回率曲线则呈现快速线性增长趋势。这表明我们的方法在检索任务中表现良好。普通用户通常只关注检索结果开头的少数几项。在这种情况下，我们的方法具有更高的精确率。专家往往会浏览比普通用户更多的结果。随着检索结果数量增加，我们的方法能够实现线性增长的召回率。专家可以期待在搜索过程中获得持续稳定的高质量结果。总之，DMMVH 能为不同用户群体提供令人满意的检索结果。

IV. 结论与未来工作

我们提出了一种新的多视图哈希框架（DM-MVH），将深度度量学习引入多视图哈希问题。我们展示了 DMMVH 能够为不同类型的用户提供令人满意的检索结果。与典型的基于图的方法相比，DMMVH 的计算开销更低。它利用上下文门控融合多视图特征，并通过深度度量学习优化表示。所提出的方法解决了多视图哈希问题的两大主要挑战。在多种实验设置下，其性能比最先进的方法最高提升了 15.28%。实验中，我们注意到一些问题。例如，随着哈希码长度增加，性能提升并不十分显著。我们将继续研究这些问题，以进一步改进所提出的方法。

致谢

本研究得到浙江省“万人计划”（2021R52007）、国家重点研发计划（2022YFB4500405）和科技创新2030——重大项目（2021ZD0114300）的部分资助。

参考文献

[1] Lei Zhu、Xu Lu、Zhiyong Cheng、Jingjing Li 和 Huaxiang Zhang，“用于大规模图像搜索的深度协同多视图哈希”，IEEE 图像处理汇刊，第29卷，第4643–4655页，2020年。[2] Lei Zhu、Xu Lu、Zhiyong Cheng、Jingjing Li 和 Huaxiang Zhang，“用于可扩展多媒体检索的灵活多模态哈希”，ACM智能系统与技术汇刊（TIST），第11卷，第2期，第1–20页，2020年。[3] Xu Lu、Lei Zhu、Li Liu、Liqiang Nie 和 Huaxiang Zhang，“用于灵活多媒体检索的图卷积多模态哈希”，载于第29届 ACM 多媒体国际会议论文集，2021年，第1414–1422页。[4] Max Welling 和 Thomas N Kipf，“基于图卷积网络的半监督分类”，载于《国际学习表征会议》（ICLR 2017）论文集，2016年。[5] Antoine Miech、Ivan Laptev 和 Josef Sivic，“利用上下文门控进行视频分类的可学习池化”，arXiv预印本 arXiv:1706.06905，2017年。[6] Kaiming He、Xiangyu Zhang、Shaoqing Ren 和 Jian Sun，“用于图像识别的深度残差学习”，载于 IEEE计算机视觉与模式识别会议论文集，2016年，第770–778页。[7] Jacob Devlin、Ming-Wei Chang、Kenton Lee 和 Kristina Toutanova，“BERT：用于语言理解的深度双向Transformer预训练”，arXiv预印本 arXiv:1810.04805，2018年。[8] Mark J Huiskes 和 Michael S Lew，“MIR Flickr检索评测”，载于第1届 ACM多媒体信息检索国际会议论文集，2008年，第39–43页。[9] Tat-Seng Chua、Jinhui Tang、Richang Hong、Haojie Li、Zhiping Luo 和 Yantao Zheng，“NUS-WIDE：来自新加坡国立大学的真实世界网络图像数据库”，载于 ACM图像与视频检索国际会议论文集，2009年，第1–9页。[10] Tsung-Yi Lin、Michael Maire、Serge Belongie、James Hays、Pietro Perona、Deva Ramanan、Piotr Doll´ar 和 C Lawrence Zitnick，“Microsoft COCO：上下文中的常见物体”，载于《欧洲计算机视觉会议》论文集。施普林格，2014年，第740–755页。[11] Jingkuan Song、Yi Yang、Zi Huang、Heng Tao Shen 和 Jiebo Luo，“用于大规模近重复视频检索的有效多特征哈希”，IEEE多媒体汇刊，第15卷，第8期，第1997–2008页，2013年。[12] Li Liu、Mengyang Yu 和 Ling Shao，“用于高效图像搜索的多视图对齐哈希”，IEEE 图像处理汇刊，第24卷，第3期，第956–966页，2015年。[13] Xiaobo Shen、Fumin Shen、Quan-Sen Sun 和 Yun-Hao Yuan，“用于高效多媒体搜索的多视图潜在哈希”，载于第23届 ACM多媒体国际会议论文集，2015年，第831–834页。[14] Xiaobo Shen、Fumin Shen、Li Liu、Yun-Hao Yuan、Weiwei Liu 和 Quan-Sen Sun，“用于可扩展多媒体搜索的多视图离散哈希”，ACM智能系统与技术汇刊（TIST），第9卷，第5期，第1–21页，2018年。[15] Xianglong Liu、Junfeng He、Di Liu 和 Bo Lang，“结合多特征的紧凑核哈希”，载于第20届 ACM多媒体国际会议论文集，2012年，第881–884页。[16] Rui Yang、Yuliang Shi 和 Xin-Shun Xu，“用于有效图像检索的离散多视图哈希”，载于2017年 ACM多媒体检索国际会议论文集，2017年，第175–183页。[17] Luyao Liu、Zheng Zhang 和 Zi Huang，“结合集体潜在特征学习的灵活离散多视图哈希”，《神经处理快报》，第52卷，第3期，第1765–1791页，2020年。[18] Xu Lu、Lei Zhu、Zhiyong Cheng、Jingjing Li、Xiushan Nie 和 Huaxiang Zhang，“用于大规模多媒体检索的灵活在线多模态哈希”，载于第27届 ACM多媒体国际会议论文集，2019年，第1129–1137页。[19] Chaoqun Zheng、Lei Zhu、Zhiyong Cheng、Jingjing Li 和 An-An Liu，“用于高效社交图像检索的自适应部分多视图哈希”，IEEE多媒体汇刊，第23卷，第4079–4092页，2020年。