

宋文韬, 姜茹月, 舒欣. 基于零样本学习的枸杞虫害识别[J]. 江苏农业学报, 2024, 40(2) : 320-330.

doi: 10.3969/j.issn.1000-4440.2024.02.014

基于零样本学习的枸杞虫害识别

宋文韬, 姜茹月, 舒欣

(南京农业大学人工智能学院, 江苏 南京 210095)

摘要: 针对农业领域缺少有效的零样本虫害识别与检索方法, 本研究提出一种基于零样本学习的枸杞虫害检索与识别方法。首先, 通过对原始数据进行深层矩阵分解获得深层次结构特征, 获取不同模态数据的特征表示, 生成各模态的哈希码。然后结合类别属性信息对生成的哈希码引入线性约束, 实现已知类别到新类别之间的知识迁移。最后, 对所提出的模型通过直接学习离散哈希码避免了连续松弛方法带来的量化误差, 提高了检索精度。在 2020 年宁夏枸杞虫害图文跨模态检索数据集及 Wiki、Pascal VOC 这 3 个公开数据集上的试验结果表明, 与现有的基于协同矩阵分解的哈希方法(CMFH)、基于潜在语义的稀疏哈希方法(LSSH)、基于迁移监督知识的哈希方法(TSK)、基于属性的哈希方法(AH)、基于跨模态属性的哈希方法(CMAH)、基于正交投影的哈希方法(CHOP)、离散非对称零样本哈希方法(DAZSH)相比, 本研究所提出的方法具有优越性。

关键词: 零样本学习; 矩阵分解; 枸杞病虫害识别; 哈希码

中图分类号: TP391 文献标识码: A 文章编号: 1000-4440(2024)02-0320-11

Identification of *Lycium barbarum* pests based on zero-shot learning

SONG Wen-tao, JIANG Ru-yue, SHU Xin

(College of Artificial Intelligence, Nanjing Agricultural University, Nanjing 210095, China)

Abstract: In order to solve the problem of lack of effective zero-sample recognition and retrieval methods in agricultural field, a zero-sample learning-based retrieval and recognition method for *Lycium barbarum* pests was proposed in this study. Firstly, the deep structure features were obtained by deep matrix decomposition of the original data, and the characteristic representations of different modal data were obtained, and the hashing codes of each modality were generated. Then the linear constraint was introduced to the generated hashing code with the class attribute information to realize the knowledge transfer from the known class to the new class. Finally, the proposed model could avoid the quantization error caused by the continuous relaxation method and improve the retrieval precision by learning discrete hashing codes directly. The experimental results on the three public datasets, 2020 Ningxia *Lycium barbarum* pest image-text cross-modal retrieval dataset, Wiki, Pascal VOC, showed that the method proposed in this study was superior to the existing collective matrix factorization hashing (CMFH), latent semantic sparse hashing (LSSH), transferring supervised knowledge hashing (TSK), attribute hashing (AH), cross-modal attribute hashing (CMAH), cross-modal hashing with orthogonal projection (CHOP), and discrete asymmetric zero-shot hashing (DAZSH).

收稿日期: 2023-11-01

基金项目: 国家自然科学基金项目(61602248); 江苏省信息技术处理重点实验室开放课题项目(KJS2164)

作者简介: 宋文韬(1997-), 男, 江苏涟水人, 硕士, 主要研究方向为计算机视觉、模式识别。(E-mail) 374267655@qq.com。
姜茹月为共同第一作者。

通讯作者: 舒欣, (E-mail) xinshu@njau.edu.cn

zation hashing (CMFH), latent semantic sparse hashing (LSSH), transferring supervised knowledge hashing (TSK), attribute hashing (AH), cross-modal attribute hashing (CMAH), cross-modal hashing with orthogonal projection (CHOP), and discrete asymmetric zero-shot hashing (DAZSH).

Key words: zero-shot learning; matrix factorization; *Lycium barbarum* pests detection; hashing code

枸杞具有增强人体免疫力、护肝明目、滋养皮肤等功效。同时,作为耐旱、耐贫瘠、耐盐碱的灌木作物,其防风固沙的能力不容小觑。由于上述特性,枸杞在中国西北地区广泛种植^[1-2]。然而,枸杞作为抗虫害能力较差的作物,容易遭受多种虫害导致产品质量和数量方面的下降。因此,准确且快速地对枸杞虫害类型进行识别与检索,及时给出防治措施从而减小因作物虫害导致的经济损失,在推动枸杞产业的发展方面尤为重要。

传统的农作物病虫害识别与检索主要依靠肉眼观察作物发生病虫害部位,将观察结果与农作物病虫害图像信息手册中的病虫害图像进行比较从而判别病虫害类别^[3]。这种方法不仅花费大量的时间和人力,由于观察所得结果具有主观性,识别与检索结果的精度也不令人满意^[4]。随着互联网的发展,基于机器学习的农作物病虫害识别方法受到了研究人员的广泛关注。机器学习方法是根据已有的数据设计并构造算法使得计算机学习获得数据特征的规律,以此进行预测^[5]。赵芸^[6]利用光谱数据建模以及数字图像处理等技术,对油菜病虫害的分类进行建模。赵建敏等^[7]利用OTSU 阈值算法,对马铃薯叶片病害区域图像进行分割,提取对应部分的颜色、形状、纹理特征,采用支持向量机进行分类识别,识别率为 92%。王佳^[8]使用计算机视觉技术,对香芋叶片提取颜色和纹理特征,以此构建病害数据库,同时采用支持向量机对香芋病害进行识别,识别的平均准确率达到 93.00%。Nettleton 等^[9]提出了 M5Rules 和递归神经网络的两种基于机器学习算法的方法,与 Yoshino 以及 WARM 这 2 种基于过程的模型进行比较,经过试验得出基于机器学习的两个模型在最大平均归一化以及平均绝对误差(MAE)结果方面更加优秀,同时具有更好的泛化性。基于机器学习方法尽管可以通过提取农作物的病害图像特征对农作物病虫害进行识别和检索,但是由于客观条件,如光照条件变化、复杂背景的存在等,难以提取有效的数据特征。此外,机器学习方法需要大量的数据支持模型训练,才可以获得较好的识别准确率,然而现实中获取大量有效的农作物病虫害图像样本比较困难。因此现有基于机器学习方法的农作物病虫害识别与检索仍然具有挑战。

随着深度学习的不断发展,近年来基于深度学习的农作物病虫害识别方法在农业领域取得了不错的成果。王国伟等^[10]对 LeNet 模型进行改进,通过使用 Adam 替换 SGD 以及添加 L2 正则项到交叉熵函数构建 10 层 CNN 网络结构,对于 5 种玉米病虫害的平均识别率达 96%。赵立新等^[11]对 AlexNet 进行改进,引入迁移学习方法辅助数据增强,在此基础上对预训练模型进行微调,识别棉花叶部病虫害的平均准确率达 97.16%。鲍文霞等^[12]针对苹果病害叶片数据集,采用选择性核卷积模块以及使用全局平均池化替代全连接层的方法对 VGG16 网络模型进行改进,对苹果病害局部病斑的识别准确率达 94.70%。冯晓等^[13]分析 MobileNetV2 模型以及迁移学习方法,通过合适的增强方法、添加 Dropout 层、改变初始学习率等对模型进行改进,改进后的模型对 3 种小麦叶部病害的识别准确率达到 99.96%。谢州益等^[14]使用 ResNet18 作为特征编码器,在解码器上融合了注意力机制的 LSTM,对 10 种水稻病虫害的 Top5 准确率达 98.48%。彭红星等^[15]改进 MobileNetV2 模型,在反向残差模块种引入坐标注意力机制,同时设计双分支特征融合模块,对 19 类葡萄病虫害的识别准确率达 89.16%。

然而现有病虫害识别与检索方法要求在训练阶段获得所有类别信息及对应的样本,不能对一些未知类别病害进行识别。为此,研究人员提出了零样本学习技术解决上述问题。零样本学习是指根据一些可见类别的数据,融合辅助信息用于训练模型,实现对未见类别的数据进行分类。目前主流的学习方式是利用嵌入模型,其主要思路是将数据特征嵌入到同一个语义空间中进行相似性度量学习^[16]。Zhong 等^[17]提出了一种多层次的网络架构,将不同模态的数据特征和类别标签以及属性信息联系起来。Ji 等^[18]构建了一个属性引导的网络结构,使用深度神经网络学习相应的数据特征。Shu 等^[19]结合数据特征及其属性信息,使用非对称相似策略构造哈希码。现有主流零样本学习方法侧重于利用语义空间来实现对未知类别的分类,在数据特征提取的过程中没有嵌入农作物病虫害标记属性信息。此外,基于零样本学习的农作物病虫害识别与检索也处于探索阶段,尚未形成有效的研究成果。为此,本研究以枸杞虫害为研究对象,提出一种基于零样本

学习的枸杞虫害检索与识别方法。

1 材料与方法

1.1 试验数据集介绍

本研究选取了 2020 年宁夏枸杞虫害图文跨模态检索数据集^[20]进行零样本以图搜文和以文搜图两种跨模态检索任务的相关试验。同时,本研究选取了现有零样本跨模态检索方法常用的 2 个公开非农业数据集 Wiki 和 Pascal VOC 数据集,进一步验证所提方法的有效性。试验所使用的数据集详情如下:

(1)2020 年宁夏枸杞虫害图文跨模态检索数据集包含了来源于 17 个类别经扩增处理后的共 9 496 对枸杞虫害图像-文本样本对。试验随机选取了每个类别的 80% 共计 7 596 对样本对组成训练集,剩余 1 900 对样本对组成测试集。对于图像模态数据,使用 512 维的 GIST 特征表示,对于文本模态数据,使用潜在狄利克雷分布方法提取为 10 维的特征向量。

(2)Wiki 数据集包含了来源于 10 个类别的共 2 866 对图像-文本样本对,其中 2 173 对为训练集数据,剩余 693 对为测试集数据。对于图像模态数据,通过 BOVW 的 SIFT 提取为 128 维特征向量,对于文本模态数据,通过潜在狄利克雷分布方法提取为 10 维特征向量。

(3)Pascal VOC 数据集包含了来源于 20 个类别的共 9 963 对图像-标签样本对,本试验只选取其中单标签的数据,共有 2 799 对训练集数据以及 2 820 对测试集数据。对于图像模态数据,使用 512 维 GIST 特征表示,对于文本模态数据,使用 399 维的词频特征描述。

对于每个数据集中的类别标签,采用了 GloVe 模型对其进行词向量转换,生成对应的 300 维特征向量作为类别属性信息。

1.2 试验设置及评价指标

本研究随机选取数据集中的类别并将其设置为不可见类进行试验。其中,对于 2020 年宁夏枸杞虫害数据集,本研究随机选取 3 个类别作为不可见类,剩余 14 个类别作为可见类。对于 Wiki 数据集,本研究随机选取 2 个类别作为不可见类,剩余 8 个类别作为可见类。对于 Pascal VOC 数据集,本研究随机选取 4 个类别作为不可见类,剩余 16 个类别作为可见类。共进行 10 次试验取平均值作为结果。在训练阶段,训练样本包含了来源于图片以及文本模

态的特征向量、one-hot 形式的类别标签以及属性信息。在评估阶段,查询样本为测试集中不可见类的全体样本,用于检索的数据库包含了训练时的可见类样本以及未参与训练的不可见类样本。对于 2020 年宁夏枸杞虫害、Wiki 以及 Pascal VOC 数据集,设置了两个中间层,其中 p_1 为 200, p_2 为 100。对于 2020 年宁夏枸杞虫害数据集,超参数 η 设置为 150, τ 设置为 4, α 设置为 6, λ 设置为 40。对于 Wiki 数据集,超参数 η 设置为 350, τ 设置为 6, α 设置为 6, λ 设置为 15。对于 Pascal VOC 数据集,超参数 η 设置为 100, τ 设置为 5, α 设置为 3, λ 设置为 20。

本研究采用 mAP 、 $PH2$ 、Top-10 准确率 3 种常见的评价指标对提出的方法进行评价。3 种评价指标的定义如下:

$$mAP = \frac{1}{Q} \sum_{q=1}^Q \frac{1}{R'} \sum_{r=1}^R P_q(r) \delta_q(r)$$

式中: Q 为待检索样本数; R 为数据库中数据量; R' 为数据库与待检索样本相关的数据量; $P_q(r)$ 为排名第 q 个检索样本的 $top-r$ 精度值, $\delta(r) \in \{0, 1\}$, 当第 r 个数据实例与第 q 个检索样本相关则为 1, 否则为 0。 $PH2$ 表示在汉明空间中由哈希对照表判定半径为 2 的检索准确度, 反应模型对局部数据分情况描述的准确性。 Top-10 准确率为检索结果中概率最大的前 10 个结果包含正确类别的占比。

1.3 模型概述

基于零样本学习的枸杞病虫害检索与识别模型。假设 $X^{(1)} = [x_1^{(1)}, x_2^{(1)}, \dots, x_n^{(1)}] \in R^{d_1 \times n}$ 为图像模态, $X^{(2)} = [x_1^{(2)}, x_2^{(2)}, \dots, x_n^{(2)}] \in R^{d_2 \times n}$ 为文本模态, n 为可见类别的数据样本数, d_1 和 d_2 分别为图像和文本模态数据特征维度, 一般情况下 $d_1 \neq d_2$ 。可见类别以及不可见类别的标签信息分别表示为 $L_s = [l_1, l_2, \dots, l_s] \in R^{l_s \times s}$ 和 $L_u = [l_1, l_2, \dots, l_u] \in R^{l_u \times u}$, s 和 u 分别为可见类别、不可见类别的数量, 且 $L_s \cap L_u = \emptyset$ 。令 $A = [a_1, a_2, \dots, a_n] \in R^{d_a \times n}$ 为类别属性, A 通常由标签信息决定, d_a 为属性特征维数。本研究可见类别的数据信息为图像和文本模态学习统一的哈希编码和每个模态对应的哈希函数, 在检索阶段将不可见类别的数据投影到汉明空间中进行相似度评估。

1.3.1 深层矩阵分解 本研究使用深层矩阵分解的方法来学习不同模态数据的深层潜在特征表示。考虑图像特征矩阵的半非负矩阵分解, 有 $X^{(1)} \approx Z_1^{(1)} H_1^{(1)}$, 其中 $Z_1^{(1)}$ 为基向量矩阵, $H_1^{(1)}$ 为潜在特征

矩阵,且 $Z_1^{(1)} \geq 0$ 。对潜在特征矩阵 $H_1^{(1)}$ 继续进行半非负矩阵分解得到 $H_1^{(1)} \approx Z_2^{(1)} H_2^{(1)}$, 其中 $Z_2^{(1)}$ 为基向量矩阵, $H_2^{(1)}$ 为潜在特征矩阵。如此循环往复直到 $H_{m-1}^{(1)} \approx Z_m^{(1)} H_m^{(1)}$, 其中 $H_m^{(1)}$ 为获得第 m 层的潜在特征矩阵。类似的可以获得文本模态第 m 层的潜在特征矩阵 $H_m^{(2)}$ 。上述分解过程可以描述如下:

$$\begin{cases} X^{(v)} \approx Z_1^{(v)} H_1^{(v)}, Z_1^{(v)} \geq 0 \\ H_1^{(v)} \approx Z_2^{(v)} H_2^{(v)}, Z_2^{(v)} \geq 0 \\ \vdots \\ H_{m-1}^{(v)} \approx Z_m^{(v)} H_m^{(v)}, Z_m^{(v)} \geq 0 \end{cases} \quad (1)$$

本研究使用如下损失函数衡量分解时的误差:

$$L_1 = \| X^{(v)} - Z_1^{(v)} Z_2^{(v)} \dots Z_m^{(v)} H_m^{(v)} \|_F^2 \quad (2)$$

式中, $Z_i^{(v)} \in R^{p_{i-1} \times p_i}$, $i \in \{1, 2, \dots, m\}$, $v \in \{1, 2\}$, p_i 为第 i 层的维数, p_0 为初始维数, $p_m = k$ 为哈希码长度。

1.3.2 统一哈希码学习 由于不同模态的样本对属于同一个类别,生成的哈希码在汉明空间中的距离应该足够小,可以近似看成同一个哈希码,故本研究将图像模态生成的哈希码和文本模态生成的哈希码进行如下约束:

$$L_2 = \| H_m^{(v)} - H \|_F^2 \quad (3)$$

通过上述约束可以获得统一的哈希码 $H \in \{-1, 1\}^{k \times n}$, 其中 k 为哈希码长度。

1.3.3 类别信息嵌入 为了学习具有判别能力的哈希码,即对于相同类别的样本,其汉明空间中的距离应该足够小,同时对于不同类别的样本则汉明空间中的距离应该足够大。本研究通过以下线性分类损失函数,获得具有判别能力的哈希码:

$$L_3 = \| W^T H - L_S \|_F^2 \quad (4)$$

式中, W 为线性分类器。

1.3.4 属性特征保持 为了适应零样本环境下的跨模态检索任务,生成的哈希码还需要具有一定的属性保持能力,从而对不可见类的样本,在检索阶段也可以通过相似的属性信息来对其进行有效的分类。因此对类别属性设置以下线性约束:

$$L_4 = \| V^T A - H \|_F^2 \quad (5)$$

式中, V 为线性分类器。

1.3.5 哈希函数学习 在检索阶段需要将不可见类的样本投影到汉明空间中进行相似性度量,本研究考虑线性哈希码生成函数,对应的损失函数如下:

$$L_5 = \| P^{(v)} X^{(v)} - H \|_F^2 \quad (6)$$

式中, $P^{(v)} \in R^{k \times d_v}$, $v \in \{1, 2\}$ 为投影矩阵。

综上所述,本研究所提出的方法的目标函数如下:

$$\begin{aligned} \min_{Z_i^{(v)}, H_i^{(v)}, H, W, V, P^{(v)}} L &= L_1 + \tau L_2 + \lambda L_3 + \alpha L_4 + \eta L_5 \\ \text{s.t. } Z_i^{(v)} &\geq 0, H \in \{-1, 1\}^{k \times n} \end{aligned} \quad (7)$$

式中, $\tau, \lambda, \alpha, \eta$ 为超参数。

1.4 优化

由于目标函数是非凸的,本研究采用交替迭代优化方法来求解所提出的模型。

1.4.1 优化 $Z_i^{(v)}$ 由公式(7)可知, $Z_i^{(v)}$ 的解可以表述为如下优化问题:

$$\min_{Z_i^{(v)}} \| X^{(v)} - Z_1^{(v)} Z_2^{(v)} \dots Z_m^{(v)} H_m^{(v)} \|_F^2 \quad (8)$$

$$\text{s.t. } Z_i^{(v)} \geq 0$$

对于 $X^{(v)}$, $v \in \{1, 2\}$ 的求解是相互独立的,为求解方便,舍去上标 v 得到:

$$\min_{Z_i} \| X - D Z_i H_i \|_F^2, \text{s.t. } Z_i \geq 0 \quad (9)$$

式中, $D = Z_1 Z_2 \dots Z_{i-1}$, $1 \leq i \leq m$ 。进一步,可以得到:

$$Z_i = \begin{cases} Z_i \odot \sqrt{\frac{[X \tilde{H}_1^T]^+ + [Z_1 \hat{H}_1]^-}{[X \tilde{H}_1^T]^- + [Z_1 \hat{H}_1]^+}}, & \text{如果 } i=1 \\ Z_i \odot \sqrt{\frac{[D^T X \tilde{H}_i^T]^+ + [D^T Z_i \hat{H}_i]^-}{[D^T X \tilde{H}_i^T]^- + [D^T Z_i \hat{H}_i]^+}}, & \text{如果 } i>1 \end{cases} \quad (10)$$

式中, \tilde{H}_i 是对第 i 层特征矩阵的重构矩阵, $\hat{H}_i = \tilde{H}_i \tilde{H}_i^T$, \odot 表示哈达玛积。

1.4.2 优化 $H_i^{(v)}$ 固定其他变量,关于 $H_i^{(v)}$ 的目标函数为:

$$\min_{H_i^{(v)}} \| X^{(v)} - Z_1^{(v)} Z_2^{(v)} \dots Z_i^{(v)} H_i^{(v)} \|_F^2$$

对 $H_i^{(v)}$ 求偏导并将偏导置为 0 可得:

$$H_i^{(v)} = (\Phi^T \Phi)^{-1} \Phi^T X^{(v)} \quad (11)$$

式中, $i < m$, $\Phi = Z_1^{(v)} Z_2^{(v)} \dots Z_i^{(v)}$ 。

1.4.3 优化 $H_m^{(v)}$ 固定其他变量,关于 $H_m^{(v)}$ 的优化问题可以为:

$$\min_{H_m^{(v)}} \| H^{(v)} - \Phi_m H_m^{(v)} \|_F^2 + \tau \| H_m^{(v)} - H \|_F^2$$

对第 m 层的潜在特征描述 $H_m^{(v)}$ 求偏导并将偏导置为 0,可以求解得出:

$$H_m^{(v)} = (\Phi_m^T \Phi_m + \tau I)^{-1} (\Phi_m^T X^{(v)} + \tau H) \quad (12)$$

式中, I 为单位矩阵, $\Phi_m = Z_1^{(v)} Z_2^{(v)} \dots Z_m^{(v)}$ 。

1.4.4 优化 H 固定其他变量,可以获得如下子问

题:

$$\min_H \sum_{v=1}^V (\tau \|H_m^{(v)} - H\|_F^2 + \lambda \|P^{(v)} X^{(v)} - H\|_F^2) + \alpha \|V^T A - H\|_F^2 + \eta \|W^T H - L_S\|_F^2$$

式中, $H \in \{-1, 1\}^{k \times n}$, 可以通过线性化方法进行求解得到:

$$H = \text{sgn} \left\{ \sum_{v=x}^y \left[\frac{\tau}{\eta} H_m^{(v)} + \frac{\lambda}{\eta} P^{(v)} X^{(v)} + \frac{\alpha}{\eta} V^T A \right] - WW^T H + WL_S \right\} \quad (13)$$

式中, sgn 为符号函数。

1.4.5 优化 $P^{(v)}$ 固定其他变量, 关于 $P^{(v)}$ 的优化问题可以为:

$$\min_{P^{(v)}} \|P^{(v)} X^{(v)} - H\|_F^2$$

对 $P^{(v)}$ 求偏导并将偏导置为 0, 不难得出:

$$P^{(v)} = H X^{(v)T} [X^{(v)} X^{(v)T}]^{-1} \quad (14)$$

1.4.6 优化 W 固定其他变量可得:

$$\min_W \|W^T H - L_S\|_F^2$$

对 W 求偏导并将偏导置为 0, 不难得出:

$$W = (HH^T)^{-1} HL_S^T \quad (15)$$

1.4.7 优化 V 固定其他变量可得:

$$\min_V \|V^T A - H\|_F^2$$

对 V 求偏导并将偏导置为 0, 不难得出:

$$V = (AA^T)^{-1} AH^T \quad (16)$$

通过交替求解上述优化问题, 可得到原问题 (7) 的最优解, 总体优化流程为:

输入: 图片和文本模态中可见类别数据为 $X^{(1)}$ 和 $X^{(2)}$, 类别标签为 L_s , 属性信息矩阵为 A , 中间层大小为 p_i , 超参数为 $\tau, \lambda, \alpha, \eta$ 。

输出: 生成的哈希码矩阵 H 以及两个模态对应的哈希函数 $P^{(v)}$ 。

- 1: 随机初始化 $[Z_i^{(v)}, H_i^{(v)}], P^{(v)}, W$ 和 V ;
- 2: 重复;
- 3: 对每个模态中所有层通过公式 (10) 更新 $Z_i^{(v)}$;
- 4: 对于每个模态通过公式 (11) 和公式 (12) 更新 $H_i^{(v)}$ 和 $H_m^{(v)}$;
- 5: 通过公式 (14) 更新 $P^{(v)}$;
- 6: 通过公式 (15) 更新 W ;
- 7: 通过公式 (16) 更新 V ;
- 8: 通过公式 (13) 更新 H ;
- 9: 直到收敛结束返回 H 和 $P^{(v)}$ 。

1.5 时间复杂度分析

设 $p = \max(p_i), 1 \leq i \leq m$, 上述优化过程中的时间复杂度更新如下:

(1) 更新 $Z_i^{(v)}$ 的时间复杂度为 $O(2p^2 n + 2p^2 d_v + n d_v p + p^3)$;

(2) 更新 $H_i^{(v)} (i < m)$ 的时间复杂度为 $O(n d_v p + p^2 n + d_v p^2)$;

(3) 更新 $H_m^{(v)}$ 的时间复杂度为 $O(n d_v p + d_v p^2 + p^3)$;

(4) 更新 $P^{(v)}$ 的时间复杂度为 $O(n d_v p + d_v^2 n)$;

(5) 更新 H 的时间复杂度为 $O(sp^2 + spn + 2pd_v n)$;

(6) 更新 W 的时间复杂度为 $O(nsp + p^2 n)$;

(7) 更新 V 的时间复杂度为 $O(n d_v^2 + p d_v n)$ 。

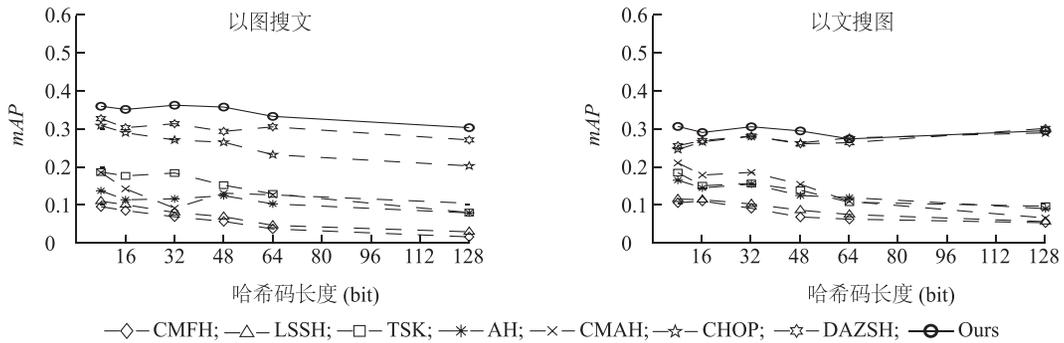
令 $d = \max\{d_v\}, v \in \{1, 2\}$, 则总的时间复杂度为 $O\{t l [2(3pdn + 2p^2 n + nd^2 + nps + dp^2 + sp^2) + p^3]\}$, 其中 t 为迭代次数, $l = m + 1$ 为总的层数。通常情况下, p, d, s 的取值都远远小于样本数 n , 易知所提出的算法的时间复杂度可以表示为 $O(n)$ 。

2 结果与分析

2.1 试验结果

将提出的方法与 7 种哈希方法进行比较。其中基于协同矩阵分解的哈希方法 (CMFH)^[21] 和基于潜在语义的稀疏哈希方法 (LSSH)^[22] 为跨模态哈希方法, 基于属性的哈希方法 (AH)^[23] 和基于迁移监督知识的哈希方法 (TSK)^[24] 为零样本单模态哈希方法, 基于跨模态属性的哈希方法 (CMAH)^[17]、基于正交投影的哈希方法 (CHOP)^[25] 和离散非对称零样本哈希方法 (DAZSH)^[19] 为零样本跨模态哈希方法。

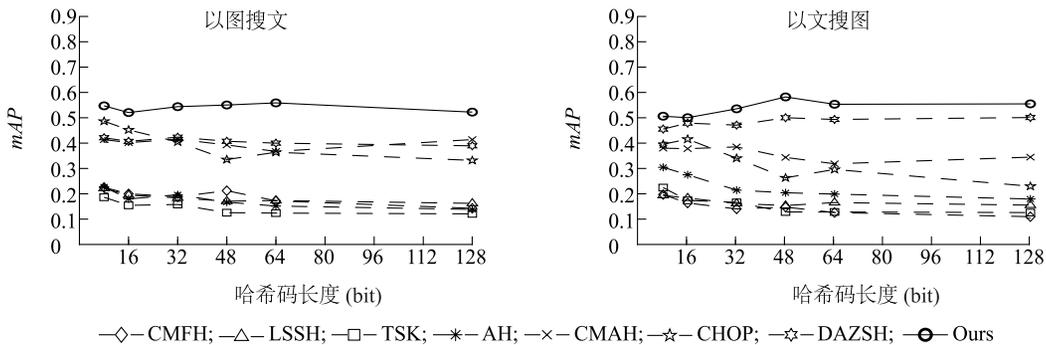
图 1、图 2、图 3 显示了本研究所提出的方法与对比方法在 3 个数据集上的 mAP 值。与 CMFH 以及 LSSH 相比, 所提出的方法由于引入了类别属性信息的监督, 可以在零样本的跨模态检索精度上取得较高的提升。而 AH 和 TSK 方法是针对单一模态的零样本检索, 虽然能够比传统的跨模态哈希方法更好, 但是在零样本跨模态检索精度上效果不佳, 所提出的方法由于考虑到了不同模态数据特征的异构性, 故能够获得更高的检索精度。与 CMAH、CHOP 及 DAZSH 相比, 由于提出的深层矩阵分解方法能够针对不同模态的深层数据特征进行表述, 所以在检索精度方面取得了很好的效果。



CMFH: 基于协同矩阵分解的哈希方法; LSSH: 基于潜在语义的稀疏哈希方法; TSK: 基于迁移监督知识的哈希方法; AH: 基于属性的哈希方法; CMAH: 基于跨模态属性的哈希方法; CHOP: 基于正交投影的哈希方法; DAZSH: 离散非对称零样本哈希方法; Ours: 本研究方法。

图1 2020年宁夏枸杞虫害数据集上不同哈希方法的 mAP 值

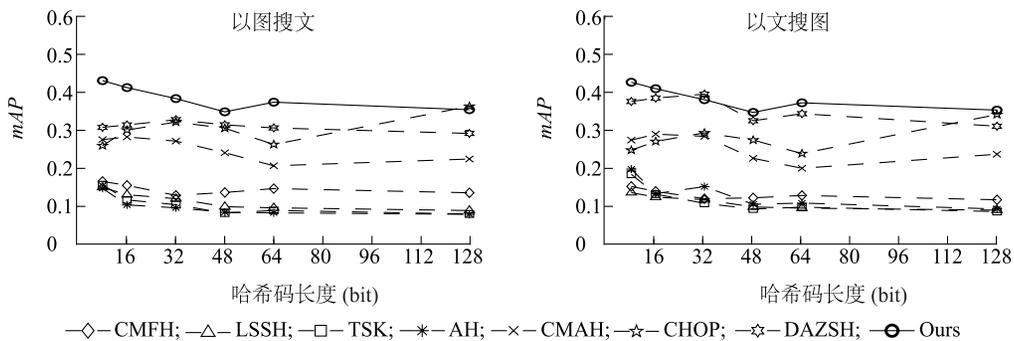
Fig.1 The mAP values of different hashing methods on 2020 Ningxia *Lycium barbarum* pests dataset



CMFH: 基于协同矩阵分解的哈希方法; LSSH: 基于潜在语义的稀疏哈希方法; TSK: 基于迁移监督知识的哈希方法; AH: 基于属性的哈希方法; CMAH: 基于跨模态属性的哈希方法; CHOP: 基于正交投影的哈希方法; DAZSH: 离散非对称零样本哈希方法; Ours: 本研究方法。

图2 Wiki数据集上不同哈希方法的 mAP 值

Fig.2 The mAP values of different hashing methods on Wiki dataset



CMFH: 基于协同矩阵分解的哈希方法; LSSH: 基于潜在语义的稀疏哈希方法; TSK: 基于迁移监督知识的哈希方法; AH: 基于属性的哈希方法; CMAH: 基于跨模态属性的哈希方法; CHOP: 基于正交投影的哈希方法; DAZSH: 离散非对称零样本哈希方法; Ours: 本研究方法方法。

图3 Pascal VOC数据集上不同哈希方法的 mAP 值

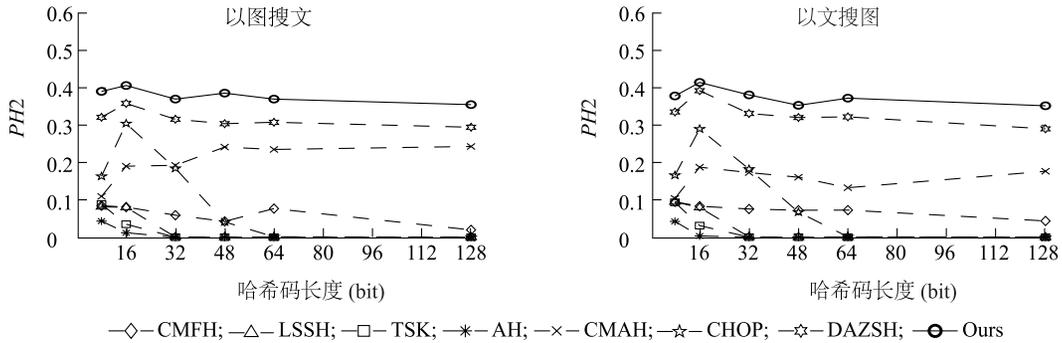
Fig.3 The mAP values of different hashing methods on Pascal VOC dataset

此外,图4、图5、图6显示了本研究所提出的方法与对比方法在3个数据集上的PH2值,从结果可以看出本研究所提出的方法在以文搜图和以图搜文任务上表现良好,验证了所提出的方法的有效性。

图7显示了本研究所提出的方法与对比方法在Wiki数据集上的Top-10准确率,从结果可以看出所提出的方法在以文搜图和以图搜文任务上表现良好,验证了所提出的方法的有效性。图8显示了所提出

的方法与融合注意力方法^[26]在 2020 年宁夏枸杞虫害数据集上的病虫害识别结果对比实例,实线边框为正确识别的结果,虚线边框为识别错误的结果。

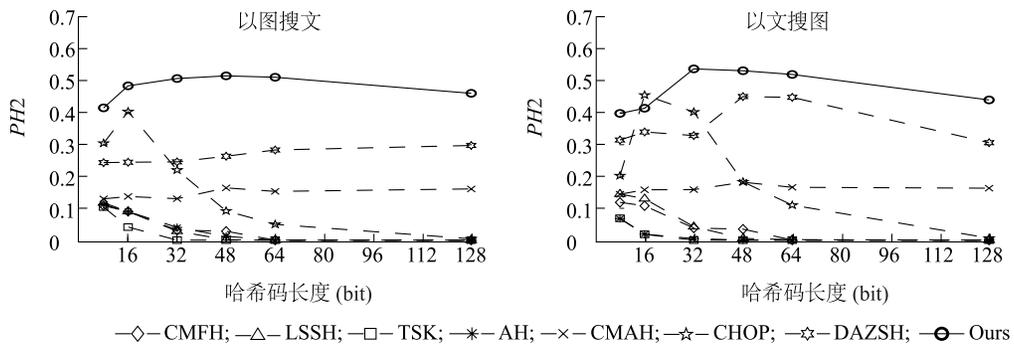
从以文搜图和以图搜文的结果可以看出,本研究所提出的方法在零样本的枸杞病虫害识别方面表现优于对比方法。



CMFH: 基于协同矩阵分解的哈希方法; LSSH: 基于潜在语义的稀疏哈希方法; TSK: 基于迁移监督知识的哈希方法; AH: 基于属性的哈希方法; CMAH: 基于跨模式属性的哈希方法; CHOP 基于正交投影的哈希方法; DAZSH: 离散非对称零样本哈希方法; Ours: 本研究方法。

图 4 2020 年宁夏枸杞虫害数据集上不同哈希方法的 PH2 值

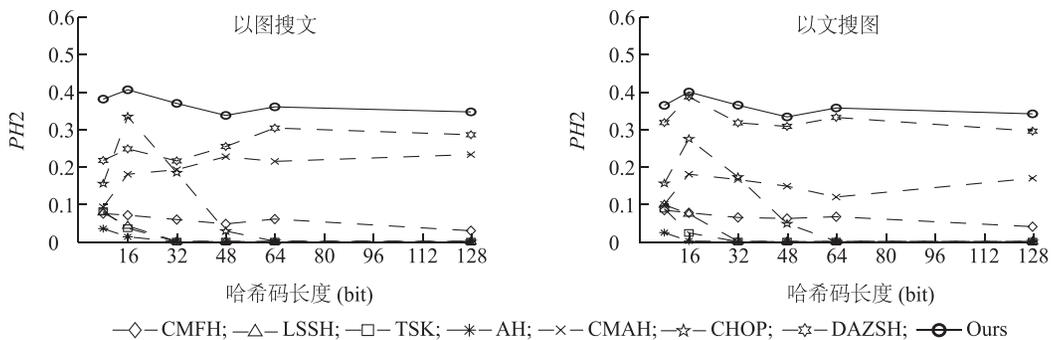
Fig.4 PH2 values of different hashing methods on 2020 Ningxia *Lycium barbarum* pests dataset



CMFH: 基于协同矩阵分解的哈希方法; LSSH: 基于潜在语义的稀疏哈希方法; TSK: 基于迁移监督知识的哈希方法; AH: 基于属性的哈希方法; CMAH: 基于跨模式属性的哈希方法; CHOP 基于正交投影的哈希方法; DAZSH: 离散非对称零样本哈希方法; Ours: 本研究方法。

图 5 Wiki 数据集上不同哈希方法的 PH2 值

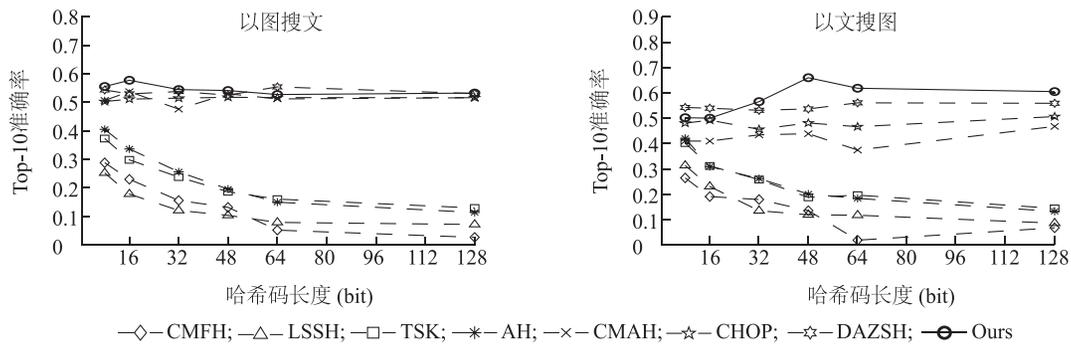
Fig.5 PH2 values of different hashing methods on Wiki dataset



CMFH: 基于协同矩阵分解的哈希方法; LSSH: 基于潜在语义的稀疏哈希方法; TSK: 基于迁移监督知识的哈希方法; AH: 基于属性的哈希方法; CMAH: 基于跨模式属性的哈希方法; CHOP 基于正交投影的哈希方法; DAZSH: 离散非对称零样本哈希方法; Ours: 本研究方法。

图 6 Pascal VOC 数据集上不同哈希方法的 PH2 值

Fig.6 PH2 values of different hashing methods on Pascal VOC dataset



CMFH:基于协同矩阵分解的哈希方法;LSSH:基于潜在语义的稀疏哈希方法;TSK:基于迁移监督知识的哈希方法;AH:基于属性的哈希方法;CMAH:基于跨模态属性的哈希方法;CHOP:基于正交投影的哈希方法;DAZSH:离散非对称零样本哈希方法;Ours:本研究方法。

图7 Wiki数据集上不同哈希方法的Top-10准确率

Fig.7 Top-10 precision of different hashing methods on Wiki dataset

表1显示在Wiki数据集上本研究所提出的方法与CHOP以及DAZSH在训练上所花费的时间,从对比结果可以看出本研究所提出的方法与CHOP相比需要花费较多的训练时间,与DAZSH相比花费较少的训练时间。综合检索精度与训练时长的结果,本研究所提出的方法能够适用于大规模数据集的零样本跨模态检索任务。

2.2 有效性分析

为验证深层矩阵分解的有效性,将本研究方法与不使用深层矩阵分解的方法进行比较。当 m 取值为0时,可以看作对模态数据只进行单层矩阵分解,这里选取了在Wiki数据集上哈希编码为32 bit的试验结果。其 mAP 与 $PH2$ 评价指标结果如图9所示。

从结果可以看出,深层矩阵分解方法在 mAP 和 $PH2$ 指标上相较于单层矩阵分解方法有较大的提升。表明深层矩阵分解方法可以提取到模态数据深层次结构特征,从而提升零样本识别的精确度。

2.3 参数敏感性

对于本研究所提出的方法中的超参数(τ 、 λ 、 α 、 η),如何选择合适的超参数值以获得更好的试验结果。以Wiki数据集下32 bit长度的哈希编码为试验环境,初始将所有的超参数的值设置为1,每次只调整其中1个超参数的值,固定其他参数,对其进行敏感性分析,结果如图10~图13所示。

由图10可以看出 mAP 值在 τ 取值为[2,7]时相对稳定,在[8,9]时下降比较明显。由图11可以看出 mAP 值在 λ 取值为[5,15]时有较大的上升,在

[15,40]时比较稳定。由图12可以看出 mAP 值在 α 取值为[1,4]时较为稳定,在[5,8]时波动且稍微下降。由图13可以看出 mAP 值在 η 取值为[50,350]时相对稳定,在 $\eta=400$ 时以图搜文的 mAP 值有所下降。

3 结论

本研究提出了一种基于零样本学习的枸杞虫害跨模态检索与识别方法,将零样本跨模态哈希方法应用于农作物虫害检索与识别领域,同时解决了现有零样本跨模态哈希方法中存在的问题。通过对不同模态的数据进行深层矩阵分解获得深层结构特征,生成具有判别能力的哈希码。利用类别标签和类别属性信息对生成的哈希码引入线性约束,从而实现不可见类到可见类的知识迁移。使用离散方法学习哈希码避免了连续松弛带来的量化误差,增加了检索时的精度。

在2020年宁夏枸杞虫害图文跨模态检索数据集以及Wiki、Pascal VOC这3个公共数据集上的试验结果显示,与7种现有零样本跨模态哈希方法相比,本研究所提出的方法在以文搜图任务上的平均精度提高了0.2106,在以图搜文任务上的平均精度提高了0.1450,优于所有对比方法。

从试验结果可以看出本研究所提出的方法作为有监督的零样本跨模态哈希方法,在检索精度方面取得了不错的效果,如何将本研究方法扩展到半监督零样本跨模态哈希方法,提升模型的泛用性将是后续研究的方向。

表 1 Wiki 数据集上不同哈希方法训练时长对比

Table 1 Training time of different hashing methods on Wiki dataset

项目	哈希码长度 (bit)					
	8	16	32	48	64	128
CHOP 训练时长 (s)	0.69	0.70	0.71	0.72	0.75	0.99
DAZSH 训练时长 (s)	30.11	32.25	48.96	57.77	54.71	71.41
本研究方法训练时长 (s)	5.52	6.00	6.33	7.09	8.20	13.21

CHOP: 基于正交投影的哈希方法; DAZSH: 基于离散非对称零样本哈希方法。

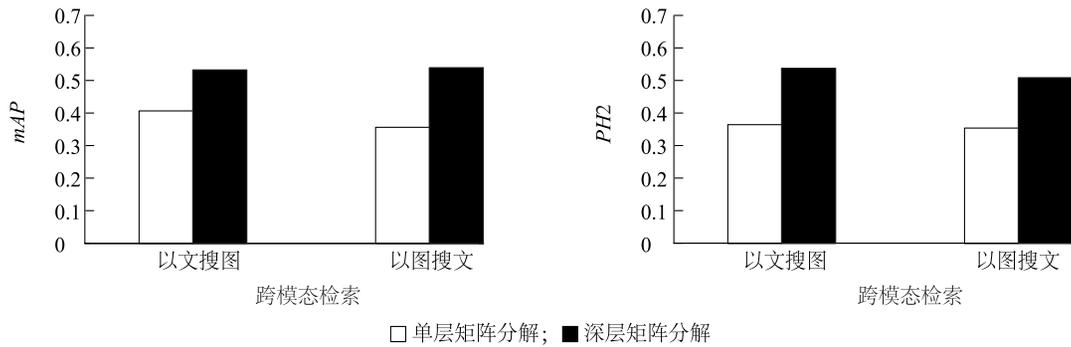


图 9 Wiki 数据集下哈希编码为 32 bit 时的 mAP 和 $PH2$

Fig.9 The mAP and $PH2$ on Wiki dataset (32 bit)

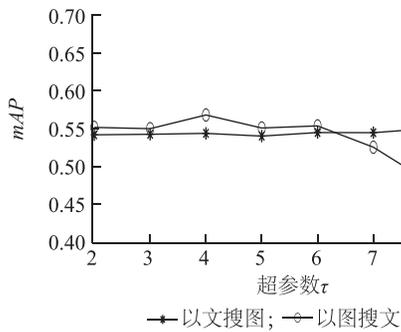


图 10 Wiki 数据集上超参数 τ 的敏感性分析

Fig.10 Sensitivity analysis of hyperparameter τ on Wiki dataset

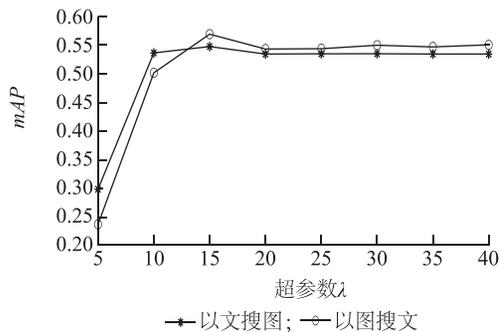


图 11 Wiki 数据集上超参数 λ 的敏感性分析

Fig.11 Sensitivity analysis of hyperparameter λ on Wiki dataset

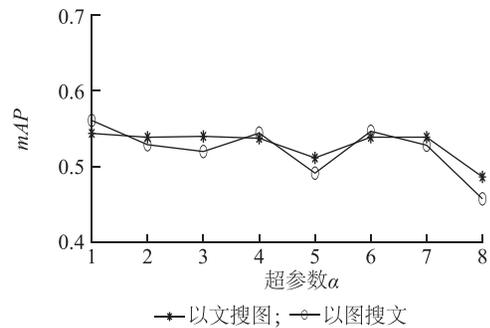


图 12 Wiki 数据集上超参数 α 的敏感性分析

Fig.12 Sensitivity analysis of hyperparameter α on Wiki dataset

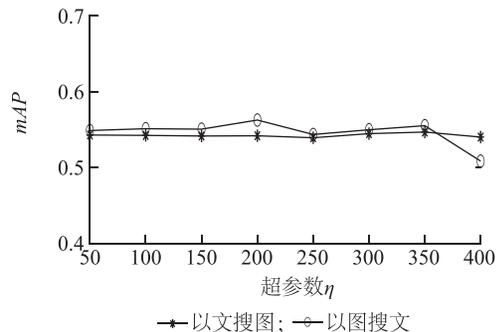


图 13 Wiki 数据集上超参数 η 的敏感性分析

Fig.13 Sensitivity analysis of hyperparameter η on Wiki dataset

参考文献:

- [1] 许盼盼. 枸杞抗盐种质资源筛选与抗盐基因的克隆鉴定[D]. 咸阳:西北农林科技大学,2018.
- [2] 徐峰. 宁夏枸杞产业竞争力研究[D]. 银川:宁夏大学,2017.
- [3] 范振军. 农作物病虫害图像检索方法研究与实现[D]. 绵阳:西南科技大学,2018.
- [4] 汪京京,张武,刘连忠,等. 农作物病虫害图像识别技术的研究综述[J]. 计算机工程与科学,2014,36(7):1363-1370.
- [5] 杭立,车进,宋培源,等. 基于机器学习和图像处理技术的病虫害预测[J]. 西南大学学报(自然科学版),2020,42(1):134-141.
- [6] 赵芸. 基于高光谱和图像处理技术的油菜病虫害早期监测方法和机理研究[D]. 杭州:浙江大学,2013.
- [7] 赵建敏,薛晓波,李琦. 基于机器视觉的马铃薯病害识别系统[J]. 江苏农业科学,2017,45(2):198-202.
- [8] 王佳. 计算机视觉在香芋病害检测中的应用研究[J]. 农机化研究,2020,42(8):241-244.
- [9] NETTLETON D F, KATSANTONIS D, KALAITZIDIS A, et al. Predicting rice blast disease: machine learning versus process-based models[J]. BMC Bioinformatics,2019,20:1-16.
- [10] 王国伟,刘嘉欣. 基于卷积神经网络的玉米病害识别方法研究[J]. 中国农机化学报,2021,42(2):139-145.
- [11] 赵立新,侯发东,吕正超,等. 基于迁移学习的棉花叶部病虫害图像识别[J]. 农业工程学报,2020,36(7):184-191.
- [12] 鲍文霞,吴刚,胡根生,等. 基于改进卷积神经网络的苹果叶部病害识别[J]. 安徽大学学报(自然科学版),2021,45(1):53-59.
- [13] 冯晓,李丹丹,王文君,等. 基于轻量级卷积神经网络和迁移学习的小麦叶部病害图像识别[J]. 河南农业科学,2021,50(4):174-180.
- [14] 谢州益,冯亚枝,胡彦蓉,等. 基于ResNet18特征编码器的小麦病虫害图像描述生成[J]. 农业工程学报,2022,38(12):197-206.
- [15] 彭红星,徐慧明,刘华鼎. 融合双分支特征和注意力机制的葡萄病虫害识别模型[J]. 农业工程学报,2022,38(10):156-165.
- [16] 冀中,汪浩然,于云龙,等. 零样本图像分类综述:十年进展[J]. 中国科学(信息科学),2019,49(10):1299-1320.
- [17] ZHONG F, CHEN Z, MIN G. An exploration of cross-modal retrieval for unseen concepts[C]//LI G L, YANG J, GAMA J, et al. Database systems for advanced applications:24th international conference, Proceedings, Part II. Cham, Switzerland: Springer International Publishing,2019:20-35.
- [18] JI Z, SUN Y, YU Y, et al. Attribute-guided network for cross-modal zero-shot hashing[J]. IEEE Transactions on Neural Networks and Learning Systems,2019,31(1):321-330.
- [19] SHU Z, YONG K, YU J, et al. Discrete asymmetric zero-shot hashing with application to cross-modal retrieval[J]. Neurocomputing,2022,511:366-379.
- [20] 陈磊,刘立波,王晓丽. 2020年宁夏枸杞虫害图文跨模态检索数据集[J]. 中国科学数据,2022,7(3):149-156.
- [21] DING G, GUO Y, ZHOU J. Collective matrix factorization hashing for multimodal data[C]//IEEE. Proceedings of the IEEE conference on computer vision and pattern recognition. Los Alamitos, USA:IEEE,2014:2075-2082.
- [22] ZHOU J, DING G, GUO Y. Latent semantic sparse hashing for cross-modal similarity search[C]//ACM. Proceedings of the 37th international ACM SIGIR Conference on research & development in information retrieval. New York:ACM,2014:415-424.
- [23] XU Y, YANG Y, SHEN F, et al. Attribute hashing for zero-shot image retrieval[C]//IEEE. 2017 IEEE international conference on multimedia and expo (ICME). Hong Kong:IEEE,2017:133-138.
- [24] YANG Y, LUO Y, CHEN W, et al. Zero-shot hashing via transferring supervised knowledge[C]//ACM. Proceedings of the 24th ACM international conference on multimedia. New York:ACM,2016:1286-1295.
- [25] YUAN X, WANG G, CHEN Z, et al. CHOP:an orthogonal hashing method for zero-shot cross-modal retrieval[J]. Pattern Recognition Letters,2021,145:247-253.
- [26] 刘立波,赵斐斐. 融合注意力机制的枸杞虫害图文跨模态检索方法[J]. 农业机械学报,2022,53(2):299-308.

(责任编辑:成纾寒)