

# Deep Learning Based Semantic Scene Image Retrieval

Haijiao Xu, Zhanhong Zhang, Jialei He, Yumin Fang

Department of Computer Science, Guangdong University of Education, Guangzhou Guangdong  
Email: guesskkk99@163.com, 206080110@qq.com, 947440047@qq.com, 190963986@qq.com

Received: July 28<sup>th</sup>, 2019; accepted: August 12<sup>th</sup>, 2019; published: August 19<sup>th</sup>, 2019

---

## Abstract

With the explosive growth of multimedia objects such as Web images over the Internet, online semantic scene image retrieval has been receiving increasing research interest. Conventional studies focus on single-concept-based image retrieval and cannot effectively retrieve semantic scene images including multiple concepts that describe characteristic semantic scene. To tackle this issue, *i.e.*, semantic scene Web image retrieval, we propose a novel approach called multi-modal deep learning based Semantic Scene Image Retrieval (SSIR) in this paper. In particular, we first train a multi-modal Convolutional Neural Network (CNN) as a concept classifier for images and texts. Second, semantic interdependencies of the sub-concepts included in the images are utilized to refine the predicted semantic scores in order to enhance holistic scene recognition. Finally, to improve the performance of retrieving rare scene concepts, a gradient descent algorithm is used for compensating the varying frequencies of concepts derived from imbalanced image datasets. The results of our experiments on MIR Flickr 2011 have shown that our proposed approach performs favorably compared with several traditional methods.

## Keywords

Semantic Scene Image Retrieval, Convolutional Neural Network, Deep Learning

---

# 基于深度学习的语义场景图像检索

徐海蛟, 张展鸿, 何佳蕾, 方钰敏

广东第二师范学院, 计算机科学系, 广东 广州

Email: guesskkk99@163.com, 206080110@qq.com, 947440047@qq.com, 190963986@qq.com

收稿日期: 2019年7月28日; 录用日期: 2019年8月12日; 发布日期: 2019年8月19日

---

## 摘要

随着互联网图像等多媒体内容的爆炸式增长, 在线Web图像的语义场景检索问题引起了学者们的研究兴

趣。传统的研究工作聚焦在基于单概念的图像检索上,未能很好检索含有复杂语义场景的图像。为解决语义场景Web图像检索问题,我们提出了一种基于多模态深度学习的语义场景图像检索方法(SSIR)。首先,使用一个多模态CNN训练网络作为概念分类器;其次,通过计算语义概念之间的依赖关系来精炼概念的语义分数,以进一步增强分类器的场景识别能力;最后,为提升对稀疏场景概念的检索性能,应用梯度下降算法来补偿在真实应用中不平衡图像集上语义概念的频率差。在MIR Flickr 2011标准图像数据集上对比了其他传统方法,结果表明我们的语义场景检索方法性能更优。

## 关键词

语义场景图像检索, 卷积神经网络, 深度学习

Copyright © 2019 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



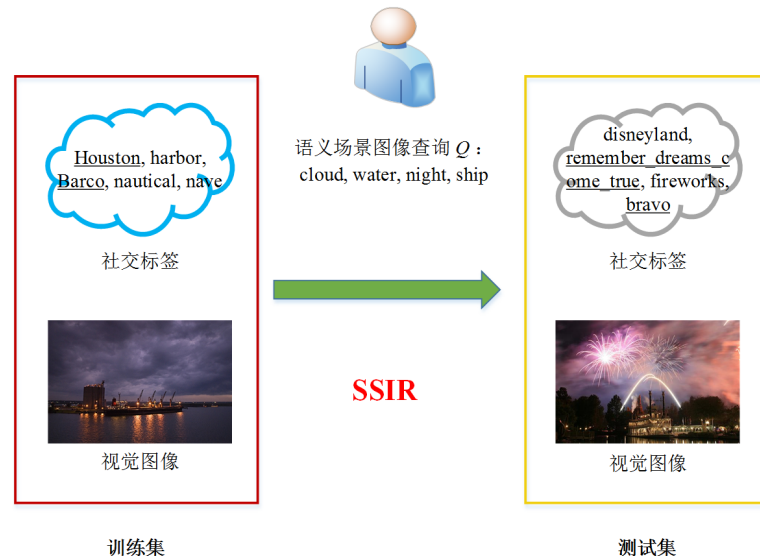
Open Access

## 1. 引言

在 Web 2.0/3.0 时代, 社交媒体获得了前所未有的发展, “互动”是图像等社交多媒体最重要的传播特征和要求, 用户常常自由地给予这些 Web 图像共享的文本社交标签(社会化标签) [1] 或者用户评论。这样的多媒体内容已经大量地出现在人们的日常生活中, 所以亟须有效的检索方法帮助人们寻找到目标多媒体内容。Web 图像往往伴随着若干个文本标签, 它们是由图像用户即兴给定的关键字或者文本描述符, 可能部分或者间接地描述了图像的内容, 是带有高噪音的弱语义信息, 存在着语义表达不完整、语义错误、语义歧义、语义重复等问题, 所以其不同于由专业人士给出的强语义标注或者类标签。一个直截了当的图像检索方法是直接在图像社交标签上执行图像检索, 然而图像社交标签往往在语义上是高噪音的, 并且可能并不会直接地与图像内容有关联, 从而导致较低的图像检索性能。如图 1 所示, 在测试图像 *I* 中存在一些带下划线的社交标签例如 “remember\_dreams\_come\_true” 和 “bravo” 并没有直接描述该图像 *I* 的视觉内容, 而训练图像的社交标签例如 “Houston” 和 “Barco” 仅部分描述了图像内容。虽然这些噪音文本社交标签不适宜直接用于图像查询概念, 但是可以把它们看作是额外的、有价值的低层文本特征即文本模态数据以用于图像的监督学习。

对于在线图像检索, 传统的单模态方法要么仅使用了图像模态数据 [2], 要么仅使用了文本模态数据 [3]。由于文本模态数据例如社会化标签可以用来增强多概念图像检索的性能, 因此一些多模态跨模态检索方法 [4] 研究了图像与文本等多个模态数据之间的关联以提升图像检索的性能。然而, 这些传统的方法聚焦于基于单概念的图像检索而较少考虑多概念语义场景图像检索场景, 这限制了其应用范围, 因为用户使用检索系统的时候往往提交多概念的语义场景检索请求。为了处理语义场景图像检索, 传统方法在单概念检索系统上执行多概念语义场景检索任务, 这在一些场景中可能是低效的, 这是因为多概念场景有其独特的视觉外观特性, 这难以仅仅依靠单概念分类器去识别。于是, 进一步研究语义场景图像检索有用且有必要。

近年来, 卷积神经网络(CNNs) [5] 作为一种重要的深度学习技术已广泛应用于基于单概念的图像检索。CNNs 占据了检索测试基线的排行榜, 这表明通过 CNNs 学习到的深度特征能较好地表征图像的内部结构。因此, 考虑使用深度学习技术去提升语义场景图像检索的性能。我们提出的方法 SSIR 整合了两类分类器: 1) 多概念场景分类器(有助于整体场景识别); 2) 单概念目标分类器(有助于单概念识别)。从实



**Figure 1.** The graphical illustration of semantic scene image retrieval  
**图 1.** 语义场景图像检索示例图

验结果看，这种设计能显著提高分类器对多概念语义场景的判断识别能力。具体来说，首先，设计一个多模态 CNN 以识别多概念语义场景。每个 CNN 分为卷积块层与全连接分类器层。学习时，图像(视觉模态)与关联文本(文本模态)被分别送入各自模态的卷积块层与全连接分类器层。全连接分类器层包含两类分类器，即适合单概念识别的单概念目标分类器以及有助于整体场景识别的多概念场景分类器。第二，语义概念之间的依赖关系被用来计算概念的语义分数，以进一步增强分类器的判断识别能力。假若一个概念  $c$  与其关联概念们以较高的频率共现在图像集中，一旦发现关联概念们频繁出现， $c$  的预测分数值将被提高。最后，通过融合运算，SSIR 方法整合了视觉模态和文本模态的预测分数值。为改善对稀疏概念的检索性能，应用梯度下降算法来补偿在真实应用中不平衡图像集上语义概念的频率差。

## 2. 基于多模态深度学习的语义场景图像检索

### 2.1. 问题描述

设多模态图像训练集为  $L = \{(I_1, T_1), \dots, (I_N, T_N)\}$ ，多模态图像测试集为  $U$ 。每个图像文本对  $(I_i, T_i) \in L$  或者  $(I, T) \in U$  表示一幅图像及其关联的若干个弱语义文本标签，分别由低层视觉特征和低层文本特征组成。给定一个包含有  $K$  个不同语义单概念的强语义单词表  $V = \{c_1, \dots, c_i, \dots, c_k\}$ ，其中每一个语义概念  $c_i \in V$  是一个语义单概念(例如“rainbow”或者“wedding”)。在训练集  $L$  中的每幅图像被标注若干个语义单概念  $c_i$ ，而测试集  $U$  中的图像没有任何强语义标注。每个语义场景多概念  $C_i = \{c_1, \dots, c_k\}$  是  $V$  的幂集中的一个元素即  $C_i \in 2^V$  或者  $C_i \subseteq V$ ，其中  $k$  是  $C_i$  的长度即  $k = |C_i|$ 。给定一个多概念语义场景查询  $Q = \{c_1, \dots, c_t\} \in 2^V$  和带强语义标注的训练集  $L$  以及未标注图像测试集  $U$ ，目标是寻找包含所有  $t$  个目标单概念的最相关图像  $I \in U$ 。

### 2.2. 系统框架

SSIR 检索框架如图 2 所示，包含 3 个部分：卷积块层、分类器层、在线检索。卷积块层学习出两类特征：深度视觉特征与深度文本特征。通过视觉卷积块层，将图像像素转换为视觉特征向量；通过文本卷积块层，将关联的文本社交标签转换为词嵌入[6]，然后学习出文本特征向量。分类器层学习出从特征到概念的映射函数，包含两种分类器：单概念目标分类器和多概念场景分类器。这里，需要生成多概念单词表  $V^*$ 。由于  $V^*$  中的概念分为语义单概念和场景多概念两类，所以共有四类映射函数：单概念视觉映

射函数、多概念视觉映射函数、单概念文本映射函数、多概念文本映射函数。在线检索根据给定的多概念语义场景查询  $Q$  从测试集中检索到最相关的图像结果。由于要使用多概念之间的语义依赖关系，因此需要生成查询上下文  $R_{rc}(Q)$ 。

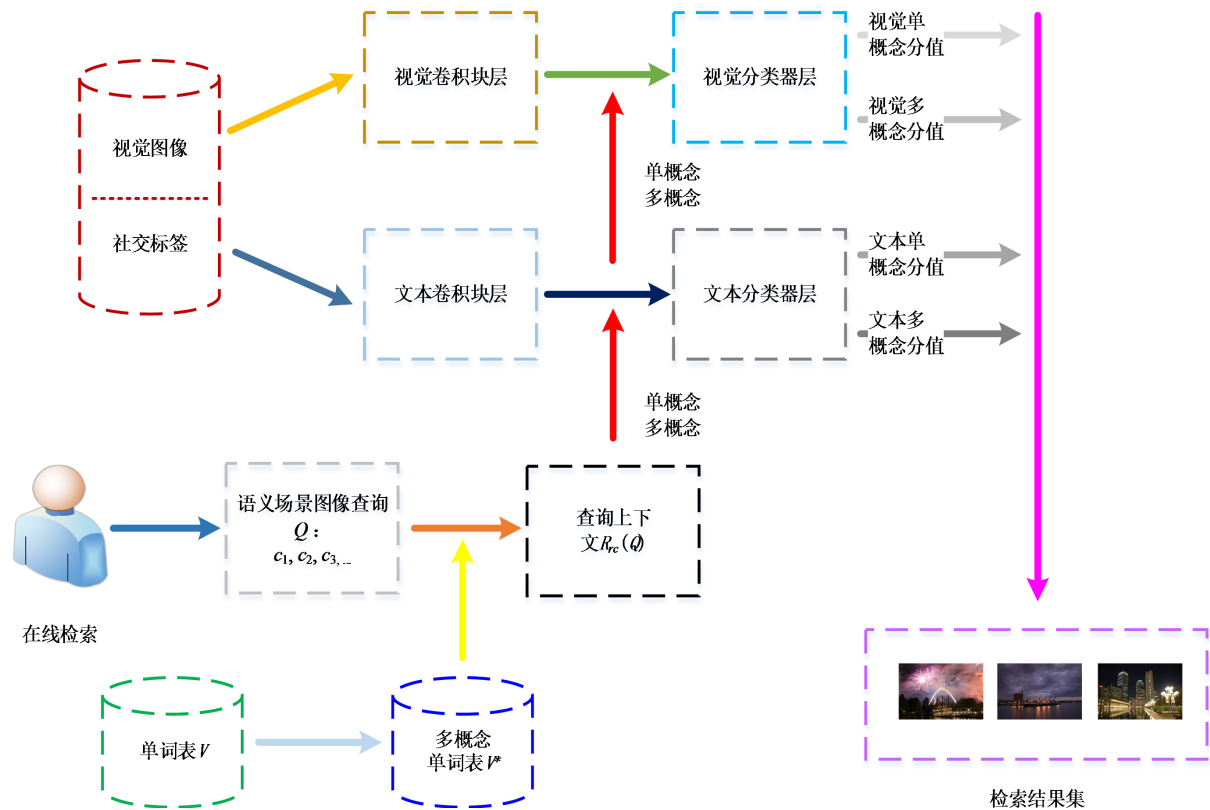


Figure 2. The SSIR system framework  
图 2. SSIR 系统框架

每个多概念  $C_i \in V^*$  可视为单个抽象场景概念  $C_i$ ，也可视为  $k$  个分离的单概念  $c_i$ 。对于多概念单词表  $V^*$  的生成，为避免产生无意义的语义概念排列组合，SSIR 模型基于训练集  $L$  上的共现规则选择出有场景语境的多概念  $C_i$  用以生成一个强语义多概念单词表  $V^*$ ：

$$|C_i| \leq t \tag{1}$$

$$|N_r(c_1, \dots, c_k)| \geq s \tag{2}$$

其中  $N_r(c_1, \dots, c_k)$  表示所有  $k$  个语义概念  $c_i$  在训练集  $L$  中的共现总数，即多概念频率。如果集合  $V^*$  太大，可以控制共现阈值  $s$  以减少训练开销。

### 2.3. 多模态 CNN 结构

任何 CNN 网络都可融入我们的检索模型，不失一般性，我们选择了近年来一个有影响力的高效 CNN 模型 ResNet [7] 来作为我们的视觉模态 CNN 分类器。由于传统 CNN 分类器网络聚焦于单概念分类，而我们的语义场景检索任务是一个多概念分类任务，因此为 ResNet 模型定义了一个新的多概念 softmax 损失函数使之适应多概念语义场景检索任务。首先，第  $i$  张图片  $I$  与第  $j$  个概念  $C_j \in V^*$  的归一化关联概率可定义为：

$$p(C_j | I) = \frac{\exp(q_j(I))}{\sum_k \exp(q_k(I))}, \quad (3)$$

其中,  $q_j(I)$  是图像  $I$  在第  $j$  个概念  $C_j$  的离散概率分布, 它由 ResNet 分类器产生。为最小化 ResNet 预测概率与真实概率的 KL 距离, 我们使用如下场景多概念 softmax 损失函数:

$$f_{\text{softmax}} = -\frac{1}{N} \sum_i \sum_j \bar{p}_{i,j} \log(p(C_j | I)), \quad (4)$$

其中,  $\bar{p}_{i,j}$  是一个图片  $I$  的指示器函数: 当概念  $C_j$  在图片  $I$  中存在则  $\bar{p}_{i,j} = 1$  否则  $\bar{p}_{i,j} = 0$ ;  $N$  是图像总数。传统的单概念 CNN 基本结构为卷积块层后接分类器层。卷积块层学习出深度特征, 而分类器层执行从  $I$  到  $C_j$  的映射, 识别出语义概念  $C_j$ 。我们设计的 SSIR 场景检索模型包含两类分类器: 单概念分类器和多概念场景分类器。当  $|C_j| = 1$  时,  $C_j$  为一个传统的单概念; 当  $|C_j| > 1$  时,  $C_j$  为一个场景多概念, 其作为一个整体概念参加学习。

对于文本 CNN 网络, 很多适合自然语言处理的 CNN 网络都可融入我们的场景检索模型, 不失一般性, 我们选择了一个有影响力的高效 CNN 模型 SentenceCNN [8] 作为我们的文本模态 CNN 分类器。为训练 SentenceCNN, 将一副图像所有的文本标签视为一个输入语句加以学习。类似视觉模态 CNN 分类器, 修改分类器层, 增加一种多概念场景分类器, 于是, 文本模态 CNN 分类器也包含两类分类器: 单概念分类器和多概念场景分类器。

#### 2.4. 语义场景查询上下文 $R_{rc}(Q)$ 的生成

首先生成语义邻居集  $R(Q) \subset V^*$ , 通过选择出概率  $p(Q|C_i) > 0$  的邻居概念  $C_i$ , 这个对称的语义概率  $p(Q|C_i)$  表示两个概念  $Q$  和  $C_i$  之间的相关性, 它可以基于多模态图像训练集  $L$  如下定义:

$$p(Q|C_i) = \frac{2 \times Nr(Q, C_i)}{Nr(Q) + Nr(C_i)} \quad (5)$$

其中  $Nr(Q)$  与  $Nr(C_i)$  表示多概念  $Q$  与  $C_i$  的共现频率,  $Nr(Q, C_i)$  表示同时包含有两个概念  $Q$  与  $C_i$  的图像数, 每一个多概念  $C_i$  被看做是自己的语义邻居并且服从约束条件  $p(C_i|C_i) = 1$ 。

其次, 选择出所有的查询零件  $C_i \in Q$  到集合  $R_{rc}(Q)$ 。最后从剩余的语义概念中选择出最相关的  $r$  个概念  $C_r$  到集合  $R_{rc}(Q)$ , 于是包含  $K_{rc}$  个元素的查询上下文  $R_{rc}(Q)$  就生成了。后续实验中, 在验证集上将  $K_{rc}$  试验取值范围为 2~20 执行交叉验证, 发现当设置  $K_{rc} = 8$  时性能最佳, 因此  $K_{rc} = 8$  作为后续实验的默认值。为了保持语义相关性的概率属性, 语义关联概率  $p(Q|C_i)$  被如下归一化:

$$p(Q|C_i) = \begin{cases} \frac{p(Q|C_i)}{\sum_{r=1}^{K_{rc}} p(Q|C_r)} & \text{if } C_i, C_r \in R_{rc}(Q) \\ 0 & \text{elsewhere.} \end{cases} \quad (6)$$

#### 2.5. 多模态场景语义映射

对于查询概念  $Q$ , 上述  $K_{rc}$  个有语义相关性的元素  $C_i \in R_{rc}(Q)$  联合参与未标注多模态图像文本对  $(I, T)$  的相关性计算, 输出相关性估值  $r(Q, (I, T))$ :

$$r(Q, (I, T)) = r(Q, I) + r(Q, T) \quad (7)$$

$$r(Q, I) = \alpha_1 \times r_s(Q, I) + \alpha_2 \times r_m(Q, I) \quad (8)$$

$$r(Q, T) = \alpha_3 \times r_s(Q, T) + \alpha_4 \times r_m(Q, T) \quad (9)$$

其中,  $r(Q, I)$ 和 $r(Q, T)$ 分别表示图像 $I$ 和文本 $T$ 各自模态关于查询概念 $Q$ 的相关性估值;  $r_s(Q, \cdot)$ 和 $r_m(Q, \cdot)$ 分别表示 CNN 单概念分类器和 CNN 多概念分类器输出的单模态相关性估值;  $\{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ 是 SSIR 场景检索方法的待优化参数, 且服从约束条件:  $\alpha_1 + \alpha_2 + \alpha_3 + \alpha_4 = 1$ 。公式(7)体现了多模态预测数据的后期融合, 公式(8)与(9)体现了单概念分类器与多概念场景分类器预测数据的后期融合。  $r_s(Q, \cdot)$ 和 $r_m(Q, \cdot)$ 相关性估值的计算如下, 体现了 SSIR 场景检索方法对于语义相关性的利用。

$$r_s(Q, I) = \prod_{i=1}^t p(c_i | I), \quad (10)$$

$$r_m(Q, I) = \sum_{i=1}^{K_{rc}} p(Q | C_r) \times p(C_r | I), \quad C_r \in R_{rc}(Q), \quad (11)$$

$$r_s(Q, T) = \prod_{i=1}^t p(c_i | T), \quad (12)$$

$$r_m(Q, T) = \sum_{i=1}^{K_{rc}} p(Q | C_r) \times p(C_r | T), \quad C_r \in R_{rc}(Q), \quad (13)$$

其中, 后验概率  $p(C_i | I)$ 和 $p(C_i | T)$ 经由视觉 CNN 多概念分类器和文本 CNN 多概念分类器计算; 语义相关性  $p(Q | C_r)$ 可以看做是后验概率  $p(C_i | I)$ 和 $p(C_i | T)$ 的权重。

## 2.6. SSIR 参数估算

为找出较好的参数  $\alpha = \{\alpha_1, \alpha_2, \alpha_3, \alpha_4\}$ , 使用在训练数据上的极大对数似然函数方法。令  $y_{Q_i} \in \{0, 1\}$  表示语义场景多概念  $Q$  是否出现在多模态图像数据  $(I_i, T_i) \in L$  中, 相关性预测值  $p(y_{Q_i})$  如下给出:

$$p(y_{Q_i} = 1) = r(Q, (I, T)), \quad (14)$$

$$p(y_{Q_i} = 0) = 1 - r(Q, (I, T)), \quad (15)$$

$$p(y_{Q_i}) = p(y_{Q_i} = 1)^{y_{Q_i}} p(y_{Q_i} = 0)^{1 - y_{Q_i}} \quad (16)$$

因此, 检索概念  $Q$  的对数似然函数被如下改写:

$$L_Q = \sum_{i=1}^N n_{Q_i} \log p(y_{Q_i}) \quad (17)$$

其中  $n_{Q_i}$  是正反例数目  $N^+$  和  $N^-$  的不平衡性惩罚因子。假若  $y_{Q_i} = 1$ , 设置  $n_{Q_i} = 1/N^+$ ; 否则,  $n_{Q_i} = 1/N^-$ 。将公式(17)中的  $p(y_{Q_i})$  项用公式(14)到(16)替换, 可以获得下列对数似然函数:

$$L_Q = \sum_{i=1}^N n_{Q_i} \log \left\{ (\alpha\beta)^{y_{Q_i}} (1 - \alpha\beta)^{1 - y_{Q_i}} \right\} \quad (18)$$

其中,

$$\beta = \left\{ \prod_{i=1}^t p(c_i | I), \sum_{i=1}^{K_{rc}} p(Q | C_r) \times p(C_r | I), \prod_{i=1}^t p(c_i | T), \sum_{i=1}^{K_{rc}} p(Q | C_r) \times p(C_r | T) \right\}.$$

通过梯度下降法[9]最大化对数似然函数  $L_Q$  求出  $\alpha$  优化值。

## 3. 实验和评价

### 3.1. 数据集

评价实验采用了公开数据集 MIR Flickr 2011 [10]。它包含有 18,000 张图像, 每张图像含有 3~26 个

标注, 单词表  $V$  含有 99 个语义概念。每个图像  $I$  与其关联的文本标签  $T$  组成了多模态图像, 即图像文本对  $(I, T)$ 。采用随机抽样, 8000 张图像作为训练集, 余下 10,000 张图像作为测试集, 约 70% 概念频率低于平均概念频率。很明显, 这是非平衡图像数据集, 增大了场景识别难度。当一幅图像  $I$  包含  $Q$  中所有的单概念  $c_i$  时, 则  $I$  是相关的, 否则  $I$  是不相关的。采用平均准确率均值(MAP)分数作为性能评价指标, 所有指标值越高表示场景检索性能越好。

### 3.2. 实验结果与分析

由于 MIR Flickr 2011 数据集中的概念最高频数是 11, 因此设定公式(1)的参数  $t = 11$ 。  $s$  控制计算  $V^*$  的计算开销, 经验设置  $s = 200$ 。于是,  $V^*$  含有 15,970 个场景多概念。为测试语义场景检索性能, 我们构建了检索测试集  $Q^*$ : 所有的单概念  $c_j \in V$  加入  $Q^*$ ; 随机生成 500 个双概念场景查询; 随机生成 500 个三概念场景查询; 随机生成 500 个四概念场景查询。这样,  $Q^*$  共计包含 1599 个语义场景查询。

表 1 列出了与最新图像标注方法的对比实验结果。对于被比较的传统单概念检索方法, 采用公式(10)和(12)计算语义多概念估值。

**Table 1.** Performance comparisons of semantic scene image retrieval

**表 1.** 语义场景图像检索性能比较

检索方法	全部概念	双概念	三概念	四概念
GResNets [11]	0.23	0.25	0.18	0.17
DANE [12]	0.22	0.24	0.16	0.18
AlexNet 串联并联[13]	0.18	0.21	0.15	0.15
VGGNet-Dropout [14]	0.18	0.20	0.15	0.14
VGGNet + KNN [15]	0.18	0.21	0.15	0.14
VGGNet + SVM [15]	0.18	0.21	0.14	0.13
MKL [16]	0.16	0.19	0.13	0.12
Autoencoder [17]	0.19	0.23	0.16	0.15
DBM [17]	0.20	0.23	0.16	0.16
SSIR (Ours)	<b>0.26</b>	<b>0.28</b>	<b>0.24</b>	<b>0.23</b>

从该表中可见, 我们的 SSIR 方法超越了其他对比方法, 获得了更好的场景检索性能。与表中最好的对比检索方法 GResNets 比较, 提出的 SSIR 方法的全部场景概念 MAP 提高了 13%。一方面, 多概念场景分类器有助于识别语义场景, 另一方面, 概念间语义依赖关联性更能准确判断多概念形成的语义场景, 缓解误判率。在图像检索任务中, 这两种类型的相关性都提供了有用的信息, 具有一定的互补性, 从这个角度上说我们的 SSIR 方法可提高图像检索的性能。此外, 应用梯度下降算法来补偿在真实应用中不平衡图像集上语义概念的频率差, 缓解了稀疏场景概念的误判率, 所以, 在不平衡数据集 MIR Flickr 2011 上, 我们的 SSIR 方法具有更好的场景检索效果。

### 基金项目

2019 年广东第二师范学院校级教学质量与教学改革工程项目(编号: 2019jxgg18); 广东第二师范学院软件工程重点学科建设项目(编号: 9030-1700207); 广东省自然科学基金项目(编号: 2018A0303130169); 广东省科技计划项目(编号: 粤财农[2017]94 号, 2016A010106007); 广东省应用型科技研发专项资金资助项目(编号: 2016B090927010); 2019 年广东第二师范学院大学生创新创业训练计划项目(编号:

201914278146)。

## 参考文献

- [1] 熊回香, 叶佳鑫. 基于同义词词林的社会化标签等级结构构建研究[J]. 情报杂志, 2018, 37(1): 126-131.
- [2] 龚彦宇, 董剑利, 侯明亮. 综合颜色特征的图像检索方法研究[J]. 计算机科学与应用, 2016, 6(10): 583-589.
- [3] Fang, Q., Xu, C. and Sang, J. (2016) Folksonomy-Based Visual Ontology Construction and Its Applications. *IEEE Transactions Multimedia*, **18**, 702-713. <https://doi.org/10.1109/TMM.2016.2527602>
- [4] Wang, W., Yang, X. and Ooi, B.C. (2016) Effective Deep Learning Based Multi-Modal Retrieval. *The VLDB Journal*, **25**, 79-101. <https://doi.org/10.1007/s00778-015-0391-4>
- [5] 杨泽明, 刘军, 薛程, 于子红. 卷积神经网络在图像分类上的应用综述[J]. 人工智能与机器人研究, 2018, 7(1): 17-24.
- [6] 叶天顺. 一种改进的社交词嵌入算法[J]. 计算机应用与软件, 2018(9): 132-137.
- [7] He, K., Zhang, X., Ren, S. and Sun, J. (2016) Deep Residual Learning for Image Recognition. *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 27-30 June 2016, 770-778. <https://doi.org/10.1109/CVPR.2016.90>
- [8] Kim, Y. (2014) Convolutional Neural Networks for Sentence Classification. *ACL International Conference on Empirical Methods in Natural Language Processing*, Doha, 25-29 October 2014, 1746-1751. <https://doi.org/10.3115/v1/D14-1181>
- [9] 陶秉墨, 鲁淑霞. 基于自适应随机梯度下降方法的非平衡数据分类[J]. 计算机科学, 2018, 45(z1): 487-492.
- [10] Nowak, S., Nagel, K. and Liebetrau, J. (2011) The CLEF 2011 Photo Annotation and Concept-Based Retrieval Tasks. *CLEF Conference and Labs of the Evaluation Forum*, Amsterdam, 19-22 September 2011, 1-25.
- [11] 李鹏, 蒋品群, 曾上游, 夏海英, 廖志贤, 范瑞. 基于分组残差结构的轻量级卷积神经网络设计[J]. 微电子学与计算机, 2019, 36(7): 43-47.
- [12] 王雅慧, 刘博, 袁晓彤. 基于近似牛顿法的分布式卷积神经网络训练[J]. 计算机科学, 2019, 46(7): 180-185.
- [13] 张涛, 杨剑, 宋文爱, 郭雁蓉. 改进卷积神经网络模型设计方法[J]. 计算机工程与设计, 2019, 40(7): 1885-1890.
- [14] 程俊华, 曾国辉, 鲁敦科, 黄勃. 基于 Dropout 的改进卷积神经网络模型平均方法[J]. 计算机应用, 2019, 39(6): 1601-1606.
- [15] Simonyan, K. and Zisserman, A. (2014) Very Deep Convolutional Networks for Large-Scale Image Recognition.
- [16] Guillaumin, M., Verbeek, J.J. and Schmid, C. (2010) Multimodal Semi-Supervised Learning for Image Classification. *The 23rd IEEE Conference on Computer Vision and Pattern Recognition*, San Francisco, 13-18 June 2010, 4307-4311. <https://doi.org/10.1109/CVPR.2010.5540120>
- [17] Srivastava, N. and Salakhutdinov, R. (2014) Multimodal Learning with Deep Boltzmann Machines. *Journal of Machine Learning Research*, **15**, 2949-2980. [https://doi.org/10.1162/NECO\\_a\\_00311](https://doi.org/10.1162/NECO_a_00311)

**Hans 汉斯**

### 知网检索的两种方式:

1. 打开知网首页: <http://cnki.net/>, 点击页面中“外文资源总库 CNKI SCHOLAR”, 跳转至: <http://scholar.cnki.net/new>, 搜索框内直接输入文章标题, 即可查询;  
或点击“高级检索”, 下拉列表框选择: [ISSN], 输入期刊 ISSN: 2161-8801, 即可查询。
2. 通过知网首页 <http://cnki.net/>顶部“旧版入口”进入知网旧版: <http://www.cnki.net/old/>, 左侧选择“国际文献总库”进入, 搜索框直接输入文章标题, 即可查询。

投稿请点击: <http://www.hanspub.org/Submission.aspx>

期刊邮箱: [csa@hanspub.org](mailto:csa@hanspub.org)