DOI：10.16667/j.issn.2095-1302.2020.09.016

基于提取标签显著性区域的深度学习图像检索方法

田少骅，胡琦瑶，蒙泽新，王炅昱
（西北大学 信息科学与技术学院，陕西 西安 710127）

摘　要：现今主流的图像检索技术需人工添加数据集标签后，方可对深度学习网络进行训练。但人为添加标签会花费大量时间及精力，并且对图片标签的描述有一定的条件性和规则性，因此添加的标签不能够很好地代表大众的视觉习惯，不利于神经网络的深度训练，得到的结果准确率不尽如人意。为解决这些问题，文中提出了一种基于提取标签显著性区域的深度学习图像检索方法。首先，对图片集进行网络标签粗过滤，去除与图片无关的噪声标签；其次，提取已知图像的显著性区域，对显著性区域标签进行向量化；将确定显著性区域图像的三元组作为 VGG16 网络的输入，通过三元组目标函数对图像进行学习；通过汉明距离的判断，得到检索的相似图片。实验对比表明，现有方法的准确率相比原有方法准确性提高了 3%。

关键词：显著性区域；标签向量化；word2vec；图像三元组；图像检索；哈希编码
中图分类号：TP391.41　　**文献标识码**：A　　**文章编号**：2095-1302（2020）09-0054-04

0　引　言

传统图像检索技术缺点明显，如人工工作量大，且一幅图像只有一个哈希编码，无法准确代表人类的视觉习惯，同时检索精确度大大降低。现有的网络用户图像分享系统在快速发展，如 Flickr，Photobucker，Picasa WebAlbums，这些系统不仅提供了大量且丰富的数据标签，更因其标签具有人类视觉的习惯特点，有利于更好地进行深度网络训练。基于此，本文改进了传统的网络检索技术，对已知数据库的图片进行初步噪声过滤，提取显著性区域的向量标签，并对其进行深度网络训练得到哈希编码，以提高检索精确度。利用数据集图片哈希编码和被检索图片之间哈希编码的汉明距离来判断图像的相似度，从而得到被检索的图像信息。这种基于提取显著性区域网络标签的深度学习图像检索方法将大大提高检索精度，减少人工的工作量。

1　粗过滤网络标签

考虑到使用 NUS-WIDE 数据集直接过滤网络标签十分困难，所以本文将网络标签的过滤问题转换为视觉内容与图像标签相关度排序问题。本文采用 Aixin Sun[1] 等人提出的内聚性和分散性方法来判断标签是否可以被剔除。该方法的核心思想：在数据集 NUS-WIDE 中，若有已知特定标签的图像集 P，那么从图像集 P 中任意抽取的一组图像集 C，他们之间的相似度一定小于图像集 P 之间的相似度。图像越相似，底层特征的特征向量距离越小，计算过程如下。

（1）使用 K-means 聚类算法分别迭代整个图像集 P、聚类中心 C_P 以及含有具体标签 d 的图像集 p 的聚类中心 C_{pd}。

（2）计算 P 的分散距离 sd_P 和内聚距离 cd_P。

（3）从 P 中随意抽取 N 个图片，并计算其内聚距离 cd_C，以及 C 与 P 的分散距离 sd_C。

（4）如果 P 的内聚距离 cd_P 小于或等于 C 的内聚距离 cd_C，且 P 的分散距离 sd_P 大于或等于 C 与 P 的分散距离 sd_C，则保留该标签；否则，剔除该标签。

内聚距离计算公式：

$$\mathrm{cd}=\sum_{i\in q}\cos\theta\left[i,C_{pd}\right] \tag{1}$$

分散距离计算公式：

$$\mathrm{sd}=\cos\theta\left[C_{pd},C_P\right] \tag{2}$$

$$\cos\theta=\frac{A\cdot B}{|A|\cdot|B|} \tag{3}$$

图 1 所示为数据集中两幅图片的标签过滤前后的对比结果。

过滤前	过滤后	过滤前	过滤后
Red Vegetable People Four Tomatoes Stilllife	Red Tomatoes Four	Guitar People Youth Lawn American Earphone	Youth Guitar Lawn

图 1　网络标签粗过滤结果对比

收稿日期：2020-03-13　　**修回日期**：2020-04-15
基金项目：西北大学大学生创新创业训练计划项目（2020387）

2 标签显著性区域提取

2.1 显著性区域集合

采用二值化赋范特性（BING）[2] 算法提取图片显著性区域。一个图片可以拥有多个显著性区域，考虑到不同的人看同一幅图片的重点会有所不同，所以多个显著性区域会大大增强检索技术的精确度，而 BING 算法很好地将整幅图像的检索技术转换为基于显著性区域的检索技术。

使用全局特征信息（GIST）[3] 算法提取每一个显著性区域的特征向量，得到其 512 维的特征向量。为了挖掘图片的相似度，通过每两个显著性区域的 512 维特征向量的欧氏距离来判断样本之间的相似度。计算公式如下所示：

$$\mathrm{dist}(X,Y)=\sqrt{\sum_{i=1}^{n}(X_i-Y_i)^2} \tag{4}$$

提取的显著性区域的欧氏距离越小则相似度越高，当其值小于设定的阈值时，则判定 2 个显著性区域之间为相似。通过计算，形成显著性区域的相似集合和不同显著区域的集合。

本文将 NUS-WIDE 数据集中的每一张图片都提取显著性区域，一张图片将被划分为多个部分。之后，将这些显著性区域提取特征向量，通过欧氏距离将显著性区域重新分类，形成新标签的显著性区域集合。

2.2 语义标签向量化

图片的语义标签一般是具有强烈鲜明特点的名词或者形容词，此时，显著性区域将更容易成为用户添加网络标签的目标，并且相似的显著性区域会更加容易出现相同的语义标签。本文采用 word2vec 算法提取标签的特征向量，使用 TF-IDF[4] 算法对已提取的标签向量进行再次量化。通过 TF-IDF 算法能够对粗过滤后的标签进行处理，得到相似的语义网络标签。具有相似语义的网络标签，其特征向量也相似。提取词语的词频 TF 和权值参数 IDF，得到显著性区域标签的特征向量。

$$\mathrm{TF}=\frac{\text{某个词出现的次数}}{\text{所有词语出现的次数}} \tag{5}$$

$$\mathrm{IDF}=\frac{\text{词料库总文档数}}{\text{包含某个词的文档}+1} \tag{6}$$

$$\mathrm{TF-IDF}=\mathrm{TF}*\mathrm{IDF} \tag{7}$$

2.3 显著性区域标签向量化

得到图片的相似显著性区域的集合以及显著性语义标签的特征向量之后，将显著性区域与语义标签建立对应关系。

对相似的显著区域所在图像的 TF-IDF 标签向量按位进行向量求和，将数值最大的索引位 index 所代表的标签向量作为这些相似的显著性区域共同的语义标签。具体的计算公式如下所示：

$$\mathrm{index}=\arg\max\sum_{N=1}^{N}\sum_{R=1}^{R}P \tag{8}$$

式中：N 表示 N 个图像有相似显著性区域；R 表示一张图中的标签数目；P 表示指定图片中的指定标签的向量；index 表示标签通过位计算后的最高索引。本文将最终得到的最大索引量作为该特定显著性区域的向量标签。

3 深度网络模型训练

近年来，哈希方法 [5-6] 是解决近似最近邻检索问题的主流方法。

无监督哈希算法通过图片数据对哈希函数进行训练，由于不包含有图片的网络标签，即图片的文字信息，所以训练的结果不尽如人意。而基于 CNN 的有监督哈希学习 [7-8] 却取得了较好的实验结果，能够处理复杂的文字信息。

在提取显著性区域的标签后，本文将显著性区域的标签作为判断图像是否相似的标准。如果几幅图像的显著性区域标签相同，则可判定为相似；否则，为不相似。

3.1 显著性三元组

建立图片的显著性三元组作为深度网络的输入信息。每 3 张图片为一组，其中两张图片为显著性区域吻合的图片，另一张图片的显著性区域完全不相同。通过提取一系列的三元组，将这些集合作为深度网络模型训练的输入。基于图片显著性区域而非整幅图像的网络训练将大大提高检索精确度。图像三元组示例如图 2 所示。

前两个为显著性区域相似的图片
第三个为显著性区域完全不相同的图片

图 2 图像三元组

3.2 深度学习模型

本文选择使用大规模图像识别的深度卷积网络 VGG16 作为深度学习的模型。对网络模型进行优化，去掉 VGG16 的最后一个全连接层，添加三个全连接层，分别为：Dense_1，共 1 024 个神经元；Dense_2，共 512 个神经元；Dense_3，共 64 个神经元。优化的全连接层减少了计算量。

为了防止过拟合，本文选择 L2 正则化约束，并在后

三个全连接层两两之间添加 dropout，对 Dense_3 输出加 L1 正则化约束，使得输出特征稀疏化，以利于之后对其进行哈希化。

将网络最后的输出图像特征当作嵌入空间，并使网络进行空间中数据语义分布的计算，使得相似的汉明距离更近，不相似的汉明距离更远。

3.3 哈希函数训练

本文选择 tanh 作为激活函数，有利于进行公式哈希化，且输出特征是零中心，有利于网络梯度下降。在 Dense_3 进行哈希化处理，使得激活函数映射为（–1，1），输出特征映射为 0 或 1，以节省存储空间。利用汉明距离计算特征向量间相似性的时间复杂度相比欧式距离更低。

$$H=\begin{cases}1, & \text{Output}>0\\ 0, & \text{Otherwise}\end{cases} \tag{9}$$

本文借鉴了 Kevin Lin[9] 提出的哈希化公式，此公式可以对已知显著性区域向量化标签进行哈希化。

当给定显著性区域的特征向量时，将 Dense_3 的输出值作为该区域的特征向量，对该特征向量的每一位都进行式（9）所示的运算，即可得到本文需要的二进制哈希编码。

4 实验结果

用户输入一张图片，首先会提取一张图片的显著性区域，接着将提取的信息输入 VGG16 网络进行学习，最终获得其特征标签向量并对其进行哈希化。图像检索流程如图 3 所示。

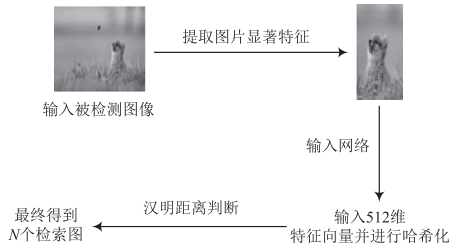

图 3 图像检索流程

分别计算此哈希编码与数据集中显著性区域的哈希编码间的汉明距离，将汉明距离最小的前 n 个哈希码所对应的图像作为最终返回的检索结果。实验中，分别以 16 位、32 位、48 位、64 位的哈希码输出。

NUS-WIDE[10] 数据集为本次显著性区域的网络标签的深度学习图像检索方法测试所用数据集。因该数据集存在的方式是图像的网络地址，而其中一部分网络地址已经失效，所以本文选择了有效部分的网络地址所代表的图像作为最终训练测试集。如果检索出来的图像与被检索的图像大于等于一个共同的网络标签，则可判定它们相似。本文采用平均精度（MAP）作为评价指标。

为了更好地检测图像检索的精确度，将本文的方法（SNDIR）与现今主流的图像检索技术 [11-13] 进行对比测试，具体如图 4 所示。

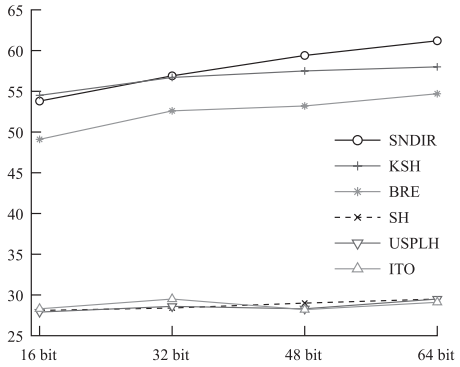

图 4 本文的 SNDIR 方法与其他方法

采用不同长度哈希码的 MAP 结果经对比可以明显得出，在使用不同长度的哈希码进行图像检索时，本文的 SNDIR 方法的 MAP 明显优于主流图像检索方法 [14-15]，证明了该检索方法能够很好地利用网络标签使检索结果达到满意的效果。SNDIR 方法在不同哈希编码长度下检索图片的效果均优于基于弱监督的图像检索方法，但在哈希编码比较短的情况下，SNDIR 的图像检索方法与传统基于有监督图像检索方法相比并未显示出优越性，而当哈希编码较长时，SNDIR 检索效果的优越性比较突出。

当哈希编码长度增长时，SNDIR 方法的 MAP 出现了转折性大幅度增长，所以在进行网络训练时应尽量选用长度较长的哈希编码，使得检索效果精确度得到提升。基于显著性区域网络标签的图像检索方法有着更加丰富的网络标签，而更加精细的标签会使得网络学习更加精确。

5 结 语

本文论述的是一种基于显著性区域网络标签的深度学习图像检索技术，利用算法直接提取图片的显著性区域并进行向量化，同时也对网络标签进行向量化处理，找出不同显著性区域最适合的特征向量。之后通过三元组的学习输出后，对显著性区域的特征向量进行哈希化处理，得到二进制特征向量。检索时，输入图片，最终根据显著性区域的哈希编码间的汉明距离求得数据集中图像的相似度，得到与被检索图像相似的图。

与传统的图片检索技术相比，本文的方法更节省人力。本文使用的训练图片有相对丰富的网络资源，且更加细致，更符合人类的视觉习惯。通过实验测试，证明了相比主流的图像检索技术，在使用长哈希编码的基础上，使用本文方法检索图像更加精确。

注：本文通讯作者为胡琦瑶。

参 考 文 献

[1] SUN A，BHOWMICK S S. Quantifying tag representativeness of visual content of social images [C]// Proceedings of the 18th ACM international. conference on Mult-imedia. ACM，2010：471-480.
[2] Ming Ming Cheng，Ziming Zhang，Wen Yan Lin，et al. BING：Binarized normed gradients for objectness estimation at 300 fps [Z]. Computational Visual Media，2018.
[3] Torralba，Murphy，Freeman，et al. Con-text-based vision system for place and object recognition[C]// Proceedings Ninth IEEE International Conference on Com-puter Vision. IEEE，2008.
[4] 唐明，朱磊，邹显春 . 基于 Word2Vec 的一种文档向量表示 [J]. 计算机科学，2016（6）：214-217.
[5] GIONIS A，INDYK P，MOTWANI R. Similarity search in high dimensions via hashing [J].Vldb，1999（6）：518-529.
[6] JAIN P，KULIS B，GRAUMAN K. Fast image search for learned metrics [Z]. 2008.
[7] LIU W，WANG J，JI R，et al. Supervised hashing with kernels [C]// Computer Vision and Pattern Recognition（CVPR），2012 IEEE Conference on. IEEE，2012：2074-2081.
[8] NOROUZI M，BLEI D M. Minimal loss hashing for compact binary codes[C]// Proceedings of the 28th international conference on machine learning（ICML-11）. 2011：353-360.
[9] LIN K，YANG H F，HSIAO J H，et al. Deep learning of binary hash codes for fast image retrieval [C]// Proceedings of the I-EEE conference on computer vision and pattern recognition workshops，2015：27-35.
[10] CHUA T S，TANG J，HONG R，et al. NUS-WIDE：a real-world web image database from National University of Singap-ore [C]// Proceedings of the ACM internat-ional conference on image and video retr-ieval. ACM，2009：48.
[11] WANG J，KUMAR S，CHANG S F. Semi-supervised hashing for scalable image retrieval [C]// 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition，2010：1-40.
[12] WEISS Y，TORRALBA A，FERGUS R. Spectral hashing [C]// Advances in neural information processing systems，2009：1753-1760.
[13] YANG H F，LIN K，CHEN C S. Supervised learning of semantics-preserving hash via deep convolutional neural networks [J]. IEEE transactions on pattern analysis and machine intelligence，2018，40（2）：437-451.
[14] XIA R，PAN Y，LAI H，et al. Supervised hashing for image retrieval via image representation learning [Z]. AAAI，2014.
[15] ZHANG R，LIN L，ZHANG R，et al. Bit-scalable deep hashing with regularized similarity learning for image retrieval and person re-identification [J]. IEEE transactions on image processing，2015，24（12）：4766-4779.

作者简介：田少骅（2000—），女，本科，主要研究方向为通信信号处理与传输。
蒙泽新（2000—），女，本科，主要研究方向为信号与信息处理、自动化设计。
王炅昱（1999—），女，本科，主要研究方向为电路优化设计和数字信息处理。
胡琦瑶（1993—），女，硕士，助理工程师，主要研究方向为深度学习、嵌入式系统设计。

（上接第 53 页）

对图 2 ～图 4 所示的数据进行分析可得：输出 1 的 R^2=0.82，输出 2 的 R^2=0.85，输出 3 的 R^2=0.94，表明所取得的预测结果较好，当 R^2 的值越接近 1，表明其预测结果越好。从 3 个输出的 R^2 对比可得，输出 3 的预测效果更好，输出 1，2 相对来说预测效果较差。

3 结 语

本文运用人工神经网络算法完成对室内自然光照度的预测，通过仿真验证可知该方法具有较高的准确性。但是人工神经网络训练速度较慢，且容易陷入局部最小值。在接下来的研究中，应该不断地优化算法来提升预测的准确性和收敛速度；同时，应把室内工作平面距离窗户的距离以及顶棚的高度对室内自然光照度的影响考虑在内，不断地优化预测模型。

参 考 文 献

[1] KAZANASMAZ Tuğçe，GÜNAYDIN Murat，BINOL Selcen. Artificial neural networks to predict daylight illuminance in office buildings [J]. Building and environment，2008，44（8）：1751-1757.
[2] COLEY D A，CRABB J A. An artificial intelligence approach to the prediction of natural lighting levels [J]. Building and environment，1997，32（2）：81-85.
[3] 章云，许锦标，谷刚，等 . 建筑智能化系统 [M]. 2 版 . 北京：清华大学出版社，2017.
[4] 孙友明，沈晓明，覃善华，等 . 基于两级定位跟踪的自然光导光照明装置设计 [J]. 科学技术与工程，2017，17（17）：72-77.
[5] 段春丽，黄仕元，于兰，等 . 建筑电气 [M].2 版 . 北京：机械工业出版社，2016.
[6] 王金光 . 智能照明控制策略的研究与仿真 [D]. 上海：同济大学，2008.
[7] 朱俊丞，杨之乐，郭媛君，等 . 深度学习在电力负荷预测中的应用综述 [J]. 郑州大学学报（工学版），2019，40（5）：13-22.
[8] 刘蕾 . 基于 Tensorflow 的循环神经网络模型在上海市空气质量预测中的应用 [D]. 上海：上海师范大学，2019.
[9] 黄志辉 . 基于卷积神经网络的量化选股模型研究 [D]. 杭州：浙江大学，2019.
[10] 冯冬青，刘丹丹 . 室内智能舒适照明控制策略研究 [J]. 郑州大学学报（理学版），2015，47（3）：99-104.