# 基于文本与语义相关性分析的图像检索

穆亚昆,冯圣威,张 静
华东理工大学 信息科学与工程学院,上海 200237

**摘 要:**为了更加有效地检索到符合用户复杂语义需求的图像,提出一种基于文本描述与语义相关性分析的图像检索算法。该方法将图像检索分为两步:基于文本语义相关性分析的图像检索和基于SIFT特征的相似图像扩展检索。根据自然语言处理技术分析得到用户文本需求中的关键词及其语义关联,在选定图像库中通过语义相关性分析得到“种子”图像;接下来在图像扩展检索中,采用基于SIFT特征的相似图像检索,利用之前得到的“种子”图像作为查询条件,在网络图像库中进行扩展检索,并在结果集上根据两次检索的图像相似度进行排序输出,最终得到更加丰富有效的图像检索结果。为了证明算法的有效性,在标准数据集Corel5K和网络数据集Deriantart8K上完成了多组实验,实验结果证明该方法能够得到较为精确地符合用户语义要求的图像检索结果,并且通过扩展算法可以得到更加丰富的检索结果。

**关键词:**图像检索;基于文本语义相关性的图像检索;语义相关度;SIFT低层视觉特征;图像扩展检索
**文献标志码:**A **中图分类号:**TP391.42 doi:10.3778/j.issn.1002-8331.1709-0209

穆亚昆,冯圣威,张静.基于文本与语义相关性分析的图像检索.计算机工程与应用,2019,55(1):196-202.
MU Yakun, FENG Shengwei, ZHANG Jing. Image retrieval based on text and semantic relevance analysis. Computer Engineering and Applications, 2019, 55(1):196-202.

## Image Retrieval Based on Text and Semantic Relevance Analysis

MU Yakun, FENG Shengwei, ZHANG Jing
College of Information Science and Engineering, East China University of Science and Technology, Shanghai 200237, China

**Abstract:** An image retrieval algorithm based on text and semantic relevance is proposed to effectively retrieve images, which can satisfy the complex requirement of users. There are two steps in this method: images retrieval based on textual semantic relevance analysis and image extension retrieval based on SIFT features. Firstly, according to the analysis of Natural Language Processing technology, it acquires text keywords and semantic correlation of users' demands, and applies them to retrieve seed images from the image dataset by semantic relevance analysis. Then image extension retrieval based on SIFT feature are used to conduct extend retrieval in Web image datasets according to the retrieval results of last step. Finally, the paper achieves the query results by combining the retrieval results of these two steps. The experiments on Standard Data Set Corel5K and Web Data Set Deriantart8K prove that this method can achieve more precise image retrieval results by semantic relevance analysis and can enrich the query results by extend retrieval.

**Key words:** image retrieval; image retrieval based on semantic correlation of text; semantic relevance; low-rise visual features of SIFT; image extension retrieval

## 1 引言

随着图像采集设备和社交网络的飞速发展,网络图片的数量正以指数级的速度增长。而如何从大量的图像中快速有效地检索到用户想要的图片,成为一个亟待解决的问题。现有的图像检索技术主要分为三种:基于文本的图像检索技术(Text-Based Image Retrieval,TBIR)[1-2]、基于内容的图像检索技术(Content-Based Image Retrieval,CBIR)[3]和基于语义的图像检索技术(Semantic-

**基金项目:**国家自然科学基金(No.61402174)。
**作者简介:**穆亚昆(1993—),男,硕士,主要研究领域为数字图像处理;冯圣威(1989—),男,硕士,主要研究领域为数字图像处理;张静(1979—),女,副教授,CCF会员,主要研究领域为数字图像处理、数据挖掘,E-mail:jingzhang@ecust.edu.cn。
**收稿日期:**2017-09-15 **修回日期:**2018-01-24 **文章编号:**1002-8331(2019)01-0196-07
CNKI **网络出版:**2018-05-19, http://kns.cnki.net/kcms/detail/11.2127.TP.20180518.0854.002.html

Based Image Retrieval,SBIR)[4]。

基于文本的图像检索产生于20世纪70年代末期,这种技术依赖于手工标注,效率低下。2011年,Li[5]等人提出使用了多实例学习的基于文本的图像检索策略,使用改进的约束正样本袋的多实例学习(PMIL-CPB)进行更加精确的图像结果检索排序。然而,基于文本的图像检索方法在海量图像数据中的检索效率已经不能满足人们的需求了。

20世纪90年代初,基于内容的图像检索的技术被提出来,大大提高了图像查询的效率。然而"语义鸿沟"(Semantic Gap)的存在,成为了CBIR进一步发展的瓶颈。为了缓解"语义鸿沟"的影响,张永库等人提出了多特征融合的图像检索算法:基于底层特征综合分析算法(CAUC),通过将全局颜色特征、局部颜色特征、形状特征和纹理特征进行融合得到的图像的多特征描述与查询图像计算相似度,从而得到查询结果[3]。而刘胜蓝等人提出多特征图融合的检索方法:分组排序融合(GRF)来提高图像检索精度[6]。然而这些图像检索方法忽略了图像标签之间的语义关联,对用户的潜在需求理解不够,所以基于语义的图像检索应运而生。

类似于TBIR,SBIR也是采用关键词来对图像进行检索,不同的是SBIR以图像语义的自动标注为基础,通过将图像底层特征与高层语义进行关联得到图像的语义描述,并实现对图像的检索[7]。所以检索结果的准确性受到图像自动标注结果的限制,而Tu等人[8]提出一种结合图像背景主题模型的图像检索方法。此方法主要通过图像的视觉描述词、文本词和背景之间的关系模型来对图像标签进行补全,然后提出一个融合标签和图像主题的图像相似度度量算法来实现图像检索。Feng等人[9]则提出一种利用图像标签共现关系建立标签间层次共现模型来提高图像检索效率的算法,其通过融合语义标签之间的全局语义共现关系(谷歌距离)、全局视觉共现关系(flicker网站上的标签共现距离)和局部视觉共现关系(图像集中的共现关系)来构建语义之间的关联,并且通过标签关联的远近来构建以小团体为单位层次结构。在图像检索时,用EMD(Earth Mover's Distance)作为查询条件与标已标注图像标签之间的距离衡量标准,将得到的图像按照与查询条件的关联度进行排序输出[9]。然而这些基于语义的图像检索算法在提高检索精度的同时,图像的丰富性大大降低了,无法满足用户各种各样的潜在需求。

由于TBIR、CBIR和SBIR各有缺陷,所以很多研究者开始研究融合这几种技术的图像检索方法[10-12],顾昕等人结合基于文本和基于内容的图像检索特性,采用稠密的尺度不变特征转换(DSIFT)构造视觉单词的方式来描述图像内容,依据基于概率潜在语义分析(PLSA)模型使用自适应的不对称学习方法融合并学习视觉模态和文本模态信息实现图像自动文本标注。在图像检索时,首先对待查询图像(测试集)进行自动标注得到其文本信息并与库中文本比较初步检索筛选出语义相关图像集,然后根据其自动获取的视觉内容对语义相关图像进行相似性度量并排序,返回检索结果[10]。这些工作虽然显著地提高了图像检索性能,然而它们在获取图像文本语义时多借助于人工的注释或是采用适当的算法提取Web中相关的文本信息,因此应用受到了一定的限制。

本文提出了一种基于文本与语义相关性分析的图像检索算法。首先,通过对文本信息分析得到基于文本的图像检索模式;通过对文本信息的关键词语义关联分析得到基于语义标签位置关联的图像检索模式。融合上述两种模式,通过分析图像标签之间的相对位置关系并在"种子"图像库中进行图像检索得到"种子"图像。然后,利用这些"种子"图像的低层视觉特征实现在网络图像库中进行基于内容的图像检索,得到准确度较高且更为丰富的图像检索结果。

## 2 图像检索模型框架

为了更加准确地匹配用户的检索信息,得到用户预期的检索结果,文中提出了基于文本与语义相关性分析的图像检索方法。如图1所示,此方法主要分为两部分:基于文本语义的图像初次检索和基于"种子"图像的扩展检索。其中基于文本语义的图像初次检索分为三部分:"种子"图像库的建立、文本信息的自然语言处理以及基于语义的图像检索。基于"种子"图像的扩展检索主要是经过对"种子"图像进行视觉特征分析以及特征匹配扩展图像检索结果。

## 3 基于文本语义的图像检索

基于文本语义的图像检索分为三部分:"种子"图像库的建立、文本信息的自然语言处理以及基于语义的图像检索。

### 3.1 "种子"图像库语义位置关系的构建

用户输入的检索信息是文本信息,是对期望图像的语义描述,所以本文自建了一个"种子图像库"来将这些语义描述进行图像化表示。

文中选择使用基于文本语义的图像检索,而在庞大的网络图像数据库中进行图像的标注和提取语义关联信息是非常费事费力的,所以文中选择建立一个小规模的"种子"图像数据库来满足图像的初次检索得到"种子"图像。

首先将"种子"图像库表示为 $\Gamma=\{I_1,I_2,\cdots,I_n\}$ ,$n$ 为图像库的大小,同时设定 $C=\{c_1,c_2,\cdots,c_m\}$ 为图像库中包含的标签集合。根据手工标注结果为图像 $I_k$ 建立一

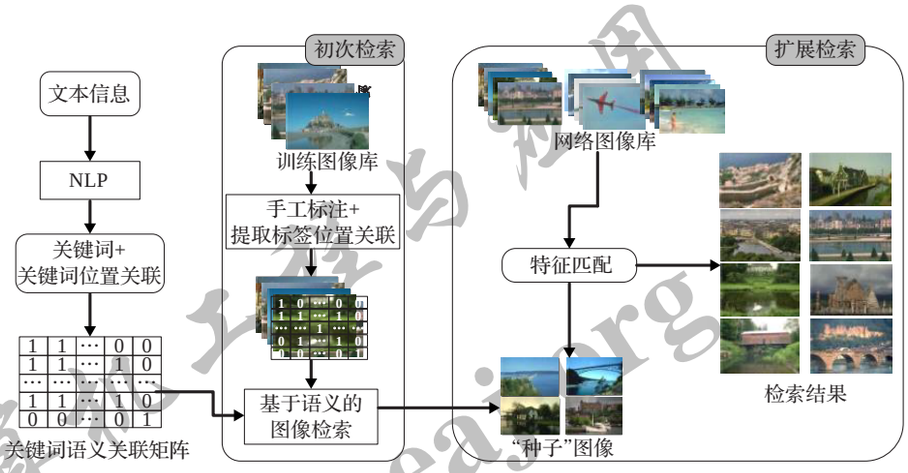


**图1　基于文本与语义相关性分析的图像检索模型框架**

个标签特征向量 $I_k \rightarrow Y_k=\{y_1,y_2,\cdots,y_m\}$，$y_i\in\{0,1\}^m$。如果 $I_k$ 包含标签 $c_i$，则 $y_k(i)=1$ 否则为 $y_k(i)=0$。

本文选择提取的语义关联是图像标签之间的相对位置关系。图片标签之间的位置关系分为拓扑关系和方向关系[13-14]。而在图像结构较为复杂的网络图像中，拓扑关系是非常复杂的且具有很大的偶然性，故此处只考虑图像标签之间的方向位置关系。相对位置关系可以分为上方、下方，左方和右方。而通常左方和右方对图像的影响很小，所以可以将方向关系分为上方、下方和旁边。如图2所示，其划分依据是根据标签区域质心连线与水平线的夹角。

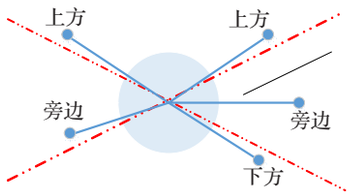


**图2　方向位置关系**

本文构建标签的上位置关系矩阵 $M_{\text{abo}}^{k}\in R^{m\times m}$、下位置关系矩阵 $M_{\text{bel}}\in R^{m\times m}$ 和旁边位置关系矩阵 $M_{\text{bes}}\in R^{m\times m}$。其中子元素 $\mu_{\text{abo}}(i,j)$ 表示标签 $i$ 在标签 $j$ 的上方，$\mu_{\text{bel}}(i,j)$ 表示标签 $i$ 在标签 $j$ 的下方，$\mu_{\text{bes}}(i,j)$ 表示标签 $i$ 在标签 $j$ 的旁边，其计算过程如下：

$$\mu_{\text{abo}}(i,j)=\begin{cases}1, \text{ if } \dfrac{\pi}{6}<\theta_{ij}<\dfrac{5\pi}{6}\\ 0, \text{ otherwise}\end{cases} \tag{1}$$

$$\mu_{\text{bel}}(i,j)=\begin{cases}1, \text{ if } -\dfrac{5\pi}{6}<\theta_{ij}<-\dfrac{\pi}{6}\\ 0, \text{ otherwise}\end{cases} \tag{2}$$

$$\mu_{bes}(i,j)=\begin{cases}\text{if } -\dfrac{\pi}{6}\leqslant\theta_{ij}\leqslant\dfrac{\pi}{6}\\ 1, \dfrac{5\pi}{6}\leqslant\theta_{ij}\leqslant\pi\\ \text{or } -\pi\leqslant\theta_{ij}\leqslant-\dfrac{5\pi}{6}\\ 0, \text{ otherwise}\end{cases} \tag{3}$$

其中 $\theta_{ij}$ 表示标签 $i$ 和标签 $j$ 区域质心连线和水平线的夹角。

本文为“种子”图像库中的每个图像生成各自的标签方向关系矩阵，以备后期基于文本语义相关性图像检索使用。

### 3.2　文本信息的自然语言处理

自然语言处理(Natural Language Processing，NLP)是研究人与计算机交互的语言问题的一门学科。这里本文使用常用的自然语言处理工具NLTK来处理用户输入的查询文本信息。

NLTK(Natural Language ToolKit)：是一个用于建立Python程序语言和人类语言交互的平台，它提供了易于使用的接口，超过50个语料库和词汇资源，如WordNet，以及一套文本处理库的分类、标记、标注、句法分析、语义推理。

本文使用NLTK来分析用户输入的含有搜索意图的文本描述信息，提取其中的关键词以及关键词之间的位置关系。比如“建筑海洋上下结构并且海洋上有船”这样的描述信息，NLTK系统首先提取到关键字“建筑”，“海洋”，“船”。然后提取关键字与关键字之间的位置关系“建筑在海洋上方”、“海洋表面上有船”，因此可得到“建筑-船-海洋”的三层上下位置关系的语义结构。或是“建筑汽车结构并且建筑旁边是汽车”，NLTK系统可以提取关键词“建筑”和“汽车”，并且得到关键词之间的位置关系。

本文用 $T$ 来表示NLP得到的关键词信息：$T^{\tau}\rightarrow Y_{\tau}=\{y_1,y_2,\cdots,y_m\}$，$y_i\in\{0,1\}^m$。如果 $T^{\tau}$ 包含标签 $c_i$，则 $y_{\tau}(i)=1$ 否则为 $y_{\tau}(i)=0$。

以图像的上下关系为例，本文构建它的关键词上方位置关联矩阵 $M_{\text{abo}}^{\tau}\in R^{m\times m}$，其中子元素 $\mu_{\text{abo}}^{\tau}(i,j)$ 表示文本描述中包含关键词 $i$ 在关键词 $j$ 上方的关系：

$$\mu_{\text{abo}}^{\tau}(i,j)=\begin{cases}1, \text{ if } c_i \text{ is above on } c_j \text{ in } T_{\tau}\\ 0, \text{ otherwise}\end{cases} \tag{4}$$

同理得到关键词下方关系矩阵 $M_{\text{bel}}^{\tau}\in R^{m\times m}$ 和旁边关系矩阵 $M_{\text{bes}}^{\tau}\in R^{m\times m}$:

$$\mu_{\text{bel}}^{\tau}(i,j)=\begin{cases}1, \text{ if } c_i \text{ is below on } c_j \text{ in } T_\tau\\ 0, \text{ otherwise}\end{cases} \tag{5}$$

$$\mu_{bes}^{\tau}(i,j)=\begin{cases}1, \text{ if } c_i \text{ is beside on } c_j \text{ in } T_\tau\\ 0, \text{ otherwise}\end{cases} \tag{6}$$

### 3.3 基于文本语义的图像检索

基于文本语义的图像检索分为两个部分:基于文本的相似度计算和基于语义的相似度计算。

本文通过NLP技术分析文本查询信息得到关键词集合 $T^{\tau}$ 和关键词之间的语义关系矩阵 $M_{\text{abo}}^{\tau}$、$M_{\text{bel}}^{\tau}$ 和 $M_{\text{bes}}^{\tau}$,将这些查询条件在"种子"图像库中进行相似度匹配,得到符合条件的相似图像,作为"种子"图像。这里使用的相似度计算模型是空间向量模型中的欧式距离。

文本关键词集合 $T^{\tau}$ 和"种子"图像 $I_k$ 标签集合的相似度:

$$sim_1(\tau,k)=\frac{1}{\sqrt{\sum_{i=1}^{m}\left(y_\tau(i)-y_k(i)\right)^2}} \tag{7}$$

文本语义关联矩阵 $M_{\text{abo}}^{\tau}$ 和"种子"图像语义关联矩阵 $M_{\text{abo}}^{k}$ 的相似度定义如下:

$$sim_2(\tau,k)=\frac{1}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\mu_{\text{abo}}^{\tau}(i,j)-\mu_{\text{abo}}^{k}(i,j)\right)^2}} \tag{8}$$

文本语义关联矩阵 $M_{\text{bel}}^{\tau}$ 和"种子"图像语义关联矩阵 $M_{\text{bel}}^{k}$ 的相似度定义如下:

$$sim_3(\tau,k)=\frac{1}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\mu_{\text{bel}}^{\tau}(i,j)-\mu_{\text{bel}}^{k}(i,j)\right)^2}} \tag{9}$$

文本语义关联矩阵 $M_{\text{bes}}^{\tau}$ 和"种子"图像语义关联矩阵 $M_{\text{bes}}^{k}$ 的相似度定义如下:

$$sim_4(\tau,k)=\frac{1}{\sqrt{\sum_{i=1}^{m}\sum_{j=1}^{m}\left(\mu_{\text{bes}}^{\tau}(i,j)-\mu_{\text{bes}}^{k}(i,j)\right)^2}} \tag{10}$$

综上,查询文本与"种子"图像之间的匹配度定义为:

$$sim(\tau,k)=\alpha\times sim_1(\tau,k)+(1-\alpha)\left[sim_2(\tau,k)+sim_3(\tau,k)+sim_4(\tau,k)\right] \tag{11}$$

其中,$\alpha$ 是决定文本关键词匹配度和标签语义匹配度在综合匹配度中权重的参数,根据后期实验结果,这里取 $\alpha=0.5$。最终,本文选取匹配度最大的前 $U$ 张图片作为初次检索的结果。

## 4 基于特征匹配的图像扩展

基于特征匹配的图像扩展是根据上一章得到的"种子"图像集,在网络图像库中进行基于内容的图像检索,得到更加丰富的图像检索结果。

本文采用SIFT(Scale-Invariant Feature Transform)特征来做图像的扩展,其尺度不变性的特点可以保证图像在拉伸、变形的情况下仍不影响其特征。它在空间尺度中寻找极值点,并提取出其位置、尺度、旋转不变量[15]。因此,针对大规模无固定尺寸的图像来说,使用SIFT特征具有很大的优势。

128维的SIFT特征向量能够非常清晰的描述图像局部特征,在图像局部特征匹配时的实用性很强。为了计算两张图像的匹配度,本文采用SIFT关键特征点之间的欧式距离来计算图像关键点的相似度。

由于一幅图像拥有成千上万的SIFT特征点,如果使用128维的全特征来计算网络图像库与"种子"图像之间的欧式距离,计算效率低下。因此,本文采取主成分分析(PCA)来对SIFT特征进行降维处理。

主成分分析是一种常见的图像数据压缩方法[16]。它主要根据原始矩阵的特征向量按递减顺序构建变换核矩阵。然后,取对图像质量影响最大的前 $K$ 个特征向量构造最终的变换核矩阵。参考文献[10]和文献[17],本文将SIFT特征降维到20维,即 $K=20$,使用SIFT特征点匹配时,在保证匹配精确度时,有效的提高了匹配效率。

本文选取"种子"图像中的某个关键点,找到网络数据库中与之欧式距离达到小于阈值的关键点作为其的匹配关键点。若两幅图像之间的匹配关键点越多,则说明幅图像的相似度越高。其计算方式如下:

$$\psi(j)=\frac{1}{U}\sum_{k=1}^{U}sim(\tau,k)\times Sim_{\text{match}}(j|k) \tag{12}$$

其中,$Sim_{\text{match}}(j|k)$ 表示"种子"图像 $k$ 与扩展集图像 $j$ 的图像相似度。

本文将网络图像库中二次检索得到的结果图像按其匹配度 $\psi(j)$ 的降序排列结果作为图像扩展结果。

## 5 实验

为了验证基于文本和语义相关性分析的图像检索算法的有效性,本文进行了一系列实验。本文在公共标准数据库Corel5K和网络数据库Deriantart8K分别做了实验。并针对图像检索准确度与几种图像检索算法做了对比实验。下面首先介绍实验过程中所选用的数据集和评估标准,然后介绍实验结果以及对实验结果的分析。

### 5.1 数据集及评估标准

Corel5K在图像标注和图像检索的研究领域中应用非常广泛,该图像库中包含50种类别的图像,每个类别子库中包含了100张图像。首先,本文Corel5K中选取1 500张图像作为"种子"图像库,并对其进行手工标注,以及标签的语义相关性分析,得到标签相关性矩阵,然

**表1　数据集Corel5K和网络数据集Deriantart5K**

| 图像库 | “种子”图像库 | 扩展数据集 | 标签 |
|---|---|---|---|
| Corel5K | 1 500 | 3 500 | “sky”,“plant”,“sea”,“sand”,“rock”,“road”,“house”,“building”,“sun”,“car”,“plane”,“boat”,“land” |
| Deriantart8K | 3 000 | 5 000 | “sky”,“plant”,“water”,“mountain”,“grass”,“sand”,“rocks”,“car”,“road”,“ground”,“castle”,“chapel”,“house”,“building”,“snow”,“train”,“track”,“plane”,“runway”,“smoke”,“cloud” |

**表2　不同图像检索方法在Corel5K和Deriantart8K的*MAP***

| 数据集 | 方法 | Top5 | Top10 | Top15 | Top20 | Top25 | Top30 | Top35 | Top40 |
|---|---|---|---|---|---|---|---|---|---|
| Corel5K | PMIL-CPB | 75.45 | 70.71 | 63.39 | 55.89 | 49.87 | 46.38 | 43.21 | 40.10 |
| | GRF | 80.74 | 75.37 | 73.25 | 68.41 | 63.51 | 59.88 | 54.27 | 50.14 |
| | 文献[9] | 98.47 | 90.31 | 84.32 | 77.51 | 72.48 | 68.43 | 64.38 | 59.37 |
| | 文献[10] | 86.45 | 84.65 | 80.15 | 75.45 | 73.56 | 68.48 | 64.38 | 58.46 |
| | 本文方法 | 95.34 | 93.59 | 87.76 | 82.62 | 79.84 | 77.25 | 75.46 | 73.55 |
| Deriantart8K | PMIL-CPB | 67.43 | 62.37 | 57.39 | 50.45 | 43.82 | 40.97 | 38.42 | 36.57 |
| | GRF | 69.32 | 64.47 | 60.46 | 55.41 | 50.73 | 47.39 | 44.72 | 40.85 |
| | 文献[9] | 87.45 | 83.42 | 80.11 | 76.43 | 70.38 | 65.43 | 62.71 | 58.42 |
| | 文献[10] | 82.74 | 79.58 | 73.57 | 67.67 | 60.70 | 54.27 | 50.14 | 57.03 |
| | 本文方法 | 85.39 | 83.01 | 78.33 | 76.46 | 74.37 | 72.35 | 70.82 | 68.94 |

后在该“种子”数据集上完成基于文本语义的初次检索，其手工标注的标签如表1所示。然后将Corel5K中剩下的3 500张图像作为网络扩展图像库，完成基于“种子”图像的扩展检索。

Deriantart8K图像数据集是在图像分享网站Deriantart上获取8 000张图像。本文在Deriantart8K中选择3 000张图像作为“种子”图像库，其手工标注的标签如表1所示。类似于Corel5K，在Deriantart8K剩余的5 000张图像中进行图像扩展检索。

为了评估实验结果，本章采用了MAP[18]评估标准，该评估标准在图像检索领域广泛使用，单次检索结果的平均准确率是检索结果准确率的平均值，整个图像库的平均检索结果是每个主题的平均准确度的平均值，*MAP*的定义如下：

$$MAP=\frac{\sum_{i=1}^{S}AP_i}{S} \tag{13}$$

其中$S$表示输入的查询文本的个数，对于一条查询文本$i$，笔者取前$U$个检索结果，则AP可以定义为：

$$AP=\frac{\sum_{i=1}^{U}\rho_i P_i}{H_{GT}} \tag{14}$$

其中：

$$\rho_i=\begin{cases}1, & \text{if the } i-th \text{ result is correct}\\ 0, & \text{otherwise}\end{cases} \tag{15}$$

$$P_i=\frac{No(i)}{i} \tag{16}$$

其中，$U$为选定的查询结果的计算范围，$H_{GT}$表示在Top $U$个查询结果中符合Ground Truth中正样本的数量，$No(i)$表示第$i$个结果在Ground Truth正样本中的排序。

## 5.2　实验结果及分析

为了验证提出的基于文本与语义相关性的图像检索算法的有效性，本文将该算法和PMIL-CPB[5]、GRF[6]、文献[9]以及文献[10]提出的方法在Corel5K上进行对比实验，在其初次检索得到的“种子”图像中，分别选取Top5、Top10、Top15、Top20、Top25、Top30、Top35和Top40的结果集进行评测，其结果如表2和图3所示。

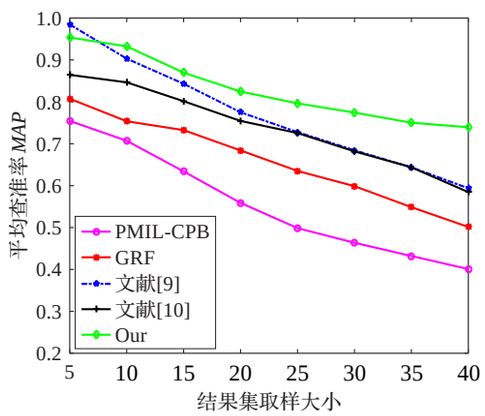


**图3　不同图像检索方法在Corel5K的*MAP***

从图3可以得出，本文提出的基于文本与语义相关性的图像初次检索的结果在所有的结果集取样中，其平均查准率都是优于PMIL-CPB、GRF和文献[10]等图像检索方法。而文献[9]提出的基于语义的图像检索算法，在结果集取样为Top5时的检索精度稍高于本文提出的图像检索方法。但是随着图像检索结果集取样容量的增加，文献[9]中图像检索方法的检索精度快速下降，当图像检索结果选取到Top40时，其检索精度已经下降到60.00%，而本文提出的图像检索方法在此过程中依旧能保证73.00%以上的平均查准率。这表明融合

文本与语义相关性的图像检索得到的“种子”图像较为准确,在很大程度上满足用户的语义需求。

从表2可以看出,本文提出的基于文本与语义相关性分析的图像检索算法,在标准数据集Corel5K上图像检索的平均查准率要高于PMIL-CPB、GRF和文献[10]等图像检索方法。并且在Top5结果集中,*MAP*达到95.37%,而且随着结果集取样容量增加时,依旧保持较好的查询准确率,明显优于表中其他图像检索方法。

除了Corel5K数据库,本文的算法还在网络数据集Deriantart8K上进行相关的实验,实验结果表明,本文提出的检索方法无论在Corel5K标注数据库还是在网络图像数据库Deriantart8K上都能取得较好的检索结果,明显优于其他相关算法。

为了验证本文提出的图像检索算法的鲁棒性,将Corel数据集上1 500张图像集$\Gamma$划分为5个容量相同的互斥子集,即$\Gamma=\Gamma_1\cup\Gamma_2\cup\Gamma_3\cup\Gamma_4\cup\Gamma_5$,$\Gamma_i\cup\Gamma_j=\varnothing(i\neq j)$。每个子集$\Gamma_i$都是从$\Gamma$中通过随机采样得到,即从Corel数据集的50种类别图像集中进行选取同等数量的图像。同时针对每种类别图像集,采取随机不重复采样的方式,这样就可以尽可能保持数据分布的一致性。每次选取4个子集的并集作为“种子”图像集,50条测试文本信息不变,这样可以得到5组测试结果,最终返回5组测试结果的均值,其结果如表3所示。从5组实验结果可以看出,本文提出的基于文本与语义相关性分析的图像检索算法在不同的数据采样集上能够表现出比较稳定的查准率,即本文提出的算法受到不同“种子”图像库的影响不大,具有很好的鲁棒性。

**表3 在Corel5K上5组实验的*MAP***

| 实验 | Top5 | Top10 | Top15 | Top20 | Top25 | Top30 | Top35 | Top40 |
|---|---|---|---|---|---|---|---|---|
| 第1组 | 94.25 | 91.35 | 88.47 | 83.21 | 80.25 | 77.27 | 74.37 | 72.79 |
| 第2组 | 93.49 | 92.43 | 86.51 | 81.27 | 78.49 | 76.43 | 75.58 | 73.39 |
| 第3组 | 97.21 | 95.41 | 90.01 | 84.97 | 81.33 | 78.57 | 76.42 | 74.93 |
| 第4组 | 95.43 | 94.03 | 88.42 | 83.27 | 80.39 | 77.92 | 76.00 | 74.25 |
| 第5组 | 96.31 | 94.72 | 85.38 | 80.37 | 78.73 | 76.07 | 74.95 | 72.37 |
| 平均值 | 95.34 | 93.59 | 87.76 | 82.62 | 79.84 | 77.25 | 75.46 | 73.55 |

从图4“种子”图像结果可以看出,本文提出的方法检索得到的图像很大程度的满足了用户的文本需求。为了得到更加多样性的图像检索结果,本文在检索得到的“种子”图像的基础上,选取符合用户文本语义关联需求的前5个查询结果集作为扩展

“种子图像”,利用基于内容的图像检索方法进一步得到更加丰富的图像检索结果。图5给出了图像扩展的检索结果样例。

如图5所示,用户文本查询条件中的语义需求在图像扩展检索结果中得到了较好地体现,扩展结果依旧能满足用户输入的相对位置关联,同时扩展图像又提供了多样性的检索结果。并且这些扩展图像并不是一幅查


**图4 检索得到的“种子”图像结果**


**图5 查询扩展的结果**

询图像延伸得到的,它们是结合多个“种子”图像的视觉特征共同检索得到,这种方式就弱化了某些“种子”图像中的噪声特征对扩展结果的影响,这就在提高图像检索结果的丰富度的同时,依旧保证图像检索的精确度。

由此可见,本文提出的基于文本与语义相关性分析的图像检索方法不仅能够得到精确的查询结果,而且通过基于“种子”图像的扩展查询为用户提供了更加准确丰富的图像检索结果,更好地满足用户的检索需求。

### 5.3 实验参数分析与实验

对于文中图像相似性度量,本文考虑了杰卡德距离、欧式距离以及余弦距离。为了选取最合适的距离计算方法,本文在选定的Corel5K以及Deriantart8K数据集上,分别选取Top10的结果集上的平均查准率来评估3种距离计算方式对图像检索性能的影响,结果如图6所示,欧式距离作为图像的相似性度量在算法中取得了更好的图像检索结果,因而本文选择欧式距离作为相似度度量标准。

在公式(11)中的各种相似度融合时,本文采取的是为4种相似度分配不同的权重,进行算术平均得到综合匹配度。假定在基于文本和语义图像检索中,文本相似性和语义关联相似度同等重要,故在综合相似度计算中

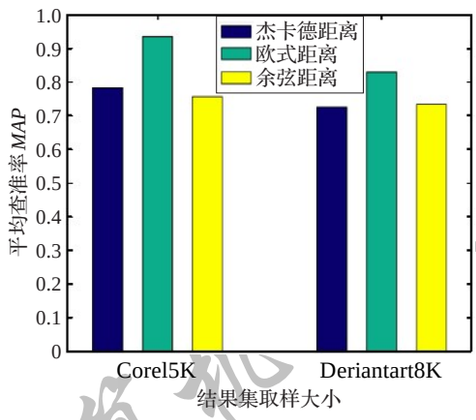


**图6　不同相似度度量下*MAP***

为其分配了的相同的权重,为了验证该假设的合理性,本文在选定的Corel5K数据集上进行实验,选取Top10的结果集上的平均查准率来评估不同参数融合对检索结果的影响,实验结果如图7所示。从实验结果可得,公式中选取$\alpha=0.5$时,图像检索结果最优。

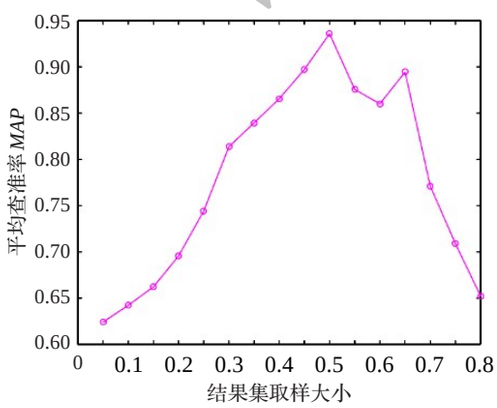


**图7　不同$\alpha$取值时的*MAP***

## 6　结论

本文提出了一种基于文本与语义相关性分析的图像检索算法,首先通过自然语言处理技术将用户输入的文本查询语句进行分析,得到其中的文本关键词和语义关联。然后,在手工标注的“种子”图像库中利用文本语义检索得到“种子”图像。最后根据“种子”图像的SIFT特征在大规模网络图像库中进行基于内容的图像扩展检索,并综合两次图像检索从而得到最佳的图像检索结果。该方法充分发挥了文本、视觉内容以及语义相关性对图像检索各自的优势,并通过实验证明,该算法在保证图像检索的准确性同时,大大提高检索结果的丰富性。

## 参考文献:

[1] Ho T,Ly N.A scene text-based image retrieval system[C]//Proceeedings of IEEE International Symposium on Signal Processing and Information Technology,2012:79-84.

[2] Li W,Duan L,Xu D,et al.Text-based image retrieval using progressive multi-instance learning[C]//Proceeedings of IEEE International Conference on Computer Vision,2011:2049-2055.

[3] 张永库,李云峰,孙劲光.基于多特征融合的图像检索[J].计算机应用,2015,35(2):495-498.

[4] Stanescu L,Burdescu D D,Brezovan M,et al.Semantic-based image retrieval[M]//Creating New Medical Ontologies for Image Annotation.New York:Springer,2012:91-102.

[5] Li W,Duan L,Xu D,et al.Text-based image retrieval using progressive multi-instance learning[C]//Proceeedings of IEEE International Conference on Computer Vision,2011:2049-2055.

[6] 刘胜蓝,冯林,孙木鑫,等.分组排序多特征融合的图像检索方法[J].计算机研究与发展,2017,54(5):1067-1076.

[7] Liu A H,Tong H,Tong Q.A method for semantic-based image retrieval[J].Proceedings of SPIE,2009,7495:1-7.

[8] Tu N A,Cho J,Lee Y K.Semantic image retrieval using correspondence topic model with background distribution[C]//Proceeedings of International Conference on Big Data and Smart Computing,2016:191-198.

[9] Feng L,Bhanu B.Semantic concept co-occurrence patterns for image annotation and retrieval[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2016,38(4):785-799.

[10] 顾昕,张兴亮,王超,等.基于文本和内容的图像检索算法[J].计算机应用,2014(S2):280-282.

[11] Wei D,Zhao Y,Cheng R,et al.An enhanced histogram of oriented gradient for pedestrian detection[C]//Proceedings of the Fourth International Conference on Intelligent Control and Information Processing,2013:459-463.

[12] Ylioinas J,Hadia A,Guo Y.et al.Efficient image appearance description using dense sampling based local binary patterns[C]//Proceeedings of ACCV 2012,2012:375-388.

[13] Aksoy S,Tusk C,Koperski K,et al.A scene modeling and image mining with a visual grammar[J].Frontiers of Remote Sensing Information Processing,2003(1):35-62.

[14] Ri C Y,Yao M.Bayesian network based semantic image-classification with attributed relational graph[J].Multimedia Tools & Applications,2015,74(13):4965-4986.

[15] 吴锐航,李绍滋,邹丰美.基于SIFT特征的图像检索[J].计算机应用研究,2008,25(2):478-481.

[16] 马莉,韩燮.主成分分析法(PCA)在SIFT匹配算法中的应用[J].电视技术,2012,36(1):129-132.

[17] Zhou W,Yang M,Wang X,et al.Scalable feature matching by dual cascaded scalar quantization for image retrieval[J].IEEE Transactions on Pattern Analysis & Machine Intelligence,2016,38(1):159-171.

[18] Zhang J,Feng S,Li D,et al.Image retrieval using the extended salient region[J].Information Sciences,2017,399:154-182.