自动图像标注技术研究进展

鲍 泓[1,2] 徐光美[2] 冯松鹤[1] 须 德[1]
(北京交通大学计算机与信息技术学院 北京 100044)[1] (北京联合大学信息学院 北京 100101)[2]

摘 要 近年来,自动图像标注(Automatic Image Annotation,AIA)技术已经成为图像语义理解研究领域的热点。其基本思想是利用已标注图像集或其他可获得的信息自动学习语义概念空间与视觉特征空间的潜在关联或者映射关系,来预测未知图像的标注。随着机器学习理论的不断发展,包括相关模型、分类器模型等不同的学习模型已经被广泛地应用于自动图像标注研究领域。现有的自动图像标注算法可以大致分为基于分类的标注算法、基于概率关联模型的标注算法以及基于图学习的标注算法等三大类。首先根据自动图像标注算法的特征提取及表示机制不同,将现有算法划分为基于全局特征和基于区域划分的自动图像标注方法。其次,在基于区域划分的自动图像标注算法中,按照学习算法的不同,将其划分为基于分类的标注方法、基于概率关联模型的标注方法以及基于图学习的标注方法,并分别介绍各类别中具有代表性的标注算法及其优缺点。然后给出了自动图像标注最新的研究进展,最后探讨自动图像标注的进一步研究方向。

关键词 自动图像标注,多示例学习,多标记学习,图学习,概率建模
中图法分类号 TP311 **文献标识码** A

Advances in Automatic Image Annotation

BAO Hong[1,2] XU Guang-mei[2] FENG Song-he[1] XU De[1]
(School of Computer & Information Technology, Beijing Jiaotong Upiversity, Beijing 100044, China)[1]
(Information College, Beijing Union University, Beijing 100101, China)[2]

Abstract Automatic image annotation has emerged as a hot topic in the field of image semantic understanding due to its potential application on Web image search. To effectively access and retrieve images, a popular solution is to tag images with meaningful semantic keywords, which is considered as automatic image annotation. Various machine learning techniques have been employed extensively in the field of image analysis, and there is no exception for automatic image annotation. Existing image annotation algorithms can be roughly divided into three categories, i. e., the classification based methods, the probabilistic modeling based methods, and the graph learning based methods, respectively. We surveyed nearly 50 key theoretical and empirical contributions in the current decade related to automatic image annotation, and discussed the spawning of related sub-fields in the process. By carefully analyzing what has been achieved so far, we also conjectured what the future may hold for automatic image annotation research.

Keywords Automatic image annotation, Multi-instance learning, Multi-label learning, Graph learning, Probabilistic modeling

1 引言

随着数码相机和可拍照手机等设备的日益普及,各种各样的图像数量呈现几何级的飞速增长。而同时互联网的快速发展也使得图像传播与共享变得更加快捷。因此,对网络多媒体信息进行有效的管理与检索成为迫切需要解决的问题。虽然基于内容图像检索(Content-based Image Retrieval, CBIR)已经取得了不少的研究成果,但由于受到"语义鸿沟(Semantic Gap)"瓶颈的制约,即低层视觉特征(如颜色、纹理、形状等)不能完全反映和匹配用户的查询意图,导致基于内容图像检索技术的研究遇到了前所未有的巨大挑战,如何真正实现基于语义的图像检索仍旧是一个难题。由于用户更加习惯于利用关键词(Keywords)这种最为直接的方式来表达查询需求,并且现有的互联网搜索引擎均提供基于文本的图像检索功能,而人工标注又是一项相当费时费力的工作,由此催生了自动图像标注技术的发展。

自动图像标注(Automatic Image Annotation,AIA)就是让计算机自动地给无标注的图像加上能够反映图像内容的语

到稿日期:2010-08-05 返修日期:2010-11-12 本文受国家自然科学基金项目(60972145),北京市教育委员会人才强教深化计划(PHR200907120)资助。

鲍 泓(1958—),男,教授,主要研究方向为图像处理,E-mail:baohong@buu.edu.cn;徐光美(1977—),女,博士,讲师,主要研究方向为图像检索与自动标注、数据挖掘;冯松鹤(1981—),男,博士,讲师,主要研究方向为图像处理;须 德(1944—),男,教授,博士生导师,主要研究方向为图像处理。

义关键词。它利用已标注图像集或其他可获得的信息自动学习语义概念空间与视觉特征空间的关系模型，并用此模型标注未知语义的图像，即它试图在图像的高层语义信息和低层特征之间建立一种映射关系，因此在一定程度上可以解决“语义鸿沟”问题。现有的大部分自动图像标注算法，都尝试着直接在图像级别实现语义关键词的标注，即算法无需在图像的区域和关键词之间建立一一对应的映射关系。但也有部分工作试图从物体识别的技术角度去解决标注问题，为一幅图像的每个区域均赋予关键词。据此，我们将前者称之为标注(annotation)，而将后者称之为区域命名(region naming，one-to-one correspondence between words and regions)。图像标注和区域命名的示例参见图1，其中右图表示的是标注，而左图表示的是区域命名。

图1 自动图像标注的两种方式

自动图像标注是图像检索研究领域中非常具有挑战性的工作，是实现图像语义检索的关键。若能实现自动图像标注，则图像检索问题就可以转化为相当成熟的文本检索问题。自动图像标注涉及计算机视觉、机器学习、信息检索等多方面的内容，具有很强的研究价值和潜在的商业应用，如个人桌面照片管理、互联网图像广告自动投放等应用[1]。

Mori等人在1999年提出的共生模型(Co-occurrence Model)[2]，开辟了自动图像标注领域的研究。此后各种新颖的自动图像标注算法不断涌现，众多的研究者从不同的角度分析和解决标注问题，期望能找到良好的检索和标注方法。这些方法从图像的特征表示机制进行分析，大致可以分为两类：一类是使用图像的全局视觉信息，采用面向图像场景语义的方法进行标注，该类方法将图像特征和文本标注词完全分离，在纯视觉层次上比较图像相似度，是有监督的学习方法。已标注的训练图像集合被用于确定图像特征和标注词间的关系，标注可以通过比较视觉特征并传播标注词实现。另一类是首先将图像划分为若干个同质区域或图像子块，再基于区域划分进行图像语义标注。该类方法采用图像分割算法，试图有效地将图像划分为若干个语义对象单元，通过寻找标注关键词与区域语义对象或整幅图像本身间的对应关系来实现自动图像标注。现有主流的标注算法大多采用基于区域划分的表示机制。

本文首先根据自动图像标注算法的特征提取及表示机制的不同，将现有算法划分为基于全局特征和基于区域划分的自动图像标注方法。其次，在基于区域划分的自动图像标注算法中，按照学习算法的不同，将其划分为基于分类的标注方法、基于概率关联模型的标注方法以及基于图学习的标注方法，并分别介绍各类别中具有代表性的标注算法及其优缺点。然后给出了自动图像标注最新的研究进展，最后探讨自动图像标注的进一步研究方向。

2 基于全局特征的自动图像标注方法

早期的基于全局特征的自动图像标注工作等同于图像场景的自动分类。Oliva等人使用面向图像场景语义的方法对图像进行自动标注[3,4]，该方法基于图像的空间属性(如平均深度，尺寸等)产生现实场景(可以是人工场景也可以是自然场景，比如可以是房间内或房间外的地方)的有意义描述。算法验证了全局统计特征(Gist)可以用于分析图像场景中对象的存在与否，从而免去了对图像进行分割和进行面向对象分析的过程。文献[5]提出的是面向显著兴趣点的方法，论文中使用显著区域的局部描述子的向量空间表示来描述图像，并通过相似的图像传播语义来实现自动标注。Yavlinsky等人[6]继续探索了单纯利用图像的全局特征进行语义标注的可能。其建模框架基于鲁棒的非参数密度估计方法，并使用核平滑技术，研究了利用各类全局图像特征对标注性能的影响，也显示EMD(Earth Mover's Distance)距离标准可以与该框架有效整合利用。结果显示其标注性能与推理网络方法和基于CRM[7]的方法性能相当。此外算法也论证了在COREL数据集上单纯利用全局的颜色信息就可以达到较好的标注性能。在图像数据集中两幅图像的视觉特征相似的情形下，全局颜色特征将是建模关键词密度的坚实基础。尽管算法将每幅图像划分为3×3的矩形区域，但该类分割方式属于硬划分(不同于基于内容的分割策略)，因而仍可以看成是基于全局特征的标注算法。

此类方法的优点是可以免除对图像的区域分割、区域聚类、三维注释和面向对象的分析等诸多过程。但通常来说，图像全局特征一般只适用于表示简单的图像或背景较为单一的图像，如纹理图像、自然场景图像、建筑物图像等。由于人眼在观察一幅图像时，总是很自然地将图像分为前景目标和背景区域，因此用户查询时更注重图像内具有一定语义信息的特定目标或者区域，而非背景区域。图像的全局特征只提供粗粒度的语义描述，未考虑到图像中前景物体与背景的差异，因而不能反映图像丰富的细节语义内容，标注的性能也不甚理想。若能将图像的前景目标区域从背景中分割出来，实现对象级的语义描述，则可以减少由于目标物体在图像中的背景变化和场景变化带来的影响，从而更接近语义检索的目标。因此提取区域级的低层视觉特征比全局的视觉特征更加贴近人对图像的语义理解，基于区域划分的图像标注技术(Region-based Image annotation)也就应运而生。

3 基于区域划分的自动图像标注方法

基于区域的自动图像标注方法的基本思想是：首先根据一定的图像分割算法将图像分成若干同质区域，并提取每个区域的低层视觉特征；然后采用机器学习算法建立图像区域和标注词间的语义关联。根据研究者采用的学习方法的不同，可以将基于区域分块的标注算法划分为：基于分类的图像标注、基于概率关联模型的图像标注、基于图学习的图像标注三类。

3.1 基于分类的自动图像标注算法

较为直观的自动图像标注的思路，是将标注问题看成是图像语义分类问题。若将每个语义关键词都看成是一个类别标记(label)，则图像标注问题就转化为图像分类问题。因此完全可以从图像分类的角度去解决标注问题。但不同于传统的图像分类问题中每幅图像只归属于某一语义类别，自动图像标注问题有其特殊性。从关键词的角度分析，在标注问题

中每幅图像同时属于多个语义类别(即标注有多个关键词),因此标注问题属于一个典型的多标记学习问题(Multi-Label Learning)[8]。从图像的角度分析,若将整幅图像看作由多个示例(即区域)组成的包,示例没有概念标记,但包有一个概念标记。如果包中至少有一个示例是正例,则该包被标记为正包,如果包中没有任何一个示例是正例,即所有示例都是反例,则该包被标记为反包。而给定的训练集上关键词均只是标注于整幅图像上,而并不知道关键词与图像区域之间的对应关系,因此标注问题的这一歧义性使得其符合典型的多示例学习(Multi-Instance Learning)问题。现有的基于分类的标注算法大多单纯从多示例学习的角度或者多标记学习的角度来描述和解决标注问题。尽管这些方法在具体表达上各有特点,但它们的核心思想却是一致的,即利用已知的标注数据建立某种模型来描述文本词汇与图像特征之间的潜在关联或者映射关系,并据此预测未知图像的标注。

文献[9,10]将自动图像标注问题看作多标记学习问题,通过将多标记学习问题转化为若干个单标记学习问题,提出了基于支持向量机(Support Vector Machine,SVM)的自动图像标注算法。算法在构造与每个关键词对应的二类分类器时,首先将所有标注该关键词的训练样本图像作为正例样本,而将所有未标注该关键词的训练样本图像作为反例;然后,分别提取正反例图像的全局颜色直方图特征,并据此为给定关键词构建 SVM 分类器;最后给定未标注图像,利用每个关键词的分类器实现对其的分类,选择分类标记结果值最高的前几个关键词作为未标注图像的最终标注结果。由于给定的训练样本图像只给出了关键词与图像的关联,但并没有关键词与图像中区域的对应关系,即训练样本图像中存在不属于该关键词语义的区域,而现有的基于多标记学习的自动标注算法未考虑到标注信息的歧义性,因此最终的标注性能并不理想。文献[11]中提出了上下文相关的关键词传播方法,该方法使用了多标记学习方法并借用线性规划方法来提高标注性能,该方法能够同时传播多个关键词。

由于训练图像集合并不提供区域级别的标注信息,即关键词是与整幅图像相关联而不是与图像中的区域关联,因此在图像标注领域,标注有某个关键词的正例样本图像中也会存在伪示例。多示例学习(Multi-Instance Learning)[12]作为一种泛化的监督学习算法,能较好地处理这种歧义性问题,因此很自然地被引入到自动图像标注问题中。文献[13]提出了多示例学习领域经典的多样性密度(Diverse Density)算法来解决标注问题。算法的基本思想是,如果特征空间中某点最能表征某个给定关键词的语义,那么正包中应该至少存在一个示例靠近该点,而反包中的所有示例应该远离该点。因此该点周围应当密集分布属于多个不同正包的示例,同时远离所有反包中的示例。特征空间中如果某点附近出现来自于不同正包中的示例越多,反包中的示例离得越远,则这点表征了给定关键词语义的概率就越大。用多样性密度来度量这种概率,具有最大概率的点即为要寻找的目标点。算法的缺点在于:首先,由于关键词语义的丰富性,很难用唯一的特征向量来表征其语义;其次,多样性密度算法需要将每一个正包示例都作为初始点进行一次搜索,且要进行多次梯度下降搜索以求解最优值,因此其训练时间开销相当大。文献[14]提出了基于非对称支持向量机的多示例学习(Asymmetrical Support Vector Machine Based Multiple Instance Learning,ASVM-MIL)算法,它将自动图像标注任务转化为监督学习。算法考虑了包的歧义性,通过最小化包的分类误差将 SVM 直接应用到多示例学习问题中。

文献[15,16]也提出了基于多示例学习思想的自动标注算法。采用层次化高斯混合模型算法(Mixture Hierarchic Gaussian Model,Mix-Hier)来估计每个关键词在特征空间中对应的特征分布。算法首先收集每个关键词所对应的正例图像集合,并将每幅正例图像以包的形式表示。然后对每幅正例图像采用高斯混合模型进行建模,在此基础上利用每个正例图像的特征分布作为输入,再次利用高斯混合模型对整个正例图像集合进行语义建模。最后选取特征空间中概率密度分布最高的视觉特征向量来表征关键词。文献[17,18]也是基于分类进行图像自动标注的尝试。文献[19]综合考虑了标注问题输入空间和输出空间的歧义性,将多示例学习和多标记问题两者融合起来完成标注算法。

基于分类的图像标注算法的基本流程如图 2 所示。

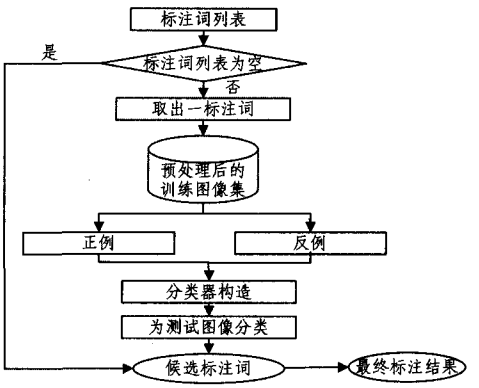

图 2　基于分类的图像自动标注算法的一般流程

3.2　基于概率关联模型的自动图像标注算法

基于概率关联模型的图像标注算法,其本质是在概率统计模型的基础上,分析图像区域特征与语义关键词之间的共生概率关系,并以此为待标注图像进行语义标注。直观地,两幅图像若具有较高的视觉相似性,则两者标注相近关键词序列的概率就越高。这种方法的特点在于,无需通过学习机制为每个语义关键词建立相应的低层视觉特征表示。换句话说,语义关键词与低层视觉特征之间不存在一一对应的映射关系。基于概率关联模型的自动图像标注算法的一般流程如图 3 所示。

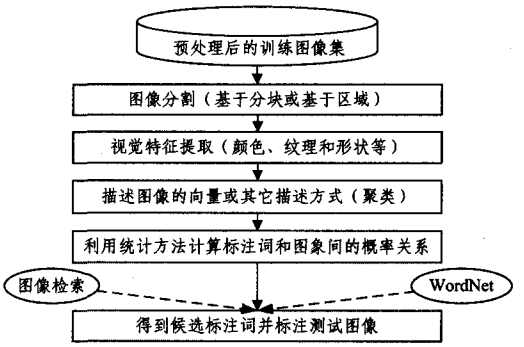

图 3　基于统计方法的图像自动标注过程

文献中各方法的区别在于计算标注词和图像间概率关系所采用的统计方法不同。虚线部分表示部分算法借助图像检

索的结果或由 WordNet 得到的标注词本身间的相互关系来决定最终标注结果。

文献[20]最早提出了基于机器翻译模型来解决图像标注问题，算法假设文本标注与视觉特征是用以描述同一图像内容的两种不同语言。基于此假设，它引入了自然语言中的双语翻译模型，将图像标注的过程视为从“视觉”语言到“文本”语言的翻译问题。其中，视觉词汇是由图像的各个分割区域经过聚类的结果，被称为“blob”；而文本词汇就是标注关键词，然后他们利用机器语言翻译的方法建立起 blob 与文本词汇之间的对应关系，进而得到图像的语义标注。由于该算法的标注结果偏重于在训练图像集中出现频率较高的关键词，因此为克服这一问题，Kang 等人先后提出了两种改进方案：一种是基于对称的翻译模型[21]，该模型将由视觉词汇到文本词汇的翻译结果和由文本词汇到视觉词汇的翻译结果进行融合；另一种则通过对翻译概率规则化来克服词频的影响[22]。

相关模型是目前基于概率关联模型的自动图像标注领域最重要的算法之一，许多后续的标注算法都是基于相关模型进行改进和提高的。其基本思想主要是建立图像和语义关键词之间的概率相关模型。算法通过为某一幅待标注图像找到与其相关性最大的一组语义关键词，来获得图像的标注结果。测试图像标注关键词的概率由该测试图像的所有分割区域共同决定，即通过乘积的方式来得到测试图像的每一个区域与训练集中每个图像的视觉相似性。而对测试图像标注结果影响较大的通常是与其相似度较高的训练图像集合，而与其相似度较小的训练图像对其标注结果的影响通常较小。

文献[23]将图像标注问题看作是跨语言检索问题，从而提出了跨媒体概率相关模型(Cross Media Relevance Model，CMRM)。由于 CMRM 模型采用图像子块(blob)来表征图像的语义内容，而 blob 是采用区域聚类后离散化的方式生成的，因而这种离散化的表示会造成视觉特征内容的损失，影响标注效果。针对这一问题，文献[7]提出了一种基于图像连续特征的相关模型(Continuous-space Relevance Model，CRM)。CRM 利用图像各分割区域的连续特征向量组合来表示图像，然后通过高斯核函数估计区域间的相似关系。CMRM 模型中对区域特征进行了离散化操作，而 CRM 直接使用连续特征建模，因此不依赖于聚类从而避免了粒度问题。CRM 与 CMRM 算法的表示形式极为相似，但最大的不同在于 CRM 在图像连续特征空间比较两幅图像的相似性，而 CMRM 则使用聚类算法生成 blob 来表示图像内容，由于聚类过程本身会带来信息缺失，因而 CRM 算法效果更好。

针对 CMRM 和 CRM 算法存在的不足，文献[24]提出了多重伯努利相关模型(Multiple-Bernoulli Relevance Model，MBRM)，算法针对前面两种模型进行了改进。首先，由于图像分割算法(Normalized-Cut)计算复杂，MBRM 采用了简单的网格划分图像的方法，将图像切分为规则的矩形区域，简化了计算复杂度，实验验证了这一改进的有效性。另外，不同于 CMRM 和 CRM 算法采取多项式分布来估计，MBRM 则引入了多重伯努利分布来估计词汇的概率分布。由于采取多项式分布，在词汇标注时暗含所有关键词出现概率之和为 1 的约束条件，导致各词汇在图像标注任务中存在排斥的关系。而在图像标注任务中，通常强调的是一个词是否应当被用来描述该图像，即强调关键词的“存在性”，因此多重伯努利分布比多项式分布更加适合描述关键词的分布概率。文献[25]给出了一种基于贝叶斯理论的图像标注和检索方法，文献[26]给出了一个融合图像内容和上下文信息的图像标注框架，并且图像标注是区域级别的标注，其标注过程并不依赖于分块的大小。文献[27,28]也是该类方法的典型代表。

3.3 基于图学习的自动图像标注算法

近年来，基于图学习(Graph Learning)的方法作为一种重要的机器学习算法，已经被用来有效地解决图像自动标注这一图像语义理解问题。基于图学习的算法是一种半监督学习算法，已知类标的训练数据和未知类标的测试数据都将参与到算法的学习过程中。与传统的有监督学习和无监督学习相比，半监督学习可以在学习阶段利用更多的信息，如数据的分布特性等，它适用于总数据量较大、已标记训练数据量相对较小的情况。若我们将每幅图像(或每个标注词)作为图节点，以图像间(或标注词间)的相似关系作为边，通过图学习算法就可以实现标注信息从已标注图像到未知图像的传播，从而完成图像标注任务。

文献[29]首次提出了一种基于图的自动标注方法(CGap)，图像、标注关键词和同质区域被分别表示为三类不同的图结点，并根据它们之间的相互关系连接成图。文献[30]中给出了一种基于流形排序(Manifold-Ranking)的图像标注方法，该方法同时考虑了视觉信息和文本信息，并用由 WordNet 获得的词间的关系来为图剪枝。在该框架下，图像标注被分为两个阶段来完成，即基本图像标注与图像标注改善。其中，前者是通过以图像间相似性为依据的图学习过程来提供图像的初始标注，而后者是通过以词汇间语义相关性为依据的图学习过程来改善前者取得的标注结果。文献[31]通过视觉相似度来标注关键词，该模型只利用图像间的相似度来构建 k-NN 相似图，而没有考虑词间的相关性。文献[32]提出了一个基于图模型的最近邻生成链(Nearest Spanning Chain，NSC)来标注图像，模型给出了图像相似性的统计估计。文献[33]中给出了一个基于图学习的图像标注算法框架，并进一步改进了现有的 NSC 方法。该框架同时考虑了训练集合中的词共生关系和 Web 上下文中的词共生关系。

基于图学习的图像标注算法的流程如图 4 所示。其中虚线部分含义是指该步骤是可选项，即表示仅有部分已有算法包含该步骤。

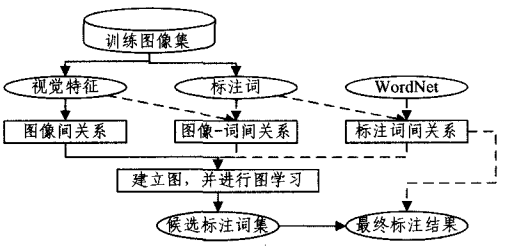

图 4 基于图学习的图像标注算法流程示意图

4 进一步的研究进展

利用图像的自动标注来实现图像的语义理解已成为当前的研究热点与重点。虽然图像标注工作已经取得了很大的发展，并提出很多的图像标注算法，但仍然不能满足用户的需求。针对这样的现状，自动图像标注衍生了两种新的研究课题：图像标注改善(Image Annotation Refinement)和基于

Search 的标注算法。图像标注改善方法是指针对在基本图像标注过程中得到的初步标注结果，通过利用标注词汇间的关联关系，去除不相关的词汇并填补上可能遗漏的词汇，从而保证最终的标注结果具有良好的语义一致性[32]。而基于 Search 的图像标注算法通过有效地融合 CBIR 技术，将未标注的图像看作是查询图像，根据检索技术找到查询图像的一些相关图像集合，然后从相关图像的标注词的集合中，应用文本分析技术挖掘出标注结果。

4.1 图像标注改善

图像标注改善是自动图像标注过程的重要步骤。基本图像标注阶段得到的标注信息虽然获得了较好的结果，但通常情况下并不能很好地反映图像的语义信息。在基本图像标注阶段得到的候选标注信息可能是不完整的，或者包含了一些与图像不相关的标注信息。这主要是由于现有的标注算法将每个语义关键词单独分析，并没有考虑到关键词之间的语义关联。而通常情况下，词汇与词汇之间的语义联系还是非常紧密的，通常词汇间包括层次关系和相关性信息。例如“tiger”和“grass”两个词的语义联系比较紧密，当一幅图像标有关键词“tiger”时，其标有“grass”的概率也相应提高。因此利用词汇与词汇之间的相互关系，从候选词汇中挑选出紧密相关的词汇，滤除那些无关的噪声词汇，是改善图像标注性能的重要手段之一。许多研究工作将词汇间的相关性融入到模型的估计过程中，文献[34]提出了一致性语言模型(Coherent Language Model，CLM)，文献[35]提出了互相关标记传播模型(Correlated Label Propagation，CLP)；另外，Wang 等人[28]尝试着先假设单词间相互独立来完成基本的图像标注，然后再结合词间相关性对前一过程得到的标注结果进行改善。文献[36]尝试着在相关模型的基础上融合全局特征、局部特征以及文本上下文信息来完成图像标注工作。

4.2 基于 Search 的自动图像标注算法

文献[37，38]提出了一种基于 Search 的图像标注算法(Search Based Image Annotation，SBIA)。算法属于一种数据驱动且与模型无关的标注算法，其基本假设是，在理想情况下若图像库足够大可使得任一待标注图像都能从图像库中找到与之完全一样或者几乎一样的图像，则相关模型的优化问题就变成简单的近邻传播问题。文献[38]提出融合检索技术进行标注的方法，即 AnnoSearch 方法，算法首先将未标注图像作为查询图像，由用户给查询图像提供一个初始的标注词。然后，根据基于文本的图像检索技术，在 Web 中检索到与查询图像相关的图像集合，同时也得到一个相关图像的标注词集合。最后，对这个标注词集合进行聚类，给出相关标注词的排序列表，从中决定查询图像的标注结果。该方法的检索精度依赖于用户提供的初始标注词，因此，在一定程度增加了用户的负担，而且还具有用户的主观性。为了简化标注过程，文献[37]改进了文献[38]的工作，该方法无须用户提供初始标注词，实现了检索与标注的全自动化。基于 Search 的标注方法避免了复杂的参数学习的过程。而且，由于通过检索找到相关的图像，因此该方法不受训练集或者标注词集合的限制。

5 展望

自动图像标注技术通过语义关键词来表征图像的语义信息，是提高图像检索性能的重要手段，在一定程度上弥合了“语义鸿沟”。虽然自动图像标注工作已经取得了很大的发展，但受限于多种困难，该领域现有研究成果与实际应用还有比较大的距离，从规模和质量上仍然不能满足用户的需求，尤其当数据集规模较大或者训练图像样本不足时大多数算法的性能会急剧下降。因此，自动图像标注的研究仍存在很大的改进空间。

进一步的潜在的研究方向包括：

(1) 更有表征性的特征提取及选择

目前，低层视觉特征的提取仍然是基于内容的图像检索及自动图像标注的基础。由于图像低层特征与其本身所包含的高级语义之间存在着巨大差距，使得基于内容的图像分析和检索还未取得令人满意的效果。而填补低层特征与高级语义之间的鸿沟是多媒体信息检索中最具挑战性的课题。因此，采取何种特征描述来表达图像的语义信息，采取何种基于语义的图像分割算法，使之能够有效地表征用户对图像内容的感知，是提高图像自动标注性能的重要手段。

(2) 基于 Web 分析的文本语义挖掘与融合

虽然图像标注工作已经取得了很大的发展，并提出了许多相关算法，但是从规模和质量上仍然不能满足用户的要求，这主要是因为现有的算法大多要求给定许多已标注图像作为训练样本，然后通过某种机器学习算法来得到相应的模型，从而完成对未知图像的标注。这种对训练集的依赖一方面需要对训练集的质量提出较高的要求，从而导致方法本身的可推广性受到很大的限制，另一方面依旧无法摆脱“语义鸿沟”问题。基于此，如何利用互联网搜索技术，在整个网络索引数据库范围内，将图像检索技术、网络搜索技术融入到图像标注任务中，用以克服传统方法对训练集数据的依赖，并从一定程度上缓解“语义鸿沟”的障碍以及巨大的图像数量带来的可推广型问题等，是下一步研究图像自动标注的新的切入点。

基于内容的图像语义分析和检索是一个跨学科的、富有挑战性的研究课题。随着成像设备的迅速普及与网络技术的飞速发展，在未来的日子里，图像数量将会呈现极速的增长，使得人们对图像进行有效组织和快速搜索的需求日渐迫切。在这个领域中，众多的研究者从不同的角度进行探索，期望能找到良好的检索和标注方法。虽然面临着一个困境，但这个领域仍有许多需要发展的技术和很多值得研究的课题。向深度挖掘的课题，如视知觉理论在图像内容分析中的应用，向广度延伸的方向，如网络多媒体搜索技术，都会丰富这个领域的研究；而研究与具体应用的结合，如为医学图像等提供专门的解决方案，也将推动这个研究领域的进步。

参考文献

[1] 赵玉凤. 图像检索中自动标注技术的研究[D]. 北京：北京交通大学，2009

[2] Mori Y, Takahashi H, Oka R. Image-to-word transformation based on dividing and vector quantizing images with words[C]// Proc. ofIntl. Workshop on Multimedia Intelligent Storage and Retrieval Management (MISRM'99). Orlando, Oct. 1999

[3] Oliva A, Torralba A. Modeling the shape of the scene: A holistic representation of the spatial envelope[J]. Int. J. Comput. Vision, 2001, 42(3): 145-175

[4] Oliva A, Torralba A B. Scene-centered description from spatial envelope properties[C]// BMCV '02: Proceedings of the Second International Workshop on Biologically Motivated Computer Vi-

sion. London, UK: Springer-Verlag, 2002: 263-272

[5] Hare J S, Lewis P H. Saliency-based models of image content and their application to auto-annotation by semantic propagation [C]// Proceedings of the Second European Semantic Web Conference (ESWC2005). Heraklion, Crete, May 2005

[6] Yavlinsky A, Schofield E, Rüger S. Automated Image Annotation Using Global Features and Robust Nonparametric Density Estimation[C] // Proceedings of the 4th International Conference on Image and Video Retrieval. Polani D, Browning B, Bonarini A, eds. Lecture Notes in Computer Science 3568, Singapore: Springer-Verlag, July 2005: 507-517

[7] Lavrenko V, Manmatha R, Jeon J. A model for learning the semantics of pictures[C]// Proc. of Advances in Neural Information Processing Systems (NIPS'03). 2003

[8] Kang F, Jin R, Sukthankar R. Correlated label propagation with application to multi-label learning[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition. 2006: 1719-1726

[9] Tang J, Lewis P H. A study of quality issues for image auto-annotation with the Corel dataset[J]. IEEE Trans. on Circuits and Systems for Video Technology, 2007, 17(3): 384-389

[10] Cusano C, Ciocca G, Schettini R. Image annotation using SVM [C]// Proc. of Int. SPIE Conf. on Imaging IV. San Jose, CA, USA, Feb. 2004: 330-338

[11] Lu Zhi-wu, Horace H S I, He Qi-zhen. Context-based multi-label image annotation [C] // Proceeding of the ACM International Conference on Image and Video Retrieval. Santorini, Fira, Greece, July 2009

[12] Maron O, Lozano-Perez T. Multiple-instance learning for natural scene classification[C]// Proc. of Int. Conf. on Machine Learning (ICML'98). Madison, Wisconsin, USA, July 1998: 341-349

[13] Yang C, Dong M, Fotouhi F. Region-based image annotation through multiple instance learning[C]// Proc. of ACM Conf. on Multimedia (ACM MM'05). Singapore, Nov. 2005: 435-438

[14] Yang C, Dong M, Hua J. Region-based image annotation using asymmetrical support vector machine-based multiple-instance learning[C]// Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'06). New York, USA, June 2006: 2057-2063

[15] Gustavo C, Nuno V. A database centric view of semantic image annotation and retrieval[C]// Proc. of Int. ACM SIGIR Conf. on Retrieval (ACM SIGIR'05). Salvador, Brazil, Aug. 2005: 559-566

[16] Carneiro G, Chan A B, Moreno P J, et al. Supervised learning of semantic classes for image annotation and retrieval[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007, 29 (3): 394-410

[17] Qi X, Han Y. Incorporating multiple SVMs for automatic image annotation[J]. Pattern Recognition, 2007, 40(2): 728-741

[18] Tang J H, Hua X, Qi G, et al. Typicality ranking via semi-supervised multiple-instance learning [C] // Proc. of ACM Conf. on Multimedia (ACM MM'07). Augsburg, Germany, Sep. 2007: 297-300

[19] Feng Song-he, Xu De. Transductive Multi-Instance Multi-Label Learning Algorithm with Application to Automatic Image Annotation[J]. Expert Systems with Applications, 2010, 37(1): 661-670

[20] Duygulu P, Barnard K, Freitas N, et al. Object recognition as machine translation: learning a lexicon for a fixed image vocabulary[C]// Proc. of European Conf. on Computer Vision (ECCV'02). Copenhagen, Denmark, May 2002: 97-112

[21] Kang F, Jin F. Symmetric statistical translation models for automatic image annotation[C]// Proc. of SIAM Conf. on Data Mining. Newport Beach, CA, Apr. 2005: 21-23

[22] Kang F, Jin R, Chai J. Regularizing translation models for better automatic image annotation[C]// Proc. of Int. Conf. on Information and Knowledge Management. Washington, D. C., USA, Nov. 2004: 350-359

[23] Jeon J, Lavrenko V, Manmatha R. Automatic image annotation and retrieval using cross-media relevance models[C]// Proc. of Int. ACM SIGIR Conf. on Research and Development in Information Retrieval (ACM SIGIR'03). Toronto, Canada, July 2003: 119-126

[24] Feng S, Manmatha R, Lavrenko V. Multiple bernoulli relevance models for image and video annotation[C]// Proc. of IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR'04). Washington DC, USA, June 2004: 1002-1009

[25] 张元清，包骏杰，况夯，等. 基于贝叶斯理论的图像标注和检索[J]. 计算机科学，2008(8)

[26] Lu Hong, Zheng Ying-bin, Xue Xiang-yang, et al. Content and context-based multi-label image annotation[C]// IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. Miami, FL, USA. June 2009: 61-68

[27] Monay F, Gatica-Perez D. Modeling semantic aspects for cross-media image indexing[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2007, 29(10): 1802-1817

[28] Wang C, Jing F, Zhang L, et al. Image annotation refinement using random walk with restarts[C]// Proc. of ACM Int. Conf. on Multimedia (ACM Multimedia'06). Santa Barbara, CA, Oct. 2006: 647-650

[29] Pan J Y, Yang H J, Pinar D. Automatic multimedia cross-modal correlation discovery[C]// The Tenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. August 2004: 653-658

[30] Liu J, Li M J, Ma W, et al. An adaptive graph model for automatic image annotation[C]// Eighth ACM International Workshop on Multimedia Information Retrieval. 2006: 61-70

[31] Tong H, He J, Li M, et al. Manifold-ranking based keyword propagation for image retrieval[J]. EURASIP J. Appl. Signal Process. Spec. Issue Inf. Min. Multimedia Database, 2006, 21: 1-10

[32] Liu J, Li M, Liu Q, et al. Image annotation via graph learning [J]. Pattern Recognition, 2009, 42(2): 218-228

[33] 卢汉清，刘静. 基于图学习的自动图像标注[J]. 计算机学报，2008, 9(31): 1629-1639

[34] Li X, Chen L, Zhang L, et al. Image annotation by large-scale content-based image retrieval[C]// Proc. of the 14th ACM Multimedia (ACM Multimedia'06). Santa Barbara, USA, Oct. 2006: 607-610

[35] Kang F, Jin R, Sukthankar R. Correlated label propagation with application to multi-label learning[C]// Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'06). New York, USA, June 2006: 1719-1726

[36] Wang Y, Mei T, Gong S G, et al. Combining global, regional and contextual features for automatic image annotation[J]. Pattern Recognition, 2009, 42(2): 259-266

[37] Wang X J, Zhang L, Li X, et al. Annotating images by mining image search results[J]. IEEE Trans. on Pattern Analysis and Machine Intelligence, 2008, 30(11): 1919-1932

[38] Wang X J, Zhang L, et al. AnnoSearch: Image auto-annotation by search[C]// Proc. of IEEE Computer Society Conf. on Computer Vision and Pattern Recognition (CVPR'06). New York, USA, June 2006: 1483-1490