DOI：10.13998/j.cnki.issn1002-1248.24-0109

融合事件和情感的图像语义描述框架研究

胡守敏[1]，董焕晴[2]
（1. 华中师范大学 图书馆，武汉 430079；2. 华中师范大学 信息管理学院，武汉 430079）

摘　要：[目的／意义]针对图像组织和检索过程中存在的语义缺失和不完整性问题，提出一个面向社会化媒体中的图像语义描述框架，旨在丰富现有的图像描述理论体系，提高图像的检索效率和利用率，为实现自动化的图像语义标注提供参考。[方法／过程]首先，调研分析国内外有关图像描述的研究进展，总结现有的图像描述和标注理论、元数据规范和相关技术方法；其次，在此理论基础上，针对社会化媒体图像领域，构建社会化媒体图像语义描述框架，并详细阐述语义层次及其相互关系。最后，通过人物图像和风景图像实例描述验证图像语义描述框架的可行性。[结果／结论]人物图像和风景图像描述实例结果表明，图像语义描述框架可通过各层之间的语义关联消除图像描述中的“语义鸿沟”，实现对图像外部特征和内容特征的多侧面、多维度、多层次的结构化和语义化描述，具有较强的可移植性和灵活性。

关键词：语义描述框架；图像特征；语义标注；Sora

中图分类号：G251.3；TP39　　**文献标识码：**A　　**文章编号：**1002-1248（2024）02-0051-10

引用本文：胡守敏，董焕晴. 融合事件和情感的图像语义描述框架研究[J]. 农业图书情报学报，2024，36(2)：51-60.

0 引　言

图像是人们日常工作、学习和生活广泛使用的多媒体信息资源类型之一。互联网技术的发展与社交网络的成熟，促使数字图像数量迅速增长，美国新兴数据分析和商业智能公司 DOMO 发布的《Data Never Sleep 4.0》信息图揭示，2016 年的每一分钟，Snapchat 用户会发布 284 722 张照片，Instagram 用户会发布 123 060 张照片[1]。海量的图像中蕴藏着巨大的商业价值，如何更好地对图像信息进行描述、组织、存储和检索引起了工业界和学术界的普遍关注。同时，随着人工智能技术的飞速进步和深度学习领域的不断拓展，大模型的数量和种类呈现出爆发式的增长。这一趋势不仅体现在自然语言处理领域，也逐渐深入到了图像和视频处理领域，为图像语义描述领域带来了机遇与挑战。在这个背景下，OpenAI 推出了文生视频大模型 Sora，其核心价值在于能够根据用户的文本提示快速生成高质量、逼真的视频内容。然而要更好地实现这一功能，Sora 需要深入理解和解析图像中的丰富信息，对图像进行高质量的描述，从而为用户提供更丰富、更精准的视觉信息处理和生成能力。

目前，图书情报领域普遍采用元数据的方法对图像进行描述、著录，但元数据主要是对图像外部特征进行描述，在图像内容信息的描述上缺乏必要的规范，

收稿日期：2024-01-11

基金项目：国家社会科学基金项目“基于事理图谱的社会化问答知识组织与服务研究”（19BTQ075）

作者简介：胡守敏（1981-），女，博士，华中师范大学图书馆，研究方向为知识组织、信息资源管理。董焕晴（1995-），女，博士研究生，研究方向为知识组织、情报分析、数据治理

DOI：10.13998/j.cnki.issn1002-1248.24-0109

再加上底层特征与高层语义之间存在明显的“语义鸿沟”，导致对图像的检索达不到理想的效果[2]。其中，对具有检索意义的特征进行揭示的图像标引工作是建立图像检索系统的前提和基础[3]。与元数据方法不同，图像的语义描述不仅能够对图像的外部特征进行描述，通常还会更多地考虑用户对图像的认知，从多个维度对图像特征进行揭示。国内外学者在图像语义描述领域都有一定的研究成果，但缺乏一套面向社会化媒体的数字图像描述规范，由此导致社会化媒体上的图像标注结果无法统一，基于机器学习的自动标引更是不能满足要求。因此，本文首先通过对数字图像描述方法进行综述，然后针对当前社会化媒体上发布的图像提出一套图像语义描述框架，再基于图像元数据理论，完善语义描述框架下的图像内容描述方法。以期规范社会化媒体数字图像描述，提高数字图像的检索效率，为 Sora 等大模型在图像理解和语义分析方面的深入研究提供理论参考。

1 相关研究

元数据是资源组织最常用的手段，国内外学者对于图像元数据的研究主要集中在对图像元数据标准的创建和选择上。目前，适用于图像描述的元数据标准主要有图像专用的元数据标准、文化资源相关的元数据标准以及通用元数据标准[3]，前者包括 IPTC 照片元数据标准，该标准由国际出版电讯委员会提出并制定，2016 年的最新版本包括标题、创作者、场景代码、主题代码、创作时间、城市等 23 个核心元素以及照片中的人物地点、照片供应商、照片注册表等 34 个扩展元素[4]；与文化资源相关的元数据标准的描述对象主要是馆藏中的实物资源以及派生的其他数字对象，常用的标准有 VRA Core、CDWA、REACH 等[3]；通用元数据标准包括面向多媒体信息的标准（MPGE-7）、面向各类信息资源的标准（DC 元数据标准）以及编目标准（RDA、AACR2R）等。另一个角度是从用户对图像的检索需求来研究图像的描述与标注，该类研究主要通过考察用户如何对图像进行标注以及用户对图像的检索需求，获悉用户对图像的深层语义理解，由于该内容主要来自于大众用户，其检索需求的描述范围相对比较全面，研究成果也较为丰富[5-9]。

此外，也有学者从图像特征的分层、分类现象对图像特征类型进行梳理。PANOFSKY[3]在研究文艺复兴时期的艺术作品时提出艺术作品在内容表现上有 3个层次，分别是前图像志（The Pre-Iconographical）、图像志（The Iconographical）和图像学（The Iconographical Interpretation）。其中，前图像志指图片主题，又分为事实和情感，前者是指根据一般的知识经验就可以解读图片表现的对象和事件，后者则指图片传达的情绪；图像志是图像所表现的可辨别出名称的客观事物，也分为可命名的客观事物及所表征的抽象、象征意义两类；图像学是指图像内容的内在含义，需要综合图像所处的艺术、文化及社会环境，甚至创作者个人特质进行理解。在 E.Panofsky 的研究基础上，S.Shatford[10]从图像语义的抽象程度和图像表现的内容两个维度共同描述图像特征，图像语义的抽象程度分为一般（Generic of）、专指（Specific of）和抽象（Abstract），图像表现的内容分为对象、事件、地点和时间，由此形成一个具有 3×4 类图像特征的模型。JAIMES 和 CHANG[11]综合心理学、图书馆学与艺术领域的研究提出了更为细致的 10 层模型，模型 1~4 层属于语法 / 知觉层，描述人们对图像的直观感受，包括技术类型、全局特征、局部特征和全局组成。5~10 层是结合 Panofsky/Shatford 模型构建的语义 / 概念层，包括一般对象、一般场景、专指对象、专指场景、抽象对象、抽象场景。FAUZI 和 BELKHATIR[12]提出了面向互联网图像的 5 层描述框架：信号层对应图像的底层视觉信息；对象层对应图像中的实体；场景层对应图像的整体特征；关系层对应图像中各对象间的关系以及创作者、创作时间、摄影师等外部关系；抽象层对应情感、质量等抽象概念。国内学者在此研究领域也有一定成果：王惠锋从用户检索的角度提出了面向对象的图像模型[13]；吴楠在深入研究图像高层语义的低层特征描述的基础上提出了图像语义的层次划分，并对每个高层语义层提出了语义抽取和检索算法[14]；王晓

光针对敦煌壁画数字图像，提出了语义描述框架和领域主题词表相结合的数字图像内容语义描述方法[2]。

通过文献分析发现，图像的元数据标准适合于图像外部特征的浅层描述，对于图像深层语义内容揭示不足；利用用户对图像的检索需求虽可进一步挖掘图像的深层语义，但太受限于用户性格、知识背景等；图像特征的分层分类描述能够更全面地对图像进行揭示，但目前的描述框架仍有不完善的地方，并且更多的是对图像内容的“粗描述”，不能覆盖当前社会媒体上发表图片的描述，表 1 所示是图像分层描述理论的层次覆盖表。另外，随着人工智能领域对情感计算和事件识别技术的发展，图像语义描述可通过进一步融合事件和情感分析的能力，这种融合不仅能够增强模型对图像内容的深层次理解，还能够提升描述的人性化和情感表达，使得生成的语义描述更加贴近人类的感知和表达习惯。因此，本文结合图像的元数据标准和图像特征的分层分类描述理论，融合图像的事件与情感特征，构建一套较为完善“细粒度”的图像语义描述框架具有一定的理论与实践意义。

2 图像语义描述框架的构建

图像一种视觉信息的载体，它通过像素点的排列组合来呈现自然景物、人物或抽象概念。它不仅是人类感知世界的重要手段，也是信息传播和文化表达的重要媒介。在数字化时代，图像以其直观性、生动性和易理解性等特点，成为信息传播和交流的重要工具。随着科技的发展，社会化媒体，作为一种新型的网络交流平台，以其开放性、互动性和即时性等特点，改变了人们的信息获取和交流方式。社会化媒体逐渐渗透了人们的日常生活，伴随而来的则是海量的社会化媒体图像。社会化媒体图像，则是图像在社会化媒体这一特定环境下的表现形式。作为社会化媒体内容的重要组成部分，它不仅包含了图像的基本属性，还融入了社会化媒体的特性，成为了一种独特的视觉表达方式。相较于普通图像，社会化媒体图像在事件性和情感性更强。

图像的外部特征能够很好地揭示图像的固有属性，如图像大小、格式、创作者与创作时间等。但仅有外部特征还不足以对图像进行完整的揭示，还需借助图像本身所具有的颜色、纹理和形状等内容特征与高层语义特征；考虑到社会化媒体图像的事件性与情感性特征，因而，在对社会化媒体图像进行语义描述时，也需要借助其事件性和情感性特征。对象、场景和事件作为图像重要的语义内容，是知识的浓缩。图像情感更是人对客观图像所产生的主观意识。基于此，本文从外部特征层、内容层、对象层、关系层、场景层、事件层、情感层构建图像语义描述框架，如图 1 所示。从图 1 框架中可知，图像的外部特征、内容特征与高层语义特征分别从不同的维度对图像进行揭示，但它们之间并不是孤立存在的，只有将其有效结合才能更好地对图像进行描述、组织和管理。

表 1 图像分层描述理论层次覆盖表

Table 1 Image hierarchical description theory hierarchical coverage

研究者	外部特征层	内容层	对象层	关系层	场景层	事件层	情感层
PANOFSKY			✓			✓	✓
SHATFORD			✓		✓	✓	
JAIMES		✓	✓		✓		
王惠锋		✓	✓	✓	✓	✓	✓
HOLLINK	✓	✓	✓		✓	✓	
吴楠		✓	✓	✓	✓	✓	✓
FAUZI		✓	✓	✓	✓		✓
王晓光		✓	✓	✓	✓	✓	✓
笔者	✓	✓	✓	✓	✓	✓	✓

DOI：10.13998/j.cnki.issn1002-1248.24-0109

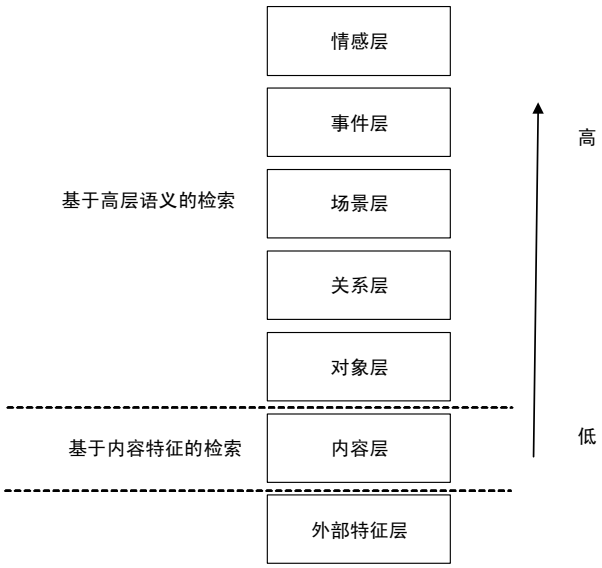

图 1　图像语义描述框架图

Fig.1　Framework for semantic description of images

2.1　外部特征层

外部特征层描述图像的固有属性，如图像大小、格式、创作者、创作时间等，该层对图像资源的获取、共享、保存和管理具有重要意义，许多元数据标准都能很好地对其进行描述。MPEG-7[15]是由 ISO 国际标准化组织针对多媒体内容描述提供的一套标准框架，从功能上分为内容管理（Content Management）、内容组织（Content Organization）、导航与访问（Navigation and Access）、用户交互（User Interaction）、基本元素（Basic Elements）5 个部分，其中，内容管理模块可处理与多媒体文档的创建、媒体所有权和描述等相关的信息，具体描述如表 2 所示。

2.2　内容层

内容层描述图像的底层视觉特征，包括颜色、纹理和形状等，随着图像学与机器视觉的发展，底层特征都能通过机器自动识别，对其描述也已形成标准，比较常用的是 MPEG-7[15]中对于视觉特征的描述符：颜色空间、直方图及结构直方图描述颜色；纹理浏览描述图像的粗糙度、方向感；对象边界区域及基于区域的形状描述形状。观察发现，社会化媒体上拥有大量带有文字信息的图像，这些文字信息往往能对图像的深层语义进行较为准确的揭示，因此，有必要对图像中的文字信息进行识别与提取。

2.3　对象层

对象层描述图像中出现的实体，是整个语义描述框架中最重要的部分。JÖRGENSEN[16]通过实证发现用户最常描述图像中包含的对象，RANSOM 和 RAFFERTY[5]通过分析 Flickr 上的图像标签发现人物、对象类标签最多。以往的研究倾向于将对象分为一般、专指、抽象，缺少更细粒度的划分。本文对象层描述图像中的对象及对象属性（表 3），社会化媒体上发布的图像包含的对象主要有生物（人、动物、植物等）、非生物（建筑物、物品等）以及从现实生活中抽象出的

表 3　对象类别及对象属性

Table 3　Object classes and object properties

对象	对象子类	对象属性示例
生物	人物	名称、性别、年龄阶段、动作
	动物	物种、动作
	植物	物种、颜色
非生物	建筑物	名称、形状、颜色
	物品	名称、形状、颜色
非现实事物	神话人物	名称、性别、动作
	漫画人物	名称、性别、动作

表 2　MPEG-7 图像外部特征的描述维度

Table 2　Description dimension of MPEG-7 image external features

描述维度	描述内容及说明
产生和制作	描述一些与内容的产生和制作相关的信息，包括标题、创作者、分类等级、创作目的等。这些信息大部分是作者产生的，而不能从内容中摘要出来
用法	描述与内容使用方面的信息，包括著作权、使用权、发行权等。这些信息很有可能会在图像管理期间发生变化
媒介	存储介质的描述，包括存储格式、多媒体内容的编码方式、媒介的鉴定方式等

实体（神话人物、漫画人物等），同时，可将对象描述为一般对象、专指对象或抽象对象；属性是指图像中的对象自身具有的属性，如人物有名称、性别、年龄、动作等属性。需要注意的是，有些对象是可以细分的，如一张桌子由一个桌面和 4 个桌腿组成，桌面和桌腿也可作为对象进一步的描述。

2.4 关系层

关系层用来描述图像中对象之间存在的关系，通过对大量图片的观测发现，对象之间的关系主要包括空间关系和人物关系（表 4）。MPEG-7[15]中关于空间关系的描述可分为对象所在区域的方向关系、对象本身的拓扑关系和对象间的语义空间关系 3 类：对象所在区域的方向关系用来描述对象在图像中的位置；对象本身的拓扑关系用来描述图像中对象与对象间的相对位置关系；对象间的语义空间关系即对象间的逻辑关系，如从属、依赖、关联、整体与部分等关系。

表 4 对象关系描述

Table 4 Object relation description

对象关系	关系子类	描述内容示例
空间关系	方向关系	上、下、左、右、左下、左上、右下、右上
	拓扑关系	分离、邻接、重叠、包含
	逻辑关系	从属、依赖、关联、整体与部分
人物关系	家庭关系	夫妻、长辈/晚辈、兄弟姐妹
	工作关系	上司/下属、同事
	朋友关系	好友、情侣
	师生关系	/
	合作关系	/
	共现关系	/

人物关系指的是在特定的社会范围内与他人之间存在和产生的关系，根据人物之间的熟悉程度和亲密关系一般将人物关系类型归纳为以下几类：家庭关系、工作关系、朋友关系、师生关系、合作关系、共现关系[17]。其中，共现关系是在人物关系分析过程中无法通过特征进行准确分类的情况。

2.5 场景层

场景层是根据图像中的对象及对象关系从全局的角度揭示图像特征，场景是有一定空间关系的目标结合体或集合[14]，是一组特定对象的特定关系的抽象，在一定程度上可以认为图像中除主要对象外其余部分都叫做场景[2]。因此，用图像的场景语义进行图像的描述和检索更接近于人的认知也更加符合检索习惯[18]。将敦煌壁画中的场景分为地理场景、时间场景、天气场景 3 类[2]，这 3 类场景能较好地覆盖社会化媒体中发布的图像，对该层具体描述如表 5 所示。

表 5 场景描述

Table 5 Scene description

场景	描述内容示例
地理场景	天空、大海、山峦、城市、乡村、室内、室外等
时间场景	春、夏、秋、冬、清晨、傍晚等
天气场景	晴、雨、阴天、多云、风、雷、电、雪等

2.6 事件层

事件是指在某个特定的时间片段和地域范围内发生的、由一个或多个角色参与、由一个或多个动作组成的一件事情，通常用一个短语或句子描述[19]。从事件性角度来看，社会化媒体图像具有极强的时效性和现实关联性。它们能够迅速捕捉并反映社会热点、时事新闻等最新动态，通过直观的图像语言，将事件现场、人物表情、环境氛围等细节呈现给用户，使用户能够迅速了解事件的进展和背景。因而，在描述一个事件的属性时必须包括对象（Who）、内容（What）、地点（Where）和时间（When）。对象在对象层已有描述，内容需要大量的背景知识和推理得到，地点和时间可由场景层中的地理场景和时间场景替代。

2.7 情感层

在社会化媒体平台上，用户通过上传、分享和评论图像，来表达自己的情感、态度和观点。这些图像往往蕴含着丰富的情感信息，能够引发用户的共鸣和讨论。同时，社会化媒体图像也常常被用来传递和表达特定的情感氛围，如喜悦、悲伤、愤怒等，通过视觉元素的运用，使得情感表达更加直观和深刻。图像情感具有两层含义，一层是图像给用户带来的主观感

DOI：10.13998/j.cnki.issn1002-1248.24-0109

受，是用户的知识背景、生活经历等与图像产生的共鸣；另一层是图像中对象本身的表情和神态，不以人的意志为转移。与传统文本信息相比，人们在检索图像时，往往具有一定的情感需求，在检索过程中会产生较强的情感反应。因此，基于情感的图像描述是基于语义的图像描述的必然发展，其意义在于推动图像描述研究从客观内容层次向主观体验层次迈进。

情感研究涉及多学科交叉，目前比较常用的模型有心理学 PAD 模型、情感计算 OCC 以及 Ekman 经典情感理论模型，Ekman 的情感理论模型描述符包括气愤、厌恶、恐惧、高兴、悲伤、惊讶，其他情感都可以由这 6 种情感构成，在不同文化传统间差异性较小[20]。目前采用较多为使用愉快的 / 忧伤的、激动的 / 平静的、紧张的 / 放松的 3 组词汇描述自然风景图像给人们所带来的情感体验[21]，因此，本文采用这些描述词进行图像情感上的描述，如表 6 所示。

表 6　情感描述

Table 6　Emotion description

情感	描述内容示例
图像中对象本身的情感	气愤、厌恶、恐惧、高兴、悲伤、惊讶
图像给人带来的情感体验	愉快的/忧伤的、激动的/平静的、紧张的/放松的

3　图像语义层次之间的关联关系

图像语义描述框架中的每一层都不是孤立存在的，而是相互关联，并且部分上层语义是可以由下层语义推理而得的。从实质来看，本文框架中的外部特征层是对图像固有特征的揭示，内容层是对图像颜色、形状、纹理进行描述，这两层都不能表达真正意义上的图像语义内容，因此图像的语义描述研究重点关注的是图像的对象层语义及其上层语义。图 2 是各层语义之间的关联关系图，对象语义是整个语义框架的核心部分，其他语义层都依赖于对象层，不同对象之间通过对象属性相互关联；关系层是根据对象之间在位置、空间方位等方面的信息推导而出的关系语义；场景层依赖于对象层与关系层，是通过从全局的角度分析对象及对象关系语义而得到的；事件依赖于对象并与场景关联，主要从一组对象间行为语义推导而得；图像中对象的情感是由对象语义分析而得，用户因图像而产生的情感则是从对象语义、场景语义与事件语义等综合推理得到的。

4　图像描述示例

图像的语义描述框架可用于网络图像的标注、组

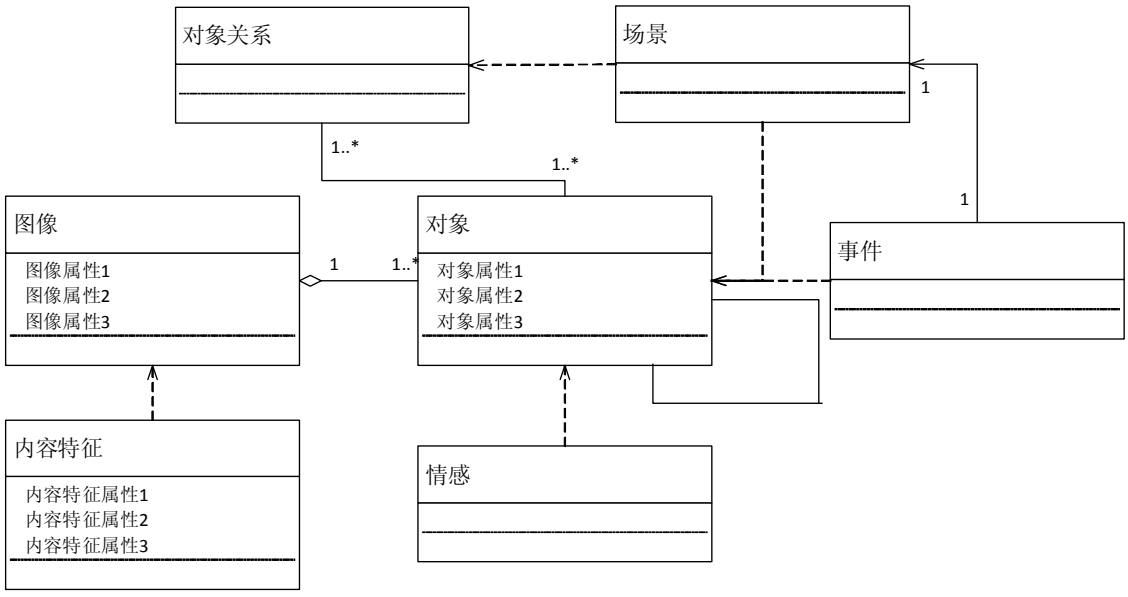

图 2　语义层次间的关联图

Fig.2　An association graph between semantic levels

织和检索，图像描述和标注的根本目的为了满足用户图像查询检索的需求，填补用户需求与图像描述之间的鸿沟，解决“查询非所得”的困境[9]。下面将使用本文提出的语义描述框架对人物图像与风景图像分别进行描述。

4.1 人物图像描述示例

图 3 中的对象有男性、女性、自行车、阳光、房屋、树等，主要对象是男性和女性，主要对象的属性是对象年龄、对象行为以及对象服饰组成；对象和图像之间的空间方向关系是“女性在图像左边、男性在图像右边”；拓扑关系是“女性骑着自行车、男性骑着自行车、女性在男生左性”；人物关系是“情侣”。图 3 中的对象都是出现在室外，且图 3 中出现了大面积的道路，表现的季节场景是“秋天”，由此确定图像场景层的描述。抽象出图像中的事件内容是“一对情侣在骑自行车”。根据图 3 中的主要色调、人物对象的面部表情和行为动作可推测图像情感是“高兴的”，图 3 给人带来的情感体验是“愉快的”（表 7）。

图 3 一对情侣骑自行车

Fig.3 A couple riding bicycles

4.2 风景图像描述示例

图 4 中的主要对象有大海、天空、太阳，对象的属性主要是对象的颜色属性和形状属性；对象和图 4 之间的空间方向关系有“天空在图像上方、红日在图像中间、大海在图像下方”；进而确定对象之间的拓扑关系：“红日在天空之内、天空在大海之上、红日在大海之上”。由于图 4 中的对象中没有人物出现，所以不存在人物关系，而图 4 中对象较少且比较简单，也不存在复杂的逻辑关系。图 4 中的对象都是出现在海边、室外，且图 4 表现的时间是清晨，由此确定图 4 场景层的描述。由对象、时间、地点抽象出图像中的

表 7 对图 3 的语义描述

Table 7 Semantic description of Fig.3

语义层次	描述			
对象层	对象	男性，女性，自行车，阳光，房屋，树		
	对象属性	男性：青年、伸腿		
		服饰组成：黑色上衣、棕色裤子、黑色靴子		
		女性：青年、伸腿		
		服饰组成：黑色帽子、白色上衣、黑色裤子、黑色靴子		
关系层	方向关系	女生在图像左边；男生在图像右边		
	拓扑关系	女生骑着自行车；男生骑着自行车；女生在男生左边		
	人物关系	情侣		
场景层	地理场景—室外，街道；时间场景—秋天；天气场景—晴			
事件层	对象	男性、女性		
	内容	一对情侣在骑自行车		
	地点	街道	时间	秋天
情感层	对象情感	高兴的		
	情感体验	愉快的		

DOI：10.13998/j.cnki.issn1002-1248.24-0109

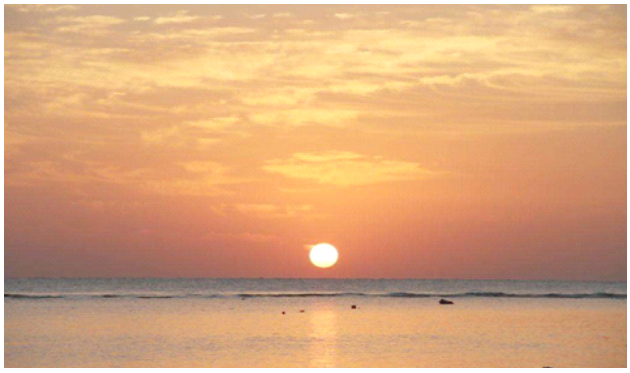

图 4　海上日出

Fig.4　Sunrise over the sea

事件内容是“海上日出”。对于图 4 情感的描述，主要是根据图 4 中的主要色调和对象之间的方位关系来确定，“海上日出”这幅图给人带来的情感是“平静的”“轻松的”（表 8）。

5　结　语

本文基于图像元数据标准、图像分层分类理论以及用户检索需求理论，针对目前社会化媒体上发布的大量标注不足甚至未标注图像，提出了一套较为完善的图像语义描述框架。融合事件和情感的图像语义描述框架实现了基于高层语义的图像描述与组织，为基于内容和语义的图像检索奠定了基础。在一定程度上，一方面，该框架能够推动 Sora 向更高层次的语义理解发展，从而对图像内容进行更为深入和全面的理解，生成更贴近人类认知和表达习惯的描述；另一方面，可以扩展 Sora 应用的广度与深度，使 Sora 可以开拓更广泛的应用场景，如情感分析、故事生成、个性化推荐等，从而在现有应用中提供更深层次的用户体验和价值。同时，图像语义描述框架和描述方法不依赖于具体的领域、应用系统以及软硬件环境，因此具有较强的可移植性和灵活性。文章同样存在一定的不足与改进：①基于本文提出的图像语义描述框架，开发基于图像标注原型系统；②用户在社会化媒体上发布的图像与情境关联密切，同时更倾向于情感的表达，未来可结合用户发布的文本信息对图像语义层进行更细粒度的研究。③未来的研究可进一步探索深度学习在图像与文本融合方面的应用，实现更精准的事件和情感识别。通过构建更复杂的神经网络结构，实现对图像中事件和情感信息的深度挖掘和融合。④在描述图像时，不仅要关注静态的视觉特征，还要考虑动态的事件进展。未来的框架可以尝试结合静态和动态信息，以提供更丰富、更生动的图像描述。

参考文献：

[1]　Data never sleeps 4.0[EB/OL]. [2017-01-08]. https://www.domo.

表 8　对图 4 的语义描述

Table 8　Semantic description of Fig.4

语义层次	描述			
对象层	对象	天空、太阳、大海		
	对象属性	颜色：天空—红色、太阳—金色、大海—红色		
		形状：红日—圆形		
关系层	方向关系	天空在图像上方、红日在图像中间、大海在图像下方		
	拓扑关系	红日在天空之内、天空在大海之上、红日在大海之上		
	人物关系	/		
场景层	地理场景—海边、室外；时间场景—清晨；天气场景—晴			
事件层	对象	天空、太阳、大海		
	内容	海上日出		
	地点	海边	时间	清晨
情感层	对象情感	/		
	情感体验	平静的、放松的		

com/blog/data-never-sleeps-4-0/.

[2] 王晓光, 徐雷, 李纲. 敦煌壁画数字图像语义描述方法研究[J]. 中国图书馆学报, 2014, 40(1): 50-59.
WANG X G, XU L, LI G. Semantic description framework research on Dunhuang fresco digital images[J]. Journal of library science in China, 2014, 40(1): 50-59.

[3] 黄崑, 王珊珊, 耿骞. 国外图像特征研究进展与启示[J]. 图书情报工作, 2015, 59(8): 138-146.
HUANG K, WANG S S, GENG Q. Research progress and enlightenment of image features abroad[J]. Library and information service, 2015, 59(8): 138-146.

[4] IPTC photo metadata standard[EB/OL]. [2017-01-08]. http://www.iptc.org/std/photometadata/specification/IPTC-PhotoMetadata.

[5] RANSOM N, RAFFERTY P. Facets of user-assigned tags and their effectiveness in image retrieval[J]. Journal of documentation, 2011, 67(6): 1038-1066.

[6] JÖRGENSEN C, STVILIA B, WU S H. Assessing the relationships among tag syntax, semantics, and perceived usefulness[J]. Journal of the association for information science and technology, 2014, 65(4): 836-849.

[7] KEISTER L H. User types and queries: Impact on image access systems[J]. Challenges in indexing electronic text and images, 1994: 7-22.

[8] TURNER J M. Comparing user-assigned terms with indexer-assigned terms for storage and retrieval of moving images[C]//Proceedings of the Annual Meeting of the American Society of Information Science, 1995.

[9] JORGENSEN C. Indexing images: Testing an image description template[C]//Proceedings of the ASIS Annual Meeting. 1996, 33: 209-13.

[10] SHATFORD S. Analyzing the subject of a picture: A theoretical approach[J]. Cataloging & classification quarterly, 1986, 6(3): 39-62.

[11] JAIMES A, CHANG S F. Conceptual framework for indexing visual information at multiple levels[C]. Proceedings of SPIE - The International Society for Optical Engineering, 2000, 3964: 2-15.

[12] FAUZI F, BELKHATIR M. Multifaceted conceptual image indexing on the world wide web[J]. Information processing & management, 2013, 49(2): 420-440.

[13] 王惠锋, 孙正兴, 王箭. 语义图像检索研究进展[J]. 计算机研究与发展, 2002, 39(5): 513-523.
WANG H F, SUN Z X, WANG J. Semantic image retrieval: Review and research[J]. Journal of computer research and development, 2002, 39(5): 513-523.

[14] 吴楠, 宋方敏. 一种基于图像高层语义信息的图像检索方法[J]. 中国图象图形学报, 2006, 11(12): 1774-1780.
WU N, SONG F M. An image retrieval method based on high-level image semantic information[J]. Journal of image and graphics, 2006, 11(12): 1774-1780.

[15] MPEG-7 overview[EB/OL]. [2017-04-07]. http://mpeg.chiariglione.org/standards/mpeg-7.

[16] JÖRGENSEN C. Attributes of images in describing tasks [J]. Information processing & management, 1998, 34(2/3): 161-174.

[17] 廉营. 基于语义角色标注的微博人物关系抽取[D]. 哈尔滨: 哈尔滨工业大学, 2013.
LIAN Y. Character relationship extraction in microblog based on sementic role labeling[D]. Harbin: Harbin Institute of Technology, 2013.

[18] (英)格列高里, 著. 彭聃龄, 杨棨, 译. 视觉心理学[M]. 北京: 北京师范大学出版社, 1986.
GREGORY L R. Visual psychology[M]. Beijing: Beijing Normal University Press, 1986.

[19] 高强, 游宏梁. 事件抽取技术研究综述[J]. 情报理论与实践, 2013, 36(4): 114-117, 128.
GAO Q, YOU H L. Summary of research on event extraction technology[J]. Information studies: Theory & application, 2013, 36(4): 114-117, 128.

[20] 刘小瑞. 基于 Mpeg-7 的图像多层次语义知识库的构建[D]. 太原: 太原理工大学, 2012.
LIU X R. Construction of image multi-level semantic knowledge base based on Mpeg-7[D]. Taiyuan: Taiyuan University of Technology, 2012.

[21] 黄崑, 赖茂生. 图像情感特征的分类与提取[J]. 计算机应用, 2008, 28(3): 659-661, 668.
HUANG K, LAI M S. Classification and extraction of image affective features[J]. Journal of computer applications, 2008, 28(3): 659-661, 668.

农业图书情报学报
DOI：10.13998/j.cnki.issn1002-1248.24-0109

Framework for the Semantic Description of Images with Integrated Events and Emotions

HU Shoumin[1], DONG Huanqing[2]
(1. Central China Normal University Library, Wuhan 430079;
2. School of Information Management, Central China Normal University, Wuhan 430079)

Abstract: [Purpose/Significance] Aiming at the semantic missing and incomplete problems in the process of image organization and retrieval, a framework for semantic description of images in social media is proposed to enrich the existing theoretical system of image description, improve the efficiency and utilization of image retrieval, and provide a reference for the realization of the automatic semantic annotation of images. [Method/Process] First, we conducted a survey and analysis of research progress related to image description both domestically and internationally, summarizing the existing theories of image description and annotation, metadata specifications, and related technical methods. Second, based on the image metadata standards and the theory of hierarchical and categorical description of image features, we constructed a semantic description framework for social media images, focusing on seven layers: external feature layer, content layer, object layer, relationship layer, scene layer, event layer, and emotional layer. We also elaborated in detail the various semantic layers and their interrelationships. Finally, we verified the feasibility of the image semantic description framework by describing the examples of character images and landscape images. [Results/Conclusions] The results of the descriptive examples of character images and landscape images indicate that the image semantic description framework can eliminate the "semantic gap" in image description through semantic associations between different layers, and achieve a multi-faceted, multi-dimensional, and multi-level structured and semantic description of the external and content features of images. It has strong portability and flexibility. However, there are also certain limitations and areas for improvement in this paper: 1) Based on the image semantic description framework proposed in this paper, a prototype system based on image annotation needs to be developed; 2) The images posted by users on social media are closely related to the situation, and they are more likely to express emotions. In the future, more research on the semantic layer of images can be conducted based on the text information posted by users; 3) Future research can further explore the application of deep learning in image and text fusion to achieve more accurate event and emotion recognition. By constructing a more complex neural network structure, the event and emotion information in the image can be deeply mined and fused; 4) When describing images, the study should pay attention not only to static visual features, but also to consider the dynamic course of events. Future frameworks could attempt to combine static and dynamic information to provide richer, more vivid descriptions of images.

Keywords: semantic description framework; image feature; semantic annotation; Sora