

# 静态图像行为标注众包系统的设计与实现

侯禹臣 吴 伟

(内蒙古大学计算机学院 呼和浩特 010021)

**摘 要** 针对静态图像行为识别研究缺乏标注数据的问题,在 Android 平台下,利用“众包”思想,设计并开发了基于静态图像的视觉行为人工标注系统。该系统主要包括分配标注任务、用户标注图像信息、评审标注信息和查看历史标注信息等功能。对于评审分数较高的标注信息,利用网络爬虫技术提取该图像的辅助文本标签,并且将标注信息转化为词向量后进行存储,以便于后期的实验研究。同时,系统应用一种基于定价机制的任务分配算法,有效地提高了用户图像标注效率。实际部署系统应用情况表明,该系统操作简洁、流畅,各算法功能模块稳定、高效,并且充分地利用移动端便捷的优势,能够顺利进行图像行为标注数据的收集和整理工作。

**关键词** 图像行为标注,众包,网络爬虫,词向量,Android

**中图法分类号** TP311.52 **文献标识码** A

## Design and Implementation of Crowdsourcing System for Still Image Activity Annotation

HOU Yu-chen WU Wei

(Department of Computer Science, Inner Mongolia University, Huhhot 010021, China)

**Abstract** According to the problem of lack of the annotation data in still image activity recognition research, under the Android platform, a visual activity manual annotation system based on still images was designed and developed with the idea of “crowd-sourcing”. The system mainly includes the functions of assigning annotation tasks, annotating image information, reviewing annotation information and examining historical annotation information. For the annotation information with high evaluation scores, the web crawler technology is used to extract the auxiliary text labels of the image, and the annotation information is converted into word embedding storage for the convenience of later experimental research. Meanwhile, the system applies a task assignment algorithm based on pricing mechanism, which effectively improves the efficiency of user image annotation. The application of the actual deployment system shows that the operation is simple and smooth, the function modules of each algorithm are stable and efficient, and the advantages of mobile terminals are fully utilized to collect and organize the data of image activity annotation.

**Keywords** Image activity annotation, Crowdsourcing, Web crawler, Word embedding, Android

## 1 引言

图像的识别与理解一直以来都是计算机视觉领域的研究热点<sup>[1]</sup>。图像的识别与理解研究目前主要局限于识别图像中不同区域的目标对象,目的是解决图像目标对象的识别与定位问题,但是却无法解决对图像所表达出各种高层语义信息的理解问题,如抽象概念,人类的行为活动、情感、事件等各种语义。就人类行为识别而言,现有的行为识别研究大部分集中在视频或图像序列、智能手机等移动传感器采集的时空序列数据的分析上<sup>[2-3]</sup>,很少有人专门从事静态图像的行为识别研究。然而在海量的图像数据中存在大量包含人类视觉行为的图像,这些行为活动既包括简单的人体动作姿势,也包括复杂的、抽象的语义行为,对这类图像中人类视觉行为的识别与理解研究有着重要的理论意义和应用价值,它可以提高计算机处理各种基于图像应用的智能化水平,在娱乐、虚拟现实、运动捕捉和运动训练<sup>[4]</sup>等领域都有应用。

由于目前国内外很少有人专门从事静态图像的行为识别研究,国际上一些主流的图像数据集(如:ImageNet<sup>[5]</sup>、MSCOCO<sup>[6]</sup>、ImageCLEF、Pascal VOC<sup>[7]</sup>等)也没有人类视觉行为的训练标注数据,这些图像集一般仅对图像的整体或区域对象标注名词或抽象名词概念,而没有标注行为语义概念。另外,国内外现有的图像标注工具,如国内的“精灵标注助手”,国际上的图像标注工具 ByLabel<sup>[8]</sup>、VGG Image Annotator<sup>[9]</sup>、Vatic 等,都只涉及标注图像中的目标对象,没有专门针对行为语义等图像高层语义的标注功能,并且这些标注工具大多数只能工作在 PC 机上,标注工作在时间和空间上有一定的限制。

针对以上问题,本文设计并开发了一款基于 Android 平台<sup>[10]</sup>的图像视觉行为众包标注系统。该系统利用基于定价机制的分配算法为广大用户分配标注任务,使分配工作更高效、更便捷;其次,评选并收集合格的标注信息,转化为词向量,存储到数据库中。另外,系统借助爬虫技术获取图像的辅

本文受国家自然科学基金项目(61763035)资助。

侯禹臣(1993—),男,硕士生,主要研究方向为智能信息处理;吴 伟(1980—),男,博士,副教授,主要研究方向为智能信息处理,E-mail:cswuwei@imu.edu.cn(通信作者)。

助文本关键词,以进一步丰富图像的语义。静态图像行为标注系统突破了时间和空间的限制,利用该系统能够有效建立视觉行为研究所需的图像行为识别标注数据库,解决采集标注行为的图像数据问题。

2 系统总体设计

2.1 系统的体系结构

静态图像行为标注系统由 Android 客户端和服务端组成,它们通过服务器接口进行通信。客户端的功能是与用户交互,完成图像标注和查看历史标注信息等操作。服务端的功能是处理客户端的请求,完成标注任务分配、评审标注信息等操作。客户端和数据库的数据传输也要通过服务器来完成。系统架构如图 1 所示。

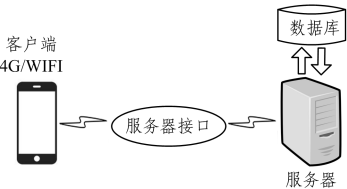


图 1 系统架构图

2.2 系统功能模块介绍

为了满足图像行为标注需求,该系统具体的功能模块划分如图 2 所示。

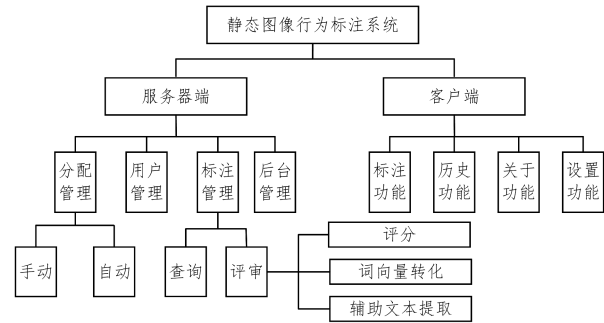


图 2 系统功能模块图

(1)服务器端的功能模块

分配管理:给用户分配标注任务(手动/自动);查询用户标注任务完成情况。

标注管理:查询标注信息;评审者对用户的标注结果进行评审,给出评分和评审意见;对于评分较高的标注信息,进行图像辅助文本标签的提取并且转化为词向量后进行存储。

用户管理:对标注者的管理,主要涉及的操作是查询和统计。

后台管理:对评审者的增加、删除、修改、查询(模糊查询)等操作。

(2)客户端的功能模块

历史功能:查看标注历史、评审者打分情况和评审意见。

关于功能:对标注要求、标注步骤、注意事项、评分规则等进行说明。

设置功能:包括系统设置和修改个人信息。

标注功能:包括目标对象的手工标注(用户在给定的目标对象范围内,针对图像中出现的目标对象用矩形框标注),行为语义标签的手工标注(用户在给定的视觉行为范围内,针对图像的行为语义进行标注),行为语义的自然语句标签手工标

注(用户对给定的图像标注自然语句)。图 3 给出了标注示例图,演示了用户要标注的各种信息。

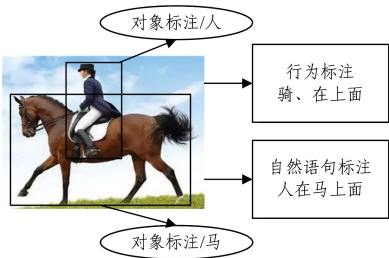


图 3 标注示例图

本文参考了视频数据的行为识别类别以及自然语言中的动词语义网络库,定义了 5 大类、100 种左右的视觉行为(如表 1 所列),基本涵盖了真实场景图像中有代表性的大多数日常行为类别。

表 1 图像中行为语义识别范围

| 编号 | 主要分类 | 具体行为语义概念                                                                                                                   |
|----|------|----------------------------------------------------------------------------------------------------------------------------|
| 1  | 单人动作 | 走 站 坐 跑 跳 蹲 躺 弯腰 跪 抬 洗澡 洗脸 做操 锻炼 上楼 下楼                                                                                     |
| 2  | 交互行为 | 扔 背 抱 踢 捡 推 拉 打架 依靠 拥抱 触摸 吃 戴 扶 喝 摔跤 砍 看 听 写 亲 交谈 会见 割(削) 叫 穿 玩耍 带 骑 剪 使用 陪 洗衣服 刷牙 梳头 滑冰 冲浪 滑雪 打 踢 接(电话) 放(风筝) 玩(电脑) 开(汽车) |
| 3  | 距离   | 在前面 在后面 在左边 在右边 在上面 在下面                                                                                                    |
| 4  | 位置关系 | 在远处 在近处 在身边 在附近 很近 很远 完全接触                                                                                                 |
| 5  | 表情   | 微笑(大笑) 哭                                                                                                                   |

本文同时参考了 MSCOCO、ImageNet、Flickr<sup>[11]</sup>等图像库的物体对象类别,限定了图像中的行为对象交互类别,指定 91 种典型的实体对象作为交互对象,它们都是人类经常交互的、有代表性的对象,如表 2 所列。

表 2 图像中行为交互对象的范围

| 编号 | 主要分类          | 具体对象名称                                             |
|----|---------------|----------------------------------------------------|
| 1  | 随身物品/<br>8 种  | 帽子 背包 手袋 行李箱 眼镜 雨伞 鞋带 钱包                           |
| 2  | 动物/10 种       | 马 牛 绵羊 鸟 狗 长颈鹿 大象 猫 斑马 熊                           |
| 3  | 交通工具/<br>8 种  | 汽车 自行车 火车 摩托车 轮船 飞机 公共汽车 卡车                        |
| 4  | 体育/10 种       | 棒球棒 网球拍 保龄球 足球 风筝 滑冰鞋 滑雪板 冲浪板 飞盘 橄榄球               |
| 5  | 室外场景/<br>6 种  | 停车标志 路标 大树 信号灯 海滩 消防栓                              |
| 6  | 食物/10 种       | 香蕉 苹果 胡萝卜 汉堡包 蛋糕 冰激凌 面包圈 披萨饼 橙子 三明治                |
| 7  | 厨房工具/<br>8 种  | 瓶子 水杯 勺子 刀 叉 酒杯 碗筷 盘子                              |
| 8  | 室内家居/<br>16 种 | 椅子 餐桌 沙发 窗户 盆栽 书桌 床 卫生间 镜子 门 微波炉 洗衣机 冰箱 洗手池 榨汁机 烤箱 |
| 9  | 室内用具/<br>8 种  | 书 表 牙刷 梳子 剪刀 玩具熊 吹风机 花瓶                            |
| 10 | 电子物品/<br>6 种  | 电视 电脑 手机 遥控器 键盘 鼠标                                 |
| 11 | 人/1 种         | 人                                                  |

2.3 小结

本节主要介绍了系统的架构设计和功能模块设计,针对系统的核心功能模块——标注功能,进行了详细描述,介绍了用户标注的 3 类信息,并且分别定义了图像标注的目标对象范围和视觉行为范围。第 3 节将重点讨论自动分配算法、网络爬虫、词向量等系统实现涉及到的关键技术。



4 系统测试

表 3 列出了本文设计的静态图像行为标注系统和现有标注工具的对比结果。

表 3 行为标注系统和现有标注系统对比

| 功能       | 行为标注系统 | 现有标注系统 |
|----------|--------|--------|
| 打开本地图像   | 有      | 有      |
| 图像辅助文本提取 | 有      | 无      |
| 词向量转化    | 有      | 无      |
| 已标数据可视化  | 有      | 有      |
| 视觉行为标注   | 有      | 无      |
| 移动端标注    | 有      | 无      |

本文首先对目前收集的包含 40 多种视觉行为类别的 12000 张图像进行标注测试。之后,继续采集和整理图像数据。图 8 分别给出了客户端的标注任务界面和图像标注界面。

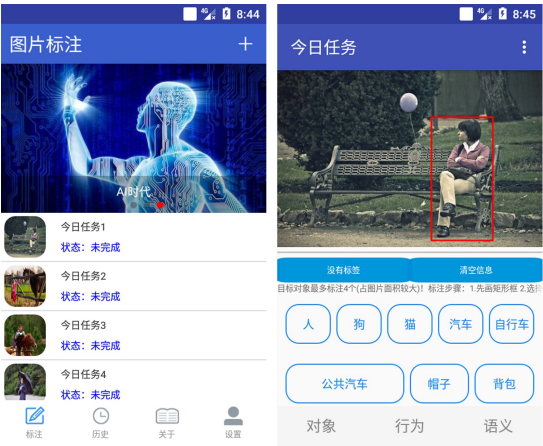


图 8 图像标注功能

**结束语** 静态图像行为标注众包系统的实现在很大程度上解决了目前缺乏行为标注数据的问题,为静态图像视觉行为的识别与理解研究奠定了实验基础。值得一提的是,系统的开发采用了“众包”思想,在今后的研究工作中,将会引入更多的众包策略,如众包奖金策略和在线激励机制算法,旨在进一步提高图像的标注效率和标注质量。

参 考 文 献

[1] HUISKES M J, THOMEE B, LEW M S. New trends and ideas in visual concept detection[C]//Proc of the 11th ACM Conference on Multimedia Information Retrieval, Philadelphia, PA, USA:ACM,2010:527-536.

[2] 单言虎,张彰,黄凯奇. 人的视觉行为识别研究回顾、现状及展望[J]. 计算机研究与发展,2016,53(1):93-112.

[3] GUO G, LAI A. A survey on still image based human action recognition[J]. Pattern Recognition,2014,47(10):3343-3361.

[4] SONG X,ZHOU H,LIU G. Human behavior recognition based on multi-feature fusion of image[J/OL]. <http://doi.org/10.1007/s10586-018-2073-7>.

[5] MUHAMMED M A E,AHMED A A,KHALID T A. Benchmark analysis of popular ImageNet classification deep CNN architectures[C]//Smart Technologies For Smart Nation (Smart-TecCon). IEEE,2017:902-907.

[6] HUANG G, HU H. c-RNN: A Fine-Grained Language Model for Image Captioning[J]. Neural Processing Letters, 2019, 49(2):683-691.

[7] LIU Z,LI X,LUO P,et al. Deep learning markov random field for semanticsegmentation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence,2018,40(8):1814-1828.

[8] QIN X, HE S, ZHANG Z, et al. ByLabel: A Boundary Based Semi-Automatic Image Annotation Tool[C]//2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE,2018:1804-1813.

[9] BERNAL J,HISTACE A,MASANA M,et al. GTCreator:a flexible annotation tool for image-based datasets[J]. International Journal of Computer Assisted Radiology and Surgery, 2018:1-11.

[10] MA Y,LIU X,LIU Y,et al. A Tale of Two Fashions: An Empirical Study on the Performance of Native Apps and Web Apps on Android[J]. IEEE Transactions on Mobile Computing,2018, 17(5):990-1003.

[11] ZHANG P, WEI Z, LI Y, et al. Automatic Image Annotation Based on Multi-Auxiliary Information[J]. IEEE Access, 2017, 5(99):18402-18411.

[12] 江志刚. 基于 SSM 框架的网上题目录入答题系统设计[J]. 无线互联科技,2017(20):62-63.

[13] 任玉刚. Android 开发艺术探索[M]. 北京:电子工业出版社, 2015:414-441.

[14] SINGER Y,MITTAL M. Pricing Mechanisms for Crowdsourcing Markets[C] // International Conference on World Wide Web. 2014:1157-1166.

[15] SINGER Y. How to win friends and influence people, truthfully:influence maximization mechanisms for social networks[C]// WSDM. 2012:733-742.

[16] DOBZINSKI S,PAPADIMITRIOU C,SINGER Y. Mechanisms for complement-free procurement[C]//EC. 2011:273-282.

[17] 余豪士,匡芳君. 基于 Python 的反爬虫技术分析与应用[J]. 智能计算机与应用,2018,8(4):112-115.

[18] 周练. Word2vec 的工作原理及应用探究[J]. 科技情报开发与经济,2015,25(2):145-148.

[19] FAUZI M A. Word2Vec Model for Sentiment Analysis of Product Reviews In Indonesian Language[J]. International Journal of Electrical & Computer Engineering,2018,7(1):244-252.