

专题综述

食物识别研究进展

复旦大学 陈静静 香港城市大学 朱斌 杨宗桦

一、引言

随着我国经济的快速发展，人民的生活水平不断提高，饮食健康问题越来越受关注。据数据显示，全球每年有 1100 万人死于不健康饮食。因此，为了提高饮食健康状况，记录饮食，分析饮食营养状况是一项十分重要的研究任务。鉴于饮食健康管理的重要性，食物识别与营养分析的相关研究受到越来越多的关注。

作为健康管理的重要组成部分，食物识别的目的是帮助用户自动构建饮食日记，避免手动记录食物名字带来的麻烦。相比于通用图像识别，细粒度食物图像识别有其特有挑战。首先，由于切菜，烹饪方法的差异导致同类食物图片视觉上的差异很大。其次，同类食物的组成原料会随着地域，季节的改变而变化，使得很多食物没有固定的组成原料。图 1 展示了三种食物由于所用食材的差异造成的食物图片的类内视觉差异大的例子。类内视觉差异大使得细粒度食物图片识别成为一个比较难的问题。

早期的食物识别研究主要基于手工特征，如 SIFT，HoG 等，在这些特征的基础上再训练 SVM 分类器进行分类。由于食物图片类内差异大，手工特征难以取得满意的分类效果。近几年内，由于深度学习模型在图像识别上取得了巨大成功，近期的食物识别研究主要基于深度卷积网络。除了常见的用在通用物体分类上的深度卷积模型（如 VGG, ResNet）外，针对食物图像分类的深度卷积模型，例如，Wide slice Residual Networks^[1] 也被提出。目前，在包含上百个类的中等大小的数据集上，深度卷积神经网络能够取得高达 80% 的识别准确率。



图 1 所用食材差异造成的食物图片的类内视觉差异。第一排，第二排以及第三排图片分别属于“黄瓜炒鸡蛋”，“香辣藕丁”以及“手撕杏鲍菇”。

尽管深度卷积网络的出现使得食物图像识别的性能有了大幅度的提升，然而现有的基于卷积神经网络的食物识别模型离实际应用仍有一段距离。首先，现有食物识别模型都是在中小数据上训练得到，仅能识别百种食物图片，难以处理实际场景中上千万种类别的食物识别。其次，现有的食物识别模型依赖于大量手工标注的食物图片进行训练。由于手工标注成本高，这些识别模型的难以通过直接扩充训练数据集的方式提高模型的泛化能力。

为了解决实际应用中的大规模食物识别问题，现有的工作主要采用两种途径。一是对食物图片进行原料识别。食物由多种原料组成，与食物类别数量相比，原料种类的数量要少很多。因此，相比于食物类别识别，识

别食物原材料对于大规模的食物识别更加有效。另外一种途径是通过跨媒体菜谱检索进行食物识别。给定一个食物图片，通过跨媒体菜谱检索找到对应菜谱。由于菜谱包含食物名字，食材，甚至营养成分，因此，通过菜谱检索能够有效的进行食物识别以及营养分析。同时，由于网络上有大量的包含食物图片的菜谱，直接利用这些菜谱数据训练跨媒体检索模型能够节省人工标注，进而有效应对大规模食物图像识别问题。

二、研究现状

本章将从三个方面对现有食物识别工作进行介绍，数据集，食物识别以及跨媒体菜谱检索。

2.1. 数据集

食物数据集是食物识别的基础和先决条件。研究者们为此建立了大量的基准数据集，包含中国菜[2,3]、日本菜[4,5]和西式菜[6,7]等。中国自古以来就是对美食有极度追求的国家，也是食物种类、跨度最大的国家之一。[1]从众多的中文菜谱网站中爬取了大量的菜谱和食物图片数据，建立了首个大规模中国菜数据集：Vireo Food-172。该数据集总共包含 172 种不同类型的食物类型和 353 中不同的食材，同时图片数量达到 110,241 张。西式菜中属 Recipe1M+^[5]最大，包含百万菜谱和超过一千万的图片，但是需要注意的是，该数据集的标签并不是人工标注得到，而是通过对菜谱标题进行聚类自动生成。因此，该数据集提供的食物种类和食材的标注包含很多噪声。

2.2. 食物识别

现有工作对食物识别的研究主要包含两个部分：食物分类和食材识别。给定一张食物图片，食物分类模型需要能够判断是那种食物类型，是一个多类问题；而食材识别则更复杂些，需要识别图片中的食物是由哪些食材制作出来的，是一个多标签分类问题。[2]提出了不同的面向食物分类与食材识别的多任务学习模型。[2]中的结果显示，同时学习食物分类和食材识别能够比单任务学习食材识别好很多。值得一提的是，中国菜食物分类的 Top-5 准确率能够达到近 96%，食材识别也取得了不俗的表现。

众所周知，食材的切法和烹饪方法对食物最终食物的视觉效果产生巨大的影响，只考虑食材来做识别显然有其缺陷性。因此，[8]对食材的切法和烹饪方法进行了进一步的标注，通过引入食材切法以及烹饪方法为辅助任务进行多任务学习，使得食材识别的效果有了进一步的提升。此外，除了多任务学习，[8]为了处理由于相机到食物距离以及不同的切菜方法造成的食材大小差异问题，引入多尺度局部原料识别，实验证明该方法能够有效提高原料识别准确率。

2.3. 跨媒体菜谱检索

除了原料识别，应对大规模食物识别的另外一种有效方法为跨媒体菜谱检索。该方法不需要人工进行数据标注，因此近几年来吸引了不少研究者。相比与其他的跨媒体检索问题，菜谱检索有其独特的挑战性。首先菜谱是结构性文本，通常包含标题、食材和准备步骤三块信息。如何对菜谱的进行有效建模，从而学习到菜谱表征是一个难题。其次，菜谱步骤是教学性语言，食物图片是该教学语言产生的结果。然而，并不是每一个步骤都会对食物造成视觉上可见效果。因此，如何对菜谱步骤的视觉相除了原料识别，应对大规模食物识别的另外一种有效方法为跨媒体菜谱检索。该方法不需要人工进行数据标注，因此近几年来吸引了不少研究者。相比与其他的跨媒体检索问题，菜谱检索有其独特的挑战性。首先菜谱是结构性文本，通常包含标题、食材和准备步骤三块信息。如何对菜谱的进行有效建模，从而学习到菜谱表征是一个难题。其次，菜谱步骤是教学性语言，食物图片是该教学语言产生的结果。然而，并不是每一个步骤都会对食物造成视觉上可见效果。因此，如何对菜谱步骤的视觉相关性进行建模以建立起菜谱到实物图片的映射关系也是一个难题。

针对跨媒体菜谱检索，现有的方法大都是针对菜谱文本和食物图像学习一个联合嵌入空间，在该联合嵌入空间中计算图片与菜谱的相似度，以此进行菜谱检索。在学习联合嵌入空间的过程中，[9][10]引入注意力机制 (Attention Model) 使得学习到的联合嵌入空间更好，

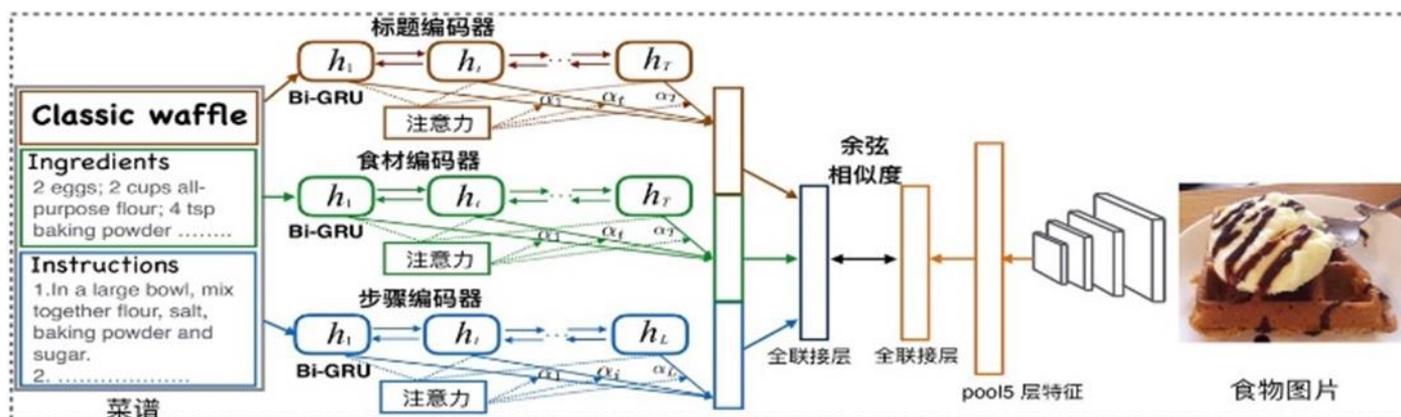


图 2 基于层次注意力机制的跨媒体菜谱检索模型

从而有效提高菜谱检索的准确率。为了对菜谱到食物图片的视觉相关性进行建模, [10]引入了层次注意力机制。如图 2 所示, 在菜谱表征学习中, 层次注意力机制会对每一个单词以及每一个句子学习到相应的权重。权重的大小表示该单词或句子的重要性。引入层次注意力能够帮助过滤掉一些视觉无关的句子或者单词, 使得学习到的联合嵌入空间的过程能够更好的进行菜谱检索。R2GAN^[11]探索了一种基于生成对抗网络(GAN)的检索模型, 通过专门设计的一个生成器两个识别器结构, 让检索模型能够学到更有效的特征表达。并且, 由菜谱特征重建出的食物图像能够协助用户更快速的找到相应的菜谱, 对检索中的排序结果也起到了解释的作用。

三、未来发展方向

尽管通过原料识别, 跨媒体菜谱检索能够一定程度上解决大规模食物识别问题。然而, 这些方法仍有一定的局限性。首先, 目前的识别模型难以应对跨地区识别以及检索。例如, 对于原料识别, 在中国菜上训练的模型难以应用到日本菜或者西方菜的原料识别上。同时, 在西式菜上训练的跨媒体检索模型也难以搬迁到中国菜的菜谱检索上。因此, 我们认为未来的发展方向包括利用迁移学习实现跨区域原料识别以及菜谱检索。此外, 小样本学习也是解决大规模食物识别的一个有效途径, 因此, 小样本食物识别也是一个值得研究的方向。

责任编辑 任桐炜

参考文献

- [1] Martinel, Niki, Gian Luca Foresti, and Christian Micheloni. "Wide-slice residual networks for food recognition." 2018 IEEE Winter Conference on Applications of Computer Vision (WACV). IEEE, 2018.
- [2] Chen J, Ngo C W. Deep-based ingredient recognition for cooking recipe retrieval[C]//Proceedings of the 24th ACM international conference on Multimedia. ACM, 2016: 32-41.
- [3] Chen X, Zhu Y, Zhou H, et al. ChineseFoodNet: A large-scale image dataset for Chinese food recognition[J]. arXiv preprint arXiv:1705.02743, 2017.
- [4] Harashima J, Someya Y, Kikuta Y. Cookpad image dataset: An image collection as infrastructure for food research[C]//Proceedings of the 40th International ACM SIGIR Conference on Research and Development in Information Retrieval. ACM, 2017: 1229-1232.
- [5] Horita D, Shimoda W, Yanai K. Unseen food creation by mixing existing food images with conditional styleGAN[C]//Proceedings of the 5th International Workshop on Multimedia Assisted Dietary Management. ACM, 2019: 19-24.
- [6] Marin J, Biswas A, Ofli F, et al. Recipe1m+: A dataset for learning cross-modal embeddings for cooking recipes and food images[J]. IEEE

- [7] Bossard L, Guillaumin M, Van Gool L. Food-101—mining discriminative components with random forests[C]//European Conference on Computer Vision. Springer, Cham, 2014: 446-461.
- [8] Chen J, Ngo C W, Chua T S. Cross-modal recipe retrieval with rich food attributes[C]//Proceedings of the 25th ACM international conference on Multimedia. ACM, 2017: 1771-1779.
- [9] Chen J J, Ngo C W, Feng F L, et al. Deep understanding of cooking procedure for cross-modal recipe retrieval[C]//2018 ACM Multimedia Conference on Multimedia Conference. ACM, 2018: 1020-1028.
- [10] Chen, Jing-Jing, et al. "Deep understanding of cooking procedure for cross-modal recipe retrieval." 2018 ACM Multimedia Conference on Multimedia Conference. ACM, 2018.
- [11] Zhu B, Ngo C W, Chen J, et al. R2GAN: Cross-Modal Recipe Retrieval With Generative Adversarial Network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 11477-11486.



陈静静

复旦大学青年副研究员。主要研究方向为多媒体内容分析和计算机视觉。

Email: chenjingjing@fudan.edu.cn



朱斌

香港城市大学博士研究生。主要研究方向为食物识别和跨媒体检索。

Email: andrewzhu1216@gmail.com



杨宗桦

香港城市大学教授。主要研究方向为多媒体，计算机视觉，模式识别。

Email: cwngocs.cityu.edu.hk

热点追踪

多模态医学图像合成与重建

上海交通大学 王乾

医学图像在临床诊断和治疗中发挥着基础性作用。而如何将多模态图像提供的互补信息有效融合，是医学图像领域的经典科学问题之一。针对不同来源的图像数据，通过信号处理、模式识别等手段捕捉与生理变化或疾病进展紧密相关的影像特征，建立面向特定临床应用的数据模型，是多模态融合方法在与医学交叉研究中常采用的技术路线。

近期，基于深度学习的图像合成在医学图像领域大显身手。例如，可以由磁共振(MR)图像合成获得 CT 图像，或者在对应不同成像序列或场强的 MR 图像之间进行转化。图像合成的发展，深刻改变了医学图像多模态融合。长期以来，多模态融合主要依赖于图像配准，即建立多模态图像之间的相似性测度并进行优化，从而估计图像之间的空间变换。但是，不同模态图像视觉特征迥异，跨模态的相似性测度设计是众所周知的难点。我们通过图像合成实现不同模态图像之间的相互转化，进而将多模态图像之间的配准问题分解为(多个)与合成图像相关联的单模态配准问题，重塑了多模态融合的概念。

同时，深度学习推动多模态融合从传统图像后处理

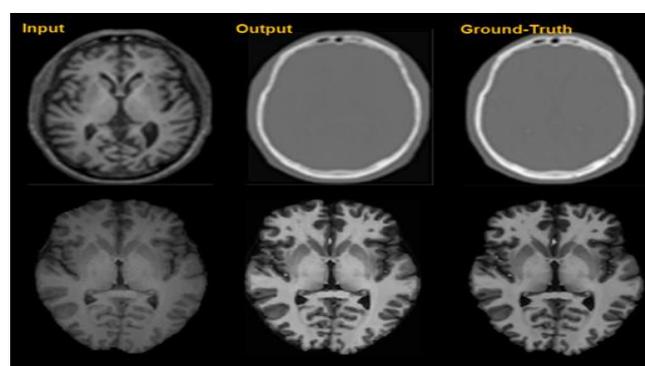
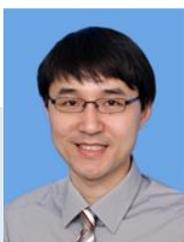


图 1 通过深度学习实现不同输入/输出之间快速合成

领域延伸至前端成像，奠定了图像智能重建基础。以 MR 图像中最常用 T1、T2 序列为例，我们通过深度学习融合不同序列图像互补视觉特征，将较快 T1 成像数据注入 T2 重建，使得在对 T2 数据 K 空间快速降采样的条件下，仍然可以重建出高质量 T2 图像。这些成果可以用于以新一代国产 PET-MR 设备为代表的高端医学影像设备中，从而推动我国医学影像装备产业链的发展。

以上工作已发表于国际会议 MICCAI 2017-2019。

责任编辑 王金甲



王乾

上海交通大学生物医学工程学院长聘副教授，医学影像先进技术研究院院长助理，主要研究方向为医学图像计算。

Email: wang.qian@sjtu.edu.cn

热点追踪

基于 AI 的新冠肺炎肺部 CT 影像量化分析



计算机视觉专委会常委
员、华中科技大学白翔教授团

图 1 华中科技大学科研助力抗疫一线

升诊断效率,减轻医生诊断
工作负荷。目前该系统已在

队联合华中科技大学附属协和医院放射科、华为云、蓝网科技等共同研发出一套基于人工智能技术的面向新冠肺炎肺部 CT 影像量化分析系统,实现 CT 影像结果秒级输出。

该系统可自动分离出毫米级微小病灶,自动计算出肺部病变面积的百分比,实现单病例全自动精准量化结果的秒级输出。可以辅助医生更高效地区分新冠肺炎的早期、进展期与重症期,有利于早期筛查与防控。同时,对于确诊病人,基于对多次复查影像数据的量化分析,医生能够有效评估病情进展及用药疗效等情况,大幅提

全国几十家医院使用。

2月29日下午,中共中央政治局委员、国务院副总理孙春兰率中央督导组到华中科技大学考察疫情防控科研攻关情况,肯定了研究团队所开发系统的精确性。同时也非常关注该系统的落地应用情况,特别是在湖北省、武汉市的应用,鼓励团队结合 CT 影像与核酸试剂等信息进行更为精准的新冠肺炎诊断。3月3日央视新闻以“科研攻关助力抗疫一线”、3月25日《焦点访谈》以“救治在一线 科研当先锋”为题分别进行了报道。

责任编辑 储璐

华中科技大学新闻:

<http://news.hust.edu.cn/info/1002/38080.htm?from=groupmessage&isappinstalled=0>

“华为云和TA的朋友们”微信公众号: <https://mp.weixin.qq.com/s/GxZ1F2qug3vUk7hbChUkQw>

中央电视台官网: <http://tv.cctv.com/2020/03/25/VIDE5rVd8uYIksbfguXqbrov200325.shtml>

顶会观察

ACM Multimedia 2019

北京电子科技学院 金鑫

ACM International Conference on Multimedia (简称 ACM MM) 是中国计算机学会推荐的计算机图形学与多媒体领域的 A 类学术会议, 一年举办一次。2019 年的 ACM MM 会议已于 2019 年 10 月 21 日至 25 日在法国尼斯举行。

一、多媒体会议主题的变化

多媒体会议主要由 Systems (系统)、Experience (体验)、Engagement (参与)、Understanding (理解) 四大主题组成。2019 年的多媒体会议给出了这四大主题更详细的诠释和更新: Multimedia Systems (多媒体系统)、Multimedia Experience (多媒体体验)、Engaging Users with Multimedia (用多媒体吸引用户)、Understanding Multimedia Content (理解多媒体内容)。

Multimedia Systems (多媒体系统) 今年包含的子主题为: Systems and middleware (系统与中间件)、Transport and delivery (传输与交付)、Data Systems Management and Indexing (数据系统管理与索引), 去掉了 2018 年的 Multimedia Telepresence and VR/AR (遥现与虚拟/增强现实)、Multimedia Scalability and Management (扩展与管理)、Mobile Multimedia (移动多媒体)。

Multimedia Experience (多媒体体验) 今年包含的子主题为: Interactions and Quality of Experience (交互和体验质量)、Art and Culture (艺术与文化)、Multimedia Applications (多媒体应用), 去掉了 2018

年的 Multimedia for Collaboration in Education & Distributed Environments (教学协作与分布式环境)、Music, Speech and Audio Processing in Multimedia (音乐、语音、声音)、Multimedia Authoring and Enrichment (多媒体创作与丰富化), 新添加了多媒体艺术与文化。

Engaging Users with Multimedia (用多媒体吸引用户) 主要包含: Emotional and social signals (情感与社会信号)、Multimedia Search and Recommendation (多媒体检索与推荐)、Summarization, Analytics, and Storytelling (摘要、分析与故事讲述), 去掉了 2018 年的 Social Multimedia (社交多媒体)。

Understanding Multimedia Content (理解多媒体内容) 今年全面改版, 新增了: Multimodal Fusion and Embedding (多模态融合与嵌入)、Vision and Language (视觉与语言)、Media Interpretation (媒体解译), 去掉了 2018 年全部的三个子主题: Deep Learning for Multimedia (多媒体深度学习)、Multimodal Analysis and Description (多媒体分析与描述)、Multimedia and Vision (多媒体与视觉)。

笔者认为在深度学习技术火了几年之后, 现在多媒体领域已经进入了后深度学习时代, 更多的注意力转向多媒体问题本身的解决, 而不是只关注特殊的方法, 多模态融合尤其是视觉和语言的融合成为了新主题, 笔者团队发表的论文^[1]正好就属于该新主题。

二、参会与论文录用情况

本次会议的参会人员共 750 名，来自 42 个国家，收到摘要投稿共 1275 篇，有效投稿 936 篇，共有 98 名领域主席以及 758 名审稿人，最终录用了 252 篇论文，录用率为 27%。

在被录用的论文中，占比最多的是 Media interpretation (媒体解译)，共 89 篇，占比 35.3%，其次是 Vision and Language (视觉与语言)，共 31 篇，占比 12.3%，排名第三的是 Multimodal Fusion & Embedding (多模态融合与嵌入)，共 25 篇，占比 9.3%。占比最少的主题是 Art and Culture 共 6 篇，占比 2.4%，其次的两个主题是 Data Systems Management and Indexing (数据系统管理与索引) 与 Summarization, Analytics, and Storytelling (摘要、分析与故事讲述)，各 7 篇，占比都是 2.8%。

与往年不同的是，今年第一次增加了一个代码可重复 Track: ACM Multimedia Reproducibility Track，收录的是 2018 年 ACM MM 录用论文中一些可重复代码的论文，以鼓励作者提供自己论文的相关代码，增加论文实验的可信度及对比实验的便利性。笔者认为这个学术氛围非常好，这两年在生物学领域出现了一些实验不可重复的丑闻，而在计算机领域，不应该出现这个现象，今年的这个 Track 为计算机领域的学术界注入了一股清风，为今后学术界的规范开了一个好头。

三、Keynote Speak

笔者今年听了两个主旨报告，第一个是来自法国国家视听研究所 (The French National Audiovisual Institute) 的 Jean Carrive，题目是：“Using Artificial Intelligence to Preserve Audiovisual Archives: New Horizons, More Questions”。法国国家视听研究所成立于 1974 年，是一个保存国家视听资料遗产的公共研究所，集科研、培训、教育、产品、咨询、编辑于一体的大型综合研究所。该研究所拥有长达 123 公里的模拟视听材料，大于 1900 万小时的电视与广播，预计在 2021 年能够完成 100% 的数字化。该研究所正在利用语音翻译、图像识别、人脸和人声识别等人工智能技术

建立视听资料的索引，同时指出了在如此庞大的视听资料中进行人工智能分析面临的两大困难：(1) 没有现成的数据集可以用于训练，(2) 资料中概念和词汇过于复杂，期望能做成一个公开项目，以得到学术界更多的技术支持。

笔者今年听的第二个主旨报告是来自比利时布鲁塞尔自由大学 (Free University Brussels) 和荷兰拉德堡德大学 (Radboud University) 的 Mireille Hidebrandt 教授，题目是：“EU Data Protection Law: An Ally for Scientific Integrity?” Hidebrandt 教授介绍了欧洲的 GDPR 组织 (General Data Protection Regulation)，是一个欧洲专门从事数据保护的公共组织，并且介绍了欧洲数据保护法的相关情况。

四、Best Paper Session

与计算机视觉的会议不同，ACM MM 会议的 Best Paper Session 需要进行现场的报告和投票，因此笔者听到了 5 篇最佳论文提名的精彩报告，分别来自中科院自动化所^[2]、山东大学^[3]、美国约翰霍普金斯大学/Snap 研究院^[4]、新加坡国立大学^[5]、山东师范大学^[6]。内容涉及基于多模态知识网络的医学问答^[2]、多模态对话系统^[3]、视听缩放^[4]、隐私保护^[5]、基于多模态哈希的大规模多媒体检索^[6]。最终，美国约翰霍普金斯大学/Snap 研究院的 “Audiovisual Zooming: What You See is What You Hear” 胜出获得了最佳论文奖，在该篇论文中，作者创造性的提出了 Audiovisual Zooming 问题，能够在视频进行缩放的同时，缩放画面所对应的声音，用户可以只听见所看见的人或物体。

五、Brave New Ideas

这个 Session 是笔者所感兴趣的版块之一，共包含了四篇论文，分别来自中国人民大学^[7]、荷兰阿姆斯特丹大学^[8]、德国马克思普朗克研究所^[9]、新加坡国立大学^[10]。内容涉及故事可视化^[7]、多模态数据超学习 (HyperLearn)^[8]、基于人眼聚散行为的内部思维检测^[9]、图像主观属性学习^[10]，其中中国人民大学的 Shizhe Chen 以画饼充饥这个故事所对应的图像序列生成为例子，生动的讲解了内容涉及故事可视化方法。

六、总结与展望

ACM MM 会议还有多个丰富多彩的 sessions，例如 workshops、Art & Culture、艺术展、Best Demo、Open Source Competition 等等，会议时间有限，笔者并没有在上文中介绍全部的内容，笔者的理解仅代表个人观点，如有偏差还请各位读者指出。在最后的 SIGMM(国际多媒体特别专家组)会议上，组委会介绍

了 2020 年的一些多媒体相关会议安排，包括 ICMR、ACM IH & MMsec、ACM Multimedia Asia(由 PCM 和 ICIMCS 合并)等。2020 年的 ACM MM 会议将在 2020 年 10 月 12 日至 16 日在美国西雅图举行，期待与 MM 的下一次相遇。

责编委 王金甲

参考文献

- [1] Jin X, Wu L, Zhao G, et al. Aesthetic Attributes Assessment of Images. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [2] Zhang Y, Qian S, Fang Q, et al. Multi-modal Knowledge-aware Hierarchical Attention Network for Explainable Medical Question Answering [C]. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [3] Nie L, Wang W, Hong R, et al. Multimodal Dialog System: Generating Responses via Adaptive Decoders. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [4] Nair A, Reiter A, Zheng C, et al. Audiovisual Zooming: What You See is What You Hear. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [5] Shen Z, Fan S, Wong Y, et al. Human-imperceptible Privacy Protection Against Machines. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [6] Lu X, Zhu L, Cheng Z, et al. Flexible Online Multi-modal Hashing for Large-scale Multimedia Retrieval. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [7] Chen S, Liu B, Fu J, et al. Neural Storyboard Artist: Visualizing Stories with Coherent Image Sequences. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [8] Ayra D, Rudinac S, Worring M. HyperLearn: A Distributed Approach for Representation Learning in Datasets with Many Modalities. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [9] Huang X, Li J, Ngai G, et al. Moment-to-Moment Detection of Internal Thought during Video Viewing from Eye Vergence Behavior. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.
- [10] Gelli F, Uricchio T, He X, et al. Learning Subjective Attributes of Images from Auxiliary Sources. ACM Multimedia (ACM MM), Nice, France, 21-25 October, 2019.



金鑫

北京电子科技学院副教授，主要研究方向为计算美学与多媒体信息处理，主页：jinxin.me。
Email: jinxin@besti.edu.cn