

热点追踪

# 基于元学习的泛化人脸识别

中科院自动化研究所 郭建珠 朱翔昱 雷震

人脸识别模型通常需要部署在未知的场景中，识别未知的人群。由于训练数据和目标场景的分布偏差，模型在未知场景的识别性能往往表现很差，比如，在网图训练的模型部署在监控场景中，识别的准确率会大打折扣，在这对模型的泛化性能带来了极大的挑战。为解决该技术难题，本文提出了一种基于元学习的泛化人脸识别框架。该框架在训练采样过程中，通过采样元训练域和元测试域来模拟场景的分布偏差，并通过一种提出的元优化的损失，使得模型同时在元训练和元测试域上都能有提升，进而提升模型在真实未知场景下的泛化性。相关成果被 CVPR2020 录取为口头报告。

近些年得益于深度学习的发展，人脸识别性能在一些通用测试集上得到了极大提升<sup>[1,2,3,4,5]</sup>，然而这些测试集与训练集有着类似的数据分布。当识别模型部署在实际场景中时，由于目标场景与训练集数据分布不一致，模型的性能会显著下降。并且目标场景通常是未知的，数据亦不可获取，无法使用目标场景的数据对模型进行微调。针对这个挑战，我们提出了泛化人脸识别问题：在目标场景未知的条件下，如何设计一个有效的训练策略或方法，让模型能在未知场景下取得较好的泛化性，

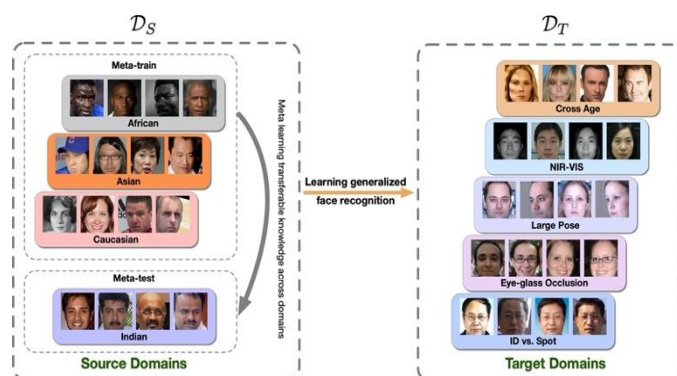


图 1 泛化人脸识别

如图 1 所示。

针对泛化人脸识别问题，我们提出了一种基于元学习的人脸识别框架 MFR (Meta Face Recognition)。MFR 主要包括三部分：(1) 跨域采样；(2) 多域分布优化；(3) 元优化。整体的框架如图 2 所示。

首先，跨域采样是为了模拟训练场景和测试场景的分布偏差，每次迭代时，根据训练集的域标签，将训练集分为元训练域和元测试域，并在两个域中分别采样一定人数。其次，在多域分布优化中，我们使用了三种损

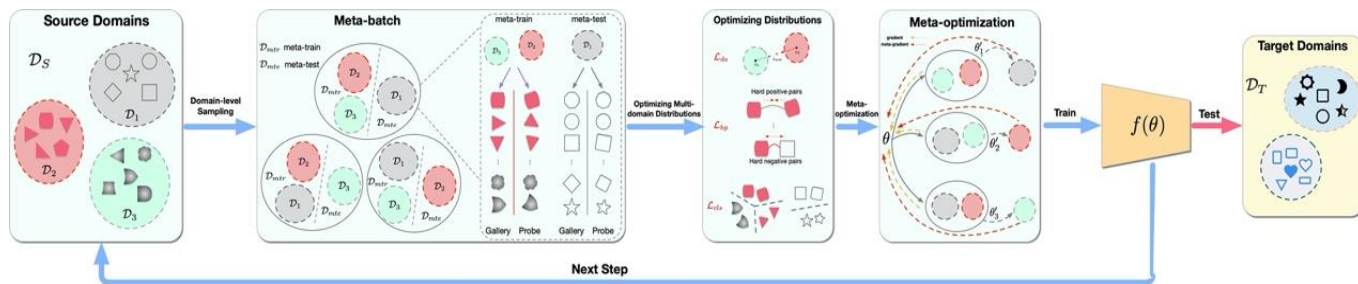


图 2 MFR 整体框架

失函数，包括难样本损失、软分类损失和域对齐损失，来学习具有判别性和域不变性的人脸表征。最后，元优化通过三个步骤对模型进行优化。(1) 元训练：对元训练域进行优化，并得到梯度更新后的模型参数；(2) 元测试：在元测试域上，对元训练更新后的模型参数进行二次更新；(3) 对元优化和元测试的损失进行加权，对原始模型的参数进行梯度反传更新。元优化的示意图如图 3 所示。

为了验证模型的泛化性，我们提出了两个不同难度的测试协议，如表 1 所示。一个是跨种族（印度人、非洲人、亚洲人，高加索人）测试协议 GFR-R；另一个是跨场景测试协议 GFR-V，其更接近实际场景，也更具挑战性。在协议中，目标域的数据在训练中是未知的，用于模拟未知的应用场景。通过表 2 可以看出，通过元学习，本方法在跨种族和跨场景的测试中均取得了最好的性能。

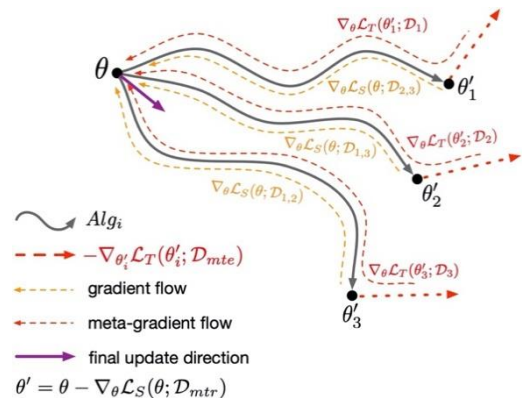


图 3 元优化

Protocol	Source Domains	Target Domain(s)	
GFR-R	I	Caucasian	Indian
		Asian	
		African	
		Indian	
	II	Caucasian	African
		Asian	
		Indian	
		Indian	
	III	Caucasian	Asian
		African	
		Indian	
		Indian	
	IV	Asian	Caucasian
		African	
		Indian	
		Indian	
GFR-V	Caucasian	CACD-VS	
	Asian	CASIA NIR-VIS 2.0	
	African	MultiPIE	
	Indian	MeGlass	
	Indian	Public-IvS	

表 1 GFR-R 跨人种测试协议和 GFR-V 跨域测试协议

Protocol	Method	VR (%)			Rank-1 (%)
		FAR=1%	FAR=0.1%	FAR=0.01%	
GFR-R I (Indian)	Base	94	82.2	64.65	80.3
	Base-Agg	94.1	80.9	65.3	81
	Base-FT rnd.	62.5	39	21.05	39.3
	Base-FT imp. [34]	87	69.9	51.2	69.6
	MLDG [12]	94.2	83	66.3	80.5
	MFR (Ours)	<b>95.4</b>	<b>86.1</b>	<b>71.4</b>	<b>83.1</b>
GFR-R II (African)	Base	91.6	74.5	55.4	73.1
	Base-Agg	90.5	74.8	56.3	74
	Base-FT rnd.	26.2	10.9	3.5	21
	Base-FT imp. [34]	78.7	56.6	36.45	57.9
	MLDG [12]	91.9	74.8	55.7	73.8
	MFR (Ours)	<b>92.3</b>	<b>79.4</b>	<b>60.8</b>	<b>75.2</b>
GFR-R III (Asian)	Base	91.89	77.98	60.86	75.98
	Base-Agg	91.49	78.08	59.41	76.28
	Base-FT rnd.	40.44	17.32	7.67	27.53
	Base-FT imp. [34]	80.58	57.56	39.79	61.86
	MLDG [12]	92.29	78.28	60.3	76.68
	MFR (Ours)	<b>93.49</b>	<b>80.7</b>	<b>62.56</b>	<b>78.68</b>
GFR-R IV (Caucasian)	Base	96.6	89.6	78.6	86.6
	Base-Agg	97	88.1	79.1	86.8
	Base-FT rnd.	61.1	36.2	18.9	36.7
	Base-FT imp. [34]	91.5	78.2	63.4	76.8
	MLDG [12]	96.8	89.6	79.15	86.3
	MFR (Ours)	<b>98.2</b>	<b>92.9</b>	<b>81.1</b>	<b>88.9</b>
GFR-V (MeGlass)	VR (%)				
		FAR=0.01%	FAR=0.001%	FAR=0.0001%	Rank-1 (%)
	Base	85.92	71.96	53.5	97.6
	Base-Agg	86.77	73.5	54.96	97.69
	MLDG [12]	85.54	69.23	49.32	<b>97.81</b>
	Face Syn. [31]	<b>90.14</b>	<b>80.32</b>	<b>66.92</b>	96.73
MFR (Ours)	<b>90.79</b>	<b>80.86</b>	<b>66.15</b>	<b>98.57</b>	

表 2 GFR-R 跨人种测试协议和 GFR-V 跨域测试协议结果

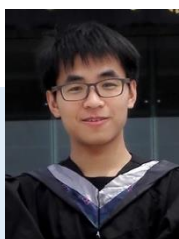
责任编辑 崔海楠

参考文献

[1] Yaniv Taigman, Ming Yang, Marc’Aurelio Ranzato, and Lior Wolf. Deepface: Closing the gap to human-level performance in face verification. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1701–1708, 2014.

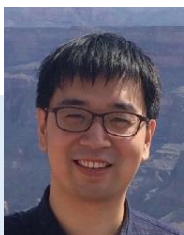
[2] Yi Sun, Xiaogang Wang, and Xiaoou Tang. Deep learning face representation from predicting 10,000 classes. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1891–1898, 2014.

- [3] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 815–823, 2015.
- [4] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Spheraface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 212–220, 2017.
- [5] Jiankang Deng, Jia Guo, Xue Niannan, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In CVPR, 2019.



## 郭建珠

中科院自动化研究所博士生。主要研究方向为人脸识别和三维人脸。  
Email: jianzhu.guo@nlpr.ia.ac.cn



## 朱翔昱

中科院自动化研究所副研究员。主要研究方向为人脸识别、人脸对齐和三维人脸。  
Email: xiangyu.zhu@nlpr.ia.ac.cn



## 雷震

中科院自动化研究所研究员。主要研究方向为人脸检测、人脸对齐、人脸识别、计算机视觉、模式识别、人工智能、图像处理等。  
Email: zlei@nlpr.ia.ac.cn

热点追踪

# 多图的离线协同与在线增量匹配

上海交通大学 严骏驰

**图** 匹配问题一直是计算机视觉领域中一个十分基础而重要的问题，其宗旨在于寻找两个或多个存在复杂关系的目标集间元素的对应关系。图匹配最经典的应用之一在于将不同图片中的特征点进行匹配，这是包括图像匹配、图像识别与理解、三维多视角重建、几何物体形状匹配、运动结构理解等很多计算机视觉问题的关键一环。另一方面，图匹配一般是指在给定相似性模型下，寻找两个或多个图的节点之间的对应关系。其二图匹配的形式一般被规约成劳勒的二阶指派问题（Lawler's Quadratic Assignment Programming），这已经被证明为 NP 难问题，而基于二图匹配提出的多图匹配形式则更加复杂，难度也更大。图匹配领域的突破也在为研究

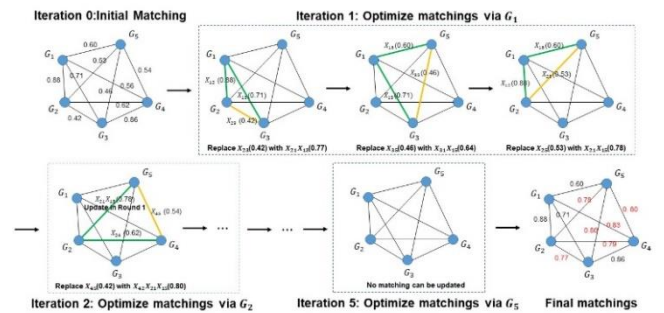


图 2 MGM-Floyd 算法流程

不错的效果。该方法遵循匹配链优化的思想来解决多图匹配问题，通过搜索匹配之间的组合来优化原有的二图匹配。通过超图的视角看待问题，超图的顶点表示需要匹配的图，超图的边表示两个图之间的匹配，搜索匹配链的组合方式相当于在超图上搜索最优的路径。研究者证明了由相似度得分和循环一致性组成的匹配质量函数具有与距离函数相似的性质，并通过严密的推理论证最短路算法与多图匹配存在结合的理论基础。基于这一点，融合了经典 Floyd 算法的离线多图求解器 MGM-Floyd 与融合了单源最短路 SPFA 算法的在线多图求解器 MGM-SPFA、FastSPFA 被提出。相较于现有

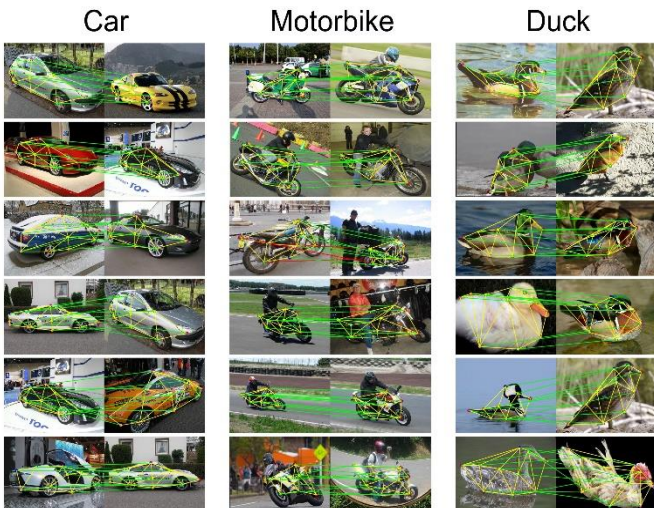


图 1 Willow Object 数据集上 Floyd 方法匹配效果

NP 难的组合优化问题作出贡献。

近期，基于超图上的最短路径优化算法引入多图匹配领域，同时在离线协同匹配与在线增量匹配领域取得

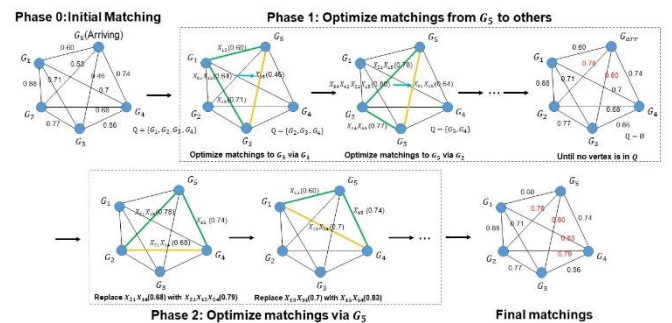


图 3 MGM-SPFA 算法流程

category	Hungarian	Floyd-pc	Floyd-uc	Floyd-c	Spectral [20]	MatchLift [21]	MatchALS [22]	MISM [5]
Car	0.503	0.844	0.840	<b>0.850</b>	0.601	0.665	0.629	0.750
Duck	0.442	<b>0.803</b>	0.800	0.793	0.485	0.554	0.525	0.732
Motorbike	0.317	0.821	0.817	<b>0.843</b>	0.255	0.296	0.310	0.653
Face	0.854	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	0.927	0.931	0.934	0.937
Winebottle	0.543	<b>0.934</b>	0.930	0.931	0.630	0.700	0.669	0.814
Time (Car)	1.263	9.377	10.960	14.981	<b>1.469</b>	17.695	2.622	3.966

表 1 不同离线多图匹配求解器在 Willow-Object 数据集上的匹配精度以及匹配速度的表现

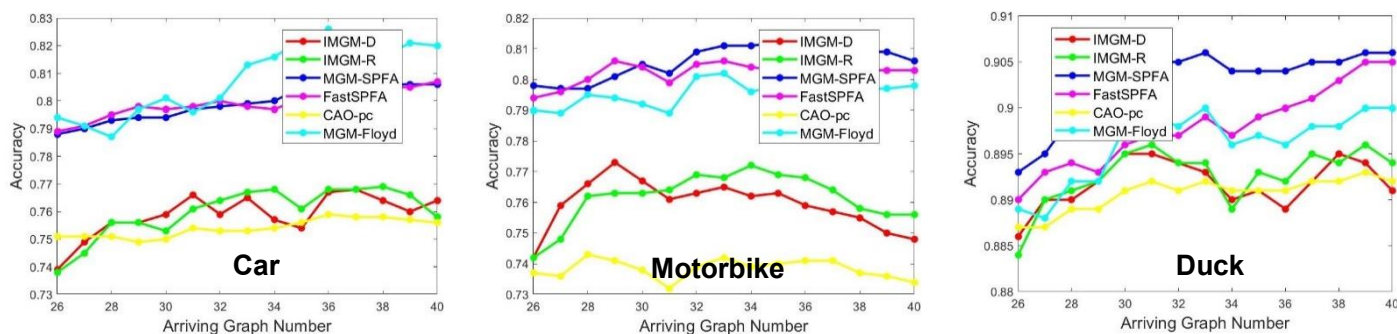


图 4 不同在线多图匹配求解器在 Willow-Object 数据集上的匹配精度的表现

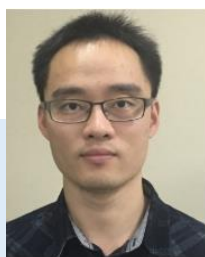
的求解器，这些算法在时间、精度等多个方面都有较大的提升。

这些算法都在仿真数据集和真实数据集上都进行了实验。仿真测试数据集包括形变测试，外点测试，遮罩测试等多种环境下的效果测试；真实数据集则包括 Willow-ObjectClass 和 Sub-Rome16k 两个真实数据集。从综合的角度来看，本文提出的方法在所有数据集的离线和在线实验中都拥有着最佳的性能。在离线模式下，MGM-Floyd 的综合表现上比较突出，相较于 SOTA，在对应模式下或是精度或是速度都有着明显的优势，在与其他连续空间离线多图匹配求解器比较的时候精度优势则更大，精度有着 10%-20% 的提升。在线模式中，MGM-SPFA 在在线多图增量匹配问题中精度最高，而

FastSPFA 精度上与 MGM-SPFA 比较接近，速度上比其他的在线多图匹配求解器快了将近 1-2 倍。与此同时，经过试验检验，以上所有算法都有着相当高的稳定性与鲁棒性，可以适应复杂的现实场景图片匹配应用。

以上工作已发表于国际期刊 IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)。具体题目为：Unifying Offline and Online Multi-graph Matching via Finding Shortest Paths on Supergraph。三位作者是蒋泽天（本科生），王天哲（本科生），严骏驰（副教授），均来自上海交通大学。

责任编辑 储璐



## 严骏驰

上海交通大学计算机系与人工智能研究院长聘教授。主要研究方向为机器学习与计算机视觉。

Email: yanjunchi@cs.sjtu.edu.cn

热点追踪

# 上下文感知的密集视频描述生成

中国人民大学 陈师哲 金琴

视频描述生成是全面理解视频语义内容的一个重要任务。传统视频描述生成任务主要是为视频生成单句话的文本描述。然而，长时间视频中往往包含多个事件，难以用单句话概括。因此，越来越多的研究开始关注于密集视频描述生成(Dense Video Captioning)任务<sup>[1]</sup>。这一任务旨在检测长时间视频中的所有事件，并分别为每一事件生成文本描述。图 1 展示了密集视频描述的一个示例。目前的方法主要将这一任务分解成两部分进行解决，分别是事件检测和事件描述生成，其中事件检测模型负责在长时视频中检测出有意义的事件，事件描述模型为检测的每一个事件生成一句文本描述。

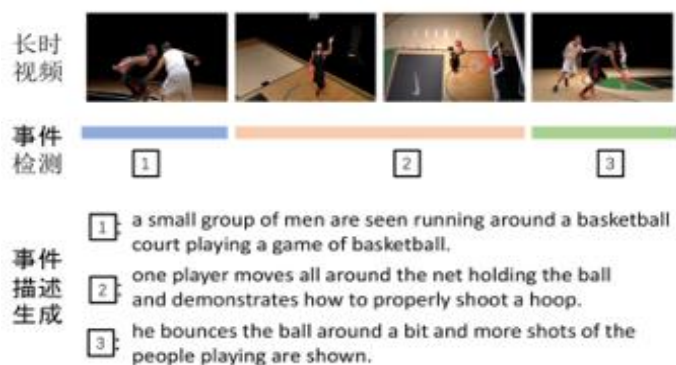


图 1 密集视频描述示例

为了推动密集视频描述这一研究任务的发展，CVPR ActivityNet Workshop 从 2017 年开始连续 4 年举办了 Dense-Captioning Events in Videos 比赛。该挑战赛由斯坦福大学组织，基于 ActivityNet 密集视频描述数据集，共有 20000 个视频，平均每个视频包含 3.7 句事件描述。该竞赛吸引了卡内基梅隆大学、北京大学、清华大学、复旦大学、浙江大学等国内外科研

院校，以及百度、京东、腾讯、三星等公司和研究院参加。我们团队在该比赛上从 2018 年到 2020 年连续 3 年获得了冠军。下面简单介绍我们提出的方案——上下文感知的密集视频描述生成模型。

上下文关系对于检测和理解长视频中的不同事件至关重要。首先，视频中的不同事件不是独立的，上下文为理解每个事件提供更全面的信息，使得事件语义理解更为准确；其次，上下文关系有利于区分不同事件，不仅可以使得检测的事件片段更为多样化，而且使得生成的事件文本描述更具有区分性。

对于事件检测，我们提出了基于时序上下文的事件检测模型<sup>[2]</sup>，图 2 展示了所提出模型的结构框架。该模型采用 GRU 网络直接进行事件时间戳的预测，输入为前一步所生成的事件片段的位置和语义特征，从而可以根据上文信息大大减少生成事件片段的冗余性。为了更有效利用事件的双向时序信息，我们融合从前往后以及从后往前的事件时间戳预测结果。实验结果表明我们提出的模型可以生成更加准确的事件片段候选。

对于事件描述生成，我们系统地探讨和比较了不同上下文特征和模型的影响<sup>[3]</sup>。我们提出五种上下文类型，分别是上下文增强的帧级特征、局部上下文、全局上下文、事件上下文和句子上上下文，以及两种上下文编码模型，分别是事件独立模型和事件相关模型。图 3 展示了上下文感知的事件独立模型示意图。实验结果表明，利用上下文增强的帧级特征和局部上下文的事件独立模型在速度和性能上达到较好平衡，在事件描述准确性和多样性两个方面均有提高。

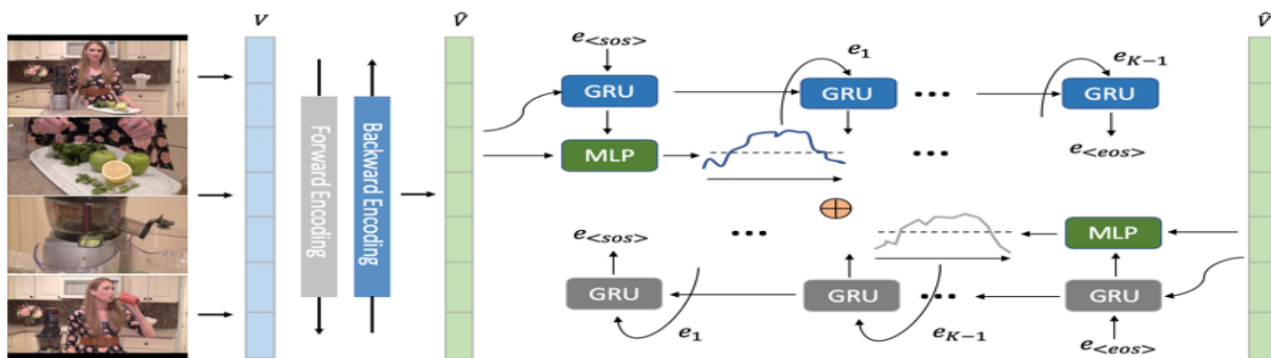


图2 基于时序上下文的事件检测模型

密集视频描述为全面和细粒度的视频语义理解提供了新的途径，但目前的研究还较为初步，还存在着许多问题值得探索。例如，使得生成的不同事件描述更具有连续性，形成逻辑连贯的段落描述，而不是独立的事件描述。现有评测标准也具有局限性，仅从检测的精度和描述准确率考虑，缺乏对事件检测的召回率和描述多样性等不同方面的综合考量。

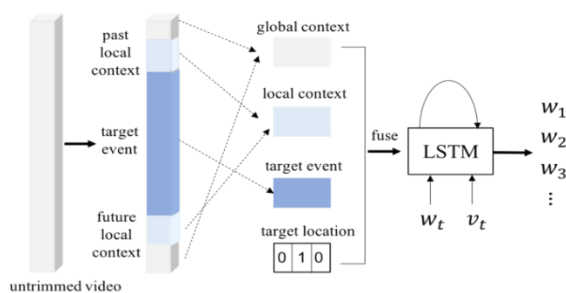


图3 上下文感知的事件描述生成模型

责任编辑 王金甲 任桐炜

## 参考文献

- [1] Krishna, Ranjay, Kenji Hata, Frederic Ren, Li Fei-Fei, and Juan Carlos Niebles. "Dense-captioning events in videos." In Proceedings of the IEEE international conference on computer vision, pp. 706-715. 2017.
- [2] Song, Yuqing, Shizhe Chen, Yida Zhao, and Qin Jin. "Team RUC\_AIM3 Technical Report at Activitynet 2020 Task 2: Exploring Sequential Events Detection for Dense Video Captioning." arXiv preprint arXiv:2006.07896 (2020).
- [3] Chen, Shizhe, Yuqing Song, Yida Zhao, Qin Jin, Zhaoyang Zeng, Bei Liu, Jianlong Fu, and Alexander Hauptmann. "Activitynet 2019 task 3: Exploring contexts for dense captioning events in videos." arXiv preprint arXiv:1907.05092 (2019).



## 陈师哲

中国人民大学信息学院博士，主要研究方向为视觉与自然语言、视频语义理解和多模态计算。  
Email: cszhel@ruc.edu.cn



## 金琴

中国人民大学信息学院教授，主要研究领域包括多媒体智能计算、人机交互等。  
Email: qjin@ruc.edu.cn

顶会观察

## CVPR 2020

中山大学 李冠彬

**国**际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR) 是计算机视觉及模式识别的顶级学术会议, 与 ICCV、ECCV 并称为计算机视觉领域三大顶会。CVPR 有着非常高的学术影响力, 在中国计算机学会推荐国际学术会议被评为人工智能领域的 A 类会议; 在 Google Scholar 发布的 2020 年学术指标中, 以高达 299 的 h5 指数雄踞人工智能领域的榜首、位列所有领域出版物的第五位。因此, CVPR 每年都吸引来自世界各地的参会者, 其中不乏学术界和工业界的泰斗大牛、广罗人才的企业机构, 以及奋战于科研一线的研究生、博士生、教授、研究员等。

CVPR 2020 原定于 2020 年 6 月 14 日至 19 日于美国华盛顿州西雅图举行, 由于 COVID-19 新冠病毒在世界各地陆续爆发, 本届 CVPR 转为完全线上虚拟会议方式进行, 给予了各个参会人员全新的参会体验。

## 一、国际计算机视觉与模式识别会议的亮点

在大会举办日期的前几个月, 新冠肺炎疫情在世界各地各个国家陆续爆发, 因此 CVPR 2020 首次采用了完全数字虚拟会议的方式, 让世界各地的参会者通过互联网参加虚拟会议。会议收录的论文作者需要准备各种形式的材料, 包括摘要、图片、论文讲解视频等, 作为虚拟会议的交流媒介。会议主办方通过邮件进行协调, 并建立了虚拟会议的网址, 以供参会人员进行展示及技术交流。每篇论文的简介及讲解视频会根据论文主题进行分类, 在网页上进行展示。且每篇论文安排了两次线上交流时间, 允许作者与感兴趣的参会者通过文本聊天或线

上会议的方式进行答疑与交流, 这为不同时区的参会者提供了非常大的便利。本次虚拟会议在保证安全的前提下, 最大限度地保障了参会者的自由度、便利程度, 让大家获得了很好的参会体验及交流经历。

## 二、论文录用情况

CVPR 2020 总共收到了 6656 篇有效投稿, 其中 1470 篇论文被接收, 接收率约为 22%, 从 CVPR 2018 开始论文接受率连续两年下降, 而今年创下近十年来 CVPR 会议的最低接收纪录。其中仅有 335 篇论文入选 oral presentation, oral 率约为 5.7%。从大会公布的数据来看, 今年大会接收到的注册及有效投稿数量都有显著的提高, 可见被接受论文的质量之高。

本次 CVPR 2020 收录论文中, 中国作者占论文作者总人数的 39.2%, 占比为所有国家中最高。从大学/机构来看, 来自清华大学的作者人数最多, 排名第一。其他贡献突出的中国高校还有上海交通大学、北京大学、浙江大学等。除了学术界的科研力量, 来自中国工业界的各大互联网企业也获得了不俗的成绩, 其中包括华为、阿里巴巴、商汤、百度、腾讯、旷视等。其中, 商汤及联合实验室共 62 篇论文入选, 华为研究团队有 34 篇论文入选, 阿里巴巴、旷视分别有 16 篇论文入选。这些被录用的论文在许多重要工业应用领域有着重大的突破, 包括网络结构搜索、三维点云理解、网络模型训练加速与量化、视频理解领域等。此外, 谷歌在本次会议表现亮眼, 共有近 70 篇论文入选, 涵盖了诸多热门领域, 如图像识别、自动驾驶、语义分割、目标追踪等。

CVPR 2020 会议涵盖的方向包括目标检测、目标

跟踪、图像分割、人脸识别、姿态估计、三维点云、视频分析、模型加速、GAN 和 OCR 等，论文关键词包括迁移学习 (Transfer Learning)、表示学习 (Representation Learning)、机器人 (Robotics)、三维数据分析、无监督学习 (Unsupervised Learning) 和自监督学习 (Self-supervised learning) 等。

### 三、主题演讲

今年 CVPR 2020 邀请到了两位重量级嘉宾进行炉边谈话，他们分别是微软 CEO Satya Nadella 与亚马逊云服务 SVP Charlie Bell。Satya 在与微软公司前执行副总裁沈向洋的对话中，分享了对计算机视觉、人工智能研究与应用前景的思考与展望。Satya 首先回顾了自己从印度生活开始，到开始在美国学习计算机科学并最终成为微软 CEO 的经历，指出“予力全球每一人、每一组织，成就不凡”这一微软使命是鞭策他每天前行努力工作的根基。Satya 也列举了计算机视觉当今的一些精彩应用，以及未来发展的热点方向。计算机视觉技术为工业界许多场景提供了有力的支持，包括计算机视觉在 3D 场景下的应用，在云和边缘设备上的创新应用，在疾病筛查与诊断场景下的应用。关于计算机视觉技术的未来发展热点，Satya 指出了三个突破方向，包括 4D Understanding (4D 理解)，Background Matting (背景替换)，Holoportation (全息瞬移)。4D 理解能够让计算机理解跟踪物品、人、互动行为及群组活动；背景替换能够在不适用绿幕情况下将演讲者的图像投影到虚拟舞台上，制造更真实的会议体验；全息瞬移则能超越时空及语言局限，使得演讲者能够用多种语言和形象进行流利的演讲。此外，Satya 还对人工智能接下来的发展方向、应用领域、社区间的国际合作、行业机遇、AI 设计的道德问题等分享了自己的观点。

值得一提的是，本届 CVPR 特意缅怀了 84 岁华人计算机视觉泰斗 Thomas S. Huang (黄煦涛)。黄煦涛教授是首位担任 CVPR 程序主席的华人，也是华人计算机视觉史上的关键人物，他为华人 CV 领域留下丰厚的学术遗产。经 PAMITC 执行委员会批准，CVPR 设立计算机视觉领域的 Thomas S. Huang 纪念奖，从 2021 年开始每年在 CVPR 上颁发。

### 四、会议热点论文与教程

本次会议涌现了许多热点论文，它们具有非常高的学术价值与应用价值。其中，CVPR 2020 的最佳论文<sup>[1]</sup>、最佳学生论文<sup>[2]</sup>一作均为华人，且均聚焦于 3D 领域。最佳论文<sup>[1]</sup>亮点在于仅给定一张单目图像，可在解构拍摄视角的同时，将其深度 (3D)、光照等分解出来，采用无监督的方式估计 3D 对象，还原物体原始的 3D 面貌。传统的 3D 重建方法需要有监督的信息 (如多视图、深度图、轮廓、关键点)；当训练数据足够多，模型训练成功以后，才可基于双目图像估计出深度图，该类方法通常需要耗费高昂的成本。而本文基于自动编码器解构给定图像的视角、深度、texture 等维度，再经过组合渲染，重构 3D 模型。该方法在多个不同类别的物体重建中证明了其潜在的应用价值，不仅可根据人脸图像进行重建，还可对绘画作品、动物等进行重建。

最佳学生论文<sup>[2]</sup>提出一种解决多边形网格生成问题的新思路。该工作受计算机图形学经典空间数据结构 Binary Space Partitioning (BSP) 的启发，对空间进行递归细分以获得凸集，从而使网络在无监督条件下可以通过凸分解学习 3D 形状高质量表示。BSP-Net 能自动生成最少的多边形，合成外形尽量完美、真实的三维物体。相比其它神经网络模型，BSP-Net 所用的多边形数量大幅减少，镶嵌效果更好。且通过 BSP-Net 训练的凸面可以轻松提取以形成多边形网格，无需进行昂贵的等值曲面处理。

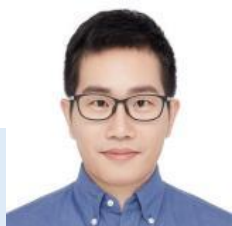
另外，来自国内学术界和工业界的多篇论文也引起了广泛的讨论。香港中文大学推出面向细粒度动作分析的高质量数据集 FineGym<sup>[3]</sup>，通过对现有动作识别方法进行分析与对比，为未来细粒度动作相关研究任务提供了方向。腾讯提出神经网络滤波器嫁接技术<sup>[4]</sup>，通过引入外部信息激活无效的滤波器，从而大幅度提高模型性能。香港科技大学与腾讯优图实验室提出基于注意力机制及多关系检测器的小样本物体检测<sup>[5]</sup>，基于注意力机制构建的 RPN、多关系检测器及三元组对比训练方法，使得网络能够不经重新训练就能用于新类别检测，并提出了 1000 类的小样本物体检测数据集。华为诺亚方舟实验室提出 AdderNet<sup>[6]</sup>，在深度学习领域开创性地提

出了加法神经网络，将卷积计算中的乘法操作替换成了轻量级的加法计算，在国内外引起了热烈的讨论。商汤 EIG 算法中台团队针对物体检测中的尺度问题，设计了更具有等变性的特征金字塔以及物体检测器结构<sup>[7]</sup>，使得单阶段检测器的性能在 COCO 数据集上大幅度提升。香港大学和中山大学合作提出了 Ref-Reasoning 数据集<sup>[8]</sup>，包含真实图像和具有不同推理布局的语义丰富的表达式，是一个可用于结构化的指称表达式推理的大规模真实数据集。该论文同时提出了一种场景图引导的模块网络，提升了基于语言描述的视觉目标定位这一问题求解过程的可解释性。

此外，CVPR 2020 也举办了很多高质量的研讨会 (Workshops) 与教程 (Tutorials)。涉及的主题包括可解释机器学习、视频建模教程、自动化深度学习 (覆盖超参搜索及网络结构搜索)、图网络结构表示学习、视觉-语言研究进展、自动驾驶等。这些研讨会与教程为大家提供了很好的学习交流的平台，起到了全面的指导作用。

## 参考文献

- [1] Wu, Shangzhe, Christian Rupprecht, and Andrea Vedaldi. "Unsupervised Learning of Probably Symmetric Deformable 3D Objects from Images in the Wild." IEEE/CVF CVPR 2020.
- [2] Chen, Zhiqin, Andrea Tagliasacchi, and Hao Zhang. "BSP-Net: Generating Compact Meshes via Binary Space Partitioning." IEEE/CVF CVPR 2020.
- [3] Shao, Dian, et al. "FineGym: A Hierarchical Video Dataset for Fine-grained Action Understanding." IEEE/CVF CVPR 2020.
- [4] Meng, Fanxu, et al. "Filter Grafting for Deep Neural Networks." IEEE/CVF CVPR 2020.
- [5] Fan, Qi, et al. "Few-shot Object Detection with Attention-RPN and Multi-relation Detector." IEEE/CVF CVPR 2020.
- [6] Chen, Hanting, et al. "AdderNet: Do We Really need Multiplications in Deep Learning?" IEEE/CVF CVPR 2020.
- [7] Wang, Xinjiang, et al. "Scale-Equalizing Pyramid Convolution for Object Detection." IEEE/CVF CVPR 2020.
- [8] Yang, Sibe, et al. "Graph-Structured Referring Expression Reasoning in the Wild." IEEE/CVF CVPR 2020.



## 李冠彬

中山大学数据科学与计算机学院副教授。主要研究方向为计算机视觉和机器学习。  
Email: liguanbin@mail.sysu.edu.cn

## 五、总结与展望

回顾近三年的 CVPR 最佳论文，我们可以发现，CVPR 越来越青睐致力于解决真实场景下存在的视觉问题的方法或工作，包括对真实场景的建模、实现各类视觉子任务 (分类、检测、语义分割) 的统一融合，从而模拟人类视觉系统对真实物理世界的认知。从其关键词来看，实现该目的的概念方法 (meta learning、life-long learning、robust learning) 也是当前及若干年内可以想见的热门领域。

从 Staya 的炉边谈话以及最近举办的 AAAI/ICLR/ICASSP 等会议可以看出，线上虚拟会议一方面能够确保参会者的交流与体验，另一方面能够较大程度上节约参会者的开销与时间。此外，这种方式无疑在新冠肺炎疫情肆虐的这段时间最大程度保障了学术活动的正常进行。在计算机视觉技术日益发展的今日，或许这种方式可以成为未来科研交流的主要形式。

责任编辑 王金甲