

主办 CCF 计算机视觉专业委员会

COMPUTER  
VISION  
NEWSLETTER

# CCCF 计算机视觉 专委会简报

02 2020

总第 24 期



CCF 计算机视觉  
专委会

# COMPUTER VISION NEWSLETTER



## 计算机视觉专委会 简报

2020年第02期  
总第24期



CCF 计算机视觉  
专委会

### 主 办 编委会

CCF 计算机视觉专业委员会

#### /专委动态/

荣誉主编	王 亮	中国科学院自动化研究所
主 编	马占宇	北京邮电大学
执行主编	李实英	上海科技大学
主 编	毋立芳	北京工业大学
编 委	黄 岩	中国科学院自动化研究所
	任传贤	中山大学
	杨巨峰	南开大学

#### /科技前沿/

主 编	王金甲	燕山大学
编 委	储 珺	南昌航空大学
	崔海楠	中国科学院自动化研究所
	任桐炜	南京大学

#### /委员风采/

主 编	余 焯	合肥工业大学
编 委	刘海波	哈尔滨工程大学
	赵振兵	华北电力大学

#### /学术资源/

主 编	李 策	兰州理工大学
编 委	樊 鑫	大连理工大学
	贾 同	东北大学
	沈沛意	西安电子科技大学

#### /海外学者/

主 编	金 鑫	北京电子科技学院
编 委	刘帅奇	河北大学
	张汗灵	湖南大学

# CONTENTS

## 简报目录

### | 专委动态

- 04 走进高校活动系列
- 05 2020 年度“CCF-CV 奖”评选活动正式启动
- 06 CCF-CV “持久影响力论文奖”评选启动
- 07 CCF-CV 走进高校系列报告会申请开放
- 08 CCF-CV 2020 年委员增选申请开始
- 09 缅怀计算机视觉泰斗 Thomas S. Huang 教授系列
- 15 CCF-CV 专委会成立国际顾问委员会

### | 科技前沿

- 16 视觉关系检测研究进展
- 20 基于稀疏量化的神经网络加速与压缩
- 22 TCNN: 利用轨迹片段进行视频目标检测
- 23 ICLR 2020

### | 委员风采

- 26 中国海洋大学董军宇教授访谈
- 29 委员好消息

### | 学术资源

- 30 医学图像配准开源代码
- 32 手势识别数据集
- 36 好文推荐

### | 海外学者

- 39 英国萨里大学宋一哲教授团队
- 44 征文通知

CCF 计算机视觉  
专委会

 [CCFCV.CCF.ORG.CN](http://CCFCV.CCF.ORG.CN)

 [CCFCVN@GMail.com](mailto:CCFCVN@GMail.com)

## CCF-CV 走进高校系列报告会

### 第 86 期 山东大学



2020年5月10日上午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，山东大学-山东省人工智能研究院联合承办，微软亚洲研究院协办的第86期CCF-CV走进高校系列报告会——“智能媒体计算”以云直播的方式成功举行。本期报告会特邀清华大学季向阳教授、中山大学赖剑煌教授、中国科学院计算所山世光研究员、西安电子科技大学石光明教授、微软亚洲研究院王井东首席研究员、西安交通大学薛建儒教授6位专家学者做报告。由山东大学聂礼强教授、甘甜副教授、吴建龙助理教授担任本次报告会的执行主席。

此次CCF-CV走进高校系列报告会持续了四个多小时，六位专家既生动展示了当前计算机视觉的前沿技术，又为与会观众开展研究指明了研究方向。

报告首次采用“微软AI讲堂”云直播的方式，哔哩哔哩平台人气峰值接近1.7万！观众通过线上弹幕积极进行互动、交流和提问，专家也给出了详细的解答，极大地鼓舞了大家的科研热情。报告会在热烈的气氛中圆满结束。

### 第 87 期 宁夏大学



2020年5月31日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，宁夏大学--宁夏大数据与人工智能省部共建协同创新中心联合承办，宁夏通信学会、宁夏电子学会协办的第87期CCF-CV走进高校系列报告会——“大数据智能计算”云直播活动圆满结束。本期报告会特邀中国科学院自动化研究所王亮研究员、北京大学田永鸿教授，中山大学郑伟诗教授，中国科学院计算技术研究所王瑞平研究员以及西安电子科技大学王楠楠教授5位专家学者做报告。由宁夏大学信息工程学院高玉琢教授、孙学宏教授和刘昊副教授担任本次报告会的执行主席。

在圆桌讨论环节，专家针对主持人提出的议题展开了热烈讨论。最后，由宁夏大学信息工程学院院长高玉琢教授进行总结。高院长再次表达对专家学者的衷心感谢。希望能以本次报告为契机，加强区内外相关领域的合作和交流，促进宁夏在人工智能领域的发展，吸引更多的人才来宁工作，共同推动宁夏新旧动能的快速转换。

责任编辑 毋立芳

## 2020 年度 "CCF-CV 奖" 评选活动正式启动

**为**发挥中国计算机学会计算机视觉专委会（简称 CCF-CV）的学术评价作用，激励、调动计算机视觉及相关领域专业工作者的积极性，促进计算机视觉及相关领域的创新和进步，自 2017 年起，CCF-CV 设立中国计算机学会计算机视觉专委会奖（简称 CCF-CV 奖），旨在奖励计算机视觉及相关领域中，为领域发展做出突出贡献的老前辈、取得公认成就的科研人员、意义重大的科技成果、影响深远的学术成绩，具备潜力的学术新锐，以及为专委会工作提供卓越服务的优秀个人。

### 评选要求

#### 一、候选人提名

CCF-CV 奖项实行委员推荐制，一名以上 CCF-CV 专委会委员即可（联合）提名推荐。每个奖项的评选标准等事宜如下：

##### 1. CCF-CV 终身学术贡献奖

授予在计算机视觉及相关领域做出了突出贡献的中国计算机科技工作者（年龄在 70 岁以上）。每年评选一次，获奖人数不超过二名。获奖者可获得奖励证书、奖杯及奖金。

##### 2. CCF-CV 杰出成就奖

最高标准评选出计算机视觉领域取得了公认的杰出成就的计算机科技工作者，其获奖主要工作应在中国境内完成，国籍不限。每年评选一次，获奖人数不超过一名。获奖者可获得奖励证书、奖杯及奖金。

##### 3. CCF-CV “中科视拓 Seeta 服务贡献奖”

专委会活动、会议组织、秘书处工作、工委小组工作有重要贡献者。每年评选一次，获奖人数不超过三名。获奖者可获得奖牌、奖励证书和奖金。

##### 4. CCF-CV “中科视拓 Seeta 学术新锐奖”

在计算机视觉及相关领域从事相关研究的研究生（含硕士研究生与博士研究生），尤其是具备一定科研潜力的低年级博士生；所有申请参评人员需要提供一项代表性工作。每年评选一次，获奖人数不超过三名。获奖者可获得奖励证书、奖杯及奖金。

##### 5. CCF-CV 持久影响力论文奖

鼓励原创性、基础性并能够产生持久影响力的创新性研究，其获奖论文应为 2010 年 12 月 31 日以前在计算机视觉主流会议和期刊上正式发表的论文，且第一单位为中华人民共和国境内（含港澳台地区）的学术机构。每年评选一次，获奖论文不超过一篇。获奖论文的作者可获得奖励证书、奖杯及奖金。获奖论文的作者需在当年 PRCV 或 CCF-CV 年度工作会议上以公开学术报告的形式介绍其工作。

#### 二、推荐方法

由推荐人联合填写推荐表 1 份，将其电子版扫描件于截止日期之前提交至电子邮箱 [ccfcvaward@sina.com](mailto:ccfcvaward@sina.com)。

### 评奖流程

#### 一、评奖办法

在 CCF-CV 常委会领导下，由 CCF-CV 提名与奖励

工作组负责评奖工作, 具体包括: 工作组负责征集提名、委员会参与评审)。所有奖项候选名单提交常委会表决, 过半数通过。

初筛与资格审查 (“持久影响力论文奖” 还将提名评审

### 三、其他未尽事宜

详见《中国计算机学会计算机视觉专委会奖励条例》。

### 二、报名截止时间

即日起—2020 年 7 月 15 日。

责任编辑 任传贤

## CCF-CV “持久影响力论文奖” 评选启动

为推动中国计算机视觉领域的科技进步, 鼓励原创性、基础性并能够产生持久影响力的创新性研究, 特设立中国计算机学会计算机视觉专委会持久影响力论文奖 (以下简称 “持久影响力论文奖”)。本奖项每年评选一次, 获奖论文不超过一篇, 由 CCF-CV 评选和颁奖。欢迎计算视觉专委会委员积极提名推荐!

### 冠名征集

为了提升 “持久影响力论文奖” 的影响力, 推动企业与学术界合作, 助力企业扩大在国内计算机视觉领域的影响力, 现诚征 “持久影响力论文奖” 的冠名单位。有意向者请与 CCF-CV 提名与奖励工作组秘书任传贤 (rchuanx@mail.sysu.edu.cn) 联系。

### 参选要求

#### 一、如何成为正式候选论文

参加 “持久影响力论文奖” 评选应具备如下条件:

1. 候选论文应为 2010 年 12 月 31 日以前在计算机视觉主流会议和期刊上正式发表的论文。
2. 候选论文第一单位为中华人民共和国境内 (含港澳台地区) 的学术机构。

3. 候选论文需由一名 CCF-CV 专委会委员推荐, 方可成为正式候选论文。

#### 二、推荐方法

1. 拟提名推荐的委员需填写《CCF-CV 持久影响力论文奖推荐表》及其电子版, 提交至电子邮箱 ccfcvaward@sina.com。

2. 每位 CCF-CV 专委会委员同一年最多推荐一篇候选论文。

### 评奖流程

#### 一、评奖办法

在 CCF-CV 常委会领导下, 由 CCF-CV 提名与奖励工作组负责评奖工作。评奖过程分初审和复审两个阶段:

1. 初审阶段由 CCF-CV 提名与奖励工作组负责接受提名、初筛和资格审查, 推荐正式候选论文与评审委员会名单, 并提交 CCF-CV 常委会审议通过。
2. 复审阶段由评审委员会负责组织, 根据正式候选论文情况, 最终推荐不超过 1 篇 CCF-CV 持久影响力论文, 提交 CCF-CV 常委会审议通过。

3. CCF-CV 提名与奖励工作组将邀请获奖论文作者在

当年 PRCV 大会或专委会年度会议以公开学术报告的形式介绍其工作。论文作者(一人)免 PRCV 大会注册费。

4.CCF-CV 在专委会年度会议上对获奖论文作者颁发获奖证书与奖金。

## 二、报名截止时间

即日起—2020 年 7 月 15 日。

## 三、其他

提名与奖励工作小组拥有最终解释权，其他未尽事宜详见《CCF-CV 持久影响力论文奖评选条例》。

责任编辑 任传贤

# CCF-CV 走进高校系列报告会申请开放中



为了更好地推动计算机视觉领域的学术与技术交流，促进国内外学者间的了解与合作，全面推动国内计算机视觉的学科发展，提升我国计算机视觉研究在国际领域的影响力，中国计算机学会计算机视觉专委会在全国范围的高校和科研院所等率先开展了 CCF-CV 走进高校系列报告会活动。

自 2015 年 11 月开始以来，本活动得到了高校、讲者、听众的大力支持。截至 2020 年 1 月，CCF-CV 走进高校系列报告会活动已成功举办了 85 期，遍及祖国大陆所有省市自治区直辖市，并在澳门、悉尼成功举办。目前，活动已邀请讲者做分享专题报告 320 余场，活动现场人均听众 200 人次，微信公众号平均阅读 1000 余次。在征得讲者同意的前提下，于爱奇艺视频网站上传共享报告会的现场视频，并于专委会网站分享讲者的报告 PPT，在国内外计算机视觉领域引起了强烈反响。在这里，向对活动顺利开展提供支持和帮助的承

办方以及分享精彩报告的讲者们表示由衷的感谢！

为了应对疫情，第 86 期活动于 2020 年 5 月 10 日线上举办，活动在哔哩哔哩平台直播，每场报告听众人数均超过 1.1 万，人气峰值接近 1.7 万。新模式、新思路、新影响，效果不错！特开放 CCF-CV 走进高校活动征集！如您愿意尝试组织线上活动，欢迎申请！

在新一届专委会领导班子的指导下，在 CCF-CV 全体委员的帮助下，在对活动进行全面总结的基础上，我们对活动内容、活动申请、讲者邀请等进行了完善，形成了 CCF-CV 走进高校系列报告会活动细则（2020.02）。

现继续面向全国的高等院校、科研院所等机构征集 2020 年度活动申请，希望能够进一步贯彻落实《新一代人工智能发展规划》纲要中有关计算机视觉领域发展的精神，继续推动计算机视觉领域的学术交流、思想碰撞，为广大师生提供更多近距离接触行业领域专家、互动沟通交流的机会。

联系方式：

毋立芳：lfwu@bjut.edu.cn

杨巨峰：yangjufeng@nankai.edu.cn

责任编辑 杨巨峰

## CCF-CV2020 年委员增选申请开始



申请时间：2020 年 6 月 6 日—2020 年 10 月 1 日

申请流程：填写申请表，发送给秘书处 (ccfcv@139.com)，主题“2020 新委员申请-姓名-单位”。(注：推荐人必须是现任专委委员，名单可以从专委网站查询。电子版申请表中需填写推荐人姓名和意见，委员增选成功后可以补签签名)。

申请资格：任职国内外学术界或企业界副教授或同等级别以上的人员，拥有计算机视觉相关领域的高水平研究成果，是 CCF 会员，且积极参加计算机学会计算机视觉专委会的各项活动。特别优秀的讲师、企业人士亦可考虑。

申请需知：现任专委委员每人可推荐最多 3 名候选人，主任委员、副主任委员、秘书长、常务委员每人可推荐最多 5 名候选人。本次申请结果将在“2020 年中国模式识别与计算机视觉大会”期间举行的专委工作年会上投票确定（申请者届时必须“注册参会”）。

责任编辑 毋立芳

自 2013 年 10 月成立以来，中国计算机学会 (CCF) 计算机视觉专业委员会 (ccfcv.ccf.org.cn) 发展迅速，举办了很多有影响力的活动，如计算机视觉前沿进展研讨会 (RACV)、CCF-CV 走进高校系列活动、CCF-CV 走进企业系列活动、CCF-CV 视界无限系列活动，与中国自动化学会模式识别与机器智能专委会、中国图像图形学学会视觉大数据专委会、中国人工智能学会模式识别专委会共同举办中国模式识别与计算机视觉大会 (PRCV)，定期出版专委简报，建设专委中英文网站。搭建了全方位、高水平、大规模的计算机视觉领域交流平台。专委会成立六年以来，已经发展委员 321 人，在 CCF 专委评估中获得“特色活动奖”、“综合进步奖”、“优秀专委奖”、“年度特别奖”等 5 个奖项。为了保持专委会的活力、促进国内外视觉领域人员的交流和合作，专委会现开放 2020 年计算机视觉专委会的委员增选工作。

## 缅怀计算机视觉泰斗 Thomas S. Huang 教授



计算机视觉与模式识别领域国际著名学者、美国伊利诺伊大学香槟分校 Thomas S. Huang (黄煦涛) 教授在美国东部时间 2020 年 4 月 25 日夜间去世，享年 84 岁。

黄教授 1956 年毕业于台湾大学，1963 年获得美国麻省理工学院博士学位后留校任教，先后担任助理教授和副教授。1973 年担任普渡大学教授兼信息处理实验室主任。1980 年加入伊利诺伊大学香槟分校，担任 William L. Everitt 杰出讲座教授。他是美国工程院院士，中国科学院和中国工程院双外籍院士。他曾荣获国际信号处理领域的最高荣誉 IEEE Jack Kilby 奖，模式识别领域的最高荣誉 King-Sun Fu 奖和计算机视觉领域的 Honda 终身成就奖。

黄教授主要从事信息和信号处理方面的研究工作，发明了预测差分量化 (PDQ) 的两维传真 (文档) 压缩方法，该方法已发展为国际 G3/G4FAX 压缩标准；在多维数字信号处理领域中，提出了关于递归滤波器的稳定性的理论；建立了从二维图像序列中估计三维运动的公式，为图像处理和计算机视觉开启了新领域。

黄教授十分关注中国科学技术事业的发展。先后访问过我国多所大学和研究所作学术报告，积极展开学术交流，并担任顾问和名誉教授等职务。他从教 50 余年，直接培养的研究生过百人，无数在计算机视觉与模式识别领域颇有建树的后辈都曾受过他的点拨。



黄教授的不幸逝世是学术界的巨大损失，我们深切缅怀他！

责任编辑 黄岩

## 怀念我的导师 Thomas S. Huang 教授



### 引言

计算机视觉与模式识别领域国际著名学者、美国伊利诺伊大学香槟分校 Thomas S. Huang (黄煦涛) 教授在美国东部时间 2020 年 4 月 25 日夜间去世, 享年 84 岁, 他的不幸逝世是学术界的巨大损失! 我们联系了一些曾与 Thomas S. Huang 教授有过交集的国内学者, 整理一些纪念性的文字以寄托对教授的哀思。本次我们联系的是 Thomas S. Huang 教授的博士生、上海交通大学的刘允才教授。

### 1 绅士风度, 睿智幽默

庚子年乃不幸之岁, 教授夫人 Margaret 和恩师 Thomas S. Huang 相继辞世。计算机视觉专委会邀我写篇文章悼念先生, 而我此刻肝肠寸断, 心绪如麻, 不能成章, 谨呈往日跟随先生的一些片段, 以寄哀思。

我于 1984 年 12 月中旬来到 UIUC。一周之后的 12 月 24 日, 黄教授邀我参加他的家庭圣诞节晚宴。一个从学校毕业不久涉世未深的青年去国际学术大师家中做客, 内心诚惶诚恐。傍晚六点左右, 下着雪, 先生带着他的小儿子开车到我的寓所接我, 使我非常感动。那是我有生以来第一次过圣诞节, 也是见到先生家的成

员最多的一次, 有教授夫人 Margaret、两个儿子、两个女儿及她们的男朋友。温馨儒雅的晚宴气氛和先生的绅士风度给我留下深刻的印象。餐后, 先生开车与教授夫人一同送我回寓所, 途中还绕道去 Champaign 观看城市圣诞夜景。对着地上的白雪和烛光, 教授夫人对我说, white Christmas 是一个好征兆。自那以后在美国学习期间, 我和家人的圣诞节大多都是在先生家快乐地度过。先生的热情和平易近人作风、文雅的绅士风度是我一生学习的楷模。

先生谦和低调, 睿智幽默, 差旅中从来不让学生帮助提包。进入大厅时总是让女士先行, 哪怕这位女士是年轻的晚辈。许多深邃的理念, 在教授那里, 常是简单的语言和幽默表达, 令人豁然开朗, 如沐春风。我记得在一篇论文的作者简介中先生这样描写自己, “爱收集书籍 (不一定全读), 爱收集音乐 (不一定全听)”, 幽默谦虚。还有一次实验室的会上, 见先生穿了一件蓝色 T-恤, 前胸印着白色大字 “Digital”, 后背印着 “Analog”。那是他刚从课堂上回到实验室。他说, “这学期讲两门课, 分别是模拟信号处理和数字信号处理。为了使学生不致混淆, 讲哪门课时就将哪个字穿在胸前”。

### 2 学术先驱, 追求真知

作为计算机视觉的学术先驱, 先生有着辉煌的研究成果。他不仅有深远的学术洞察力和严谨的科研态度, 而且有坦荡开放的学术胸襟。我读博士时的研究课题是基于二维线元的三维刚体运动分析。在 1985 年, 这个课题还是一个原创性的研究题目, 当时除了我们的视觉实验室外, Texas 大学的 J. K. Aggarwal 教授也在攻关

这一课题。作为学者，每个人都希望将自己的研究成果在第一时间最先发表。在我们取得了研究结果准备发表论文之际，先生将 Aggarwal 教授请到 UIUC，在实验室的会议上我和 Aggarwal 教授分别介绍自己的数学推导和实验结果。最终，我们的研究成果和 Aggarwal 教授的成果同时发表在 Workshop on Motion (1986)会议上，随后 ICPR' 86，展示了不同的研究思路和相同的科学结论。先生这种追求真知的科学精神成为了我后来学术研究的指路航标。

在我的工作生涯中，一直没有离开先生的热心关怀。博士毕业后，我作为博士后继续跟随先生两年，随后是先生亲自将我推荐给日本住友电气公司的高层主管到公司工作。2000年，在先生亲切关怀下，我有幸到上海交通大学任特聘教授。自此至2014年，先生多次应邀来上海交通大学讲学与学术指导。

### 3 伉俪相伴，相濡以沫

先生和夫人一生伉俪相伴，相濡以沫，不管先生到哪里开会访问，夫人总是陪伴在侧。记得2002年先生应聘上海交通大学顾问教授，发言后回到座位时，夫人给了先生一个深情拥抱。这祝贺，是激情，是爱。会场中的师生无不羡慕。先生原籍上海，夫人出生于杭州，对上海有着深厚的感情，他们来上海时喜欢入住外滩的和平饭店，品尝上海的地方菜肴。有一次，上海交通大学图像处理与模式识别研究所所长施鹏飞教授和夫人在一家优雅的餐馆招待黄教授夫妇，我与爱人陪同，大家请先生点菜。先生点的菜肴精美而不奢靡，绝无鲍鱼海参燕窝之类；量轻而不致馁，让人餐后尚有回顾之念。先生和夫人用餐优雅，较大块的菜肴必定用餐刀切小、面包用手撕成小块才放入口中，举手投足间流露出高尚的绅士风范。教授夫人热爱生活，喜爱上海旗袍，每次来到上海都约我的爱人陪同去南京路上的丝绸大王绸店庄挑选衣料定制服装，随挑选随讲述服装在不同年代的缝制款式。我们总被教授夫人深厚的中华文化底蕴所深深感染。

晚年，由于身体的原因，先生夫妇住进了护老院。微软沈向洋先生是这样描述的：“一个清晨，在美国中部伊利诺伊州香槟镇护老院里的走廊上，老先生在老太太

身边走着。通常，老太太坐在轮椅上被老先生推着。但今天，老太太搀着她的双扶手助步器，缓慢而坚定地走了起来。一位院友在他们经过的那一刻，高高举起她紧扣的双手，做出胜利的姿势。「真是太好了！」，他兴奋地对我说，「你可以走起来了！」午饭后，他们俩回到房间。原打算要小憩片刻，但休息前，他们情不自禁跳起舞来，尽管老太太身体的平衡还不是太好。一、二、三、四，一、二、三、四，他们一边数着，一边一步一步挪动着身体，走起了四方步。跳着跳着，他们搂着的身体逐渐拥抱起来。拥抱着，他们深情地吻了起来”。



### 4 巨星陨落，承继遗志

Thomas S. Huang 教授在数字信号处理、计算机视觉、人机交互和多媒体领域做出了举世公认的奠基性贡献，是这些领域的拓荒者和一代宗师。在先生的执教生涯中，培养的博士和博士后120多人，分布世界各地，大都成为教育界与工业界的栋梁。他著有理论书籍23部，学术论文上千篇，是世界上最高贡献的计算机科学家之一。他与夫人相继离开我们，是我们学术界的重大损失，巨星陨落，带走了一个时代。真诚希望我们这些后继者在专委会的组织下，继承先生的事业，对计算机视觉学术领域做出新的贡献。

愿恩师 Thomas S. Huang 和夫人 Margaret 天堂相伴安息。

刘允才

2020年5月4日于上海

责任编辑 黄岩

## 学识传世，至人如常，先生之风，山高水长



### 引言

计算机视觉与模式识别领域国际著名学者、美国伊利诺伊大学香槟分校 Thomas S. Huang (黄煦涛) 教授在美国东部时间 2020 年 4 月 25 日夜間去世，享年 84 岁，他的不幸逝世是学术界的巨大损失！我们联系了一些曾与 Thomas S. Huang 教授有过交集的国内学者，整理一些纪念性的文字以寄托对教授的哀思。本次我们联系的是 Thomas S. Huang 教授的博士生、文远知行 (WeRide) 执行总监徐迅 (Jason Xu) 博士。

2020 庚子年，从这个世界带走了太多。4 月下旬的一个下午，和部分同门前辈、师兄弟与病榻上的先生短暂视频。次日一早，收到了他驾鹤仙去的消息，心情异常沉重。短短三个月前，师母 Margaret 刚刚离开了我们，让人不愿相信这一切竟发生得这样突然。对于学术界，Tom 是图像处理和计算机视觉等领域众人景仰的宗师，是中美学术交流的积极倡导者和引路人之一。对于我们这百余名弟子，他则是恩师和家。回想起和 Tom 相处的点滴，很多印象浮现在眼前：他为学的执着勤勉，为人的谦和诚挚；他思想的睿智深刻，语言的简练幽默；他对名利的澹泊，对生活的热爱，对家庭的珍视。这一切如春风化雨，感染着他身边所有学生、同事和朋友。

记得和 Tom 的初见，是在 2002 年。彼时我在微软亚洲研究院 (MSRA) 多通道用户界面组实习，正遇上 Tom 在研究院做访问。说是初见，其实我只看到他在办公室中的背影，同事告诉我说，这就是世界知名的 Thomas Huang 教授。满头银发，全神贯注于工作中的 Tom，让我顿时心生景仰。过了一段时间，Tom 恰好又来清华计算机系访问，我的硕士导师张长水教授和我前去拜访，介绍了我们当时的一些工作。当时没有想到，一年之后我会成为他的博士研究生。回想跟随他学习工作的那些年，实是人生大幸，我想 Tom 所有的学生，也无一不有同感。

2003 年来到 UIUC 之后，上的第一门研究生专业课，就是 Tom 的数字图像处理。Tom 讲课风格鲜明，充实深刻而轻快幽默。在课上除了听他举重若轻地将各种理论娓娓道来，还可以听到他亲历的轶事趣闻，诸如香农、维纳和“Lena”。我在课上还得了个“大”奖：一块硕大的曲奇饼。除了上课之外，就是在 IFP (Tom 在 UIUC 创立的研究组) 里从事课题研究。Tom 一生追求真知，涉猎的学术方向众多，而在各个阶段又有不同重点。我在 IFP 的那几年，研究组的主题是多媒体内容分析和人机交互。在这个主题指引下，Tom 给了我们很大的自由去探索人脸、手势、动作、语音识别等各个方向。我当时在人脸课题组，从事的一项人脸分析工作引起了媒体关注，被 National Geographic (国家地理) 频道拍摄进纪录片。之后学校新闻办来采访，在我桌前为 Tom 拍了这张照片，屏幕上显示的 demo 还是我当时写的。这张照片承载着和 Tom 工作那些年的珍贵回忆，于我有特别的意义。



Tom 为人极平易谦和，却律己甚严。当时他常年旅行不断，在世界各地访问和讲学，但令人惊讶的是，时差和疲惫似乎对于他是不存在的。无论是刚从亚洲还是欧洲回到学校，Tom 都会准时出现在下一场例会或项目会上，饶有兴味地听汇报，敏锐中肯地提问题。相比之下，倒是我们这些学生在会上显得有些不够积极。记得有一次会上，Tom 看着我们半开玩笑地说：你们应该再努力点，看我还在记笔记呢。他早已功成名就、著述等身，尚且勤勉如此，我们有什么理由，不再努力一点呢？直至今日，我还时常想起当时的情景，感受到激励和鞭策。

不知从何时起，学术导师开始被冠以“老板”的称呼。我想，Tom 的学生没有人会有这样的印象。他总是谦虚地把贡献和荣耀归于他的学生。而且，对于我们这些弟子而言他不仅仅是老师。毫不夸张地说，Tom 和 Margaret 在 Urbana-Champaign 给了我们一个温暖的家。他们二人伉俪情深，非常重视家庭。同时真诚地希望学生们也能如此。对于学生们订婚结婚、添丁增口等喜事，他们总是发自内心地高兴和祝福。我和太太在学校筹办婚礼的时候邀请他们参加，Tom 回复说那时他们在国外访问，将在我们婚礼当天下午才飞回 Champaign，不一定能赶上。当婚礼那天，看到 Tom 和 Margaret 身着正装，手持礼物出现在婚礼会场的时候，我们心里的惊喜和感激之情，无以言表。Tom 在婚礼上的致辞，一如以往地儒雅幽默，丝毫不见旅途劳顿。后来才得知，先生和师母是刚下飞机，在机场换上正装

就直接赶来参加婚礼，为的就是在第一时间给我们以见证和祝福。Tom 和 Margaret 一生对学生和晚辈的关爱提携不可胜数，这在其中或许只是一件小事，却足以让我们长久怀念，心存温暖和感激。

我博士毕业之后离开了伊利诺伊，和 Tom 除了在学术会议上短暂相会之外，还有两次较大规模的 IFP “家庭”聚会，一次是 2013 年 Tom 和 Margaret 访问旧金山湾区，另一次是 2016 年为庆贺 Tom 八十岁生日在 UIUC 举办的研讨会 Huang Symposium。2017 年因为工作的缘故，有幸回 Champaign 一次，和恩师、师母和一众年轻师弟们共进晚餐。我从加州带去一瓶不错的 Cabernet，Tom 是红酒品鉴家，很高兴地与大家一起分享，举杯细品，相谈甚欢。他对人工智能的各种工业应用仍然兴趣浓厚，思维活跃，问了不少问题。散席后，一如往常，Tom 和 Margaret 牵着手缓缓走出饭店。望着他们的背影，我不禁感动泪目。只是当时怎么也想不到，那竟是最后一次陪他们用餐。如今回首，空留遗憾，真应该多回学校几次看看他们……

师母 Margaret 出身名门，优雅而睿智，慈爱而干练。在一次和他们同去芝加哥访问的路上，谈及人生，Margaret 说了一句非常短却意味深长的话：“人生很短暂，have fun”。我知道她当然不是教我们及时行乐，而是因为人生短暂，需要努力让这个过程更加丰富而有意义。Tom 和 Margaret 携手度过了令人羡慕和尊敬的一生，为世界留下了宝贵的精神遗产。不仅如此，他们还提携、关怀了包括我在内的许许多多的人，使我们能达到更好的生命状态。先生和师母的恩情、慈爱与教诲，我们将永铭于心，唯有努力工作，认真生活，关爱他人，回报世界方能不负。

Tom and Margaret, we miss you, may you rest in eternal peace.

Jason Xu 徐迅

2020 年 5 月于加州

责任编辑 黄岩

## 忆往昔，历历在目，思先生，感激敬佩



### 引言

**计**算机视觉与模式识别领域国际著名学者、美国伊利诺伊大学香槟分校 Thomas S. Huang (黄煦涛) 教授在美国东部时间 2020 年 4 月 25 日夜间去世，享年 84 岁，他的不幸逝世是学术界的巨大损失！我们联系了一些曾与 Thomas S. Huang 教授有过交集的国内学者，整理一些纪念性的文字以寄托对教授的哀思。本次我们联系的是曾经在 Thomas S. Huang 教授团队工作过的、南京理工大学杨静宇教授。

看到我尊敬的老师和朋友 Tom 去世的讣告，犹如晴天霹雳，悲痛万分。我 1982 年至 1984 年曾有幸在伊利诺伊大学 (UIUC) 的 CSL 实验室工作、学习，成为 Tom 领导的研究 Team 的一员。我至今还清晰地记得和 Tom 的第一次见面。1982 年 2 月我乘飞机从上海到达美国西海岸的旧金山，由于航班晚点，原订接续的航班已经起飞，只好临时改签了一张飞往 Peoria 的机票。第一次出国遇到这样的情况，人生地不熟，英语语言表达能力也很有限，真是不知所措。情急之下，我请机场服务人员给 Tom 家里打了一个对方付款的电话，告知情况。Tom 接到我的电话后，要我不要着急，告诉他我的航班号，他会来接我。飞机到达 Peoria 机场已经

接近深夜。走出机场，我认出了 Tom 和一位女士站在出口处。经 Tom 介绍我认识了他的夫人 Margaret。Tom 亲自开车，行驶了 150 多公里，到达我在 Urbana 预订的住处，此时已是第二天的凌晨。每当想起此事，我心中充满了感激，同时也深感内疚和不安，占用了工作异常繁忙的先生很多宝贵的休息时间。



Tom 学业精深，在学术界声望极高，但他为人谦和，平时言语不多。有一次和 UIUC 计算机系 C. L. Liu (刘炯朗) 教授一起聊天，他和 Tom 是老朋友，一起共事多年。他告诉我，Tom 是 1956 年台湾 10 万高考学生的状元，为人“虚怀若谷，大智若愚”。至今回想起来，都觉得刘教授对 Tom 的评价非常值得我们这些在他身边工作过的人体会。

Tom 工作十分繁忙，他每个学期都要上课。不仅给研究生上课，还给本科生上大课。我听过 Tom 讲授的一门有关图像分析的研究生课程。他对讲授的内容非常熟悉、精通，有时会手拿一杯咖啡，在讲台上边踱步边讲授，如数家珍。记得有一次讲授图像重构的内容，在

讲到 CT 投影重构的原理和方法时，他很兴奋地提到他

采用 Fourier 分析的原理也提出了一种图像重构的方法，提出的时间和获得 1979 年诺贝尔医学奖的英国学者 Hounsfield 在同一时间段。讲到这，他开玩笑地说，他和诺贝尔奖是擦肩而过。

Tom 携夫人 1985 年首次来中国浙江大学讲学，期间我邀请他来我所在的华东工学院（即现在的南京理工大学）作学术报告，先生愉快地接受了邀请来到南京。Tom 报告的内容是有关立体视觉的内容，十分生动、精彩，吸引了南京各高校的近百位师生。Margaret 十分怀念她读中学的苏州，我们从他们下榻的金陵饭店包了一辆车，我陪同 Tom 夫妻一路开往苏州寻访。可惜，因时间太久，她读书的中学已不在原地。

Tom 参加各种学术活动比较频繁，我有机会多次在国内外见到他。有一年，他来北京访问。结束后，教育部副部长韦钰陪同他访问东南大学，Tom 想到了在南京工作的我。他请东南大学一位领导和我联系，问我是否有时间见一面。得知先生来到南京，我立即赶到东南大学，参加了学校为 Tom 举办的欢迎晚宴，有机会见到先生。没想到，这是和先生的最后一次见面。

Tom 和 Margaret 已经离开了我们，但他们的音容笑貌将永远留在我的记忆中。

杨静宇

2020 年 5 月于南京

责任编辑 黄岩

## CCF-CV 专委会成立国际顾问委员会

为了推进专委会的国际化建设，促进海内外计算机视觉领域的交流合作，CCF-CV 专委会成立“国际顾问委员会”。专委会将邀请国际顾问委员会委员参加专委会的学术活动，定期向顾问委员汇报专委会的发展情况，并就专委会开展的重要活动进行咨询等。

经过提名并报常委会审议，日前通过了国际顾问委员会名单如下（按姓氏拼音排序）：

龚邵刚 教授 Queen Mary University of London

华刚 博士 Wormpex AI Research

罗杰波 教授 University of Rochester

马毅 教授 University of California, Berkeley

沈春华 教授 University of Adelaide

陶大程 教授 The University of Sydney

张正友 博士 Tencent Robotics X and AI Labs

朱松纯 教授 University of California, Los Angeles

责任编辑 黄岩

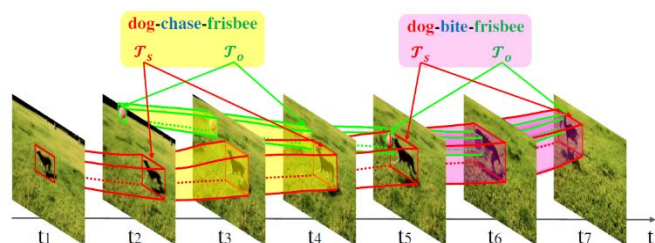
专题综述

## 视觉关系检测研究进展

南京大学 任桐炜 孙旭 于凡 武港山

## 一、引言

视觉关系检测旨在以<主语, 谓语, 宾语>形式的三元组来表示图像和视频中的物体及其交互关系<sup>[1]</sup>。图 1 和图 2 分别展示了面向图像和面向视频的视觉关系检测示例。对于给定的源图像, 视觉关系检测方法会生成一定数量的三元组, 以及每个主语、宾语所对应的物体的边界框(示例中用相同颜色标识主语/宾语和其对应物体边界框); 对于给定的源视频, 视觉关系检测方法则需要为每个视觉关系三元组中的主语和宾语生成对应的物体边界框轨迹。

图 1 面向图像的视觉关系检测示例<sup>[1]</sup>图 2 面向视频的视觉关系检测示例<sup>[2]</sup>

相比于物体检测、动作识别等技术, 视觉关系检测提供了更加丰富的图像和视频语义描述, 更好地连接视觉内容和自然语言描述, 为图像和视频的语义结构化提

供了新的探索途径, 能够为内容检索、描述生成、视觉问答等任务提供更加有力的支撑, 因而在近年来广受研究者的关注。

作为将计算机视觉与自然语言相结合的复杂语义理解任务, 视觉关系检测有着独特的挑战。首先, 视觉关系的类别由主语、谓语、宾语三者的类别组合而成。假设主语和宾语的类别空间为  $O$ , 谓语的类别空间为  $P$ , 理论上视觉关系的类别空间规模为  $|O|^2|P|$ , 因此视觉关系的类别数量会随物体和谓语类别数量的增加呈爆炸式增长。其次, 海量的视觉关系类别在现实世界数据中仅有一小部分会以较高的频率出现, 其余的大多数类别则只会在特定上下文出现, 这导致数据集的分布呈现严重的长尾现象。极不平衡的数据分布为视觉关系的分类带来了巨大的挑战, 尾部数据的稀疏分布则对模型的小样本学习能力提出了要求。此外, 少量视觉关系类别由于很罕见, 难以获得训练数据, 但这类关系在实际场景中仍有可能出现, 因此零样本学习也是视觉关系检测会遇到的关键问题。

## 二、面向图像的视觉关系检测研究现状

## 2.1. 代表性方法

视觉关系检测研究起源于面向图像的视觉关系检测。由于面向图像的物体检测已经形成了很好的技术积累, 早期的图像视觉关系检测大多采用独立的物体检测器, 得到图像中的物体类别和位置后, 再输入关系预测网络预测物体对之间的关系三元组。例如, Lu 等人<sup>[1]</sup>采用深度网络对每个物体对的联合区域提取特征; 并且考虑到视觉关系中主语、谓语和宾语在语义上具有很高的

关联性，引入了语言模型使得相近的对象有相近的关系分布。Zhang 等人<sup>[3]</sup>通过将对象的视觉特征映射到低维的关系空间中，然后用对象间的转移向量来表示对象之间的关系。

为了提升视觉关系预测效果，端到端的视觉关系预测模型被提出。这些模型将在预测视觉关系时利用物体检测中所提取的视觉特征和空间特征，而非仅使用检测结果。例如，Zheng 等人<sup>[4]</sup>采用多线索注意力机制，更多关注有意义的视觉关系，并结合视觉特征、空间特征和语言特征，从而让不同的视觉关系更好地利用关联度更高的特征。由于视觉关系可以构成图结构，部分研究者采用图模型来引入物体和关系信息作为上下文，使得视觉关系的预测更加准确和完善。例如，Hu 等人<sup>[5]</sup>在视觉特征、空间特征和语言特征的基础上，采用交互图对物体及其关系建模，并通过信息传递的方式引入上下文更新物体和关系的特征。

由于视觉关系数据集的标注工作量远大于物体检测等任务的数据集，很难为视觉关系检测任务构建出大规模并具有很好一致性、完备性的数据集。为此，研究者从不同方向开展努力。为了解决训练数据规模的问题，Peyre 等人<sup>[10]</sup>将弱监督学习引入视觉关系检测，利用对整幅图像标注的视觉关系三元组即可作为训练数据，降低了训练数据标注的负担。同时，为了解决数据集中标注完备性、一致性较低的问题，Sun 等人<sup>[11]</sup>提出了层次式视觉关系检测，鼓励在无法产生准确的关系预测时，产生语义更加抽象但置信度更高的三元组来提高预测结果的正确率，以解决现有视觉关系检测中由于仅关注召回率而引入过多错误结果的问题。

## 2.2. 数据集

目前，有 2 个代表性的面向图像的视觉关系检测数据集：VRD 数据集<sup>[1]</sup>和 VG 数据集<sup>[6]</sup>。其中，VRD 数据集包含了 5,000 张图像、100 类物体、70 类视觉关系谓语句和 37,993 个视觉关系。VG 数据集包含了 108,077 张图像、33,877 类物体、42,374 类视觉关系谓语句和 230 万个视觉关系。由于这两个数据集存在较为明显的分布不均衡问题，导致所训练的模型容易依赖数据偏向，更多地预测占比大的视觉关系谓语句。因此，Liang 等人基

于 VG 数据集提出了 VrR-VG 数据集<sup>[7]</sup>，包含了更有视觉相关性的视觉关系谓语句，并调整了每类关系的比例。此外，部分数据集专注于人和物体之间的交互关系，如 HICO-DET 数据集<sup>[8]</sup>和 HCVRD 数据集<sup>[9]</sup>。

## 三、面向视频的视觉关系检测研究现状

与面向图像的视觉关系检测相比，面向视频的视觉关系检测除了输入输出形式上的差异外，具有其自身的难点<sup>[2]</sup>。首先，面向视频的视觉关系检测使用边界框序列而非单个边界框来表示主语/宾语所对应的物体，需要准确检测出每个物体出现和消失的视频帧以及期间其在每帧上对应的边界框。其次，视频中包含了动态的视觉关系类型，如物体之间相对位置的变化、运动速度的相对快慢等，增加了视觉关系中谓语句预测的复杂性。最后，视频中同一对主语和宾语所对应物体之间的关系可能会随时间推移而变化，因此需要准确判定每个视觉关系实例所持续的时间范围。

### 3.1. 代表性方法

针对面向视频的视觉关系检测特点，研究者们提出一些有效解决方案。例如，Shang 等人<sup>[2]</sup>将完整视频切成固定长且有重叠的短视频片段，利用视频内容在短时间内能够保持相对稳定的特性，在每个视频片段上进行物体轨迹抽取和短期视觉关系识别，最后将从连续视频片段中检测到的短期视觉关系实例合并产生完整视频视觉关系实例。Tsai 等人<sup>[12]</sup>构建了基于条件随机场的门限时空能量图模型，同时从时间维度和空间维度进行关系推理。Qian 等人<sup>[13]</sup>将连续视频片段中检测到的物体连接为 3D 时空图，利用图卷积网络融合时间和空间上下文进行视觉关系识别，并采用孪生网络匹配合并相邻视频片段中检测到的短期视觉关系实例以提升性能。

### 3.2. 数据集

第一个面向视频的视觉关系检测数据集 ImageNet-VidVRD<sup>[2]</sup>基于 ILSVRC2016-VID 数据集构建，包含了 1,000 个视频，对其进行了稠密物体轨迹和视觉关系的标注，涉及 35 个物体类别、132 个谓语句类别、3,219 个视觉关系类别，平均每个视频包含 9.5 个视觉关系实例。

由于 ImageNet-VidVRD 数据集中的视频场景相对简单且视频长度较短, Shang 等人从社交媒体收集了 10,000 个用户生成、总时长 84 小时的日常视频, 对其进行了稠密物体轨迹和视觉关系的标注, 构建了更具普适意义的 VidOR 数据集<sup>[14]</sup>。VidOR 数据集涉及 80 个物体类别、50 个谓语类别、6,553 个视觉关系类别, 平均每个视频包含 38 个视觉关系实例。与 ImageNet-VidVRD 数据集相比, VidOR 数据集中的视频内容复杂多样, 存在较为明显的物体遮挡、镜头抖动等现象, 使得视觉关系检测更具挑战性也更贴近真实应用场景。

#### 四、技术挑战赛

现有的视觉关系检测相关的技术挑战赛主要有 2 个: 以人为中心的视觉关系竞赛(Person In Context, PIC) 和视频视觉关系理解竞赛 (Video Relation Understanding, VRU), 分别为面向图像和面向视频的视觉关系检测。

PIC 竞赛于 2018 年首次举办, 依托 ECCV 会议, 要求参赛者分割出图像中的人和物体并预测以人为主语的视觉关系, 由南京大学获得冠军。第 2 届 PIC 竞赛于 2019 年依托 ICCV 会议举办, 分为关系分割和人-物交互检测两个任务, 均由天津大学获得冠军。

VRU 竞赛于 2019 年首次举办, 依托 ACM MM 会议, 分为视频物体检测和视觉关系检测两个任务, 分别由深兰科技和南京大学取得第一名。第 2 届 VRU 竞赛仍然依托 ACM MM 会议, 目前正在进行中。

#### 五、未来发展方向

视觉关系检测为理解图像和视频语义提供了新的途径, 但目前的研究还较为初步, 仍有不少问题有待解决。例如, 图像和视频中包含大量的视觉关系, 但仅有一小部分能够反映图像和视频的内容主旨。研究如何有效检测出能够反映内容主旨的视觉关系, 有助于更好地支撑描述生成等上层应用。再如, 当前的视觉关系检测数据集在规模上与真实世界中的视觉关系存在着很大的差距, 所涉及到的物体类别、谓语类别数量依然十分有限, 使得已有的视觉关系检测模型不足以产生足够丰富且准确的视觉关系。构建面向真实场景的更大规模、更丰富类型的视觉关系检测数据集将带来更多全新的挑战, 包括更大容量的学习模型、多模态特征的融合、更大规模数据的收集、感知与逻辑推理的结合以及使用有限的数据进行知识迁移等。

责任编辑 崔海楠

#### 参考文献

- [1] Lu C, Krishna R, Bernstein M, et al. Visual Relationship Detection with Language Priors[C]. European Conference on Computer Vision. 2016: 852-869.
- [2] Shang X, Ren T, Guo J, et al. Video Visual Relation Detection[C]. ACM International Conference on Multimedia. 2017: 1300-1308.
- [3] Zhang H, Kyaw Z, Chang S F, et al. Visual Translation Embedding Network for Visual Relation Detection[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2017: 5532-5540.
- [4] Zheng S, Chen S, Jin Q. Visual Relation Detection with Multi-Level Attention[C]. ACM International Conference on Multimedia. 2019: 121-129.
- [5] Hu Y, Chen S, Chen X, et al. Neural Message Passing for Visual Relationship Detection[C]. ICML Workshop on Learning and Reasoning with Graph-Structured Representations, 2019.
- [6] Krishna R, Zhu Y, Groth O, et al. Visual Genome: Connecting Language and Vision Using Crowdsourced Dense Image Annotations[J]. International Journal of Computer Vision, 2017, 123(1): 32-73.
- [7] Liang Y, Bai Y, Zhang W, et al. VrR-VG: Refocusing Visually-Relevant Relationships[C]. IEEE International Conference on Computer Vision. 2019: 10403-10412.

- [8] Chao Y W, Liu Y, Liu X, et al. Learning to Detect Human-Object Interactions[C]. IEEE Winter Conference on Applications of Computer Vision. 2018: 381-389.
- [9] Zhuang B, Wu Q, Shen C, et al. HCVRD: a Benchmark for Large-scale Human-centered Visual Relationship Detection[C]. AAAI Conference on Artificial Intelligence. 2018: 7631-7638.
- [10] Peyre J, Sivic J, Laptev I, et al. Weakly-supervised learning of visual relations[C]. IEEE International Conference on Computer Vision. 2017.
- [11] Sun X, Zi Y, Ren T, et al. Hierarchical Visual Relationship Detection[C]. ACM International Conference on Multimedia. 2019: 94-102.
- [12] Tsai Y H, Divvala S, Morency L P, et al. Video Relationship Reasoning Using Gated Spatio-temporal Energy Graph[C]. IEEE Conference on Computer Vision and Pattern Recognition. 2019: 10424-10433.
- [13] Qian X, Zhuang Y, Li Y, et al. Video Relation Detection with Spatio-temporal Graph[C]. ACM International Conference on Multimedia. 2019: 84-93.
- [14] Shang X, Di D, Xiao J, et al. Annotating Objects and Relations in User-generated Videos[C]. ACM International Conference on Multimedia Retrieval. 2019: 279-287.



## 任桐炜

南京大学副教授。主要研究方向为视觉媒体分析。  
Email: rentw@nju.edu.cn



## 孙旭

南京大学硕士研究生。主要研究方向为视觉关系检测。  
Email: sunx@smail.nju.edu.cn



## 于凡

南京大学硕士研究生。主要研究方向为视觉关系检测。  
Email: yf@smail.nju.edu.cn



## 武港山

南京大学教授。主要研究方向为多媒体，计算机视觉，模式识别。  
Email: gswu@nju.edu.cn

热点追踪

## 基于稀疏量化的神经网络加速与压缩

中国科学院自动化研究所 王培松 程健

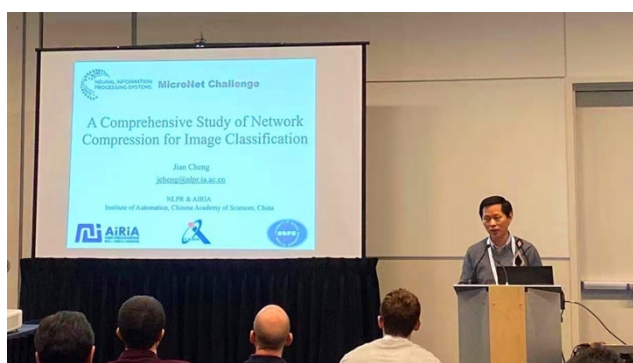


图1 本工作受邀在 NeurIPS 2019 进行展示

深度学习在越来越多的领域取得飞跃性进展，然而模型存储和计算复杂度却越来越高，严重妨碍了深度学习模型在实际场景中的大规模部署。因此，以模型压缩和加速为代表的深度学习计算优化技术成为最近几年学术界和工业界最为关注的焦点之一，对深度学习的进一步落地具有重要意义。

目前主流的神经网络加速与压缩方法包括网络剪枝、网络量化和知识蒸馏等。针对任意一种技术，均有大量的相关工作，例如网络剪枝，可以实现 90% 以上的稀疏度而不会对精度产生影响；例如网络量化，可以实现 2~4 比特无损压缩，甚至对于某些网络和任务可以实现 1 比特压缩。一方面，单纯依赖某一种压缩手段，都很难实现更高比率的加速和压缩；另一方面，这些不同的优化技术具有一定的互补性，如何综合运用这些技术手段，以达到更高压缩率和加速率，成为网络压缩领域的热点研究问题之一。

MicroNet Challenge 竞赛是由 Google、Facebook 和 OpenAI 等机构在 NeurIPS 2019 上共同

主办的，旨在通过优化神经网络架构和计算，达到模型高精度、计算效率和硬件资源占用等方面的平衡，实现软硬协同优化发展，启发新一代硬件架构设计和神经网络架构设计等。此次竞赛吸引了 MIT、加州大学、KAIST、华盛顿大学、京都大学、浙大和北航等国内外著名前沿科研院校，以及 ARM、IBM、高通和 Xilinx 等国际一流芯片公司参加。我们团队参加了此次比赛的图像分类任务，并获得 ImageNet 和 CIFAR-100 两个赛道的双料冠军。下面简单介绍我们提出的综合利用稀疏、量化以及蒸馏的网络压缩方案。

首先，对于网络剪枝，不同的层具有不同的敏感性，因此需要使用不同的稀疏率进行剪枝。我们提出一种简单的鲁棒性分析方法，对每一层权值进行不同稀疏度剪枝，并观察对精度的影响，从而确定每一层的稀疏度。图 2 展示了 MixNet-S 网络每一层的敏感度，可以看出

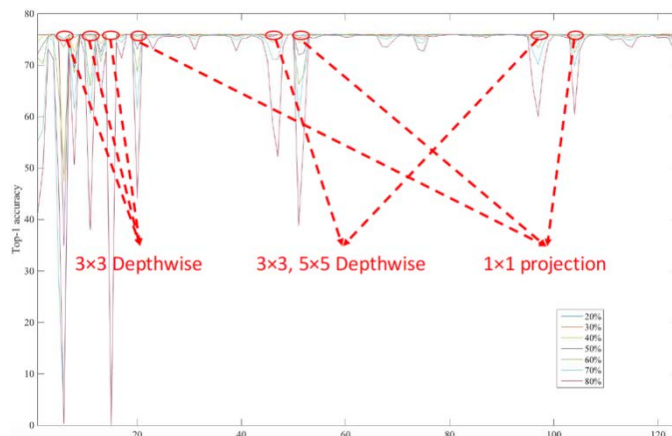


图2 网络不同层鲁棒性分析

Student	Teacher	Teacher Acc.	Top-1 Accuracy
MixNet-s-pruned	-	-	74.4 %
MixNet-s-pruned	MixNet-s	75.9 %	74.6 %
MixNet-s-pruned	MixNet-m	77.2 %	74.9 %
<b>MixNet-s-pruned</b>	<b>MixNet-l</b>	<b>78.9 %</b>	<b>75.0 %</b>
MixNet-s-pruned	SENet154	81.3 %	74.7 %

表 1 不同性能教师网络对知识蒸馏的影响

对于不同层，网络的敏感度差异非常大，因此对于这些比较敏感的层需要使用更低的压缩率进行压缩。

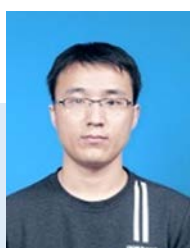
对于量化，我们采用了两阶段量化方式，以实现激活量化和权值量化的解耦。具体而言，我们首先对激活进行量化，而保持权值为全精度浮点数。此时，网络仅仅学习中间层的低比特特征表达，浮点数权值能够保障学习到更好的中间层特征。在低比特特征学习之后，再对权值进行量化，通过激活量化和权值量化的解耦，降低网络量化的难度，从而达到更高的精度。另一方面，与网络剪枝类似，不同的层对于量化的敏感度不一样，我们采用了混合精度量化方案，即对于不同的层选取不同的位宽，达到精度和压缩率的平衡。

知识蒸馏使用一个高精度的教师网络去监督一个低精度的学生网络，往往可以大幅度提升学生网络的性能。然而，教师网络的选择对于学生网络的影响非常大。

我们通过大量实验发现，通过使用更高性能的教师网络作为监督，并不能保证学生网络的性能更高。表 1 展示了不同性能的教师网络去监督 MixNet-S 网络的结果，可以看出，教师网络的性能过高或者偏低都不利于学生网络的学习。

通过综合运用以上网络稀疏化技术、网络量化技术以及知识蒸馏技术，可以将深度学习算法模型进行轻量化和计算提速，大幅度降低算法模型对算力、功耗以及内存的需求。最终，相比于主办方提供的基准模型，我们在 ImageNet 任务上取得了 20.2 倍的压缩率和 12.5 倍的加速比，在 CIFAR-100 任务上取得了 732.6 倍的压缩率和 356.5 倍的加速比，并获得上述两个任务的第一名。

责任编辑 王金甲



## 王培松

中国科学院自动化研究所助理研究员，主要研究方向为机器学习、计算机视觉、神经网络高效计算。

Email: peisong.wang@nlpr.ia.ac.cn



## 程健

中国科学院自动化研究所研究员，人工智能和先进计算联合实验室主任，自动化所南京人工智能芯片研究院常务副院长，主要研究方向为机器学习、计算机视觉、芯片设计。

Email: jcheng@nlpr.ia.ac.cn

热点追踪

# TCNN：利用轨迹片段进行视频目标检测

悉尼大学 欧阳万里

近年来图像目标检测准确率有了很大提升。除了更强大的卷积神经网络架构，新型检测算法框架也扮演了十分重要的角色。但是在视频目标检测领域，相关框架性研究仍然十分少，视频时序和上下文信息仍然没有被很好地利用起来。TCNN(Tubelets with Convolutional Neural Network)视频目标检测框架的提出，能够较较好地将视频的时序和上下文信息利用起来，极大地提升了将单图目标检测算法应用在视频中的准确率。

该框架提出利用轨迹片段(Tubelets)信息来提升视频目标检测的准确率。主要模块包括多上下文抑制(Multi-Context Suppression)及运动导向的帧间传播(Motion-Guided Propagation)、轨迹片段二次估计(Tubelet Re-scoring)等。为了得到更好的结果，该方法还使用了多模型组合增强的技术。

多上下文抑制模块考虑了整个视频检测目标的置信度，综合排序后筛去低置信度目标，有效地降低了误检率。运动导向的帧间传播模块使用了光流信息估计一个检测目标在相邻若干帧图像中的运动位置变化，并将该检测目标向相邻帧传播，以此降低漏检率，提高召回率。轨迹片段二次估计利用了视觉目标跟踪的思想，将高置信度的目标跟踪形成轨迹片段，并在轨迹片段内重

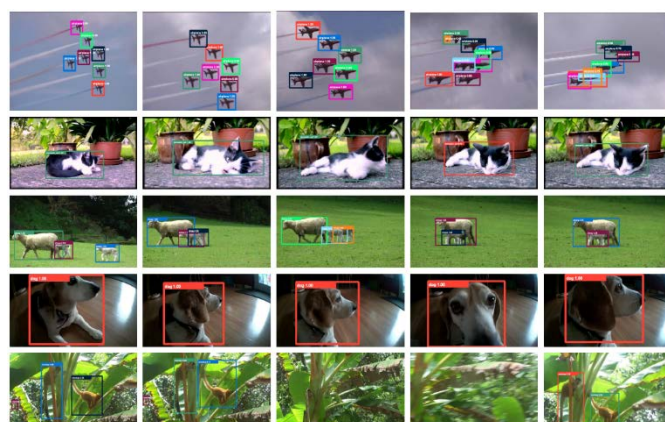
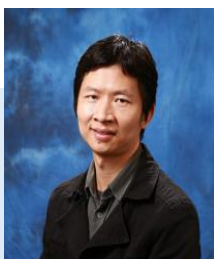


图1 TCNN 算法效果示例

新估计所有目标的置信度，从而得到更可靠的估计。相比此前的视频目标检测的尝试，该框架充分分析了视频目标检测中的问题，系统地提出了如何利用视频时序和上下文信息的解决方法，为许多后来工作提供了重要的启示和借鉴。

该算法框架在2015年的ImageNet大规模视觉识别竞赛中取得视频目标检测赛道冠军，相关论文投稿于TCSVT期刊。2020年，论文第一作者康恺博士被授予IEEE CAS 杰出青年作者奖。该奖项每年在电路与系统协会五个期刊发表的所有论文中选择一篇。

责任编辑 储珺



## 欧阳万里

悉尼大学高级讲师，IEEE 高级会员，主要研究方向为基于深度学习的物体检测和跟踪。  
Email: wanli.ouyang@sydney.edu.au

顶会观察

## ICLR 2020

燕山大学 王金甲

**国**际表征学习大会(International Conference on Learning Representations, ICLR)是人工智能(AI)分支表征学习,特别是深度学习领域的顶级学术会议。ICLR是由深度学习三巨头中的Yoshua Bengio和Yann LeCun主导创办。ICLR是世界上增长最快的一个人工智能会议。ICLR与会者背景广泛,从学术界和工业界的研究人员,企业家和工程师,到研究生和博士后。

2020年4月27日至30日第八届ICLR原定在埃塞俄比亚首都亚的斯亚贝巴举行,由于COVID-19新冠病毒疫情的全球肆虐,成为史上首个完全线上虚拟会议。

### 一、国际表征学习会议的亮点

由于新冠病毒疫情,非洲大陆史上第一次承办的大型人工智能顶会被迫采用完全数字虚拟会议,这成为了本次会议最大的一个亮点。虚拟会议网站的开发人员将每一篇论文的摘要、演讲、论文、代码、评论全部集中在一起。ICLR会议接收论文的作者需要预先录制好自己的视频、音频或文本。每篇论文分两次播放,且每篇论文都有一个讨论聊天功能,可以通过文本进行交互。会议提前召集了500名志愿者对系统的各个方面进行了压力测试。虚拟会议减少了研究人员差旅费、注册费和时间成本,1300多名演讲者吸引了来自近90个国家的5600多名参与者,参加人数暴增了一倍。根据大会官方统计数据,线上演讲视频观看次数突破10万+,相关Zoom讨论会议多达1400余次。大会主席康奈尔大学技术学院Alexander Sasha Rush居功甚伟。会议线上Q&A环节也十分精彩。全部会议内容免费开放。

### 二、论文录用情况

ICLR 2020共收到了2594篇论文投稿,最终有687篇被接收,录取率为26.5%,低于去年31.4%。接收论文包括48篇oral,108篇spotlight和531篇poster。每篇oral的演讲时间10分钟以上,每篇spotlight的演讲时间4分钟。

据统计,412篇被接收论文中有华人学者参与,占比60.0%。2566位全部作者中共有655位华人学者,占比25.5%。入选ICLR 2020论文最多的学者是UC Berkeley的副教授Sergey Levine,共有13篇论文。入选论文较多的华人学者中有三位最为瞩目,分别是清华大学计算机系朱军教授,佐治亚理工学院计算科学与工程系终身副教授宋乐和北京大学的王立威教授。

根据清华大学计算机系研发的AMiner学术数据库统计显示,谷歌入选论文80余篇,其中12篇Oral,7篇满分论文。在入选高分论文中,如华为、腾讯、快手、字节跳动等国内企业,清华大学、北京大学、南京大学、哈尔滨工业大学、西安电子科技大学等国内高校都榜上有名。可以看到,无论是华人学者,还是国内企业、高校和研究所在人工智能领域发挥越来越大的作用。

ICLR2020会议论文关键词包括深度学习(Deep Learning)、强化学习(Reinforcement Learning)、表征学习(Representation Learning)、图神经网络(Graph Neural Network)、生成模型(Generative Model)等。深度学习、强化学习持续占领最受欢迎关键词王冠,图神经网络则新贵上位。

### 三、主题演讲

会议一共邀请了8位全球高影响力的学者作主题演讲,包括FAIR首席科学家Devi Parikh、“卷积网络之父”Yann LeCun、深度学习先驱Yoshua Bengio、美国三院院士Michael I. Jordan等。这些特邀报告都是提前录制的,并当天开放,每个报告有Q&A讨论环节。

Devi Parikh讨论了计算机视觉和自然语言处理交叉AI系统和挑战,解释了为什么视觉和语言的交集问题是令人兴奋的。Yann LeCun看好自监督学习,“我们作为人类学到的大多数知识以及动物学到的大多数知识都是自监督模式,而不是强化模式。”自监督学习算法不依赖注释,而是通过揭示数据各部分之间的关系从数据生成标签。Yann LeCun认为不确定性是阻碍自监督学习成功的主要障碍,自监督表示学习的未来在于正规化的隐变量能量模型。Yoshua Bengio总结了深度学习先验和意识处理的思想、因果结构和注意力机制的最近进展,目的是为了获得更好的分布外泛化能力。Michael I. Jordan认为机器学习(ML)的决策将是未来的重点,从动力、统计和经济角度给出了机器学习的决策进展。

### 四、满分论文和新颖想法的论文

ICLR会议采用Open Review的评审制度。所有提交论文都公开论文姓名等信息,并且接受所有同行的评审以及提问(open peer review)。会议出现了多达34篇满分论文,涉及的内容有神经结构搜索、生成对抗网络、多智能体学习、神经机器翻译、语言生成、信息瓶颈、稀疏表示、字典学习、图神经网络、强化学习、元学习、贝叶斯深度学习、无监督学习、对抗学习、最优化、推理、因果发现和自动微分框架等。可以看出这些满分论文的关键词也体现出了深度学习的主要趋势。这里分享几篇满分论文的精彩报告,分别来自北京大学<sup>[1]</sup>、清华大学<sup>[2]</sup>、图宾根大学<sup>[3]</sup>、斯坦福大学<sup>[4]</sup>、麻省理工学院<sup>[5]</sup>、西安电子科技大学<sup>[6]</sup>、南京大学<sup>[7]</sup>、DeepMind<sup>[8]</sup>和谷歌<sup>[9]</sup>等,论文内容分别涉及复杂系统框架的设计<sup>[1]</sup>、

门控可学习ISTA稀疏编码<sup>[2]</sup>、自动微分框架的反向传播算法<sup>[3]</sup>、用不动点概念来描述ReLU函数<sup>[4]</sup>、生成对抗网络和隐空间插值用于模型泛化<sup>[5]</sup>、贝叶斯深度学习的神经语言模型<sup>[6]</sup>、生成网络的神经机器翻译<sup>[7]</sup>、元学习与扭曲梯度下降<sup>[8]</sup>、多智能体学习广义训练方法<sup>[9]</sup>等。

本次会议涌现出了许多令人感兴趣的论文,这里提及其中的几篇论文。Yoshua Bengio等人提出了元学习因果结构;DeepMind和哈佛大学研究了神经网络控制虚拟小白鼠;斯坦福大学和谷歌给出了ELECTRA预训练模型实现更高效的自然语言处理;谷歌发布了元数据集,学会从少样本中学习的数据集;DeepMind和麻省理工学院等联合发布CLEVRER数据集,推动视频理解中的因果逻辑推理;谷歌和UC伯克利推出了单个GPU上可运行更高效的Transformer模型Reformer;斯坦福大学和哈佛大学给出了预训练图神经网络模型。

### 五、研讨会

本次会议共有15个研讨会(workshops)<sup>[10-11]</sup>,涉及的主题有AI时代基础科学、AI在癌症护理克服全球差异的能力、地球科学中的AI、AI和认知科学、深度神经网络和微分方程、重新思考ML的安全和隐私、ML应对气候变化、现实生活中ML、神经结构搜索、非洲自然语言处理、负担起的医疗AI、不断变化的环境和任务中的强化学习、农业中的计算机视觉、用于决策的因果学习,以及发展中国家受限/低资源场景下的ML。

### 六、总结与展望

AAAI 2020实现了半线上会议,ICLR 2020实现了完全线上虚拟会议。ICLR2020成为了在线举行的最大人工智能学术研讨会,成为了完全线上虚拟会议的开端。ICASSP 2020和计算机视觉顶会CVPR 2020也实现了完全线上虚拟会议。ECCV 2020、ICML 2020和NeurIPS2020也宣称将转为完全线上虚拟会议。或许线上虚拟会议将是未来学术交流形态发展的必然趋势?

责任编辑 任桐炜

### 参考文献

- [1] Yilun Xu, Shengjia Zhao, Jiaming Song, Russell Stewart, Stefano Ermon. "A Theory of Usable Information under Computational Constraints", ICLR2020, <https://openreview.net/forum?id=r1eBeyHFDH>.

- [2] Kailun Wu, Yiwen Guo, Ziang Li, Changshui Zhang. " Sparse Coding with Gated Learned ISTA ", ICLR2020, <https://openreview.net/forum?id=BygPO2VKPH>.
- [3] Felix Dangel, Frederik Kunstner, Philipp Hennig. "BackPACK: Packing more into Backprop", ICLR2020, <https://openreview.net/forum?id=BJlrF24twB&noteId=EHmfrSLVu>.
- [4] Vaggos Chatziafratis, Sai Ganesh Nagarajan, Ioannis Panageas, Xiao Wang. "Depth-Width Trade-offs for ReLU Networks via Sharkovsky's Theorem", ICLR2020, <https://openreview.net/forum?id=BJe55gBtvH>.
- [5] Ali Jahanian, Lucy Chai, Phillip Isola. "On the "steerability" of generative adversarial networks", ICLR2020, <https://openreview.net/forum?id=HylsTT4FvB>.
- [6] Dandan Guo, Bo Chen, Ruiying Lu, and Mingyuan Zhou. "Recurrent Hierarchical Topic-Guided Neural Language Models", ICLR2020, <https://openreview.net/forum?id=Byl1W1rtvH>.
- [7] Zaixiang Zheng, Hao Zhou, Shujian Huang, "Mirror-Generative Neural Machine Translation", ICLR2020, <https://openreview.net/forum?id=HkxQRTNYPH&noteId=HmFOnYDNdx>.
- [8] Sebastian Flennerhag, Andrei A. Rusu, Razvan Pascanu, Francesco Visin, Hujun Yin, Raia Hadsell. " Meta-Learning with Warped Gradient Descent ", ICLR2020, <https://openreview.net/forum?id=rkeiQIBFPB&noteId=zwVZBPeDT>.
- [9] Paul Muller, Shayegan Omidshafiei, Mark Rowland. "A Generalized Training Approach for Multiagent Learning", ICLR2020, <https://openreview.net/forum?id=Bkl5kxrKDr&noteId=XEgYShJzuK>.
- [10] [https://iclr.cc/virtual\\_2020/index.html](https://iclr.cc/virtual_2020/index.html).
- [11] <https://www.aminer.cn/conf/iclr2020/>.



## 王金甲

燕山大学信息科学与工程学院教授，主要研究方向为信号处理和模式识别。  
Email: [wjj@ysu.edu.cn](mailto:wjj@ysu.edu.cn)

## 中国海洋大学董军宇教授访谈

2020年4月28日,《CCF-CV专委简报》在线采访了中国海洋大学海德学院院长、视觉实验室负责人、博士生导师董军宇教授。下面是采访实录。

董老师,您好!首先,请您介绍一下您的个人学习和研究经历。

我本科和硕士都是在青岛海洋大学读的,也就是现在的中国海洋大学,当时学习的专业是应用数学。硕士毕业之后,我去了英国 Heriot-Watt 大学读博,并在2003年取得了博士学位。博士期间我的主要研究方向就是计算机视觉,主要研究了三维表面纹理的合成等课题。2004年回国后我创办了计算机视觉实验室,之后在计算机视觉及海洋大数据分析等方面进行了有针对性的研究。

您在计算机视觉和海洋大数据领域有很深入的研究,能否介绍一下您在这两个领域的研究情况和相关研究成果?

不敢说有深入的研究,只是进行了针对性的探索。在计算机视觉领域,我们主要围绕水下高分辨率三维重建开展研究,由于水下环境对成像造成的影响,涉及了多项研究内容,比如成像模型、特征提取与匹配等。我们研发出了国内外首个定点式海底三维微地形高分辨率获取设备,已在国家深海基地管理中心应用;目前主持承担了国家自然科学基金重大仪器项目,研制移动式水下高分辨率三维成像仪,解决该类设备我国长期依赖

进口的困境。

在海洋大数据领域,主要研究数据驱动的海洋环境要素及过程的预测方法,我们的工作国际上最早提出采用深度学习方法预测海表温度的研究之一,相关方法多次被国内外同行作为数据驱动的海表温度预报基准对比方法。

上述工作也在国际相关领域的顶尖及权威期刊与国际会议上发表了50余篇论文,并获得了10余项授权专利。

您认为水下视觉和海洋大数据分析方面研究的特殊之处在于什么地方?这方面我们国家现在的研究进展如何?相比国际,我们的研究地位如何?

由于水下光线不足造成的暗光视觉问题和由于水的混浊和色度等情况造成的颜色视觉问题等,使得水下图像具有更加复杂的噪声信息,为目标识别和三维重建均带来极大的挑战。海洋大数据则具有不平衡性、多源性及时空过程性,关键问题在于如何将传统物理海洋模型方法与数据驱动方法相结合,精细化发现并预测海洋过程。

我国在水下视觉领域进展快速,自2017年起,国家自然科学基金委每年主办水下机器人目标抓取大赛,国内的顶尖研究团队通过比赛进行交流,技术水平不断进步,在水下目标识别等领域正逐步占据国际领先地位。海洋大数据领域起步较早,但是国际上进展迅猛。

科研方面，您获批了多项国家级项目，如国家重大科研仪器研制和联合基金项目等，能跟大家分享一下您成功申请的经验和体会吗？

申请项目的时候，我们首先考虑的自身研究兴趣与国家需求相结合，将科研创新建立在满足国家重大战略发展服务的基础上，以提高国家尖端技术领域的国际竞争力为目标，进行学术界的前沿性探索。与此同时，申请材料的内容结构和格式规范也是非常重要的，我们会听取不同领域的专家意见，多次会议讨论，仔细修改完善申请材料。

您曾去过英法两国多个大学进行学术访问，能跟大家分享一下您的访学经历，及您对各个访学大学学术环境的看法吗？

从英国 Heriot-Watt 大学博士毕业后，我在 2007 年 5 月和 10 月受英国 British Council 资助，在英国剑桥 Sanger 研究所从事合作研究；2007 年 6 月及 2011 年 5 月受聘在法国南特大学担任访问教授；此外，还在 2008 年 10 月到 2012 年 7 月在英国 Heriot-Watt 大学纹理实验室兼任博士生导师。各个大学的学术环境都有自己的特色，比如剑桥大学人才济济，学术上以打造世界顶尖研究为目标；Heriot-Watt 大学重视新兴学科及专业特色；法国大学的学术氛围融洽而且宽松，注重工程实际应用。

您不仅是 CCF-CV 专委委员，还是 CCF 青岛分部和 ACM 青岛分会主席，能介绍一下您作为 CCF 青岛分部主席，在组织 CCF 活动时的一些主旨吗？您认为 CCF 对学术、企业界最大的贡献在哪里？

CCF 青岛分部主要组织的活动，包括“走进名企”系列活动、“海洋大数据智能分析研讨会”、“走进科学号海洋科考船”、“Tech-Fair”系列活动、“崂山论坛”系列活动等等，这些活动的主旨都是为学术界和企业界搭建一个沟通交流的平台，其中的“Tech-Fair”平台是我们 CCF 青岛分部的特色活动，活动的目的就是推

动学术界和产业界的交流合作。

我认为 CCF 是一个良好的平台，能够为计算机领域专业人士的学术和职业发展提供很好的帮助，促进学术界和企业界的交流合作，学术界的人士可以向企业界介绍学术前沿成果，有助于推动相关研究成果的孵化；来自企业界的人士也可以通过创新技术，提出实际需求，为学术研究工作提供更多的新思路。

您作为中国海洋大学-澳大利亚阿德莱德大学合作设立的海德学院院长，并兼任信息学院副院长，每天需要处理众多日常事务，您是如何做好行政工作与科研工作的平衡呢？能分享一下您的经验吗？

我认为平衡行政工作和科研工作最重要的一点就是时间管理，比如我会每天拿出固定的时间进行科研，每周参与每个科研小组的会议讨论，定时与学生进行沟通，讨论科研中的问题和进展，利用零碎的时间去处理零碎的任务，比如等待开会的前几分钟，或者在外出差的时候，可以在火车上处理邮件，或者在飞机上阅读修改论文等。做好时间的管理，利用好零碎的时间去处理一些事情，可以帮助我们更高效地完成我们的任务。

作为中国海洋大学视觉实验室负责人，您是如何带领及管理此团队的？

我们实验室针对理论研究及科研项目设有几个研究小组，每个研究小组都由老师、博士研究生和硕士研究生组成，小组会定期进行会议汇报科研进展并讨论问题。我主要根据项目及理论研究进展协调各个研究小组，并主持或参与小组的讨论，确定大的方向及时间进度要求。我认为一个团队需要的是宽松的科研环境以及能够打刻苦攻坚战的凝聚力，因此日常科研中大家会畅所欲言，团队经常会组织各种各样的团建活动，比如爬山、跑步，这些活动在增强团队凝聚力的同时也锻炼了大家的身体，很受同学们的喜欢，当遇到难题时，大家可以团结奋战。

能介绍一下中国海洋大学视觉实验室的人员结构、研究方向和未来愿景吗？

我们实验室成立于 2004 年，主要开展海洋大数据深度学习算法方面的研究。团队有成员 100 余人，包括 1 位教授，7 位副教授，3 位讲师，2 名博士后、17 名博士研究生和 80 多名硕士研究生。团队的主要研究内容包括：水下视觉、机器学习、海洋大数据分析及复杂网络等。从成立至今，实验室面向国家重大需求开展研究，得到了国家重点研发计划项目、国家自然科学基金面上项目、国家科技支撑计划项目以及山东省重点研发计划项目等资助，研发了国内外首个分辨率达到毫米级的定点式水下目标三维重建系统。未来希望实验室能够在高分辨率水下三维成像方面取得重要的突破和创新，能够在解决国家面临的关键问题上出一份力。

请问下您是怎么管理研究生的？您对他们的要求是什么？

我们实验室高度重视研究生培养：一方面，为学生

提供一流的硬件设施和丰厚的科研奖励；另一方面，在科研训练中积极帮助学生成长，相当多的硕士在毕业时学术成果和水平可以达到某些中等博士水准。我们要求学生具有一定的编程能力和有一定的英文阅读和书写能力，这对于学习和科研是一个比较好的基础。此外，很重要的一点，就是希望学生能有毅力。我们都知道做科研不是一件很容易的事情，在完成一个目标的过程中可能会经历很多次失败，希望学生能够在学习科研的时候沉下心，不要因为一时的失败而气馁，坚持努力，相信每个人最终都能够有所成就。

如果吐露研究工作者的心声，您最想说的是什么呢？

我觉得国内的计算机视觉研究工作者最好针对国家或社会需求发力，需要大量且耐心的文献阅读、想象力、好奇心以及锲而不舍的钻研精神。

责任编辑 赵振兵 余烨



## 董军宇

教授，博士生导师，现任 CCF 青岛分部主席，ACM 青岛分会主席，中国海洋大学与澳大利亚阿德莱德大学合作设立的海德学院院长，兼任信息科学与工程学院副院长。分别于 1993、1999 年在中国海洋大学获得学士和硕士学位，2003 年在英国 Heriot-Watt 大学获得博士学位。2008 年被列入教育部新世纪人才计划。担任 KSII Transactions on Internet and Information Systems 期刊编委，法国 Interdisciplinary Graduate School for the Blue Planet (ISBlue) 国际学术委员会委员。多次担任国际会议主席或者主要组织者。主要研究方向为计算机视觉、水下视觉及海洋大数据分析。主持承担了科技部国际合作项目 1 项、国家自然科学基金项目 6 项，包括重大仪器（自由申请）项目 1 项、NSFC-山东联合基金 1 项，此外还主持承担了多个省部级项目。已在多个主流国际期刊及国际学术会议上发表论文 200 余篇，获得授权国家发明专利 6 项。

## 委员好消息

❖ 2020年4月23日，2019年度河北省科学技术奖评选结果揭晓，CCF-CV专委会委员、华北电力大学赵振兵副教授参与完成的“基于多光谱成像和旋翼飞行器的输电线路巡检关键技术及其应用”项目荣获科技进步一等奖。

❖ 2020年5月19日，2019年度上海市科学技术奖获奖项目出炉，CCF-CV专委会委员、复旦大学姜育刚教授被授予上海市青年科技杰出贡献奖。本年度共授奖308项（人），其中，2人获上海市科技功臣奖，10人获青年科技杰出贡献奖，43项成果获自然科学奖，31项成果获技术发明奖，205项成果获科技进步奖，15项成果获科学技术普及奖，2位外籍专家获上海市国际科技合作奖。

❖ 2020年5月21日，中国电子学会公布了第七届中国电子学会优秀科技工作者表彰名单，CCF-CV专委会副秘书长、北京工业大学毋立芳教授获“中国电子学会优秀科技工作者”荣誉称号。

❖ 2020年5月30日，人力资源社会保障部、中国科协、科技部、国务院国资委公布了《关于表彰第二届全国创新争先奖获奖者的决定》。CCF-CV专委会委员、西安电子科技大学高新波教授（现任重庆邮电大学校长）入选第二届全国创新争先奖状表彰名单。全国创新争先奖是继国家自然科学奖、国家技术发明奖、国家科学技

术进步奖之后，中国批准设立的又一个重要的科技奖项，是国家科技奖励体系的重要组成部分和补充，是国家科技奖项与重大人才计划的有机衔接，是仅次于国家最高科技奖的一个科技人才大奖。该奖每三年评选表彰一次，每次表彰10个科研团队授予奖牌，表彰不超过30个科技工作者授予奖章，享受省部级劳模待遇，表彰不超过300名科技工作者授予奖状。本年度授予10个创新团队创新争先奖牌、28名科技工作者创新争先奖章、258名科技工作者创新争先奖状。

❖ 2020年5月30日，陕西省“特支计划”“三秦学者”创新团队支持计划拟入选名单公布，共192人拟入选第二批陕西省“特支计划”、47个团队拟入选第二批陕西省“三秦学者”创新团队支持计划。CCF-CV专委会委员、西安电子科技大学邓成教授和中科院西安光机所卢孝强研究员入选“特支计划”。

❖ 2019年12月2日，中国图象图形学学会首批会士名单公布，CCF-CV专委会顾委主任、中科院自动化所谭铁牛院士和CCF-CV专委会常务委员、中山大学赖剑煌教授入选。

责任编辑 刘海波

# 医学图像配准开源代码

大连理工大学 李孜 樊鑫

**可**变形配准问题是医学图像分析的关键技术之一，用于对齐不同医学图像。它旨在在图像对之间建立密集的对对应关系（也称为配准字段），从而将不同的图像转换至公共坐标系。它在医学图像分析中具有广泛的应用，例如群体建模，解剖结构变化诊断和多模态图像融合。

传统方法旨在求解特定于图像对的优化模型，计算成本非常高。深度学习用网络模型学习代替昂贵的数值优化，可以提供快速的变形估计。本文介绍基于深度学习的可变形医学图像配准工作及其开源代码。

## 1、An Unsupervised Learning Model for Deformable Medical Image Registration

**工作：**该论文将配准定义为参数函数并使用卷积神经网络对该函数建模。给定一对新扫描，使用学习的参数直接评估函数，从而实现快速配准。该方法借助空间转换函数实现无监督学习，流程图如图 1。相关工作发表在 CVPR-2018 上。

**论文：**<https://arxiv.org/pdf/1802.02604v3>

**代码：**<https://github.com/balagk/voxelmorph>

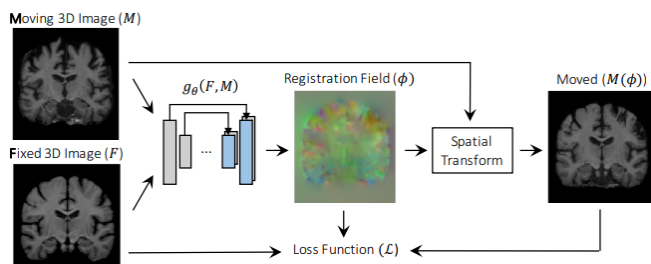


图 1 无监督学习框架

## 2、Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration

**工作：**该论文提出一个概率生成模型，并推导了一种基于无监督学习的推理算法，同时基于平方和缩放实现微分同胚的配准。该方法可以提供拓扑结构保留保证和不确定性估计。图 2 给出了模型示意图。该方法利用经典配准方法的领域知识，以及卷积神经网络的最新发展，建立了经典方法与深度学习方法之间的联系。相关工作发表在 MICCAI-2018 上。

**论文：**<https://arxiv.org/abs/1805.04605>

**代码：**<https://github.com/voxelmorph/voxelmorph>

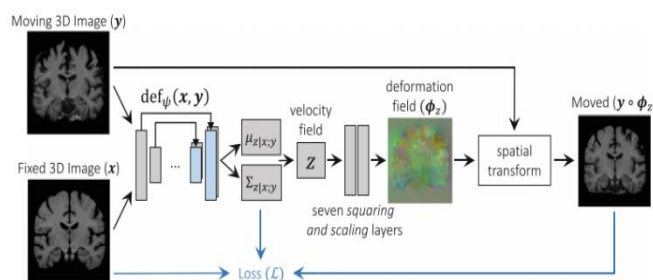


图 2 基于概率生成的微分同胚学习模型

## 3、Region-specific Diffeomorphic Metric Mapping

**工作：**该论文介绍了一种特定于区域的微分同胚度量映射 (RDMM) 配准方法。该模型使用时空正则化器来捕

获大位移，跟踪区域变形。此外所提出的学习方法可实现快速的 RDMM 估计。图 3 给出了相关网络结构。相关工作发表在 NIPS-2019 上。

论文: <https://arxiv.org/pdf/1906.00139v2>

代码: <https://github.com/uncbiag/registration>

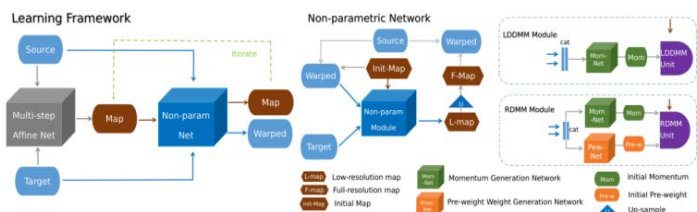


图 3 学习框架及其非参数配准组件

#### 4、Generation Recursive Cascaded Networks for Unsupervised Medical Image Registration

**工作:** 该论文通过深度递归级联结构实现渐进式配准。运动图像通过每个子网络的预测流场依次变形，最后与固定图像对齐，每个子网络学习对当前的变形图像的渐进式配准。图 4 给出了深度递归级联网络的结构。通过深度的递归迭代，最终的流场可以被分解为简单轻微的渐进变化，从而降低每个子网络的学习难度。相关工作发表在 ICCV-2019 上。

论文: <https://arxiv.org/pdf/1907.12353v3>

代码: <https://github.com/microsoft/Recursive-Cascaded-Networks>

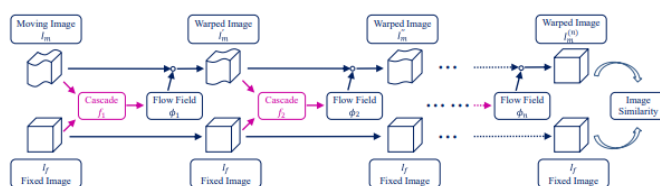


图 4 递归级联的神经网络结构

#### 5、Bi-level Probabilistic Feature Learning for Deformable Image Registration

**工作:** 该论文提出了一种新颖的双层优化配准模型，可以协同求解变形场和特定于配准任务的特征。该模型根据上层子问题交替求解变形场，并根据下层子问题调整特征表示，从而为前端特征提取阶段自适应地强制执行配准约束。此外，该项工作针对两个优化子问题设计了可学习的深度网络。图 5 给出了交替合作的双模型优化示意图。相关工作发表在 IJCAI-2020 上。

论文: <https://github.com/DUT-RegNet/BiLevelReg>

代码: <https://github.com/DUT-RegNet/BiLevelReg>

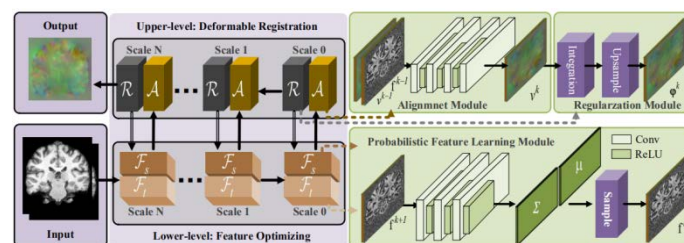


图 5 交替合作的双模型优化机制

责任编辑 沈沛意 李策



李孜

硕士研究生，大连理工大学国际信息与软件学院，研究方向为医学影像分析。



樊鑫

博士生导师，大连理工大学国际信息与软件学院从事教学与科研工作，担任中日国际信息与软件学院院长。研究方向为计算机视觉与图像处理、医学影像分析。

个人主页: [http://faculty.dlut.edu.cn/Xin\\_Fan/zh\\_CN/index.htm](http://faculty.dlut.edu.cn/Xin_Fan/zh_CN/index.htm)

## 手势识别数据集

西安电子科技大学 史媛媛 苗启广 李宇楠

**手**势是一种交流的形式，它是指利用人的肢体动作来说明其意图或态度的行为。由于在视频监控控制、标志语言理解、虚拟现实和人机交互等领域有着巨大的应用前景，越来越多的研究人员开始研究手势识别算法，以实现将人类手势解释给机器的目标。近年来，为了评价手势识别方法的性能，大量的公开数据集被创建。使用公开可用的数据集有两个主要优点。一方面，它们节省了时间和资源，也就是说，不需要录制新的视频序列或付费，因此科研人员可以专注于其特定的算法和实现。另一方面，使用相同的数据集可以方便地比较不同的方法，并了解不同方法的能力。但目前在手势识别领域还没有一个对公开数据集的完整描述。

手势识别数据集一般分为静态手势数据集和动态手势数据集，在视觉手势识别算法方面，静态手势一般基于图像帧，静态手势操作时手的姿态不随时间变化，而动态手势则是基于连续的图像序列。本文分别从静态手势数据和动态手势数据两方面分别介绍几个在手势识别领域常用的公开数据集，具体包括 ASL、11K Hands、LaRED、ChaLearn LAP IsoGD、SKIG 和 EgoGesture。最后将给出的公开数据集进行对比，以便于研究者选择适合的数据集。

### 1、静态数据集

#### ● ASL 数据集

**介绍:** 美国手语字母数据集 (American Sign Language, ASL) 是由 5 个手势表演者在不同的环境光照及背景条件下使用深度传感器 Kinect 采集完成，数据集集中的每

个手势样例均包含了相应的彩色图像与深度图像。

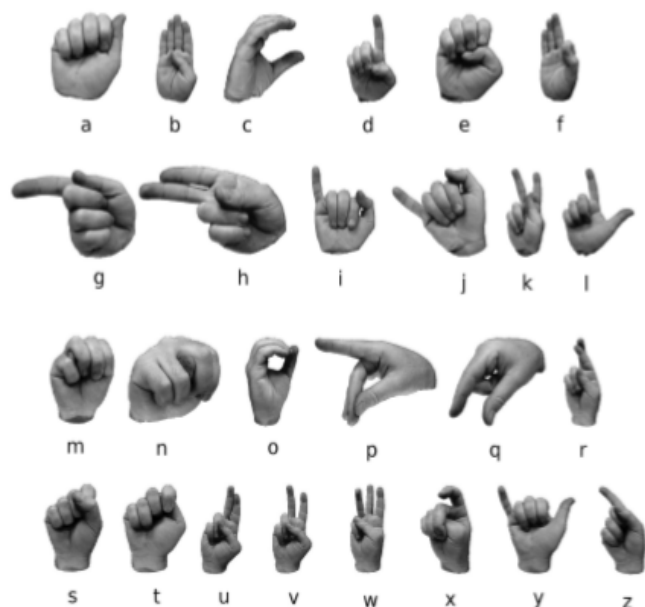


图 1 ASL 数据集

ASL 数据集中总共包含 26 种英文字母手势，其中包括了 24 个静态字母手势及 2 个动态字母手势。针对每个操作者，每种手势字母均包含了 500 张左右的彩色图与深度图。因此，ASL 数据集中的图片总数约为 120000(500×2×24×5)张。图 1 展示了 ASL 数据集中部分对象。

#### ● 11K Hands 数据集

**介绍:** 11K Hands 数据集涵盖了来自 190 个 18-75 岁之间的实验对象的 11076 幅手部图像(1600×1200 像素)。图 2 展示了 11K Hands 数据集中部分对象，数据集中每个表演者都张开或握紧左右手的手指，然后在统一的白色背景，且手距离相机位置大致相同的情况下，

分别从每只手的背侧和手掌侧进行拍摄。每个图像相关的数据记录包括：(1) 实验对象 ID、(2) 性别、(3) 年龄、(4) 肤色、以及 (5) 有关所捕捉到的手的一组信息，即右手或左手、手的部位(手背或手掌) 以及逻辑指标，即指示手部图像是否包含配饰、指甲油或不符规则处。所提出的数据集具有大量附加更为详尽的元数据的手部图像。

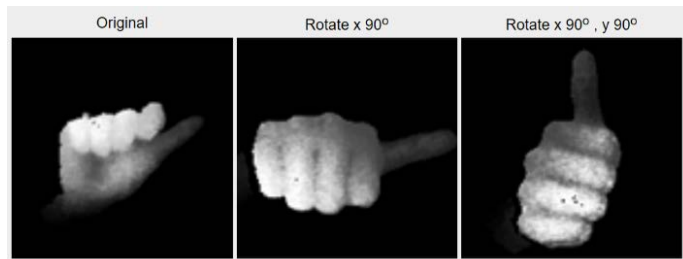


图 4 LaRED 数据集同一手势的三个朝向

对于每个类, 有 10 个表演者(5 名男性与 5 名女性) 执行相应的手势, 10 名表演者编号为 M001-M005(男性)与 F001-F005(女性)。对于每个表演者的每个手势类, 采集 300 个样本, 每个样本包含同步的彩色图像与深度图像。同时, 该数据集给出二值化的掩膜图像来定位手的边界。因此, 在 LaRED 数据集中, 共包含 242900 个样本(由于表演者 M003, 手势编号 G013, 朝向 O002 的类只有 20 个样本)。每个样本由彩色图像, 深度图像, 和对应手势的掩膜图像。对于每个类别手势, 手相对相机有轻微旋转和小范围移动。LaRED 数据集共包含 10 个人共计 242900 张图片, 平均每个人的每个手势约包含 300 张图片。

### 数据集地址

<https://www.kaggle.com/grassknotted/asl-alphabet>  
<https://github.com/mahmoudnafifi/11K-Hands>  
<http://mclab.citi.sinica.edu.tw/dataset/lared/lared.html>

## 2、动态数据集

### ● ChaLearn LAP IsoGD 数据集

**介绍:** ChaLearn LAP IsoGD 数据集是中国科学院自动化研究所 Wan 等人在整理 ChaLearn Gesture Challenge(CGD)2011 数据集的基础上发布的, 是一个大型的操作者独立的动态手势数据库, 适合在深度学习方面使用, 图 5 展示了 ChaLearn LAP IsoGD 数据集的样例。



图 2 11K Hands 数据集

### ● LaRED 数据集

**介绍:** LaRED (Large RGB-D Extensible Hand Gesture Dataset) 数据集为 2014 年 Hsiao 等人公开的大型可扩展手势识别数据集, 采用英特尔的短距离深度相机进行记录。LaRED 数据集使用 27 个手势(记为 G001~G027)作为基础手势, 大多数手势取自美国手语, 图 3 展示了 LaRED 数据集中的 27 个基础手势。



图 3 LaRED 数据集中的 27 个基本手势

对每个基础手势, 取三个朝向(记为 O001-O003)。三个朝向分别为: 基础手势、基础手势绕 x 轴旋转 90 度、基础手势绕 x, y 轴各旋转 90 度。因此数据集总共 81(27×3)个类, 三个手势朝向视图如图 4 所示。

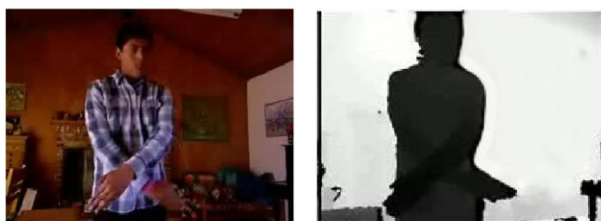


图 5 ChaLearn LAP IsoGD 数据集中的样例

ChaLearn LAP IsoGD 数据集一共包含 47933 个视频序列，每个视频有 RGB 和深度两种数据，单个视频中只包含一个手势实例，采集人员在视频起始和结束期间保持站立姿势，中间部分完成指定手势动作。视频长度不固定，最短视频只有 9 帧图像，最长视频序列长度高达 405 帧。数据集一共有 249 种手势动作，由 21 个人完成，分成训练集、验证集、测试集三个子集，每个子集的详细分布情况参考表，各自包含了 35878、5784、6271 个视频序列，其中根据训练、验证、测试的顺序分别配备了 17、2、2 个表演者负责对应子集的手势动作采集。

● SKIG 数据集

**介绍:** 动态手势数据集 Sheffield Kinect Gesture Dataset (SKIG) 是由 Sheffield 大学的 L.Liu 和 L.Shao 在 2013 年收集和发布的。SKIG 一共有 2160 段手势序列，包括 1080 段 RGB 序列和 1080 段 Depth 序列，其中每一段 RGB 和 Depth 序列都是使用 Kinect 深度摄像头同时拍摄的。SKIG 数据集在 6 种不同的主题场景中采集了 10 种不同类型的动态手势，每种手势如图 6 所展示，并且 10 种不同手势之间的分类依据是手臂的运动轨迹，以手的运动信息居多。

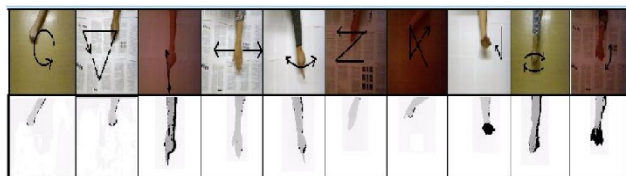


图 6 SKIG 数据集

10 种手势在采集时都使用了三种不同的手掌形态：握紧拳头、食指张开、五指张开。为了增加同一种类型手势的差异性，每种手势都采集了 2 种光照（强光和弱光）和 3 种背景（木板背景、白板背景和纸张背景）的

样本。总的来说，SKIG 共收集了 360 段手势序列样本。

SKIG 数据集比较适合作为视觉动态手势识别研究的实验数据，因为 SKIG 数据集采集时兼顾了不同的光照、不同的背景条件以及不同的手掌形态。

● EgoGesture 数据集

**介绍:** EgoGesture 数据集是一个公开的动态手势识别数据集，该数据集是人自我视角下进行拍摄的数据集，共包含 83 个手势，图 7 展示了这些类型的手势。

Label	Illustration	Instruction	Label	Illustration	Instruction	Label	Illustration	Instruction
1		Scroll hand towards right	22		Measure (distance)	43		Palm to fist
2		Scroll hand towards left	23		Photo frame	44		Fist to Palm
3		Scroll hand downward	24		Number 0	45		Trigger with thumb
4		Scroll hand upward	25		Number 1	46		Trigger with index finger
5		Scroll hand forward	26		Number 2	47		Hold fist in the other hand
6		Scroll hand backward	27		Number 3	48		Grasp
7		Cross index fingers	28		Number 4	49		Walk
8		Zoom in with fist	29		Number 5	50		Gather fingers
9		Zoom out with fist	30		Number 6	51		Snap fingers
10		Rotate fist clockwise	31		Number 7	52		Applaud
11		Rotate fist counterclockwise	32		Number 8	53		Dual hands heart
12		Zoom in with fingers	33		Number 9	54		Pat two fingers together
13		Zoom out with fingers	34		OK	55		Take two fingers apart
14		Rotate fingers clockwise	35		Another number 3	56		Turn over
15		Rotate fingers counterclockwise	36		Pause	57		Move fist upward
16		Click with index finger	37		Shape C	58		Move fist downward
17		Sweep diagonal	38		Make a phone call	59		Move fist towards left
18		Sweep circle	39		Wave hand	60		Move fist towards right
19		Sweep cross	40		Wave finger	61		Bring hand close
20		Sweep checkmark	41		Knock	62		Push away
21		Static fist	42		Beckon	63		Flourish upward

图 7 EgoGesture 数据集

EgoGesture 数据集包含来自 50 个不同主题的 2081 个 RGB-D 视频，24161 个手势样本和 2953224 帧。这些视频是从 6 种不同的室内和室外场景中收集的，还考虑了人们在行走时执行手势的情况。图 8 展示了这 6 个场景中的手势样例。

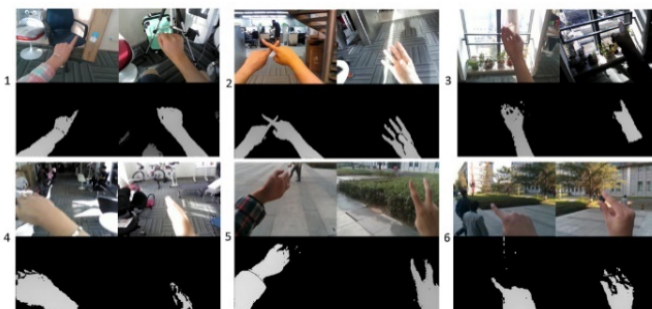


图 8 EgoGesture 数据集的 6 个场景

### 数据集地址

<http://gesture.chalearn.org/2016-looking-at-people-cvpr-challenge/isogd-and-congd-datasets>

[https://docs.opencv.org/3.0-beta/modules/dataset/doc/datasets/gr\\_skg.html](https://docs.opencv.org/3.0-beta/modules/dataset/doc/datasets/gr_skg.html)

<http://www.nlpr.ia.ac.cn/iva/yfzhang/datasets/ego-gesture.html>

### 3、对比总结

近年来，手势识别领域发布的数据集数量大幅增加，与此同时，人们对手势识别的兴趣也越发高涨。本节将从以下几方面对手势数据集进行对比：

(1) 从彩色相机和立体相机到能够获得的颜色和深度图像等多种模态的手势信息也越来越受到研究者的关注。以上提供的数据集中大多数数据集提供了两种模态以上的信息，少数只提供单模态信息。这些数据集部分是由于提出时间较早，深度数据的提取技术不够完善，例如 ASL、11K hands 数据集等。

(2) 手势识别数据集的规模也逐渐在增大。由于研究者对该任务的进一步探索，手势数据集中手势的类型也变得更加丰富。如 SKIG 数据集只有 10 类手势，ASL 数据集由 26 个手势组成，而 EgoGesture 数据集和 ChaLearn LAP IsoGD 数据集分别由 83 类、249 类手势数据组成，丰富的手势类型为手势技术的发展提供了强有力的支撑。

(3) 随着手势数据集数量的增加，数据面向的应用场景变得丰富多样。如 EgoGesture 数据集和 LaRED 数据集倾向于自动驾驶中的手势识别，ASL 数据集面向聋哑人群的手语的分析。

(4) 随着手势识别技术的方法，手势识别数据的背景也变得多样化。不同的数据集针对不同的应用场景，背景情况也不尽相同。SKIG 数据集和 11K hands 数据集的背景相对比较简单，没有背景环境的干扰；而 ChaLearn LAP IsoGD 数据集的背景环境比较复杂，对手势识别提出了一定的挑战性；而 Ego Gesture 数据集针对不同复杂程度的背景构建了 6 种不同场景下的数据集。

责任编辑 李策 沈沛意



### 史媛媛

博士研究生，西安电子科技大学计算机科学与技术学院，研究方向为计算机视觉。



### 苗启广

博士生导师，西安电子科技大学计算机科学与技术学院从事教学与科研工作，西安电子科技大学计算机科学与技术学院副院长，学院教授委员会主任。研究方向为计算机视觉、机器学习、高性能计算、大数据分析。个人主页：<https://web.xidian.edu.cn/qgmiao/>



### 李宇楠

讲师，西安电子科技大学计算机科学与技术学院，研究方向为计算机视觉与机器学习。

## 好文推荐

悉尼大学团队“动作定位”最新成果发表在 IEEE TPAMI-2020。

论文: Rui Su, Dong Xu, Luping Zhou, Wanli Ouyang. Progressive cross-stream cooperation in spatial and temporal domain for action localization, IEEE TPAMI, Early Access, 2020.

计算机视觉技术通常涉及图像采集, 然后通过对采集图像的分析实现对视频中目标的行为进行定位和分析。在一些场景中, 计算机视觉技术还涉及基于行为分析和跟踪进行非法行为的预警等操作。计算机视觉技术中常利用深度学习模型进行图像处理, 但是可能存在的问题是, 目标被遮挡或者运动幅度较大的问题会产生目标丢失或者精确度不够的现象。有效利用互补的视觉特征是解决该问题的方式之一, 而互补的视觉特征可以通过多模态信息获取。例如, 外观特征可以通过 RGB 图像来表达, 而动作特征可以通过光流图像来捕捉。当目标被遮挡的时候, 外观特征无法提供足够的信息, 但是动作特征可以提供额外的信息来提高检测结果。当检测目

标的动作幅度较小时, 此时无法通过动作特征来定位动作而外观特征可以帮助模型定位动作。

长期以来, 绝大多数工作都是简单的把单独依靠外观特征或者动作特征产生的结果融合起来, 但是这种方式并没有充分地利用互补的视觉信息。本工作基于双分支 (即 RGB 分支和光流分支) 的架构, 在空间和时序领域上, 同时在特征层面和区域提名层面, 使用由一个分支 1 (如 RGB 分支) 产生的特征和区域提名去帮助分支 2 (如光流分支) 产生更好的特征和区域提名。然后分支 2 产生的特征和区域提名可以返回去帮助改善分支 1 得到的特征和区域提名。按照这个思路进行多次迭代后能够逐步提高每一个分支产生的动作定位结果从而提高最后的结果。这种有效利用互补视觉信息的方式能够同时应用到空间和时间的动作定位任务当中, 从而提高在视频中对目标进行时空动作定位的准确度。

本工作在两个常用的时空动作定位任务公开数据集上进行了模型测试。试验结果表明有效的利用外观和动作的互补信息能够明显的提高时空动作定位的结果。

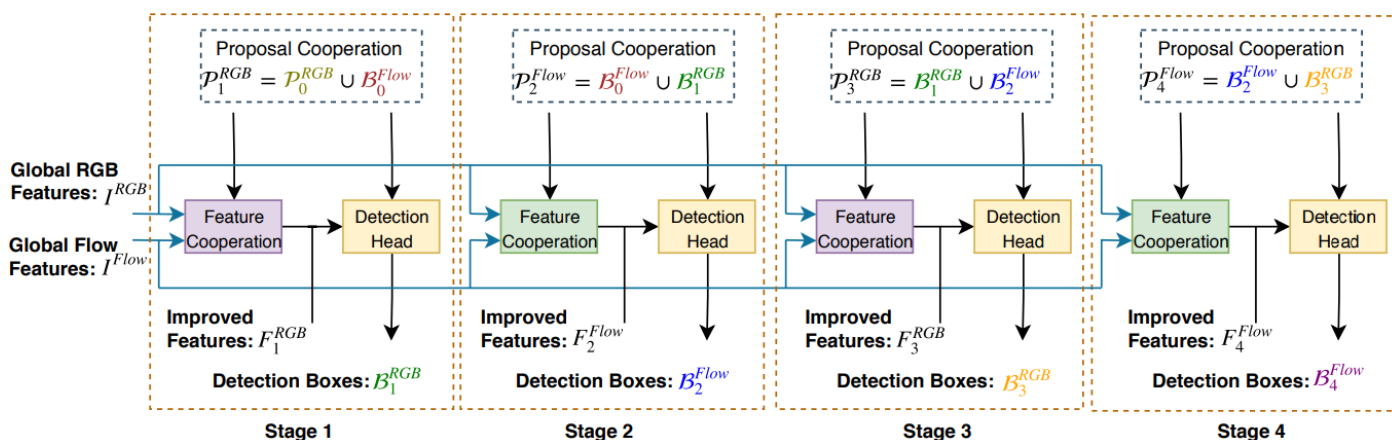


图 1 累进跨分支合作算法

责任编辑 贾同 樊鑫

## 好文推荐

南开大学和新加坡科技研究局团队等多家单位联合发表“有序或无序：基于视频的行人重识别”最新成果在 IEEE TPAMI-2020。

论文: Le Zhang, Zenglin Shi, Joey Tianyi Zhou, Ming-Ming Cheng, Yun Liu, Jia-Wang Bian, Zeng Zeng, Chunhua Shen. Ordered or orderless: a revisit for video based person re-identification, IEEE TPAMI, Early Access, Feb. 2020.

在行人重识别任务中, 递归网络(RNN)是常用的一种算法。但是, RNN 对学习基于视频的行人重识别的视觉表示真的是必要的吗?

基于这一问题, 本文首先证明 RNN 并不是聚合时间和空间特征的最优选择。同时, 通过实验发现, 循环结构并不能像大家所期待的那样有效的学习时间依赖性, 而是会产生一个无序的表达。

从图 1 中可以看到, RNN 过分强调人的运动特征与时间的依赖性, 而这些行为生物特征受到很大的类内/类间变化的影响, 使得现有的 RNN 算法泛化能力较差。基于这一发现, 假设单个静态图像的外观信息在 VPre-id 任务中发挥了更重要的作用, 作者将 VPre-id 看做一个无序集和的排序任务。其中, 每个排序者都是单个图像帧的行人重识别器。多个排序结果通过已有算法进行集成学习, 得到最终结果。所提算法通过基于图像的行人重识别算法的准确度和视频中图像帧的差异, 保证了集成学习中算法的强度和多样性。

相较于传统算法, 所提算法提供了比视频对更加多样化的图像对。已有算法对这些图像对进行了更准确的排序, 有效提高了最终的行人重识别精度。此外, 在独立同分布的假设下, 作者设计了一个误差边界, 来帮助提高 VPre-id 的精度。

实验表明, 所提算法能够与基于图像的行人重识别算法相结合, 在广泛使用的数据集(iLIDS VID、PRID 2011 和 MARS)上取得了优异的成果。同时, 所提算法为缩小视频和图像的行人重识别问题间的差距提供了一条新的思路。

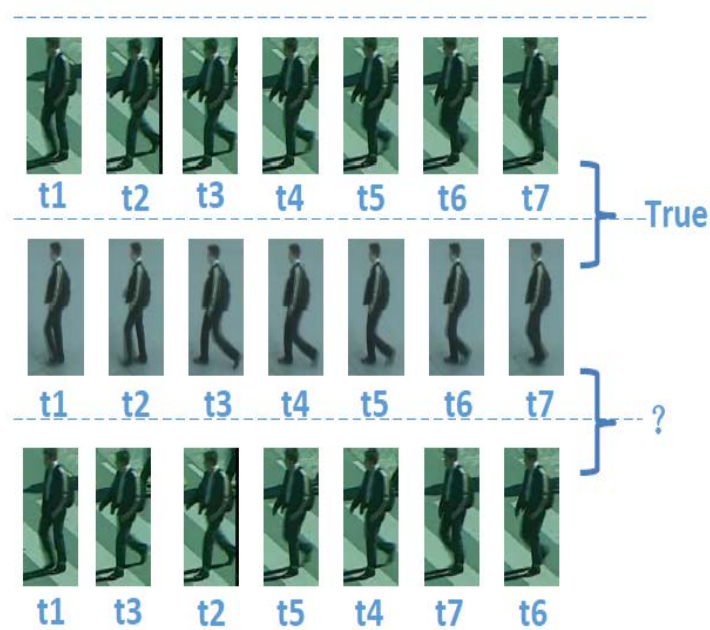


图 1 动机。前两行表示采用 RNN 建模行人重识别问题中的时序依赖, 后两行表示人类可以根据乱序图像进行行人重识别。

责任编辑 贾同 樊鑫

## 好文推荐

西安交通大学团队“图像压缩传感的深度学习方法”最新成果发表在 IEEE TPAMI-2020。

论文: Yan Yang, Jian Sun, Huibin Li, Zongben Xu. ADMM-CSNet: a deep learning approach for image compressive sensing, IEEE TPAMI, vol. 42, no. 3, pp. 521-538, Mar. 2020.

压缩传感(CS)是从少量采样数据中重建图像的有效技术,被广泛应用于医学成像、遥感、图像压缩等领域。传统的压缩传感方法基于压缩感知物理背景建立能量模型并进行优化求解。该论文及其会议版本(NeurIPS 2016)提出融合压缩传感物理机制建模与深度学习的深度压缩传感方法,并应用于核磁共振成像和一般自然图像压缩传感重建问题。该类网络框架被称为 ADMM-CSNet,包含两个深度学习模型。网络具体构建过程如下。首先,考虑建立一个广义的压缩传感模型,包括了物理约束的重建项和具有待定变换域和待定正则子的正则项。然后,提出了两种基于交替方向乘子法(ADMM)的优化求解器对模型进行优化。最后相应的 ADMM 优化算法被推广定义为相应的两种深度网络结构。压缩传感模型中的待定变换域和待定正则子,以及 ADMM 算法中的超参数均转换为网络中的权重参数,并通过采样/真实重建数据对进行网络参数的端到端训练学习。

ADMM-CSNet 的数据流图如图 1 所示,具体的求解步骤如下:

1、构建广义的 CS 模型;

2、以不同的方式引入辅助变量,得到两个不同的 ADMM 迭代算法:

1) 在变换域中引入辅助变量;

2) 在图像域中引入辅助变量;

3、用不同的 ADMM 算法求解 CS 模型,并进一步推导出两个不同的 ADMM-CSNets,分别为: Basic-ADMM-CSNet 和 Generic-ADMM-CSNet。

图中的  $X^{(n)}$  为重建层,  $Z^{(n)}$  为辅助变量更新块,  $M^{(n)}$  为乘法更新层。在 Basic-ADMM-CSNet 中,  $Z^{(n)}$  是由一个卷积层  $C^{(n)}$  和一个非线性变换层  $S^{(n)}$  构成的。在 Generic-ADMM-CSNet 中,  $Z^{(n)}$  可以分解为由卷积层  $(C_1^{(n,k)}, C_2^{(n,k)})$ , 非线性激活层  $H^{(n,k)}$  和加法层  $A^{(n,k)}$  组成的  $N_i$  次迭代过程。实验结果表明, ADMM-CSNet 在核磁共振图像和自然图像的 CS 成像任务中有效提高了重建精度和速度。

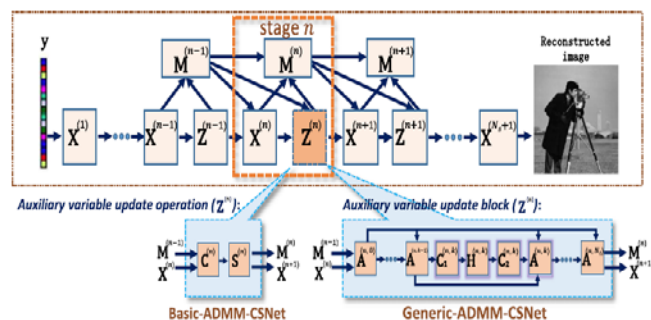


图 1 ADMM-CSNets 的数据流图。该模型通过在不同的域中引入辅助变量,进一步推导出了 ADMM-CSNet 和 Generic-ADMM-CSNet。

责任编辑 樊鑫 贾同

## 英国萨里大学宋一哲教授团队

**SketchX**  
图鸦 视觉实验室

英国 SketchX 实验室由宋一哲教授 (Yi-Zhe Song) 于 2011 年创立, 并与国际知名的计算机视觉与机器学习专家 Tao Xiang 教授、Timothy Hospedales 教授共同领导。目前隶属于欧洲最大的人工智能研究机构之一: 萨里大学视觉、语音和信号处理中心 (CVSSP)。

SketchX 成立的愿景是通过对人类手绘数据更好地理解, 使计算机视觉和认知科学受益, 并最终转化为对人类视觉系统更深的认知 (Seeing Explained by Drawing)。多年来, SketchX 秉承初心, 专注于对人类手绘的研究。在识别、搜索、生成和建模等领域取得了重要的进展, 并不断在顶级会议和期刊上发表。SketchX 还积极致力于人类手绘相关科研社区的构建, 并主导了人类手绘在国际顶级会议的首次专题研讨会 VASE@ECCV'2016。目前有 12 名博士生与 2 名博士后在 SketchX 工作。更多团队成果与信息可以在实验室网站 <http://sketchx.ai> 获取。

从远古时期开始, 人类便开始用与草图类似的岩画或者石雕来渲染他们的视觉世界, 早于语言的出现数千年。直到今天, 这种与生俱来的独特意象表达能力依然被应用于各个方面 -- 从具有教育意义的早教时期, 帮助孩子更好地认知, 到具有娱乐意义的“你画我猜”益智游戏; 再从具有法制意义的精确犯罪画像, 到具有商业意义的作为一种替代文字的搜索引擎输入方式。随着

数字化时代触控设备的普及, 手绘的获取变得更简单、更快捷, 实际应用场景也将更广泛。



1500 BC

AD 1800

AD 1970

图 1: 手绘自古以来是一种普世的人类视觉交流方式

人类视觉系统理解手绘与自然图像并无显著差异, 但是对于机器而言会产生很大的不同。挑战主要体现在以下几方面: (1) 手绘中由于人主观因素的加入而带来的抽象和失真的视觉特性; (2) 手绘是由稀疏的黑色线条与白色背景构成, 与自然图像中大量的色彩和纹理形成鲜明的对比; (3) 手绘数据模态表达的双重性。手绘既可以是栅栏化的像素值 (RGB 图像), 也可以是渲染过程中保存的时序信息 (离散的点)。

### 研究方向 1: 手绘理解 (Sketch Understanding)

SketchX 在 2015 年, 首先重温了传统的识别任务, 提出了首个针对手绘类别识别的深度学习模型 Sketch-a-Net<sup>[1]</sup>, 超越了人类在相同任务下的表现。Sketch-a-Net 荣获了英国机器视觉会议 BMVC 2015 的最佳论文奖并获得了包括 BBC 在内的主流媒体的关注。SketchX 对手绘的理解进一步扩展到了感知分组 (perceptual

grouping) 和语义分割 (semantic segmentation) 任务<sup>[2,3]</sup>, 并贡献了迄今为止最大的相关数据集。受到 AlphaGo 在围棋领域取得突破性进展的激励, SketchX 在 2018 年提出让机器智能在“你画我猜 (Pictionary Game)” 的游戏中挑战人类的愿景。并在文献<sup>[4][5]</sup>的

工作中迈出了坚实的一步。图 2 展示了其中的一些成果。值得注意的是, 算法如何有目的地对目标类别的物体进行刻画, 以及对具有类别区分性的视觉特征进行表现, 就像人一样。

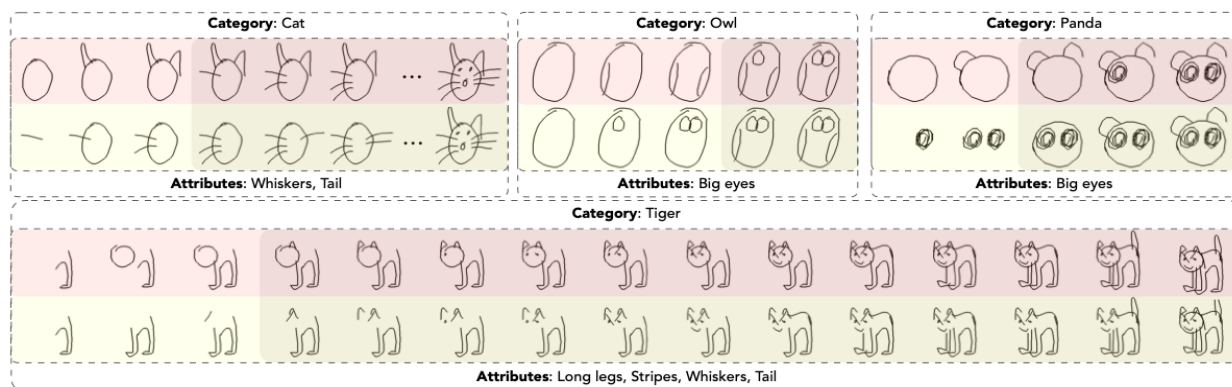


图 2: 机器基于不同的奖励回馈类型, 以不同的方式绘画<sup>[5]</sup>

## 研究方向 2: 从手绘到图像 (Sketch-Photo Gap)

手绘的优势在于“所画即所想”。有些原本难以用语言描述的视觉特点在手绘下变得异常简单。这使手绘成为了一种优越的检索模式并吸引了大量的学术兴趣。在 2014 年, SketchX 首次提出了基于手绘的细粒度图像检索 (fine-grained sketch-based image retrieval, FG-SBIR) 的概念<sup>[6]</sup>, 并于 2016 年在深度模型框架下实现了突破性的进展<sup>[7]</sup>。相关的成果在 CVPR 2016 上进行了口述报告, 在随后的 demo 环节也引起了广泛的关注。所贡献的深度三元模型以及数据库为当今 FG-SBIR 领域的研究奠定了基础。SketchX 此后持续深耕 FG-SBIR 领域, 在可靠性<sup>[8,9]</sup>, 实用性<sup>[10,11]</sup>, 泛化性<sup>[12]</sup> 等方向上取得了大量进展。

## 研究方向 2: 手绘的其他 (Sketch for X)

在坚持手绘是对现代计算机视觉技术及相关应用的重要补充的理解下, SketchX 多年来还深入地探索了以下方向: (1) 领域适应与泛化中的手绘 (Sketch for

Domain Adaptation and Generalisation)<sup>[13,14]</sup>。手绘稀疏和严重形变的视觉特性更容易让模型学到关键与鲁棒的视觉信息。SketchX 在 ICCV 2017 提出的 PACS 数据集是该领域最重要的基准数据集之一。(2) 手绘对于司法领域的帮助 (Sketch for Forensics)。其中, SketchX 首次在 CVPR 2016 提出了目击证人在案发后不同时间段内, 对嫌疑人画像的不同记忆能力的建模方法<sup>[15]</sup>。(3) 手绘对于传统视觉问题的帮助 (Sketch+Vision)。例如, 在使用少量样本 (few-shot learning) 学习识别模型这一问题中, SketchX 摆脱了过去必须用语言无监督模型中学习到的类别向量 (word vector) 作为唯一先验信息这一单一范式, 实现了用最少仅一个手绘来代表类别信息, 并取得了不错的结果。相关工作被 CVPR 2018 接收为 spotlight 汇报<sup>[16]</sup>。在近期, SketchX 还将类似的想法拓展到了图像语义分割领域。

手绘识别代表性论文: [1] Qian Yu, Yongxin Yang, Yi-Zhe Song, Tao Xiang, T. M. Hospedales, Sketch-a-Net that Beats Humans, BMVC 2015

手绘识别是手绘理解的基本任务之一，其本质是从抽象的线条组合中提取语义信息。在文献[1]的工作中，课题组充分考虑手绘在视觉表达上的稀疏性以及特有的时序信息，设计了首个针对手绘的深度神经网络 Sketch-a-Net，实现了在基准数据集上超越人类的识别准确率。

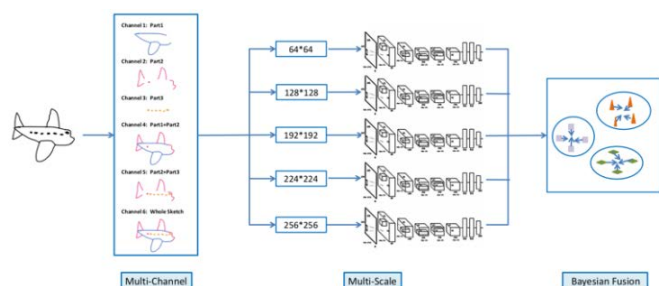


图 3：首个针对手绘的深度网络 Sketch-a-Net<sup>[1]</sup>

如图 3 所示，Sketch-a-Net 是一个多通道多尺度神经网络。其中，多通道的输入数据有效地建模了手绘的时序信息。在分析了大量手绘数据的基础上，课题组发现人们在绘画的时序上遵循相似的规律，即先轮廓后细节。基于这一发现，团队设计了相应的算法对手绘进行分解，并作为不同通道送入网络。其次，多尺度的网络设计有针对性地解决了手绘不同绘画风格和抽象程度的问题。将输入数据缩放到不同尺度以模拟不同的抽象程度，每个尺度的数据都由一个专门的网络负责。然后不同尺度的网络提取的特征经过贝叶斯融合得到一个更有效的特征表达，最终用于分类任务。

在该工作的拓展工作<sup>[17]</sup>中，课题组进一步对时序信息进行探索。通过结合手绘的时序和笔画长度信息，以及借鉴图像领域的变换方法，课题组设计了两个针对手绘数据的扩充算法。这一方面有效地扩充了训练数据集，另一方面也简化了多通道训练数据带来的运算量大的问题。该拓展工作在基准数据集上的识别准确率相比人类高出近 5 个百分点。这两个工作中的网络设计理念以及对手绘时序信息的探索启发了大量后续的研究。（北京航空航天大学于茜副教授撰稿）

手绘匹配代表性论文: [7] Qian Yu, Feng Liu, Yi-Zhe Song, Tao Xiang, T.M. Hospedales, Chen Change Loy, CVPR 2016; [11] Ayan K.Bhunia, Yongxin Yang, T.M. Hospedales, Tao Xiang, Yi-Zhe Song, CVPR 2020; [12] Kaiyue Pang, Ke Li, Yongxin Yang, Honggang Zhang, T.M. Hospedales, Tao Xiang, Yi-Zhe Song, CVPR 2019.

不同于传统的基于类别的手绘匹配，课题组在 2016 年提出了基于深度尺度学习框架的手绘细粒度图像检索<sup>[7]</sup>（图 4），旨在利用手绘草图的细节信息来精确匹配目标图像。所贡献的深度三元模型 -- 将相似的手绘图像在高维向量空间内推近，并将相异的拉远 -- 迄今为止还被认为是最有效的 FG-SBIR 模型之一。此外，团队还首次贡献了一个包含 1432 个手绘草图和对应图像的精粒度手绘匹配数据集（鞋子和椅子），并提供了包括 32000 个三元组排序手工标注信息，使 FG-SBIR 的深度学习成为可能。



图 4：基于手绘的细粒度图像检索(FG-SBIR)<sup>[7]</sup>

然而上述的 FG-SBIR 模型要求一个完整的手绘输入。这也就需要用户有一定的耐心和绘画功力，并在一定程度阻碍了其实用性。因此，课题组又对如何降低手绘门槛进行了深入探索，提出了一个基于增强学习的跨域检索模型<sup>[11]</sup>，旨在用尽量少的手绘笔画即能取得最优的检索效果。课题组为此设计了新颖的奖惩回馈机制，使在整个绘画周期每一笔都被考虑，并给予早期笔画以更大的权重。同时在相邻笔画间建立联系，使时序上的渐进搜索更加协调。图 5 展示了该模型相对于文献 [7] 的优越性。此工作被录为 CVPR 2020 的口述报告。

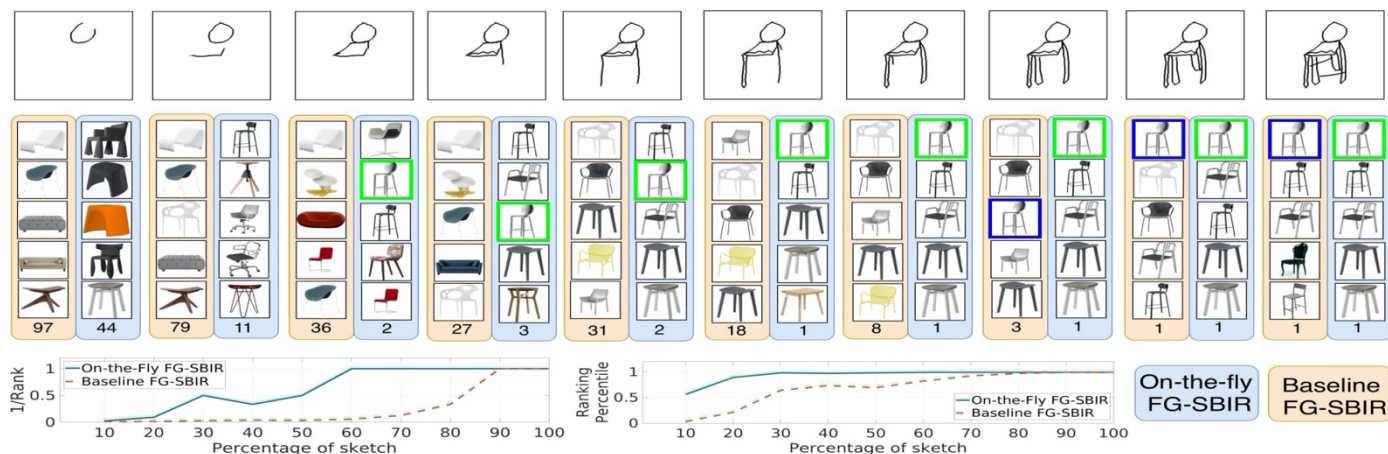


图 5: 基于增强学习的实时 FG-SBIR 模型<sup>[11]</sup>可以使用户在最少仅需几笔即可以完成细粒度搜索

此外, FG-SBIR 问题由来已久的一个痛点是缺少有标注的数据并且收集成本极高。因此, 课题组进一步研究了如何增强模型的域泛化能力<sup>[12]</sup>, 期望模型在已有的标注数据集上训练三元模型后, 在新的目标数据集上不需要任何标注数据即可进行模型参数更新。这一工作最关键的想想法是学习一个辅助的无监督嵌入网络, 将任何给定的草图映射到原型草图特征或多重嵌入的通用字典中, 这样能得到一个更加鲁棒的草图特征表达。大量实验证明, 这一特征提取方法可以有效地使三元模型的泛化能力得到提升, 以至应用于新的目标类别匹配时取得了领先的效果。(北京邮电大学齐勇刚副教授撰稿)

手绘+视觉代表性论文: [13] Da Li, Yongxin Yang, Yi-Zhe Song, T.M. Hospedales, ICCV 2017; [14] Da Li, Jianshu Zhang, Yongxin Yang, Cong Liu, Yi-Zhe Song, T.M. Hospedales, ICCV 2019

在机器学习中, 领域泛化是为了让模型能学习到训练和测试数据分布差异的鲁棒性。先前的领域泛化致力于从图像到图像领域的泛化研究, 对算法评测有着局限性。在从手绘到图像的检索课题中, 课题组已知手绘与

图像之间的差异给检索带来了挑战。因此, 为了更全面地评估机器模型的识别能力在领域泛化中的表现, 课题组将手绘引入到了领域泛化中, 提出了 PACS 数据集<sup>[13]</sup>。在 PACS 上几乎所有之前的最先进的算法, 在从图像泛化到手绘的识别中性能都有所下降。为了提升泛化性能, 不同于现有算法使用预提取的数据特征, 课题组首次使用了端到端的训练方式, 提出了从训练领域模型中学习根本领域模型的低秩参数化算法<sup>[13]</sup>。该方法的特点是:

(1) 根本领域模型学习到了独立于训练领域的类别信息, 泛化到未知的测试领域。(2) 低秩的参数化基础模型降低了训练参数量。最近, 受到小样本 (few-shot) 学习的启发, 课题组提出了基于模拟领域泛化情节的 (episodic training) 学习方法, 取得了 PACS 数据库上最好的结果<sup>[14]</sup>。该算法在特征及分类器学习中分别加入了模拟领域泛化的学习约束, 使训练的模型对领域差异有很好的鲁棒性。同时, 该方法也适用于不同的基础模型, 并有着比先前方法更好的训练效率。(英国剑桥三星人工智能中心李达研究员撰稿)

责任编辑 张汗灵

## 参考文献

- [1] Yu Q, Yang Y, Song YZ, Xiang T, Hospedales T. Sketch-a-net that beats humans. In Proceedings of the British Machine Vision Conference (BMVC), 2015 (Best Science Paper)
- [2] Qi Y, Song YZ, Xiang T, Zhang H, Hospedales T, Li Y, Guo J. Making better use of edges via perceptual grouping. In

- Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR) 2015
- [3] Li K, Pang K, Song J, Song YZ, Xiang T, Hospedales TM, Zhang H. Universal sketch perceptual grouping. In Proceedings of the European Conference on Computer Vision (ECCV) 2018
  - [4] Riaz Muhammad U, Yang Y, Song YZ, Xiang T, Hospedales TM. Learning deep sketch abstraction. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018
  - [5] Riaz Muhammad U, Yang Y, Hospedales TM, Xiang T, Song YZ. Goal-driven sequential data abstraction. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2019
  - [6] Li Y, Hospedales TM, Song YZ, Gong S. Fine-grained sketch-based image retrieval by matching deformable part models. . In Proceedings of the British Machine Vision Conference (BMVC), 2014
  - [7] Yu Q, Liu F, Song YZ, Xiang T, Hospedales TM, Loy CC. Sketch me that shoe. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 (Oral Presentation)
  - [8] Song J, Yu Q, Song YZ, Xiang T, Hospedales TM. Deep spatial-semantic attention for fine-grained sketch-based image retrieval. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2017
  - [9] Pang K, Li D, Song J, Song YZ, Xiang T, Hospedales TM. Deep factorised inverse-sketching. In Proceedings of the European Conference on Computer Vision (ECCV) 2018
  - [10] Pang K, Yang Y, Hospedales TM, Xiang T, Song YZ. Solving mixed-modal jigsaw puzzle for fine-grained sketch-based image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020
  - [11] Bhunia AK, Yang Y, Hospedales TM, Xiang T, Song YZ. Sketch Less for More: On-the-Fly Fine-Grained Sketch Based Image Retrieval. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2020 (Oral Presentation)
  - [12] Pang K, Li K, Yang Y, Zhang H, Hospedales TM, Xiang T, Song YZ. Generalising fine-grained sketch-based image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2019
  - [13] Li D, Yang Y, Song YZ, Hospedales TM. Deeper, broader and artier domain generalization. In Proceedings of the IEEE international conference on computer vision (ICCV) 2017
  - [14] Li D, Zhang J, Yang Y, Liu C, Song YZ, Hospedales TM. Episodic training for domain generalization. In Proceedings of the IEEE International Conference on Computer Vision (ICCV) 2019 (Oral Presentation)
  - [15] Ouyang S, Hospedales TM, Song YZ, Li X. Forgetmenot: Memory-aware forensic facial sketch matching. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2016 (Spotlight Presentation)
  - [16] Hu C, Li D, Song YZ, Xiang T, Hospedales TM. Sketch-a-classifier: sketch-based photo classifier generation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2018 (Spotlight Presentation)
  - [17] Yu Q, Yang Y, Feng L, Song YZ, Xiang T, Hospedales TM. Sketch-a-Net: A Deep Neural Network that Beats Humans. International Journal of Computer Vision (IJCV) 2017



## 宋一哲 (Yi-Zhe Song)

宋一哲是英国萨里大学 (University of Surrey) 视觉语音信号中心 (CVSSP) 教授 (Reader), SketchX 实验室创始人。分别于 2003 年和 2009 年在英国巴斯大学获得理学一等学位和计算机科学的博士学位, 在取得博士研究学位之前, 宋教授在英国剑桥大学获得了计算机科学硕士学位, 其毕业论文被计算机实验室 (Computer Laboratory) 授予当年的杰出论文奖。主要研究领域包括计算机视觉, 计算机图形学, 模式识别, 机器学习。发表论文 90 余篇, 如 ACM TOG, IJCV, IEEE TIP, IEEE TVCG 等国际知名期刊, 和 CVPR, ICCV, ECCV, AAAI 等顶级国际会议, 曾获 2015 年 BMVC 最佳论文奖 (Best Science Paper Award)。宋教授还是英国萨里大学和英国玛丽女王大学人工智能研究生课程的发起人和领头人。目前, 宋教授是 Higher Education Academy 会士, IEEE 高级会员, 以及 BMVA (英国机器视觉协会) 的成员。宋教授还是英国工程和物理科学基金委 (EPSRC) 评审委员会委员, 主持了累计价值超过 200 万英镑的多个项目。

## 征文通知

### 1 会议征文

计算机视觉领域相关国内外会议的征文通知如表 1 所示。同时，可继续关注每个会议举办的 workshop 或 special session。

### 2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 Image and Vision Computing 和 Pattern Recognition。

### 3 会议简介

亚洲计算机视觉会议 (Asian Conference on

Computer Vision) 是亚洲计算机视觉联盟(AFCV)举办。1993 年举办第一届，每两年举办一次。ACCV 为中国计算机学会 CCF 推荐人工智能会议。论文录取率 20%~25%，是仅次于计算机视觉三大顶会的会议，近年学术水平及等级进一步提高。

2018 年 ACCV 在澳大利亚珀斯举行，2020 年和 2022 年将在日本京都和中国澳门举行。本届会议将主要汇聚亚洲从事计算机视觉理论与应用研究的广大科研工作者及工业界同仁，共同分享计算机视觉领域的最新理论和技术成果，为大家提供精彩的学术盛宴。

责任编辑 刘帅奇

**表 1 计算机视觉领域相关国内外会议**

会议名称	会议时间	会议地点	截稿日期	会议网站
ACCV 2020	2020.11.30-04	Kyoto, Japan	2020.07.08	<a href="http://accv2020.kyoto/">http://accv2020.kyoto/</a>
ICPR 2020	2020.01.10-15	Milan, Italy	2020.07.11	<a href="https://www.micc.unifi.it/icpr2020/">https://www.micc.unifi.it/icpr2020/</a>
3DV 2020	2020.11.25-28	Fukuoka, Japan	2020.08.01	<a href="http://3dv2020.dgcv.nii.ac.jp/index.html">http://3dv2020.dgcv.nii.ac.jp/index.html</a>
WACV 2021	2021.01.05-06	Waikoloa, USA	2020.08.26	<a href="http://wacv2021.thevcf.com/home">http://wacv2021.thevcf.com/home</a>
AAAI 2021	2021.02.02-09	Vancouver, Canada	2020.09.10	<a href="https://aaai.org/Conferences/AAAI-21/">https://aaai.org/Conferences/AAAI-21/</a>

**表 2 计算机视觉领域相关国内外期刊专刊**

期刊名称	专刊题目	投稿网址	截稿日期
PR	Deep Relational Learning In Visual Recognition Understanding and Reasoning	<a href="https://www.journals.elsevier.com/pattern-recognition/call-for-papers/deep-relational-learning-in-visual-recognition-understanding">https://www.journals.elsevier.com/pattern-recognition/call-for-papers/deep-relational-learning-in-visual-recognition-understanding</a>	2020.7.1
PR	Conformal and Probabilistic Prediction with Applications	<a href="https://www.journals.elsevier.com/pattern-recognition/call-for-papers/conformal-and-probabilistic-prediction-with-applications">https://www.journals.elsevier.com/pattern-recognition/call-for-papers/conformal-and-probabilistic-prediction-with-applications</a>	2020.7.30
IVC	Cross-Media Learning for Visual Question Answering (VQA)	<a href="https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/cross-media-learning-visual-question-answering">https://www.journals.elsevier.com/image-and-vision-computing/call-for-papers/cross-media-learning-visual-question-answering</a>	2020.8.31
CVIU	Recent Advances in Modeling, Methodology and Applications of Action Recognition and Detection	<a href="https://www.journals.elsevier.com/computer-vision-and-image-understanding/call-for-papers/modeling-methodology-and-applications-of-action-recognition">https://www.journals.elsevier.com/computer-vision-and-image-understanding/call-for-papers/modeling-methodology-and-applications-of-action-recognition</a>	2020.9.15

# COMPUTER VISION NEWSLETTER

02 2020  
总第 24 期



## 计算机视觉专委会简报



CCF 计算机视觉  
专委会