主办 CCF计算机视觉专业委员会

COMPUTER VISION NEWSLETTER

CCF 计算机视觉专委会简报

01 2022

总第31期

CCF 计算机视觉专委会

COMPUTER VISION NEWSLETTER

计算机视觉专委会简报

2022 年第 01 期

总第 31 期

主　办　CCF 计算机视觉专业委员会

编委会

CONTENTS
简报目录

CCF 计算机视觉专委会
CCFCV.CCF.ORG.CN
CCFCVN@GMail.com

CCF-CV 走进高校系列报告会

第 111 期 武汉理工大学

2022 年 1 月 7 日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、武汉理工大学承办的 CCF-CV 走进高校系列报告会第 111 期活动以线上线下结合方式成功举办。本次活动邀请了中国科学院自动化所研究院胡卫明研究员、北京大学彭宇新教授以及北京邮电大学杜军平教授做特邀报告与学术研讨交流。武汉理工大学计算机与人工智能学院熊盛武院长、朱安娜副教授担任本次活动执行主席。

活动伊始，武汉理工大学计算机与人工智能学院熊盛武院长致欢迎辞，他首先代表学院感谢 CCF-CV 专委会对武汉理工大学的信任，能够再次提供这样的交流平台；然后对三位嘉宾表示了热烈的欢迎和衷心的感谢，接着介绍了武汉理工大学与学院的基本情况；最后表示本次活动的开展一方面是为学习先进的计算机视觉相关领域的前沿技术和知识，另一方面是为了学术交流，希望各位专家能够帮助该领域的老师和学生提升技术水平，答疑解惑，指导他们如何深入学习和掌握计算机视觉相关前沿技术和创新应用，希望这类高水平、高创新、高收获的交流会能够多在武汉理工大学举办，并预祝本次报告会取得圆满成功！

会后，武汉理工大学计算机与人工智能学院袁景凌副院长进行了活动总结，再次感谢了三位专家的精彩报告与学术交流分享，其次感谢了线上和线下听众的热情参与和高质量提问，最后再次感谢 CCF-CV 专委会和武汉理工大学计算机人工智能学院对活动的大力支持！并欢迎各位专家学者能再次来武汉理工大学进行学术指导和交流！祝贺本次活动取得了圆满成功！

责任编委 毋立芳

第 12 期 智能内容创作（AIGC）前沿进展与未来趋势

CCF-CV 视界无限系列研讨会

2022 年 1 月 5 日，由中国计算机学会计算机视觉专委会联合百度视觉技术部、增强现实技术部举办的第 12 期 CCF-CV “视界无限” 系列活动——“智能内容创作（AIGC）前沿进展与未来趋势” 研讨会在百度科技园成功举办。研讨会邀请了百度集团副总裁吴甜、专委会副主任中科院自动化所王亮研究员致辞，中科院自动化所赫然研究员、北京大学刘家瑛副教授、北京航空航天大学刘偲教授、清华大学刘烨斌副教授、清华大学刘永进教授和微软亚洲研究院杨蛟龙研究员做主题报告。百度视觉技术部首席架构师王井东及以上六位讲者参与了深度研讨。计算机视觉专委会 B 站公众号对本次会议进行了全程直播，直播人气峰值达到 2000+。

百度集团副总裁吴甜在开场致辞中回顾并梳理了内容生产的发展历程，无论在互联网出现之前还是之后，内容的载体一直在升级变化，但人类社会对内容的需求是一直存在的。AIGC 的内容生产方式近年来备受关注，背后主要有两大推动力：一是需求层面，数字世界对大量高质量、个性化、有创意的内容需求日益增加；二是 AI 相关技术发展，如超大规模知识图谱、计算机视觉、语音、自然语言处理、生成大模型等技术的融合创新发展。吴甜表示，目前在内容创作过程中，AI 更多发挥的是辅助创作的作用，要想完全运用 AI 技术自主制作内容还有很长的路要走，也还有很多可以探索的工作。她提出对于 AI 技术而言，AIGC 目前面临着三大挑战，即模仿和超越人类的创意、增加对人类认知和逻辑理解多样性的准确判断、降低制作高精良内容的成本并缩短周期以实现规模化。

王亮副主任向大家介绍，本次研讨会是计算机视觉专委会举办的 “视界无限” 系列活动第十二期，既是 2022 年开年的第一场活动、也是第一次由企业承办的 “视界无限”，非常有意义。王老师指出， AIGC 智能内容创作是当前最前沿和活跃的研究方向，具有非常广

阔的应用前景，近年来该领域的研究和应用也开展得如火如荼。不过这一领域一方面发展得很快，另一方面也存在很多实际的困难和挑战。借由本次活动，大家可以与该领域的优秀专家学者就目前研究现状做一次梳理和交流，一起探索未来可能的发展趋势。最后，王亮副主任还对百度公司、研讨会组织者以及各位青年学者表达了感谢，并预祝本次会议成功举办。

赫然研究员的报告题目是“人脸图像合成与鉴别”。人脸图像合成是指使用深度学习等智能化技术对人脸图像数据进行修改、编辑和替换，进而创造出从身份内容或表观纹理上完全不同的图像。人脸图像合成及其鉴别是机器学习和计算机视觉等领域的重要研究内容之一，被广泛应用于影视媒体娱乐和人工智能安全，在国家公共安全领域具有重要研究意义。赫然研究员结合人脸图像合成的实际应用需求，介绍“合成”与“鉴别”相辅相成、“攻击”与“防御”相互促进的对抗博弈机制；从信息理论角度探寻深度合成的信息交换本质，介绍表象最优传输和信息瓶颈解表达等生成模型，以及全脸生成、属性编辑（年龄、光谱、样式）、身份交换、人脸重演（表情、姿态）、语音驱动人脸等合成方法。

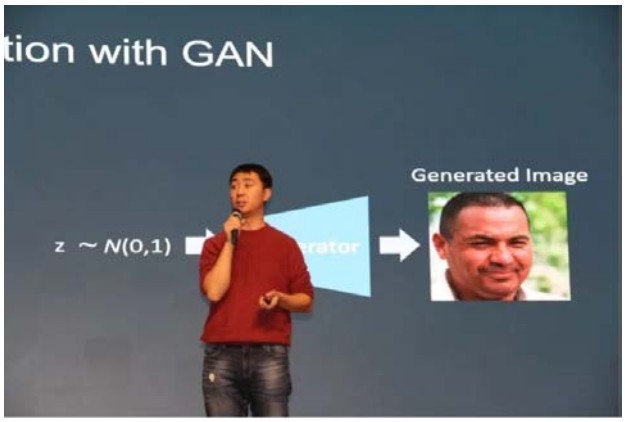
刘家瑛副教授的报告题目是“智能艺术生成与美学计算”。数据驱动下基于生成技术的影像与美学计算研究飞速发展，本次报告以图像风格化和生成技术在智能影像编辑与生成建模领域的应用为主线，分享了刘家瑛副教授研究小组围绕文字风格化艺术生成、基于手绘的人脸编辑等方面的探索，重点介绍基于生成对抗网络的文字风格化与去风格化技术、形状可变的风格化文字生成技术等一系列相关研究工作。

刘偲教授的报告题目是“多模态内容生成”。在报告中，刘偲教授介绍了视频自动配乐、语言指导的图像编辑两种多模态生成方法。在视频自动配乐方向上，刘偲教授团队首先建立了视频和背景乐的韵律关系，并提出了可控的音乐 Transformer 生成模型，能够按照用户指定的音乐流派和乐器，实现音乐和视频的一致性。在语言指导的图像编辑方向上，团队提出的方法主要包含 2 个模块：编辑描述模块，即可以给定任意一对图像，输出其之间的编辑嵌入表述；图像-语言注意力模块，即能够根据语言自适应地编辑图像。在这一方向上，刘偲教授团队也提出了一种新的评价标准。

刘烨斌副教授的报告题目是“数字人技术——交互性、沉浸性及创造性”。在元宇宙和人工智能热潮下，基于神经网络的数字人重建与生成技术受到学术界和工业界的广泛关注。围绕真实人物对象的三维重建、运动捕捉和智能生成成为构建现实世界和虚拟世界间的桥梁技术。此次报告围绕智能数字人的 3I 技术，即人的行为感知实现交互性（Interaction）、人的外观重建实现沉浸性（Immersion）、赋予人的思想实现创作性（Imagination）分别介绍了刘烨斌副教授在人体运动捕捉、人体动态三维重建、人体视频高质量生成等三方面的科研工作，涵盖人体、人脸、人手等相关视觉图形学前沿技术。报告同时对沉浸式全息通信技术、AI 数字人等前沿热点进行了展望和探讨。

刘永进教授的报告题目是“虚拟人智能体的人机共情研究”。共情是社会互动中个体理解他人立场、并产生与他人相似情感的心理推理形式，是人类合作行为的主要动机来源。目前虚拟形象技术研究在计算机图形学、计算机视觉和社交多媒体领域受到越来越广泛的重视，新颖的虚拟形象技术正在不断的提出。此次报告就人能否准确理解虚拟形象的情绪，并从让虚拟人智能体能够产生和人一致的情绪反应与让用户能够产生和机器一致的情绪反应两个角度开展研讨。

杨蛟龙研究员的报告题目是“跨越维度鸿沟——三维可控的图像生成研究”。二维图像是三维空间中物体的投影。目前生成对抗模型已经可以生成逼真的二维图



像，然而其生成过程并不感知所生成内容的三维形态，对生成内容也不能进行三维空间中的控制，如视角控制、形状变化等。本次报告中，杨蛟龙研究员以三维可控图像生成和编辑为主题，分享了其团队在基于解耦合 2D GAN 的三维可控人脸图像生成、基于神经辐射场（NeRF）的三维视角可控图像生成、单幅人像视角和表情编辑、多视图人体图像实时任意视角合成等一系列相关研究方向上的工作成果。

在 Panel 环节，与会嘉宾就“人脸编辑、迁移、风格化的技术发展、现状和挑战”“数字人相关技术应用趋势与难题”“视频配乐生成原理与发展方向”“对人机交互领域技术前景的预测”“智能创作领域的客观评价标准”“AI 驱动虚拟人到真人驱动虚拟人的距离”等问题展开热烈讨论，各位专家就上述问题分享各自的观点。

责任编委 杨巨峰

CCF-CV 走进高校系列报告会活动细则发布

自 2015 年 11 月以来，CCF-CV 走进高校系列报告会活动得到了高校、讲者、听众的大力支持。截至 2022 年 1 月底，CCF-CV 走进高校系列报告会活动已成功举办了 111 期，遍及中国大陆所有省、自治区、直辖市，并在中国澳门、澳大利亚悉尼等地成功举办。目前，已进行专题报告 456 场，活动现场平均听众 200 人/次，并于 2020 年开始启动线上活动，B 站人气峰值最高达 2.3 万，微信公众号平均阅读 1000 余次。在征得讲者同意的前提下，于专委会 B 站账号共享报告会的现场视频，并在专委会网站分享讲者的报告 PPT，在领域内引起强烈反响。CCF-CV 走进高校第 100 期特别活动得到学会和专委会领导的大力支持，活动纪念视频已于专委会 B 站账号发布。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！

结合前期活动的经验及各方反馈，我们拟进一步优化活动申请流程，加强讲者与承办单位的深度交流，充分发挥领域资深讲者对承办单位的指导作用。以下是修订后的活动细则，欢迎各位同行踊跃申请

1．宗旨：

为了更好地推动计算机视觉学科专业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拟在全国范围的高校和科研院所等开展 CCF-CV 走进高校系列报告会活动（以下简称“活动”）。

2．活动内容：

1）邀请报告（可选）：

邀请 2-4 名讲者作前沿学术报告；

2）研究交流（可选）：

a. 研究点评：承办单位若干名在读博士生或青年教师汇报自己的研究工作，专家进行点评，并提出研究建议；

b. 答疑解惑：讲者和承办单位师生以座谈会的方式进行交流，专家就承办单位师生提出的人才培养、科学研究等方面的问题进行解答并给出建议；

c. 现场参观：专家现场参观承办单位的研究工作，并提出研究建议；

d. 其他活动：承办单位根据实际需求提出，和秘书处协商确定。。

3．活动形式：

a. 活动由 CCF-CV 主办，CCF-CV 秘书处负责协调相关工作；

b. 每次活动由一家单位承办，主要负责安排论坛主题、联系讲者、推广宣传、场地预定、通知听众、撰写新闻稿等；

c. 每次活动邀请 2-4 名专家，活动内容包括邀请报告、研究交流两个环节；

d. 每月原则上组织不超过 2 次活动；

e. 原则上同一城市举办两次活动的间隔须在三个月以上（线上活动可以不受此限制）。

4．活动申请：

a. 活动采取自愿申请的方式，有意承办活动的单位须向秘书处提交书面申请；

b. 承办方需提前至少三个月向秘书处提出申请（活动申请可通过下面链接:

https://www.wjx.top/vj/tbZKkkL.aspx 或者扫描下面二维码填写，秘书处将在接到申请两周内商定并给予答复；

关于活动申请有任何问题，请联系 CCF-CV 秘书处活动联系人：

毋立芳 lfwu@bjut.edu.cn

杨巨峰 yangjufeng@nankai.edu.cn。

5．讲者邀请：

a. 活动讲者采取邀请为主、推荐辅助的方式，由秘书处与承办方共同完成，讲者确定后，由秘书处统一给讲者发正式邀请邮件；

b. 为了更好地宣传计算机视觉及相关领域的原创性工作，鼓励国内外学者将自己的工作以 “附件 1: CCF-CV 走进高校活动讲者信息表” 的形式发给秘书处活动联系人，我们将根据活动主题进行匹配推荐；

c. 最终讲者名单由活动执行主席确定。

6．活动准备与宣传：

a. 活动开始日期前至少 1 个月启动活动准备工作；

b. 活动的具体时间和地点由秘书处和承办方协商确定；

c. 秘书处联系讲者填好附件 1：CCF-CV 走进高校活动讲者信息表和个人照片（尺寸为 240*180 像素）等资料发给承办方，承办方参考附件 2: CCF-CV 走进高校宣传材料模板制作本次活动宣传材料；

d. 承办方须提前两周将制作完成的宣传材料和一张承办单位的横向照片发至专委会秘书处，以便在专委会网站和微信公众号上统一宣传。

7．活动现场：

a. 活动现场悬挂横幅 “CCF-CV 走进高校系列报告会（第 x 期）” ，电子横幅也可；

b. 活动开始前，承办方负责提醒讲者填写 “附件 3：CCF-CV 走进高校系列报告会讲者报告视频公开知情书”，其中甲方由活动执行主席代表专委签字；

c. 活动开始前，参会人员扫描以下二维码签到，签到信息根据承办方需要提供给承办方；

d. 活动开始前播放 CCF-CV 专委宣传片（见附件 4：CCF-CV 专委宣传片）、CCF-CV 走进高校活动 100 期纪念视频（见附件 5：CCF-CV 走进高校活动 100 期纪念视频）和 CCF-CV 宣传页 PPT（见附件 6：CCF-CV 宣传信息）。活动休息（茶歇）期间播放 CCF-CV 宣传页 PPT；

e. 活动期间，承办方须按照讲者的意愿安排录制现场活动的音视频。

8．活动总结：

a. 活动结束后两日内承办方须提交以下材料给秘书处：1）活动新闻稿；2）讲者签字版知情书扫描件；3）填写 CCF-CV 走进高校系列报告会活动总结，链接如下：https://www.wjx.top/vm/PRo8SF4.aspx；

b. 活动结束后两周内承办方须提交以下材料给秘书处：1）讲者 PPT；2）录像资料（与知情书一致），录像格式为 MP4 压缩格式，大小为 1280*720；3）签到表电子版和学生、教师、CCF 会员、CCF-CV 专委委员统计数据；

c. 秘书处负责把讲者同意分享的 PPT 放到专委网站 http://ccfcv.ccf.org.cn，并将讲者同意分享的报告视频放到专委 B 站账号：https://live.bilibili.com/22339632。

9．注意事项：

a. 每次活动的时间和地点由秘书处和承办方协商确定；

b. 承办方需承担活动场地、宣传等相关费用，支付讲者课时费以及承担讲者活动期间的本地交通和餐饮费用。对于有条件的承办方，可在上述费用基础上，额外承担讲者的城市间往来交通、本地住宿等费用。原则上讲者自己承担交通和住宿费用。具体费用承担情况，承办方、讲者和秘书处在活动前协商一致即可；

c. 承办方有义务为讲者出具官方的邀请信等证明材料，以便讲者报销差旅费用；

d. 活动开始至少一个月之前，秘书处与承办方确定讲者、场地等事宜；

e. 承办方负责录制现场活动的音视频，并在活动结束后将论坛的影音资料和活动报道等发送给秘书处，经审核后发布。

10．其他：

a. 该活动以学术交流为主要目的，严禁以活动名义进行非法集会，传播非学术内容；

b. 如果有赞助等商业活动，须与秘书处提前商定。

11．附加材料：

附件 1：CCF-CV 走进高校活动讲者信息表；

附件 2：CCF-CV 走进高校宣传材料模板；

附件 3：CCF-CV 走进高校系列报告会讲者报告视频公开知情书；

附件 4：CCF-CV 专委宣传片；

附件 5：CCF-CV 走进高校活动 100 期纪念视频。

附件 6：CCF-CV 宣传信息；

附件下载链接地址：

附件 1-3 可从 CCF-CV 专委网站下载：http://ccfcv.ccf.org.cn/ccfcv/xgzy/zjgxhdxl/。

附件 4 下载链接：

https://www.bilibili.com/video/BV1A64y187Hj

附件 5 下载链接：

https://www.bilibili.com/video/BV1nq4y1X7DV

附件 6 下载链接：

http://ccfcv.ccf.org.cn/ccfcv/xgzy/zwxczl/。

责任编委 黄岩

专题综述

基于 Transformer 的行人重识别研究与展望

大连理工大学人工智能学院 张平平

行人重识别（Person Re-identification, Re-ID）技术的目的是依据不同时间和地点拍摄的图像或视频内容，检索场景中的特定行人。由于该技术在安全社区、智能监控和刑事侦查等前沿应用中的重要性，近年来已经成为计算机视觉领域研究的热点问题之一。然而，在现实复杂场景下，由于其易受到摄像机视角变化、行人姿态变化、物体区域遮挡、图像低分辨率、行人图像未对齐等诸多因素的影响，精确高效的行人重识别仍是一项极具挑战性的研究课题。

在过去的十余年里，行人重识别的研究取得了很大的进展，并产生了一系列的相关任务，如基于图像的行人重识别(Image-based Re-ID)、基于视频的行人重识别 (Video-based Re-ID)、行人搜索 (Person Search) 等。这些任务的完成离不开图像或者视频信息的鲁棒特征表示和检索度量(策略)的提升。如何获得更加鲁棒的视觉表征是制约着行人重识别性能提升的关键。早期的行人重识别算法主要关注手工特征的提取和相似性度量的设计。随着深度学习技术的发展，越来越多的工作聚焦端到端地学习更具判别性的深度特征，主要瞄准设计更加复杂的深度卷积神经网络（Convolutional Neural Network, CNN)。然而，深度 CNN 是通过逐层堆叠卷积操作实现的，而卷积是一种局部操作，一个卷积层通常只会建模邻域像素之间的关系，并不能实现信息的全局建模，这严重制约着行人重识别的性能。近期，基于 Transformer 的模型[1]无论是在自然语言处理领域还是在计算机视觉领域均取得了优异的表现，其核心原因在于 Transformer 是基于自注意力的全局操作，可以建模所有元素之间的关系，从而普遍提升模型的全局感知能力。得益于此，已经有一些工作尝试使用 Transformer 模型完成行人重识别，并取得了优异的性能。本文将重点介绍近期基于 Transformer 的行人重识别相关研究进展和未来发展趋势。

一、基于纯Transformer的行人重识别

纯 Transformer 模型在行人重识别领域的代表性工作是发表在 ICCV2021 上的 TransRe-ID[2]，该工作直接借鉴 ViT[3]模型处理图像数据的思路，将行人图像分成多个图像块(例如 16x16 像素大小)，并把这些图像块作为序列输入标准的 Transformer 编码器中，提取行人的判别性特征用于行人重识别。此外，为了进一步增强特征的鲁棒性，该工作还设计了两个新的模块：(i)拼图模块(Jigsaw Patch Module, JPM)，通过移动和混洗操作对图像块的嵌入进行重新排列，使区域覆盖范围更加多样化，提高了行人重识别能力。(ii)侧信息嵌入(Side Information Embeddings, SIE)，通过插入可学习嵌入来融合非视觉线索，从而减轻对相机/视图变化的特征偏差。该工作的具体框架如图 1 所示。

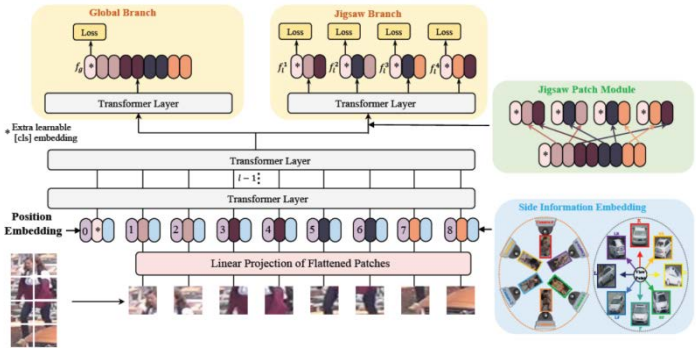


图 1　TransRe-ID 模型框架

总体而言，TransRe-ID 框架对 Transformer 的应用比较简单直接，没有考虑行人本身的结构化特点和空间的连续性，因而在行人出现遮挡或者不对齐时表现不佳。为此，学术界开展了一些列的改进工作。如，Sharma 等人[4]提出了一种局部感知 Transformer(Locally Aware-Transformer, LAT)，将全局增强的局部特征聚合到集成分类器中实现行人重识别，其具体框架如图 2 所示。为了提升对遮挡行人的表示能力，Zhao 等人[5]设计了一种基于局部特征的 Transformer(Partial Feature Transformer, PFT)用于行人重识别。其主要贡献是构建了图像块全维增强模块、融合重建模块和空间切片模块，显著提高了遮挡下的行人重识别性能。为了处理行人图像不对齐问题，Zhu 等人[6]首次在 Transformer 体系结构中引入了一种对齐方案，并提出了自动对齐的 Transformer(Auto-Aligned Transformer, AAformer)用于在图像块级别上自动定位行人和非行人部件。同时，将部件对齐集成到自注意模块中，输出的部件特征可以直接用于行人检索，其具体框架如图 3 所示。大量的数值实验验证了所学部件的有效性。此外，为了克服行人重识别任务对大量标注数据的依赖，Cao 等人[7]结合多标签分类法，将 ViT 应用于无监督行人重识别任务。实验结果表明，增强的 ViT 模型的性能普遍优于传统方法和大多数基于 CNN 的方法。为了降低初始伪标签的噪声，Xia 等人[8]基于 Transformer 设计了特征提取模型 Trans-Encoder。与传统的 CNN 相比，Trans-Encoder 提取的特征对跨域迁移具有更强的鲁棒性。在此基础上，可以提高特征聚类的置信度。

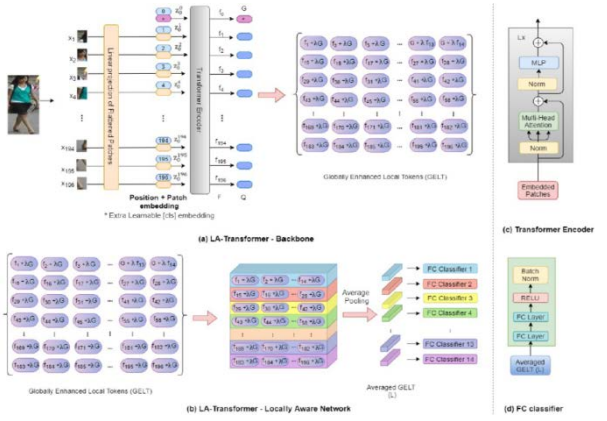


图 2　基于局部感知 Transformer 的行人重识别框架

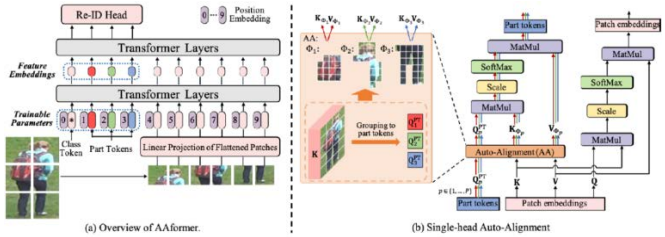


图 3　基于自动对齐 Transformer 的行人重识别框架

尽管基于纯 Transformer 的行人重识别方法取得了较大的成功，但是这类方法仍存在一些明显问题，如模型复杂度高、难以快速训练、在一定程度上忽略了行人图像的空间结构信息、无法在主流计算机设备上部署等。因而，基于纯 Transformer 的行人重识别模型仍有较大的提升空间。

二、CNN和Transformer耦合的行人重识别

在行人重识别任务中，利用判决性部件特征往往能带来额外的准确率提升，因而如何更好地学习行人的部件特征也是行人重识别成功的关键。众所周知，CNN 模型更注重局部特征的提取，因而天然地适合提取部件信息。然而，受制于有限的接受域，通常 CNN 提取的特征并不具有较强的全局特性，这又制约着 CNN 在行人重识别这一语义检索类任务上效能的发挥。而 Transformer 模型对空间和序列数据具有很强的长距依赖关系建模能力，天然地适合提取全局语义信息。因而，如何充分结合这两类模型的优势，构建更加精确和鲁棒的行人重识别模型也是当前研究的热门方向之一。而目前基于 CNN 和 Transformer 耦合的行人重识别方法可以大致分为如下三类：

2.1. 底层 CNN+高层 Transformer 架构

针对 CNN 和 Transformer 提取特征的不同，最直接的融合提升方法就是首先使用 CNN 提取行人的底层视觉特征，然后利用 Transformer 的全局建模能力汇聚得到具有高级语义信息的检索特征。沿着这一思路，目前大部分基于 Transformer 的行人重识别算法均取得了优于主流 CNN 模型的性能。如在基于图像的行人重识别中，Zhou 等人[9]首先利用 ResNet-50 提取多尺度视觉特征，然后通过 Transformer 编码器聚合查询样本

的 k 近邻上下文信息，最终设计了一种重排序网络来预测查询样本和排名靠前的邻居样本之间的相关性。在 6 个常用的行人和车辆重识别数据集上进行了实验，验证了该方法的有效性。此外，由于目标行人经常被各种障碍物或其他人遮挡，为了解决这些问题，Li 等人[10]提出了一种端到端的部件感知 Transformer(Part-Aware Transformer)，通过 CNN 提取图像视觉信息，同时构建像素上下文 Transformer 和部件原型 Transformer 挖掘不同的部件信息，实现了对被遮挡行人的不同部位的重识别，其具体框架如图 4 所示。Jia 等人[11]利用 CNN 和 Transformer 架构，通过对被遮挡行人图像的局部特征进行全局推理，实现了无对齐的行人重识别。为了实现跨域的行人重识别，Waseem 等人[12]提出了一种使用 CNN 混合视觉 Transformer 的域适应方法，并将聚类损失函数和广泛使用的三重损失函数合并在一起，改善了现有的无监督领域自适应行人重识别方法的性能。

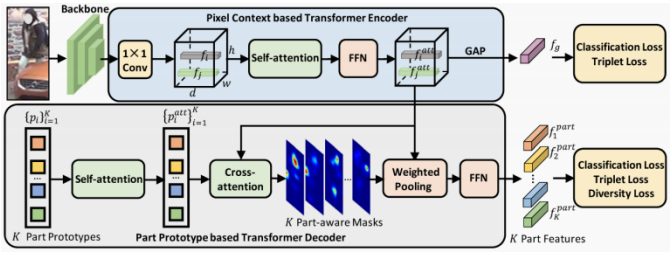


图 4 基于部件感知 Transformer 的行人重识别框架

针对基于视频的行人重识别，为了获取更丰富的感知信息和提取更全面的视频表示，Liu 等人[13]提出了一种基于多视角学习的三叉 Transformer(Trigeminal Transformer，TMT)框架，如图 5 所示。具体来说，该工作首先将原始视频数据通过 CNN 联合转换为空间、时间和时空域特征，然后利用三种自我视图的 Transformer 来增强空间、时间和时空域的信息。此外，还提出了一个交叉视图 Transformer 来聚合多视图特征，以实现全面的视频表示。实验结果表明，在公开的三个数据集上，该方法可以获得比其他同时期最先进的方法更好的性能。此外，He 等人[14]发现有效地提取多尺度细粒度特征并构建它们之间的结构交互是基于视频行人重识别成功的关键。因此，他们提出了一个混合框架，即密集交互学习(Dense Interaction Learning, DenseIL)，它综合利用了 CNN 和 Transformer 架构的主要优点。如图 6 所示，DenseIL 包含一个 CNN 编码器和一个密集交互 Transformer 解码器。CNN 编码器负责有效地提取空间特征，而 Transformer 解码器被设计成密集地模拟跨帧的时空固有交互作用。

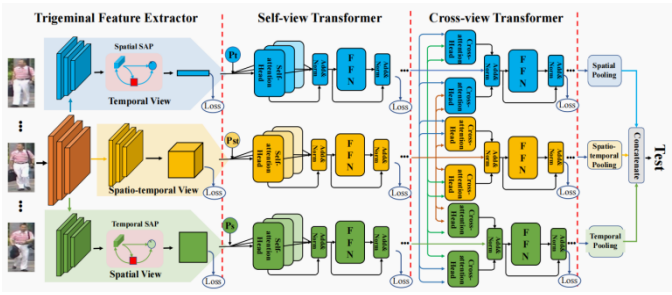


图 5 基于三叉 Transformer 的视频行人重识别框架

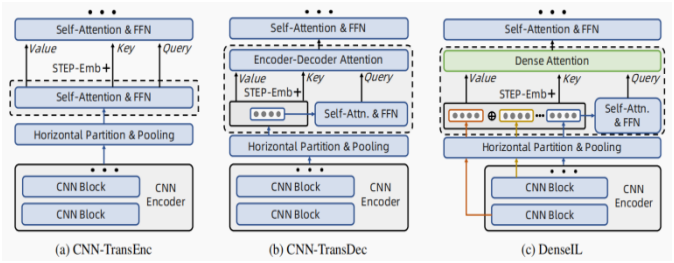


图 6 基于密集交互学习的视频行人重识别框架

2.2. 深度多层次耦合

前面所述的行人重识别方法主要是将 CNN 和 Transformer 作为独立的功能模块实现鲁棒的特征表示。事实上，它们忽略了 CNN 和 Transformer 均是层次化的表示模型，不同的卷积层和 Transformer 层可以表示不同的抽象信息。为了实现更加丰富和充分的信息融合，一些研究者尝试从深度多层次耦合的角度实现 CNN 和 Transformer 的互补，并提升行人重识别的性能。如，Tahir 等人[15]直接采用了三种不同的 CNN 网络结构，将 Transformer 模块插入到 CNN 的不同层，构建了新的主干网实现行人重识别。Zhang 等人[16]利用 CNN 和 Transformer 的层次化特点，提出了一种基于层次化聚合的 Transformer(Hierarchical Aggregation Transformer, HAT)框架用于图像的行人重识别，其结构如图 7 所示。为了解决行人裁剪后的轨迹时空偏差，Liu 等人[17]利用改进的 Transformer 构建了逐层的由粗到细轴向注意网络。该工作不仅能显著降低计算量，而且无需考虑空间和时间对齐以及数据集噪声的影响。

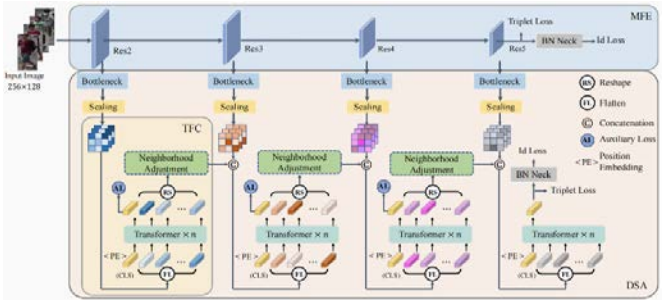


图 7 基于层次化聚合 Transformer 的行人重识别框

高性能的行人重识别要求模型同时关注行人的全局轮廓和局部细节。为了提取更具有代表性的特征，一种有效的方法是利用具有多个分支的深度模型。然而，大多数基于多分支的方法都是通过部分骨干结构的重复来实现的，通常会导致计算量的显著增加。为此，Zhang 等人[18]借鉴目前流行的特征金字塔网络(Feature Pyramid Network, FPN)，同时使用 Transformer 结构从不同的网络层提取全局特征，并将它们聚合成一个双向金字塔结构应用于行人重识别任务中。实验表明该方法取得了较好的识别效果。

2.3. 额外信息引导耦合

在行人重识别的现实应用中，额外的关联信息如摄像机信息、行人姿态、语言描述、属性标签等也能起到关键的作用。因而，如何利用这些额外信息或者在这些信息的引导下实现 CNN 和 Transformer 的耦合，进而提升模型的性能也是目前研究的一大趋势。为了从跨摄像机非配对训练数据中学习摄像机的不变表示，Ge 等人[19]提出了一种基于摄像机引导的 Transformer 行人重识别框架。该工作通过 CNN 变换伪跨摄像机正特征对，最小化伪特征对的距离，从摄像机特定的特征分布中挖掘出跨摄像机的自监督信息。此外，同步利用 Transformer 实现局部特征的自动定位和提取，进而实现超远距离的行人重识别。鉴于行人的姿态信息在识别过程中也扮演着重要作用，Ma 等人[20]提出了一种姿态引导的部件间和部件内关系 Transformer(Pose-guided Inter- and Intra-part Relational Transformer, PIRT)，其框架如图 8 所示。该工作通过引入 CNN 来生成姿态掩码，Transformer 来建立局部感知的长距相关性，实现了模型耦合性能的提升。类似地，Wang 等人[21]构建了基于姿态引导的特征解纠缠方法。该工作利用 Transformer 提取全局视觉特征，在姿态引导的特征聚合模块中利用匹配和分布机制，初步将姿态信息与图像块信息分离，实现了遮挡条件下的行人重识别。

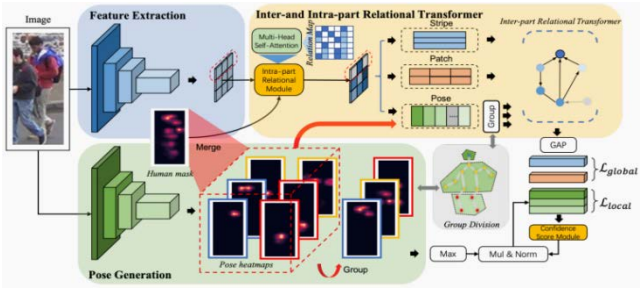


图 8 基于姿态引导耦合的行人重识别框架

预训练是计算机视觉的主导范式。为了寻找传统的预训练替代方法，Xiang 等人[22]提出了一种基于文本引导的行人重识别骨干网络预训练方法。该方法使用 CNN 提取图像的视觉特征，同时利用 Transformer 从文本标注中学习视觉表示，从而实现了 CNN 和 Transformer 的耦合学习。在基准测试集上进行的实验表明，与在 ImageNet 上预训练的模型相比，该方法可以取得极具竞争力的性能，揭示了它在行人重识别任务上的潜力。

三、总结与展望

本文介绍了近期基于 Transformer 的行人重识别方法，包括基于纯 Transformer 的模型以及基于 CNN 和 Transformer 耦合的模型。相关工作表明，有效地构建 CNN 和 Transformer 耦合方法对于提升行人重识别的性能至关重要。未来的发展方向包括以下几个方面：如何将 Transformer 这一全局/长距建模能力很强的模型应用于其他视觉模态以及与服装无关的生物特征上，从而实现多模态、跨媒体、超长时的行人重识别；如何将行人重识别与目标检测、多目标跟踪、行人分割等相关任务进行联合建模，从而发挥多任务学习的优势，促进视觉任务的有机融合;如何设计精确且高效的轻量化 Transformer 模型以及寻找经济和高效的训练方式(包括无监督、半监督、自监督等)；如何开发面向任务和用户的可解释 Transformer 网络，实现模型的可解释性。

责任编委 王金甲

参考文献

[1] Vaswani A, Shazeer N, Parmar N, et al. Attention is all you need[J]. NeurIPS 2017.
[2] He S, Luo H, Wang P, et al. Transreid: Transformer-based object re-identification[C]. ICCV 2021.
[3] Dosovitskiy A, Beyer L, Kolesnikov A, et al. An image is worth 16x16 words: Transformers for image recognition at scale[C]. ICLR 2021.
[4] Sharma C, Kapil S R, Chapman D. Person re-identification with a locally aware transformer[J]. arXiv:2106.03720, 2021.
[5] Zhao Y, Zhu S, Wang D, et al. Short Range Correlation Transformer for Occluded Person Re-Identification[J]. arXiv:2201.01090, 2022.
[6] Zhu K, Guo H, Zhang S, et al. Aaformer: Auto-aligned transformer for person re-identification[J]. arXiv:2104.00921, 2021.
[7] Cao G, Jo K H. Unsupervised Person Re-Identification with Transformer-based Network for Intelligent Surveillance Systems[C]. ISIE 2021.
[8] Xia L, Yu Z, Ma W, et al. Refining Pseudo Labels for Unsupervised Domain Adaptive Person Re-Identification[J]. IEEE Access 2021.
[9] Zhou Y, Wang Y, Chau L P. Moving Towards Centers: Re-ranking with Attention and Memory for Re-identification[J]. arXiv:2105.01447, 2021.
[10] Li Y, He J, Zhang T, et al. Diverse part discovery: Occluded person re-identification with part-aware transformer[C]. CVPR 2021.
[11] Jia M, Cheng X, Lu S, et al. Learning Disentangled Representation Implicitly via Transformer for Occluded Person Re-Identification[J]. IEEE TMM 2022.
[12] Waseem M D, Tahir M A, Durrani M N. Hybrid Vision Transformer for Domain Adaptable Person Re-identification[C]. ICCCI 2021.
[13] Liu X, Zhang P, Yu C, et al. A Video Is Worth Three Views: Trigeminal Transformers for Video-based Person Re-identification[J]. arXiv:2104.01745, 2021.
[14] He T, Jin X, Shen X, et al. Dense Interaction Learning for Video-based Person Re-identification[C]. CVPR 2021.
[15] Tahir M, Anwar S. Transformers in Pedestrian Image Retrieval and Person Re-Identification in a Multi-Camera Surveillance System[J]. Applied Sciences 2021.
[16] Zhang G, Zhang P, Qi J, et al. Hat: Hierarchical aggregation transformers for person re-identification[C]. ACM MM 2021.
[17] Liu C T, Chen J C, Chen C S, et al. Video-based Person Re-identification without Bells and Whistles[C]. CVPR 2021.
[18] Zhang S, Yin Z, Wu X, et al. FPB: Feature Pyramid Branch for Person Re-Identification[J]. arXiv:2108.01901, 2021.
[19] Ge W, Pan C, Wu A, et al. Cross-Camera Feature Prediction for Intra-Camera Supervised Person Re-identification across Distant Scenes[C]. ACM MM 2021.
[20] Ma Z, Zhao Y, Li J. Pose-guided Inter-and Intra-part Relational Transformer for Occluded Person Re-Identification[C]. ACM MM 2021.
[21] Wang T, Liu H, Song P, et al. Pose-guided Feature Disentangling for Occluded Person Re-identification Based on Transformer[J]. arXiv:2112.02466, 2021.
[22] Xiang S, Zhang Z, Guan M, et al. VTBR: Semantic-based Pretraining for Person Re-Identification[J]. arXiv:2110.05074, 2021.

张平平

大连理工大学人工智能学院副教授。主要研究方向是：计算机视觉、深度学习。目前已经在计算机视觉和人工智能相关领域的国内外著名学术期刊和会议发表论文 40 余篇，担任多个国际学术期刊和会议的审稿人及程序委员会委员。

Email: zhpp@dlut.edu.cn

热点追踪

抗姿态和遮挡的人脸生成与识别

重庆大学 段青言　张 磊

姿态变化和面部遮挡是影响人脸识别最主要的两个因素。对于姿态变化，通常采用抗姿态的特征表达和基于生成对抗网络（Generative Adversarial Net，GAN）的人脸正面化这两种方式进行解决。对于面部遮挡，基于 GAN 模型的人脸修复方法也层出不穷，这些方法更多地关注正面或近正面的人脸面部结构以及像素细节，而非身份判别性。可见，姿态变化和面部遮挡通常被当作两个单独的任务来分别加以解决。然而，在实际生活中，这两种情况常常同时发生，且逐渐演变成为一种富有挑战且有待研究的问题。如图 1 中，第一行和第二行的输入图片姿态角分别为 45°和 60°，其中，(a) 为输入的侧面遮挡人脸，(b) 为我们提出的 BoostGAN 的生成结果，(c) 和 (d) 分别为现有两种人脸正面化方法的生成结果，(e) 为真实的正面人脸。从图 1 可以看出，当侧面人脸出现部分遮挡时，(c) 和 (d) 作为仅针对姿态变化的人脸正面化方法出现了不同程度的生成误差。

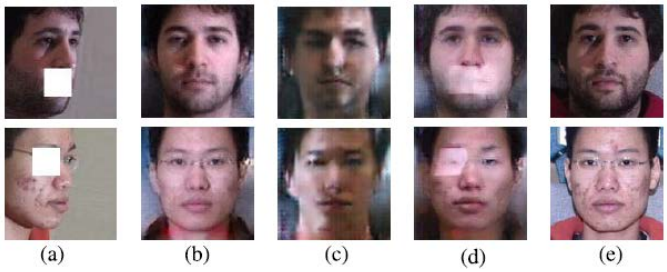


图 1 遮挡对现有人脸正面化方法的影响

解决姿态变化和面部遮挡的混合问题，直接的想法是分两步进行处理，即先采用人脸修复方法去遮挡，然后人脸正面化。然而，两步法容易产生较大的误差且依赖大量训练样本。我们提出的 BoostGAN 是一种端到端的集成式生成模型，如图 2 所示，采用多张局部遮挡的图片作为模型的输入，来完备身份和纹理信息。该网络由一个深度的编-解码器 (即粗糙网络) 和一个浅层的集成网络（即精细网络）组成。粗糙网络用于在多重遮挡和大姿态变化的人脸图像上实现粗糙的正面化和去遮挡生成。而精细网络旨在通过集成多个中间输出的互补信息，生成干净、正面的人脸图像并保持身份特异性。

更进一步，考虑到噪声作为先验知识被广泛应用至图像修复中，以及人脸修复和人脸正面化两个任务之间的协同作用，我们提出一种遮挡掩模引导下的两阶段生成对抗网络（TSGAN），如图 3 所示。该网络主要包含

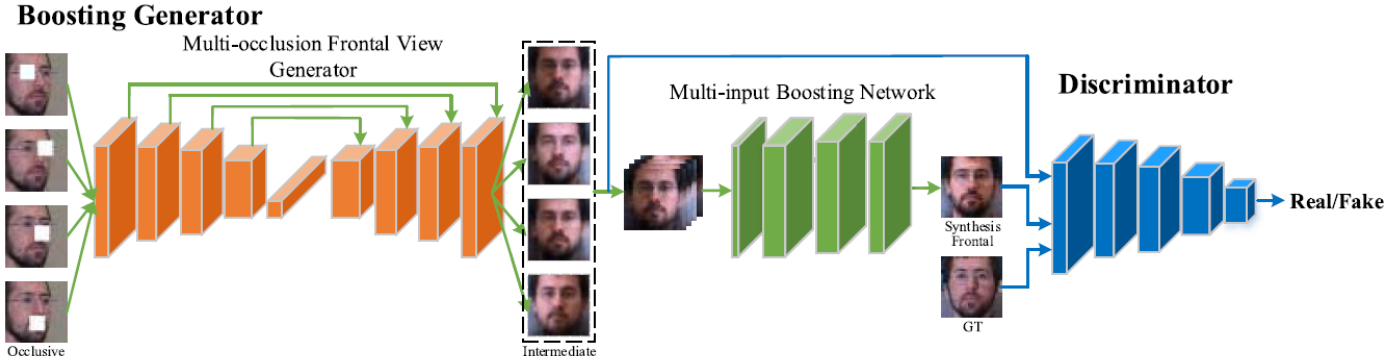


图 2 端到端由粗糙到精细生成的 BoostGAN 框架

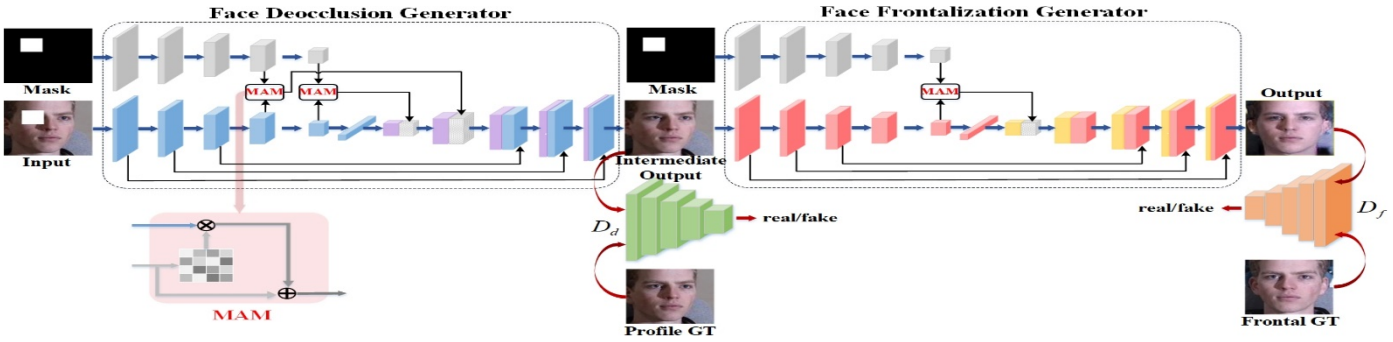


图 3 两阶段生成对抗网络（Two-Stage GAN，TSGAN）

3 个模块，即人脸去遮挡模块、人脸正面化模块和掩模注意力模块（Mask Attention Module, MAM）。前两个模块分别被设计用于不同的阶段，而 MAM 则在两个阶段中均被部署。在第一个阶段，作为一种重要的先验知识，引入遮挡掩模来拟合输入图像中的噪声，作为辅助信号帮助 TSGAN 完成人脸修复。MAM 使得人脸去遮挡模块更多的关注和更好地填充侧面人脸图像上的“空洞”，如图 4 所示。在第二阶段，由第一阶段生成的无遮挡侧面人脸作为人脸正面化模块的输入，通过 MAM 进一步获得最终逼真的正面图像。值得注意的是，TSGAN 模型是一个端到端的结构。此外，为了更有效地分别监督两个阶段保持身份的一致性和提高身份相关特征的判别性，提出针对去遮挡和正面化的双重三元损失来联合训练 TSGAN 的两个阶段。

在约束和非约束的人脸图像数据集上，定量和定性

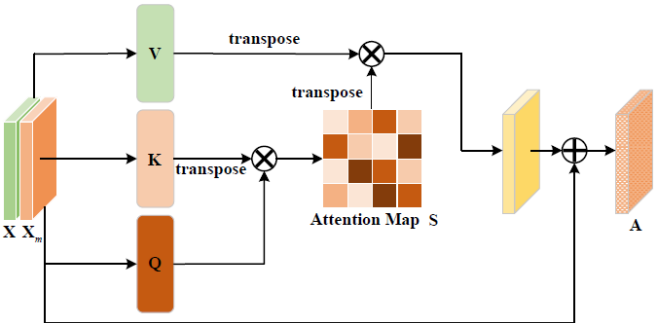


图 4 遮掩模注意力模块（MAM）

实验表明了提出的两个模型 BoostGAN 和 TSGAN 对遮挡、侧面人脸生成和识别的优越性，达到 SoTA。

以上 2 个成果分别被国际期刊 IEEE Transactions on Neural Networks and Learning Systems (2020) 和 IEEE Transactions on Circuits and Systems for Video Technology (2021)接收。

责任编委 储珺

段青言

2021 年 6 月博士毕业于重庆大学，现为重庆邮电大学讲师，主要研究方向为人脸识别、人脸生成、深度学习。

Email: duanqy@cqupt.edu.cn

张磊

重庆大学教授，博士生导师，IEEE/CCF 高级会员。主要研究方向为开放环境视觉感知、深度学习、迁移学习等。

Email: leizhang@cqu.edu.cn

顶会观察

NeurIPS 2021

南京大学人工智能学院副研究员 叶翰嘉

神经信息系统大会（Conference on Neural Information Processing Systems, NeurIPS）是机器学习领域的顶级会议之一，在国内外有广泛影响力，被评为 CCF-A 类会议。由于疫情影响，第 35 届 NeurIPS 大会于 2021 年 12 月 6 日至 14 日通过线上会议形式举办，包括 1 天的 Tutorial、4 天的正式会议以及 2 天的 Workshop。

一、 会议亮点

线上会议形式：NeurIPS 已连续两年通过纯线上会议形式举办。NeurIPS 2021 为注册者提供与其他参会者互动的权限，例如向演讲者提问、参加现场海报环节、与其他与会者聊天和交流等。与 2020 年类似，NeurIPS 2021 线上海报环节使用 GatherTown 系统，通过 Rocket.Chat 实现参会人员和作者的交流。2021 年度会议免费直播 Tutorial 和主题报告 Keynote，并在会议结束后提供所有视频。会议相关动态也随着会议的逐步进行在其官方博客（https://blog.neurips.cc）上同步。

审稿流程变化：首先，不同于往年使用 CMT 系统进行投稿、审稿，今年 NeurIPS 使用了 OpenReview 系统。为促进更加透明的审稿过程，所有录用论文的审稿意见公开（包括匿名审稿人的意见、最终 Meta-Reviewer 的意见以及作者的回复），并接收非匿名的评论，未录用论文的审稿意见由作者决定是否公开；其次，NeurIPS 2021 审稿去除了提前拒稿（Desk Reject）阶段。NeurIPS 2020 会议在开始审稿前，领域主席会对部分论文阅读后进行直接拒稿，而 2020 年审稿实验表明约有 6%被提前拒稿的论文在进行完整的审稿流程后能够被录用。因此 2021 年只针对违反投稿规则的论文进行提前拒稿；最后，2021 年审稿过程中允许审稿人和作者之间进行多轮回复。往年使用 CMT 系统时，作者在 Rebuttal 阶段后，无法继续跟进审稿人的后续问题，而在 OpenReview 系统中，作者和审稿人之间可以使用类似论坛的形式针对某些疑问进行反复讨论。

检查列表（Checklist）要求：不同于往年在投稿系统中进行一些检查项的选择，今年投稿中需要作者在文末附上检查列表。其设计主要考虑到论文的可复现、透明性，论文的研究伦理和社会影响等方面。目的是鼓励作者记录、思考论文的潜在问题和局限，并在写作时尝试有针对性地解决这些问题。希望作者填写对论文结果复现步骤的说明，是否附上代码和数据。Checklist 仅在投稿中使用，论文录用后不强制要求作为附录出现。

伦理审查改进：伦理审查造成独立于技术问题的挑战，NeurIPS 也不断尝试对伦理审查的过程进行改进。NeurIPS 于 2020 年已试行了伦理审查流程，而 2021 年伦理审查规模进一步扩大，有 100 多名伦理审稿人对论文审稿人筛选出的 265 篇论文（2.9%）进行了伦理审查。伦理审查目的是促进论文在伦理方面的思考，主要包括偏见和公平性问题、研究诚信问题、法律合规性、不适当的潜在应用和影响（如人权问题）等。

数据集投稿 Track：考虑到以往大量 NeurIPS 研究工作以算法为主，在寻找有效数据进行算法评估上有一定困难，且大量算法评估时只考虑到小型数据集，会产生有偏评估结果，NeurIPS 2021 新增了数据集和基准（Datasets and Benchmarks）投稿 Track，鼓励论文

对数据和任务基准进行高质量评估。作者投稿时需要在附录中说明数据集的收集、获取、维护、伦理等问题。对于数据集的评审和对于算法的评审有所差别，例如对于数据集的评审是单盲（single-blind）审稿，并且共分两轮投稿，审稿会专门考虑数据集是否能完整地对算法和问题进行评估和描述。数据集 Track 录用的论文也会在会议中进行报告，并被收录到 NeurIPS 单独文集。

二、 录用情况

NeurIPS 2021 共收到 9122 篇有效投稿，最终接收 2344 篇论文，接收率为 26%，达到近几年来最高水平。相较于 2020 年，2021 年投稿量降低 3.5%，录用率提升 23.4%。录用论文中有 56 篇 Oral 论文（2.39%），282 篇 Spotlight 论文（12.03%），其占比相对于 2020 年度有所降低。所有评审意见的平均分为 6.36（审稿在 1-10 区间打分，10 分为满分），Oral、Spotlight 和 Poster 论文的平均得分分别为 7.56、7.01 和 6.36。

录用论文的研究热点方向包括强化学习、深度学习、表示学习、优化、图神经网络、Transformer、对抗/鲁棒性等。谷歌（Google）、斯坦福大学、麻省理工学院、卡耐基梅隆大学、加州大学伯克利分校、微软等科研机构、高校院所录用论文较多。NeurIPS 2021 继续进行审稿实验，共随机抽取 298 篇最终录用的论文（包括录用为 Oral、 Spotlight 和 Poster 的论文），先后进行了两组评审。两组评审后，只有 77 篇（26%）论文得到了一致的评审意见，而剩余的 221 篇论文第二组评分均有下降，其中有 199 篇被第二组评审评为拒稿。

数据集 Track 共有两轮征稿，收到 484 篇论文，录用 174 篇（录用率 35.9%），其中关于计算机视觉的论文最多（占 20%），其他热门主题包括自然语言处理（15%）、强化学习和模拟环境（15%）、语音识别（7%）和多模态数据（6%），也有约 15%的论文关于元分析以及 AI 公平性。所有论文中，有 55%的论文提出新的数据集，20%的论文提出评测基准，25%的论文两者皆有。

三、 邀请报告

NeurIPS 2021 邀请了普林斯顿大学名誉教授、美国科学院院士、诺贝尔经济学奖得主 Daniel Kahneman 进行关于人类和机器智能的访谈，也邀请了 7 位专家进行报告（Invited Talk），具体内容如下：

How Duolingo Uses AI to Assess, Engage and Teach Better. 多邻国（Duolingo）联合创始人和首席执行官 Luis von Ahn 介绍了 Duolingo 的 AI 使用情况。Duolingo 是世界上最受欢迎的语言学习工具之一，报告介绍了多邻国是如何利用人工智能技术增加用户粘性，提高学习效果。例如，应用 AI 技术向用户推荐练习测试以提升语言的学习效率，以及应用 AI 对语言翻译进行评分、针对不同方面进行语言能力测试等。

The Banality of Scale: A Theory on the Limits of Modeling Bias and Fairness Frameworks for Social Justice (and other lessons from the Pandemic). 微软研究院的高级首席研究员 Mary L. Gray 介绍了为社会正义构建偏见和公平性框架的局限性理论。报告通过为北卡罗来纳州边缘化黑人和拉丁裔社区提供服务的社区组织（CBO）的实践案例，分析机器学习模型在公正和包容性抽样方面的局限性。目前在衡量数据和决策系统中的偏见和公平性方面理论研究的不足，使我们更加关注于收集数据的价值，而非对社区的价值。报告主要论证关注收集社区成员数据的需求，并观察计算上难以衡量但在质量上无价的社会作用对于推进社会公正的机器学习的必要性，以及对如何将计算机科学和机器学习重新定位为更明确的数据权力共享理论和实践提出建议。

Do We Know How to Estimate the Mean? 巴塞罗那庞培法布拉大学经济与商业系的 ICREA 研究教授 Gábor Lugosi 在报告中讨论统计学中一个基本的问题——基于独立观测的均值估计。随着机器学习和数据科学应用的发展，可以从统计和计算的新角度看待这个问题。通过回顾最近关于平均估计器的统计性能的一些结果，发现这些估计器允许高维数据中存在严重的长尾和对抗性污染。

Benign Overfitting. 加州大学伯克利分校电子工程系和计算机科学系教授、澳大利亚科学院院士 Peter Bartlett 在报告中探讨了模型过拟合的问题。深度学习

即使没有任何显式的努力来控制模型的复杂性，模型仍然完美地拟合有噪声的训练数据，并能展现出优秀的性能。报告介绍了在概率环境下如何进行准确预测的方法。对于线性回归问题，最小范数插值预测具有近似最优的预测精度。在这种情况下，过度参数化对于良性过拟合是必不可少的，即参数空间中对预测不重要的方向的数量必须显著超过样本大小。报告还讨论了对抗样本对鲁棒性的影响，并介绍了岭回归的扩展以及通过依赖于模型的泛化界限来分析良性过拟合的障碍。

Optimal Transport: Past, Present, and Future. 苏黎世联邦理工学院 FIM 数学研究所的讲座教授和主任、2018 年菲尔兹奖得主 Alessio Figalli 报告了关于最优运输的相关内容。在 18 世纪末，Gaspard Monge 为了得到如何将材料从一个地方运输到另一个地方来建造防御工事的最优解，提出了最优运输问题。在过去 30 年里，这个理论衍生出不同形式，在数学的许多领域都有应用。目前最优运输问题在机器学习诸多领域中大放异彩。报告内容概述了最优运输问题及应用，详细介绍了最优运输在金融、随机矩阵、生成式对抗网络、单细胞基因组等问题的应用，以及最优运输的复杂度分析。

Gender, Allyship & Public Interest Technology. 纽约大学 Arthur L. Carter 新闻学院的教授 Meredith Broussard 探讨了为什么大型计算机系统被困在 1950 年代关于性别的观念中，以及更新社会技术系统需要什么。考虑到从 2021 年 10 月开始，X 正式成为美国护照上的性别选项，而为了适应这种更具包容性的性别选择，需要对如何进行计算的更改进行探讨。报告也同时探讨了如何利用公益科技来超越性别二元思维，审查软件系统，并创造对社会有益的代码。

The Collective Intelligence of Army Ants, and the Robots They Inspire. 哈佛大学计算机科学系教授、ROOT Robotics 联合创始人 Radhika Nagpal 在报告中介绍了如何基于军蚁的集体智慧启发机器人。在自然界中，数以千计的个体组成的群体纯粹通过局部相互作用来创造复杂的结构，在这些系统中，尽管个体能力有限，但足以实现巨大的复杂性集体。报告讨论了怎样才能创造出达到自然所能达到的规模和复杂性的人工集体。还介绍了 Eciton Robotica 的一个项目，该项目为了研究这个问题，利用来源于集体生物的灵感创造机器人系统。受军蚁筑巢的启发，该项目想创造一个可以自组装的软攀登机器人群体，这项工作横跨软机器人、新的自组织自组装理论模型和新的生物学领域实验。

四、 热点论文

2021 年度共有以下 6 篇论文获得最佳论文奖（Outstanding Paper Awards）。

A Universal Law of Robustness via Isoperimetry. 本文提出了一个理论模型来解释为什么许多最先进的深度网络需要非常多的参数以平滑地拟合训练数据。特别地，在关于训练分布的某些正则性条件下，$O(1)$-Lipschitz 函数对标签噪声下的训练数据进行插值（interpolate）所需的参数数目为nd，其中n是训练样本的数目，d是数据的维数。这一结果与传统结果形成了鲜明对比。传统结果指出，函数需要n个参数来插值训练数据。而为了平滑插值，额外的因子d似乎是必需的。该理论与一些对 MNIST 分类具有很强泛化能力的模型的实验观察一致。这项工作还提供了关于开发用于 ImageNet 分类的鲁棒模型所需模型大小的可信预测。

On the Expressivity of Markov Reward. 马尔可夫奖赏函数是应对不确定序列化决策和强化学习的主要框架。本文详细阐述了马尔可夫奖赏何时足以或不足以使系统设计人员根据他们对特定行为的偏好、或对状态和动作序列的偏好来指定任务。作者用简单例证说明，存在一些特定任务，它们不能指定马尔可夫奖赏函数来推导出期望的任务和结果。作者也证明了在多项式时间内可以确定对于期望的设置是否存在相容的马尔可夫奖赏。如果存在的话，那么在有限决策过程中构造这种马尔可夫奖赏的多项式时间算法也就存在。这项工作揭示了奖赏设计的挑战，并可能为研究马尔可夫框架何时及如何足以实现人类所希望的性能开辟新的研究途径。

Deep Reinforcement Learning at the Edge of the Statistical Precipice. 本文提出了针对深度强化学习算法精确性评估比较的更加严格的方法。具体而言，对新算法的评估应该提供分层的自助法（Bootstrap）

置信区间、不同任务和不同批次下的性能曲线、以及四分位数之间的均值。本文强调过去的一些文献报告中深度强化学习在多个任务和多次运行中的结果会导致很难评估新算法是否比过去的方法具有一致和可观的进步，并用实例说明了这一点。本文提出的性能总结方式能在每个任务少量运行的情况下进行计算，这可能是许多计算资源受限下的科研所需要的。

MAUVE: Measuring the Gap Between Neural Text and Human Text using Divergence Frontiers. 本文提出了 MAUVE 方法，用于衡量模型生成的文本分布和人类生成的文本分布的偏差。论文想法很简单，对于需要比较的两个文本的量化表示，使用连续的(软) KL 散度进行度量。MAUVE 实质上是对一系列测量的综合，其目的是同时捕捉第一类错误(生成不真实的文本)和第二类错误(未捕捉所有可能的人类文本)。实验表明，与以往的度量相比，MAUVE 识别了模型生成文本的已知模式，并且与人的判断更加相关。在开放场景文本生成快速发展的背景下，论文的结论有一定的借鉴意义。

Continuized Accelerations of Deterministic and Stochastic Gradient Descents, and of Gossip Algorithms. 本文描述了 Nesterov 加速梯度法(Nesterov’s accelerated gradient method)的“连续”版本，其中两个独立的向量变量在连续时间内联合演化。这一场景类似使用微分方程来理解加速度的方法，但更新时使用的是由泊松点过程（Poisson point process) 确定的随机时间下产生的梯度。这种新的方法是一种随机的离散时间方法，具有以下特点：(1) 拥有与 Nesterov 方法相同的加速收敛；(2) 提供了利用连续时间论点的清晰且透明的分析，比以前的加速梯度方法更容易理解；(3) 避免了离散化连续时间过程的额外错误，这与之前那些试图理解并使用连续时间过程的加速方法的尝试形成了对比。

Moser Flow: Divergence-based Generative Modeling on Manifolds. 本文提出了一种训练黎曼流形上连续规格化流（Continuous Normalizing Flow, CNF）生成模型的方法。其核心思想是利用 Moser (1965) 的一个结果，这个结果刻画了一个 CNF 的解，它使用具有几何正则性条件的一类受限常微分方程，并使用目标密度函数的散度来显式定义。本文提出的 Moser Flow 方法利用这一解的概念，提出一种基于参数化目标密度估计器 (可以是神经网络) 的 CNF 方法。训练相当于简单地优化密度估计器的发散性，绕过了运行 ODE 解算器(而这正是标准反向传播训练所必需的)。实验表明，与以前 CNF 工作相比，该方法训练时间更短，测试性能更好，并且能够在具有非常数曲率的隐式曲面（如 Stanford Bunny 模型）上对密度进行建模。这种利用几何正则性条件绕过昂贵反向传播训练的概念很有价值。

NeurIPS 的时间检验奖（Test of Time Award）一般授予 10 年前 NeurIPS 会议的论文。由于在 2020 年委员会考虑了更广泛的论文且选择了 2011 年而不是 2010 年的获奖者，因此，本年度 NeurIPS 时间检验奖委员会同时考虑 2010 年和 2011 年的论文，且由于 2010 年度未有论文获奖，本次时间检验奖重点考虑 2010 年度的论文。最终，2021 年度的时间检验奖授予 NIPS2010 年论文《Online Learning for Latent Dirichlet Allocation》，作者为 Matthew Hoffman、David Blei 和 Francis Bach。这篇论文提出了一种基于随机变分梯度的推断方法用于在大规模数据集上训练 Latent Dirichlet Allocation （LDA）模型。在理论方面，它表明训练过程收敛到局部最优值，随机梯度更新对应于证据下限（Evidence Lower Bound, ELBO）目标的随机自然梯度；实验方面，作者首次表明 LDA 可以在含数十万个文档的文本语料库上训练，使其成为解决“大数据”的实用技术。该想法在机器学习社区产生了巨大的影响，推进了一般化的基于随机变分梯度的推断方法在更多模型上的使用。

数据集和基准 Track 共有 2 篇论文获得最佳论文奖（Best Paper Awards）。

Reduced, Reused and Recycled: The Life of a Dataset in Machine Learning Research. 该工作分析了数千篇论文，研究了不同机器学习社区中数据集使用的演变，以及数据集采用和构建之间的相互作用。论文发现：在大多数社区中，数据集随着时间的推移越用越

少，且这些数据集来自少数机构。这种演变是有害的，使得基准变得不普遍，数据集来源中存在的偏差可能会被放大，且新数据集更难被接受。这对整个机器学习社区来说是一个重要的“警钟”，提醒研究者要更加批判性地思考哪些数据集用于基准测试，并更加重视创建新的和更多样化的数据集。

ATOM3D: Tasks on Molecules in Three Dimensions. 本文收集了一组具有小分子和生物聚合物 3D 表示的基准数据集，用于解决单分子结构预测、分子功能的设计和工程任务等问题。 通过与最先进的 1D 或 2D 表示模型进行比较，论文发现简单鲁棒的 3D 模型具有更好的性能。这项工作为如何为给定任务选择和设计模型提供了重要见解。论文不仅提供了基准数据集，还提供了基线模型和开源工具，大大降低了机器学习人员进入计算机生物学和分子设计的门槛。

五、 总结展望

NeurIPS 2021 中强化学习、图神经网络、表示学习以及注意力机制等领域仍保持热度。而对公平性、可解释性的关注有所提升。本年度 NeurIPS 的重要变化包括审稿方式的变化以及数据集 Track 的加入。一方面通过这些尝试使得审稿过程更加透明化，作者和审稿人之间更容易交流，同时也使整个机器学习研究社群关注数据集、基准的构建。

在会议的形式上，NeurIPS 2021 继续通过纯线上会议形式进行，通过 2 年的尝试，线上会议模式已有多项改进。根据目前的信息，NeurIPS 2022 将在新奥尔良会议中心(美国)以线下会议的形式举办，但同时也会保留线上会议的形式。希望这种混合模式能够让 NeurIPS 的参与者有新的体验。

责任编委 **魏秀参**

叶翰嘉

南京大学人工智能学院副研究员。主要研究方向为表示学习、模型复用等。
Email: yehj@nju.edu.cn

顶会观察

AAAI 2022

清华大学计算机系研究员 兴军亮

国际人工智能大会（AAAI Conference on Artificial Intelligence）是机器学习领域的顶级会议之一，在国内外具有广泛的影响力，被评为 CCF-A 类会议。由于疫情因素影响，第 36 届 AAAI 大会于 2022 年 2 月 22 日至 2022 年 3 月 1 日通过线上会议形式举办，包括 1 天的 Tutorial，4 天的正式会议以及 2 天的 Workshop。

一、 会议亮点

线上会议形式：AAAI 已连续两年通过纯线上会议形式举办。2022 年的 AAAI 会议全程使用 GatherTown 系统，为注册者提供与其他参会者互动的权限，例如向演讲者提问、参加现场海报环节、与其他与会者聊天和交流等。AAAI 的相关动态在其官方网站（https://aaai-2022.virtualchair.net/index.html）上同步，包含大会报告视频、接收论文等。

审稿表格更新：AAAI 期望提高审稿质量，对审稿表格进行了修改。首先，审稿表格中对每一项打分添加了明确的评判标准，如论文的社区影响小问中优秀（Excellent）得分的解释为“论文可能会在 AI 的多个子领域上产生影响”；其次，审稿表格中对论文总分添加了清晰的文本解释，如接收（Accept）得分的解释为“技术扎实的论文，至少对 AI 的一个子领域有较高的影响，或对 AI 的一个以上的领域有中到高的影响，有良好到优秀的评价、资源、可复现性，没有未解决的道德问题”；最后，审稿表格要求审稿人对自己的专业程度进行打分，该得分只有大会主席可见，避免审稿人因不熟悉审稿论文的领域导致不合理的打分。

审稿人分配程序修改：审稿人和作者之间可能会存在利益冲突或利益相关，审稿人为了拒绝或者接收某篇论文，审稿人可能会故意修改自己的审稿兴趣来增加分配得到该论文的概率。为了减少此类现象的发生，AAAI 修改了审稿人分配程序，如使用更彻底的冲突检查、随机分配审稿人等。

主观性减轻措施：审稿人在对论文进行打分时，需要从多个指标对论文进行评估，如创新性、技术合理性、社区影响性等，在权衡各项指标后得出最后的论文分数，然而审稿人对于各个指标的权重具有主观性，进而带来审稿偏差。为了减轻此类现象，AAAI 提出了一些措施，如利用算法对审稿人的分数进行识别，通过审稿人之间的广泛讨论来决定是否接收论文等。

二、 录用情况

2022 年度 AAAI 大会共收到 9020 篇有效投稿，最终 1370 篇论文被接收，接收率仅为 15.2%，达到近几年来最低水平，接收论文中有 416 篇 Oral 论文(30.36%)。2022 年 AAAI 投稿量较 2021 年提升 10.7%，接收率下降 6.2%；录用论文数量较 2020 年降低 13.89%。AAAI 启动了快速投稿通道，NerurIPS 2022 的拒稿修改后可以重投并跳过第一个审稿阶段，共收到了 590 篇投稿，接收率为 26.9%。录用论文的研究热点方向主要包括机器学习(29.3%)、计算机视觉 (28.8%)、自然语言处理(10.5%)、数据挖掘(4.89%)、博弈论(4.82%)、多智能体系统(1.2%)等。针对论文中作者所在国家进行统计，中国学者共提交论文 4230 篇，接收率为 11.30%，美国学者共提交论文 1479 篇，接收率 19.13%，韩国学者共提交论文 373 篇，接收率为 16.09%。

三、 邀请报告

2022 年度 AAAI 共举办了 8 场邀请报告（Invited Talk），具体内容如下：

The State of AI. 康奈尔大学工程和计算机科学系教授、AAAI 大会主席 Bart Selman 介绍了人工智能领域的现状和未来发展方向。报告回顾了人工智能领域的发展历史和现状，随着深度学习技术的兴起，人工智能的各个核心领域，如计算机视觉、自然语言处理、机器翻译、博弈论、强化学习等，发生了巨大的变化，同时各个核心领域之间也开始出现了融合交互，涌现出具有多项能力的智能体，报告预测未来数十年领域的统一和模式的整合是 AI 研究的核心驱动力。报告推测下一层次的人工智能将需要结合数据驱动范式、知识驱动方法、人机交互，形成真正鲁棒、可信赖、可理解的智能系统。

The Data-Centric AI. 斯坦福大学教授、DeepLearning AI 和 Landing AI 创始人吴恩达教授介绍了以数据为中心的 AI 研究。AI 可以分解为模型和数据，传统 AI 研究是以模型为中心的研究，通过优化模型来提高数据集上的准确率，而新型的以数据为中心的研究将关注点转移到数据上。报告讨论了以数据为中心的 AI 的多个研究方向：数据集竞赛、数据质量评估、数据迭代、数据管理工具、数据中报、数据增广和数据合成、错误数据识别与修复等。

Interpretable Machine Learning: Bringing Data Science Out of the “Dark Age”. 杜克大学计算机科学与工程教授 Cynthia Rudin 介绍了可理解的机器学习算法的重要性及前沿研究。报告以纽约的井盖检测为例，指出在数据来源混乱、决策具有高风险的场景下，如刑事司法、医疗保健、金融借贷等，可理解的机器学习算法具有更高的性能。报告介绍了团队在可理解的机器学习算法研究中的两项最新工作：广义线性和加性模型的快速稀疏分类法，可解释的罪犯再犯预测模型。

Toward an AI Network for Trustworthy AI. 美国退伍军人事务部国家人工智能研究所的所长 Gil Alterovitz 博士介绍了值得信赖的 AI 研究。值得信赖的 AI 需满足三个条件：合法的，尊重所有适用的法律和规定；道德的，尊重道德准则和价值观；鲁棒的，同时考虑技术角度和其社会环境。报告分享了来自退伍军人事务部国家人工智能研究所的几个使用案例，包括授权退伍军人控制药物依从性的试点，医生评估 COVID-19 相关的预后和需求，以及 VHA 工作人员基于文本的输入以快速识别和协助处于危机的退伍军人等。

Advancing Agricultural Genome to Phenome Research. 爱荷华州立大学客座教授 Pat Schnable 介绍了农作物表征类型预测方向的前沿研究。为了适应气候变化，在有限的土地上获得更高的粮食产量，植物科学研究所希望开发统计模型以预测农作物在不同环境中的表现。农作物的表型，如产量和耐旱性，是由基因型、环境及其相互作用控制的。然而，必要的表型数据量仍然是有限的，对基因型和环境之间的互动的理解也是有限的。为了解决这一限制，研究所正在建立新的传感器和机器人来自动收集大量的表型数据。报告还分享了应对热应激反应、优化根系统架构、优化作物冠层结构三个案例，

Safety and Robustness for Deep Learning with Provable Guarantees. 牛津大学计算系统教授 Marta Kwiatkowska 介绍了深度学习的安全性和鲁棒性。在一些强调安全的深度学习应用中，如自动驾驶汽车和医疗诊断，算法的安全性和鲁棒性至关重要。由于深度学习容易受到对抗扰动的影响，因此需要有严格的软件开发方法来确保算法决策的安全性和鲁棒性。报告分享了基于学习的软件组件开发自动认证技术的进展，从最大安全领域出发，介绍了基于搜索的、基于博弈对抗的、基于可达性验证的等安全性验证算法。报告还讨论了贝叶斯学习和因果关系所发挥的作用。

Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report. 布朗大学计算机科学系教授 Michael L. Littman 分享了《人工智能百年研究》研究结果。报告分享了人工智能领域的 14 个关键问题和研究结果，包括 AI 领域的里程碑、挑战性问题、通用人工智能前景、AI 技术的危机与机遇等。报告总结 AI 领域取得了巨大的进展，但也需要考虑决策风

险、不平等数据等问题，需要政府认识到 AI 的重要性，支持广泛的教育，需要研究机构分享研究结果、避免炒作、讨论危机和机遇、将 AI 纳入整个社区系统中，以实现 AI 领域的进一步发展。

Thinking Fast and Slow in AI. IBM 研究员 Francesca Rossi 分享了“快慢思考”决策理论。传统符号主义算法模仿人类推理能力，决策速度较慢，需要可控的环境，而新兴数据驱动算法利用原始数据进行预测，决策速度较快，但是缺少可理解性、可推广性、鲁棒性等能力，报告认为未来的 AI 决策应结合数据驱动算法和符号推理算法，实现 AI 能力的下一步突破。报告分享了该方向的一项最新成果，基于快/慢求解器的元认知组件通用架构，快求解器根据过去经验对问题做出快速反应，慢求解器对问题进行实时推理求解，元求解器根据问题规模、求解收益等判断采用哪一个求解器。报告介绍该框架在受限格子环境中的实验结果，系统随着时间的推移不断发展，在有足够经验的情况下逐渐从慢速思维过渡到快速思维，这对决策质量、资源消耗和效率有很大帮助。

四、 热点论文

2022 年度 AAAI 有 1 篇论文获得杰出论文奖 (Outstanding Paper Awards) 。

Online Certification of Preference-based Fairness for Personalized Recommender Systems. 本文提出了一种新的审查推荐系统公平性的评价指标“无嫉妒性”，即每个用户都应该更喜欢自己的推荐，而不是其他用户的推荐。为了计算推荐系统的“无嫉妒性”，本文将寻找更好的用户推荐建模为一个多臂老虎机问题，探索在相同场景下当前用户是否会更喜欢其他用户的推荐。为了减轻探索行为损害用户体检，本文采用保守的探索算法，保证审查中的推荐系统的实际性能和原推荐系统相近。本文通过实验证明推荐系统的“嫉妒性”来自于过强的模型假设和相等用户收益约束。

2022 年度 AAAI 有 2 篇论文获得杰出论文提名奖。

Bayesian Persuasion in Sequential Decision-Making. 本文研究具有更多全局信息的委托人如何为只具有局部信息的代理人设计合理的建议策略，使得在代理人最大化自身利益的情况下委托人可以实现自己的目标。该研究问题在现实生活中有很多的应用场景，如导航 APP 为用户提供导航策略的同时希望优化全局交通时间。本文根据代理人优化即时奖励和长期奖励，将代理人分成短视类型和远视类型。本文提出了一种多项式时间算法在短视类型的代理人情况下求解最优策略，证明在远视类型的代理人情况下不存在多项式时间算法。本文还提出在远视类型的代理人情况下可以设计一种基于威胁的建议策略，其效果和短时类型的代理人情况相同。

Operator-Potential Heuristics for Symbolic Search. 本文提出了一种结合符号搜索和启发式搜索的经典规划方法。符号启发式搜索算法有效的关键在于满足两个性质，有效评估二进制决策图表示的状态集合以及产生一个良好的划分，虽然过去已经提出了几种符号启发式搜索算法，但它们仅能满足性质一，因此算法性能未能超过朴素的符号搜索算法。本文利用潜在的启发式方法可以被编码为潜在算子的特性，将启发式信息直接整合到过渡关系函数中，避免了其他符号启发式算法 经常引起的二进制决策图的大量分化。实验表明该算法满足了性质二，同时算法性能超过了目前最优的符号搜索算法。

2022 年度 AAAI 有 6 篇论文获得卓越论文奖 (Distinguished Papers) 。

AlphaHoldem: High-Performance Artificial Intelligence for Heads-Up No-Limit Poker via End-to-End Reinforcement Learning. 本文提出了一种高水平轻量化的两人无限注德州扑克 AI 程序 AlphaHoldem。AlphaHoldem 整体上采用一种精心设计的伪孪生网络架构，并将一种改进的深度强化学习算法与一种新型的自博弈学习算法相结合，在不借助任何领域知识的情况下，直接从牌面信息端到端地学习候选动作进行决策。AlphaHoldem 使用了 1 台包含 8 块 GPU 卡的服务器，经过三天的自博弈学习后，战胜了 Slumbot 和 DeepStack。在每次决策时，AlphaHoldem 仅需不到 3 毫秒，比 DeepStack 速

度提升超过了 1000 倍。同时，AlphaHoldem 与四位高水平德州扑克选手对抗 1 万局的结果表明其已经达到了人类专业玩家水平。

Certified Symmetry and Dominance Breaking for Combinatorial Optimization. 在组合搜索和优化问题中，一个至关重要的步骤是添加多项约束条件来打破对称性和支配性，然而验证约束条件的正确性是十分困难的。本文提出了一种优化问题的认证方法，给定原公式和能处理的支配性的目标函数，算法提出了一种支配解的显式构造方法，使得验证器可以核对该构造是否在满足原问题的条件下打破了支配性。实验表明该方法可以有效地验证布尔可满足性求解中完全通用的对称性突破，从而首次提供了一个统一的方法来认证一系列先进的 SAT 技术。本文还将该方法应用于最大群组求解和约束性编程，证明该方法适用于更广泛的组合问题。

Online Elicitation of Necessarily Optimal Matchings. 本文研究在房屋分配模型中，仅知晓代理人的前 K 偏好情况下，如何询问最少次数的代理人偏好来实现必要帕累托最优匹配或必要顺序最大化匹配。本文首先考虑传统的询问代理人偏好的方式：下一个最佳查询模型，在该模型下提出了一种在线算法，可以以 1.5 竞争比率实现必要顺序匹配最大化匹配，且证明该算法是最优的。本文又提出了两种询问代理人偏好的方式：混合查询模型和集合比较查询模型，为这两种模型提出了在线算法并给出了计算复杂度。

Sampling-Based Robust Control of Autonomous Systems with Non-Gaussian Noise. 本文提出了一种规划算法，在具有未知加性噪声的线性动态系统中，以高置信度概率控制无人机避开某些不安全的区域，安全到达给定区域。本文首先将连续动态系统抽象为离散状态模型，通过有限样本采样评估状态间的转移概率，采用场景优化算法计算近似正确的转移概率的上下限。本文接着利用区间马尔可夫决策过程形式化抽象模型，并计算一个鲁棒的策略来最大化安全到达目标状态的概率，若该概率不满足预先设定的阈值，算法会收集更多的样本来减少转移概率区间的不确定性。实验表明本文算法可以获得更加鲁棒的结果。

Subset approximation of Pareto Regions with Bi-objective A*. 本文提出了一种高效求解双目标优化问题帕累托最优近似子集解的算法。本文引入两个实参数，将原双目标优化问题转化为新目标问题，保证新目标问题的子集解是原优化问题的子集解，且保证原优化问题的启发式信息仍然满足。本文采用双目标 A*算法求解新目标问题，实验表明算法能够在比求解原问题少一个数量级的时间内，获得一个包含大约 10%的解决方案的多样化解决方案集。本文还证明通过以适当的实参数序列运行该算法，可以收敛得到原优化问题的完整解集。

The Soft Cumulative Constraint with Quadratic Penalty. 资源调度问题需要将有限的资源分配给若干任务，该问题的研究通常约束资源不能超出限制，而本文研究允许资源超出限制但会带来惩罚的情形。本文提出了检查器算法和过滤算法，可以求解线性惩罚约束和二次惩罚约束的问题。实验表明该算法比现有算法更通用，且性能超过分解约束的算法。

2022 年度 AAAI 有 1 篇论文获得杰出学生论文奖 (Outstanding Student Paper) 。

InfoLM: A New Metric to Evaluate Summarization & Data2Text Generation. 本文作者引入了一种叫做 InfoLM 的新指标用于自动评估文本摘要和 data2text 生成。InfoLM 主要包含两个关键组件：(1)一个预训练掩码语言模型被用来分别计算在给定候选句子和参考句子的情况下观察词汇每个标记(token)的离散概率分布。(2)一个用于测量前面两个离散概率分布之间差异性的对比函数。InfoLM 直接依赖于 token 的统计数据，因此也可被看作是一种基于字符串的度量标准。同时，PMLM 的引入允许为释义分配高权重并可以捕获长程依赖关系，因此 InfoLM 不存在基于字符串指标中的常见缺陷。这个指标还利用了信息度量，使 InfoLM 有可能适应不同的评价标准。作者通过大量的实验证明了与现有指标相比，InfoLM 在文本摘要和 data2text 生成任务中都取得了具有统计学意义的显著改进。

2022 年度 AAAI 有 2 篇论文获得杰出学生论文提名奖。

Compilation of Aggregates in ASP Systems. 回答集编程（ASP）作为一种声明式人工智能形式主义，被广泛用于知识表示和推理。目前最先进的 ASP 实现采用了 ground&solve 方法，并成功地应用于工业和学术问题。然而，有一类 ASP 方法，由于 grounding 步骤带来的组合爆炸问题，其评估效率并不高。最近的研究表明，基于编译的技术可以缓解 grounding 的瓶颈问题。然而，对于包含聚合（aggregate）的 ASP 程序，还没有开发出基于编译的技术，而聚合是 ASP 中最相关和最常用的结构之一。本文作者为带有聚合的 ASP 程序提出了一种基于编译的方法。作者在当前最先进的 ASP 系统上实现了这一方法，并在公开的基准上进行了性能评估。实验表明，该方法针对 ground-intensive 型的 ASP 程序是有效的。

Entropy estimation via normalizing flow. 熵估计是信息论和统计科学中的一个重要问题。现有的熵估计器在维度快速增长的情况下会出现估计偏差，这使得它们不适合于高维问题。本文作者提出了一种基于变换的高维熵估计方法，它由以下两个主要成分组成。首先，基于已有的 k-NN 熵估计器，作者提出了一个新的估计器，它对接近均匀分布的样本具有较小的估计偏差。其次，作者设计了一个基于标准化流（normalizing flow）的映射，可以将样本推向均匀分布，并推导出了原始样本的熵和转换后的熵之间的关系。因此，本文方法通过首先将样本转化为均匀分布，然后对转化后的样本应用改进后的 k-NN 估计器来解决高维熵估计问题。数值实验的结果证明了本文方法可以解码复杂的高维分布，并获得了熵值的准确估计。

2022 年度 AAAI 有 1 篇论文获得最佳演示奖（Best Demonstration Award）。

A Demonstration of Compositional, Hierarchical Interactive Task Learning. 本文作者展示了一个交互式任务学习型的智能体在模拟的军营环境中通过情景化的自然语言指导学会巡逻的过程。在此过程中，该智能体建立了一个由先天和后天任务组成的大型层次结构。这些任务可以被描述为实现某个目标或者遵循一定的规则，其中包含有条件分支和循环，并涉及智能体的通信和心理活动。该智能体是在 Soar 认知框架下实现的，该框架使用了声明式任务网络来表示任务，并通过分块将其编译为程序性规则。实验证明，智能体从单一训练场景中学习到了复杂任务的解决方法。

五、 总结展望

2022 年 AAAI 的接收投稿数量相对提升，但接受率和录用论文数量均相对下降，表明 AAAI 期望进一步控制接收论文的数量和质量。本年度 AAAI 的重要变化主要在于审稿方式的变化，通过这些尝试使得审稿过程更加公平客观。本年度继续采用线上会议的方式，整体参会体验良好，也使得更多人有机会参与到会议中。目前越来越多的审稿会议采用 OpenReview 审稿系统，如果 AAAI 未来也能使用 OpenReview 审稿系统，应该可以促进更加透明的审稿过程。至于如何通过改进多轮评审机制同时保证投稿质量和审稿质量，仍将是需要不断探索的过程。

责任编委 **魏秀参 崔海楠**

兴军亮

清华大学计算机系研究员。研究方向为计算机博弈、计算机视觉和人机交互学习。
谷歌学术主页：https://scholar.google.com/citations?user=jSwNd3MAAAAJ
Email: jlxing@tsinghua.edu.cn

西安电子科技大学苗启广教授访谈

2022 年 2 月 17 日，《CCF-CV 专委简报》在线采访了西安电子科技大学网络与继续教育学院院长、博士生导师苗启广教授。下面是采访实录。

苗老师，您好！首先，请您分享一下您的个人学习和研究经历。

我是在 2001 年来到西安电子科技大学的，师从计算机学院前院长王宝树教授攻读硕士学位，一年后硕博连读攻读博士学位。当时主要从事图像融合方面的研究。当时人工智能还没有像现在这么火热，因此主要基于传统的金字塔融合、小波融合、Shearlet 融合等方法进行研究，但也是在这个过程中，让我积累了大量的计算机视觉和模式识别领域的知识，为后续研究工作的展开打下基础。同时，王老师有许多国防预研项目，对于图像融合的研究正是依托这些项目展开。这要求算法结果能够真正应用于实际场景当中，实质上对研究有着更高层次的要求。可以说，这一段时间的学习和研究，为我后来的研究积累了相当丰富的经验，让我后续的科研工作更加得心应手。

在 2005 年博士毕业后，出于对科研的憧憬、对教育事业的热爱和对母校深深的眷恋，我选择了留校任教。作为一名高校教师，我对整个教学、科研及体系的认识也更为完善。一方面，以图像融合为起点，我带领研究生对三维点云数据融合、彩色地形图智能解译与优化、图像去雾霾及人体行为/手势识别等多个方面展开了深入研究；另一方面，认识到计算机视觉是一门高速发展的学科，我又进一步将科研工作当中了解到的前沿知识和研究成果融入教学当中。

在 2013 年前后，了解到深度神经网络的发展情况，我带领团队对其在计算机视觉领域的应用展开了深入研究，这当时在西电也是比较领先的。在这个过程中，我们对深度神经网络的理论有了更为深刻的认识，我们后续在图像去雾霾和人体行为/手势识别等领域的研究成果都跟这一段时间的积累是分不开的。

2019 年后，我调到网络与继续教育学院担任院长，在这个过程中我也尝试将计算机视觉的研究融入到学院的工作当中。针对网络教育的性质，我带领团队积极探索以技术赋能智能教育新模式，打造面向全校本科生、研究生、留学生和继续教育学生“四位一体”的西电 SPOC 学习平台，该平台在疫情期间向学校整体本科教育加以推广，取得了良好的效果。

您在计算机视觉、机器学习和大数据分析等多个领域有很深的造诣，能否介绍一下您在这些领域中最突出的几项研究成果？针对这些领域的研究者，您有什么建议？

主要的研究成果包括彩色地形图智能解译与优化、人体行为/手势识别理解和智能教育几个方面。其中，针对彩色地形图的要素提取与识别问题，提出了以地理信息智能解译为目标的一系列优化理论和模型，解决了地形图中要素分类算法受噪声数据和非均衡数据影响、深

度网络模型受高复杂性地理要素样本影响、地理要素分离提取受混淆色彩信息影响以及多时地形图差异检测受非对齐数据影响等关键问题。相关成果在 TEC、TIP、TGARS 等国际顶级期刊发表，并获得了陕西省自然科学一等奖；针对人体行为/手势识别领域，易出现背景、光照等无关因素干扰的问题，提出了一系列基于多模态数据和三维卷积神经网络的识别算法，相关成果蝉联两届 Chalearn 国际大规模独立手势识别竞赛的冠军，并已应用到海信智慧家居整体解决方案当中，取得了较好的实际应用效果；在智慧教育方面，则主要基于实际应用需求，研发了西电 SPOC 智慧教育平台，推进人工智能+教育、互联网+教育等创新模式，促进线上线下混合式教学模式发展。

针对相关领域的研究者，我的建议，也是我最大的体会是，一方面要时刻跟上研究前沿的发展脚步。如上所述，计算机视觉是一个发展较为迅速的学科，特别是近些年随着人工智能等学科的崛起，相关研究无论是在数量上还是质量上都有着质的飞跃。正所谓“学如逆水行舟，不进则退”，一个月不跟进研究前沿的内容就有可能被抛下，因此就需要我们保持一颗探索的心，关注各大顶会，甚至是 arXiv 上的研究进展，以活跃我们的思维，激发我们的创新灵感；另一方面则要沉下去，做出一些真正经得起时间考验的成果。不可否认近些年人工智能、计算机视觉领域的研究呈井喷式增长，各大顶会的投稿量持续上升，但作为研究者，我们需要思考的是，我们的研究是否真的具有应用价值？当前正逢国家大力发展人工智能的关键时期，作为一线研究者，我们更应该沉下心，把眼光放长远，面向“卡脖子”难题，真正做出一系列具有应用前景和现实意义的成果，为中国智造添砖加瓦。

您获批了多项国家级课题，如核高基国家重大科技专项课题等，能跟大家分享一下您成功申请的经验和体会吗？

对于项目申报，我想从下面三个方面分享我的一些经验和体会。

首先项目研究应以国家发展规划的重大需求为导向，解决其中存在的关键性科学问题。我们研究团队先后主持的核高基国家重大科技专项课题和国家自然科学基金面上项目的研究内容是紧扣国家对新一代人工智能技术发展的指导意见而提出的。而团队获得陕西省自然科学技术奖一等奖的项目内容（地理要素提取与识别）也是针对国家对地理、地形信息快速获取和更新的实际需求展开。在相关项目的支持下，通过多年的课题研究，解决、突破了相关内容中的一些关键性科学问题。

其次，重视深层次的技术创新。针对项目研究内容中所存在的诸多科学问题，提出创新性的解决方案。这些解决方案应该是能够从深层次的解决相关问题，而非浅层次的模型或方法改进。这就需要深入分析问题的本质，提出能够实质解决问题的网络、模型或者方法。例如：在地理要素提取方面，针对线要素提取问题，我们提出能量密度的概念，并以此作为区分像素特征的度量标准，解决了地理要素提取研究领域中长期存在的线要素提取不准确的难题。以此，通过核心的技术创新来解决项目研究的关键科学问题是申报和执行项目的一个关键。

最后，我们应该能够在相关项目研究方面具有扎实的研究基础。任何项目研究都需要有充分的前期准备，对相关研究内容和现有技术研究现状进行全面的调研，深入分析其中存在的关键性科学问题，对相关研究内容进行前期的探索性研究等。如果是跨学科研究，则建议在前期就有深入的合作，相互了解对方的技术和需求，碰撞交叉研究的结合点，解决不同学科相关研究领域的关键性问题等。这些都是为项目研究奠定前期的研究基础，以能够保障项目能够顺利开展。

您获得了 2 项省部级科技奖励，能否跟大家分享一下您申报奖励的经验？您认为奖励申报成功的要素有哪些呢？

对于申报奖励，可分享的经验主要有：科学研究过程和奖励申报。对于科学研究，首先我们应该面向实际应用需求，重视基础研究，紧扣研究热点。而对于具体的问题，则应该鼓励原创性技术创新，以实质解决科学研究过程中遇到的关键技术问题。其次，我们应该选定一个长远的研究方向和目标，坚持在同一个研究领域持续进行科研工作，突破相关技术难题，相关研究能够形成一个较为系统性的成果。最好不要多点式短暂性的科学研究，这样难以形成一个完备的研究体系。然后，还需要认真考虑理论研究成果的实用化、产业化等。另外申报奖励的选题应该具有重要的研究和实用意义，且建议是团队研究最为深入的课题，已经产出了较多的高质量的研究成果(包括理论成果和实际应用成果)。合理组织奖励申报团队。最后，认真准备申报材料，反复打磨，言简意赅地突出核心贡献、解决的关键技术难题、研究成果国际影响力及其实际应用情况等。

奖励申报成功的要素主要在于有重要科研价值的选题，具有重要技术突破的研究成果，高质量的申报材料以及答辩时的良好表现。

您发表了百余篇国内外重要期刊和会议学术论文，在论文发表方面取得了突出成绩，能跟大家分享一下您是如何做到持续产出高水平论文的呢？

我认为学术论文写作与发表是科研工作的重要组成部分，在基础理论研究领域更是如此。论文写作主要是对自己科研成果或学术创新的真实记录、整理和规范化的书面表达，论文发表则是为了确认科学发现，进行学术交流和科研成果的推广应用。我觉得我们团队可以持续产出高水平论文的关键是我们团队的研究时刻面向国家重大科研需求与卡脖子问题，目标是国际最先进水平，这样产出的科研成果就是高水平的、新颖的、有延续性的。

您担任 CCF 的理事、西安分部主席等多个学术职务，您是如何兼顾学会工作和学校工作的呢？CCF 对您的帮助有哪些呢？您认为在组织 CCF 活动时要注意什么呢？您认为 CCF 对学术界、企业界最大的贡献是什么？

CCF 是极好的科研交流平台，而我是科研工作者，很多学会的工作都对我的科研都有很大的帮助，让我学习到了很多学术前沿，认识了很多国内外的同行，更重要的是让我得到了充分的锻炼。CCF 组织活动的关键就是建设好的平台，提供有意义的内容，一定要让参加的会员有所收获，这样才能不断的提高。我认为 CCF 对学术界、企业界最大的贡献就是给科研工作者提供了一个非常完善的舞台，大家可以在这个舞台上面学习和进步，为祖国的科技事业出上自己的一份力。

您担任了多个行政职务，每天需要处理众多日常事务，能跟大家分享一下您是如何兼顾管理工作和学术研究工作的吗？能分享一下您的经验吗？

我认为平衡行政工作和科研工作的关键在于时间的充分利用，比如我会每天都坚持进行学习与了解学术的前沿，每周都会与每个研究生进行 2-3 次的会议讨论，每次行政会议的间隔都会去看学生的最新进展，在外出差的时候，等火车或者候机的时候都是很好的办公时间，在火车上或飞机上都是阅读、写材料和修改论文。经常自省：醒着，就抓紧时间工作!

您担任过多个国际学术会议的大会主席，能介绍一下您参与组织这些会议的经验吗？另外，在与国内外同行交流时，怎样才能获得更好的效果呢？

学术会议是科研工作者学习与交流的平台，参会者可以分享自己的研究成果，提升自己的价值，了解最新的研究动向与学术前言。而会议的组织者最重要的职责就是搭建好这样的平台，其中的关键就是成立会议组织工作组和确定会议的主题。

在与国内外同行交流时，关键在于可以有效表达出自己的科研成果与研究态度，这样双方都可以很好地找到共鸣，学习对方的长处，共同进步。

您获得过西电“十佳师德标兵”称号，您是教书育人的楷模，请问您是如何平衡教学和科研的？如何将您的科研经验及成果融入教学之中的呢？您是如何跟学生相处的？

我一直认为，教学与科研像是一枚硬币的两面，各具特色，永远紧贴，不离不弃，相辅相成。学生的知识面和能力已经不能和自己上学时同日而语，而且计算机科学领域知识更新换代非常快，只有不断提升自己，才有底气站在讲台上“学高为师”。而科研恰好处于领域知识的最前沿，因为在这个过程中你看到的、做出的都是最新的东西。我尝试结合课堂教学的内容对中科院计算所、自动化所、以及我们自己在涉及图像融合、图像目标检测与识别、图像水印、视频伪造/防伪等方面的研究成果向学生进行展示，学生对这些内容有着非常浓厚的兴趣，相比单纯的理论，这些是可以和他们自身的生活相结合的。通过这样图文并茂的展示，我想学生会更容易理解所学的内容及其应用价值。

而对待学生，我觉得一方面要不忘教师初心，本着“要对得起每一名学生”的信念来从事教育工作，这是教师的责任。教学的过程，是发自内心的一种“爱的付出”的过程，当你真正把学生都当作自己的孩子一样，时时刻刻铭记将“真心”融入到备课、讲课、反思中，看到学生有所成长，有所收获，你也会乐在其中；另一方面，也要把学生作为朋友来看待。现在的学生因为获得信息的渠道很多，其认识的广度和我们当学生的年代不可同日而语。如果一味以老师的权威“压人”，反而不利于师生之间的沟通。因此需要像朋友一样，跟学生推心置腹，了解学生内心的真实想法，这样才有助于和学生增进互信。特别是在科研过程中，学生有的时候有很多新颖的思路，这个时候不能一味打压，而是要积极引导，让学生自己动手去解决问题，获得新的理解与感悟，对我们老师自己也是一种提升。

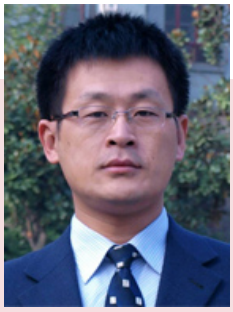
您领导着非常优秀的团队，请问您是如何管理和运作您的团队的？您是如何管理研究生的？您对他们的要求是什么？

团队将科学研究和人文关怀融会到团队建设和学生培养的各个方面之中，制定了“一学二进三指导，四训五抓六文建，七奖八论九不许”的实验室规范，依照每个年级学生的科研能力和水平进行任务划分，在科研上严格训练，狠抓落实，保证学生论文过程的规范，培养学生的学术能力，提升团队的科研水平；同时也组织包括学术交流会、趣味运动会、迎新/欢送会等各位文体活动，丰富学生的业余文化生活，使学生能够更好地融入团队当中，也促进了学生的全面发展。

如果吐露研究工作者的心声，您最想说的是什么？

希望所有研究工作者都可以做出有影响力的成果，实现自身的价值，解决国家重大卡脖子问题。

责任编委 赵振兵 余烨



苗启广

苗启广，博士，教授，博士生导师，西安电子科技大学网络与继续教育学院院长。2012 年入选“教育部新世纪优秀人才支持计划”。中国计算机学会(CCF)理事、CCF 西安分部主席、CCF YOCSEF 主席（2017-2018），ACM 西安常务理事，陕西省计算机学会常务理事（兼常务副秘书长），CCF 计算机视觉专委会委员，CCF 大数据专委会委员，CCF 人工智能与模式识别专委会委员，CCF 青年工作委员会委员，CCF 分部与会员工作委员会执委，CCF 杰出会员，IEEE Senior Member。主要从事计算机视觉、机器学习、大数据分析等方面的研究。主持在研和完成核高基国家重大科技专项课题、国家重点研发计划课题、国家自然科学基金、省自然科学基金、国防预研、国防 863、武器装备基金项目 30 余项。获省部级奖 2 项。在 IEEE TNNLS/TIP/TGRS/TEC/TIST/TVCG/TCYB、IJCV、ICCV、AAAI、IJCAI、ICCV、软件学报、计算机学报、电子学报、光学学报等国内外重要学术期刊、国际会议上发表 SCI/EI 收录论文 100 余篇。担任 ChinaVIS2020 大会主席、FG2018、NPC 2016 国际会议组织委员会主席、NPC 2020 出版主席、ICYCSEE 2018、IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IEMCE) 2015/2019 国际会议程序委员会主席。2008/2011/2014 年分别获西安电子科技大学“十佳师德标兵”称号，2018 年被评为陕西省高教工委优秀共产党员。

委员好消息

✪ 2021 年 12 月 15 日，中国 21 世纪议程管理中心公布了国家重点研发计划“国家质量基础设施体系”重点专项 2021 年度项目立项结果，CCF-CV 专委会委员、上海海事大学**周日贵**教授牵头负责的“基于谱学和显微成像的产品品质检测坚定技术研究与应用”项目获得立项，这是周日贵教授第二次担任国家重点研发计划项目的首席科学家。

✪ 2022 年 1 月 5 日，2021 年 CCF 优秀博士学位论文奖评选结果公布，CCF-CV 专委会 4 位委员指导完成的论文入选。南开大学**程明明**教授指导完成的《认知规律启发的显著性物体检测方法与评测》、东南大学**耿新**教授指导完成的《机器学习中的标记增强理论与应用研究》获 CCF 优秀博士学位论文奖，北京大学**林宙辰**教授指导完成的《基于冲量的加速优化算法》、西北工业大学**王庆**教授指导完成的《多视光场光线空间几何模型研究》获 CCF 优秀博士学位论文奖提名。

✪ 2022 年 1 月 6 日，CCF-CV 专委会委员、北京师范大学**黄华**教授获《自动化学报》2021 年度优秀编委奖。

✪ 2022 年 1 月 7 日，中国教师发展基金会公示了 2021 年度高校计算机专业优秀教师奖励计划评选结果，CCF-CV 专委会 3 位委员入选，他们是国防科技大学**李健**老师、同济大学**何良华**教授、浙江大学**章国锋**教授。本年度共 55 人入选。

✪ 2022 年 1 月 9 日,2021 年度中国指挥与控制学会(CICC)科学技术奖颁奖典礼在京召开，CCF-CV 专委会委员、中国科学院空天信息创新研究院**孙显**研究员获得 CICC 青年科学家奖，同时，其参加的“基于数字地球的联合应急与指挥控制系统应用”项目获 CICC 科技进步一等奖。

✪ 2022 年 1 月 11 日，2021 年 CCF-阿里巴巴创新研究计划青年科学基金评审结果公布，来自 9 所高校的 9 个项目入选。CCF-CV 专委会委员、杭州电子科技大学**俞俊**教授的“跨境电商国家差异化搜索关键技术研究”项目获得资助。

✪ 2022 年 1 月 12 日，2021 中国电子学会科学技术奖公告发布，CCF-CV 专委会 6 位委员参与的项目获奖：中国科学院自动化研究所**雷震**研究员参与完成的“多模态高鲁棒细微情感分析关键技术与系统”、西安电子科技大学**董伟生**教授参与完成的“时空谱编码耦合与深度网络解耦超限成像技术”获技术发明一等奖，浙江大学**李玺**教授和**赵洲**副教授参与完成的“超大规模高性能图神经网络计算平台及其应用”获科技进步一等奖，北京邮电大学**明悦**副教授参与完成的“边端协同技术及其在安保指挥通信系统中的应用”、中科院自动化所**朱翔昱**副研究员参与完成的“基于多模态身份识别的智能金融终端及跨域云服务平台”获科技进步二等奖。

✪ 2022 年 1 月 14 日，2021 年 CCF 卓越服务奖评选结果公布，CCF-CV 专委会委员、中科院计算所**陈熙霖**研究员获奖。陈熙霖研究员长期服务于 CCF，参与了 CCF 多种学术服务工作，为推动 CCF 学术平台进步做出了重要贡献。

✪ 2022 年 1 月 15 日，广东省科学技术厅公布了 2021 年度广东省科学技术奖拟奖公示，CCF-CV 专委会委员、南方科技大学**于仕琪**副教授参与完成的“高光谱和高空间分辨率遥感数据信息提取与定量反演”项目拟授自然

科学二等奖。

✪ 2022 年 1 月 17 日，2021 年度 CCF 杰出演讲者评选结果公布，CCF-CCV 专委会委员、北京大学**彭宇新**教授和中科院自动化所**张兆翔**研究员入选。

✪ 2022 年 1 月 27 日，2021 年度吴文俊人工智能科学技术奖获奖名单公布，CCF-CV 专委会 7 位委员获奖：清华大学**苏航**副研究员等参与完成的“鲁棒高效的深度学习理论与方法”获自然科学一等奖，西北工业大学**韩军伟**教授主持的“高空无人机对地精准观测技术”获技术发明一等奖，北京邮电大学**明悦**副教授主持的“跨模态数据协同识别技术与应用”获技术发明二等奖，中科院自动化所**董晶**副研究员参编的《哇塞！机器人——中小学机器人科普读本》获科技进步奖（科普项目），中国科学院信息工程研究所**任文琦**副研究员、重庆大学**张磊**研究员、天津大学**朱鹏飞**副教授获优秀青年奖。

✪ 2022 年 1 月 29 日，第二十四届中国科协求是杰出青年成果转化奖拟获奖人员名单公示，CCF-CV 专委会委员、北京航空航天大学**徐迈**教授入选。

✪ 2022 年 2 月 15 日，教育部公布了首批国家级虚拟教研室建设试点名单，CCF-CV 专委会委员、哈尔滨工程大学**刘海波**教授作为带头人的智海 AI 课程虚拟教研室和西安电子科技大学**邓成**教授作为带头人的数字逻辑与微处理器课程群虚拟教研室入选。

✪ 2022 年 2 月 24 日，AAAI2022 公布了获奖论文，CCF-CV 专委会委员、中科院自动化所**兴军亮**副研究员指导的论文 AlphaHoldem: High-Performance Artificial Intelligence for Heads-Up No-Limit Poker via End-to-End Reinforcement Learning 获卓越论文奖。本次大会评出杰出（Outstanding）论文奖 1 篇及提名奖 2 篇，卓越（Distinguished）论文奖 6 篇，杰出学生论文奖 1 篇及提名奖 2 篇，最佳演示奖论文 1 篇。

✪ 2022 年 3 月 3 日，第二十四届茅以升科学技术奖—北京青年科技奖评选结果公示，CCF-CV 专委会委员、北京师范大学**邬霞**教授入选。

✪ 2022 年 3 月 4 日，CCF-CV 专委会委员、中国科学院心理研究所**王甦菁**副研究员参加了在北京市残疾人文化体育指导中心点位进行的冬残奥会火炬传递。该点位的主题是自强不息，共有 49 名火炬手参与活动，王甦菁副研究员因在提倡“AI+辅具”理念中做出突出贡献，作为北京市代表入选火炬手。

责任编委 刘海波

基于 Transformer 的图像生成开源代码

大连理工大学 付陈平 樊鑫

Transformer 技术起初应用于自然语言处理(Natural Language Processing, NLP)任务中，极大推动了该任务的发展。Transformer 技术在该任务中的优秀表现，越来越多的研究人员希望将 Transformer 推广到计算机视觉的应用当中。其中，底层视觉图像生成任务获得了较好的研究。本文将从较早的 Transformer 图像生成方法 TransGAN 开始，重点介绍一些基于 Transformer 技术图像生成研究成果。

1、TransGAN

工作：该文提出 TransGAN 结构用于图像生成，通过 Transformer 构建生成对抗网络(Generative Adversarial Networks, GAN)结构。TransGAN 包括一个内存空间占用小且基于 Transformer 的生成器，该生成器可以逐步提高特征分辨率。此外，TransGAN 还包括一个多尺度判别器，用于捕获语义信息和低层纹理。在该生成-判别器的基础上，该文还提出了一种新的网格自注意模块，该模块将用于进一步减少模型的内存占用空间，从而可将 TransGAN 应用到高分辨生成之中。与此同时，该文开发了一系列高效可行的训练策略，可解决 TransGAN 在训练过程中出现的一些不稳定的问题。这些训练策略包括数据增强、模型归一化和位置编码均等方面内容。该文在各种基准数据集(例如：CIFAR-10、STL-10、CelebA 等)上进行了大量实验，这些实验证明了 TransGAN 在图像生成任务中的有效性和潜力。TransGAN 的网络结构图如图 1 所示。

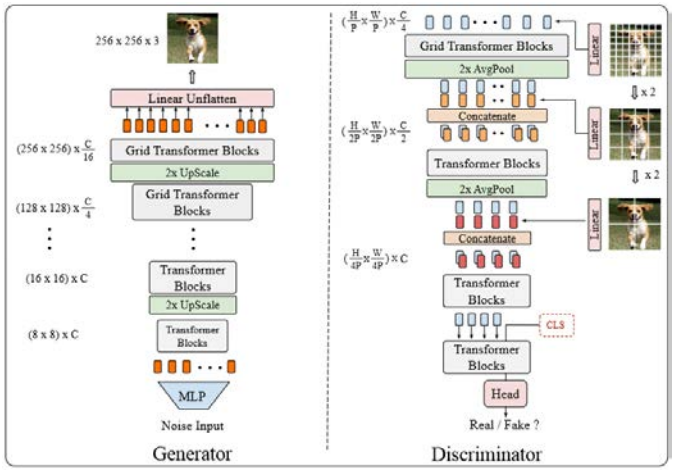


图 1 TransGAN 结构图

更多有关 TransGAN 的详细内容可参考发布该方法的论文“TransGAN: Two Pure Transformers Can Make One Strong GAN, and That Can Scale Up”。

论文地址：https://arxiv.org/abs/2102.07074
代码地址：https://github.com/VITA-Group/TransGAN

2、ViTGAN

工作：Vision Transformers (ViTs) 在图像识别领域取得较大的成功，且视觉特异性诱导偏差方面需求较少。因此该文试图将 ViTs 应用到图像生成领域之中。然而，ViTs 与 GAN 的集成(即 ViTGAN)会导致训练过程中出现严重的不稳定性。该问题是由现阶段正则化方法与自注意力交互作用较差导致的。为解决这个问题，该文设计了一种新的正则化方法，以实现较为平稳地训练 ViTs

与 GAN 的集成结构 ViTGAN。ViTGAN 的判别器与生成器都是基于 ViT 设计的，判别器的评分是由分类嵌入得到，生成器基于区域嵌入逐区域生成像素。ViTGAN 的网络结构图如图 2 所示。更多有关 ViTGAN 的详细内容可参考发布该方法的论文“ViTGAN: Training GANs with Vision Transformers”。

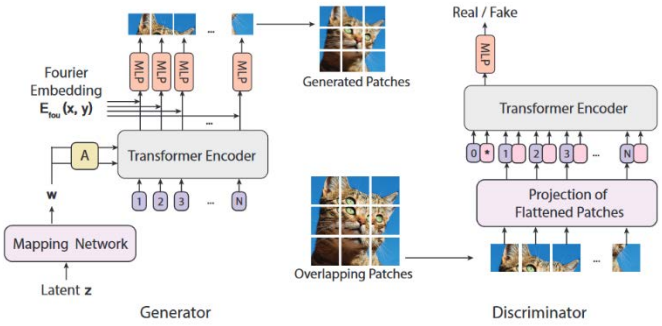


图 2 ViTGAN 结构图

论文地址： https://arxiv.org/pdf/2107.04589.pdf
代码地址： https://github.com/wilile26811249/ViTGAN

3、Taming Transformers

工作： 该文面向高分辨率图像的生成提出了 Taming Transformers。Taming Transformers 将卷积网络的诱导偏差有效性与 Transformer 的良好表达能力相结合，使其统一建模从而可用于高分辨率图像的合成。一方面，Taming Transformers 可以通过卷积网络来学习图像丰富的上下文信息；另一方面，Taming Transformers 利用 Transformer 在高分辨率图像中有效的将这些上下文信息进行建模表达。该文所提方法可推广到各类条件合成任务中，例如非空间信息（如目标分类）和空间信息（如目标分割）都可以控制生成的图像。Taming Transformers 主要包括两部分：VQGAN 和 Transformer。VQGAN 是 VQVAE 的一种变体，它使用判别器和感知损失来提高视觉质量。通过 VQGAN，一系列上下文丰富的离散向量可用于表示图像，从而，Transformer 可通过自回归的方式对这些向量进行预测。接下来，为生成高分辨图像，Transformer 模型将以此学习长程交互作用，其网络结构图如图 3 所示。更多有关 Taming Transformers 的详细内容可参考论文“Taming Transformers for High-Resolution Image Synthesis”。

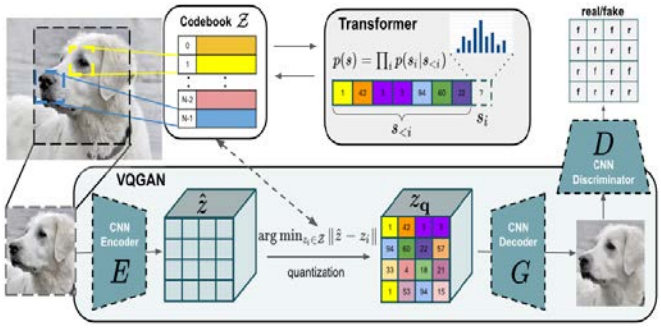


图 3 Taming Transformers 结构图

论文地址： https://openaccess.thecvf.com/content/CVPR2021/html/Esser_Taming_Transformers_for_High-Resolution_Image_Synthesis_CVPR_2021_paper.html
代码地址： https://git.io/JLlvY

责任编委 贾同 李策

付陈平

博士研究生，大连理工大学国际信息与软件学院，研究方向为计算机视觉。

樊 鑫

教授，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

RGB-D 点集数据集

西安交通大学 杜少毅 万腾

随着视觉传感器的快速研发，一系列能够同时采集目标结构和颜色信息的 RGB-D 成像技术，已被广泛用于空间测绘、目标检测与追踪、骨架与面部识别、点集配准、物体和场景重建、相机及目标姿态估计、实时定位与建图等技术领域。RGB-D 传感器的工作原理主要包含 3D 结构光和飞行时间(Time of flight, TOF)两种方法。3D 结构光通过将红外光投射到对象物体，并使用相应的红外接收传感器对含有结构信息的光线进行采集。飞行时间方法通过传感器发射并接收光束，并计算该光束从发出到被物体反射回来后的时间，从而推算相机到物体的距离。常用的 RGB-D 传感器为微软 Kinect、华硕 Xtion、奥比中光 Astra 以及英特尔 RealSense 等，如图 1 所示。

本文中重点介绍点集配准和三维重建所用的 RGB-D 数据集，包括 RGB-D Object、NYU Depth V1、SUN RGB-D、RGB-D SLAM and Benchmark 以及 CoRBS 数据集。此类数据集可用于多源配准，例如，我们最近工作中尝试使用基于显著目标的配准方法(RGB-D Point Cloud Registration Based on Salient Object Detection, TNNLS 2021)，获得了较好的结果。

(a) Kinect (b) Xtion

(c) Astra (d) RealSense

图 1 RGB-D 传感器

1、RGB-D Object 数据集

介绍： RGB-D Object Dataset 是一个包含 300 个常见室内物体的大型数据集。数据包含 51 个类别，使用 WordNet 上位词-下位词(Hypernym-hyponym)关系排列，并使用类 Kinect 风格的 3D 相机记录，该相机以 30Hz 同步记录与对齐 640x480 RGB 和深度图像。每个对象都被放置在转盘上，并在整个旋转过程中捕获视频序列。每个对象的数据都包含 3 个视频序列，每个视频序列都是用安装在不同高度的摄像机记录的，以便从不同的角度观察对象。

除了 300 个物体的孤立视图之外，该 RGB-D 目标数据集还包括带注释的 22 个自然场景视频序列，同时包含该数据集中的目标。这些场景涵盖了常见的室内环境，包括办公室工作区、会议室和厨房区域。对象从不同的视点和距离可见，并且在某些帧中可能部分或完全被遮挡。该数据集的部分样例数据，如图 2 所示。

图 2 RGB-D Object Dataset 数据集

数据集地址： http://rgbd-dataset.cs.washington.edu/dataset/

2、NYU Depth V1 数据集

介绍：NYU-Depth 数据集由来自各种室内场景的视频序列组成，这些视频序列由 Microsoft Kinect 的 RGB 和深度相机记录，该数据集由以下几个组成部分：

标签数据：标签数据集是原始数据集的子集，它由成对的 RGB 和深度帧组成，这些帧已被同步，并为每张图像添加了密集标签，标记数据被保存在 Matlab.mat 格式的文件中。

原始数据：包括由 Kinect 提供的原始 RGB 图像、深度图像和加速度计数据。RGB 相机和深度相机采样率介于 20 到 30FPS 之间(随时间变化)。虽然数据帧不同步，但每个 RGB、深度和加速度计文件的时间戳都包含在每个文件名中。

操作工具:提供了用于操作数据和标签的 API 函数。

该数据集的部分样例数据，如图 3 所示。

图 3 NYU Depth V1 数据集

数据集地址：https://cs.nyu.edu/~silberman/datasets/nyu_depth_v1.html

3、SUN RGB-D 数据集

介绍：SUN RGB-D 数据集为实现基于 RGB-D 数据的高级场景理解而设计，其中包含用于训练的 3D 注释和用于评估的 3D 指标。

该数据集由四个不同的传感器采集场景所得，包含 10,000 张 RGB-D 图像，并且全部数据均经过密集注释，包括 146,617 个 2D 多边形和 58,657 个具有准确目标方向的 3D 边界框，以及场景的 3D 房间布局和类别。该数据集适合场景理解任务训练需要大量数据的算法，使用明确的 3D 数据评估指标对其进行评测，避免过度拟合到小型测试集。此外，该数据集也可用于研究跨传感器偏差。该数据集的部分样例数据，如图 4 所示。

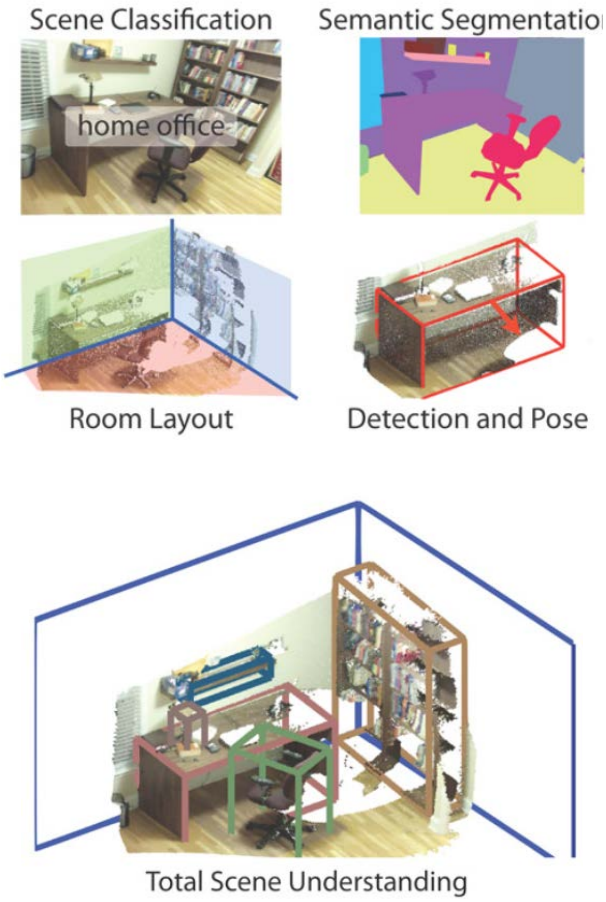


图 4 SUN RGB-D 数据集

数据集地址：http://rgbd.cs.princeton.edu/

4、RGB-D SLAM Dataset and Benchmark

介绍：RGB-D SLAM Dataset and Benchmark 数据集是一个包含 RGB-D 数据和地面实况数据的大型数据集，旨在为视觉里程计和视觉同步定位与重建系统的评估建立一个新的基准。该数据集包含由 Microsoft Kinect 传感器采集所得的传感器真实轨迹、颜色和深度图像。数据以全帧速率(30Hz)和传感器分辨率 640×480 记录。地面实况轨迹是从具有八个高速跟踪摄像机(100Hz)的高精度运动捕捉系统获得的。此外，该数据集提供了来自 Kinect 的加速度计数据，并提出用于衡量视觉 SLAM 系统估计的相机轨迹质量的评价标准。该数据集部分样例数据，如图 5 所示。

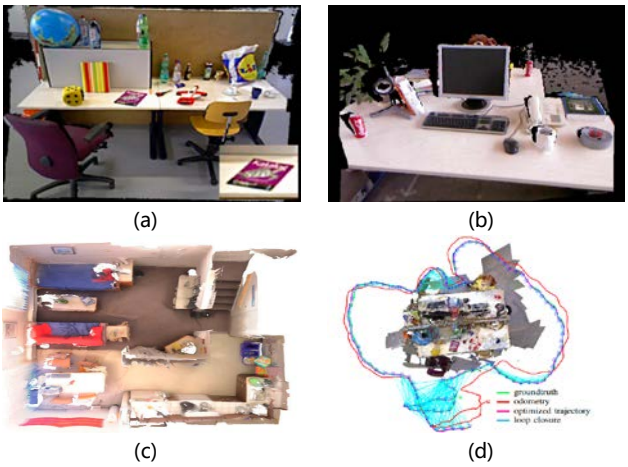


图 5 RGB-D SLAM Dataset and Benchmark 数据集

数据集地址：https://vision.in.tum.de/data/datasets/rgbd-dataset

5、CoRBS 数据集

介绍：CoRBS 数据集为同步定位与重建系统提供新的综合 RGB-D 基准。该数据集是场景真实深度和颜色数据的组合，并包含相机的地面真实轨迹以及场景的地面真实 3D 模型。该数据集使用外部运动捕捉系统获得了轨迹的基本事实，并通过外部 3D 扫描仪获得了场景几何的基本数据，每个都具有亚毫米精度，数据集包含使用 Kinect v2 捕获的四个不同场景的二十个图像序列，并在全局坐标系中提供所有数据，无需任何进一步的校准或校准即可直接评估，数据集部分内容如图 6 所示。

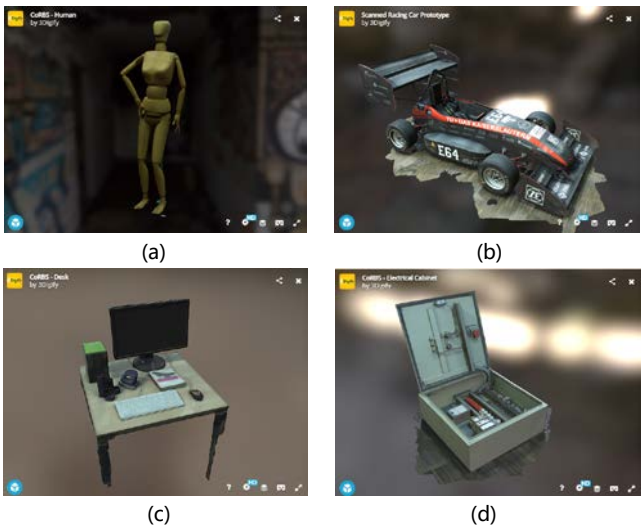


图 6 CoRBS 数据集

数据集地址：http://corbs.dfki.uni-kl.de/?pagerd_5fiilg

责任编委 樊鑫 沈沛意

杜少毅

西安交通大学人工智能学院教授、博士生导师。研究方向为图像点集配准、智能驾驶、医学影像处理及人脸识别等。
电子邮箱：dushaoyi@xjtu.edu.cn

万腾

博士研究生，西安交通大学人工智能学院，研究方向为 RGB-D 点集配准
电子邮箱：wanteng2017@stu.xjtu.edu.cn

好 文 推 荐

重庆大学、中科院和重庆邮电大学团队联合研究的“任务集成网络：联合检测和检索的图像搜索” 最新成果发表在 IEEE TPAMI 2022。

在许多现实场景中，目标(如人、车辆等)很少被准确地检测或定位。因此，作者设计了一个集成网络 I-Net 在没有标注的情况下，解决多任务集成的图像级检索(联合检测和检索的)问题。I-Net 的主要贡献如下：

论文：Lei Zhang, Zhenwei He, Yi Yang, Liang Wang, Xinbo Gao. Tasks Integrated Networks: Joint Detection and Retrieval for Image Search, IEEE TPAMI, vol. 44, no. 1, pp. 456-473, July. 2022.

1) 设计了孪生结构，并对给定图像中的相似和不相似目标设计了在线配对策略。通过孪生结构，I-Net 可以同时学习目标检测和分类任务的共享特征表示。

2) 提出了一种新的在线配对(OLP)损失，通过动态特征字典自动生成一定数量的负样本对来限制相似样本对，从而缓解了多任务训练停滞问题。

3)为了提高分类任务的鲁棒性，提出了一种基于 hard example 先验(HEP)的 softmax 损失。I-Net 的共享特征表示可能会限制检测和检索任务之间特定任务的灵活性和学习能力。因此，基于分而治之的理念，提出了一种改进的 I-Net 网络 DC-I-Net。

DC-I-Net 的贡献主要有两点：1)两个模块在集成框架中分别处理不同的任务，从而保证了不同任务的规范性；2)利用存储的类中心，提出了一种类中心引导的 HEP 损失，得到类内相似性和类间相异性，以便于进行最终的检索。

所提算法的整体框架如图 1 所示。I-Net 和 DC-I-Net 的本质区别在于两个方面。1) I-Net 的检测和重识别在不同的层中分别进行处理。2) DC-I-Net 进行重识别时，用 two-stage 来细化目标。

实验表明，在图像级的 CUHK-SYSU、PRW 等行人检索数据集和大规模的纹身搜索数据集 WebTatto 中，DC-I-Net 优于最先进的任务集成和任务分离图像搜索模型。

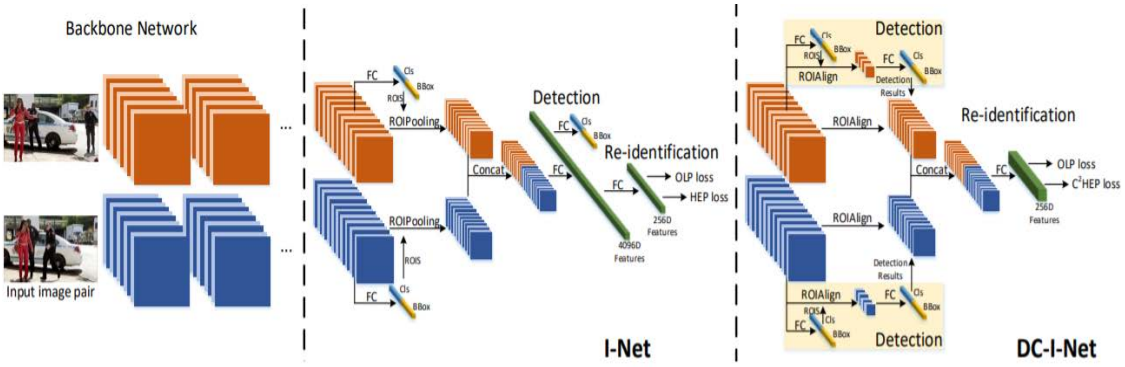


图 1. 所提算法框架。

利用包含相同对象的图像对进行训练，I-Net 和 DC-I-Net 用共享权重的同一骨干网络(左侧)进行特征提取。I-Net 用两个区域建议网络来获取图像中目标的建议框，并将 ROI 池化层生成的建议特征连接起来送到全连接层中进行检测和检索。DC-I-Net 将每个分支的特征都输入到全连接层，从而获得精确的检测结果。然后，将基于细化检测结果的 ROI 对齐层生成的目标特征串联起来，送入另一个全连接层进行检索。

责任编委 李策 樊鑫

好 文 推 荐

上海交通大学团队“使用不可靠伪标签的半监督语义分割”最新成果发表在 CVPR-2022。

论文：Yuchao Wang, Haochen Wang, Yujun Shen, Jingjing Fei, Wei Li, Guoqiang Jin, Liwei Wu, Rui Zhao, Xinyi Le. Semi-supervised Semantic Segmentation Using Unreliable Pseudo-Labels. CVPR, 2022.

语义分割是计算机视觉领域的一项基本任务，现有的基于有监督的语义分割方法都依赖于大规模带标注的数据，但获取数据的成本太高。针对上述问题，人们进行了许多尝试来实现半监督语义分割，即学习只有少量标记样本和大量未标记样本的模型。在这种情况下，如何充分利用未标记的数据变得至关重要。目前主流的解决方案是为未标记的像素分配伪标签，使用置信度高的预测结果作为伪标签。然而，仅使用可靠的预测会导致潜在问题是，在整个训练过程中可能永远无法学习某些像素，这可能导致训练类别严重失衡。

为此，上海交通大学团队提出了一种使用不可靠伪标签的替代方法。该方法框架如图 1 所示。该模型采用经典 self-training 框架，由 teacher 和 student 两个结构完全相同的网络组成，teacher 通过 EMA 的形式接受来自 student 的参数更新。损失函数通过交叉熵作为度量标准，将像素点分为可靠和不可靠像素两组，其中，所有可靠的预测都作为推导正样本的伪标签，而不可靠的像素则被推入负样本的记忆库。为了避免所有的负样本伪标签只来自类别的子集，该方法对每个类别使用一个队列，这样的设计保证了每个类的负样本数量是平衡的。同时，考虑到伪标签的质量随着模型精度的提高而提高，为此，本文还提出了一种自适应调整可靠和不可靠像素划分阈值的策略。

本文框架的性能优于许多现有的先进方法，为半监督学习研究提供了一个全新的范例。在半监督语义分割领域中具有较为明显的优势，为该领域发展提供了一定的推动力。

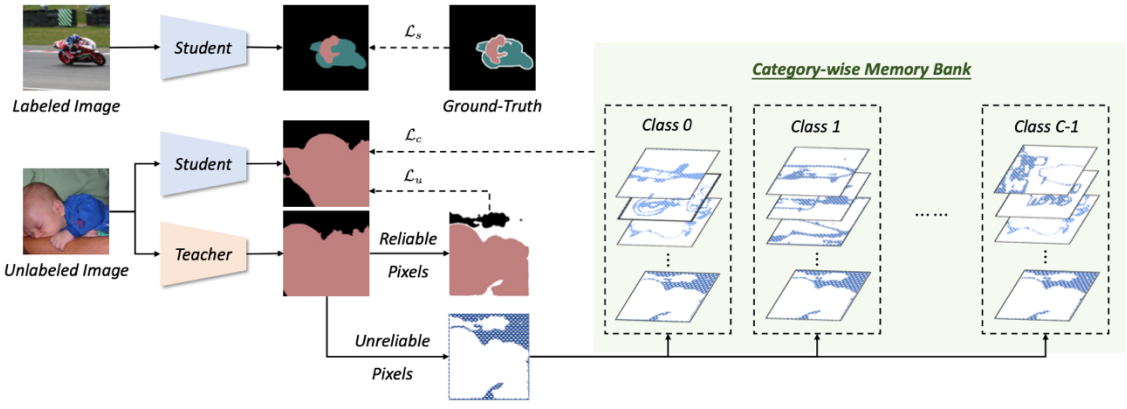


图 1 基于不可靠伪标签的半监督语义分割算法流程图

责任编委 沈沛意 贾同

好 文 推 荐

清华深圳国际研究生院团队“基于重点与全局知识蒸馏的目标检测”最新成果发表在 CVPR-2022。

论文：Zhendong Yang, Zhe Li, Xiaohu Jiang, Yuan Gong, Zehuan Yuan, Danpei Zhao, Chun Yuan. Focal and Global Knowledge Distillation for Detectors. CVPR, 2022.

目标检测是计算机视觉十分重要的研究方向之一，其广泛应用于机器人导航，工业检测，航空航天等诸多领域。为了获得更好的性能，通常需要使用更深层的网络，但需要大量的计算资源和推理时间。为解决上述问题，知识蒸馏的概念被提出。该方法在不增加额外开销的情况下获取较强的性能。然而，类别不均衡极大影响了知识蒸馏在目标检测中的效果。

为此，清华深圳国际研究生院团队提出一种针对目标检测的重点与全局知识蒸馏的方法，如图 1 所示。其中，知识蒸馏旨在使学生学习教师的知识，以获得相似的输出从而提升性能。通过分析特征层面的可视化结果，在空间与通道注意力上，教师模型与学生模型均存在较大的差异。其中在空间注意力上，二者在前景中的差异较大，在背景中的差异较小，这使得在蒸馏中学生模型的学习带来不同的难度。

针对学生与教师注意力的差异，前景与背景的差异，本文提出了重点蒸馏模块来对前景与背景进行分离，并利用教师的空间与通道注意力作为权重，指导学生进行知识蒸馏，计算重点蒸馏损失。然而，重点蒸馏模块在将前景与背景分开进行蒸馏的同时，也同时割断了前背景的联系，缺乏全局蒸馏特征信息。为此，本文还提出了全局蒸馏模块，利用该模块分别提取学生与教师的全局信息，并进行全局蒸馏损失的计算。在此基础上，提出利用重点和全局蒸馏对学生模型检测器进行引导的方法。通过在 COCO 上的大量实验，本文验证了方法在各种检测器上的有效性，包括单级、两级、无锚方法，实现了最先进目标检测性能。

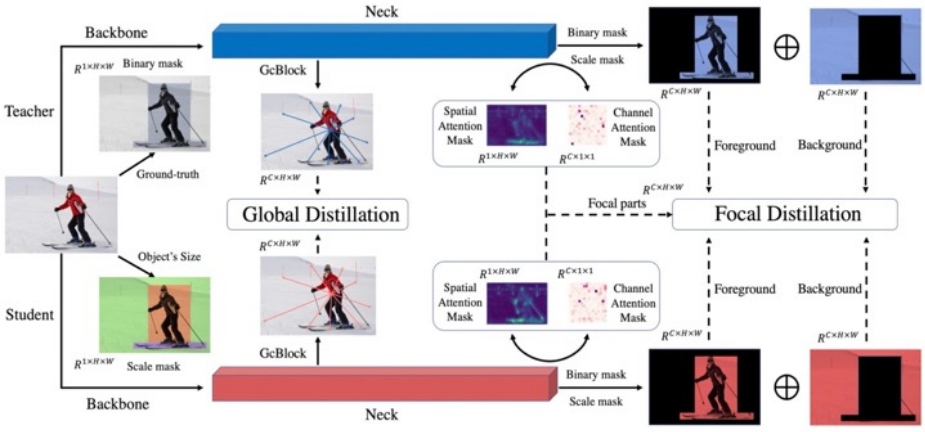


图 1. 基于重点蒸馏与全局蒸馏的目标检测框架流程图

责任编委 沈沛意 樊鑫

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 Machine Learning（ML），Pattern Recognition Letters（PRL），Journal of Computational and Applied Mathematics (JCAM) 和 IEEE Journal of Biomedical And Health Informatics（JBHI）。

3 会议简介

国际多媒体博览会（IEEE International Conference on Multimedia and Expo）是 IEEE 年度举办的学术性会议，是多媒体研究领域旗舰类国际学术会议之一，会议的主要内容是多媒体信号处理技术。

2022 年 ICME 在中国台北举行，本届会议将汇聚全世界从事多媒体理论与应用研究的广大科研工作者及工业界同仁，共同分享多媒体研究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
MM 2022	2022.10.10-14	Lisbon, Portugal	2022.04.08	https://2022.acmmm.org/
PRCV 2022	2022.10.14-17	Shenzhen, China	2022.04.15	http://www.prcv.cn/
ICML 2022	2022.11.07-11	Baltimore, India	2022.05.14	https://icml.cc/Conferences/2022
NeurIPS 2022	2022.09.28-10.9	Louisiana, USA	2022.05.20	https://neurips.cc/Conferences/2022/

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
ML	Special Issue on Imbalanced Learning	https://www.springer.com/journal/10994/updates/19536896	2022.04.04
PRL	Face-based Emotion Understanding	https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/face-based-emotion-understanding	2022.04.20
JCAM	Applied Analysis, Computation and Mathematical Modelling in Engineering	https://www.journals.elsevier.com/journal-of-computational-and-applied-mathematics/call-for-papers/special-issue-on-applied-analysis-computation-and-mathematical-modelling-in-engineering	2022.05.01
PRL	Recent Advances in Deep Learning Model Security	https://research.com/special-issue/recent-advances-in-deep-learning-model-security	2022.06.20
JBHI	AI-driven Informatics, Sensing, Imaging and Big Data Analytics for Fighting the COVID-19 Pandemic	https://www.embs.org/jbhi/special-issues-page/ai-driven-informatics-sensing-imaging-and-big-data-analytics-for-fighting-the-covid-19-pandemic/	2022.06.30

心底无私视界宽 – 施鹏飞教授专访

本栏目是期望从计算机视觉及相关领域的老前辈那里，获取一些历史回忆，从而使本领域的研究人员和爱好者能够了解计算机视觉在中国的发展历程以及老前辈们的贡献，让专委会积累一些历史资料。同时，也希望基于他们的经验和视角，来探讨计算机视觉及相关领域的发展现状、优势与不足。以及，分享他们在教书育人方面的成功经验。

本次专访的，是上海交通大学的施鹏飞教授。我是负责本次专访的采访人，复旦大学张军平。我跟施老师很早就认识。记得我第一次与施老师见面，是 2004 年 5 月 17-19 日去韩国首尔参加第六届自动人脸与姿势识别国际会议(IEEE Conference on Automatic Face and Gesture Recognition)。我在从上海到首尔的飞机上，见到了施老师，那时的他已经是赫赫有名。第二次则是约两年后，2006 年 1 月 5-7 日在香港理工大学召开的生物认证国际会议(IAPR International Conference on Biometrics)。 我们还有了一张珍贵的合影。合影里，除了施老师，还有北京大学的迟惠生教授、中科院计算所的陈熙霖研究员，都是计算机视觉的领军人物。后面在上海，在复旦，都见过多次。

图 1 上海交通大学施鹏飞教授

施老师是我一直很敬仰尊敬的，所以这次也非常荣幸由我来进行专访。为了这次采访，施老师特地从家乘车来到复旦大学江湾校区。本次的专访内容，是通过问答方式，经录音整理后，由施老师核对后完成的。

施鹏飞教授简介:

施鹏飞，1939 年 12 月生，上海市人。教授、博士生导师。 1962 年毕业于上海交通大学电机系，1963 年考入上海交通大学电器专业研究生，毕业后在上海交通大学电工系及计算机系任教。1979—1980 年赴美国王安电脑公司及匹兹堡大学进修。1993 年赴加拿大麦吉尔大学合作研究，1994 年赴俄罗斯莫斯科动力学院、德国柏林工业大学访问，1996 年赴法国巴黎第六大学访问，1998 年赴比利时根特大学访问。1981 年任计算机图象处理研究室主任，曾任电子信息学院副院长。曾任上海交大图像处理与摸式识别研究所所长、电子信息学院副院长、上海市模式识别专委会主任、中国人工智能学会常务理事、国际 IEEE 高级会员，受聘为国家自然科学基金评议组成员、科技部 973 项目信息领域咨询专家。曾获国家、省、部级科技奖 5 项，编著教材 5 本，发表学术论文百余篇，培养博士 60 名、硕士 50 余名。

施鹏飞教授主要从事数字图像处理，模式识别及人工智能领域的教学和研究工作。

图 2 第六届生物认证国际会议，香港，合影。由左到右：陈熙霖、迟惠生、施鹏飞、吴仲城、张军平、甘俊英

为能更好地帮助我们回顾本次采访，以下**我们采用了问答加书面回顾的形式来表述。**

张军平(采访者，后缩写为张)：您是 1965 年于上海交通大学电机系研究生毕业，然后直接留校任教。想请您介绍一下，60 年代读研究生的人并不多，您为什么选择大学毕业继续读书。您觉得您那个时候研究生学习和现在有什么区别？对现在求学的学子您能给些建议吗？

施鹏飞（后缩写为施)：我的本科毕业是 1962 年，上海交大电机系。

我把这个情况简要地回顾一下。我是 1957 年考上交大。1957 年，在我们国家高考历史上是一个马鞍型发展的阶段，即前后两年招生人数都多，但 57 年极少。1957 年招生，我的印象就十万八千。所以，57 年能上大学的，都很不错。

我 62 年大学毕业的时候，当时我们都是国家分配的，全国都是统一招生，统一分配。那么，我觉得因为统一分配，当然服从国家需要，这个我有思想准备。但是呢，当时也不是完全你自己决定。当时正好研究生开始恢复招生。所以我觉得，留学校继续读研究生也可以考虑。因为，我也希望能够再学一点东西，也喜欢学校的环境。在这样两个前提下，我就选择了读研究生。

图 3 施鹏飞教授和张军平的合影（2022 年 3 月 4 日）

确实，当时招研究生的数量不多。因为本科生本身就很少，加上研究生刚才恢复。还要国家有需要，也不可能都像现在这样招大批研究生。以我们交大为例，我们当时招收电机类研究生，现在叫做信息类，就只招 6 个研究生，而整个交大也才 100 多个研究生。

第二个原因为什么这么少呢？因为刚恢复研究生制度，当时教育部呢，对导师要求很严格，必须是教授指导。当时，我们还是硕士，不是博士，而研究生的导师必须是正教授。在 1962 年的时候，正教授在大学里面是很少的。我就选择继续在交大读研究生，读了三年。

那么，这三年呢对我来说，现在回忆起来，对我一生的成长、到后面能够搞新的领域，都打下了比较坚实的基础。特别因为我们那会的学生，从中学开始，一直是学习俄语。然后在研究生的时候，要学习英语了。所以，这个英语的基础呢，就从大学到研究生阶段我就把它补上了。

另外，对数学方面，尤其一些电的相关理论基础，学得跟大学就不一样了。当时像数学方面，我们学到了泛函分析、概率统计等等，这些都为后面的研究打下了基础。当然，现在学校的环境、硬件条件，那跟以前完全不一样了，可以说是天壤之别。

至于说有什么建议，可以参考我上面讲的这些。当然，也不是每个本科生都要读研究生，还是得根据你的情况，根据你的兴趣，根据你的爱好。有条件的，我觉得能够读一下研究生，就应该很好地利用这个机会。比如现在很多领域，希望解决关键技术也好，基础研究也好，本科可能还不够，应该在这些方面继续加以深造。可利用这个阶段加强国际交流跟合作，对提升我们培养

的质量，也是有帮助的。

张：您是什么时侯开始从事医学影像领域？是什么样的契机开始从事这一领域？从事交叉领域和从事经典图像领域，您认为研究方法或者研究重点有什么异同吗？您认为这一领域未来发展重点和方向是什么，有什么建议给这一方向的研究者？

施：我从图像处理说起。1980 年，我到美国访问。因为刚改革开放不久，交大校友建议我们，应该开展计算机方面的研究。当时给交大提了四个大方向，第一个是计算机网络，第二是大规模集成电路，第三是光纤通信，还有一个就是图像处理跟模式识别。

现在看来，这四个方向我认为还是对的。考虑到我的基础跟条件，因为我是电气系毕业的，所以做图像处理，我觉得比较适合的。因为无非从信号处理，从数学来说无非学代数和一些变换，是比较容易转的。而网络，因为当时计算机还没有，我觉得很难开展工作。通信光纤是刚刚才起步，而且我的基础也不是通信专业，所以我觉得也有困难。还有一个是集成电路。集成电路当然也可以考虑，但是它要制造芯片，要有外部环境的条件。所以我当时就选了图像处理。

那么图像处理，我认为原来做信号处理是一维的，图像无非两维的。所以我很容易上手。但是当时我不是马上就做医学图像，最早我还是做文字识别，做中文 OCR，还有就是指纹识别。这些应用，可以做到工业检测方面。最早的文字设备在邮政编码。邮政编码就是 0-9，简单。后来医学图像开始了，当时去国外访问交流。而国内当时计算机设备都没有，医院里面更加没有，开展研究确实有困难。

但是当时我觉得，模式识别里面的图像分割、图像检测值得做。另外，乳腺癌国外研究的也多，比如检测早期的小结节。虽然当时图像分辨率比较有限，但是可以做。所以这个也尝试了一下。当时也比较容易得到数据，国外的新病例很多。国内因为当时还不是太重视，数据还不是太多。

后来，通过和大家交流，特别国外的访问交流，我觉得医学图像是计算机视觉里面一个重要的应用领域。

第一，对健康来说，医学的问题，治疾病的问题，它是永恒的主题。不管社会制度怎么样，不管人的条件怎么样，人类要健康，那都跟医学影像有关系，所以我觉得医学图像是很重要的方向。

第二，它是交叉学科，跟经典的图像处理不一样。它可以是功能支点的，即根据需求出发，如医院的临床。当然，其他的应用如工业检测，工业上应用也要根据他的设计需要。但是，医学图像更加突出，一定要跟医院结合，跟医生结合，深度交流，这样才有共同的语言，才能把我们能够解决的问题告诉医生。医生需要我们解决的，我们也能够知道，所以这个我觉得很关键。要做的好，前提要深度交叉。如果闭门造车，方法当然也有参考价值，但是我觉得真正要应用的话，还是有很大距离。

张：您认为这个领域的重点发展方向会是哪些呢？

施：这个问题呢，要回到我为什么要做。因为这个题目，本身在不断地发展。有些问题还没有解决。特别随着医学技术的进步，医学影像分析也在进步。以前我们都叫用形象，譬如说这个肿瘤。它的形状怎么样？表面是不是光滑？它的大小怎么样？它的纹理结构怎么样？从这些方面，那还是比较粗的。那么现在呢，随着这个影像设备的提升，我们一般的从 CT，从 MR 图像从 P ET 图像，到现在发展到分子医学，一直到基因。那么，这些技术的发展对我们提出更高的要求，必然要求我们的医学图像处理，它的精度、准确性应该更高，处理的速度应该更快。另外呢，比如有些设备对人类的伤害要小，譬如低剂量 CT 图像。这又分两方面，一方面，他们专门搞 CT 设备，可以从硬件来解决。另一方面，从计算机视觉角度，我们的软件要做得更好，算法要设计得更好，所以我觉得这也是一个方向。

所以，不用发愁这个方向会不会到头了，要不断地为它做下去。我也希望大家能够进来医学图像领域接着做。

张：您刚开始从事计算机视觉研究时，有没有特别有趣和艰苦的事值得分享？我看您的介绍，当时还去过彭加木走过的路线。您觉得做实际应用，和理论研究之间有哪些不同呢？

施：上次你说我们学校采访过我，那里比较详细，所以有兴趣的可以去看看。关于计算机视觉领域，从广义的如人工智能这个角度来看，我一直认为，在这个领域里面，包含两大方向，一个是计算机视觉。第二个就是自然语言处理，因为从感知的角度来说也是这样。一个是视觉，一个是听觉。再从人工智能来说，人有认知首先要有感知，没有感知怎么认知呢。所以，它需要计算机视觉。我从 80 年代初开始做图像处理研究，虽然不完全是计算，但是应该也算计算机视觉。

国内还有不少，比如复旦的吴立德老师，他最早在 PAMI 上发表文章。我觉得他数学基础好，那这也是计算机视觉的一个方面。

另外，我以前走过彭加木的路线。这个细节，我可以讲一讲。当时，交大的王震，就是新疆的王胡子。他跟小平改革开放。当时他推动这个，我觉得他起了很大的作用。他到交大说，你们不敢搞，我来。你们搞个董事会，我来当董事长。那时，在学校搞董事长，人家都不大理解。另外，他说新疆需要有水资源，没有水，新

图 4 施鹏飞老师在塔里木骑马

图 5 当时在塔里木的科研团队合影

疆谁去也不好办。所以，当时首先要把新疆的这个水资源搞清楚，换句话说，去沙漠看看里面的自然资源到底怎么样？要摸清楚，但因为塔里木湖很长，又在塔克拉玛干沙漠里面，那么怎么做这个问题呢？当时，正好我们搞图像处理。那么利用航空遥感，这个是低空的。然后，加上图像处理。当时我们从法国引进了一套设备，这个机器很小。我记得，即使 256*256 像素 8 个比特的图像也要处理很长时间，但是，能用。通过彩色处理后，我会打印出来。但是这中间，还有很多环节是人机交互的。也不完全像现在的计算机，所以需要通过我说的这些流程处理。

另外，这样的项目一定要跟实际结合。所以，当时我们跟新疆地理研究所、新疆土壤研究所、新疆测绘所一起，当地政府也都大力支持，前后花了三年时间。做完了一个阶段后，我们还要到实地考察图像处理的结果跟实际是否符合。比方说在沙漠里面的胡杨林，它是一种耐干的植物。他到底怎么样、分布在什么地方？这些我们都要实地考察。那么，实地考察怎么去呢？当时确实有当地的农民带队，要派骆驼，要卡车。住的条件也没有很好，都在帐篷里面，都在沙漠里面。沙漠里面天又热，这个环境条件那是可想而知的。

当年彭加木其实也走了这条路。所以，在那的三年时间，我觉得因为当时我比较年轻，当时大概 40 多一点，能够承受这样的环境和锻炼。当然，我也拍了一些

照片，如图 4、5 所示。

这个与实际问题相关的项目，是国家需求的，到现在为止也确实很重要。但这种项目不是个人想做就能做的。第二呢，确实需要个团队，而不能是一个人，因为当时我们这个团队，都是多学科交叉，是计算机学科和地理学科结合的。要下去考察，这是一个团队性的项目，跟我们一般的基础研究或者我们学校的还是不一样。

所以，我建议我们高校里面，或者现在我们很多研究机构有条件的，能够承担一些国家的需求跟基础研究，这两个方向我觉得都是值得大家考虑的。

张：您在美国、加拿大、法国、瑞典等国家均做过研究工作，能帮我们分析一下不同国家开展科研的特点和特色有哪些么？其中有哪些是可以结合我们的国情进行推广和应用的？

施：改革开放以后，我就在交大，从毕业到现在没有离开过。我觉得当老师比较灵活。每年有假期，有很多学术会议，经常可以走一走。在改革开放后，80 年初，我有了第一次美国访问，那是 1979 年出去的第一批。然后我回来以后，做新疆的项目，在那做了几年。在完成新疆项目以后，我还得做基金。那需要做基础方面比较多一点，就做计算机视觉。那时，几乎每年都有机会到国外去。另外呢，我也争取了很多国际合作的项目，比方说，美国我去的次数比较多，大概有 8-10 次。而欧洲几乎主要的国家我都去了，还有澳大利亚，亚洲的日本、韩国等。俄罗斯、伊拉克这些国家都很少有人去的，我都走了一走。

但是我每一次去的时间都不长，最长的访问也是半年去到加拿大。所以，跟那些长期在国外的学者和现在都回国的学者相比，我还不能说的很好，因为我毕竟没有很长时间在国外待过。和那些有了七年、八年、十年以上的体会是不一样的。所以我的看法还是比较粗浅的。

但第一点我觉得，我们搞研究，无论哪个领域，国际合作跟交流是很重要的。因为知识和科学应该没有国

图 6 （左）施老师在巴比伦；（右）施老师在瑞典查尔姆斯理工大学

界，是人类共同的财富，所以我们要相互学习。特别我们这个领域，虽然改革开放以后，取得了很大进步跟发展，但是我们应该承认，在国际最先进和前沿的领域上，我们应该要有清晰的认识，还是有差异的。所以这是第一点，交流的必要性。

第二，应该要学习人家一些创新的精神。现在我们要能够独立思考，不能人家做什么，我们也做什么；人家做这个热点，我也做这个热点。当然有些学生刚入门，向人家学习一下，参考一下，这个也是应该的。但是我们一定要至少在人家中间，找出他们还有什么问题没有解决的。我们应该在这个方面有所改进，有所突破，这点我觉得是最重要。

还有一个，学生和我们的交流可能有时候还不够。我们很多国内的学生，开会也很少提问，我们上课的时候也很少提问。学生都不愿意提问，这个我觉得应该要改进。当然，交流是相互的。我们也有我们的长处，我们要有自信，不是说我们一味地学习人家。但是，我们首先要尊重人家，要加强这方面、抓住各种机会，才能够不断提升自己。

张：人才培养方面，您培养出很多优秀的学生，如上科大的沈定刚教授。如何成为相关领域的栋梁之材，您能给后辈们一些建议么？另外，如何因材施教，如何最大限度地挖掘学生潜力？

施：医学图像这块，我觉得沈定刚教授其实做的很不错。为了这次采访，有些问题我怕把握不准，所以我专门到沈定刚那边去了一次。我跟他也像今天这样，也交流了大概半个多小时，有些问题我也听听他的意见。

沈定刚呢，本科硕士博士，全是在上交大这读的。他在交大的时侯，我们那叫图像处理模式识别研究所。他是我们所的博士生，所以对他的情况比较了解。当他交大毕业以后，我们给他提供机会到香港城市大学，然后到新加坡，通过三年时间，他把国内跟国外的差异填平了。然后呢，他去了美国。首先到霍普金斯大学做医学图像，然后到宾夕法尼亚大学继续做医学图像，最后到北卡罗来纳大学。那么 2020 年，他回到上海科技大学，他现在的 title，我问他你的 title 是什么呢？他说，在学术界，就是教授、博导。另外是 BME，即生物医学工程的创始院长。这是上科大的体制，叫创始院长，因为是他建立了这个学院。在工业界呢，他是医疗 AI 公司联影智能的联席 CEO。另外，沈说可以看看他们的环境和条件，他认为是世界一流的。

他还深度参与了一个很大的医学影像计算相关的会议，叫 MICCAI （International Conference on Medical Image Computing and Computer Assisted Intervention），是 MICCAI 的 Fellow。另外值得一提的是，MICCAI 的 Fellow 之一叫杨广中，他是交大的医疗机器人研究院的创始院长。他也是原上交大图像所硕士生，被选派到英国帝国理工大学深造。目前他是英国皇家院士。

那么沈定刚是怎么成功的呢？我觉得他的成功，一个是因为他很刻苦，他的基础当然也很好。他是浙江慈溪人，来上交大后，从本科电子系开始。除了基础好和刻苦外，他也很关注学术方面的一些进展。所以说，他的成功，更主要还是他自身的努力。第二，改革开放给他提供了很好的学习机会和环境，第三，他的研究方向选得好，机会抓得牢。我认为这些都是必不可少的。有基础，有这个条件，自己还要抓住机会。

我认为这几个方面使得他走到了今天这样的位置。现在的他，可以说是在医学影像处理、人工智能领域的国际知名学者。那我问他，从国际的视野来看，能够跟你相提并论的人多不多。他说，不能绝对说没有，但很少，至少现在我已经超过我的导师。那我觉得应该是这样，青出于蓝而胜于蓝。

另外，怎么样发挥学生的潜力呢，他也很重视。我会在后面的问题里再说说怎么样培养。

张：您怎么看目前国内计算机视觉领域的发展现状？有哪些优势和不足？

施：关于这个问题，我做了很多思考。要讲很容易，但是要讲得好也不容易。

当前，我们计算机视觉似乎就是深度学习。这当然是一个重要的方面。因为机器学习在人工智能里面，它是一个突破点。有了深度学习以后成为转折了，特别是人工智能有了深度学习以后，因为深度学习在应用领域得到了很好的验证。现在，很多人重视这方面的研究。

但是，从计算机视觉角度，从人工智能角度，关键还是要解决黑箱的问题。到底它的理论在什么地方？但是，包括我们国内的姚期智先生，他几次提过这个问题。

另外，我觉得，我们重视计算机视觉的人可能还不够，要多学科交叉，特别要从数据角度入手。一定要与数学领域的一些人士进行更好的合作。我们要从更高的角度，从视觉感知这个角度。也一定要跟那个脑科学、神经科学和心理学的专家做交叉结合。这样才能让我们国内计算机视觉领域有更宽广的事业，我觉得这些都很重要的。

当然，应用一些深度学习，搞个好的算法，或搞个开放的平台，能够让这些算法更加通用，也不错。而更具解释性，我认为还有很多工作好做。

另外，这几年我们国内的学者在国际顶尖的计算机视觉会议上面，如 ICCV，ECCV，ACCV，CVPR 和 PRCV 上，我们国内的学者都发表了很好很多的文章，也希望能够继续。但更关键的是，希望我们能够做出更有影响力的贡献，能够在关键性的技术方面有更大的突破，也要提高我们的创新性。

图 7 （左）施老师和马颂德老师合影（左一）；（右）施老师和查红彬老师合影（右一）

张：最近朱松纯教授在 PRCV 会议上说，国内计算机视觉都在做深度学习，但却太不愿意碰一些困难的问题，您是怎么看的?

施：朱松纯教授，他是华中那边过去的。很多年前他还在 UCLA 时，我有次去洛杉矶时拜访过他，他回国后我还没见过他。他是搞统计的，在计算机视觉方面有一些独到的见解。他的这个观点，我也赞同。

张：在您刚开始从事计算机视觉研究的时期，国内还有哪些您觉得做得不错的前辈呢?

施：我回顾一下，因为计算机视觉都是从我们这个年代开始的。而且是在改革开放 80 年初开始的，所以这个历史还比较清楚。

这些前辈大部分都还健在，而且有些以前都有很多交往。所以前几届的 PRCV 大会分别给袁保宗、徐光祐、吴立德教授及马颂德、杨静宇老师颁发了 CCF-CV 的 " 终身杰出贡献奖” 。

这些都是我们同时代的，我们也都有交往。

另外，90 年代初，国家确定在北京中关村建 3 个信息类的国家实验室，分别是清华（张钹老师）牵头的智能技术与系统实验室，北大（石青云老师）牵头的视听实验室，还有中科院自动化（马颂德老师）牵头的模式识别实验室。这三个实验室不仅促进了模式识别和计算机视觉领域的发展，而且推动了国际的合作和交流，至今还发挥着重要的作用。除了三位牵头的老师外，这些实验室人才云集，我也通过不断的交流访问，认识了许多中、青年的后起之秀。

还有，中科院的陆汝钤老师，我印象也很深。他的人工智能的教材书是写得最厚的，还有上下两册。但是这些年搞人工智能，我印象中，陆老师有些场合也很少出来了。

我们上交的李介谷教授和北交的袁保宗老师是同辈的，如今已 90 高寿了。不过，李老师退的比较早一点。另外，因为他年龄大了，所以在国内呢，大家可能不太知道，他也写过计算机视觉的书，培养了好多学生。他也是我的老师辈。

另外呢，我觉得还有原来西安交通大学的、现在同济大学宣国荣教授。他应该是郑南宁院士的老师，他爱人是研究自然语言处理的，现在也在同济大学。他跟黄泰翼、原中科院沈阳自动化研究所所长蒋新松（中国工程院首批院士）都是上海交通大学同班同学。他们都是前辈里面做得很好的。当然，南理工的杨静宇老师在模式识别和无人驾驶方面也做了很多成果。

张：您带出了不少优秀的学生，能否分享下指导学生的经验?

施：关于学生培养，我觉得当老师的呢，重要的职责是培养人，学校的职责也是培养人，这是第一位的。这个叫立德树人，就是品德，还有就是要业务方面能力。

那么怎么样培养的？首先外部环境要好。这个我想大家都具备，有相似的地方。但是对老师来说，除此以外，第一是要选好苗子。因为人都是不一样、有区别的，总有他的优点跟不足。有些学生智商高一点，这个也不可否认，但是后天的努力更重要。

其次呢，是教师的责任心。按照沈定刚教授的讲法，他认为培养学生呢，一定要用心思，要把心思放在学生的身上。那么，怎么放在学生身上的？他说，我的学生进来前，我挑选的时候，就是一定要对他了解。比如这个学生来自哪个学校，他有哪些特长、哪些爱好，然后你要充分挖掘他的长处，避免他的不足的地方，加以改进。通俗来讲，就叫因材施教，不能一个模子，做成统一的产品，这是第一点。

那老师也要做得好。比方说我的学生来了后，我得了解，你从哪里来的，甚至问你是哪个中学毕业的，你的高考成绩到底有什么原因获得的，你是怎么来的，你的家庭是怎么样？虽然从国外讲这个隐私不好问，但国内可能还不太在意这个。

因为这个背景我要了解，比如从农村出来的，他就比较吃得起苦，有些条件好一点的，可能这方面就差一点。但是有些从城市出来的，他的外语可能好一些，有些山区的，可能条件限制，外语可能能力差一点等等。所以第一我觉得你要了解学生。

然后来了以后，读研究生，你要给他挑选一个比较好的题目跟方向。当然这个我们也要考虑导师原有的基础条件，因为一般导师可能不是统一做一个方向，会有一定的面。那么，根据老师做的方向，和根据项目资助的情况来确定。假如这个项目是基金资助的，那么更适应做些基础性的研究。假如是合作的，那你可以让他做一些技术性强一些的项目。我觉得，老师要把这个做好，挑选合适的学生。还要不断调动他的积极性，才能使他通过比较短的时间，在半年、或者一年就能够走上科研的路。

那么到了一年两年后，比如博士到了两年以后，可能有一些小的文章就会出来了。那么，此时可以再努力一下。我想博士毕业的时候，作为一个博士来说，他至少应该在国际的一些顶尖会议或者杂志上面能够发表论文。我说我们好的学生，现在可以在我们计算机视觉的顶尖会议上面，比如最近的 CVPR 录用已经比较常见了，当然要在 PAMI 上面发表文章，那还是有点难度，但是我觉得也可以做到了。

张：对从事计算机视觉领域研究的青年学者和学生有没有寄语？

施：我们要培养人，首先要立德树人，在研究方面一定要有创新的思想。另外，希望青年学者能够扎入感兴趣和国家有需求的方向，在一两个方面做出突出贡献，并坚持下去。眼光放远一点，希望大的环境给青年人才更多的渠道、更多的机会，让他们能够更快成长。也希望学生们能多多利用和珍惜学校的多学科资源，能学到很多文化，多听讲座，多去图书馆。学好专业，开拓视野，广泛交流。

责任编委 张军平 明悦 贾熹滨

COMPUTER VISION NEWSLETTER

01 2022
总第 31 期

计算机视觉专委会简报

CCF 计算机视觉专委会