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 2023

总第35期

CCF 计算机视觉专委会

COMPUTER VISION NEWSLETTER

计算机视觉专委会简报

2023 年第 01 期

总第 35 期

CCF 计算机视觉专委会

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 走进企业系列交流会

第 24 期 中电科智能院

2022 年 12 月 28 日，由中国计算机学会计算机视觉专委会主办的第二十四期 CCF-CV 走进企业系列交流会——“CCF-CV 走进中电科智能院”在线上成功举行。本期研讨会由中电科智能院李明强博士、计算机视觉专委会副秘书长、南京理工大学潘金山教授等共同组织。

会议邀请了计算机视觉专委会副主任、南京信息工程大学刘青山教授和中电科认知与智能技术实验室张峰副主任致辞，清华大学苏航副教授、香港中文大学（深圳）吴保元副教授、复旦大学陈静静副教授、中科院自动化所张俊格研究员作了主题报告并参与圆桌讨论。本次会议由中电科智能院李明强博士主持。

随后的圆桌讨论环节由李明强博士主持，圆桌讨论主要围绕：人工智能算法在开放非平稳环境下的应用面临的问题、如何提升现有人工智能算法的鲁棒性以及图像或视频防御与攻击等问题，展开了热烈讨论。

第 25 期 航天宏图

2023 年 3 月 15 日，CCF-CV 专委会与航天宏图举办了“AI 赋能遥感新视界”Club 研讨会。中国计算机学会专职副秘书长王新霞、中国计算机学会计算机视觉专委会查红彬主任、航天宏图董事长王宇翔等领导，以及来自北京大学、中科院自动化所、中科院计算所、天津大学、北京邮电大学、航天 514 所、北京慧点数码科技有限公司、航天 514 所、北京维视智能科技有限公司、北京中科软科技有限公司、云知万象科技创新（北京）有限公司、蚂蚁集团等 40 余位高校和企业专家参加了本次活动，共同探讨计算机视觉在卫星遥感领域的“产学研”创新技术应用。

在茶歇环节，参会人员进行了参观合影，了解了航天宏图 SAR 卫星、PIE 无人机、实景三维，智慧城市灾害应急系统，PIE 软件服务等技术，体验了大屏数字人客服，混合现实数字孪生等相关产品。

责任编委 毋立芳

第 15 期 低质图像内容增强和感知

CCF-CV 视界无限系列研讨会

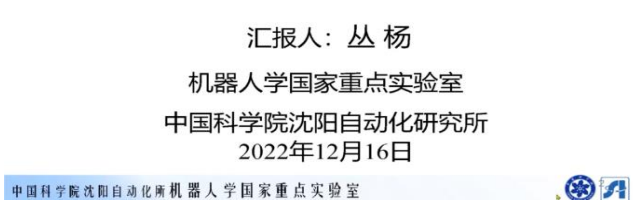
低质视觉内容增强与感知

2022年12月16日

2022 年 12 月 16 日，由中国计算机学会计算机视觉专委会（CCF-CV）举办的第 15 期 CCF-CV “视界无限” 系列活动——"低质图像内容增强和感知"研讨会在线上举办。研讨会邀请了专委会主任北京大学教授查红彬和中山大学网络空间安全学院院长操晓春致辞。中国科学院沈阳自动化研究所丛杨 、浙江大学李玺、北京航空航天大学徐迈、北京理工大学付莹、鹏城实验室杨文瀚做主题报告。中山大学任文琦副教授以及以上五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 900+。

查红彬主任介绍了活动的情况。本次研讨会是计算机视觉专委会举办的 “视界无限” 系列活动第十五期。这种座谈会的模式受到了许多的高校的欢迎，这种成功离不开广大同行的支持，查红彬主任对研讨会组织者、参加报告的讲者和各位参会老师同学同行表达了感谢，并预祝本次会议圆满成功。

机器人视觉感知及自主操作



丛杨教授的报告题目是 “机器人视觉感知与自主操作”。机器人感知和认知能力是智能机器人自主行为的关键，而视觉和机器学习是机器人感知和认知的重要手段。虽然近年来涌现出许多令人兴奋的进展，但机器人感知和认知中的一些核心问题仍然没有得到很好解决，导致机器人还无法完成很多人类看似简单的工作。这其中尚待解决的两个问题是机器人泛化能力较差和自主在线学习能力不足。对于深海 3D 感知难点，介绍了基于物理机制的深海高清成像方法、跨介质深海精细化测量三维测量。对于自主在线学习能力不足方面，介绍了自主在线学习框架，例如在线样本学习、特征增减学习、任务增量学习等。报告结合机器人学国家重点实验室的背景和特点，主要阐述针对机器人感知和认知中的视觉识别和在线学习问题所开展的探索性研究工作。最后对多元融合，任务驱动，人机安全做出展望。

付莹教授的报告题目是“噪声建模与图像增强”。从低信噪比的观测数据重建清晰图像是底层视觉研究中的一项重要任务，其在遥感、生物光学、诊断医学等领域有着广泛的应用。图像增强质量很大程度上取决于所采用的噪声模型以及图像先验的准确性，从成像传感器的物理特性出发，建模从光子到数字信号的电子成像物理过程中所涉及到的各类噪声源，例如，高斯噪声、泊松噪声、暗电流噪声、源随器噪声和行噪声等。介绍了噪声建模方法，噪声建模在图像增强的应用以及它们的未来发展方向。

徐迈教授的报告题目是“数据驱动下的压缩视频质量增强”。近年来，随着智能终端的发展以及在线视频等新型多媒体业务的普及，网络中所传输的图像视频数据量呈爆炸式增长的趋势，网络带宽供求矛盾日益尖锐，视频编码是网络带宽供求矛盾的关键技术。然而，高压缩比的视频压缩导致视频质量差，极大降低了视频用户体验。在这次报告中介绍了视频质量增强方面的研究工作。并且介绍了在解码端提升视频质量的方法，基于多帧联合优化的压缩视频质量增强技术，面向盲质量增强的动态高效深度网络模型。这些工作取得了 CVPR NTIRE 质量增强的双赛道冠军。

李玺教授的报告题目是“视觉结构建模和特征学习”。互联网和物联网时代催生了海量视频大数据，从这些海量视频数据中有效提取知识迫切需要各种人工智能的技术和手段。因此，如何进行人工智能驱动的视觉计算已经成为当今知识经济时代亟待解决的核心技术问题。这个报告围绕数据驱动的人工智能学习方法，进行大规模图像/视频数据的视觉特征学习，从目标视觉感知特性、视觉特征表达、深度学习器构建机制、高层语义理解等多维度视角进行了深入剖析，并引入了大规模视觉特征学习所涉及的主要研究问题和技术方法。并且介绍了图像生成的过程中保留的结构信息以及人类视觉的选择性和整体性。系统地回顾了视觉特征表达和学习领域的不同发展阶段，介绍了近年来如何利用视觉特征学习进行视觉语义分析和理解所做的一系列代表性的研究工作及其实际应用，并将应用于车道场景，结构场景以及对目标检测。最后指出现在的挑战，数据和人类智能存在着鸿沟，数据空间更多的是基于观察，而人类智慧是基于逻辑、符号和知识。

杨文瀚老师的报告题目是“暗光影像增强计算”。在低光照场景下进行图像/视频拍摄会导致一系列的视觉降质问题，例如暗光、欠曝、低对比度以及强噪声等。

这些视觉降质既对人眼主观视觉体验造成干扰，又对计算机视觉应用构成影响。报告中介绍了系统地探究低光照增强方法在这两类情境中面临的挑战以及如何通过暗光照增强提升两类应用的可用性和鲁棒性，例如介绍了使用的将视网膜模型和深度学习结合的方法，并在过程中构建的低光照数据集和构造方法。并且还对低光照增强的其他方向进行了探索，介绍了多任务低光照人脸检测方法，并且在 NIQ 上取得优秀排名。

Panel 实录：

任文琦（主持人）：底层视觉和高层视觉连通性研究不够多，底层视觉是否一定会处境高层视觉，如何保证底层视觉和图像处理可以有效的促进高层视觉性能？

丛杨：可能根据不同的任务/需求都不太一样，通用的可能会比较难。任务又太多，计算量代价都需要权衡。想要计算机更好地理解视觉，现在的技术是解决一个点的问题，并不是解决通用问题，所以未来还需要对此深入思考。由于人的智能涉及到的领域很多，如想象，联想，情感等，是一个复杂的事情。

李玺：大家已经遇到一些瓶颈了，应该往认知方面去走？还是当作一个信号处理？大家都又不同的观点。偏研究，还是偏应用，这两个思路有时候是不能统一的。但是，研究人员需要有很多专业知识，这需要花费很长时间，并需要很多思考。对于感知和图像质量确实很重要，主要在于，我们做的图像和信号分析，随着传感器的革命，可能会改变后端的算法处理。尤其是传感器的性能精度更高，对算法的影响很大，很多时候传感器就能解决问题。让后端算法变简单了，后端就容易建模了。对于 Benchmark 的问题，大模型对应大参数，如果没有数据支撑，这个参数很难获得的。是否需要这么大的模型，还是做精一个小任务，或者是否需要通用人工智能？我发现，现有的通用人工智能在过拟合数据空间，在解决问题时，模型是在记忆这个数据空间，是一个统计推理，而不是知识推理了。未来怎么发展，算法是否重要？需要新的模式进行适配。未来是否还有底层和高层的区别？底层是传统的模式识别，如果建立了大一统，是否还需要底层，或者底层是传感器的一个属性，高层是否是真正的高层，还是我们的方法？

任文琦（主持人）：李老师对大模型还是持有一个观望态度，具体任务具体设计。

徐迈：对于大模型的必要性，这是一个值得思考的。底层视觉任务会增强高层的视觉任务。结合我的研究，分辨率比较低的设备，获取的医学图像是不够清晰难以识别的，直接进行识别精度会很低。如果先做超分辨，会提高识别的性能。同时，用高层的语义，也可以对底层的视觉有增益。二者是正相关的。底层需要发展，端到端的底层到高层也是一个重要的发展方向。

任文琦（主持人）：我的理解是能促进的，但是怎么处理还是需要进一步研究的。

付莹：在有些任务上有促进，在有些任务上不能。这要看数据怎么用。对于退化的模型，要看这些先验。并且还需要看对这个有多深的了解，如果很擅长，可能有很大的促进。如果不擅长，端到端的可能更有效。对于底层视觉是否存在统一。如果追求极致体验，很难统一，因为用一个统一的会掉点。对于底层视觉的 Benchmark，它有它的难点，也有优势，优势在于更容易合成。弱势在于与高层视觉结合，可能难以标记。这都是开放问题，还需要进一步讨论。

杨文瀚：我们之前做了一些低光照的，不同问题之间需要解耦合。不同任务的结论可能会不一样，解耦成不同的因素可能会更有效。从对抗攻击可以看出，做一点改动就可以改变高层视觉，而底层的很多工作并没增益高层。在统一框架中，多任务肯定会掉点，先验肯定是有用的，尤其是 CNN 这种简单的。但是模型变大之后，尤其 Transformer，加入先验之后增益变小了。对于 Benchmark，和工业界结合建立可能会更好一点。

任文琦（主持人）：对于视觉感知是否需要根据不同环境变化做出动态反应能力，那如何能够使感知学习像人一样来适应新的环境和任务？

李玺：这是启发式的，我认为还是要把属性定义清楚，是为了应用还是当成一个学科。理想状态是结合到一起的，但是实际上还是从应用出发，看应用需要什么

能力。和人一样，一些通识的还是需要学习的。做一些视觉的东西，反馈是非常短期的，因为我们现在的模型都是基于优化的，需要有反馈。

任文琦：通用的框架处理全部退化类型可能还比较难，具体任务还是需要具体对待。从退化的本质的角度来说，有什么不同？

徐迈：从视频编码角度，视频压缩的失真不止一个，纹理复杂的区域容易出现模糊，其原因是高频分量会丢失。视频编码是块分割，分割的方式不一样，也会产生块效应，也会增加了高频信息。所以这是由多个不同情况的造成的。不同的压缩编码会有差异，是多个层面的失真。更多时候，数据驱动是有好处的，并且数据容易生成。失真类型也在研究，结合做可解释的可能会更好。通用框架很难，因为自然界噪声类型太多了，一个通用模型很难做到，但是一个方向。目前还没有一个高效的方法，来处理这些问题。

任文琦（主持人）：水下深海也有很多退化的问题，能说明一下区别，难点和解决方案吗？

丛扬：水下深海退化问题和视频采集的噪声不太一样，有些是物理现象引起的，例如散射。可以通过算法解决。光学方面物理模型，还可以通过让器件采集的更好。另一方面用物理模型帮助恢复，再者可以利用数据和数学的方式来进一步提高成像质量。总体来说，应该从多个角度入手，算法和器件需要综合考虑。

任文琦（主持人）：数学模型，学习模型和传感器都需要考虑，很多方面值得探讨。成像模型的构建、图像先验的准确性、成像传感器的物理特性，是否可以在一个方法中都进行考虑？

付莹：可以都考虑的。如果成像问题涉及多个方面，是有必要一起来考虑的。虽然会增大模型，但可以做小模型，轻量化的，这相对大模型会更加显著。例如简单和复杂的用不同的层级，这可以同时考虑，并且对实验结果有利。如果用的比较合理的话，还可以同时增强图像的属性。

任文琦（主持人）：同时来建模，共同增强，是值得研究的。数据库是基于数据驱动的深度学习的一个关键，应该如何建立数据库？

杨文瀚：从应用需求出发，将实际问题抽象成学术问题。工业界数据很丰富，学术界可以和工业界联合开发。可以更好地利用真实环境的大数据。

任文琦（主持人）：对于全景视觉，数据方面是否有值得做的事情？

徐迈：数据库是很重要的。深度学习发展快的原因，就是数据在推动的。底层任务例如压缩，不需要人工标注，比较好建库。但是建库工作量大，只关注建库容易忽略算法能力，还会忽略可解释性，算法和可解释性也需要关注。全景视觉是新的问题，需要建库才能去了解这个感知行为，这个是非常有必要的。发展数据库的过程要有反思，模型和参数是否越大越好。这需要一部分人做小模型可解释性，另一部分人做大数据驱动。

任文琦（主持人）：如何精准对各类退化因素建模，尤其水下？

丛扬：这是很复杂的。水下不确定因素很多，精准建模还挺难。有一些先期工作，需要考虑到水的杂质等，我觉得精准建模，如用数学模型或神经网络很难。

任文琦（主持人）：VR 场景下的视听多模态评价有意义吗？

徐迈：有。这是新型的媒体方式，需要质量评价，判断用户的体验，才有优化的目标，这是非常关键的。但前提是行为感知模型是比较精准的，这个评价指标的构建才是比较有意义的。

责任编委 杨巨峰

2023 年度 CCF-CV 特色系列活动申请开始征集啦！

CCF CV

2023 年度中国计算机学会计算机视觉专委会(CCF-CV)特色系列活动已经开始征集，详情链接如下，欢迎从事计算机视觉领域研究与应用的人员积极申请！

CCF-CV
走进高校系列报告会

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学科发展，提升我国计算机视觉研究在国际上的影响力，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。活动采取请进来的方式邀请高水平学者以学术报告等形式宣传高水平研究，同时对承办单位青年研究人员等进行面对面指导，帮助青年学者快速成长。

自 2015 年 11 月开展以来，CCF-CV 走进高校系列报告会活动得到了高校和研究所、讲者、听众的大力支持。截至 2022 年 12 月底，CF-CV 走进高校系列报告会活动已成功举办了 121 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近 500 场，活动现场人均听众 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。同时在征得讲者同意的前提下，于专委 B 站账号分享部分讲者的报告视频，并于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，向对活动的顺利开展提供支持和帮助的承办单位、组织者以及分享精彩报告和观点的讲者嘉宾们和听众表示由衷的感谢！

活动内容包括如下：

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。

欢迎各高校和科研院所从事计算机视觉领域研究与应用的人员积极申请！

2023 年度 CCF-CV 特色系列活动申请开始征集啦！

从 2015 年底至今，CCF-CV 企业交流活动得到了企业、讲者、听众的大力支持。截至 2022 年底，CCF-CV 企业交流活动已成功举办了 24 期，相继走进了多家国内外知名企业。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！为了进一步加强企业与专委的深度交流，推动高质量的产学研合作，我们采取了不同类型活动方案，以下是具体活动细则，欢迎各位踊跃申请，也感谢大家长期以来的关注和支持！

活动分为以下三个类型，每次择一种开展：

专委走进企业：高科技企业举办交流会，邀请研发主题相关的专委委员介绍他们的研究成果，并与企业技术研究人员探讨产品研发过程中遇到的实际问题及解决思路，活动关键点是需要提前明确企业的需求和痛点。

企业走进专委：专委组织的走进高校等活动中，邀请企业研究人员介绍他们的研发成果和产品，以及需要解决的瓶颈和研发需求，可以给予高校自主权选择邀请哪个企业参加，然后专委帮助协调。

专题合作论坛：在 PRCV 或 RACV 等专委主办/承办的大会期间，设立互动专区或专题竞赛，专委委员及企业研究人员现场讲解各自的研究成果，探讨研究升级思路及合作机会。

CCF-CV 企业交流活动采取自愿申请的方式，面向计算机视觉高科技企业及计算机视觉领域研究者开放申请。如您有意申请并组织承办企业交流活动，请与计算机视觉专委会秘书处联系。联系人：

潘金山：jspan@njust.edu.cn

黄　　岩：yhuang@nlpr.ia.ac.cn

“视界无限”是中国计算机学会计算机视觉专委会（CCF-CV）的品牌学术活动。该活动旨在聚焦计算机视觉领域某一重要研究主题，组织相关领域的资深研究者与优秀青年学者进行全方位的深入研讨，总结该主题前沿进展与未来趋势，助推相关研究主题的青年学者快速成长。该活动每季度举办一次，全年四次。每次活动由 1-2 位计算机视觉专家学者负责召集和组织。

视界无限”活动自 2019 年 1 月开展以来，目前已经成功举办了 15 期，活动的主题涵盖了诸如生成式对抗网络、同时定位与地图构建、行人重识别、视觉目标跟踪、虚拟现实/增强现实、视觉行为理解、底层视觉等计算机视觉领域的热点研究主题，在领域内引起强烈反响。在此，向对活动的顺利开展提供支持和帮助的承办单位、组织者以及分享精彩报告和观点的讲者嘉宾们和听众表示由衷的感谢！

“视界无限”活动面向计算机视觉专委会委员及计算机视觉领域研究者开放申请，如果您有意申请并组织承办“视界无限”活动，请通过链接 https://wj.qq.com/s2/11720924/96d8 或者扫描二维码填写申请简表，并与计算机视觉专委会秘书处联系。

责任编委 黄岩

专题综述

生物特征模板保护研究与展望

安徽大学 张慧 王华彬 金哲 李学俊

一、引言

随着互联网和硬件设备快速发展，生物特征识别技术的识别精度和速度已满足实际应用需求，被广泛应用于交通监管、门禁管理和移动应用等领域。相比于传统密码、智能卡等方式，生物特征作为身份凭证，具有唯一性、稳定性、便捷性、安全性等优势，但与此同时，生物特征被视为个人敏感信息，一旦泄露或被非法使用，将给用户带来不可挽回的后果。生物特征模板保护（Biometric Template Protection, BTP）作为可信生物特征识别（Trustworthy Biometrics）的关键技术是目前的研究热点。

1.1 生物特征识别

生物特征识别是基于个人的生理(例如人脸、虹膜、指纹和掌纹等)或者行为特征 (例如步态、语音和笔迹等)进行用户身份的识别。通用的生物特征识别框架主要包括 5 部分，如图 1 所示。传感器（Sensor）用于读取用户的生物特征信号，例如相机、指纹采集器等，通过传感器采集的信号质量影响着最终的识别性能。生物信号经过特征提取器(Feature Extractor)转换成生物特征并存储在数据库(Database)中。匹配器(Matcher)用于对认证时产生的查询生物特征与数据库中注册得到的生物特征模板进行比对，比对结果输入决策模块(Decision Module)，从而得到决策结果。

由于深度学习、硬件加速设备等技术的快速发展，目前生物特征识别的精度和速度已达到了很高的水平，并应用于很多刚需场景中。因此，关注生物特征识别的安全和隐私性问题，提高生物特征安全，保障用户隐私，始终是推动生物特征识别技术持续发展的关键之处。

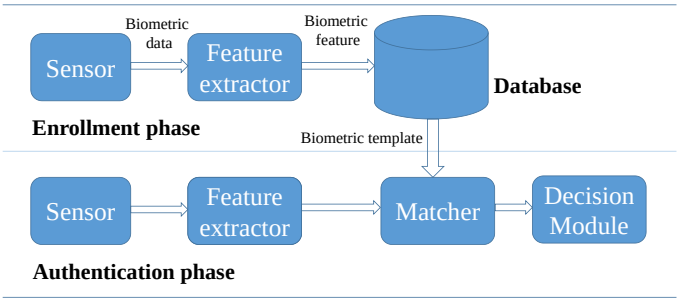


图 1 通用的生物特征识别框架示意图

1.2 生物特征模板保护

生物特征模板保护技术通过不可逆变换或者加密原始生物特征以生成安全的生物特征模板，这有利于避免直接存储原始生物特征数据所带来的安全和隐私泄露风险。在注册阶段，用户的生物特征以安全模板的形式存储在数据库中；在认证阶段，用户的实时生物特征经过同样的模板保护过程生成特征模板，并与数据库中存储的模板进行匹配操作，根据系统预先设置的阈值和二者的相似性程度返回用户身份的认证结果。

生物特征模板保护主要分为可撤销生物特征(Cancelable Biometrics，CB)和生物特征加密(Biometric Cryptosystems，BC)两个方向。CB 方案中，原始生物特征经过不可逆变换得到可撤销的模板，并在变换域内完成模板的匹配操作，但基于可撤销模板不可能或很难还原出原始生物特征，从而保护生物特征安全。BC 方案则结合密码学原语从生物特征生成或绑定密钥，

通过验证恢复密钥的准确性实现用户合法身份的认证。

1.3 亟待解决的关键问题

虽然目前已经有很多模板保护方案提出，但生物特征模板保护技术的发展仍受到生物特征的内在特性和实际应用场景的限制，生物特征的安全性和识别系统的实用性无法保证。主要概括为以下三个关键问题：

(1) 噪声干扰致使不可完全再现：受光照、角度和环境等因素的影响，用户的同一生物模态因大量随机噪声的存在致使每次提取的特征都存在差异，不能完全再现。所以用户在认证时输入的生物特征和数据库中注册存储的特征模板不完全相同，二者进行匹配操作时得到的相似度可能低于阈值，从而降低生物特征识别的性能。而基于生物特征继而进行的模板保护操作，往往以性能降低为代价提高生物特征的安全性，这会进一步增大系统性能降低的程度。因此克服生物特征的噪声干扰，平衡识别性能和生物特征安全，设计具有容错性的安全生物特征识别系统是推进该领域研究的一个重要挑战。

(2) 外部因子泄露威胁系统安全：为赋予生物特征模板如密码一样灵活重置的能力，降低生物特征泄露致使用户永久不可复用该生物特征的风险，一般采用生物特征和外部因子(如令牌)结合的方式生成不可逆的、可撤销的生物特征模板。但这种可撤销方案使得生物特征的安全与外部因子的安全息息相关，当系统受到令牌丢失攻击(Lost token attack)时，攻击者可能从丢失的令牌中分析出原始生物特征或者伪装成合法用户，威胁系统安全，降低识别性能。因此，如何设计一种更具鲁棒性的可撤销生物特征模板方案至关重要。

(3) 攻击与对抗技术的此消彼长：针对生物特征识别系统的攻击无处不在，攻击方法层出不穷，例如字典攻击(Dictionary Attacks)、关联攻击(Correlation Attacks)、爬山攻击(Hill Climbing Attacks)和重建攻击(Reconstruction attacks)等。即使始终有对抗攻击的模板保护方案提出，但面对复杂的应用环境和无法预知的攻击威胁，生物特征的保护方案在攻击者面前无能为力。攻击与对抗技术此消彼长的状态致使生物特征模板保护仍处于不断发展的阶段。

二、可撤销生物特征方案研究现状

根据所采用的变换方式是否可逆，BC 方案分为两类：不可逆变换和生物特征盐析。基于不可逆变换的 BC 方案通过变换原始生物特征模板，使转换模板无法反转。经典算法有 IoM 哈希[1]、布隆过滤法(Bloom Filter)以及基于布隆的改进算法。生物特征盐析则通过设置一些人工模式(例如随机噪声)与生物特征模板进行混合操作，从而保护原始生物特征。经典算法有生物哈希(Biohashing)以及基于生物哈希的扩展。这些方案在一定程度上提高了生物特征安全，但大多数可撤销技术都以严重的性能下降为代价提高安全性，无法保持与原始生物特征相比的合理精度。基于此，下面介绍近期的代表性工作，以解决目前 CB 方案中普遍存在的难点问题。

2.1 LIoM 哈希算法

针对系统对高性能和匹配速度的要求，Dong 等人[2]提出了用于大规模开集人脸识别的基于 IoM 哈希的模板保护方案：LIoM（Index-of-max hashing by learning）哈希。LIoM 哈希是基于 IoM 哈希设计的紧凑人脸特征表示算法，通过汉明距离实现高效的匹配操作，IoM 哈希的不可逆变换过程确保用户的隐私性保护。为验证算法的有效性，结合多种融合策略在大规模人脸数据集上进行了全面评估。

2.2 免对齐的可撤销虹膜特征识别方案

特征对齐是存在于 CB 方案中的一个重要问题。由于很多生物特征(如虹膜)无法精准对齐，为保持合理精度匹配过程需要进行移位打分操作。但 CB 方案的使用转变了原始特征空间，该匹配策略不再适用。基于此，Lee 等人[3]基于方向梯度直方图提出了一种随机增强梯度直方图算法（Random Augmented Histogram of Gradients，R·HoG）用于虹膜特征模板保护，将未对齐的 irisCode 转换为对齐健壮的可撤销模板，提高了模板匹配的效率和生物特征系统的安全性。

2.3 无令牌的可撤销生物特征识别方案

由于 CB 方案通常被设计用来保护具有两个输入因子的生物特征模板，即：生物特征识别和用于模板替换

的令牌，一旦令牌丢失，受保护模板极易遭受安全攻击和隐私侵犯。Lee 等人[4]提出了一种单因子可撤销生物特征识别方案，即扩展特征向量(Extended Feature Vector，EFV)哈希算法，该算法只需要一个生物特征作为输入，并利用与生物特征数据分离的置换密钥作为匹配的标识符，从而实现安全的单因子认证。

2.4 单因子可撤销生物特征认证方案

双因子可撤销的生物特征认证方法引入额外因子即令牌化因子带来了隐私和安全威胁问题。针对这一问题，孔小景等人[5]提出了一种唯一二值数据生物特征作为输入因子的单因子可撤销生物识别方法，即 WSE 哈希算法。WSE 哈希算法满足不可逆性，可撤销性，不可链接性以及精确性这 4 个可撤销的生物特征模板保护标准，也抵御了 3 种方式的安全性攻击测试。同时 WSE 哈希算法也可以扩展到二值向量形式表示的虹膜、面部特征、掌纹和静脉等生物特征识别。另外，算法安全性，如碰撞攻击、差分攻击等攻击方式，是未来研究方向。

三、生物特征加密方案研究现状

BC 方案的设计思想是出于对密钥的保护，同时也保证了生物特征的安全。BC 方案主要分为两类：密钥绑定和密钥生成。密钥绑定是采用生物特征与密钥进行绑定，产生公开的辅助数据，认证时结合生物特征与辅助数据以释放密钥。目前具有代表性的方案有模糊承诺(Fuzzy Commitment)，模糊保险箱(Fuzzy Vault)等。密钥生成则是基于生物特征生成或提取出密钥，认证时比对密钥以认证用户身份。代表方案有模糊提取器(Fuzzy Extractor)、安全骨架(Secure Sketch)等。但所有的 BC 方案都存在一个问题，就是提取的生物特征质量影响着恢复密钥的正确率。

3.1 基于密钥绑定的指纹细节点保护方案

基于生物特征类内差异大的特性，目前 CB 方案都依赖纠错码来提高系统的容错能力。但纠错码的纠错能力有限，很难平衡安全性和性能要求。因此，Jin 等人[6]提出了一种非纠错码密钥绑定方案以及基于指纹细节点的可撤销的不可逆变换。该方案不局限于二元特征和匹配器，可应用于多种生物特征表示。具体的密钥绑定和密钥恢复过程如图 2 所示。

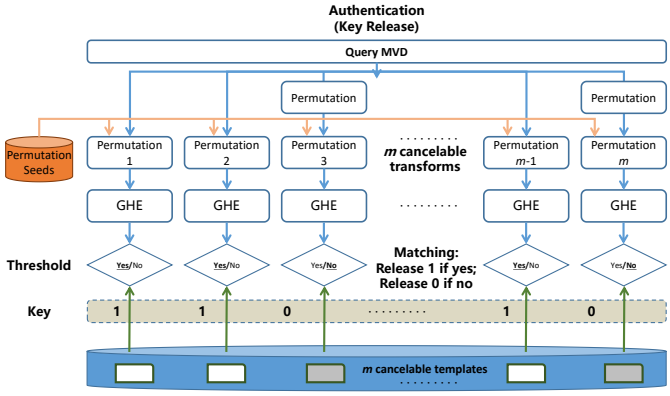


图 2 密钥绑定和密钥恢复示意图

3.2 基于对称密钥环加密的 BC 方案

由于生物特征提取时存在各种噪声干扰，生物特征识别系统出现模糊性而识别性能下降或不工作。为提高系统的容错性和安全性，Lai 等人[7]把密钥绑定视为对称的加解密问题，提出了基于对称密钥环加密(Symmetric Keyring Encryption，SKE)的 CB 方案。SKE 方案由 RV 密钥对、过滤机制和 Shamir 的秘密共享方案组成，理论分析和实验结果显示该方案可扩展到其他生物特征，并可以抵抗多种安全攻击。具体的方案概述如图 3 所示。

3.3 基于深度哈希网络的多光谱掌纹模糊承诺方案

在生物识别密码系统中，所生成的生物密钥通常结合单向函数进行严格保护，但这很难平衡模板的大小和准确性。Wu 等人[8]采用深度哈希网络，生成存储量小、匹配复杂度低的二值模板。根据鉴别能量，对模板中的比特优选后，减少了模板数据量和构造模糊承诺的计算复杂度。

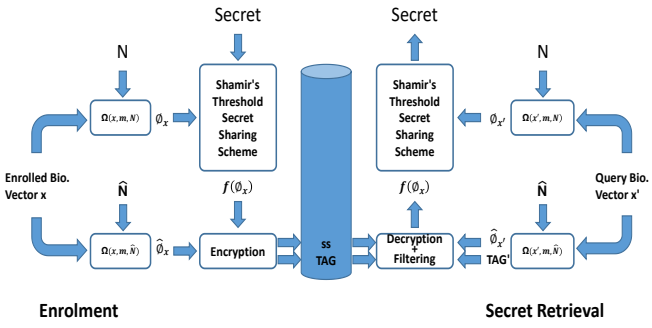


图 3 SKE 方案概述图

四、生物特征模板的攻击与对抗

随着黑客技术的不断进步，针对生物特征模板的攻击呈现出层出不穷的趋势。为增强生物特征识别的安全性和用户隐私保护，研究攻击算法并设计对抗方案尤为重要。

4.1 基于 GAN 生成器的深度人脸特征重构

基于深度卷积神经网络(CNN)的人脸识别目前已经达到了很高的识别精度，但从深度学习模型中提取的特征(深度特征)的安全性和私密性问题常常被忽视。基于此，Dong 等人[9]提出了在不访问 CNN 网络配置的情况下基于 GAN 生成器从深层特征重构人脸图像的方案，具体的框架概述如图 4 所示。该方案使用 GAN 生成器同时发挥优化目标的人脸分布约束和人脸生成器的作用，实现了高相似度和高视觉质量的人脸图像重构。该工作中所伪造的人脸图像揭示了当前人脸识别系统的安全和隐私风险，对手可能伪造人脸图像非法访问人脸识别系统。因此，需要结合模板保护方案和反欺骗检测手段保护生物特征，规避隐私数据泄露风险。

4.2 增强的掌纹重构攻击

目前很多重建攻击重构的原像普遍存在自然性、完整性和视觉质量差等问题。基于此，Sun 等人[10]采用了两种策略，一是邻域范围修改约束，减少图像质量的恶化；二是挑选重要的像素，打包修改，既增强了修改对优化的影响，同时降低了引入过多的图像不自然痕迹，从而得到两种增强的重构攻击用于掌纹识别。

4.3 基于风格迁移技术的掌纹重构攻击

攻击者基于跨数据库攻击重建的图像可以有效的攻击其他生物特征系统。为实现在线跨数据库攻击，并保证重构图像的高质量，Yang 等人[11]提出了两种新颖的风格转移技术，从特征模板重构恢复原始图像，并用于攻击基于编码的掌纹识别系统。提出的重构方法揭示了基于纹理编码的掌纹识别方法的脆弱性，因此，为了提高生物特征识别系统安全和用户隐私，有效的攻击防御技术和重构检测方法成为不可或缺的一项研究。

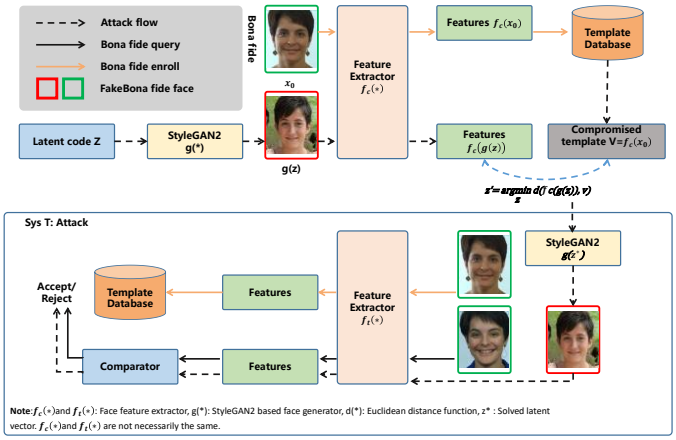


图 4 人脸图像重构框架图

五、总结与展望

本文主要论述了生物特征模板保护技术在可撤销生物特征和生物特征加密两个方向的研究进展，并针对目前该领域所存在的关键问题，重点介绍了具有代表性的研究工作。随着生物特征识别技术的不断普及，生物特征模板保护作为一项必要且有意义的工作必将持续发展。基于此，本文对未来的发展进行展望：

（1）无令牌的可撤销生物特征模板保护方案设计

目前无令牌 CB 方案比较欠缺，可以借鉴各学科领域研究方法设计安全无令牌 CB 方案，打破目前生物特征模板保护的发展瓶颈，提高生物特征认证系统安全性。

（2）多模态生物特征融合策略设计

多模态生物特征识别系统已被证明具有更高的性能和安全性优势，但特征级融合作为系统中的关键环节仍面临挑战，需要设计兼容不同生物特征信号的融合策略，提高系统对不同类型和维度生物特征的使用能力和常见攻击的抵抗能力。

（3）重构攻击和对抗方案设计

为了维持性能和安全的平衡，可撤销生物特征模板保护方案具有相对距离保持属性，但这容易招致基于相对距离保持的攻击，攻击者试图基于模板重构原像。因此，需要研究基于相对距离保持的攻击框架，设计对抗策略，从而在复杂多变的应用环境保证生物特征识别系统的安全运行。

责任编委 储珺

参考文献

[1] Jin, Z., Hwang, J. Y., Lai, Y. L., Kim, S., & Teoh, A. B. J. (2017). Ranking-based locality sensitive hashing-enabled cancelable biometrics: Index-of-max hashing. IEEE Transactions on Information Forensics and Security, 13(2), 393-407.

[2] Dong, X., Kim, S., Jin, Z., Hwang, J. Y., Cho, S., & Teoh, A. B. J. (2020). Open-set face identification with index-of-max hashing by learning. Pattern Recognition, 103, 107277.

[3] Lee, M. J., Jin, Z., Liang, S. N., &Tistarelli, M. (2022). Alignment-Robust Cancelable Biometric Scheme for Iris Verification. IEEE Transactions on Information Forensics and Security, 17, 3449-3464.

[4] Lee, M. J., Jin, Z., & Teoh, A. B. J. (2018, December). One-factor cancellable scheme for fingerprint template protection: Extended Feature Vector (EFV) Hashing. In 2018 IEEE international workshop on information forensics and security (WIFS) (pp. 1-7). IEEE.

[5] 孔小景，李学俊，金哲，周芃，陈江勇. 一种单因子的可撤销生物特征认证方法.自动化学报, 2021, 47(5): 1159-1170.

[6] Jin, Z., Teoh, A. B. J., Goi, B. M., & Tay, Y. H. (2016). Biometric cryptosystems: a new biometric key binding and its implementation for fingerprint minutiae-based representation. Pattern Recognition, 56, 50-62.

[7] Lai, Y. L., Hwang, J. Y., Jin, Z., Kim, S., Cho, S., & Teoh, A. B. J. (2019). Symmetric keyring encryption scheme for biometric cryptosystem. Information sciences, 502, 492-509.

[8] Wu, T., Leng, L., & Khan, M. K. (2022). A multi-spectral palmprint fuzzy commitment based on deep hashing code with discriminative bit selection. Artificial Intelligence Review, 1-18.

[9] Dong, X., Miao, Z., Ma, L., Shen, J., Jin, Z., Guo, Z., & Teoh, A. B. J. (2022). Reconstruct Face from Features Using GAN Generator as a Distribution Constraint. arXiv preprint arXiv:2206.04295.

[10] Sun, Y., Leng, L., Jin, Z., & Kim, B. G. (2022). Reinforced Palmprint Reconstruction Attacks in Biometric Systems. Sensors, 22(2), 591.

[11] Yang, Z., Leng, L., Zhang, B., Li, M., & Chu, J. (2022). Two novel style-transfer palmprint reconstruction attacks. Applied Intelligence, 1-18.

王华彬

安徽大学计算机科学与技术学院副教授，研究方向：模型识别与信息处理，医疗图像处理，虚拟现实。

Email: wanghuabin@ahu.edu.cn

金哲

安徽大学人工智能学院教授，研究方向：可信人工智能，模式识别与安全，深度学习。

Email: jinzhe@ahu.edu.cn

李学俊

安徽大学计算机科学与技术学院教授，研究方向：边缘计算与智能软件，医学人工智能，工业互联网。

Email: xjli@ahu.edu.cn

热点追踪

基于自监督学习的单目深度估计方法

中科院自动化研究所 周正铭 董秋雷

一、摘要

单目深度估计旨在从单幅输入图像中估计场景的深度，其对于三维重建、场景理解等任务有着重要的意义。由于在真实场景中获取稠密而准确的深度真值是很困难的，基于自监督学习的单目深度估计受到了广泛的关注。近年来，尽管自监督单目深度估计方法取得了较好的表现，但如何进一步缩小自监督与有监督方法之间的差距并提升其精度仍然是一个开放性的问题。针对这一问题，本文一方面从训练约束的角度，考虑如何更有效地同时利用两种现有的自监督深度约束训练模型；另一方面从网络结构的角度，考虑如何使网络学到对于深度估计更有效的特征。具体地，我们提出了一种感知遮挡的由粗到细的自监督单目深度估计方法，称为 OCFD-Net；提出了一种自蒸馏特征聚合模块，并在此基础上提出自蒸馏聚合网络，称为 SDFA-Net。相关成果分别被 ACM MM 2022 和 ECCV 2022 录取。

二、引言

基于自监督学习的单目深度估计旨在使用没有深度真值的样本训练一个深度神经网络，使其可以从单幅输入图像中预测稠密的深度图。现有方法一般采用不同视角拍摄的同一场景的多幅图像作为训练数据，将深度估计任务转化为图像重建任务，并使用重建损失训练网络(如图 1 所示)。根据训练数据的来源，现有自监督单目深度估计方法可以分为采用视频序列训练的方法和采用双目图像训练的方法。其中，采用视频序列训练的方法[1,2]在训练阶段以视频序列中的连续帧作为训练样本。由于连续帧之间的相机运动是未知的，这些方法在训练阶段除了估计深度图之外，还需要估计图像之间的相机运动情况。采用双目图像训练的方法[3,4]在训练阶段以双目相机拍摄的图像对作为训练样本。由于拍摄双目

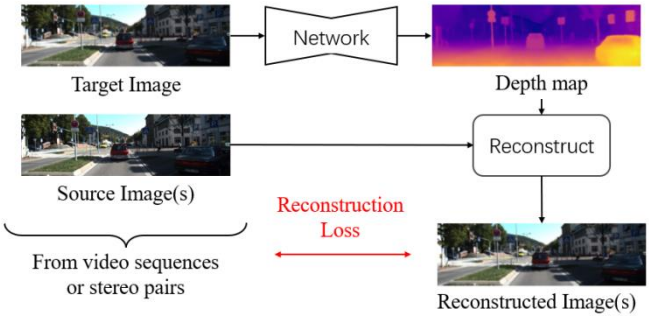


图 1 基于自监督学习的单目深度估计基本原理

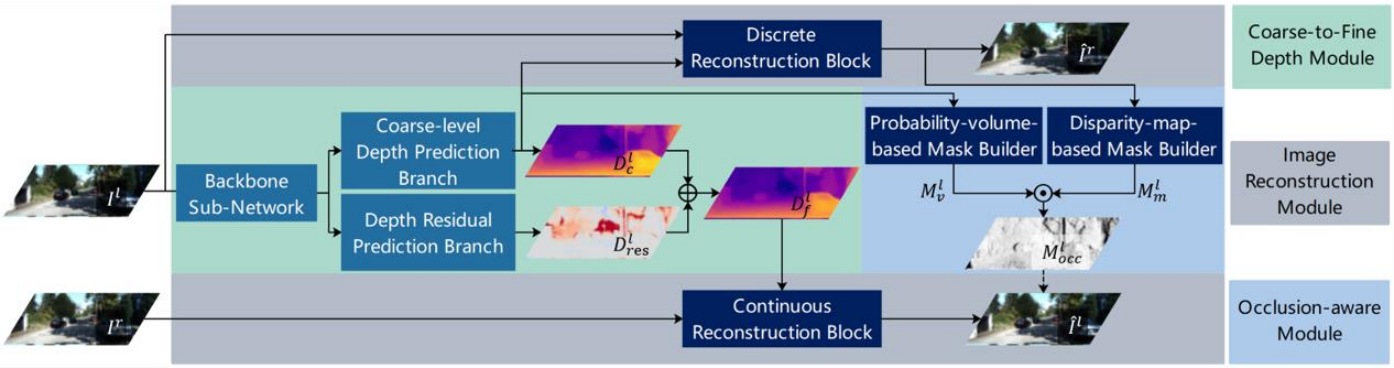


图 2 感知遮挡的由粗到细自监督单目深度估计网络 OCFD-Net 的结构示意图

图像的相机相对位置是固定的，且双目系统中的视差和深度具有确定的转换关系，这些方法只需要预测图像对应的深度图或视差图。

尽管近年来基于自监督学习的方法得到了广泛的研究且取得了较好的表现，但如何进一步提升自监督单目深度估计方法的精度仍然是一个开放性的问题。针对这一问题，一方面，本文提出了一种感知遮挡的由粗到细自监督单目深度估计网络，称为 OCFD-Net[5]。该模型通过分别估计粗粒度深度和场景深度残差的方法，结合了连续[2]和离散[4]两种深度约束的优势，并通过一个遮挡感知模块缓解训练中的遮挡问题对深度结果造成的负面影响。另一方面，本文提出了一种基于自蒸馏的特征聚合模块，并基于此模块设计了一种新的单目深度估计网络，称为 SDFA-Net[6]。在自蒸馏特征聚合模块中，采用三个分支分别预测三个特征偏移图，用于在自蒸馏条件下对待融合的多尺度特征进行细化。实验结果表明，我们提出的 OCFD-Net 和 SDFA-Net 在室外驾驶场景数据集上的表现超越了绝大多数现有的方法。

三、正文

为了有效地在自监督单目深度估计中利用连续和离散两种深度约束，我们首先通过对比实验分析了两种深度约束各自的优点和不足。分析结果表明：离散深度约束有助于保留更多深度细节信息，且可以使模型取得较高的精度，但使用离散深度约束训练的模型难以在平坦区域生成平滑的深度估计结果；连续深度约束有助于保持深度结果的平滑性，但其使得估计结果的精度相对较低。基于上述分析，我们提出了感知遮挡的由粗到细自监督单目深度估计网络 OCFD-Net(如图 2 所示)。该

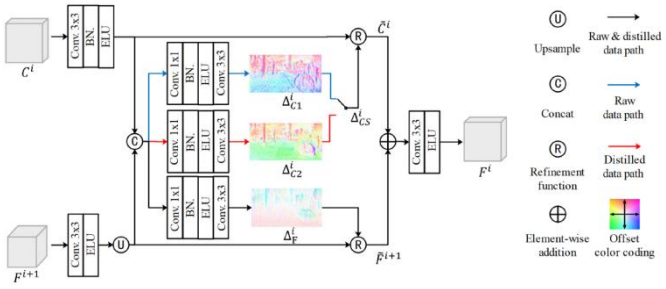


图 3 自蒸馏特征聚合模块

网络分为三个模块，其中由粗到细的深度估计模块用于从输入图像估计深度图。该模块通过一个骨干子网络从输入图像中提取特征。以该特征为输入，分别用粗粒度深度预测分支预测一个离散表示的粗粒度深度图，用深度残差预测分支预测一个连续表示的场景深度残差图。最终将两部分相加得到细粒度的深度图。图像重建模块通过预测的深度结果进行图像重建，从而对网络进行自监督训练。具体地，通过离散形重建的方式，引入离散深度约束对粗粒度深度进行训练；通过连续形重建的方式，引入连续深度约束，对深度残差进行训练。遮挡感知模块用于通过估计的深度结果计算图像中潜在的遮挡区域，并输出表示遮挡概率的 Mask。具体的，该模块分别基于深度概率体和视差图生成两个遮挡 Mask，并将两个遮挡 Mask 逐像素相乘得到遮挡概率 Mask。在训练时，我们采用重建损失和平滑正则同时训练粗粒度和细粒度的深度结果。其中，在对细粒度深度结果进行训练时，我们基于遮挡概率 Mask 减小遮挡区域的重建损失，并加大相应区域的平滑正则，来缓解遮挡问题对深度估计结果造成的负面影响。

为了使得网络能在自监督方式下学到对于深度估计更有效和准确的特征，我们提出了自蒸馏特征聚合模

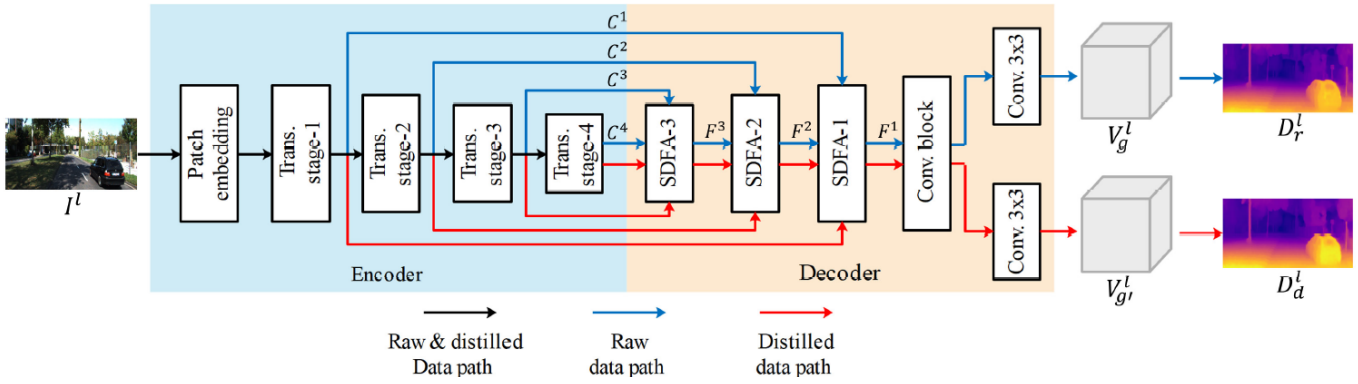


图 4 自蒸馏特征聚合网络 SDFA-Net 的结构示意图

块(如图 3 所示)用于融合两个不同尺度的特征，并保持其上下文一致性。该模块受到语义分割任务中特征对齐模块[7]的启发，使用可学习的偏移向量来融合不同尺度的特征。考虑到在自监督学习的过程中，图像重建损失存在一定的歧义性，可能使得模块无法通过自监督方式学习到准确的特征偏移图，进而造成深度结果的误差，自蒸馏特征聚合模块采用了三个分支来学习三个不同的特征偏移图。其中一个特征偏移图被用于细化小尺度的特征；其余两个特征偏移图被共同用于细化大尺度的特征，并分别通过图像重建损失和自蒸馏损失进行训练。进一步地，我们以改进的 Swin-transformer 作为编码器，以自蒸馏特征聚合模块作为解码器设计了用于单目深度估计的自蒸馏特征聚合网络 SDFA-Net(如图 4 所示)。SDFA-Net 的解码器中存在两条前向传播的数据通路，分别称为原始数据通路和蒸馏数据通路。在不同的数据通路中，会使用自蒸馏特征聚合模块中相应的偏移图预测分支来学习特征偏移图。为了有效地训练所提出的模型，我们将训练中的每次迭代分为三个步骤：第一步使用编码器提取多尺度特征，并用解码器中的原始数据通路预测深度图，采用图像重建损失和平滑正则作为损失函数；第二步使用网络中的蒸馏数据通路从编码器得到的多尺度特征中估计深度，并从第一步估计的深度结果中选择置信度较高的部分作为伪标签，计算蒸馏损失。第三步对损失进行反向传播来训练网络。

四、实验结果

表 1 展示了我们提出的 OCFD-Net 和 SDFA-Net 在 KITTI 室外驾驶场景数据集中的深度估计结果。可以看出在绝大多数指标下本文所提出的两个方法都取得了最好的表现。图 5 进一步展示了两个方法深度估计的可视化结果。从图中可以看出两个方法都能较好地保留场景中的深度细节信息，例如在细小的物体处以及物体的边缘处等。

为了验证本文所提出方法的有效性，图 6 中展示了 OCFD-Net 预测的粗粒度深度图(第二行)，深度残差图(第三行)和细粒度深度图(第四行)的可视化结果。对于深度残差图，红色表示残差为正，蓝色表示为负。可以看到深度残差使得细粒度深度在物体边缘处更加准确，在图像中的平坦区域结果更加平滑。图 7 展示了 SDFA-

表 1 OCFD-Net，SDFA-Net 和其他算法在 KITTI 数据集上的深度估计结果

Method	PP.	Abs. Rel. ↓	Sq. Rel. ↓	RMSE ↓	logRMSE ↓	A1 ↑	A2 ↑	A3 ↑
Monodepth2	✓	0.107	0.849	4.764	0.201	0.874	0.953	0.977
MonoResMatch	✓	0.111	0.867	4.714	0.199	0.864	0.954	0.979
DepthHints	✓	0.096	0.710	4.393	0.185	0.890	0.962	0.981
DBoosterNet-e		0.095	0.636	4.105	0.178	0.890	0.963	0.984
SingleNet	✓	0.094	0.681	4.392	0.185	0.892	0.962	0.981
FAL-Net	✓	0.093	0.564	3.973	0.174	0.898	0.967	**0.985**
EPCDepth	✓	0.091	0.646	4.207	0.176	0.901	0.966	0.983
EdgeOfDepth	✓	0.091	0.646	4.244	0.177	0.898	0.966	0.983
PLADE-Net	✓	0.089	0.590	4.008	0.172	0.900	0.967	**0.985**
OCFD-Net		0.091	0.576	4.036	0.174	0.901	0.967	0.984
OCFD-Net	✓	0.090	0.563	4.005	0.172	0.901	0.967	0.984
SDFA-Net		0.090	0.538	3.896	0.169	0.906	**0.969**	**0.985**
SDFA-Net	✓	**0.089**	**0.531**	**3.864**	**0.168**	**0.907**	**0.969**	**0.985**

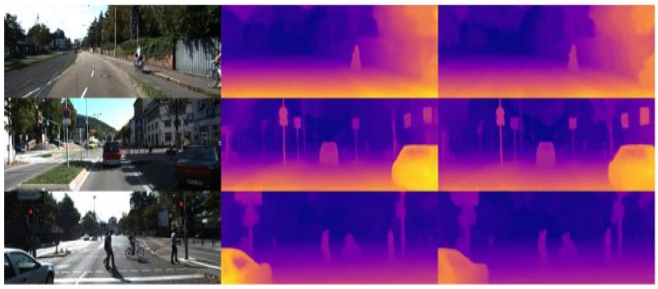


图 5 OCFD-Net 和 SDFA-Net 在 KITTI 数据集上的深度估计可视化结果

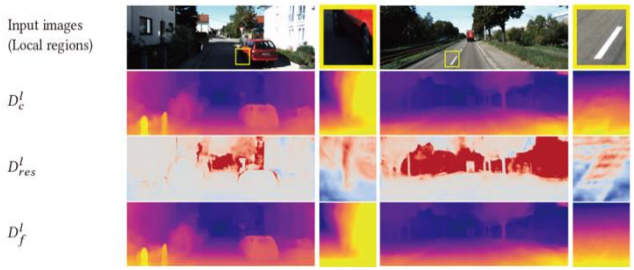


图 6 OCFD-Net 预测的粗粒度深度，深度残差和细粒度深度的可视化结果

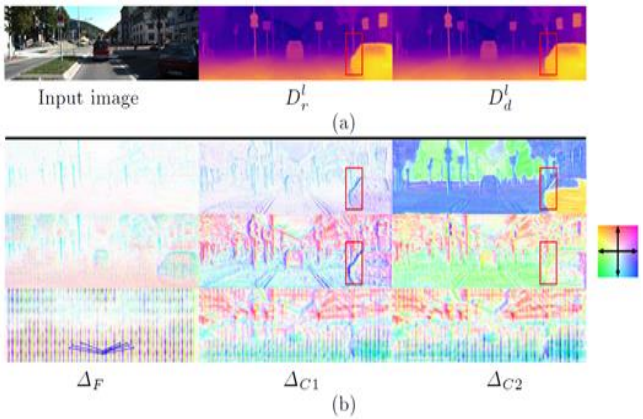


图 7 (a)SDFA-Net 通过不同数据通路预测的深度可视化结果 (b)自蒸馏特征聚合模块中特征偏移图可视化结果

Net 中使用不同数据通路预测的深度结果，可以看到使用蒸馏通路预测的深度图((a)中第三列)更加准确，尤其是在物体的边缘处。此外，我们可视化了 SDFA-Net 中多个自蒸馏特征聚合模块中学习到的特征偏移图，其中第一列的偏移图用于细化小尺度特征，第二、三列的偏移图用于分别在原始和蒸馏通路中细化大尺度特征。可以看到相较于原始通路中的特征偏移图，蒸馏通路中的的特征偏移图在物体边缘处更准确。

责任编委 崔海楠

参考文献

[1] Zhou, Tinghui, Matthew Brown, Noah Snavely, and David G. Lowe. "Unsupervised learning of depth and ego-motion from video." In Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 1851-1858. 2017.

[2] Godard, Clément, Oisin Mac Aodha, Michael Firman, and Gabriel J. Brostow. "Digging into self-supervised monocular depth estimation." In Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 3828-3838. 2019.

[3] Garg, Ravi, Vijay Kumar Bg, Gustavo Carneiro, and Ian Reid. "Unsupervised cnn for single view depth estimation: Geometry to the rescue." In European conference on computer vision, pp. 740-756. Springer, Cham, 2016.

[4] GonzalezBello, Juan Luis, and Munchurl Kim. "Forget about the lidar: Self-supervised depth estimators with med probability volumes." Advances in Neural Information Processing Systems 33 (2020): 12626-12637.

[5] Zhou, Zhengming, and Qiulei, Dong. "Learning Occlusion-Aware Coarse-to-Fine Depth Map for Self-Supervised Monocular Depth Estimation" In Proceedings of the 30th ACM International Conference on Multimedia, pp. 6386–6395. 2022.

[6] Zhou, Zhengming, and Qiulei Dong. "Self-distilled feature aggregation for self-supervised monocular depth estimation." In European Conference on Computer Vision, pp. 709-726. Springer, Cham, 2022.

[7] Huang, Zilong, Yunchao Wei, Xinggang Wang, Wenyu Liu, Thomas S. Huang, and Humphrey Shi. "Alignseg: Feature-aligned segmentation networks." IEEE Transactions on Pattern Analysis and Machine Intelligence 44, no. 1 (2021): 550-557.

周正铭

中科院自动化研究所硕士研究生。主要研究方向为深度估计、三维计算机视觉等。
Email: zhouzhengming2020@ia.ac.cn

董秋雷

中科院自动化研究所研究员。主要研究方向为三维计算机视觉、模式识别等。
Email: qldong@nlpr.ia.ac.cn

热点追踪

基于图注意力双线性池化的鲁棒性 RGB-T 跟踪

南京邮电大学 江晨风 康彬 周全

一、研究背景

随着多媒体技术的蓬勃发展，热红外摄像机已经成为一种经济实惠的摄像机。该摄像机可以捕捉温度高于绝对零度的目标发射的热红外辐射，适用于夜间监视。将 RGB 与热红外摄像机联合使用优点如下：(1)热红外摄像机具有很强的抗照度变化能力，可以在光照条件较差的情况下为 RGB 摄像机提供强有力的支持；(2)RGB 摄像机将有助于解决基于热红外摄像机的监控所面临的热交叉难题。因此，结合 RGB 和热特征的 RGB- T 跟踪可以有效应对恶劣天气挑战[1]。在 RGB-T 跟踪中，RGB 和热视频序列是成对获得的。其关键思想是利用 RGB 和热信息的互补性进行高效的多模型融合。

近些年来，研究者们开发了许多先进的方法进行多模型融合，例如基于粒子融合的 RGB- T 跟踪器[2, 3]，建立多图融合模型[4, 5]，求解统一优化问题 [6, 7]，用于 RGB- T 跟踪的密集卷积神经网络[8]，多适配器卷积网络[9]等等，其中后两种方法采用了深度卷积神经网络技术。与手工特征相比，深度卷积神经网络可以更好的提取深度语义信息，对目标进行鲁棒表示。因此，近年来深度学习技术在 RGB-T 跟踪方面表现出了巨大的潜力。然而，现有的基于 CNN 的 RGB-T 跟踪器通常将多层卷积特征图视为层次上的整体特征，忽略了 RGB 与热目标之间的部分特征相互作用。这可能会明显降低具有挑战性的视频对的跟踪精度。更严重的是，RGB 与热目标的少量有用信息可能在空间域上部分匹配甚至不匹配。在这种情况下，简单地将多个深层特征作为整体特征进行多模型融合，可能会产生不可避免的负面影响。

针对上述问题，本文提出了一种简单有效的面向四流的 Siamese 网络(FS-Siamese)用于 RGB-T 跟踪，其中四流的特征嵌入可分为范例嵌入对和候选嵌入对。通过基于图注意力的双线性池化模块，可以分别融合两个嵌入对，生成增强样本和增强候选，用于生成后续的相似图。对于双线性池化，其在异构部分信息融合方面表现出优于传统线性融合策略的性能。尽管双线性池化获得了一定的性能提升，但它无法区分深度特征图中元素的重要性。鉴于这些观察结果，我们在双线性池化中引入了共同注意力机制，将多模型池化描述为一个多图学习问题。由于目标外观可能发生剧烈变化，因此有必要在基于图注意力的双线性池化模块中引入一种有效的更新策略。目前最先进的更新策略[10, 11]只关注于探索当前目标特征与先前目标特征之间的时间相关性，而忽略了一个事实，即在线探索目标与其周围背景环境之间的空间相关性对于定位相似度最高的候选对来说是非常重要的。因此，我们设计了一种基于元学习的更新策略，以有效地更新基于图注意力的双线性池化模块的全连接层。这为利用类别信息在线更新样例语义表示提供了途径。本文的主要贡献如下:

(1)将基于注意力的双线性池化问题描述为一个多图学习问题。我们将图注意力网络和外积整合为一个统一的结构，使多个图在联合学习的同时实现有效局部信息交互。这样可以有效地消除目标对融合过程中的干扰。

(2)传统的面向多流跟踪网络只融合不同流的目标回归结果，在融合目标嵌入时没有探究成对关系。为了克服这一限制，我们提出了一种基于图注意的双线性池化的四流导向网络结构，用于有效融合多源嵌入对。

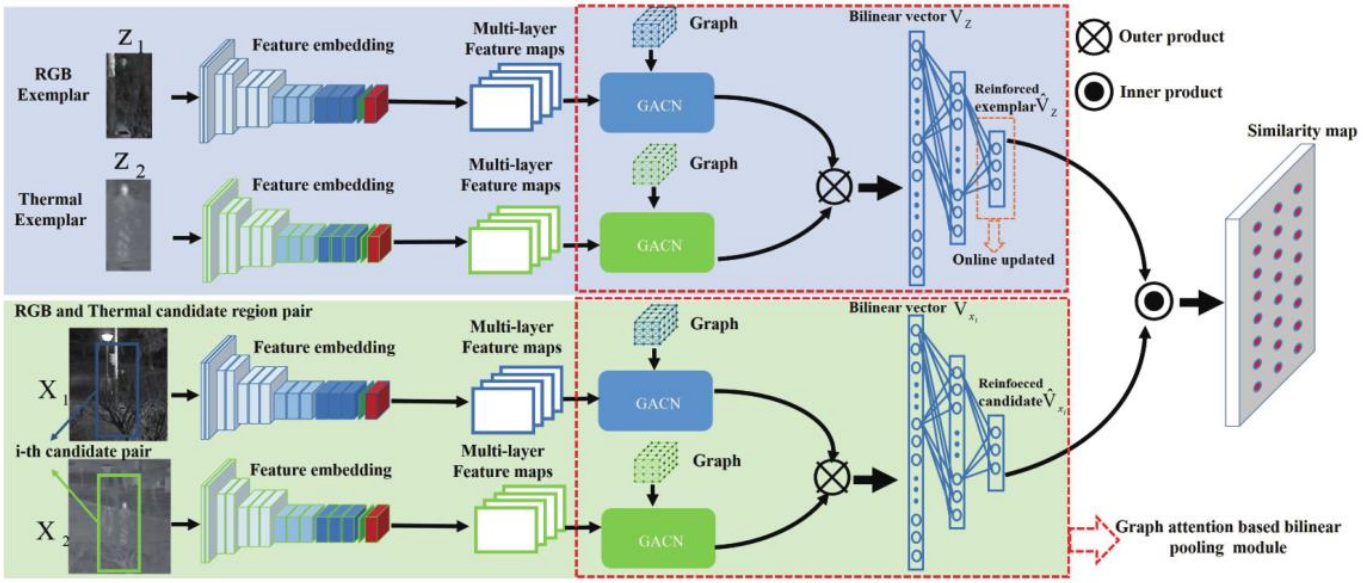


图 1 所提出面向四流的 Siamese 网络(FS-Siamese)结构示意图。

(3)将元学习应用于基于双线性池化的图注意力更新，利用类别信息在线限制样本学习到与当前跟踪结果相似的语义表示，有助于区分增强样本和增强候选样本。

(4)在 GTOT、RGBT234、CUB-2002011、FGVC-aircraft 和 Cars 数据集上的大量实验表明，基于图注意力的双线性池化模块可以有效地融合 RGB-T 跟踪中的多域多层特征图，同时可以扩展到其他多模型融合任务。

二、FS-Siamese网络结构介绍

1. 网络结构概述

我们的面向四流的 Siamese 网络(FS-Siamese)结构如图 1 所示，整体网络包含四个嵌入流。两个流用于嵌入目标范例(目标模板)对$\boldsymbol{Z_1}$和$\boldsymbol{Z_2}$。另外两个流用于在搜索区域内嵌入候选对($\boldsymbol{X_i^1}$和$\boldsymbol{X_i^2}$)。特征嵌入后，通过基于图注意力的双线性池化对样本嵌入对和第i个候选嵌入对分别进行强化融合。这可为内积计算提供一个局部强化的目标外观表征。在传统的 Siamese 网络中，目标位置的准确性依赖于样本和候选目标之间的相互关联。相比之下，我们的网络结构可以给出更准确的相似度计算结果。这是因为我们采用了基于图注意力的双线性池化模块，充分利用了多源嵌入对中固有的部分特征交互。

2. 基于图注意力的双线性池化模块

双线性池化是一种很有前途的模型，它可以克服线性池化的局限性，因为它使用外积来探索特征通道之间的成对相关性。假设我们得到两个域特征映射张量$\boldsymbol{F^1} \in \mathbb{R}^{N\times K\times C}$，$\boldsymbol{F^2} \in \mathbb{R}^{N\times K\times C}$ (N和K为单个特征映射的长度和宽度，C为特征映射通道的个数)。利用外积将两个张量的位置相乘，并将所有积集合在一起，最终可以得到双线性向量$\boldsymbol{V} \in \mathbb{R}^{c^2\times 1}$。由于特征图中的单个元素对应原始图像中的某个块，如果将目标块视为局部形式，双线性池化中的外积实际上可以探索两个图像域中局部形式之间的结构关系。这样我们就可以使用条件部分信息来表示目标外观。将张量$\boldsymbol{F^1}$和$\boldsymbol{F^2}$重新化为矩阵形式$\tilde{\boldsymbol{F}}^{\boldsymbol{1}} \in \mathbb{R}^{NK\times C}$和$\tilde{\boldsymbol{F}}^{\boldsymbol{2}} \in \mathbb{R}^{NK\times C}$，则双线性池化向量可以表示为:

$$\boldsymbol{V} = bilinear\left(\tilde{\boldsymbol{F}}^{\boldsymbol{1}}, \tilde{\boldsymbol{F}}^{\boldsymbol{2}}\right) = vec\left(\left(\tilde{\boldsymbol{F}}^{\boldsymbol{1}}\right)^{\boldsymbol{T}}\tilde{\boldsymbol{F}}^{\boldsymbol{2}}\right)$$

其中$\tilde{\boldsymbol{F}}^{\boldsymbol{1}} = \left[\tilde{\boldsymbol{f}}_{\boldsymbol{1}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{1}}, \ldots, \tilde{\boldsymbol{f}}_{\boldsymbol{C}}^{\boldsymbol{1}}\right]$，$\tilde{\boldsymbol{F}}^{\boldsymbol{2}} = \left[\tilde{\boldsymbol{f}}_{\boldsymbol{1}}^{\boldsymbol{2}}, \ldots, \tilde{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{2}}, \ldots, \tilde{\boldsymbol{f}}_{\boldsymbol{C}}^{\boldsymbol{2}}\right]$，向量$\boldsymbol{V}$中的第$(j-1)C+i$个元素记为$\boldsymbol{V}_{(j-1)C+i} = \left(\tilde{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{1}}\right)^{\boldsymbol{T}}\tilde{\boldsymbol{f}}_{\boldsymbol{j}}^{\boldsymbol{2}}$。$bilinear(\cdot)$表示双线性运算。向量$\tilde{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{1}}$ (或$\tilde{\boldsymbol{f}}_{\boldsymbol{j}}^{\boldsymbol{2}}$)中的每个元素表示图像块的条件局部形式。上式表明，每个局部表示具有同等的重要性，但却忽略了一个事实，对于多模型融合的$\tilde{\boldsymbol{F}}^{\boldsymbol{1}}$和$\tilde{\boldsymbol{F}}^{\boldsymbol{2}}$，其贡献实际上是不同的。

基于此，我们设计了一个基于图注意力的双线性池化模块来开发共同注意力机制。$\boldsymbol{V}$的元素被重新表述为:

$$\boldsymbol{V}_{(\boldsymbol{j-1})\boldsymbol{C}+\boldsymbol{i}} = \left(\tilde{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{1}}\right)^{\boldsymbol{T}}\boldsymbol{W}_{ij}\tilde{\boldsymbol{f}}_{\boldsymbol{j}}^{\boldsymbol{2}}$$

其中，共同注意权重矩阵$\boldsymbol{W}_{ij}$的目的是表明$\tilde{\boldsymbol{f}}_{\boldsymbol{i}}^{\boldsymbol{1}}$和$\tilde{\boldsymbol{f}}_{\boldsymbol{j}}^{\boldsymbol{2}}$中元

素之间的相关性。

本文的研究动机是将目标嵌入、共同注意权重矩阵估计和特征嵌入融合集成到一个统一的端到端网络结构中。为此，本文提出的基于图注意力的双线性池化模块将图注意力卷积网络和外积相结合，可以有效地利用消息传递在 RGB 和热图像中定位信息块，且计算复杂度低。基于图注意的双线性池化模块有关公式描述如下。

基于矩阵分解，$\boldsymbol{W}_{ij}$可分解为: $\boldsymbol{W}_{ij}=\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{D}_{ij}\boldsymbol{P}$，其中$\boldsymbol{D}_{ij}$是对角矩阵，可以进一步分解为两个对角矩阵$\boldsymbol{D}_{ij}=\boldsymbol{S}_i^{\,T}\boldsymbol{S}_i$。定义$\boldsymbol{D}_i=\boldsymbol{S}_i\boldsymbol{P}$，$\boldsymbol{D}_j=\boldsymbol{S}_j\boldsymbol{P}$。因此:

$$\boldsymbol{V}_{(\boldsymbol{j}-\boldsymbol{1})\boldsymbol{C}+\boldsymbol{i}}=\left(\tilde{\boldsymbol{f}}_i^1\right)^{\boldsymbol{T}}(\boldsymbol{D}_i)^{\boldsymbol{T}}\boldsymbol{P}^{\boldsymbol{T}}\boldsymbol{P}\boldsymbol{D}_j\tilde{\boldsymbol{f}}_j^2=\left(\boldsymbol{P}_i\tilde{\boldsymbol{f}}_i^1\right)^{\boldsymbol{T}}\left(\boldsymbol{P}_j\tilde{\boldsymbol{f}}_j^2\right)$$

定义$\tilde{\boldsymbol{f}}_i^1=\boldsymbol{P}^{\boldsymbol{T}}\hat{\boldsymbol{f}}_i^1$，所以:

$$\boldsymbol{P}_i\tilde{\boldsymbol{f}}_i^1=(\boldsymbol{P}\boldsymbol{D}_i\boldsymbol{P}^{\boldsymbol{T}})\hat{\boldsymbol{f}}_i^1。$$

$\boldsymbol{D}_i$是方阵，可以进一步分解。假设$\boldsymbol{P}$是拉普拉斯矩阵的特征向量，$(\boldsymbol{P}\boldsymbol{D}_i\boldsymbol{P}^{\boldsymbol{T}})\hat{\boldsymbol{f}}_i^1$可以看作是图卷积。同样的，$\boldsymbol{D}_j$也可以使用图卷积进行更新。

基于上述分析，设$\boldsymbol{G}\left(\hat{\boldsymbol{F}}^1,\hat{\boldsymbol{A}}^1\right)$和$\boldsymbol{G}\left(\hat{\boldsymbol{F}}^2,\hat{\boldsymbol{A}}^2\right)$分别为 RGB 和热特征映射张量的属性图，其中$\hat{\boldsymbol{F}}^i$ (i=1,2)中的行表示为第i个图中的节点，$\hat{\boldsymbol{A}}^{\mathrm{i}}$为编码节点对之间的成对相似度的相邻矩阵。基于双线性池化的多图学习问题表述为:

$$\boldsymbol{V}=bilinear\left(\boldsymbol{G}\left(\hat{\boldsymbol{F}}^1,\hat{\boldsymbol{A}}^1\right),\boldsymbol{G}\left(\hat{\boldsymbol{F}}^2,\hat{\boldsymbol{A}}^2\right);\boldsymbol{\Theta}\right)$$

其中图$\boldsymbol{G}\left(\hat{\boldsymbol{F}}^1,\hat{\boldsymbol{A}}^1\right)$和图$\boldsymbol{G}\left(\hat{\boldsymbol{F}}^2,\hat{\boldsymbol{A}}^2\right)$可以被图卷积神经网络学习，$\boldsymbol{\Theta}=\{\boldsymbol{\Theta}^1,\boldsymbol{\Theta}^2\}$定义为图卷积神经网络的参数集，$bilinear(\cdot)$是指利用外积动态聚合两个图卷积神经网络的双线性算法。

我们建立图注意力卷积网络，以实现在没有任何先验知识的情况下进行图学习。具体来说:

$$\boldsymbol{P}_i\tilde{\boldsymbol{f}}_i^1=\sigma\left(\Sigma_{k\in N(i)}\eta(i,k)\hat{\boldsymbol{f}}_k^1\right)$$

其中$\eta(i,k)$表示节点i和k之间边界的权值，$\sigma(\cdot)$是激活函数，$N(i)$表示节点i的邻域集。根据$\boldsymbol{V}$的表达式，我们自适应地学习$\eta(i,k)$来估计$\boldsymbol{P}_i\tilde{\boldsymbol{f}}_i^1$。

同样，$\boldsymbol{P}_j\tilde{\boldsymbol{f}}_j^2$也可以用相似的方法估算。$\eta(i,k)$的计算方法如下:

$$\eta(i,k)=\frac{exp\left(LeakyReLU\left(\boldsymbol{\beta}^T\left[\boldsymbol{U}\hat{\boldsymbol{f}}_i^1\parallel\boldsymbol{U}\hat{\boldsymbol{f}}_k^1\right]\right)\right)}{\Sigma_{s\in N(i)}exp\left(LeakyReLU\left(\boldsymbol{\beta}^T\left[\boldsymbol{U}\hat{\boldsymbol{f}}_i^1\parallel\boldsymbol{U}\hat{\boldsymbol{f}}_s^1\right]\right)\right)}$$

式中$\boldsymbol{\beta}$为单层前馈神经网络的参数向量，$\boldsymbol{U}$为表示$\tilde{\boldsymbol{A}}$和$\hat{\boldsymbol{A}}$关系的参数矩阵。$\parallel$为串联算子，$LeakyReLU(\cdot)$为激活函数。

3. 更新策略

我们希望将基于图注意的双线性池化结果的更新重新表述为一个单样本学习问题。因为当前的跟踪结果实际上是正样本，与样本相似度较低的候选样本可视为负样本。无论当前的候选样本发生了多大的变化，范例和当前的跟踪结果应该仍然具有相同的类别。因此，我们可以将类别信息纳入到$\hat{\boldsymbol{V}}_z$的在线更新中，其中$\hat{\boldsymbol{V}}_z$是样本生成双线性向量后的全连通层。

为了实现元学习的目标，我们定义与$\hat{\boldsymbol{V}}_z$相似度最高的第i个候选池化结果$\hat{\boldsymbol{V}}_{x_i}$作为正样本$\boldsymbol{C}_1$，与$\hat{\boldsymbol{V}}_z$相似度最低的第$j$个候选池化结果$\hat{\boldsymbol{V}}_{x_j}$作为负样本$\boldsymbol{C}_2$。在分类中，我们引入参数向量$\boldsymbol{\phi}$，用于微调语义表征来更新样本。其在线训练损失函数定义为:

$$J(\boldsymbol{\phi})=-log\,\mathcal{P}\left(y=1\mid\hat{\boldsymbol{V}}_z\right)$$

式中$\mathcal{P}\left(y=1\mid\hat{\boldsymbol{V}}_z\right)$定义为:

$$\mathcal{P}\left(y=1\mid\hat{\boldsymbol{V}}_z\right)=\frac{exp\left(-\left\|f\left(\hat{\boldsymbol{V}}_z;\boldsymbol{\phi}\right)-\boldsymbol{C}_1\right\|^2\right)}{\sum_{k=1}^{2}exp\left(-\left\|f\left(\hat{\boldsymbol{V}}_z;\boldsymbol{\phi}\right)-\boldsymbol{C}_k\right\|^2\right)}$$

其中，$f(\cdot)$表示经过整个网络结构处理后的输出函数，参数向量$\boldsymbol{\phi}$包含特征嵌入模块，基于图注意力的双线性池化模块以及最后的内积计算这三部分所涉及到的所有参数。

三、实验结果

1. 定量跟踪实验

为了测试我们的网络结构的效率，我们在两个广泛使用的 RGB-T 数据集上进行了大量的实验:GTOT 和 RGBT234。具体结果可见于图 2 和图 3。主要是通过两个客观指标进行定量评估：精度图和成功图。精度图表示不同定位误差阈值下的累计位置误差，定位误差定义

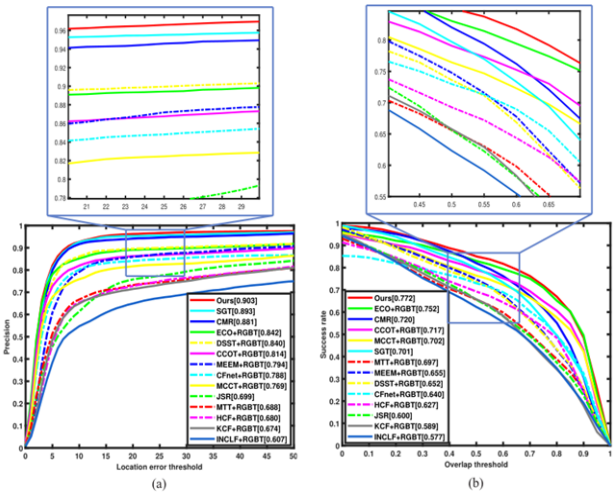


图 2　在 GTOT 数据集上的总体跟踪性能
(a)精度图，(b)成功图。

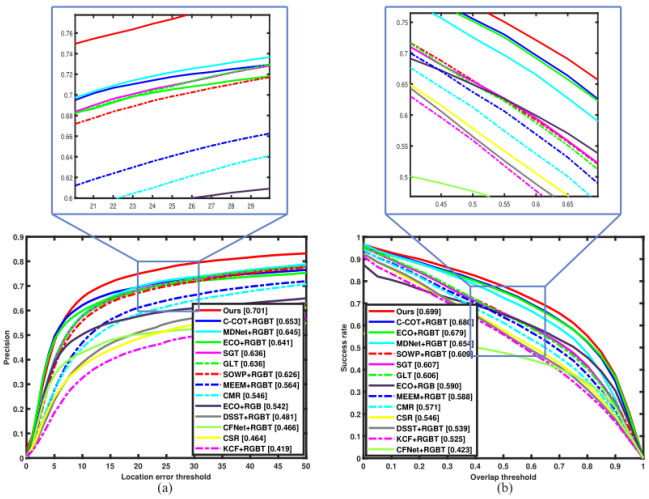


图 3　在 RGBT234 数据集的总体跟踪性能
(a)精度图，(b)成功图。

表 1　RGB-234 数据集中不同视频子集成功率的均值，最好的两个结果分别用红色和蓝色标注。

Attr. \ Meth.	Ours	ECO+RGBT	ECO+RGB	GLT	CFNet+RGBT	CMR	DSST+RGBT	CSR	KCF+RGBT	C-COT+RGBT	MDNet+RGBT	MEEM+RGBT	SOWP+RGBT	SGT	CMPP	self-SDCT+RGB
BC	56.1	52.0	49.9	50.7	28.8	37.6	42.5	34.1	37.5	50.2	48.5	52.8	50.7	50.3	53.8	44.3
CM	57.3	50.8	47.0	43.1	26.9	37.5	33.6	31.8	30.8	49.7	45.6	41.6	42.6	42.9	54.1	37.6
DEF	53.2	46.6	46.9	41.2	28.9	40.1	31.4	33.6	31.1	46.4	47.4	41.8	42.1	43.6	54.1	35.4
FM	54.6	45.6	45.3	41.9	28.6	42.2	30.3	34.7	27.3	45.3	47.6	46.3	45.5	45.2	50.8	36.3
HO	58.1	50.7	49.7	43.6	26.5	39.5	34.1	32.5	30.7	50.8	48.2	45.2	44.9	45.0	50.3	37.0
LI	58.6	52.6	54.0	48.2	32.9	34.4	43.2	34.7	39.1	53.9	43.8	48.1	49.6	45.8	58.4	48.3
LR	63.2	58.4	56.3	58.6	35.4	48.2	53.3	45.4	49.0	64.9	58.6	58.8	57.3	58.8	57.1	56.5
MB	54.5	55.2	49.5	43.3	23.6	37.9	34.1	29.5	29.9	49.9	46.7	37.8	43.2	41.1	54.1	36.1
NO	71.3	66.7	66.4	45.7	50.0	43.1	34.1	47.1	34.9	66.0	59.7	41.3	42.9	47.0	67.8	37.2
PO	62.7	62.2	61.6	50.7	42.6	45.5	43.0	43.1	40.8	62.1	57.6	49.2	52.2	51.3	60.1	43.5
SC	59.7	61.2	58.6	37.3	40.7	38.8	30.7	41.4	28.7	60.4	55.0	36.3	36.8	40.0	57.2	35.5
TC	67.2	70.0	62.4	51.5	41.1	55.0	34.2	37.6	34.7	71.9	59.5	51.9	51.6	57.2	58.3	38.8
Average	59.7	56.0	54.0	46.3	33.8	41.7	37.0	37.1	34.5	56.0	51.5	45.9	46.6	47.4	57.5	40.5

为跟踪包围框中心位置与人工标记的真实值之间的欧几里得距离，成功图反映了不同重叠阈值下的累计成功率(重叠分数大于 0.5 时的视频帧数)。由图 2 可以清晰看出，本方法的距离精度评分在 GTOT 数据集上比 ECO-RGBT 高出 5%以上。由于 ECO-RGBT 涉及热信息，其距离精度得分略高于 ECO-RGB。由图 3(a)可知，我们的方法的距离精度评分明显高于其他相比较的方法。同样，由图 3(b)，我们的方法在成功图中也获得了最好的结果。综上，得以验证我们的 FS-Siamese 网络在两个数据集上均表现出良好的性能。

此外，我们还在 RGB-234 数据集上测试了 12 个具有挑战性因素的影响，其结果如表 1 所示。从这次测试中我们可以清楚地看到，我们的方法在大多数有挑战性因素中获得了第一名。具体来说，重遮挡(HO)非常具有挑战性，因为只能从 RGB 和热目标中提取少量有用的信息。由于这个原因，最先进的跟踪方法，如 ECO, CMR 和 GLT 在这种情况下的跟踪性能很差。与传统方法相比，我们的方法成功率比顶级方法 CMPP 高 10%以上。除了 HO 之外，背景杂波(BC)、摄像机运动(CM)、快速运动(FM)、低照明(LI)和部分遮挡(PO)通常被认为是具有挑战性的场景，可作为验证跟踪精度的代表性测试。显然，与 CMPP 相比，我们的方法也能提高 6%以上的成功率。由于热交叉(thermal Crossover, TC)会严重干扰热目标的外观，池化模块在探索块关系时可能会受到更大的负面影响。由于这个原因，整体深度特征导向跟踪器(ECO+RGBT)提供了最好的成功率。表 1 的测试结果可知我们的方法可以有效地使用基于图注意力的双线性池化模块来增强有挑战性场景下的跟踪性能。

2. 定性跟踪实验

我们在图 4 中展示了定性跟踪性能。其中每个场景

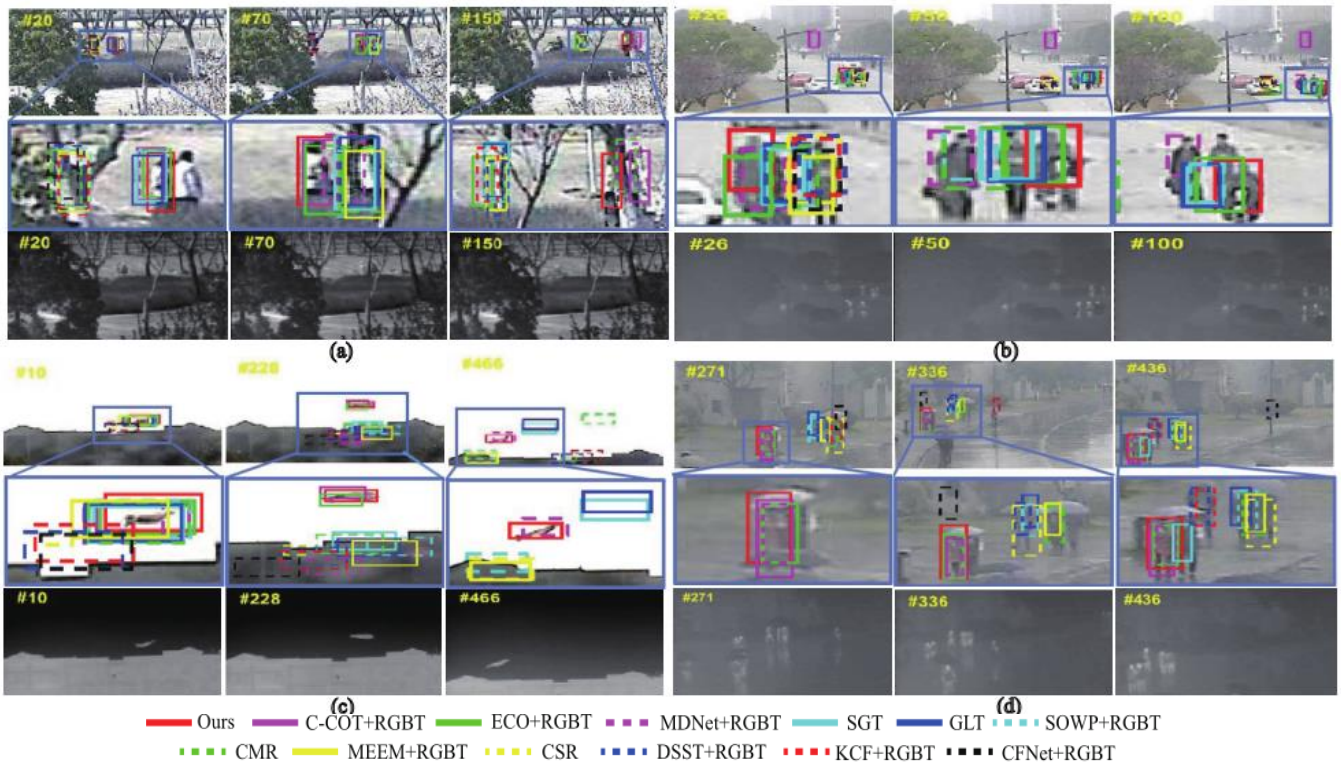


图 4 视频对定性结果(a) Diamond 视频对 (b) Elecbike3 视频对 (c) Kite4 视频对 (d) Manafterrain 视频对。

表 2 基于图注意力的双线性池化模块的消融实验设置。

Backbone	Outer product	GACN	Updating module	Methods
VGG-16	√	√	√	**Ours**
VGG-16	√	√		**Ours I**
VGG-16	√			**Ours II**

随机选择 3 个视频序列。运动目标在 diamond 序列中常被树干遮挡，即使是最先进的方法往往也会在严重遮挡后失去目标。从图 4(a)可以看出，无论局部遮挡还是重度遮挡，我们的方法都能跟踪目标。目标和相邻行人一起移动，造成图 4(b)中产生严重的背景杂波。在这种情况下，我们的方法可以做到与 ECO-RGBT 同样的效果，提供了良好的跟踪性能。如图 4(c)，在 kite 序列中，其他方法在第 300 帧后会开始有一定程度的漂移，而我们的方法仍然可以在整个视频帧中跟踪目标。图 4(d)为下雨情况下光照较低，通过该图可以看出，我们的方法可以有效地利用热信息来补充 RGB 序列。

3. 消融实验

基于图注意力的双线性池化模块是我们的 FS-Siamese 网络的核心模块，主要包括三个部分:图注意力卷积网络(GACN)、外部乘积和更新模块。在测试中，我

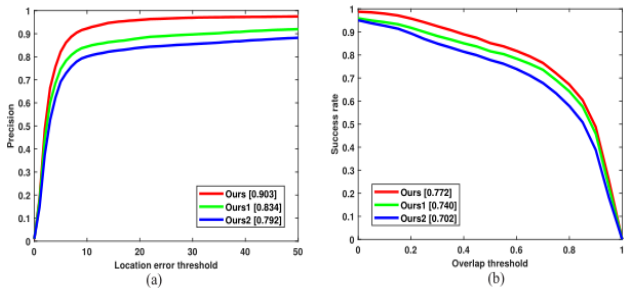


图 5 GTOT 数据集上基于图注意力的双线性池化模块消融试验。其中红色曲线对应实验设置中的方法 ours，绿色曲线对应实验设置中的方法 Ours I，蓝色曲线对应实验设置中的方法 Ours II。

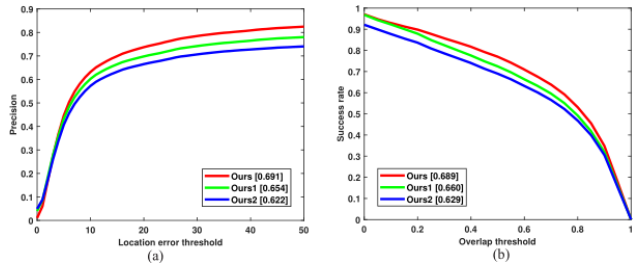


图 6 RGBT234 数据集上基于图注意力的双线性池化模块消融试验。不同颜色曲线实验设置规则同图 5。

们对 GTOT 和 RGBT234 数据集进行了消融实验，实验设置如表 2 所示，以展示不同部分的有效性。具体结果如图 5 和图 6 所示。从图 5 和图 6 中，我们可以看到两

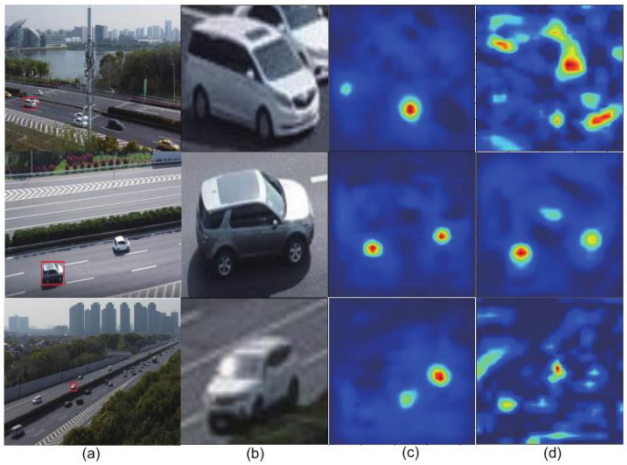


图 7　细粒度分类测试实验。 (a)测试图像，(b)局部放大图像，(c) MA-CNN+GACN 评估掩码，(d) MA-CNN 评估掩码。

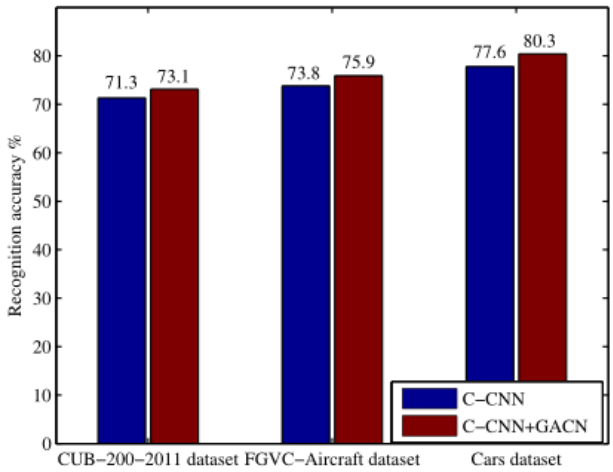


图 8　测试 GACN 普适性的细粒度识别。

个数据集上的精度图和成功图表明了我们基于图注意力的双线性池化模块的有效性。

在基于图注意力的双线性池化模块中，GACN 是及其重要的一部分，它可以通过探索局部特征交互来突出重要的图像块。基于此，我们设计了一个细粒度分类测试实验来展示 GACN 的有效性。具体来说，我们在 MA-CNN 网络的 Conv 层的末尾添加了 GACN。这样，目标嵌入就会更加关注信息丰富的图像块。详细测试实验如图 7 所示，图 7(c)和图 7(d)中评估的子区域掩码可以表明 GACN 的有效性。例如，第一行和第三行的局部放大图像分辨率较低，但 MA-CNN+GACN 仍能定位到信息子区域(如图 7(c))。相比之下，原始方法可能会在掩码中包含无信息的背景噪声(如图 7(d))。

此外，B-CNN 是细粒度识别中比较著名的一种方法，可以使用双线性池化融合两个网络结构的特征图。因此，我们将基于图注意力的双线性池化模块扩展到该方法，即“B-CNN+GACN”，以验证 FS-Siamese 创新的普适性。测试在三个细粒度识别数据集上进行:CUB-200-2011, FGVC-aircraft 和 Cars。从图 8 中我们可以清楚地看到，与原始 B-CNN 方法相比，B-CNN+GACN 可以明显提高 3%以上的识别精度。

四、总结

在本文中，我们提出了一个面向四流的 Siamese 网络(FS-Siamese)来有效地融合 RGB 和热信息。我们的网络得益于提出的基于图注意力的双线性池化模块，该模块可以采用共同注意力机制来探索 RGB 和热目标之间的部分特征相互作用。此外，我们采用元学习对双线性池化结果进行更新，可以对目标与其周围背景的空间关系进行在线更新。

在 GTOT 和 RGBT234 数据集上的大量实验表明，与最先进的 RGB 和 RGB- T 跟踪器相比，所提出的 FS-Siamese 网络可以提供更好的性能。

责任编委　王金甲

参考文献

[1] Y .Choi, N. Kim, S. Hwang, K.Park, J. S. Yoon, K.An, and I. S.Kweon, “Kaist multi-spectral day/night data set for autonomous and assisted driving,” IEEE Transactions on Intelligent Transportation Systems,vol. 19, no. 3, pp. 934–948, 2018.

[2] A.Leykin and R.Hammoud, “Pedestrian tracking by fusion of thermal-visible surveillance videos,” Machine Vision and Applications, vol. 21,no. 4, pp.587–595, 2010.

[3] M. Talha and R.Stolkin,"Particle filter tracking of camouflaged targets by adaptive fusion of thermal and visible spectra camera data," IEEE Sensors Journal, vol. 14, no. 1, pp.159–166, 2013.

[4] C.Li, N.Zhao, Y.Lu, C.Zhu, and J.Tang, "Weighted sparse representation regularized graph learning for RGB-T object tracking," in Proc.of the ACM international conference on Multimedia, pp.1856-1864,2017.

[5] C.Li, C. Zhu, J.Zhang, B.Luo, and J.Tang, "Learning local-global multi-graph descriptors for RGB-T object tracking," IEEE Transactions on Circuits and Systems for Video Technology, vol. 29, no. 10, pp.2913–2926, 2019.

[6] C.Li, H.Cheng, S.Hu, X.Liu, J. Tang, and L. Lin, "Learning collaborative sparse representation for grayscale-thermal tracking," IEEE Transactions on Image Processing, vol. 25, no. 12, pp.5743–5756, 2016.

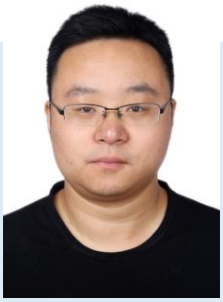
[7] X. Lan, M.Ye, R.Shao, B. Zhong, P . C. Y uen, and H. Zhou, "Learning modality-consistency feature templates: A robust rgb-infrared tracking system," IEEE Transactions on Industrial Electronics, vol. 66, no. 12, pp.9887–9897, 2019.

[8] Y . Zhu, C. Li, B. Luo, J. Tang, and X. Wang, "Dense feature aggregation and pruning for rgbt tracking," in Proc. of the ACM International Conference on Multimedia, pp.465-472,2019.

[9] C. Li, A. Lu, A. Zheng, Z. Tu, and J. Tang, "Multi-adapter rgbt tracking," in Proceedings of the IEEE International Conference on Computer Vision Workshops, pp.2262-2270,2019.

[10] Q.Wang, Z.Teng, J.Xing, J.Gao, and S.Maybank, "Learning attentions: residual attentional siamese network for high performance online visual tracking," in Proc. of IEEE Conference on Computer Vision and Pattern Recognition, pp.4854-4863,2018.

[11] Z. Zhu, Q. Wang, B. Li, W.Wu, J. Yan, and W. Hu, "Distractor-aware siamese networks for visual object tracking," in Proc. of the European Conference on Computer Vision, pp.101–117,2018.



周全

南京邮电大学副教授，硕士生导师。研究方向包括深度学习、模式识别和计算机视觉。江苏省"青蓝工程"青年骨干教师。中国计算机学会，图像图形学会高级会员，江苏省自动化学会模式识别专委会常务委员。IEEE 和 IAPR 高级成员。已主持国家自然科学基金，江苏省自然科学基金等 10 余项。已发表学术论文 70 余篇，包括 IEEE TIP、IEEE TITS、IEEE TMI、IEEE TNNLS 等。目前担任 70 多个 SCI 期刊审稿人，并担任 IEEE/SPIE ISAIR2019-2023、IEEE ICME2019 和 PRCV2022 区域主席。同时担任 Computer and Electrical Engineering 期刊编辑，以及 IEEE TMM、PR、MMTA 和 Visual Intelligence 等期刊的首席客座编辑。

Email: quan.zhou@njupt.edu.cn

康彬

南京邮电大学副教授，硕士生导师，一直从事计算机视觉、深度学习理论及应用研究工作，具体研究方向包括目标跟踪、细粒度识别及多模态信息融合等。主持参与了与深度学习应用相关多项科研项目，如：国家自然科学面上、青年基金以及军委科技委基础研发等。截至目前，共发表学术论文 40 余篇，其中发表的高水平论文包括 IEEE TIP、IEEE TNNLS、 IEEE TCSVT、IEEE TITS、AAAI 等。目前担任 IEEE TMM, IEEE SPL, IET Image Processing 等学术期刊审稿人，除此之外担任网络多媒体专委会秘书、IEEE ICC, GlobleCom 以及 WCSP 等通信领域知名国际会议技术委员会委员。

Email: kangb@njupt.edu.cn

顶会观察

NeurIPS 2022

河北工业大学　许铮铧

国际神经信息处理系统大会(Neural Information Processing Systems，简称 NeurIPS)是机器学习和计算神经科学领域的顶级国际会议，是人工智能领域水平最高，录用难度最大，论文影响力最强的会议之一，与 ICML、ICLR 并列为全球机器学习三大顶会。NeurIPS 不仅属于中国计算机学会推荐的人工智能领域 A 类国际学术会议，而且是 Core Conference Ranking 推荐 A*类会议，H5-index 高达 198，Impact Score 达到 33.49。本届 NeurIPS 大会于 2022 年 11 月 28 日至 2022 年 12 月 9 日举行，为期两周，其中第一周为线下会议，在美国新奥尔良 Ernest N. Morial 会议中心召开，第二周则为线上会议，以在线交流、录制视频和异步聊天等多种方式为无法到线下会场的与会者提供线上讨论与互动的机会。参会人员身份被划分为 Full-time Student、Academic 和 Non-academic，其中学生身份的参会人员会获得注册等费用上减免。

一、 会议亮点

多样的会议展现：1、线下海报展示：为了方便线下的快速交流，本次会议优先考虑以海报会议为中心的面对面互动讨论的方式。2、线上论文会议：除了直接从与作者交流了解他们工作外，每篇论文在 NeurIPS 网站上都有一个单独页面，线上与会人员可以在网页中找到一个 5 分钟的视频和一个专门聊天频道来异步讨论该论文相关工作。3、期刊展示：今年 NeurIPS 引入了从期刊到会议的轨道，与会者可以在其中了解被 NeurIPS 相关期刊接受的论文的工作。本次会议共有 2,672 篇会议论文被接收，此外还有 41 篇来自 JMLR 期刊和 33 篇来自 ReScience 的期刊论文在大会期刊轨道进行了展示。

严格的委员会评审：本届 NeurIPS 2022 大会由 Sanmi Koyejo(斯坦福大学)和 Shakir Mohamed(DeepMind)担任 General Chair，由 Alekh Agarwal(谷歌研究院)、Alice Oh(韩国科学技术院)、Danielle Belgrave(DeepMind)和 Kyunghyun Cho(纽约大学)担任 Program Chair。同时，各个分区的主席均来自著名的大学和科研机构，如纽约大学、牛津大学和普林斯顿大学，或者主流的人工智能企业，如 DeepMind、微软和谷歌等。本届 NeurIPS 的绝大部分论文都经过不少于四名专家的评审，并且通过 rebuttal 的形式让作者与评审专家能够对审稿意见中的相关问题进行充分讨论，以最大程度解决审稿意见中的分歧。但从大会组织方发布的相关公告中仍发现，虽然经过了严格的评审和充分的讨论，论文作者和审稿人之间仍存在一定的分歧；因此大会组委会还专门对审稿分歧这一问题展开调查，通过分析 NeurIPS 2022 中记录的关于论文质量的分歧(包括共同作者之间、作者和审稿人之间以及审稿人委员会成员之间的分歧)。组委会的调查结果表明论文质量的评估是一项从本质上具有挑战性的复杂任务，即使通过严格的评审流程，依然没有客观正确的答案，因此，组委会也呼吁研究人员应对论文提交结果持保留态度。

伦理审查程序：为了激发整个领域的伦理研究、实践和反思，本次会议将伦理审查程序作为提高 NeurIPS 2022 作者和审稿人伦理意识和参与度的一个重要环节。通过结合前几版伦理审查流程的成功经验，组委会在本次会议中着重讨论了新版伦理审查流程可靠性和适用性，从而实现进一步巩固伦理审查政策，建立一个向前

推进式的审核流程。组委会表示伦理审查目的不是惩罚性或排他性的；相反，伦理审查旨在告知、教育和阐明道德问题，以便作者可以通过公开讨论的方式来解决冲突问题。今年，组委会还发布 NeurIPS 临时道德准则初稿，旨在为社区提供更全面的道德准则和会议的期望。

二、 录用情况

NeurIPS 2022 收到的有效投稿和录用数量都有显著提高，大会共收到了 10411 篇投稿，最终接收了 2672 篇论文，接收率约为 25.6%。相较于 2021 年，今年 NeurIPS 2022 的投稿量提升了 12%(1289 篇)，整体录用论文数量也在去年的 2334 篇的基础上提升了 338 篇，增长率约为 14%。在接收的论文中，有 199 篇论文录用为 Oral Presentations，比去年增加了 143 篇，Oral 率约为 7.4%(比去年提升 3%)。今年 Spotlight 文章数量为 523 篇，相较于去年提升了 263 篇。NeurIPS2022 会议涵盖的方向包括：识别(检测、分类与检索)、3D 视觉、图像与视频的生成、深度学习与表示学习、视觉与语言、迁移学习、图神经网络、场景分析和理解、无监督学习、数据集与评估等方向。值得关注的是，在本次会议中，标题出现词汇最多的前五名依次为 Multi、Training、Self、Optimal 和 Learning，摘要中出现最多的五个词汇依次为 Learning、Model、Data、Models 和 Methods。来自中国的学术机构在本次大会取得了相当不俗的成绩。据大会公开的文章接受列表统计(以第一作者单位为准)，清华大学以 85 篇论文位列榜首，北京大学排名全球第六名，上海交大和中科大分列全球第九和第十二名。此外斯坦福大学和谷歌在本次会议中仍有不俗表现，分别有 72 篇和 65 篇论文被录用。此外，华人学者也占据了个人录用情况统计排名的前列，悉尼大学刘同亮(Tongliang Liu)教授，香港浸会大学韩波(Bo Han)教授、上海交通大学严骏驰(Junchi Yan)教授和伊利诺伊大学厄巴纳-香槟分校李博(Bo Li)教授均以 11 篇论文被录用的优异成绩共同位列排行榜的第三位。

三、 主题报告

本届 NeurIPS 2022 会议邀请了七位主题报告讲者(Keynote Speakers)，报告主题内容涵盖了包括语言模型、脑启发研究、扩散模型、图神经网络等方面，其中具有较高讨论度和影响力的报告简介如下。

The Forward-Forward Algorithm for Train-ing Deep Neural Networks. 图灵奖得主、深度学习先驱多伦多大学 Geoffrey Hinton 教授带来了关于一种新的神经网络学习方法——前向-前向算法(Forward-Forward Algorithm，FF)的介绍。该方法不同于反向传播算法，它包含两个前向传递过程，其中一个使用正数据，另一个使用网络自身生成的负数据。Hinton 认为，FF 算法的优点在于：它更好地模拟了大脑皮层的学习过程，并且能够在使用低功耗硬件模拟的同时节省大量计算资源。他主张放弃计算机软硬件分离的形式，而是将未来的计算机设计为“非永生的”(mortal)，以提高计算资源的利用率。因此，FF 算法是最佳的学习方法之一，能够高效地运行在这种硬件中。

Are Large Language Models Sentient? 纽约大学神经科学家 David Chalmers 教授带来了 GPT-3、LaMDA 2 和相关的大型语言模型是否具有感知能力的报告。近年来，随着自然语言处理领域的蓬勃发展，关于大型语言模型是否具有轻微意识的问题呈现了不同的讨论。Chalmers 从生物学，感官系统角度出发，结合全局模型的理念对 Large Language Model(LLM)进行多角度的分析，他认为关于 AI 意识的相关问题不能简单的作出判断，并预测在十年内，即使我们没有人类水平的 AGI，我们也可能拥有非常智能化意识的系统。

Conformal Prediction in 2022. 来自斯坦福大学的 Barnum-Simons 数学与统计学讲席教授 Emmanuel Candes 为与会者报告了共型预测在 2022 年的发展现状和主要的研究工作。共型预测近年来在学术界和工业界引起了广泛的讨论和研究，大有流行之势。共型预测主要为未来的数值观测提供了准确的预测区间，并且无需进行任何分布假设。本次演讲回顾了共型预测的基本原理，并回顾了过去 2-3 年内的一些主要贡献。Candes 还讨论了适用于定量和分类标签增强一致性的方式，并尝试将共型预测应用于金融经济学，市场行为等领域。最后，Candes 借用预测 COVID19 病例轨迹的示例进一步说明共型预测的具体应用方向。

Interaction-Centric AI. 来自韩国科学技术院计算学院的 Juho Kim 教授介绍并讨论了以交互为中心的人工智能。虽然现有的模型算法在很多人工智能领域都取得了卓越的性能，但这些模型上的进步很少转化为对现实世界用户的影响。当前，很多模型由于忽视了人类与人工智能交互的动态性和复杂性，从而进一步限制人工智能在实际环境中的应用。因此，报告者认为人机交互应该被视为设计人工智能应用程序的首要对象。在本次演讲中，Kim 展示了一些使用 AI 来支持复杂现实生活任务的新型交互系统，并讨论了人机交互设计中的复杂难题和解决方案，分享模型设计经验。最后，Kim 呼吁要为“以交互为中心的 AI”建立稳健的区块——一种设计和工程化人机交互的系统方法，进而补充并克服以模型和数据为中心的人工智能模型的局限性。

The Data-Centric Era: How ML is Becoming an Experimental Science. 来自谷歌大脑的 Isabelle Guyon 教授带来了《以数据为中心的时代：ML 如何成为一门实验科学》的主题报告。Guyon 表示，最近的分析显示，使用更多的数据和更深层的网络并不是一种可持续的进步方式。同时，其他指标表明机器学习越来越依赖于良好的数据和基准。这不仅可以训练更强大/更紧凑的模型，还可以合理评估新想法并强调测试模型的可靠性、公平性和安全性。在 2021 年，NeurIPS 启动了数据集和基准测试轨道，并诞生了以数据为中心的 AI 计划，开启了“以数据为中心的时代”。与设计和训练 ML 模型相比，数据科学家将花费更多时间在理解问题、设计实验和选取工程数据集上。Guyon 通过挑选一些引起争议并发人深省的热门话题进行讨论，并在报告结尾强调当前需解决的一些紧迫的问题。

四、 热点论文

本届 NeurIPS 2022 共有 13 篇论文获得杰出论文奖，1 篇论文获得时间检验奖，2 篇论文获颁杰出数据集和基准论文奖。获奖论文涵盖了梯度评估，最优学习，梯度下降，文档检索等方面。

时间检验奖：ImageNet Classification with Deep Convolutional Neural Networks. AlexNet 因首次将卷积网络(CNN)应用于 ImageNet 训练挑战赛，并获得远超当时的其他方法的优越表现而获得了该奖项。AlexNet 的成功对机器学习社区产生了巨大的影响，标志着卷积神经网络(CNN)成为图像分类的核心模型，随后 ImageNet 冠军的获胜者都采用了卷积神经网络结构，使得深度学习迎来了新的高潮。随着卷积网络在分类任务上的成功，CNN 网络随后也在目标检测任务，图像分割任务等一系列深度学习领域中继续发扬光大。

杰出论文：ProcTHOR: Large-Scale Embodied AI Using Procedural Generation. 海量数据集和高容量模型推动了计算机视觉和自然语言理解的许多最新进展，来自华盛顿大学的 AI2 PRIOR 团队受此启发提出了一种基于 AI 环境的生成框架 ProcTHOR。该框架能够对各种交互式、可定制式虚拟环境的任意大数据集进行采样，在交互和操纵任务中训练和评估模型。团队通过使用 10000 套 AI 生成的房屋样本和一个简单的神经模型，展示了 ProcTHOR 的威力和潜能。在 ProcTHOR 框架下使用 RGB 图像进行训练获得的模型，在没有明确的映射和人工任务监督的前提下，能够在 6 个 AI 基准任务上取得最先进成果，证明 ProcTHOR 框架强大。

杰出论文：High-dimensional limit theorems for SGD: Effective dynamics and critical scaling. 来自纽约大学的 Courant 团队研究了在高维区域中具有恒定步长的随机梯度下降(SGD)的尺度极限并且证明了有限维函数的轨迹在维数达到无穷大时的极限定理。该方法允许选择跟踪的有限维函数的轨迹和步长，并产生了 ODE 和 SDE 极限。在从多峰时间尺度收敛，以及从随机初始化收敛到概率远离零的次优解等方面展示了惊人的效果提升。

杰出数据集和基准论文奖：MINEDOJO: Building Open-Ended Embodied Agents with Internet-Scale Know-ledge. 受人类在开放世界中不断学习和适应的启发，来自 NVIDIA 的团队提出了构建通用 Agent 的三位一体要素：(1)支持多种任务和目标的环境；(2)大规模多模态知识数据库；(3)灵活可扩展的 Agent 体系结构。该团队在此基础上提出了 MINEDOJO，这是一个基于流行的 Minecraft 游戏构建下的新框架，它具有一个模拟套件，其中包含数千种不同的开放式任务，以及一个包含 Minecraft 视频教程、维基页面和论坛讨论的

互联网规模知识库。该方法能够解决各种以自由形式指定的开放式任务，而无需任何手动设计的密集奖励。

杰出论文：Is Out-of-Distribution Detection Learnable? 在 out-of-distribution (OOD)领域，来自悉尼科技大学澳大利亚人工智能研究所的团队提出了 OOD 检测的一种新的学习理论：probably approximately correct (PAC)。该团队首先找到 OOD 检测可学习性的必要条件；然后借助该条件，证明了在某些场景下 OOD 检测的可学习性的几个不可能性定理。基于这一观察，团队进而给出几个充分必要条件来表征 OOD 检测在一些实际场景中的可学习性，从而为监督学习提供新的优化思路。

杰出论文：Gradient Estimation with Discrete Stein Operators. 梯度估计是对分布参数的期望梯度进行近似，也是当前解决许多机器学习问题的核心。然而，当分布是离散的时，最常见的梯度估计器会出现方差过大问题。为了提高梯度估计的质量，来自斯坦福大学和清华大学的团队引入了一种基于 Stein 算子的离散分布方差缩减技术。通过使用这种技术为 REINFORCE leave-one-out 构建灵活的 control variates。并且该团队提出的 control variates 可以在线调整以获得最小化方差，并且不需要对目标函数进行额外评估。

其余获奖文章还涉及到的领域有图像文本模型，扩散生成模型，文档检索网络，梯度优化和数据修剪等方面的创新成果。其中加州大学伯克利分校的论文 LAION-5B: An open large-scale dataset for training next generation image-text models 提出了一个由 58.5 亿个 CLIP 过滤图像文本对组成的数据集 LAION-5B，使得大规模多模态模型的研究更加民主化。谷歌的论文 Photorealistic Text-to-Image Diffusion Models with Deep Language Understanding 提出了一种文本到图像的扩散模型 Imagen；该模型建立在大型 transformer 语言模型理解文本的能力上，并依赖于高保真图像生成中扩散模型的强度，达到了前所未有的照片真实性和深度的语言理解。剑桥大学的获奖论文 Gradient Descent: The Ultimate Optimizer 展示了如何通过对反向传播进行简单修改来自动计算超参数，并能够轻松地将该方法应用于其他优化器和超参数。

五、 讲习班，研讨会和竞赛

NeurIPS 2022 还开展了 13 场讲习班(Tutorials)，62 场研讨会(Workshops)和 25 项专业竞赛(Competitions)。讲习班主要聚焦机器学习及其相关新型领域的主题。部分关注度和影响力较高的讲习班如下。

- The Role of Meta-learning for Few-shot Learning
- Foundational Robustness of Foundation Models
- Advances in NLP and their Applications to Healthcare
- Incentive-Aware Machine Learning: A Tale of Robustness, Fairness, Improvement, and Performativity
- Theory and Practice of Efficient and Accurate Dataset Construction

研讨会主要是对正在进行的前沿工作和未来发展方向进行提供了一个讨论平台。今年的研讨会主要关注大型语言模型，深度学习与化学生物学科交叉等方面的问题。部分关注度和影响力较高的研讨会列举如下。

- Second Workshop on Efficient Natural Language and Speech Processing (ENLSP-II)
- Progress and Challenges in Building Trustworthy Embodied AI
- AI for Science: Progress and Promises
- Synthetic Data for Empowering ML Research
- AI for Accelerated Materials Design (AI4Mat)
- Order up! The Benefits of Higher-Order Optimization in Machine Learning

在竞赛方面，今年的竞赛同样聚焦于 Computer Vision，NLP 和 Multi-modal 等主流任务。从机器学习方法的角度出发，今年强化学习相关的竞赛显著增加。在应用方面，生物医学和健康领域的竞赛数目最多最高；与过去几年相比，竞赛数量和重要程度正在逐年增加。

部分关注度和参与度较高的竞赛列举如下。

- Causal Insights for Learning Paths in Education
- The MineRL BASALT Competition on Fine-tuning from Human Feedback
- IGLU: Interactive Grounded Language Understanding in a Collaborative Environment
- Real Robot Challenge III -Learning Dex-terous Manipulation from Offline Data in the Real World
- Second AmericasNLP Competition: Speech-to-Text Translation for Indigenous Lan-guages of the Americas
- Multimodal Single-Cell Integration Across Time, Individuals, and Batches
- Weakly Supervised Cell Segmentation in Multi-modality High-Resolution Micros-copy Images

六、 总结展望

本届 NeurIPS 2022 大会中围绕时间序列预测、分类、异常检测、表征学习以及在医疗、生物、交通、金融等方向的应用文章数量大大增加，竞争热度逐渐提升。3D 视觉、图像与视频的生成、Transformer 等领域依旧保持之前的高热度。相比于 2021 年强化学习、深度学习和表征学习等领域的火热，今年图神经网络和时序预测的文章占比有明显提升，未来可能成为更加火热的研究方向。更值得关注的是，本次会议的最佳论文中有 3 篇属于随机梯度方向，相比于模型上的改进和优化，对于底层数学算法的优化更容易获得 NeurIPS 评选委员会的青睐。与会专家认为，未来人工智能技术的发展趋势包括更强大的模型和更高效的计算机架构；在模型方面，深度学习将继续成为研究和应用的主流，同时也会出现新的架构和方法来解决现有模型的限制和缺点；在计算机架构方面，人们将继续寻求更快、更节能、更可靠的硬件来支持更大、更复杂的模型；但同时也需要关注人工智能的伦理问题和安全问题，加强监管和规范，确保人工智能的安全可控和可持续发展。笔者认为回答好上述问题将会机器学习的新机遇，从而更好地迈向更高层级更具有创造力的智能化机器学习时代。

责任编委 魏秀参

参考文献

[1] https://neurips.cc/Conferences/2022

许铮铧

河北工业大学共建电工装备可靠性与智能化国家重点实验室教授、博士生导师，河北省海外高层次人才计划入选者，河北省“优青”获得者，英国牛津大学计算机系博士、博士后、客座研究员、外聘博导。主要研究方向为：深度学习、强化学习、医学影像智能计算等。
Email:zhenghua.xu@hebut.edu.cn

上海海事大学周日贵教授访谈

2023 年 2 月 11 日，《CCF-CV 专委简报》在线采访了上海海事大学博士生导师周日贵教授。下面是采访实录。

问题 1：周老师，您好！首先，请您分享一下您的个人学习和研究经历。

我于 1997 年毕业于山东大学光电子信息工程专业，毕业后入职南昌航空航天大学，成为一名讲师。并于 2003 年在南昌航空航天大学计算机应用专业获得硕士学位，师从叶水生教授。为了能够在计算机研究领域有一些更大的突破，我于 2007 年在南京航空航天大学攻读计算机应用技术专业并获得了博士学位，师从丁秋林教授，主要研究利用量子计算来提升经典神经网络模型。这一年我离开南昌航空航天大学，入职了华东交通大学的计算机系。并于 2008 年有幸加入到潘建伟院士的团队做博士后研究；之后我先后于 2010 年在加拿大卡尔顿大学完成了博士后研究，于 2014 年受到美国北卡罗来纳州立大学的邀请成为访问学者。

在 2015 年我选择了具有鲜明特色专业的上海海事大学，并一直着力于将量子智能计算与航运大数据处理、物流监管、物流优化等学校特色相结合，为高效的海上运输提供保障。

问题 2：您致力于智能信息处理与量子智能计算方面研究，能否对我们国家在此方面的研究情况及国际地位进行一些分析和评价？您认为我们国家在此方面的研究还有哪些需要突破之处呢？

2018 年 5 月 28 日，习近平总书记在两院院士大会的讲话中强调了“以人工智能、量子信息、移动通信、物联网、区块链为代表的新一代信息技术加速突破”。在国家政策的大力支持下，我国对量子计算的投资、科研力度不断加大，先后启动“自然科学基金“、”863“计划和重点研发项目和科技创新 2030 重大专项等科研项目，推动量子计算的技术研发和产业化落地。

我国在量子计算领域研究发展较快，但过去主要以理论研究为主，最近加大了在实验研究方面的投入，参与者主要是科研机构、高校以及少数互联网企业，如阿里巴巴、腾讯和百度。

在核心论文数量、研究机构数上我国处于世界前列，基础研究能力仅次于美国，但在专利产出方面，我国明显弱于美国、英国、德国、日本等，基础研究成果转化有待加强。工程化及应用推动方面，我国与美国差距明显，国内企业要落后于 IBM、谷歌和微软等跨国企业。

对于实用化量子计算机的研发，目前普遍认为需要经过实现量子优越性、实现具有应用价值的专用量子模拟系统和实现可编程的通用量子计算机三个发展阶段。当前还处于第一阶段，实现量子优越性，也称为量子霸权。2020 年 12 月潘建伟、陆朝阳团队研制的“九章”光量子计算机率先完成了高斯玻色取样任务的量子计算优越性，之后在 2021 年 6 月，潘建伟和朱晓波研究团队又成功研制了 66 比特可编程超导量子计算原型机“祖冲之二号”，利用其中的 56 比特完成了量子计算优

越性实验。当前我国是全球唯一一个在两个技术路线上完成量子计算优越性的国家。

量子计算目前还处于原型机研发阶段，在技术上仍面临诸多挑战，比如相干时间较短，由于量子计算机容易受外界环境的影响而导致退相干，所以为了保证运算结果的可靠性，量子计算必须在其发生退相干之前全部完成。目前相干时间的上限一般为 100 微秒，在这极短的相干时间内需要量子计算机完成全部逻辑操作，就意味着要对量子逻辑门之间的切换速度提出非常高的要求；此外还有一个典型挑战就是去相干纠错编码，量子计算机通常难以避免量子比特相干出错，从而引入了纠错机制。但是关于退相干的纠错机制，目前还无法实现一个真正的能够容错且满足量子计算的逻辑比特。

问题 3：能否介绍一下您所带领的团队“智能信息处理与量子智能计算研究中心”的使命？

我们团队由青年教师、博士生和硕士生组成，团队一方面对专注于科研的成员努力创造优质科研环境，激发创新活力，鼓励成员静心于学术，夯实基础。另一方面坚持产教研相结合的方法，鼓励编程能力强的成员加入到横向项目中去，从实际项目中发现问题并解决问题以培养高素质技术技能型人才。

问题 4：能否介绍几项您认为最骄傲的科研成果？

首先是依托于现已结题的国家重点研发计划项目---特殊生物资源监测与溯源技术研究，我们团队建立了特殊生物资源信息数据库及特征谱图专用数据库；提出了基于特殊生物资源特征谱图的物种识别模型；开发了面向海关及出入境单位的特殊生物监测与溯源系统，得到了海关相关技术人员的一致认可。

其次是依托于上海市科技计划项目，我们团队开发了面向医药文本和影像资料智能解析存储大数据平台及设备建设，通过将医药的一些 pdf 格式转化成电子文档以构建数字资料库，能够帮助医院以及一些医疗机构方便快捷的检索到具体药品，包括药品名称、药品生产日期、药品的注意事项等等。

最后是团队研发的”CPS+EAM+PHM”的轨道交通供电系统智能运维一体化平台，解决了轨道交通“能源、业务、数据”多流合一及“管、用、修”业务融合的系统管控热点难题，实现了对设备、资源和业务的动态监测、优化配置、精准调度和协同运转。

问题 5：您曾两次牵头承担国家重点研发计划项目，也多次受邀作“国家重点研发计划重点专项（申报交流）”方面的报告，能否谈谈您对国家重点研发计划项目申报、承担方面的经验和建议？

在项目申报上首先要把握国家需求，针对国家现阶段受限的技术考虑是否能通过多个团队协作以突破瓶颈；其次就是项目应该整体申报，需覆盖相应指南方向的全部考核指标。并且在指南不是自己团队提的时候，也需要站在指南专家的角度，对项目做整体的策划和分解；再者申报者需要慎重对待每一份提交的报告，因为申报项目受理后原则上是不允许更改申报单位和负责人的。

在项目承担上主要体现在三点：1. 针对研究内容要合理划分时间，梳理各研究任务之间的关联和优先级，以确保项目能够在规定时间内结题；2. 牵头单位要与各课题单位定期保持沟通，互相汇报进展，对项目的难点进行联合攻关；3. 科研经费的使用要有理有据，否则尽管项目完成度高，但经费管理若出现纰漏一样会导致无法正常结题。

问题 6：您的科研项目、奖项、高水平论文都非常多，可以说是“天花板”级的水平，请问您能否对科研工作者，尤其是还在“挣扎”的科研工作者提出一些您的建议？

我认为在这个学科交叉的时代，跨学科技能的重要

性日益凸显，首要的就是不要怕跨专业，通过学习不同领域的知识，不仅可以扩充自己的知识面，更有可能再次激发科研热情。

其次，要多与实验室成员交流，针对某个棘手的问题，大家共同探讨甚至于辩论，会促使你对问题产生不同的见解。

再者，不要为琐事烦心，踏踏实实地专注于自己的课题，不与周围的人进行过度比较，真正做到不乱于心，用科学来充实自我，那你定能守得云开见月明。

最后，要把身体素质放在首位，不要为了职业而牺牲自己的健康和幸福。平时要控制工作时间，合理分配时间锻炼身体。

问题 7：您牵头制定了行业标准 2 项，团体标准 1 项，能否跟大家普及一下标准制定方面的入门建议、注意事项？

针对标准制定，想要快速入门，首先要了解行业标准以及团体标准的相关法律，不盲目制定；其次就是要提前学习一些标准的参考模版，了解标准制定的基本步骤；最后就是要掌握标准的编写方法，参照标准的编写模版，就可以看到各类系统分级的标准需要进行哪些准备以及处理哪些素材。

针对标准制定的注意事项，主要包括：1. 科学性，在全国范围内使用的标准，涉及的关键数据一定要有据可循，要有前期科研和调查做基础。此外，要防止简单的材料罗列、堆砌，标准是智慧、科学技术以及经验的结晶，一定要升华、提炼；2. 实用性，标准一定是成熟的，不能滞后也不能过度超前，尤其是一些强制性指标一定要考虑全国的整体情况。此外，标准中应只列入那些被证实的要求，不需要证实的、不宜证实的、或不便证实以及短期内无法证实的要求不应列入标准。

问题 8：您的社会兼职工作也非常多，在繁忙的科研之余，请问您是如何来平衡科研、教学、社会兼职及家庭的呢？您有什么业余爱好？不谈工作，私底下您认为自己是什么样的一个人？

对于如何平衡科研、教学、社会兼职和家庭，我个人觉得这之间并不存在什么对立的关系，而是一个相互支持的关系。例如通过教学能够加深专业知识，对遇到的问题思考启发，很容易发现一些科学研究方向，这对科研工作具有很大的帮助；而社会兼职则是响应了国家的产教研融合，将科研成果用于解决一些实际应用中较棘手的问题，然后再通过现实问题激励我们的科学研究；至于家庭，是辛苦工作后的避风港，在节假日也会抽出时间和家人一起郊游，享受生活的美好之后可以更好的投身到事业中去；对于行政工作，确实会占用很多时间，只有靠自己有计划地分配时间，管理时间，提高工作效率。

虽然现在每天都有做不完的事情，但只要闲下来我还是很乐意去羽毛球馆打一打羽毛球，锻炼锻炼身体。

如果不谈工作，私底下的我可能相对安静一点，喜欢养一些花草，读一些书籍。闲下来的时候也会去实验室与学生们做简单的交流，关注学生的身体和精神状态；在疫情防控期间，我也克服困难加入到志愿者行列，努力让自己成为更有担当的人。

问题 9：如果吐露研究工作者的心声，您最想说的是什么？

合抱之木，生于毫末；九层之台，起于累土；千里之行，始于足下。踏上科研这一条路，就要万事巨细，不放过任何一个疑难杂症，同时还应具备坚强的毅力，持续不懈才有可能在未来有所收获。

责任编委 余烨 赵振兵

周日贵

现为上海海事大学教授，博士生导师，信息工程学院院长。博士毕业于南京航空航天大学，清华大学博士后，加拿大卡尔顿大学（Carleton University）博士后，美国北卡罗来纳州立大学（North Carolina State University）访问学者。“十三五”和“十四五”国家重点研发计划项目负责人（首席科学家），教育部新世纪优秀人才，交通运输部中青年科技创新领军人才，交通运输青年科技英才，入选全球前 2% “顶尖科学家”榜单（2022 年）。上海海事大学“领航计划”入选者，省主要学科学术和技术带头人，省新世纪百千万人才，上海计算机学会计算机视觉专业委员会副主任，上海市人工智能学会理事，中国人工智能学会模式识别专业委员会委员，中国自动化学会模式识别与机器智能专业委员会委员，中国计算机学会理论计算机科学专业委员会和视觉专业委员会委员。主持国家重点研发计划项目和课题各二项、国家自然科学基金四项、教育部新世纪优秀人才资助计划一项、教育部科学技术重点项目一项、中国博士后科学基金一项、上海市“科技创新行动计划”二项、中国（上海）自由贸易试验区临港新片区管理委员会项目一项，省高等学校科技落地计划项目一项，省级科研项目二十多项；近年来在 IEEE TIP、IEEE TNNLS、IEEE TFS、IEEE TETC、IS、QIP、PRA、《电子学报》等国际期刊、国内一级学报及行业顶级会议上已发表第一及通信作者学术论文 165 篇，其中 SCI 收录 111 篇，二区及以上论文 42 篇，SCI 他引 1500 余次；获上海市科技进步二等奖一项（2020），上海市浦东新区科技进步二等奖一项（2020），江西省自然科学奖二项（2019,2014），第八届吴文俊人工智能科学技术进步奖，第六届吴文俊人工智能科学技术创新奖，南昌市科技进步奖一项，获第二十二届中国国际工业博览会高校展区“优秀展品奖”（2020），出版学术专著四部，其中三部全英文学术专著；获国家授权发明专利 5 项和软件著作权 6 项；牵头制定行业标准 2 项，团体标准 1 项；担任第二届智能科学国际会议（ICIS2017）大会执行主席（Organization Chair）、第一届量子世界大会（CQW2017）主题论坛主席和分论坛主席，是 IEEE Transactions On TIP 和 IEEE Transactions On VLSIS 等国际期刊的审稿人。

委员好消息

✪ 2022 年 11 月 23 日，四川省科技厅公示了 2022 年度四川省科学技术奖拟奖项目，CCF-CV 专委会执行委员四川大学**彭玺**、**胡鹏**等完成的“基于结构不变性的表示学习理论和方法”拟授自然科学奖二等奖，CCF-CV 专委会执行委员电子科技大学**李文**参与完成的“面向工业制造的缺陷智能检测与分类关键技术及应用”拟授科技进步奖三等奖。

✪ 2022 年 11 月 29 日，2022 年度教育部-华为“智能基座”优秀教学资源遴选结果发布，CCF-CV 专委会执行委员、北京大学**刘家瑛**主编的《计算机视觉理论与实践》和 CCF-CV 专委会执行委员、重庆邮电大学**高新波**、西安电子科技大学**王楠楠**合著的《人脸图像合成与识别》获 2022 年度教育部-华为“智能基座”优秀教材获奖。

✪ 2022 年 12 月 5 日，北京市自然科学基金委员会办公室发布了首批北京市自然科学基金奖励项目拟资助项目公告，CCF-CV 专委会执行委员、北京航空航天大学**史振威**主持的“多时相遥感影像语义变化检测方法研究”项目入选。

✪ 2022 年 12 月 26 日，PRCV 2022 举办颁奖典礼，CCF-CV 专委会执行委员、北京航空航天大学**王蕴红**等完成的 JoinTW: A Joint Image-to-Image Translation and Watermarking Method 获最佳论文奖，CCF-CV 专委会执行委员、江南大学**宋晓宁**等完成的 KITPose: Keypoint-Interactive Transformer for Animal Pose Estimation 获最佳学生论文奖，CCF-CV 专委会执行委员、联想研究院**师忠超**等完成的 Semi-supervised Medical Image Segmentation with Semantic Distance Distribution Consistency Learning 获最佳论文提名奖，CCF-CV 专委会执行委员、北京科技大学**殷绪成**等完成的 Anchor-Free Location Refinement Network for Small License Plate Detection 获最佳学生论文提名奖，CCF-CV 专委会执行委员、大连理工大学**刘日升**等完成的 Video Deraining via Temporal Discrepancy Learning 获最佳学术海报奖。

✪ 2022 年 12 月 29 日，中国科学技术信息研究所线上举办本年度“中国科技论文统计报告发布会”，报告公布了本年度百篇最具影响国际学术论文和中国百篇最具影响国内学术论文，CCF-CV 专委会常务委员、南开大学**程明明**等的论文 Res2Net: A New Multi-Scale Backbone Architecture 和 CCF-CV 专委会执行委员、中山大学**郭裕兰**、国防科技大学**刘丽**等的论文 Deep Learning for 3D Point Clouds: A Survey 获本年度百篇最具影响国际学术论文。

✪ 2022 年 12 月 29 日，陕西省教育厅公示了第十三届陕西本科高等学校教学名师奖拟获奖教师名单，CCF-CV 专委会执行委员、西安电子科技大学**邓成**和**苗启广**入选。

✪ 2023 年 1 月 4 日，2022 年 CCF 优秀博士学位论文激励计划评选结果发布，共 10 篇论文入选 2022 年“CCF 优秀博士学位论文激励计划”、4 篇论文获得 2022 年“CCF 优秀博士学位论文激励计划”提名，CCF-CV 专委会执行委员、中国科学院大学**黄庆明**指导中科院信工所杨智勇完成的《面向复杂场景的 AUC 优化理论、方法及应用》入选 2022 年“CCF 优秀博士学位论文激励计划”。

✪ 2023 年 1 月 6 日，2022 年度江苏省科学技术奖

综合评审结果公示，CCF-CV 专委会 3 位执行委员参与完成的项目入围：南京理工大学**肖亮**等完成的“高分辨率光谱智能感知与解译系统关键技术及应用”、江南大学**李朝锋**等完成的“图像质量评价与提升关键技术”拟授二等奖，中国科学院空天信息创新研究院**孙显**等完成的“城市重大基础设施星载 SAR 多维信息感知关键技术与应用”拟授三等奖。

✪ 2023 年 1 月 11 日，中国计算机学会公布了 2022 年“CCF 卓越服务奖”评奖结果公告，CCF-CV 专委会执行委员、北京师范大学**黄华**入选。黄华教授长期服务于 CCF，先后担任 CCF 理事、副秘书长、YOCSEF 主席、学术工作委员会副主任、青年工作委员会主任等职，对学会的发展做出了重要贡献。此外，他在创立青年科学家奖、青竹奖及思想秀等活动中也发挥了关键作用。

✪ 2023 年 1 月 14 日，湖北省科技厅公示了 2022 年度湖北省科学技术奖拟奖项目，CCF-CV 专委会执行委员华中科技大学**尤新革**、武汉大学**荆晓远**等完成的“面向语义理解的复杂图像表征与计算理论及方法研究”拟授自然科学一等奖，CCF-CV 专委会执行委员、武汉大学**涂志刚**等完成的“开放环境视频人体动作识别实用性理论与方法研究”拟授自然科学二等奖。

✪ 2023 年 1 月 17 日，2022 中国电子学会科学技术奖拟授奖项目公示，CCF-CV 专委会 6 位执行委员的项目或所在团队入围：中国科学院计算技术研究所**陈熙霖**作为成员的“北京大学视频编解码技术创新团队”拟授创新团队奖，厦门大学**纪荣嵘**参与的“视觉信息复杂关联计算”和南京理工大学**唐金辉**、**李泽超**、**舒祥波**、中国科学院自动化研究所**刘静**等完成的“语义关联驱动的多媒体大数据智能感知理论与方法”拟授自然科学一等奖。

✪ 2023 年 1 月 17 日，CCF-CV 专委会 3 位执行委员、中国科学院自动化研究所**谭铁牛**和**王亮**、重庆邮电大学**高新波**当选中国人民政治协商会议第十四届全国委员会委员。

✪ 2023 年 1 月 19 日，中国图象图形学学会公布了 CSIG 新晋杰出会员名单，CCF-CV 专委会 9 位执行委员晋升为 CSIG 杰出会员：中山大学**操晓春**、天津理工大学**陈胜勇**、南开大学**程明明**、中国科学院自动化研究所**何晖光**、中国科学院自动化研究**赫然**、南京信息工程大学**刘青山**、华中科技大学**桑农**、北京航空航天大学**史振威**、东南大学**郑文明**。

✪ 2023 年 2 月 7 日，Machine Intelligence Research 评出 MIR 2022 年度优秀编委 5 名，CCF-CV 专委会常务委员、南开大学**程明明**入选。

✪ 2023 年 3 月 2 日，2022 年度吴文俊人工智能科学技术奖拟授奖项目公示，CCF-CV 专委会 10 位执行委员上榜：西北工业大学**王琦**等完成的视觉影像智能分析理解的理论与方法、中科院自动化所**兴军亮**和军事科学研究院**赵健**等完成的无约束人像目标智能感知与理解拟授自然科学一等奖，北京科技大学**殷绪成**参与完成的基于规则与机器学习的 EDA 布局布线新技术拟授技术发明一等奖，中国科学院合肥物质科学研究院**汪增福**参与完成的多语种复杂场景图文识别关键技术及产业化拟授科技进步一等奖，中国科学院大学**黄庆明**参与完成的复杂互联网环境下内容治理的关键技术与应用拟授科技进步二等奖，苏黎世联邦理工学院**范登平**和南京理工大学**魏秀参**拟授优秀青年奖，上海交通大学**卢策吾**指导完成的《知识驱动的行为理解》、东南大学**郑文明**指导完成的《基于图神经网络的情感识别研究》拟授优秀博士学位论文奖。

✪ 2023 年 3 月 08 日，百度发布首份 AI 华人女性青年学者榜，CCF-CV 专委会执行委员、北京航空航天大学**刘偲**、上海科技大学**汪婧雅**、电子科技大学**姬艳丽**入选。

责任编委 刘海波

基于 Diffusion 的图像生成开源代码

大连理工大学 付陈平 樊鑫

扩散模型（Diffusion Models）已经成为最新的深度生成模型家族。他们打破了生成对抗网络（GANs）在图像合成任务中的长期主导地位，被广泛应用于计算机视觉任务当中。本文将以时间发展为依据，串讲扩散模型的主要论文与代码，包括：DDPM、DDIM、CFDG、Palette 和 DALLE 2。

1、DDPM 模型

介绍：DDPM 是扩散模型的经典模型之一， 其给出了严谨的数学推导过程以及可复现的代码。DDPM 提出的“前向加噪-反向降噪”的训练模式被后来的工作延续继承。如图 1 所示，扩散模型包括两个过程：前向扩散过程和反向生成过程，前向扩散过程是对一张图像逐渐添加高斯噪音直至变成随机噪音，而反向生成过程是去噪音过程，将从一个随机噪音开始逐渐去噪音直至生成一张图像，这是模型要求解或者训练的部分。图 2 展示了 DDPM 模型逐渐生成过程，图 3 展示了 DDPM 的生成效果图。

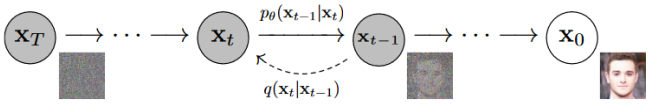


图 1 DDPM 模型框架图

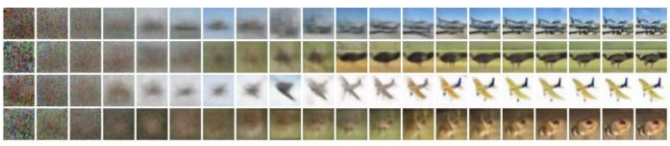


图 2 DDPM 图像逐渐生成过程

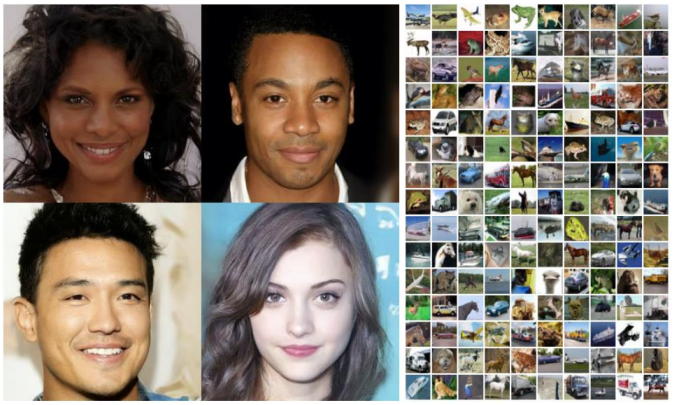


图 3 DDPM 图像生成效果

论文地址：https://arxiv.org/pdf/2006.11239.pdf
代码地址：
https://github.com/hojonathanho/diffusion

2、DDIM 模型

介绍： DDPM 在不进行对抗性训练的情况下可实现高质量的图像生成，然而该模型的生成过程需要模拟马尔可夫链采样过程，导致生成效率低下。针对这一问题，文章提出了去噪扩散隐式模型(DDIM)模型，该模型是一种更有效的迭代隐式概率模型。在 DDPM 中，生成过程被定义为特定马尔可夫扩散过程的反向过程。如图 4 所示，文章通过一类导致相同训练目标的非马尔可夫扩散过程来概括 DDPM。这些非马尔可夫过程可以对应于确定性的生成过程，从而产生高质量样本的隐式模型,提高生成效率。实验表明，与 DDPM 相比，DDIM 可以以 10 倍到 50 倍的速度生成高质量的样本，允许研究人员权衡计算和样本质量，直接在

潜在空间中执行语义上有意义的图像插值，并以非常低的代价重建观测。图 5 展示了 DDIM 在不同生成次数下的生成效果。

图 4 扩散(左)和非马尔可夫(右)推断的图形模型

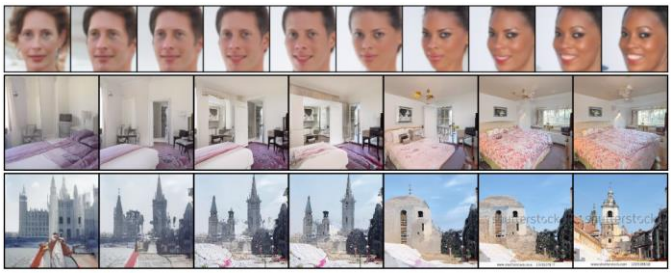


图 5 DDIM 不同生成次数的生成效果

论文地址： https://arxiv.org/pdf/2010.02502.pdf
论文代码： https://github.com/ermongroup/ddim

3、CFDG 模型

介绍： 分类器引导是最近引入的一种方法，用于在训练后的条件扩散模型中权衡模式覆盖和样本保真度，与其他类型生成模型中的低温采样或截断具有相同的出发点。分类器指导将扩散模型的得分估计与图像分类器的梯度相结合，因此需要训练与扩散模型分开的图像分类器。分类器指导方式面临一个严重的问题，即在没有分类器的情况下是否可以执行指导。文章表明，在没有分类器的情况下，指导可以由纯生成模型执行：在文章所谓的无分类器指导中，研究人员联合训练了一个有条件和无条件的扩散模型，并将得到的有条件和无条件的评分估计值结合起来，以获得样本质量和多样性之间的权衡，从而获得类似于使用分类器指导获得的结果。图 6 展示了所提无分类指导模型的生成结果图。

图 6 所提无分类指导模型生成效果

论文地址： https://arxiv.org/pdf/2207.12598.pdf
论文代码： https://github.com/lucidrains/classifier-free-guidance-pytorch

4、Palette 模型

介绍： 文章提出了一个基于条件扩散模型的图像到图像的通用转换框架（即 Palette），在四种具有挑战性的图像转换任务(即着色、修补、裁剪和 JPEG 恢复)上对所提框架进行了评估。Palette 在所有任务上不需要特定于任务的超参数调优、结构设计、任何辅助损失或其他复杂新技术就可以很好实现。此外，文章还揭示了去噪扩散目标中 L_2 和 L_1 损失对样本多样性的影响，并通过实证研究证明了自我注意在神经结构中的重要性。大量实验表明，Palette 通用模型比特定任务的专家模型表现更好。图 7 是 Palette 模型在 4 种代表性图像到图像的任务结果展示。

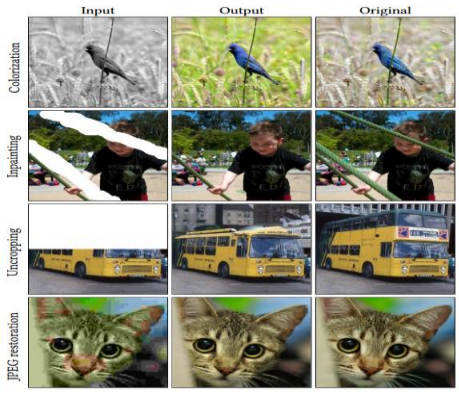


图 7 Palette 在 4 种任务的完成效果

论文地址： https://arxiv.org/pdf/2111.05826.pdf
论文代码： https://github.com/lucidrains/classifier-free-guidance-pytorch

5、DALLE 2 模型

介绍： 如图 8 所示，CLIP 可捕捉文字语义和图像风格进而生成新图像的鲁棒表示。为了利用这些表示进行图像生成，文章提出一个两阶段模型：给定文本标题生成剪辑图像嵌入的先验模型，以及以图像嵌入为条件生成图像的解码器。显式地生成图像表示提高了图像多样性，在逼真度和标题相似性方面损失最小。以图像表示为条件的解码器还可以产生图像的变体，保留其语义和风格，同时改变图像表示中缺少的非必要细节。此外，DALLE 2 的联合嵌入空间能够以零样本的方式进行语言引导的图像操作。文章将扩散模型用于解码器，并对先验的自回归模型和扩散模型进行了实验，发现后者在计算上更有效，并产生更高质量的图像例子。图 9 为 DALLE 2 生成效果展示。

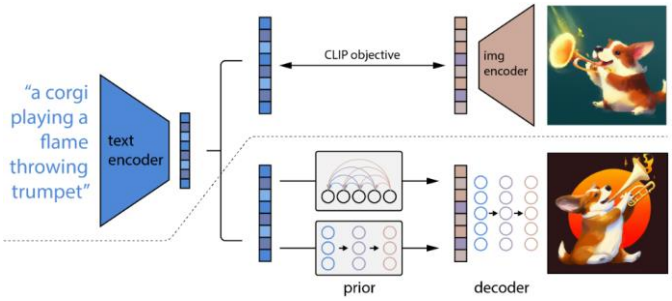


图 8 DALLE 2 流程图

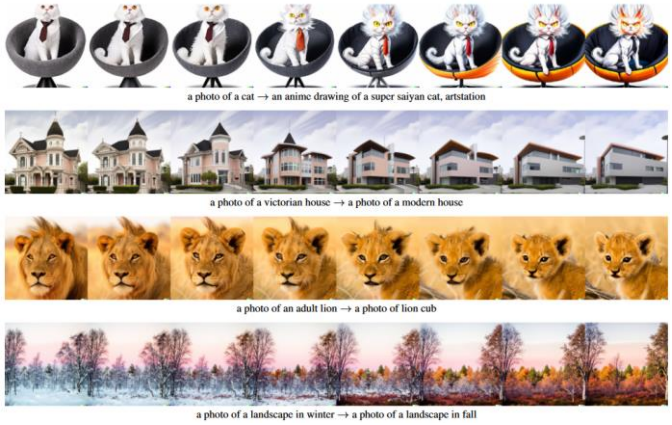


图 9 DALLE 2 生成结果展示

论文地址： https://arxiv.org/pdf/2204.06125.pdf
论文代码： https://github.com/lucidrains/DALLE2-pytorch

责任编委 贾同 沈沛意

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

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

4D 物体感知数据集

中国科学院自动化研究所 王宇琪 张兆翔

物体感知一直是计算机视觉领域的核心问题之一。近年来，随着标注数据的丰富以及对网络模型设计的探索，物体感知不断向更高维度和更深层次的方向发展，也从2D 的图像感知发展到 3D 的空间感知，再到 4D 时空一体的感知。如图 1 所示，最初，物体感知的研究主要关注于 2D 图像感知，侧重于像素级别的物体定位与语义理解。随着技术的不断发展，基于雷达或相机的 3D 空间感知也发展迅速，重点在于对空间的建模和理解，实现更精确和全面的物体感知。而最新的 4D 感知则通过挖掘物体的时序信息，实现对物体时空一体的感知能力，从而在更多应用场景中实现更精准的感知与理解。

本文重点介绍可用于 4D 物体感知的数据集，包括 WOD、nuScenes、KITTI 等代表性的数据集。

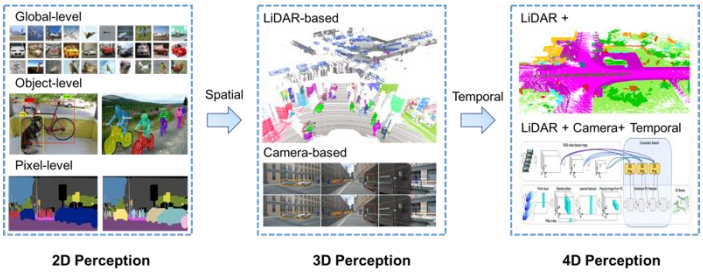


图 1 从 2D 感知到 4D 感知

1、WOD 数据集

介绍： WOD（Waymo Open Dataset）是谷歌旗下的自动驾驶公司Waymo发布的一个包含多种传感器数据的自动驾驶场景数据集。这也是目前规模最大、场景最多样化的多模态自动驾驶数据集。数据采集范围涵 盖凤凰城、柯克兰、山景城、旧金山等地区，以及各种驾驶条件下的数据，包括白天、黑夜、黎明、黄昏、雨天和晴天。数据集包含 1150 个驾驶片段，每一个片段包含 20 秒的连续画面，采样频率是 10Hz，其中拥有高质量且同步校准的激光雷达和相机数据，并提供了精细的标注:包括 2D 图像和 3D 点云的物体边界框。其中 1000 个驾驶片段用于训练集（798）和验证集（202），剩余 150 个片段用做测试集。

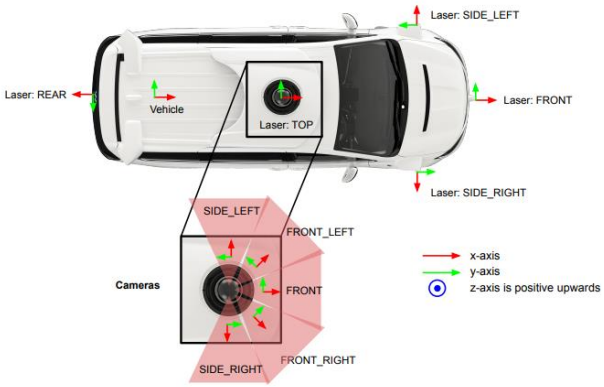


图 2 Waymo 数据集的传感器分布

2D 相机数据：如图 2 所示，数据集包含 5 个摄像头的数据：前向、左前、右前、左侧和右侧。其中 3 个前向相机采集的图像分辨率是 1920x1280，而侧向相机采集的图像分辨率是 1920x1040。数据集提供了约 9M 的 2D 物体目标框标注，254k 个物体追踪 ID。标注类别包含车辆、行人和骑自行车的人。

3D 雷达数据：数据集包含 5 个激光雷达传感器的数据：主雷达在车顶（Top），同时配备 4 个辅助雷达，前 (Front)，后 (Rear)，左侧 (Side-Left)，右侧 (Side-Right)。其中主雷达的感知范围是 75 米，辅助雷达的感知范围是 20 米，都提供了激光脉冲的两次回波数据。平均每一帧点云的点数量在 177k。数据集提供了约

12M 的 3D 目标检测框标注，113k 个物体追踪 ID，每个 3D 标注框拥有 7 维信息，即物体中心 xyz 坐标，长宽高 lwh 以及朝向角θ。标注类别包含车辆、行人、骑自行车的人以及标志牌。

时序运动数据： WOD 数据集既提供了连续的 2D 视频序列以及 3D 点云序列，可用作时序信息的探索，同时还提供了场景流的数据标注。场景流信息丰富了场景中物体的运动信息，即对每一帧点云的点上标注了 xyz 方向的速度信息。

利用 4D 数据的丰富特性，不仅可以提升物体感知的准确性和全面性，也可以探索无监督的物体发现，从而减少昂贵的人工标注。

数据集下载地址： https://waymo.com/open

相关论文链接：
https://arxiv.org/pdf/2210.04801.pdf

2、nuScenes 数据集

介绍： nuScenes 数据集是一个大规模自动驾驶数据集，由 nuTonomy 创建。如图 3 所示，该数据集以波士顿和新加坡地区的城市街景为基础，采集了 1000 个全天候开放场景的数据，每段序列大约为 20 秒，包含 40 个关键帧数据。其中 700 个序列作为训练集，150 个作为验证集，150 个作为测试集。该数据集拥有完整的自动驾驶车辆传感器套件，包括 6 个摄像头、5 个雷达和 1 个激光雷达。其中相机、雷达和激光雷达的数据同步校准，成为多模态物体感知算法研究的重要基准数据集。

3D 激光雷达:数据集包含 1 个 32 线的激光雷达，20Hz 的采样频率。数据集提供了 2Hz 频率的关键帧标注（40k），标注类型包含语义类别（23 类）、属性（可见性、活动和姿态）以及 3D 目标检测框。其中 3D 目标检测框包含物体中心 xyz，以及物体长宽高 lwh 和朝向角θ的信息。数据集提供了约 1.4M 的 3D 目标框标注。

2D 相机数据：数据集包含了 6 个相机数据，分布在前、左前、右前、后、左后和右后，每个相机视角的图像分辨率都是 1600x900，覆盖了 360 度的感知视野。由于具有和激光雷达相同的 360 度感知视野，该数据集也成为目前基于多目相机 3D 感知的主要基准数据集。

雷达数据：相比于激光雷达，雷达拥有更远的感知范围（200-300m）以及对速度的感知能力。该数据集包含了 5 个雷达传感器的数据，采集频率是 13Hz，拥有约 1.3M 的雷达点云数据。雷达包含的物体速度也为 4D 感知提供了更丰富的感知信息。

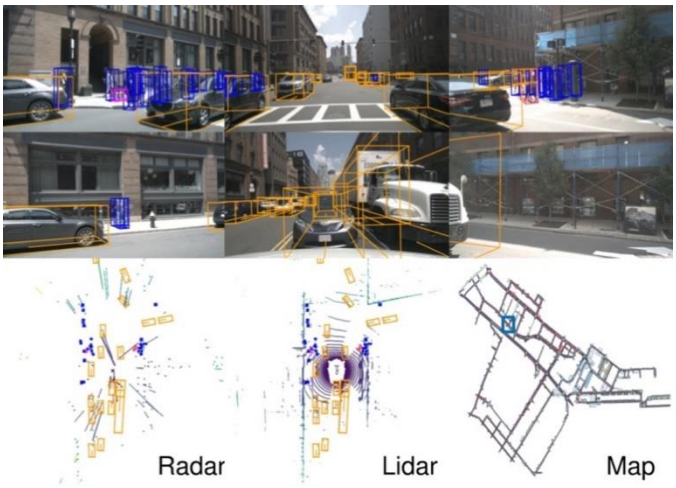


图 3 nuScenes 数据集示例

数据集下载地址： https://nuscenes.org

算法评测平台： https://eval.ai

3、KITTI 数据集

介绍： KITTI 数据集由卡尔斯鲁厄理工学院和丰田技术中心联合发布。它在 2012 年首次公开发布，是最早的致力于自动驾驶算法研究的数据集，并拥有各个任务的评测基准：立体视觉、光流、深度、视觉里程计、3D 检测、3D 追踪、2D 追踪等。该数据集包含了 22 个不同城市和高速公路上的真实场景，提供了由激光雷达和相机获取的 15k 张图片数据和 15k 点云数据。

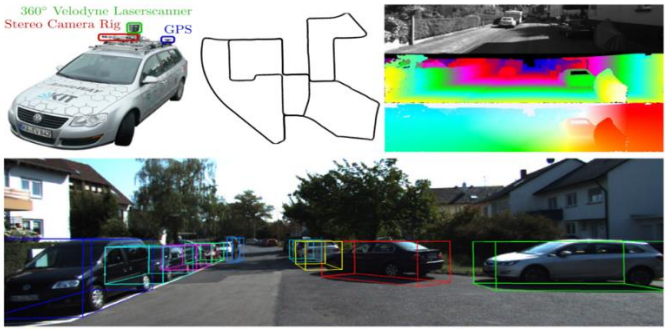


图 4 KITTI 数据集示例

3D 激光雷达：如图 4 所示，数据集包含一个 64 线 HDL-64E 激光雷达，采样频率为 10Hz，范围达到 100 米。数据集提供了约 15k 个标注数据，大约 200k 个 3D 目标检测边界框的标注。目标框的注释类别包括车辆、卡车、货车、电车、行人、坐在车上的人、骑自行车的人和其他，而评估通常只关注车辆、行人和骑自行车的人这三类。此外，每个目标框还具有可见性属性，分为可见、半遮挡、全遮挡或被截断四类。

2D 相机数据：数据集包含两个彩色和两个灰度的视频摄像头，都分布在前向，用于单目和双目视觉的算法研究。图像分辨率是 1392x512。双目相机的基线是 54cm。该数据集也作为目前单目 3D 检测的基准数据集。

稠密语义标注： SemanticKITTI 是基于 KITTI 提供的视觉里程计数据所扩展标注的数据集，其目的是更全面地理解 3D 场景中的语义信息。该数据集包含一对前向的激光雷达和相机采集的原始数据，以及相应的时序同步标注数据，拥有 23021/20351 帧用于训练/测试的点云数据，对每个点提供了 28 个标签类别，包括路面、建筑物、交通标志、汽车、行人等。如图 5 所示，由于拥有连续的场景信息和稠密的语义信息，该数据集也可以用于 4D 的场景理解任务，如 4D 全景分割。随着基于相机的 3D 场景感知技术的不断进步，SemanticKITTI 数据集已经成为评估和探索 Occupancy（3D 占用）模型的基准数据集之一。由于该数据集具有丰富的场景信息和大量的语义标注数据，因此可以用于深入探索和研究 3D 场景的位置占用情况。

物体运动信息：KITTI 数据集提供了 400 帧不同动态场景的 2D 光流和 3D 场景流的标注，用于研究场景中运动物体的感知。2D 光流记录了物体在图像中的运动信息，而 3D 场景流则体现了物体在 3D 空间里的运动信息。这是第一个基于真实环境的场景流数据集，其中 200 帧用于训练，200 帧作为测试。4D 物体感知也更加注重于物体运动信息的探索和挖掘，该数据集为 4D 的物体感知提供了丰富的基础。

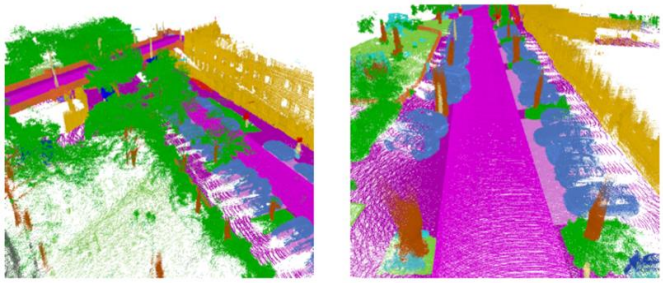


图 5 稠密语义标注示例

数据集下载地址：

https://www.cvlibs.net/datasets/kitti

http://www.semantic-kitti.org

相关论文链接：

https://www.cvlibs.net/publications/Geiger2012CVPR.pdf

https://arxiv.org/abs/1904.01416

责任编委 沈沛意 李策

王宇琪

博士研究生，中国科学院自动化研究所模式识别国家重点实验室与智能感知与计算研究中心。研究方向为计算机视觉与深度学习。

张兆翔

教授，博士生导师，中国科学院自动化研究所模式识别国家重点实验室与智能感知与计算研究中心研究员。研究方向包括模式识别、计算机视觉与深度学习。

个人主页：https://people.ucas.ac.cn/~zhangzhaoxiang

好 文 推 荐

厦门大学团队关于大规模数据的多视角聚类研究发表于 IEEE TNNLS 2022。

论文：Zhang Y, Yuan X, Li C, Wu Z and Qu Y, Learning all-in collaborative multiview binary representation for clustering, IEEE Transactions on Neural Networks and Learning Systems, 2022, 1-14.

越来越多的现实世界数据是从不同的来源收集的或通过不同特征提取器获得，数据量也非常庞大。这对无监督数据分析也提出更高的需求：高性能和高效率。基于二进制表示的多视图聚类通常忽略二值表示学习中非常重要的潜在高阶相关性。作者提出 AC-MVBC，一个视角内和视角间协作多视图二值表示聚类框架，其中多视图协作二进制表示和聚类结构以联合方式学习。所提算法的整体框架如图 1 所示。

所提方法的主要贡献如下：

1) 提出了一个新的多视角二值聚类模型，通过建模跨视图和视图内的协作，以联合学习的方式来探索更好的二进制表示和聚类分配。

2) 多视角协作建模是在张量学习的框架下构建，其中，引入了一种新型的低秩张量约束和 Bregman 散度，保证优化学习方向。

3) 为提出的方法制定优化算法，以有效且快速的求解提出的目标函数。

在四个挑战性的大规模数据集上的实验结果表明，与先进的多视图聚类方法相比，所提出的方法在保证较低计算复杂度和内存需求的同时，取得更高的聚类性能。

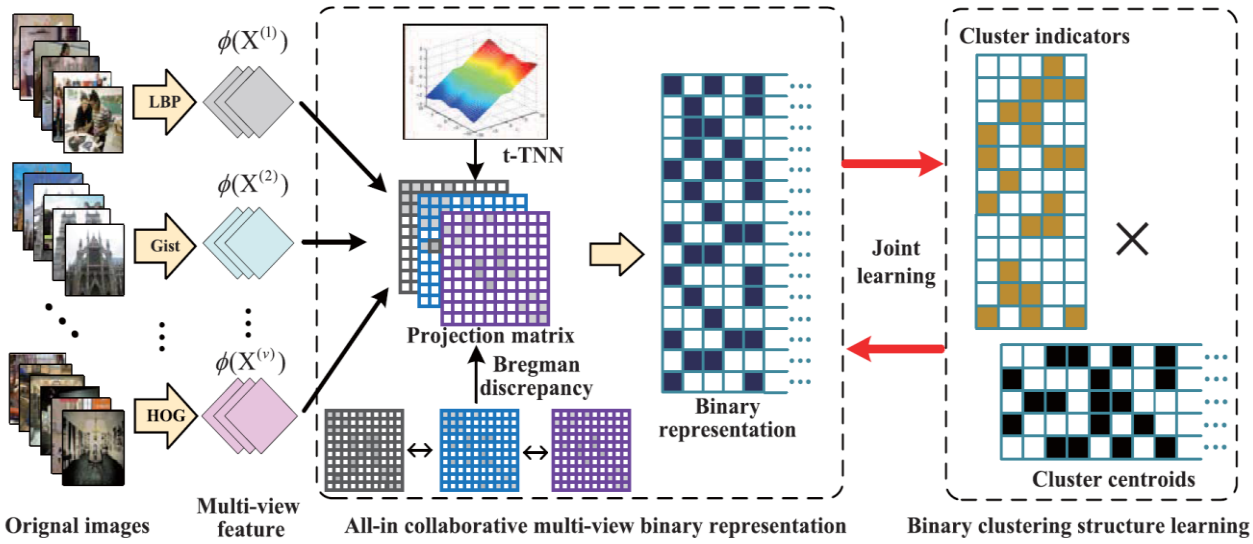


图 1. 所提算法框架。大规模图像被建模为多视角特征，然后在 t-TNN 和 Bregman 散度共同约束下挖掘不同视图的高阶一致信息，并抑制多视图中存在的噪声。对于单一视角，这种低秩约束可以促进建模相似样本的特征相似性。此外，联合学习二进制表示和聚类分配可以使得二值表示学习更具有针对性，从而获得更好的聚类性能。

责任编委 樊鑫 贾同

好 文 推 荐

清华大学、北京航空航天大学、北京交通大学和海军研究所团队“Detecting prohibited objects with physical size constraint from cluttered X-ray baggage images”最新成果发表在 Knowledge-Based Systems 2022。

论文：Chang A，Zhang Y，Zhang S，et al. Detecting prohibited objects with physical size constraint from cluttered X-ray baggage images. Knowledge-based systems,2022:237.

X 射线图像检查旨在检测行李、包裹中的违禁物品，是维护公共场所（如机场、火车站、地铁站）安全使用的最广泛的手段。为了分担安检人员人工检查 X 线图像的重复性工作，解决漏检率高、检测效率低等问题，目前许多研究人员致力于使用计算机视觉技术实现快速、准确、自动的 X 线包裹检查。目前提出的基于深度学习的违禁品检测方法致力于解决 X 射线安检图像通常存在的重叠、遮挡、类内差异、类别不平衡等问题，但它们往往忽略了违禁品的实际物理尺寸，这会导致许多检测错误的情况。

为了解决这个问题，文章提出了一个两阶段 X 射线图像违禁品检测网络，以从背景严重杂乱的 X 射线包裹图像中识别违禁品，网络总体结构如图 1 所示。首先为考虑 X 射线图像中违禁品的物理尺寸，显著减少由形状相似但物理尺寸不同引起的错误，所提出的网络在训练过程中将违禁品的物理尺寸约束公式化为正则化项。其次，由于当前 X 射线数据集仅提供违禁品（正样本）的注释而忽略了非违禁品 (负样本)，负样本的数量不足可能导致测试期间出现许多误检，因此该文还提出了一种难负样本选择方案，以从分割的前景区域中生成非违禁品的候选框，利用前景区域的难负样本提高检测器性能。总体而言，FPN 主干、基线 Faster R-CNN、难负性样本选择方案和物理尺寸约束共同构成了文章提出的针对 X 射线包裹图像的违禁品检测方法。

文章在 SIXray 和 OPIXray 数据集上进行了大量实验，实验结果表明在从杂乱的 X 射线包裹图像中识别违禁品方面，本文提出的方法优于目前最先进的目标检测方法，证明了所提物理尺寸约束和难负样本选择方案的有效性。

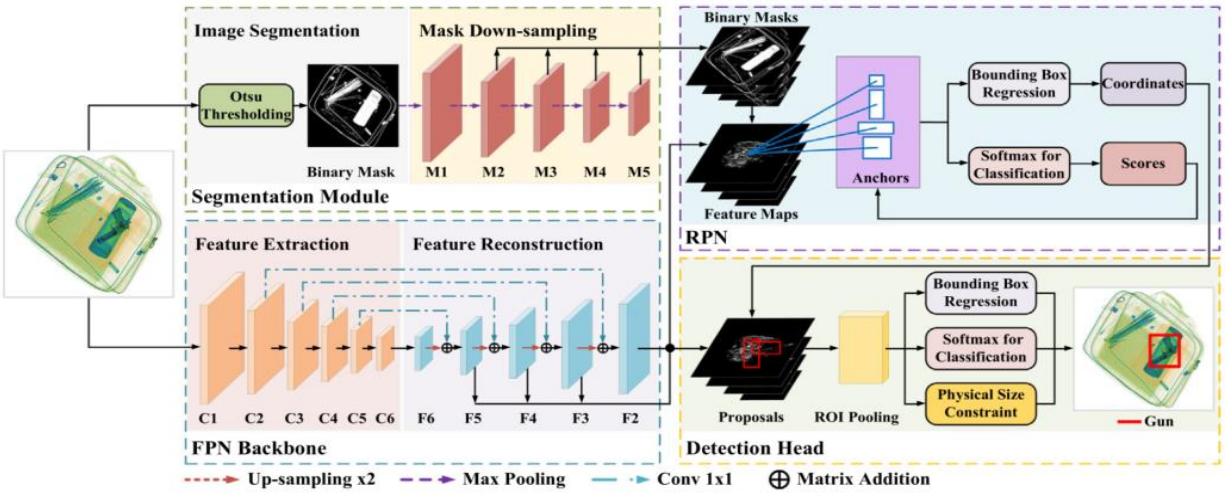


图 1 所提出的违禁品检测网络结构图

责任编委 贾同 李策

好 文 推 荐

南加州大学和 Adobe Research “Point-NeRF: Point-based neural radiance fields” 的最新成果发表在 CVPR-2022。

论文：Xu Q, Xu Z, Philip J, et al. Point-nerf: Point-based neural radiance fields. The IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 5438-5448.

新视角合成 (Novel View Synthesis) 主要目的是从图像数据中建模真实场景并渲染逼真的新视图，通过给定源图像、源姿态以及目标姿态，渲染生成目标姿态对应的的图片。新视角合成在 3D 重建、AR/VR 等领域有着广泛的应用。NeRF 系列的论文在这一方面获得了很好的效果，这些方法通常使用 MLP 网络通过射线进行重建整个空间的辐射场，但这个过程非常消耗时间，对很多大片的空旷区域进行了不必要的采样。因此，文章提出了一种高质量场景重建和渲染的新方法 Point-NeRF，通过使用其他方法获得初始的点云来指导 NeRF。

文章提出的网络架构一共分为两个部分，神经点生成模块和基于点云的体渲染模块。与传统 NeRF 模型不同，文章结合了 NeRF 利用体渲染合成高质量视图以及 MVS 快速重建场景这两种方法的优点，通过使用神经 3D 点云和相关的神经特征来模拟辐射场。首先利用 MVSNet 方法快速得到一个初始点云，然后利用点云和神经特征来构建一个基于 Point 的辐射场，在基于 Point 的辐射场上再对点云进行进一步的渲染。由于初始点云通常会包含降低渲染质量的空缺值和异常值，直接优化现有点的位置会使训练不稳定。为了解决这个问题，文章提出了点剪枝和生长技术，利用点置信度来修剪不必要的离群值，以提高几何建模和渲染质量。Point-NeRF 架构如图 1 所示。

文章在 DTU、NeRF Synthetic、ScanNet 和 Tanks & Temples 数据集上的实验表明，Point-NeRF 可以超越现有的方法和实现最先进的结果。并且 Point-NeRF 在几十分钟内可以达到超过 NeRF 的重建质量，其重建的时间要比 NeRF 提高了 30 倍。

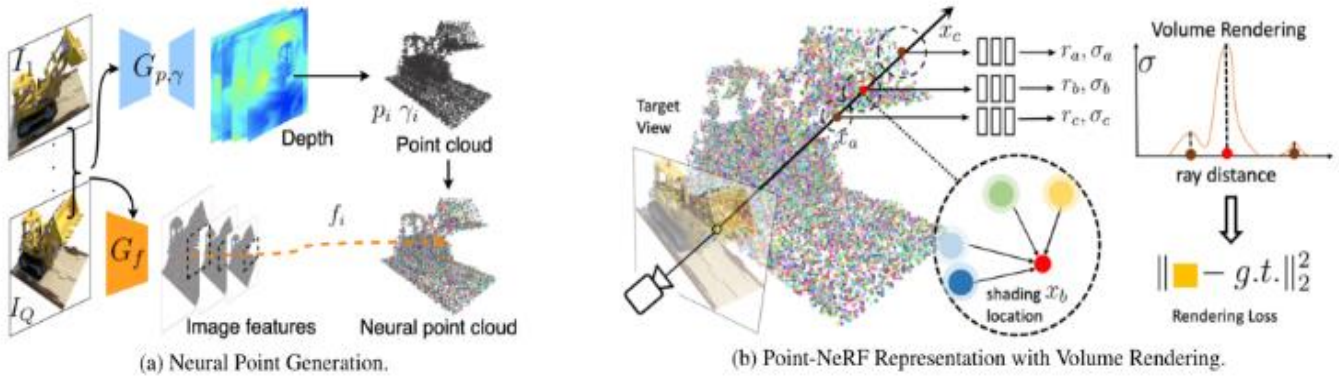


图 1 Point-NeRF 架构

责任编委 樊鑫 沈沛意

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 ACM Transactions on Multimedia Computing, Communications, and Applications, IEEE Journal of Biomedical and Health Informatics 和 Pattern Recognition Letters。

3 会议简介

中国模式识别与计算机视觉学术会议 PRCV (Chinese Conference on Pattern Recognition and Computer Vision)，由中国计算机学会（CCF）、中国自动化学会（CAA）、中国图象图形学学会（CSIG）和中国人工智能学会（CAAI）联合主办，定位国内顶级的模式识别和计算机视觉领域学术盛会。

第六届 PRCV 将于 2023 年 10 月 13 日至 10 月 15 日在厦门举办，由厦门大学承办。会议旨在汇聚国内国外模式识别和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行，共同分享我国模式识别与计算机视觉领域的最新理论和技术成果。通过此次会议，进一步加强本领域的同行与东南沿海地区的学者和企业进行学术交流和技术碰撞，从而促进模式识别与计算机视觉领域的协同合作与融合创新。

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
ACM MM 2023	2023.9.29-11.3	Ottawa, Canada	2023.05.01	https://www.acmmm2023.org/
NeurIPS 2023	2023.12.10-16	New Orleans, USA	2023.05.18	https://neurips.cc/Conferences/2023/
BMVC 2023	2023.11.20-24	Aberdeen, UK	2023.05.23	https://www.bmvc2023.org
PRCV 2023	2023.10.13-15	厦门，中国	2023.06.20	https://prcv2023.xmu.edu.cn/

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
TOMM	Integrity of Multimedia and Multimodal Data in Internet of Things	https://dl.acm.org/pb-assets/static_journal_pages/tomm/pdf/CfP-TOMM-SI-Integrity-Multimedia-Multimodal-Data-IoT-1672776274950.pdf	2023.05.15
JBHI	Real world data processing in real time for smart healthcare	https://www.embs.org/jbhi/wp-content/uploads/sites/18/2023/03/Updated-CFP-for-Real-world-data-processing-in-real-time-for-smart-healthcare.pdf	2022.06.27
JBHI	Machine Learning Technologies for Biomedical Signal Processing	https://www.embs.org/jbhi/wp-content/uploads/sites/18/2023/03/Updated-CFP-for-Real-world-data-processing-in-real-time-for-smart-healthcare.pdf	2023.06.15
PRL	Advances in Disinformation Detection and Media Forensics (A2DMF)	https://www.journals.elsevier.com/pattern-recognition-letters/call-for-papers/advances-in-disinformation-detection-and-media-forensics	2023.06.20

心底无私视界宽–丁晓青教授专访

自 50 年代以来，我国在计算机视觉领域展开了相关的科研工作。而今，我国已经拥有了一支庞大的、在该领域辛勤耕耘且能与世界一流水平并驾齐驱的科研队伍。在这一过程中，有一批见证了视觉领域发展、为我国计算机视觉领域的奠基做出了重大贡献的先驱者。

至今，本栏目已经采访了 9 位计算机视觉领域的资深教授，本次采访的是清华大学丁晓青教授，也是《视界专访》专栏创建以来第一位女性教授。丁老师获得 2022 中国计算机学会计算机视觉专委会（CCF-CV）终身学术贡献奖，作为计算机视觉领域著名学者之一，她在汉字识别、人脸识别等方面取得了一系列国际领先性成果，突破了诸多技术转化的重大瓶颈问题。通过丁老师的专访，不仅从她的个人求学工作经历、科研和教学历程中，让领域研究者更为深切地了解文字识别和生物特征识别领域在中国的发展历程，汲取其丰富科研教学经验，更感受到女性科学家在推进计算机视觉领域发展、解决国家急需解决的重要问题方面做出的巨大贡献。

图 1 丁晓青教授

我是负责本次专访的主要采访人，北京工业大学贾熹滨。本次采访通过微信交流完成，相关问题由 CCF-CV 专委会的《视界专访》组提供。为能更好地帮助我们回顾本次采访，我们采用了问答加书面回顾的形式来表述。以下是丁晓青教授的简介和专访内容。

贾熹滨（采访者，后缩写为贾）：您是 1962 年毕业于北京清华大学无线电电子学系，获优秀毕业生金质奖章并直接留校任教。能分享一下您求学期间，对您未来学术有影响的经历或有趣轶事吗？您觉得是什么样的动力促成您成为优秀学生？对现在求学的学子您能给些建议吗？

丁晓青（后缩写为丁）：我的父母是教师。在抗战时期，逃难到大后方贵州湄潭，后父母分别在浙江大学和浙大附中任教。1944 年我五岁时进入湄潭小学，抗战胜利后，46 年 10 月返回江浙。当时无法上学，46 学年在家自学了三年级课程，后在江苏南通和镇江镇师附小就读，50 年小学毕业。时为支援西北，随父迁至陕西。11 岁独自一人转赴西安入读西安省女子中学，三年后初中毕业。因不服盛传之“女不如男”说，独自转考入男女合校的省西安高级中学，至 56 年高中毕业。面临毕业后的高考，学校为我提供了令人羡慕的免试入学优待条件 (重点兰州大学的重要学科)。我为能够进入清华大学学习，申请放弃了此免试录取优待，而争取到能自主参加全国高考入学竞争的高考资格。通过全国高考的检验，最终如愿以偿，获得了在西安难得的清华录取名额。56 年 9 月，17 岁的我开始了在清华大学学习生活的重要

篇章。我感觉的是，内心永不满足的上进心始终在督促自己，不断学习努力提高，才能不断向前。

在清华大学无线电电子学系六年半的学习生涯，是我一生重要阶段，得到全面的提高和成长，最重要的是帮助我树立起为国为民、求真求实、独立自主、不务虚名的思想作风。

贾：我们了解到，您是文字识别和生物特征识别两个领域的知名学者，在图像处理、模式识别领域取得了众多突破性成果。您的成果包括“THOCR-1997 综合集成汉字识别系统”（1999 年）、“高性能东方文字文档智能全信息数字化系统”（2003 年）、“TH-ID 人脸和笔迹生物特征身份识别认证系统”（2008 年），“多字体多字号印刷体汉字识别系统”（1992 年）等一系列国家级奖励及很多省部级奖励。您能否和我们分享回顾一下您认为的重要研究成果及其影响？能分享一下这些成果背后的故事吗？

丁：在文革动乱结束后，我们从江西鲤鱼洲劳改农场返校，才为参与教学科研提供了可能。我们面对着的是，为国奋战的急迫决心与一无所有的残酷贫困现实的巨大反差。巨大的反差现实逼迫我们寻求出路：唯一的出路只有：自己动手、自己创造。我们要进行计算机图像处理研究，却没有显示设备。我们参与了彩色电视机和工业电视的研制；没有计算机图像处理设备，我们就从解决图像数字化、与计算机接口、数字图像的计算机输入、显示等入手，研究和解决从硬件到软件的各种问题。不仅在小型计算机上，而且在广泛、便宜的微型计算机上开发实现计算机图像处理设备。根据需要，我们研发成功我国首个的小型和微机两种数字图像处理系统。这些系统和技术获得了北京市和电子部的多项科技进步二等奖奖励。这第一步打基础的研发成功鼓励了我们，重要的体会是，只有认真学习和实践，才是取得进步发展的法宝。

科研工作一方面能促进科技发展，更重要的是要解决国家急需解决的重要问题，推广应用，促进生产力发展。在 80 年代，国家遇到的严重问题是信息化壁垒问题。西方发明发展的计算机，适合于西方文字和文明的需要，促进西方信息化文明的迅速发展。反观我国，汉字进不了计算机就是首先遇到的最大拦路虎。解决汉字的计算机输入就是当时立马需要解决的巨大困难问题。了解到这一情况，为祖国做贡献的理想促使我当即痛下了决心，必须要解决汉字识别汉字自动输入问题！

汉字自动识别输入计算机，不仅必须，而且十分困难，是我们不熟悉的、极端困难的模式识别问题。困难还在于汉字数量多，成千上万，结构复杂，最多笔画多达 36 画；字形变化巨大，印刷体有不同字体，手写字形因人、因时而异；还因设备而异，如铅笔、毛笔、签字笔等。汉字识别的实际应用，更要求在严重干扰噪声条件下能够准确高精度的识别，识别率要求 90-99%，甚至以上。这使得即使我们研究成功了一套在标准字样上达到高识别率的结构分析识别算法，在对实际样本的识别也败下阵来。总之，尽管在世界上和全国众多研究者一起不分昼夜，投入了研发，但收效甚微。大家都在惆怅于解决困难汉字识别的关键在哪里？

在这关键困难时刻，我们认识到，解决汉字识别困难问题，不能只是盲目地依赖于无休止的各种实验，我们必须从理论上找武器，必须从分析和理论上理解分析问题的实质。在这陷于苦难深渊的苦思冥想中，忽然出

图 2 2008 年国家科技进步奖颁奖大会

现一丝光线，那就是重要通讯问题的解决是依赖于信息论，其核心思想是，与传输信息相关的互信息的正确传输决定了通讯的质量与性能。信息论从理论到实践根本解决了通讯问题。从中我们受到启发：识别和通讯本质同样是信息的传输、变换等问题，二者有其相异、也有其相同之处。将信息论利用到汉字识别，发展了模式识别信息熵理论。从模式识别熵理论分析我们发现，从观测提取的特征与识别类别的互信息同样决定了识别的性能，而不是其他。过去有的是利用提取笔画作为特征信息识别汉字，其根本问题是，受到笔画提取准确度和有限笔画特征数量的限制。从汉字笔画提取的识别特征所具有的信息量太小，数量 8000 以上汉字类别不确定信息熵达 13 比特以上，识别特征必须的最小信息熵必须大大超过才有可能。否则是完全不足以满足克服汉字识别不确定所必须的信息量！至此，茅塞顿开，过去问题就出在汉字识别的特征提取信息不足的问题上。解决的唯一办法就是采取提供与类别相关高度信息熵的特征提取办法，提出了高维微结构统计识别算法，实现高互信息，就能保证汉字识别的基本条件。这样，就基本上和整体上解决了汉字识别这样一类巨大数量识别问题，一通百通，这就使得我们掌握了解决各种形式（印刷、手写、联机手写等）、各种文字（汉、日、韩、蒙、藏、维哈柯阿等）文字识别的钥匙，进而也帮助我们解决了许多后来遇到的生物特征识别、人脸图像识别等问题。从中，我们最大的体会是，揭示事物本质理论的重要性和强大力量！在任何科研创新中都必须放在首位。

图 3 2008 年钱伟长中文信息处理科学技术奖一等奖颁奖典礼

贾：您在理论和应用上都做出了众多创新性成果，您是如何在科研工作中发现关键科学问题，形成独创性理论方法的，能分享一下您的经历和经验，有什么建议给现在的青年学者？

丁：在我们从事的科研工作中遇到的问题，形形种种，各式各样，难以应付。但如果能够更深入更本质地进行分析，往往不难看出某些极其不同问题间其内在的一致性，有可能从中开辟出解决问题的新洞天。例如，在我们解决了汉字识别等的关键问题后，对社会智能和安全至关重要的人脸识别问题也摆在世界人民的面前。一般人看来，汉字识别和人脸识别是完全不同的两类问题，不可同日而语。但我却看到二者本质都是图像识别问题，具有相当的共性。因此，我们最早就将成功解决汉字识别的高清晰度高维微特征统计识别方法的思路用于解决复杂变化的人脸图像识别问题上，并获得显著的成功。其研究结果在 2004 年国际模式识别会议 ICPR2004 举办的世界人脸认证竞赛 FAT2004 中，以各项指标远超所有后者，而获颁发“人脸认证算法全面最优性能奖”，这是在世界范围首次将人脸识别技术达到实用水平。

图 4 在 FAT2004 颁奖仪式上

自 2005 年开始，我们的人脸识别算法系统就在深圳罗湖口岸的自助通关系统中得到大规模实用考验，取得了最优成绩而令世人惊叹。他们认为，你们搞文字识别的团队竟然在人脸识别这一看似不相关、不熟悉的领域拔得头筹，这是出人意料的。我们为什么能够取得这样的成功，是很自然的。因为我们在文字识别领域取得

图 5 丁晓青教授担任第 11 届文档分析与识别国际会议主席

了实质性的突破和进展的基础上，而我们又看到截然不同的文字和人脸识别问题，内在本质却是完全相似的图像识别问题。这一认识，促使我们采取不同于当时传统人脸识别结构分析方法，而采取类似于文字识别的高维微特征的统计识别方法，使人脸识别也取得了超乎寻常的进展。（其实，在 2014 年李飞飞在 Stanford 利用巨大图像数据库训练识别物体的思想建立视觉模型，和我们识别文字和人脸的思路和方法，是同出一辙的，可她利用神经网络识别系统的成功却晚于我们十年以后）。从这件事给我们最大的启示是，看问题，不能仅看表面的区别，更要从表及里，要看到内在同属图像识别的本质，抓住本质核心，才能找到认识和解决问题的真谛。

贾：能通过您的科研经历，回顾一下中国在文字识别、生物特征领域学术发展历史吗？有什么值得回忆的关于该领域有趣的轶事？您如何看待这一领域的发展：过去，今天与未来。

丁：文字是所有文明的基础，从 80 年代我国政府狠抓中文信息化工程开始，汉字识别（汉字自动输入技术）也得到极大的重视与关注。我们汉字识别研究 86 年取得进展，89 年汉字识别技术得以突破，92 年成功完成以 THOCR-92 简/繁多功能汉字识别系统为代表的汉字识别实用化系统。从此文字识别研究开启了持续、全面、深入发展的开端，及至研发成功各种高性能文字识别实

图 6 丁晓青教授担任第 11 届文档分析与识别国际会议主席，在闭幕式上颁奖

用系统，授权推广至国内外包括微软在内的重要公司和重要实用单位，对文字识别成功应用和发展打下了坚实的基础。

汉字识别研究从印刷体识别，发展到多种形式输入文字的识别：即多字体简/繁汉字、联机手写、脱机手写汉字识别也同时取得成功，进而还有摄像三维文字识别等等。文字识别从汉字识别向扩大文种文字识别发展，即汉、英、汉英混排，中日韩东方文字识别、蒙藏维哈克阿等主要民族文字、中国古代文字识别等，扩展至几乎世界所有重要的文字识别；而从文字识别进而发展到文档识别，则是广义文字识别发展更为重要的一环。我们知道，文档才是利用文字、符号等实际记载人类各种信息的手段和工具，其中包括各式各样的文字、符号和图案，有多种大小、字体、字形和横竖斜方向编排、有丰富多彩的图文结构，形成表达丰富信息的页面。文档识别包括对文档版面全信息的识别和理解，包括排版的行列块、篇章结构的分析和切割。对所包含内容文字符号块的行列和字符切分、内容符号识别以及排版结构的识别理解等，称为文档全信息识别理解和重构（已有文档的电子文档的无失真再现和电子出版）。这些极大促进了我们对我国古代的、历史的和现代的文档全信息识别理解，全信息的存储、传输和交流，以及当今普遍利用的电子文档的出版发行等等，以实现我们对人类文明的电子数字化的最好传承和发展。

几十年的发展已经证实，以文档的输出显示和文档识别输入为中心的信息化工程，极大地解决和推动了我国信息化事业、新闻出版、电子出版、信息存储传输等的发展。古今中外信息搜集存储，建立巨型数据库知识库，为智能化世界信息处理发展打下基础。在此基础上，还将从对已有文档的自动识别理解向未来新生文档信息的创造生成和自动创新的更高阶段发展，例如：自动回答问题、写文章小说编故事等等。

应当说，在智能信息处理范畴进行的研究工作，除了文档识别理解以外，越来越多的重视是对图像识别理解，包括生物特征识别和视频监控等等。图像识别首先在生物特征识别，尤其人脸身份认证识别对社会信息管理、社会安全保障的重要性估计再高也不为过。实际予以了证实。图像识别也经历了较长期的徘徊，在从静止图像识别到视频序列目标检测跟踪识别和理解等技术突破、以及对三维人物识别重构、表情姿态、年龄性别等深度信息识别和目标跟踪等研究的进展，在视频监控、自动驾驶等应用的推动下获得极为快速的发展。智能信息处理在方法论上，从结构方法的不足，发展到利用统计辨识方法，再进一步将统计方法和结构方法有机结合起来，解决实际世界复杂变化的对象和目标；更由于 2016 年以来深度卷积神经网络及学习算法在大规模人脸识别上取得傲人成绩以来，深度神经网络成为智能信息处理，以及人工智能发展的最强有力的武器，并普遍应用和促进各式项目的进一步提高性能和应用发展。

贾：想问一下，在您的科研学术经历里是否遇到过根据国家发展需要调整科研方向的问题？还是您会主动地去发现热点，去调整研究方向呢？

丁：应该说，我的研究经历中，总的研究课题和方向是根据和符合国家发展需要的。具体有两种情况：一种是确知国家发展需要和项目计划的，一定会调整研究方向到国家需要的项目方向，一定会想尽一切办法争取参加进去。例如汉字识别研究就是如此，一旦知道了，国家 863 计划要求进行汉字识别研究，我们立即调整方向进行研究，最后 863 以择优录取的原则，被选择进入 863

图 7 2011 年 3 月 7 日吴邦国委员长视察清华大学电子系时观看人脸识别身份认证系统

计划，并一直保持取得最优秀的成绩，对我国汉字识别发展做出了贡献。这是最好的一种情况和取得最好结果的情况。

我遇到的第二种情况是，我们难以了解到国家计划项目要求，即使了解到也没有把握能获得项目的支持，当然无从谈起调整研究方向问题。但我会根据自己对科技发展的认识和理解，认为某方向或课题，将一定会是国家急需的重要研究方向，并且我们已经有也有了一定的准备。这时我不会等待，我会根据我的理解、根据我自己的条件，自力更生，尽量立即动手进行研究，尽快获得成果。人脸识别的研究就是这样情况的一个典型例子。当时没有国家项目的支持，经费来源靠的是文字识别产品产业化取得的财务积累。我们坚信项目一定是国家发展的需要，所以经过我们的努力，一旦取得世界领先的研究成果后，就会逐渐纳入国家应用项目发展计划，得到一定的支持。事实说明了这一估计的准确性。开玩笑说，这也算是一种曲线救国吧！

总的体会是，进行科学研究，最重要和最基本的是，从国家发展的根本需求出发，但并不总是从国家发布计划项目出发。必须根据自己对问题的深入理解和持续努力，即使在困难条件下，也要创造条件，要自立，设法自己来支持自己，坚持进行研究工作。我们许多研究工作就是这样在国家有关发展项目的同时，也进行自选的重要项目相互协同完成的。

我们的研究工作进行得如此快速，令人吃惊。上世纪 90 年代我们完成了有关各种文字文档上识别的相关工作，并将其产业化市场化，在世界范围市场销售。在 20 世纪初开始、到 2004 年在人脸识别获得世界领先成果的基础上，立即开始大规模的实际应用和产业市场化齐头并进。工作如此迅速，如此广泛地在世界范围推广应用，除了依靠全体师生员工在创新的激情鼓舞下努力工作外，还有一点是，我们不是处在申请等待状态，而是时刻处在行动前进状态。一旦我认为方向是对的，有价值的，而且有可能的，我会立即动手开始工作。实际上我们为研究工作还做了许多“储备”，都是为了迎接新的任务的到来。

贾：你认为一个科研工作者以什么样的精神或态度从事科研，才能跟上时代的发展需要？您认为要做出有价值、影响力的工作，目前研究人员需要做出哪些努力呢？

丁：我们认为，科学无国界，许多研究成果发表在国际期刊，显示我们的研究成果，提供给世界，服务于全人类；科学家是有国界的，每一个科学研究者必须考虑，她的研究成果必须，首先能够服务于国家，服务于国家的复兴、国家的发展、百姓生活的提高，不这样做，就失去我国科学家的基本职责。

我们还认为，实践是检验真理的唯一标准。科学研究的成果不是仅仅发表几篇论文了事的。你的研究成功与否？能否解决实际中的问题？这些都牵涉到研究成果能否得到社会的承认、研究成果的价值等大问题。不可掉以轻心。因此，我们不仅将研究成果发表论文（论文达 600 余篇），还将成果鉴定成为可实现的系统，进而，还将成果产品化和产业化，用市场对研究成果进行实践考验。

目前科技迅速发展，成群参与科研工作，成果喷出。但真正有价值成果较少。我认为要做出真正有价值、有影响力的工作，首先就是选题问题，是选轻的、容易出成果的？还是选择重且困难的、对解决国家发展卡脖子问题中的核心困难问题有重要意义的？只有选择表面现象背后的实质和关键性基本问题，对于国家发展、科学进步、国际民生的重要关键问题。这种问题的解决，将会逐渐克服我国科技发展的倒三角缺陷，才能带来问题的根本性实际解决。在困难问题的解决中，才有可能做出有价值、有影响力的工作。

贾：您是我们视界专访栏目设置以来采访的第一位女性科学家，非常荣幸采访您，也想请分享一下，在您的成功历程中是否遇到过与男性工作者不一样的待遇或特有的问题，您是如何处理，有什么印象深刻的故事？

丁：虽然我们国家在男女平等和平权方面有了长足的进步，和其他许多国家相比，有很大的进步。作为一个女性工作者感到十分的庆幸；但说句老实话，重男轻女在中国也有几千年的历史了。在深刻的潜意识中，重男轻女仍然存在和大有市场。举一个个性化私人的例子就可说明。我的母亲，她应该是一个妇女独立的先驱者，她率先走出家庭，走进洋学堂，一辈子坚持工作，对待子女，也是男女平等，培养女孩子。但我还是感到在她的思想深处还是更偏爱男孩子的。当然不能怪她，从她的时代来说，她已经做得很好了，否则就不会有我的今天！

同我们国家一样，我当然很能体会到在深层次，依然有重男轻女的影响严重普遍存在。对女性工作者的培养和重视明显的要比对男性的要差，尽管你的各方面条件更好时也是如此。我当然遇到过这种情况。没有办法，一切只有靠自己的努力，做自己该做的，别无其他出路。我在心里的确感到，如果我是个男性科技工作者，情况可能会比现在要好些。尽管如此，我并不奢求。就像我在最后一轮，放弃了院士申请一样。因为我看到，院士评选，不仅仅是学术水平和学术成就，而且需要善于处理各种关系的全方位、综合的影响力。而一名负重的基层女性科研人员，无时间和精力去梳理难度之大的方方面面关系，更难获得如此的竞争力。

贾：作为学术界特别是工学领域的杰出女性科学家，您认为女性工作者有什么独有的优势？女性科研者面临的最大困难是什么？目前女性科研人员比例渐增，国家

也给予各种政策支持，能否请您给现在的科研领域女性工作者建议，更好发挥女性科研人员的作用。

丁：从我自己的体会看，起码在我生活的年代，没有感受到女性工作者有什么独有的优势，相反的，会遇到较多的歧视和困难。即便是工作，因为全身心投入使各种矛盾聚于一身，为了工作只能牺牲自己对孩子的照顾，这始终是心中的痛。对困难的科研工作，还会遇到中国根深棉固的传统重男轻女习俗对待，这是我的切身体会。所以要做出一点成绩，女性工作者需要付出比男同志更多的努力。如果说对现在的科研领域女性工作者的建议就是，所有一切需要依靠自己的努力，其他的，都是靠不住的。

作为女性科技工作者，我想多说几句，增强自信何其重要。中国几千年重男轻女的潜在影响依然存在，领导、公众不会关注到你。你自己再显自弱，就更没人理你了。我其实是一个只知往前闯和冲，并不很自信的人，其实没人给你信心，只有你自己。回想起来，当稍有点自信，就会坚持直至取得成功，一旦失去自信，往往会选择放弃，甚至会在关键时刻放弃而失去机遇。我想会有人和我有同感的。因此，自信是取得胜利的立足点。女性科技工作者们，珍惜自己，爱护自己，相信自己，增强自信，做出成绩，定会有更多的成功女性科技工作者出现。

我认为，要求对女性科技工作者特殊的照顾，即不可能，不现实，也不合理。只要能尽量克服潜在的重男轻女旧习俗的影响，不搞两种标准，真正男女平等对待就可以了。

贾：作为一名学术领域带头人和优秀的导师，能谈谈您的团队和学生培养经验吗？您认为什么样的导师才是一名合格的导师？一名优秀的学生应该具备什么样的素质？您认为在科研活动中导师和学生什么样的角色和关系更利于科研活动的开展，有利于整个团队科研发展的同时提升学生的学术能力？

丁：所有的工作、取得的每一项成果，都是团队中的老师和同学们大家共同努力的结果。因此发挥和调动每个人的积极性和主动性是关键。而每一位教师和同学，发挥作用，做出成绩也是他们最关心的事情。作为团队的带头人，将二者很好结合起来的关键就是使每个人都明白自己工作对自己的要求和职责，明白自己的创新任务的重要性。使每个人自觉发挥自己的主动性和创造性。一个团结朝气蓬勃的团队就会出现在你的面前。这是我们课题组的基本状况。


图 8 丁晓青教授与实验室师生演示获 2008 年国家科技进步二等奖的 TH-FaceID 人脸识别系统

我们课题组的确培养了众多优秀的学生，他们大多工作在国家重要的岗位、或自己创业发展经济、或是在国内外的公司工作。但是，我要坦白地说，他们的成长确实不是只我所为所致，而主要是依靠他们自己的努力，在学校的良好环境下，他们受到环境的影响，他们和老师同学相互切磋、互相学习，集体成长所致。

我以为一名导师应当做到的是：第一，不断学习、努力提高自己；以身育人，教师的品德修养，是培养学生的第一要素；第二，以身作则，以自己的行动教育学生；第三，在第一线解决困难和问题。绝不推诿、坐享其成。优秀的学生应当具备的素质，诚实、勤奋、同理心。在科研活动中，导师的角色是计划者和参与者；学生的角色是创新者与参与者；二者应当是平等的关系，发挥所长，共同完成任务。

贾：我们了解到您在企业承担首席科学家的职责，能通过您的经历，谈谈科研工作应该更多面向学术还是面向

市场需要，两类研究有什么异同和联系吗？对现在科研工作者如何选择研究方向有什么建议吗？

丁：以我的体会，在高等院校，科研工作首先是学术问题。在学校的科研工作，首先需要站在学科研究的前端，针对科研发展、国家发展急需提出和做出解决的方案，解决的具体方法，才有可能进行具体的市场推广，满足市场的需要。当然，市场需要的确反映了国家和民众的具体需要，科研工作也必须加以关注。但是对于高等院校或重点科研机构，首先应当站在科研第一线，解决更具有前瞻性、更具有全局性的困难问题，进而将其推广市场应用。即前者还是第一位的，解决市场需要还应是第二位的。

贾：除了科研领域，您在学科建设、人才培养方面也获得众多的成果，您能分享一些您的教学方面的经历和经验吗？能不能给年轻老师如何协调科研和教学工作一些建议？

丁：在科研领域，进行学科建设、人才培养外，我也很热爱教学工作。我很乐于将科研中的点滴体会，在解决科研难题时学习的体会、在学习中新的体会整理出来的新认识、将其系统化为一些新的学科思想和新学科内容，将其系统化教给学生，甚至写书成册，我不认为教学只是搬弄一些已有的旧东西，必须有创新，写书也不是当搬运工，从这一本，或这一篇搬到新的一本或新的一篇，教学和写作的灵魂是创新。因此，我非常乐于从事教学工作，为此也做了一些尝试，但可惜的是，由于时间的限制，做的太少。尝试有限，十分遗憾，作为弥补，在最近，我还在做一些尝试，希望对年轻的学子和同事们有所帮助。

贾：对 CCF-CV 刊物的读者寄语。

丁：珍惜祖国发展的大好年华，自强自立、目标远大、立足现实。

丁晓青

清华大学教授、博士生导师。1962 年，毕业于清华大学无线电电子学系，获优秀毕业生金质奖章并留校任教。长期从事智能图文信息处理和模式识别研究工作。在提出的模式识别信息熵理论指导下，采取微结构特征统计识别方法，对数量大变化巨干扰重的汉字和人脸等大规模复杂模式识别难题的突破做出重要贡献。率先突破汉字识别屏障、成功研制的领先识别性能的识别系统在世界范围销售推广；汉字识别覆盖了从印刷、联脱机手写多输入方式，识别文种包括世界主要中日韩英，和蒙藏维哈柯阿等文字和文档的识别系统，文档识别包括文档内容和版面结构的全信息识别理解和重构系统。率先将微特征统计识别方法应用于人脸识别等身份认证研究，取得重大突破。在 ICPR 举办的 FAT2004 国际评测中以全部超前成绩获“人脸验证算法全面性能最优成就奖”，在国际权威的 FRVT2006 人脸识别评测中获领先成绩。并国内外产品化推广。自 2005 年至今，应用于深圳罗湖口岸等四百多条出入境“旅客自助查验通道”，成为最早世界人脸识别技术大规模成功应用的重要范例。她先后荣获国家科技进步二等奖 3 次（2008、2003 和 1999 年）、三等奖 1 次（1992 年），以及 10 多项省部级奖励。发表论文 600 篇，合作专著 7 本，发明专利 33 项。由于在多文种文字识别、人脸识别等领域所取得的杰出成就和贡献，被选为国际模式识别协会会士(IAPR Fellow)、国际电气和电子工程师协会终身会士（IEEE Life Fellow)。

心底无私视界宽–焦李成教授专访

自 50 年代以来，我国在计算机视觉领域展开了相关的科研工作。而今，我国已经拥有一支庞大的、在这一领域辛勤耕耘且能与世界一流水平并驾齐驱的科研队伍。在这个过程中，有一批见证了视觉领域发展、为我国计算机视觉领域的奠基做出了重大贡献的先驱者。

《视界专访》栏目希望通过对计算机视觉研究历史、进展的见证者作一个系列专访，以帮助从事计算机视觉及相关领域的科研工作者或爱好者，全方面地了解 50 年代以来信息技术、信号处理技术以及计算机视觉相关的一些历史发展及进步，也希望能帮助我们在见证这段历史的同时，展望计算机视觉领域的未来。

我是负责本次专访的主要采访人，北京邮电大学明悦。本次采访通过微信交流完成，相关问题由 CCF-CV 专委会《视界专访》组提供。为能更好地帮助我们回顾本次采访，我们采用问答加书面回顾的形式来表述，西安电子科技大学刘旭老师负责提供了采访整理后的材料。以下是焦李成教授的简介和专访内容。

焦李成教授的主要研究方向为智能感知与计算、图像理解与目标识别、深度学习与类脑计算，培养的十余名博士获全国优秀博士学位论文奖、提名奖及陕西省优秀博士论文奖。研究成果获包括青年科技奖、吴文俊人工智能杰出贡献奖、国家自然科学奖二等奖及省部级一等奖以上科技奖励十余项，出版了国内第一部《神经网络系统理论》、《免疫优化计算、学习与识别》、《图像多尺度几何分析理论与应用》、《深度学习、识别与优化》《深度神经网络 FPGA 设计与实现》等专著二十余部，五次获国家优秀科技图书奖励及全国首届三个一百优秀图书奖。所发表的论著 H 指数为 95。

明悦（采访者，后缩写为明）：焦老师，您是如何走上计算机视觉和人工智能这个研究方向的?

焦李成（后缩写为焦）：我于 1978 年 2 月进入上海交通大学就读，1982 年 1 月毕业后我进入西安交通大学，硕士、博士期间在非线性和人工智能领域分支之一的神经网络方向开展研究。1990 年博士毕业之后，我进入西安电子科技大学工作，此后一直都在人工智能领域进行科学研究。

硕士的时候，我的研究方向是非线性电路、混沌，因此比较关注非线性方向相关的学术动态。其实不管是做什么方向的研究，关注学术前沿都是必须的。1983 年，我听加州大学伯克利分校非线性与神经网络三巨头之一的蔡少棠 (L.O. Chua) 教授在成都为期三个星期的讲学。蔡少棠教授是蔡氏混沌吸引子与细胞神经网络的创

图 1 讲课中的焦李成教授

图 2 神经网络系列丛书

立者，提出了蔡氏混沌电路（Chua's Circuit），促进了非线性电路理论的发展。那场讲学的主要内容就是非线性、混沌和神经网络，听完之后我受到了很大的启发，由此开始了解并进入人工智能这个充满魅力的领域。

博士的时候，其实我的博士论文并不是神经网络，而是《超大规模集成电路非线性和非理想效应分析》，就是集成电路系统中如果元器件出现非线性的时候，它会产生什么现象、怎么样去设计电路使它避免非理想效应。这个工作首先是要分析，就是用非线性工具去分析。虽然这篇论文是没有神经网络的，但其实我一直都在学习和研究神经网络，包括神经网络和非线性电路结合的相关研究。所有的一切都是非线性的，数学的非线性基础、物理的非线性基础，混沌也是非线性动力学，然后到用计算机编程、网络结构来解决问题，理论和实际应用、几个学科有机地交叉在一起，就会产生新的方向。期间我也在交大发表了一系列相关的文章，很多同学看了也很感兴趣。

85-87 年间，我参加了很多相关的国际和国内的会议，了解学术前沿。神经网络是典型的非线性复杂系统，也是当时的前沿方向之一，我对这个方向的兴趣也越发浓厚。因为兴趣和热爱，之后我就一直在神经网络、进化计算等人工智能领域开展研究。进入 2000 年后，我多次到美国、英国、日本等国家学习交流，开拓眼界，跟踪神经网络、人工智能学术前沿。

因为硕、博期间在神经网络方向上的一些积累和工作成果，1990 年，我获邀在第一届神经网络大会上作大会特邀报告，并成为中国神经网络委员会委员，当时 50 岁以下的委员只有我一个。1992 年，第二届国际神经网络联合大会（IJCNN）在北京举办，我也在会上做

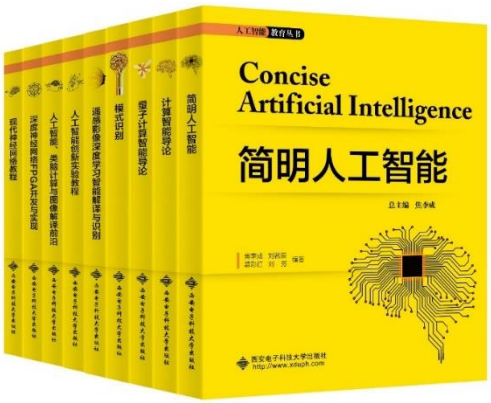


图 3 人工智能系列丛书

了主题报告。到了第三届，我们就把神经网络大会邀请到了西电，当届会议主席是保铮院士，组委会主席是谢维信副校长，我担任大会程序委员会主席，会议有近四百人参会。此后连续十余年我都作为中国神经网络大会的特邀报告人参会。一路走来，我国人工智能的发展能有今天都是因为在这个领域还属于“无人区”的时候，有许许多多的前辈给予支持，努力开拓，我们现在也要传承他们的精神，努力让我国的人工智能事业有更大的发展。

明：我们了解到您获得过诸多青年奖项和称号，能否分享您在早期的科研历程中最难忘的经历和心路？

焦：我开始读研究生的时候，研究方向是非线性电路、混沌。解析的方法，比如非线性常微分方程、偏微分方程、代数方程这些的解析解是很难求解的，常常是推了几十页的公式最后求不出来。那时凡是非线性相关的会议、讲习班等，不管是在北京、桂林还是昆明，全国各地我都会去听，在不断的学习和探索中，自然而然关注到了神经网络、遗传算法、小波等。虽然我的专业不是数学，但我在西安交大辅修了一年多数学专业的主干课程，特别是非线性和神经网络相关的数学理论，为我当时以及后来的研究打下基础。我认为科学研究要做得持久，打好基础、关注前沿、兴趣热爱这三点非常重要。上世纪 80 年代的神经网络对大家来说是一个非常新的方向，是一个“学习-识别-计算”的新思路，我也有一些对新方向的敏感性，再加上我对神经网络有浓厚的兴趣，所以在那个时期展开了一段“孤独”的研究之路。

因为方向新，所以感兴趣的人很多，但真正做神经网络的人却不多，当时整个领域的研究环境也不像现在那么方便、繁荣。那个时候没有网络，市面上也根本没有唾手可得的参考资料，只有高校的图书馆里才有那种专供的油印的文章。整个博士期间，为了查阅神经网络的文章，我一个人骑着自行车去遍了西安市各大高校的图书馆，所以感受就是“孤家寡人”，但是兴趣和热爱又在这个过程中给了我很多快乐。我的导师邱关源老师也给了我一个比较自由、宽松的科研环境，鼓励我在新方向上探索，所以硕、博期间我在神经网络方向上就有了积累，为今后研究工作打下了基础。

在西电从事博士后研究期间，我的导师保铮院士同样给了我非常宽松的科研环境，鼓励我开展神经网络、小波分析、进化计算等新方向的研究与探索，同时成立了智能处理研究小组。面向国家的实际需求，我们把这些方法应用于雷达目标识别和 SAR（合成孔径雷达）影像解译等研究中， 在国内我们是第一个这样去做 SAR 的，也一直坚持做了下来。在保老师的大力支持下，陕西省成立了国内第一个跨校的神经网络交叉研究中心，包括西电、西安交大、西工大，当时的第四军医大都参与了。

学科和领域要发展，要跟得上国际水平，就要降低学习门槛，让更多人参与进来，所以我就撰写了《神经网络系统理论》。该书是这个方向的国内首部专著，书籍内容全部基于我在神经网络方向上的积累和学习。这本书也得到了西电出版社的大力支持，那时写书就是纸、笔，手稿，当时出版社的 6 位打字员全部投入了《神经网络系统理论》的出版工作，就是打我的手稿。确实也是国内太缺少这个方向的书籍，出版之后卖得很火，被国内三百余所高校选为了本科生和研究生的教材或参考书，同时台湾儒林出版社也出版了这本书的繁体版本。非常感谢老一辈的老师、领导们对新方向的支持，也有了后面的《神经网络的应用与实现》与《神经网络计算》。基于我的博士工作，我还撰写了《非线性传递函数理论与应用》。这一系列专著得到了学界和业界的认可，获得

图 4 国家自然科学奖二等奖合影

了中国图书奖、国家教委优秀学术专著奖、国家优秀科技图书奖等国家奖项。之后我们团队又陆续出版了《免疫优化计算、学习与识别》《图像多尺度几何分析理论与应用》《深度学习、优化与识别》等二十余部专著，五次获国家优秀科技图书奖励及全国首届“三个一百”优秀图书奖。

之后有了团队支撑，我们在基于自然计算的智能学习与优化理论及方法方面进行了长期、系统的深入研究。21 世纪初，我国智能信息处理领域基础理论相对薄弱，理论体系有待完善，许多应用瓶颈问题也有待突破。我们主要做了四个方面的工作：一，针对海量、高维、非结构化信息处理中的优化与学习问题，建立了神经网络的非线性动力学模型和连接稳定性判据，提出了多子波神经网络模型；二，面向相对“小样本”和海量大规模数据学习对鲁棒、快速学习方法的需求，构造了满足 Mercer 条件的尺度核和父子波正交投影核，把 Mercer 核推广到经验映射函数，建立了隐空间支撑向量机和隐空间主分量分析模型，提出了快速稀疏逼近最小二乘支持向量机；三，面向大规模、多目标 NP-Hard 优化对高效、鲁棒优化方法的迫切需求，构造了免疫协同进化计算理论框架，建立了个体协同与竞争的智能体网络信息交互模型，进一步建立了协同认知免疫动力学计算框架，为解决实际工程中的大规模优化问题提供了高效的方法；四，建立的免疫协同进化和子波神经计算理论模型对数值优化问题、欺骗问题、组合优化问题、约束满


图 5 焦李成教授在第九届中国智能产业高峰论坛做报告

足问题等基准测试问题的求解结果都优于当时国内外文献的结果。这些成果是非常具有突破性的，也产生了广泛的国际学术影响，因此获得了 2013 年国家自然科学奖二等奖，还有省部级科学技术奖一等奖 3 项。我们实验室在人工智能领域总共获得了三项国家自然科学二等奖，还面向国家重大需求、面向社会实际应用成功研制了我国首套类脑 SAR 系统、首套基于面阵 CCD 的光谱视频成像系统、首个人脸画像识别系统等重大应用平台。2020 年我获得了吴文俊人工智能杰出贡献奖，我想也是对我在这个领域从“人迹罕至”走到“百花齐放”的一种肯定，所以科研这条路就是要坚定、坚持。

明：您是国内最早写《神经网络》相关教材的老师之一。您如何看待目前深度学习及 ChatGPT 的发展，以及它是否存在局限性呢？

焦：深度学习的基本思想是模拟人脑的信息处理机制，构建人工神经网络，基于数据驱动，希望能够对自然信息，尤其是声音、语言、图像等进行很好的处理，已经取得了传统计算机方法难以取得的突破。但是目前深度学习对于人脑的知识处理机制和推理机制实现得还不够，我们确实还有一些基本问题值得再认识与再思考。第一，如何有效地模拟人脑的稀疏性、选择性、方向性、学习性、多样性、记忆遗忘机制，对数据和知识进行学习、优化和识别。第二，如何 Beyond Data-driven，建立起 knowledge-based，physic-informed 和 brain-inspired 的机制，去解决复杂的、开放的场景问题、物理问题。第三，Beyond BP。在深度神经网络和学习中，我们主要运用 BP 算法进行优化，但是 BP 算法存在容易陷入局部最优解、梯度弥散和消失等问题。应该把全局达尔文进化学习和局部的拉马克、班德温学习结合起来，研究更有效的优化方法。学习和优化的基本数学问题实际上与高维几何动力学密切相关，我们怎么能够从高维的几何动力学角度去对神经网络的学习、逼近进行再认识、再学习，这是一个重要的问题。第四，Beyond Sigmoid。Sigmoid 函数的表征具有一定的局限性，它在稀疏层次表征、选择性、方向性、正则项、正交、紧支性上是有明显缺陷的。怎样能够更有效地表征仍然是一个需要进一步研究的问题。第五，Beyond Perception。我们现在模拟的更多是感知问题，但对认知的问题应该怎么样去做？感知与认知的协同建模与优化这个问题十分重要，需要考虑知识嵌入建模，知识与学习发现，归纳与推理，自学习，自组织，自演化，自推理等等。第六，DL for Science 的一般框架与范式。深度神经网络抑或人工智能，如何与科学紧密结合起来，真正地去解决科学的问题，而不仅仅是一种数据处理工具。深度学习理论的发展。我们还有很长的路需要去走，需要我们扎扎实实地工作。

ChatGPT 也是一个深度模型，而且是大数据驱动的大模型，其中参数量和数据量带来的收益必然是不可忽视的，同时，要训练如此体量的模型，必须有先进的训练方法、扎实的工程化技巧，ChatGPT 无疑是是神经网络应用于自然语言处理的一个成功实例，代表了深度学习成功的一个方面。但如我前面所说，深度学习的理论还有很大发展空间，ChatGPT 距离“感知+认知”的协同应用等高级智能也还有一段路要走。对于我们来说，要看到国外领先的地方，同时牢记这既是考验，也是机遇，还是要努力推动基础理论和关键技术的研究。


图 6 智能感知与图像理解教育部重点实验室十周年合影

明：焦老师早期从事的学科是电路与系统，与现在研究的智能感知和人工智能有什么区别和联系，如何看待当前智能处理方面的快速发展？

焦：电路与系统具有信号与信息处理、通信、控制、计算机等多个学科的理论基础，而我当时主要做的是非线性电路，也学了很多数学，这些都是人工智能学科的理论基础。从算法层面来说，不管是神经网络还是进化计算，都是从迭代的角度来求解非线性系统，给非线性系统求解提供了有力的工具。而这种在实际问题、实际数据、实际系统中的应用，其实就是把神经网络和物理机理结合了起来，拓展了神经网络研究的内涵和外延。人工智能是一个高度交叉的学科，理论的积累、多学科的知识和应用、融会贯通的能力对学习人工智能是非常有益的。

同时，人工智能也是普适的技术，它可以提升各个产业、行业、专业的能力和潜力，智能处理的发展重点是人工智能技术在各个行业的融合交叉，例如人脸检测、图像识别、机器翻译、语音识别与合成等相对成熟的技术今后应该要能高度交叉、深度融合的共同应用于某一领域，实现多技术协同的类人智能应用。智能技术的发展也离不开计算芯片的升级更新，传统的计算架构无法支撑人工智能算法大规模并行计算的需求，未来仍然需要“计算革命”来加速计算过程，满足产业需求。智能处理的发展还有很大空间，我们甚至希望看到它发展得更快速一些。

明：您培养了大量的学生，不少继续在国内从事科研工作。据说西安高校有不少老师是您的学生？在学术传承这块，您是怎么思考和培养的呢？

焦：西电在人工智能领域人才培养工作方面起步很早，1986 年就展开了博士和硕士研究生的人工智能教育，1991 年成立智能信号处理与识别研究小组，同年成立国内第一个神经网络研究中心，西电的人工智能领域研究生培养工作三十多年来从未间断。2003 年我们成立了智能信息处理研究所和教育部留学回国人员实验室，

图 7 智能感知与图像理解教育部重点实验室建设验收

2005 年，西电智能科学与技术本科专业开始招生，2008 年被评为国家级特色专业，2017 年成立部属高校首个人工智能学院，之后首批获准建设人工智能专业。主要得益于西电人工智能人才培养的前瞻眼光，领域内确实涌现出了许多西电的杰出校友，包括 973 首席科学家、国家领军人才，阿里达摩院专家、商汤科技 CTO 等。

对于学生来说，学校的气质是最有营养的土壤，导师和校友是最鲜活的榜样。西电既延伸着中国高校中最长的红色根系，又是“西迁”中的重要一员，这种家国情怀、不畏艰难、勤学求真的优良传统对学生是很好的熏陶，导师要以身作则，所以我到现在也是周一到周日每天都去实验室的。与此同时，我认为要贯彻产学研结合、协同育人的宗旨，我们构建的培养体系就是用产学合作、产教融合、科教协同、国际合作、本硕博衔接与协同的新工科方式，呈阶梯式、有针对性地完成让学生“会做、敢想、能创新”的培养，开发学生的创新思维、激发学生的创新潜力、提升学生的创新能力。西电的智能学科教育在多学科交叉的教学和科研实践中形成了“本硕博一体化”、“西电特色+国际化”的培养体系和理念。本科教育是学科教育的基础，人工智能学院的本科教育实行跨学科、跨学院、跨学校的三级选课和学分互认机制，本硕博一体化拉通计算，课程设置符合人工智能高度交叉的学科特性，兼顾通识教育和能力培养。本科后期就开始接触比赛、实训、实习、科研，硕士生除了实践能力更强调科研能力和创新能力，博士生强调独立性，独立地发现问题、解决问题，独立地自我管理。不论本、硕、博，都要以国际标准要求自己，站在国际学术前沿审视自己。

明：您对实验室的管理有哪些心得和高招可以分享呢？

焦：实验室的管理，目的就是保障我们的科研人员和教师们能够进行有效率、有成果的研究，同时希望大家在这个环境中能够尽量多的身心愉快。拿西电智能感知与图像理解教育部重点实验室来说，我们的原则是围绕实验室的主要研究方向，以国家战略需求和学科前沿为驱动，鼓励大家积极承担重大课题。具体做法包括几个方面，首先是制度保障，像人员管理、资源管理这些都要动态地不断完善规章制度，让管理工作更加规范化。其次是开放，要室务公开，重大事项决策要公开透明。三是各司其职，服务到位，我们设立了学术委员会、咨询委员会、管理委员会、行政办公室等，统筹兼顾科研中各个环节的需求。四是文化建设，既要严肃也要活泼，要对老师和科研人员们有人文关怀。最后，方式要先进，充分利用现在的信息技术，提高效率。

明：对于教学和科研的关系，想问问焦老师的建议，如何帮助青年教师实现教学和科研的相互补充，提升学生培养质量？

焦：教是育人的核心，研是创新的基础，高水平的科研应用于教学才能出高水平的、创新型的教学，两者是相辅相成、辩证统一的关系。从青年教师自身来说，要时刻牢记“教”和“研”都是基础，打好这两个基础，才能行稳致远。困难都是可以以时间和努力为代价去克服的，青年教师要坚定意志、增强信心、充分调动自己的能量，做国际水准的科研、内容扎实的教学，以培养出具有国际水准的学生。

从学校层面来说，首先是制度保障，建立和健全多元动态的教师发展评价机制，不单以科研或教学某一方面的成果论英雄，而是确立多元目标、多元标准、多元主体和多样方式的教师评价机制，引导青年教师在科研与教学的辩证统一中形成良性循环。其次是多维度搭建青年教师成长发展平台，发挥老教师的传帮带作用，大到成长规划、职业规划，小到项目申报、课程设计，让青年教师的疑难都有能找到获取帮助的平台。还有尽可

图 8 焦李成教授做下一代深度学习的学术报告

能创造多元的青年教师展示平台，营造好“小青年，大角色”的氛围，让他们各具特色的工作和才华被看见、被肯定，对学校和岗位有归属感，走上“正反馈”的成长之路。

从团队层面来说，要让青年教师的成长融入团队的成长。我们的团队于 2006 年获批教育部“长江学者支持计划”智能感知与图像理解创新团队及智能信息处理国家创新引智基地；2007 年，我们被批准成立智能感知与图像理解教育部重点实验室；2012 年又获批国家创新引智基地 (二期)，获批教育部“长江学者支持计划”智能感知与图像理解创新团队，获批 2 个陕西省首批重点科技创新团队，同年，我们成立了智能感知与计算国际联合研究中心；2013 年获批科技部国家级国际联合研究中心和陕西省 2011 协同创新中心，“视觉计算与协同认知”团队入选教育部创新团队；2014 年，与学校相关领域共同联合申报并获批“信息感知技术”国家 2011 协同创新中心；2015 年获批教育部智能感知与计算国际合作联合实验室、影像处理与安全传输科技部重点领域创新团队以及教育部智能感知与图像理解创新团队（滚动支持）；2017 年，智能感知与图像理解教育部重点实验室顺利通过五年一次的评估，去年我们的评估结果刚公示结束，是优秀。所有这些成果都是我们的青年教师们融入团队，和团队共同成长取得的。

明：您的学术成果得到了国际同行的广泛认可，当选了欧洲科学院外籍院士、俄罗斯自然科学院外籍院士，能分享一些您国际合作的经历和经验吗？我们看到您也

担任教育部科技委国际学部委员，能否谈谈您对推进国内和国外人工智能领域合作的想法或建议？

焦：西电智能感知与图像理解教育部重点实验室一直高度重视国际科技交流合作建设工作，积极拓展国际合作平台、高端引智、国际化人才培养等方面的途径。

其一，我们全力汇聚各方特色优势，推进一流学术平台建设。我们联合了国内外创新力量，建立了国家级的智能感知与计算国际联合研究中心、国家智能信息处理创新引智基地、智能感知与计算国际合作联合实验室等多个国际合作平台，为实验室国际化搭建桥梁。

其二，打铁还需自身硬，我们通过强化队伍建设来提升国际影响力，为国际合作夯实基础。实验室在国际顶级期刊担任主编/编委的人次逐年增加，将近 20 人次在国际组织机构任职，40 余人次担任国际期刊主编、副主编、编委。

其三，引智引课。我们引进了美国约翰斯•霍普金斯大学、美国中央密苏里大学、加拿大麦吉尔大学、英国伯明翰大学、英国诺丁汉大学等高校的计算机视觉、遥感数据处理与分析、高光谱数据处理、人工智能搜索算法导论等近 20 门海外优质课程。每年举办两次完全开放的“学术春秋”学术交流活动，已经坚持了十多年。近五年就邀请国际知名学者来校讲学 350 余次，共计报告 200 余场，其中海外院士、IEEE Fellow 做学术报告 50 余人次，国际知名期刊主编、副主编 80 余人次。促进海外学术大师与国内科研骨干和学生的深度交流。

其四，主动融入。实验室大力支持师生出国进行联合培养和交流，拓展国际学术视野。我们与美国麻省理工学院、美国哥伦比亚大学、美国加州大学洛杉矶分校、英国诺丁汉大学、荷兰莱顿大学、西班牙巴斯克大学等院校展开了常规化的交流访问，举办了 MIT 暑期访学之旅项目、斯坦福＆加州大学洛杉矶分校创新探索之旅项目、香港大学“人工智能与未来科技”访学实践项目等假期游学项目。近五年，实验室教师及研究生在 AAAI、IJCAI、ICCV 等国际会议上作特邀报告 70 余人次。

明：对从事计算机视觉的年轻科研工作者有没有一些寄语？能否在如何保持科研创新活力，在科研中发现科学问题，给青年学者一些建议？

焦：人工智能的发展方兴未艾，有志于人工智能事业的年轻学子们一定要趁势而上，更要踏踏实实。我个人经历了这一领域的起起落落，感触最深的还是要“坐得住”，不论这个领域当前的研究是冷是热，都要坚定信念，付出时间，坚持做能够推动学术发展、社会发展的研究。要保持科研创新活力，就要持续关注领域世界学术前沿，关注上游学科，具备国际视野，学会感受探索的快乐。

焦李成

欧洲科学院外籍院士，俄罗斯自然科学院外籍院士，IEEE Fellow，西安电子科技大学华山学者杰出教授。现任西安电子科技大学计算机科学与技术学部主任、人工智能研究院院长、智能感知与图像理解教育部重点实验室主任、智能感知与计算国际联合研究中心主任、智能感知与计算国际合作联合实验室主任、智能信息处理科学与技术国家创新引智基地主任、教育部科技委学部委员、教育部人工智能科技创新专家组专家、“一带一路”人工智能创新联盟理事长、陕西省人工智能产业技术创新战略联盟理事长、西安市人工智能产业发展联盟理事长、中国人工智能学会第六-七届副理事长、全国高校人工智能与大数据创新联盟副理事长、IET 西安分会主席、IEEE 西安分会奖励委员会主席、IEEE CIS 西安分会主席、IEEE GRSS 西安分会主席，担任 IEEE TCYB、IEEE TNNLS、IEEE TGRS、Research 等期刊 AE，当选 IEEE/IET/CAAI/CCF/CIE/CAA Fellow，连续八年入选爱思唯尔高被引学者榜单。国务院学位委员会学科评议组成员、人社部博士后管委会评议组专家、曾任第八届全国人大代表。1991 年被批准为享受国务院政府津贴的专家，1996 年首批入选国家级领军人才，陕西省首批“三五人才”第一层次。当选为全国模范教师、陕西省突出贡献专家和陕西省师德标兵。

COMPUTER VISION NEWSLETTER

01 2023
总第 35 期

计算机视觉专委会简报

CCF 计算机视觉专委会