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

04 2022

总第34期

CCF 计算机视觉专委会

COMPUTER VISION NEWSLETTER

计算机视觉专委会 简报

2022 年第 04 期

总第 34 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 走进高校系列报告会

第 116 期 天津大学

2022 年 9 月 17 日，由中国计算机学会计算机视觉专委会主办、天津大学承办的 CCF-CV 走进高校系列报告会活动，在 CCF 计算机视觉专委会 B 站官方账号成功举办。本次活动邀请了西安交通大学孟德宇教授、清华大学黄高副教授、同济大学赵才荣教授以及南京大学叶翰嘉副研究员四位计算机视觉领域专家学者做特邀报告。天津大学智能与计算学部李克秋教授、胡清华教授和 CCF-CV 专委会副主任、上海科技大学信息学院虞晶怡教授出席会议。天津大学智能与计算学部朱鹏飞副教授和王旗龙副教授担任本次会议的执行主席。

最后，活动执行主席、天津大学智能与计算学部朱鹏飞副教授对本次活动进行总结。首先感谢了四位讲者准备丰富，带来了十分精彩的学术盛宴。活动整体环环相扣、互有关联、精彩纷呈，为参会的老师同学们展示了一个较为全面的学术图景。此外，感谢参会的老师和同学的细心聆听，感谢中国计算机学会(CCF) 计算机视觉专委会、天津大学智能与计算学部给予本次活动的大力支持！

第 117 期 西北工业大学

2022 年 9 月 29 日，由中国计算机学会计算机视觉专委会 (CCF-CV) 和亚太信号与信息处理联合会杰出讲者计划（APSIPA-DL）联合主办，西北工业大学承办的 CCF-CV 与 APSIPA-DL 联合走进西北工业大学“智能视觉前沿技术”报告会，通过腾讯会议和哔哩哔哩成功召开，1200 余人次线上线下参加了会议。本次活动邀请了北京大学查红彬教授、上海科技大学虞晶怡教授、电子科技大学朱策教授、澳大利亚国立大学 Hongdong Li 教授、清华大学刘烨斌教授、西安交通大学兰旭光教授、澳大利亚悉尼大学 Zhiyong Wang 副教授和浙江大学李玺教授等八位智能视觉领域专家学者做特邀报告。

最后，本次活动执行主席、西北工业大学戴玉超教授感谢八位特邀讲者的精彩报告，打造了国庆节之前的一场云端学术盛宴。感谢中国计算机学会计算机视觉专委会、亚太信号与信息处理联合会杰出讲者计划、西北工业大学电子信息学院及陕西省信息获取与处理重点实验室暨国际联合研究中心给予本次活动的大力支持。

CCF-CV 走进高校系列报告会

第 118 期 太原理工大学

2022 年 11 月 20 日，由中国计算机学会计算机视觉专委会（CCF-CV）主办，太原理工大学承办的 CCF-CV 走进太原理工大学“计算机视觉前沿技术及应用”报告会，通过腾讯会议和哔哩哔哩成功召开，1600 余人次线上线下参加了会议。本次活动邀请到了江西财经大学方玉明教授、大连理工大学刘日升教授、厦门大学严严教授、东南大学耿新教授、浙江大学李玺教授和中国科学院计算技术研究所山世光研究员六位计算机视觉领域专家学者做特邀报告。与会成员有太原理工大学信息与计算机学院主持工作副院长李海芳教授，信息与计算机软件学院副院长曹锐副教授，信息与计算机学院学科带头人相洁教授，以及太原师范学院教务部部长穆晓芳教授，太原师范学院计算机科学与技术学院副院长亓慧教授等。

最后，活动执行主席、太原理工大学信息与计算机学院邓红霞副教授进行了活动总结，首先感谢了六位专家的精彩报告与学术交流分享，同时感谢了线上老师和学生听众的热情参与和高质量提问，最后再次感谢中国计算机学会(CCF) 计算机视觉专委会、学校和学院对活动的大力支持！祝贺本次活动取得了圆满成功，并期待下一次更精彩的报告！

第 119 期 复旦大学

2022 年 11 月 25 日，由中国计算机学会计算机视觉专委会 (CCF-CV) 主办，复旦大学承办的 CCF-CV 走进复旦大学“深度学习前沿技术及应用”报告会，通过腾讯会议和哔哩哔哩成功召开。本次活动邀请到了北京大学林宙辰教授、东南大学耿新教授和华中科技大学白翔教授三位计算机视觉与深度学习领域的专家学者做特邀报告。复旦大学人事处处长姜育刚教授，上海科技大学副教务长、信息科学与技术学院执行院长、CCF-CV 专委会副主任虞晶怡教授出席本次活动并发表致辞。本次活动的执行主席是复旦大学大数据学院青年研究员张力博士。

在研讨与交流环节中，参会成员首先对高校在大模型时代应该如何参与相关研发进行了踊跃交流，充分交换了意见，提出了许多具有建设性的论断，展示了资深学者对视觉领域发展方向高瞻远瞩的思考和精准的判断。活动致辞嘉宾、嘉宾、CCF-CV 专委会执行委员、复旦大人事处处长学姜育刚教授进行了活动总结，对三位与会专家的精彩报告与学术交流分享表达了衷心的感谢，同时邀请各位专家学者在疫情好转后前来复旦大学交流指导。

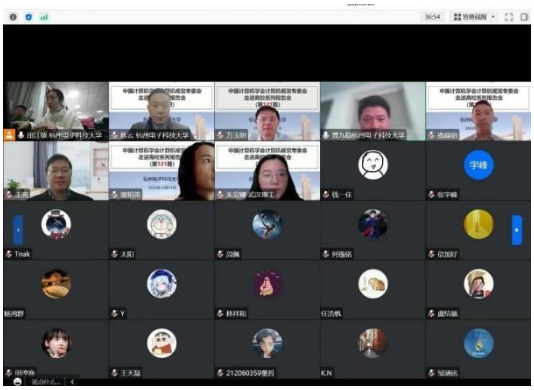
CCF-CV 走进高校系列报告会

第 120 期 北京工商大学

2022 年 12 月 4 日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、北京工商大学承办的 CCF-CV 走进高校系列报告会第 120 期活动以线上方式成功举办。北京工商大学党委常委、副校长刘敏华，CCF-CV 专委会主任、北京大学教授查红彬为报告会致辞。本次活动邀请了清华大学胡事民教授、丁贵广研究员，中科院计算所山世光研究员，北京理工大学付莹教授，和中科院自动化所刘静研究员做特邀报告，并进行了学术研讨交流。北京工商大学人工智能学院于重重院长担任本次活动的执行主席。

活动最后，于重重院长进行了简短总结。她首先感谢了五位专家的精彩报告，做报告的每位专家都在各自的领域，从理论和应用两个方面拓展了听众的视野；她表示此次活动为外界人士了解北京工商大学人工智能学院提供了很好途径，也为学校相关做视觉领域的老师、同学们提供了深入交流的机会。最后她再次感谢 CCF-CV 专委会对北京工商大学人工智能学院承办此次活动的大力支持，以及五位专家的精彩分享，并欢迎各位专家学者能再次来北京工商大学进行学术指导和交流！

第 121 期 杭州电子科技大学



2022 年 12 月 11 日上午，由中国计算机学会计算机视觉专委会（CCF-CV）主办、杭州电子科技大学承办的 CCF-CV 走进高校系列报告会第 121 期活动以线上方式成功举办。本次活动邀请了中科院自动化所王亮研究员，江西财经大学方玉明教授，深圳大学雷柏英教授三位计算机视觉与深度学习领域的专家学者做特邀报告。杭州电子科技大学自动化学院院长陈云教授，CCF-CV 专委会副主任、上海科技大学大学虞晶怡教授出席本次活动并发表致辞。杭州电子科技大学自动化学院副院长曹九稳教授担任本次活动执行主席。

活动最后，曹九稳教授进行了简短总结。他首先感谢了三位专家的精彩报告，指出做报告的每位专家都在各自的领域，从理论和应用两个方面拓展了听众的视野，最后他再次感谢 CCF-CV 专委会对杭州电子科技大学自动化学院承办此次活动的大力支持，以及三位专家的精彩分享，并欢迎各位专家学者能再次来杭州电子科技大学进行学术指导和交流！

责任编委 毋立芳、朱安娜

第14期 视觉质量评价前沿进展与未来趋势

CCF-CV 视界无限系列研讨会

视觉质量评价前沿进展与未来趋势

2022年9月16日

2022年9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14期CCF-CV“视界无限”系列活动——“视觉质量评价前沿进展与未来趋势”研讨会在线上举办。研讨会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致辞，北京航空航天大学徐迈教授、江西财经大学方玉明教授、西安电子科技大学吴金建教授、电子科技大学吴庆波副教授、北京电子科技学院金鑫副教授、字节跳动张亚彬研究员作报告。西安电子科技大学李雷达教授主持会议，并与以上六位讲者进行了深度研讨。计算机视觉专委会的B站公众号对本次线上会议进行了全程直播。

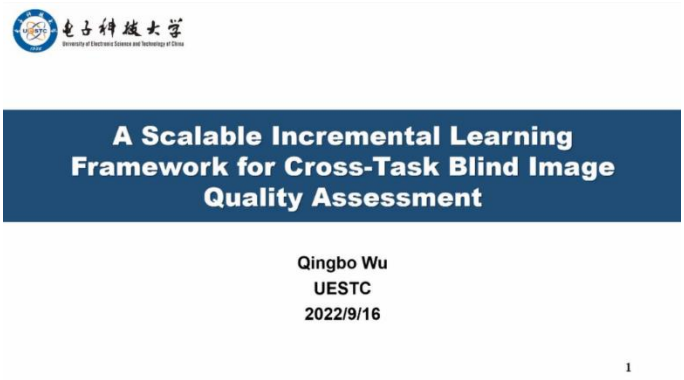
刘青山副主任首先代表CCF-CV专委会对西安电子科技大学李雷达教授团队承办本期研讨会表示感谢。他指出，“视界无限”系列研讨会是专委会举办的重要学术活动之一，本期主题——视觉质量评价具有重要的科学意义和应用价值，汇聚领域一线专家学者就该问题的前沿进展与未来趋势进行深入探讨，一定能吸引学界和工业界的关注，对促进领域发展起到巨大作用。刘青山副主任对六位讲者表示感谢，希望大家多支持计算机视觉专委会的工作，希望专委会的活动越办越好。

徐迈教授的报告题目是“面向未来沉浸式通信的全景视频质量评估”。在报告中，徐迈教授指出现有的全景视频质量评价方法并未充分考虑人类观看全景视频的用户行为，其一是观看者在观看全景视频时仅能看到视场中的内容，而非全景视频所有内容，其二是在视场内，观看者仅在关注区域看到清晰的高分辨率图像。针对上述问题，徐迈教授介绍了其团队提出的基于深度强化学习的全景视频感知模型。该模型通过引入一种基于视场的卷积神经网络首先对全景视频的视场进行预测，进而构建一个多任务学习的框架，以同时实现全景视频质量评估的主任务，以及视场内显著性检测的辅助任务。此外，徐迈教授还介绍了用于全景视频视场预测与质量评估的大规模数据库，可用于训练全景视频的视场预测模

型以及质量评估模型。最后，徐迈教授分享了其团队近些年来构建的数据集的下载链接和一些工作的开源链接，供大家学习使用。

方玉明教授的报告题目是“面向真实失真的视频图像质量评价研究”。方玉明教授在报告中首先介绍了图像质量评价研究的分类与主流方法，概述了面向真实失真的图像质量评价相关研究进展。紧接着，方玉明教授介绍了其团队近年来在该领域的一些研究工作，包括针对手机成像图像的质量评价研究、多曝光图像融合的质量评价研究和全景视频的质量评价研究等。最后，方玉明教授介绍了质量评价算法在感知优化中的应用、未来可能存在的一些应用场景以及发展趋势。

吴金建教授的报告题目是“脑启发式图像视频质量评价技术”。在本次报告中，吴金建教授首先分析了客观质量评估过程中存在的大数据集匮乏、与主观感知一致性差等难题，介绍了其团队构建的百万量级图像及视频质量评价数据集；随后，吴金建教授从大脑认知机理出发，介绍了其团队提出的基于语义衰减的客观图像质量评价模型和基于异质知识集成的视频质量评价模型，并且通过一些实验结果证明了这些模型的得到的客观质量预测更加符合人眼的主观感知。最后，吴金建教授展示了其团队近些年开发的多套图像质量评价系统在不同实际场景下的成功应用。



吴庆波副教授的报告题目是“面向跨任务盲图像质量评价的可伸缩增量学习框架”。吴庆波副教授指出，近些年来受限于网络结构的固化以及端到端的学习策略的流行，现有盲图像质量评价模型无法很好地应用于实际测试场景中。为了解决失真类型与评价原则不断变化的跨任务盲图像质量评价场景，吴庆波副教授提出了一种可伸缩增量学习框架。通过渐进式地更新局部的模型参数，从而在执行多种不同任务的盲图像质量评价时，可以避免在新任务的学习过程中对旧任务的灾难式遗忘。同时，为了避免模型参数饱和造成的学习中断问题，模型中进一步引入了可伸缩记忆单元。该记忆单元通过对属于旧任务的参数子集中神经元进行裁剪，遗忘不重要的既往经验，释放记忆存储空间，从而达到扩大模型的任务承载容量的目的。

金鑫副教授的报告题目是“视觉美学质量评估与应用”。金鑫副教授认为视觉美学质量评估是视觉质量评估的重要组成部分，如何从美学角度评估图像和视频的质量是计算机视觉、美学、认知科学、心理学、神经科学的交叉前沿热点。在本次报告中，金鑫副教授从图像与视频的美学质量评估研究动机出发，介绍了图像美感分类、美感评分、美感分布预测、美学属性评估、美学描述、美学问答、美学指导、视频美感分类和美感评分等主要研究任务，以及视觉美学质量评估前沿进展与发展趋势。最后，结合美学评价算法的实际应用，金鑫副教授介绍了其团队所提出的模型在智能手机拍摄指导、服饰搭配美感评分、广告智能辅助设计等场景的应用。

张亚彬研究员的报告题目是“画质评估算法在点播和直播的实践分享”。在本次报告中，张亚彬研究员首先针对UGC视频中场景多样化和难点问题进行深入探讨，并分析了产业界实际应用环境和学界对于画质评估算法的侧重点区别。随后，张亚彬研究员介绍了以VQScore为代表的多维度画质评估算法体系，分享了画质评估算法如何在点播和直播场景大规模地应用，以及画质评估算法如何服务于画质监控分析、画质增强和转码等业务。最后，张亚彬研究员分析了目前的画质评价算法在产业界面临的三个关键问题，并总结了自己对现有问题的思考与展望。

在Panel环节，参会讲者就“视觉质量评价前沿进展与未来趋势”、“视频质量评估相较于图像质量评估的最大难点”、“先验信息是否可以在全景视频中发挥作用”、“视觉质量评估在后面几年内最值得关注的问题”等问题展开深入讨论，并针对B站直播室的观众在会议期间提出的部分问题分享了各自的观点与见解。

Panel实录：

李雷达教授（主持人）：本次会议在最后有一个panel的讨论环节，邀请六位讲者一起来讨论一下视觉质量评估领域的现存挑战问题与未来的研究方向。第一个问题，咱们的视觉质量评价任务从2004年SSIM方法开始，已经涌现出了大量的包括全参考、半参考以及无参考的视觉质量评价方法，但是实际上在业界使用的时候还是大多数时候采用PSNR、SSIM这些通用的传统方法。那么现在到底还存在哪些问题导致我们这些设计的算法还无法很好地应用起来？

方玉明教授：我就先抛砖引玉了，在我做视觉质量评价研究的十几年中，其实包括在很早期的时候就有专家问过我，做这个任务到底有什么用？其实这也是我在本次的汇报中强调视觉质量评价在多媒体系统性能评价以及在优化方面的应用的原因。确实在近二十年来，工业界还是最喜欢用PSNR以及SSIM这些方法，我觉得工业界和我们学术界的侧重点还是不同的。之前我经常会和Wang Zhou老师（SSIM作者）进行探讨，他其实一直信奉的一句话就是“simple is beautiful”，越简单的东西越好。那么为什么他的SSIM方法受到如此多的关注以及应用的原因就在于实现起来非常简单。尽管如此，SSIM方法的计算复杂度其实还是要比PSNR要大不少，这也限制了它在工业界应用的进一步拓展。尽管现有的很多算法设计地非常fancy，但是实在是太复杂了，工业界包括我们学术界其实都不太愿意去用。其实在欧洲有些流派喜欢用人脑机理去设计算法，但是使用起来并不是很方便，也因此没有像SSIM那样被使用地那么广泛。在最近流行的深度学习方法中也存在这个问题，包括它们的可解释性到现在也不是特别清楚。其实我个人比较推崇把图像或者视频作为信号从而进

行特征提取的这类传统方法在视觉质量评价方面的应用。以上是一些我个人的比较浅显的见解，感谢。

李雷达教授（主持人）：感谢方老师的回答。那我们顺便就请教一下亚彬博士，您在工业界已经很多年了，那么在工业界进行画质评估的时候更看重哪些特性或者说更关注模型的哪些方面呢？

张亚彬研究员：这个方面其实刚才方老师已经说的很全面了，我们看重的其实一个是如何进行大规模地应用，因为我们需要大量地进行数据积累才能进行后续的优化分析。特别像我们现在的业务场景，如果没有大量的数据的话很难进行一些定制化的策略。所以说目前特别像 ToB 的场景中，VMAF、SSIM 和 PSNR 依然是主要参考指标，因为它们简单好用。VMAF 很多平台都支持，但是像深度学习的方法需要很多的依赖，很难进行很好地支持。另外一个点，深度学习方法所依赖的训练数据不够广，相较于 VMAF 和 SSIM 来说，泛化性不够强，在特定的应用场景中往往效果不够理想。解决的方法来说，可能未来需要建立各种大量的数据进行多数据的融合，包括不同的应用场景。所以其实总结起来就是两个点，一个是应用性，另一个就是泛化性。

李雷达教授（主持人）：确实，对工业界来说，简单有效的方法才能更好地进行使用。那么我们做了很多年的算法研究以后，可能会反过来去想如何去设计这种简单有效的方法，因为可能这有这种的方法才能够真正地在业界得到广泛的使用。刚才亚彬博士谈到了很多视频质量评估的工作，那么徐迈老师其实在视频质量评估方面已经做了很多的工作了，尤其是全景视频的研究，所以想问一下徐迈老师您认为相较于图像质量评估来说，视频质量评估最大的难点是什么？

徐迈教授：我觉得最大的难点在于人对于图像以及视频感知本身的机理。其实对于人对图像以及视频感知的探索还处于起步阶段，还没有充分地挖掘。包括刚刚一些讲者谈到了一些脑启发式的研究，但我觉得目前视觉心理学方面的研究还是不能给予我们足够的支撑。那么对比于图像的话，视频质量评价更大的难点在于视频的时序信息。人眼对于到底是哪一帧对于质量的影响更加关键或者说哪些关键的信息对于视频整体的影响比较大，这些信息实际上还是没有完全充分地进行挖掘。当然我们可以使用 LSTM 或者 Transformer 这种结构去建模连续帧对整体视频质量的影响机制，但事实上来说对于视频本身的时序信息建模还是很困难的。这就是一些我的看法。

李雷达教授（主持人）：另外一点，徐老师，您做的全景视频质量评估相比于自然场景的视频质量评估有哪些最大的区别，此外您认为自然场景的一些先验信息是否可以在全景视频中发挥作用？

徐迈教授：是这样的，人眼对于平面视频的关注，总会有一些重点关注的区域，那么对于全景视频来说其实也一样，因此这种视觉聚焦或者说感兴趣区域是可以作为视频质量评价一个非常重要的基础来对质量评价研究进行引导。但是和传统的平面视频不同的地方在于，全景视频的关注区域首先是一个视窗，在视窗以外的区域是无法观察到的，对质量没有任何的影响，而平面视频中的其他区域同样会对感知质量产生影响。举例来说，某个人脸区域非常重要，吸引了大量的关注，但是如果背景区域失真非常严重的话也会吸引人眼的关注。其实对于全景视频来说，这方面来说还是相对简单、容易进行处理的，这也是目前全景视频研究的预测效果相对来说比较好的原因之一。当然两者之间也是存在着共通之处，全景视频的视窗内同样存在着这样的感兴趣区域。

李雷达教授（主持人）：好的，徐老师，还有一个问题，您觉得目前来说视频质量评价模型的泛化性如何？

徐迈教授：其实不仅仅是视频质量评价任务，在其他的一些领域包括上层的图像处理任务（类似于识别），包括底层的视觉任务（增强之类的任务），都会存在泛化性的问题，这也是端到端的模型不可避免的一个问题。所以我个人觉得视频质量评价研究在泛化性方面确实也是遇到了很大的挑战。举个例子来说，针对人脸的场景进行训练的模型，很难在自然场景进行使用，因为它们的先验存在很大的区别。当然现在也有很多的工作使用类似于迁移的一些方法，但是目前对于解决泛化性问题还有很长的路要走。所以说如果我们能够了解人眼的工作机理，在网络的设计中进行结合，甚至是做一些可解释性的工作，或许可以很好地解决这个问题。

李雷达教授（主持人）：好的，谢谢徐老师。刚才谈到了模型泛化的问题，那么在刚刚的报告中包括吴金建老师和吴庆波老师都在关注这个问题，我想先请教一下吴金建老师，您认为对于脑启发式的工作研究中，后续还有哪些值得关注的点。

吴金建教授：其实我们在过往做研究的过程中其实也一直在思考，对于质量评价到底是一个什么问题没有定义清楚。因为像类似于识别、检测等任务来说，它们的目标都很明确，但是质量评价的定义很含糊和笼统。我们现在更多地是效仿检测、识别的网络来进行设计，但是它们究竟合不合适质量评价任务还是不确定的。我们团队最近几年做的脑启发的工作，是希望从认知的角度来探索从底层到中层到高层的层级认知结构，而这种结构应该是十分合理的。此外，我们还在每个层级究竟发生了哪些衰减，以及它们对感知质量或者后续任务到底有哪些影响进行探索。但是实际上人的认知是很复杂的，如何关联先验知识和人脑的记忆，以及如何解决泛化性或者是小样本的问题，其实还不是特别清楚。

李雷达教授（主持人）：谢谢吴老师，这也是后续的研究工作者可以探索的东西。那么接着请教一下庆波老师，最近几年您也是一直从跨任务的角度来进行质量评价的研究，请问您在做研究的时候在泛化性的评估方面如何准确地对不同任务和不同的域进行定义？

吴庆波副教授：刚才的报告里其实也提到了这个问题，在我提到跨任务的时候其实想强调的是不同的打分标准，即感知偏好。质量评价无论是在何种场景进行应用，最终得到的都是一个归一化的从 0 到 1 或者从 0 到 100 的一个数值。但是随着应用场景的变化，我们给这个数值赋予的语义信息是完全不一样的。这就是我们常说的，一个打分可以有多个模糊性，也即是我强调的不同任务的区别。那么在域上的区别的话，更多的是一个数据分布的差异问题。在域差异的问题上，其实前段时间我们也在思考如何解决泛化性的问题。原本是考虑用解耦的方法来做，那么对于图像来说，不同的域可能是自然图像、卡通图像或者说一些特效图像，它们都有自身的核心内容，并且往往与语义相关。所以我们也关注了一些无监督去做解耦的工作。能否从特征的层面上就分别出哪些特征是和内容相关，哪些是和纹理或者颜色相关。有些方法是需要一些额外的属性标签，当然现在也有很多无监督的方式，我觉得可能是一个比较好的解决泛化性问题的一个思路。

李雷达教授（主持人）：谢谢庆波老师。今天金鑫老师谈到的美学评估，其实最近我也在关注，其实从广义上来说美学质量以及我们研究的失真的度量都是质量，那么金鑫老师您觉得这两个方面是否可以放到一块儿进行研究？

金鑫副教授：我觉得雷达老师提出的这个问题非常好。就像刚才庆波老师所说的，可能美学的评分其实也就是质量评估当中的一个特殊任务。那么庆波老师刚刚提到的思路就非常好，我们能不能把美学评估看成一个特殊的任务加到庆波老师的那套理论中去，可以做一个这类的多任务学习框架。事实上，我自己之前在做评估模型时已经把方老师的SPAQ数据集一并用作模型的训练，那么在我们称为技术质量的指标不过关的时候，美学分数肯定也是非常低，只有在技术质量达标的时候才会往上去提及美感的评分。当然在实际应用过程中，有一些高美感的图片本身用了一些特殊的拍摄手法。它们的技术质量不是很高，但是总体的美感还是不错的，所以希望在后面希望是不是有机会可以和庆波老师合作进行研究这一问题。

吴庆波副教授：说起来其实我们之前有一段时间做一个编码的应用时就有这么一个情况，我们知道码流有的时候会出现信道失真，会产生连续的块效应，但是这样的块效应好像看起来变成了一个在画面里进行流动的形状，甚至现在有些人专门将它作为一个特效。本来这只是一个失真的现象，但是他们觉得这样很有美感，还有专门的编辑软件去生成这种信道失真。所以说美学可能和我们所说的常规的技术质量最大的区别在于主观性太强，很难用一个单一的标签去衡量。

李雷达教授（主持人）：感谢吴老师的回答。此外，我们线上的老师和同学也有一些问题，我们挑选了一些问题请专家们进行回答。第一个问题是请问多数据集混合训练在不断补充新的训练数据时是否需要从头进行训练？我们先请吴庆波老师来解答一下。

吴庆波副教授：这个问题和我做的内容其实比较相关，是这样，我做的这种参数隔离的方法实际上是不需要的，因为初始的数据只用来训练了一部分参数，这部分参数随后就固定不需要更新了。但是在下游任务进行测试时其实是可以进行重用的，那么当新的数据进来时不需要对原本的参数再进行训练，而是对额外的任务重新训练一部分参数。当然这是一个思路，此外我们在做few shot 或者 zero shot 的任务时，用到了现在更流行的方法，就是用一些很大的模型比如说 CLIP 或者 BERT 这种预训练的大模型。它们这些模型甚至不需要任何的 finetune 性能就很好，唯一要更新的可能就是大模型后面的抽头。这其实也是一个如何把超大规模的数据驱动的预训练模型应用起来的很好的思路。

李雷达教授（主持人）：最后一个问题我想让各位在线的专家各抒己见，聊一聊你们认为视觉质量评估在今后几年内最值得关注的问题是什么？

徐迈教授：我觉得最需要关注的第一个点还是怎么用，比如说怎么用在图像压缩以及增强的任务上。现在 PSNR、SSIM 还是主流，那么如何去将这些指标在比如说压缩的任务上进行融合，指导压缩的过程，不仅仅是简单的评价，这是第一点。第二点是用在哪，未来社交媒体的类型也比较多样化，包括全景、广场、多模态融合，对于质量的影响，因此用在哪儿的对象是比较关键的问题。第三点更重要的，还是要从脑启发、人的视觉记忆出发，结合现有深度学习比较重要的问题，就是可解释性的问题，亟待解决。

吴金建教授：我跟徐老师关注的问题也是比较像，还是要结合具体的应用，像我们团队另外一个研究方向是事件相机，它的数据格式和传统的格式不相同，那么如何让观察者看得更好，以及为了后续识别或者认知需求的时候又需要做怎样评价，这些需求的不同导致了设计的方法的差别。

吴庆波副教授：我也和前面两位老师有一样的感觉，就是你的方法到底怎么 work。其实之前 PSNR 和 SSIM 这么火的原因是它们可以直接放在 codec 上，但是目前所有无参考的方法除了用来做一些参数的调整工作，在很多图像处理的应用上面，如何告诉我们怎么把一个不好的质量的图像变成一个好的质量的图像这方面是匮乏的，这也是为什么这么久以来大家都对这个领域有所质疑的原因。另外一点，很多时候质量评价强调的是感知，但是当图像是给机器看的时候，它的需求是认知，感知和认知之间的 gap 如何关联起来是很重要的。

金鑫副教授：我和各位老师的观点在某一个方向上是非常一致的，那就是需要知道人脑是怎么样进行评价的，探求相关的规律，如何让质量预测模型变得具有可解释性是需要解决的。第二个点就是，模型需要在实际应用中进行发挥，来界定它是否有效，也就是强调模型的实际应用。

张亚彬研究员：其实刚才各位老师已经讲的非常全面了，我这边可能更多的是从工程应用方面来说。我认为我们需要细分垂类，因为深度学习的方法可能都不是很完善，最好是把它们做的可用性极高，在各个设备上都可以使用，包括说手机端、视频会议场景等。最好用的方法都是简单粗暴直接的，这也是工业界非常关注的点。

李雷达教授（主持人）：从各位专家的分享中也可以得知，近十几年以来，我们质量评估的领域得到了飞速的发展，也产生了各种各样的模型，那么接下来的重点可能要放在如何将这些模型真正地应用起来，使得我们领域更健康地发展，更多地和业界相结合，了解他们的需求。

李雷达教授（主持人）：由于时间的关系，本次研讨会到此就结束了。感谢各位专家的分享，感谢在线观众的耐心守候，谢谢大家。

责任编委 杨巨峰、潘金山

2022 年度 CCF-CV 专委工作会议顺利举办

计算机视觉专委会（CCF-CV）年度工作会议于 2022 年 12 月 23 日晚在线上顺利举办。来自全国高校、科研院所、企业的现任执行委员和新申请委员共计 340 多位参加了工作会议。会议由专委会秘书长、北京邮电大学马占宇教授主持。

会议首先由专委会主任、北京大学查红彬教授致辞。查主任代表 CCF 计算机视觉专委会欢迎并感谢专委会创始主任谭铁牛院士、CCF 学会代表熊盛武教授等特邀嘉宾及各位执行委员在疫情严重的特殊时期线上参会，肯定了专委会在过去一年中取得的丰硕成果，鼓励委员们聚焦学术前沿、做出高质量的研究工作，希望专委会加强产学研的互动交流合作，扩大与国际同行的交流合作，将各项学术活动做深做实，进一步提升活动的品牌质量，为推进计算机视觉学术研究与产业发展继续发挥积极的引领作用。

接下来，专委会创始主任、顾问委员会主任、南京大学党委书记谭铁牛院士致辞。谭院士感谢专委会查红彬主任的邀请，为专委会的发展规模和发展后劲感到欣慰，感谢各位委员在专委主任带领下为创建和发展专委平台付出的辛苦努力。谭院士还结合国家科技战略需求展望了计算机视觉的发展前景，鼓励委员们继续开展扎实的创新性研究，在国家制造业转型升级过程中提供强大的技术支撑；充分用好和发展专委平台，为年轻委员提供机会，推动计算机视觉的发展，为国家科技自立自强和现代化建设做出贡献。

随后，中国计算机学会（CCF）专委工委委员、CCF武汉主席、武汉理工大学人工智能学院院长熊盛武教授代表学会致辞。熊教授对专委工作会议的召开表示祝贺，对专委开展的走进高校、走进企业、视界无限等学术交流活动以及网站和简报等宣传方式给予了积极评价，特别肯定了专委在学术界与产业界、前沿与科普，以及人才培养方面的特色活动，期望专委会继续提升我国计算机视觉研究在国内外的影响力。

接下来，专委会秘书长、北京邮电大学马占宇教授向与会嘉宾和执行委员做了 2022 年度工作报告。报告简要介绍了专委会的组织结构、党的工作小组和专委顾问委员会，通报了 6 月份专委年度常委会议、2 月和 9 月秘书处工作会议所作出的若干新举措，总结了过去一年专委工作的重要成就，以及专委委员获得的各项奖励和荣誉，全面回顾了专委过去一年的学术交流活动，指出了工作中存在的问题和改进方案，最后介绍了专委会未来工作计划。

根据会议日程，工作会议进行了执行委员的新增选举工作，由专委会副主任、中国科学院自动化研究所王亮研究员主持。本年度共收到 63 位候选执行委员申请，王主任介绍了各位候选委员的基本情况。经现任常务委员投票表决，38 位候选人当选。

接下来，进入 CCF-CV 颁奖环节。颁奖仪式由专委会副主任、提名与奖励工作组组长、南京信息工程大学刘青山教授主持。刘教授详细介绍了本年度设立的奖项评选范围与评选规则，包括：终身学术贡献学者、杰出成就学者、学术新锐学者、持久影响力工作和服务贡献

学者。本年度的终身学术贡献学者为清华大学丁晓青教授。专委会创始主任谭铁牛院士高度评价了丁晓青教授在模式识别和计算机视觉等诸多领域的重要贡献，并在线为丁教授颁奖。丁教授发表了获奖感言，感谢专委会对自己几十年工作的肯定和奖励，希望年轻学者取得更加杰出的研究成果。

本年度的持久影响力工作获得者为南京信息工程大学张开华教授作为第一作者发表于 ECCV 2012 的论文 Real-Time Compressive Tracking，学术新锐学者为大连理工大学严彬、上海交通大学杨学和北京科技大学张世学，服务贡献学者为兰州理工大学李策教授、西安电子科技大学沈沛意教授、南京理工大学魏秀参教授、南方科技大学于仕琪教授和复旦大学张军平教授。五个奖项的获奖人既有早年投入我国计算机视觉奠基性研究的老一代科学家，又有新一代崭露头角的学术新星，也有为专委发展尽心尽力、无私服务的中生代科研工作者，充分展示了计算机视觉专委会大家庭的繁荣兴盛。

委员建言献策环节由专委会副主任、上海科技大学虞晶怡教授主持。虞教授介绍了走出疫情后专委将开展的主要国际交流活动计划，包括现场参加 CVPR、ICCV 等重要会议。专委会常务委员、南开大学程明明教授希望对 2023 年度线下活动进行规划并及时发布通知。南方科技大学于仕琪教授表示非常遗憾大家因疫情影响未能在深圳线下参加 PRCV2022 大会，邀请各位执行委员 2023 年来深圳组织活动。专委常务委员、航天宏图首席科学家王涛博士建议加强学术界和工业界的交流互动、合作研究，期待企业委员组织走进企业的专委活动、在专委平台发挥更大作用。专委会主任、北京大学查红彬教授赞同王涛博士的意见，在走进企业交流活动的基础上加强产学研交流合作，扩大执行委员的线下活动参与程度。虞晶怡教授就企业界和学术界各为彼此能够提供什么，指出企业可为学术研究提供应用场景和大规模数据、学术界可为企业开发新算法和提供高水平人才。太原理工大学邓红霞教授建议专委会为年轻委员提供项目申请书撰写方面的指导性帮助。

专委会主任查红彬教授对会议进行了总结。查主任指出 2023 年度的主要工作将围绕三件事情：一是如何使各项学术活动更加有深度、有成就，二是如何促进学术界与工业界之间更深层次的互动合作，三是如何加强疫情后国际同行间的交流合作。查主任还提出了对专委会组织结构进行调整的设想，加深专委会内部执行委员之间、与其他专委会之间，以及与其他学会之间的交流和联系。借明年专委会承办 PRCV2023 的机会，希望能够学习 CVPR、ICCV 等国际会议的办会方法，在 PRCV 大会期间组织一些特色活动，促进国内计算机视觉研究领域更深层、更高水平的学术交流。另外，明年是专委会的换届年，需要做好各项换届准备工作，保证专委会有序发展。

查主任再次感谢在线参加会议的特邀嘉宾和执行委员，并祝愿大家在新的一年里万事如意、身体健康。最后，专委会 2022 年度工作会议圆满结束，期待明年在厦门再聚！

责任编委 毋立芳

2022 年度 CCF-CV 秘书处第二次工作会议召开

2022 年 9 月 17 日，中国计算机学会计算机视觉专委会（CCF-CV）秘书处本年度第二次工作会议于北京召开。专委会副主任王亮研究员参会指导工作，秘书处全体成员参加了会议，会议由秘书长马占宇教授主持。

会议首先欢迎潘金山教授、朱安娜副教授作为新成员加入秘书处，并向他们介绍了秘书处的主要情况。然后，基于近年来工作开展情况和未来工作规划，优化调整了秘书处成员的工作分工。接着，为切实落实常委会决议，围绕学术活动、组织建设、宣传推广等议题展开深入讨论，包括以出版物等形式在国际上宣传国内视觉领域的优秀工作、推进《计算机视觉十讲》教材出版、开展示范性计算机视觉科普活动等，形成了具体可行的执行方案。

未来，秘书处全体成员将继续积极推进各项工作，提升专委会的组织活力与特色，用心服务各位委员及计算机视觉相关领域的专家学者。

责任编委 黄岩、任传贤

专题综述

基于低秩张量的多视图聚类相似性学习

中山大学 陈曼笙 王昌栋 赖剑煌

研究问题是基于低秩张量学习多视图样本间的相似性，并实现最终一致的聚类结果[1]。现有面向图的多视图聚类方法通常是使用多视图数据中隐藏的关系和复杂结构来实现令人深刻的聚类效果。然而，它们仍然存在以下两个常见问题：（1）他们以研究视图之间的共同表示或成对相关性为目标，忽略了多个视图间的全面性和更深层次的高阶相关性。（2）它们没有在统一的图构建中考虑特定视图表示的先验知识，并在统一的聚类框架中获得共识聚类指示矩阵。为了解决这些问题，我们提出了一种新颖的基于低秩张量的相似性学习方法用于多视图聚类(LTBPL)，在统一的框架中共同研究了多个低秩概率相似性矩阵和反映最终性能的共识聚类指示矩阵。具体来说，将多个相似性表示堆叠在一个低秩约束的张量中，以恢复它们的全面性和高阶相关性。同时，联合构建携带不同自适应置信度的特定视图表示和共识指示聚类矩阵的关系。在九个真实世界数据集的广泛实验表明了和最先进的聚类方法相比，LTBPL有明显的优越性。

一、研究背景

一个物品通常由来自多个视图的不同特征来表示，特别地不同特征之间是互补的，这直接推动了多视图学习的发展。多视图学习能够整合所有视图的不同特征，并利用它们之间的相关性去获得更精炼和更高层次的信息。多视图学习的成功来源于两个重要原则，即共识原则和互补性原则。其中，共识原则的目的是最大化多个视图之间的一致性；互补性原则意味着数据的某一视图包含了一些其他视图没有的信息。在这个工作中，我们关注于多视图聚类，其缺乏指导学习过程的真实类标。

近些年来，研究者们投入了许多的努力去设计多视图聚类算法。由于图的形式可以表征数据结构，现有的基于图的多视图聚类方法占多数，例如基于相似性的多视图谱聚类[2, 3, 4]，基于图的多视图子空间聚类[5,6]等等。其中，有几种常见的图构造方法，即k-最近邻图[7]，局部线性相似性图[8]、成对相似性图[9]和用子空间[10]学习的图。此外，基于张量核范数的张量奇异值分解(t-SVD)这个新兴策略，一些基于张量的多视图聚类方法被设计来发现多视图数据的空间结构和高阶信息[11,12]，这很好地改善了它们的聚类能力。但不幸的是，尽管这些方法取得了很大的成功，它们大多数旨在研究一个共同的表示或视图之间的成对相关性，导致了多视图数据之间全面性和更深层次高阶相关性的缺失，因此错过了重要的底层语义信息。此外，图的构建独立于聚类研究，且不关注学习相似性图趋向于聚类指示矩阵的先验信息，最终导致次优聚类效果。

针对上述问题，本文提出了一种新颖的基于低秩张量的相似性学习方法用于多视图聚类(LTBPL)，在一个统一的框架下联合研究低秩概率相似性矩阵和反映最终聚类结果的共识聚类指示矩阵。具体来说，在图学习的基础上，首先根据多视图样本点间的距离构造概率邻居图。为了全面探索所有视图之间的高阶相关性，多个视图的概率相似性矩阵被堆叠成张量，其中，本文利用基于 t-SVD 的加权张量核范数来恢复来自多个视图样本的潜在互补性和高阶相关性，在张量学习时考虑了矩阵不同奇异值之间的显著线索。此外，根据共识原则，本文联合学习了共识聚类指示矩阵和视图特定的概率相似性矩阵，其中利用多个视图概率相似性表示的不同

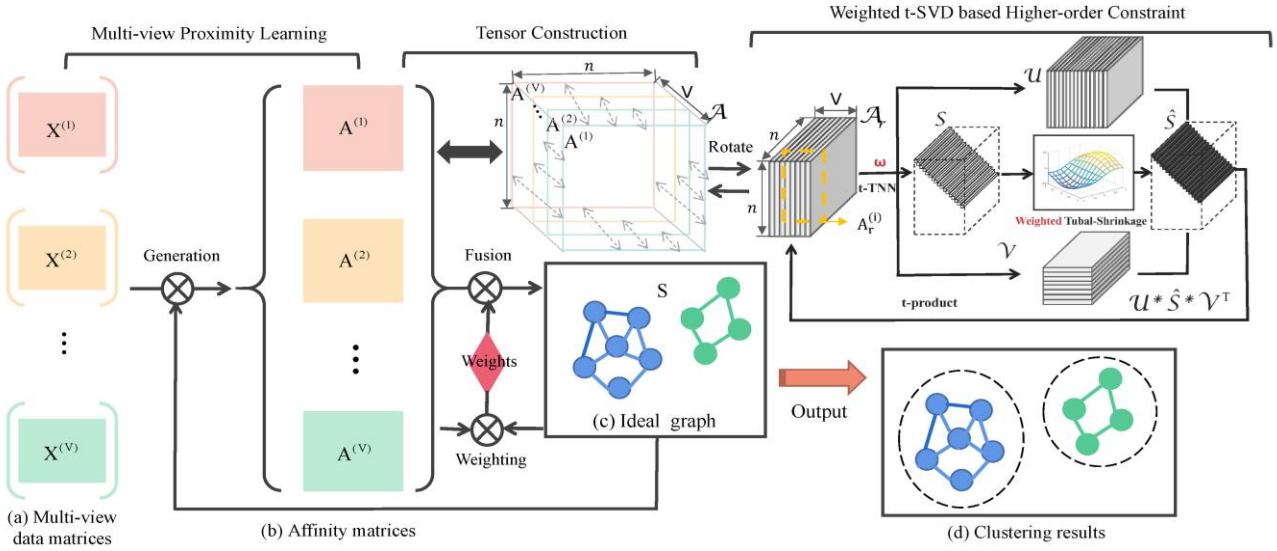


图 1 所提出 LTBPL 模型的示意图

置信度自适应地研究共识聚类指示矩阵。因此，所学习到的低秩概率相似性矩阵能够更好地表征数据结构潜在的互补和高阶相关性，而且反映最终聚类结果的共识聚类指示矩阵是同时通过了多个低秩概率相似性矩阵的不同贡献研究得到的。现将主要贡献概括如下：

- 一个新颖的框架被构造，以同时研究基于 t-SVD 加权张量核范数约束的低秩概率相似性矩阵和整合的共识聚类指示矩阵。
- 不同视图奇异值的先验信息通过基于 t-SVD 加权张量核范数能够被显示地考虑。同时，趋向最终共识聚类的不同低秩概率表征的贡献可以自适应学习得到。
- 在九个真实世界数据集上的广泛实验表明和最先进的多视图聚类方法相比，我们的方法有明显的优越性。

二、LTBPL方法介绍

1. 基于低秩张量的相似性学习

给定一个多视图数据集$X = \{X^{(1)}, \dots, X^{(V)}\}$，包含了V个视图，其中$X^{(v)} = \left[x_1^{(v)}, \dots, x_n^{(v)}\right] \in R^{d^v \times n}, \forall v = 1, \dots, V$表示第v个视图的特征空间。对于多视图聚类任务，探索数据的局部连通性是一种成功的策略，$x_i^{(v)} \in R^{d^v \times 1}$ 的邻居通常被描述为数据集中与$x_i^{(v)}$挨着的k个近邻样本点。特别地，在本文中，概率邻居简单地通过运用欧几里得距离作为距离度量来考虑，然后数据样本的相似性可以根据他们基于图学习的距离得到。具体而言，对于第v个视图的一个数据样本$x_i^{(v)}$，可以将所有数据样本点$\left[x_1^{(v)}, x_2^{(v)}, \dots, x_n^{(v)}\right]$作为连接到$x_i^{(v)}$的邻居，对应的概率为$a_{ij}^{(v)}$，其中当有一个较小的距离$||x_i^{(v)} - x_j^{(v)}||_2^2$时，可以得到一个较大的概率$a_{ij}^{(v)}$。因此，本文定义了一个基础的框架去学习相似性$a_{ij}^{(v)}$：

$$\min_{A^{(v)}} \sum_{v=1} \sum_{i,j=1}^{n} ||x_i^{(v)} - x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha ||A^{(v)}||_F^2,$$
$$\text{s.t. } 0 \le a_{ij}^{(v)} \le 1, \left(a_i^{(v)}\right)^T \mathbf{1} = 1,$$

其中α是一个权衡参数，$a_i^{(v)} \in R^{n \times 1}$表示一个列向量，它的第j个元素是$a_{ij}^{(v)}$。第一项是被用于确定概率相似性，第二项引入正则化项去避免平凡解$A^{(v)} = I$。尽管取得了显著的效果，大多数现有的方法旨在研究一个共同的表示或视图之间的成对相关性，导致多视图数据间全面性和更深层次的高阶相关性的缺失，从而错过了重要的底层语义信息。 此外，它需要一个单独的后处理步骤以获得最终的聚类结果，并且无法在一个统一的框架中考虑不同视图的多个概率相似性矩阵和最终聚类指示矩阵的联系，从而导致随后次优的聚类性能。

针对上述问题，本文提出了一个新颖的基于低秩张量的多视图聚类相似性学习方法(LTBPL)，其中每个具有高阶相关性的视图特定的相似性矩阵和最终的聚类指示矩阵以相互作用的方式实现联合优化。为清晰起见，

所提出的 LTBPL 方法的流程图如图 1 所示。依据图示，从多个特征子集或者源头获取得到的数据样本$X^{(1)},\dots,X^{(V)}$首先作为输入。基于多视图相似性学习策略，可以得到每个视图对应的相似性矩阵$A^{(1)},\dots,A^{(V)}$。为了捕捉到不同视图中多个样本点之间的高阶相关性，本文运用了张量构造技术，其中被构造的张量 $\mathcal{A}\in R^{n\times n\times V}$ 是由多个相似性矩阵构成。然后，所提出 LTBPL 方法的模型表达可以进一步构造如下：

$$\min_{\mathcal{A},A^{(v)}} \sum_{v=1}\sum_{i,j=1}^{n} ||x_i^{(v)}-x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha||A^{(v)}||_F^2 + \beta C(\mathcal{A}),$$

$$\text{s.t. } 0\le a_{ij}^{(v)}\le 1, \left(a_i^{(v)}\right)^T \mathbf{1}=1,\ \mathcal{A}=G(A^{(1)},\dots,A^{(V)}),$$

其中β是一个惩罚因子。$C(\cdot)$表示在构造张量 $\mathcal{A}$ 上的特定约束，$G(\cdot)$通过整合多个相似性矩阵$A^{(v)}$成一个三阶张量。具体地，本文采用了基于 t-SVD 加权张量核范数约束去恢复隐藏在多视图相似性矩阵里的高阶相关性，它的构造可以进一步改写如下：

$$\min_{\mathcal{A},A^{(v)}} \sum_{v=1}\sum_{i,j=1}^{n} ||x_i^{(v)}-x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha||A^{(v)}||_F^2 + \beta||\mathcal{A}||_{\omega,*}$$

$$\text{s.t. } 0\le a_{ij}^{(v)}\le 1, \left(a_i^{(v)}\right)^T \mathbf{1}=1,\ \mathcal{A}=G(A^{(1)},\dots,A^{(V)}).$$

其中$||\cdot||_{\omega,*}$表示基于 t-SVD 加权张量核范数约束。特别地，通过加权张量核范数最小化，$\mathcal{A}$ 的所有奇异值会被不平等地正则化，且软阈值函数可以用不同的加权参数来收缩所有不同的奇异值。在进一步详细计算前，需要对构造的张量 $\mathcal{A}$ 进行旋转，以便更好地捕捉视图间的低秩属性，并显著降低计算复杂度，维度从$n\times n\times V$变化为$n\times V\times n$，其变换步骤可见于图 1。

2. 自适应加权的共识整合

基于改良的带有高阶信息的相似性矩阵，可以推导出直接反映最终聚类结果的共识理想相似性矩阵：

$$\min_S \sum_{v=1} ||S-A^{(v)}||_F^2,$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1, \text{rank}(L_s)=n-c.$$

在上述的公式中，$\text{rank}(L_s)$表示的是拉普拉斯矩阵$L_s=D_s+(S+S^T)/2$的秩，其中$D_s\in R^{n\times n}$是一个对角矩阵，它的第j个元素是$\sum_j(s_{ij}+s_{ji})/2$。

注意到上述模型平等地对待每个相似性矩阵去学习一致的图表达，这忽略了多个视图的不同贡献度，并导致最终次优的性能。因此，本文设计了一个更合理的自适应加权策略去整合多个相似性矩阵[13]，它的目标函数可以表达如下：

$$\min_S \sum_{v=1} \gamma^{(v)}||S-A^{(v)}||_F^2,$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1, \text{rank}(L_s)=n-c,$$

其中$\gamma^{(v)}$被定义如下：

$$\gamma^{(v)}=\frac{1}{||S-A^{(v)}||_F}.$$

明显地，可以看到$\gamma^{(v)}$依赖于S。如果第v个视图是良好的，对应的$||S-A^{(v)}||_F$应该是小的，那么权重$\gamma^{(v)}$应该是大的。反过来，一个较差的视图会被赋予较小的权重，这表明了本文自适应加权学习策略的意义。然而，解决上述模型的优化问题是十分困难的，因为L_s依赖于目标变量S，且秩约束$\text{rank}(L_s)=n-c$是非线性的。

依据文献[14]，让$\theta_i(L_s)$表示L_s的第i个最小特征值。由于L_s是半正定的，那么有$\theta_i(L_s)\ge 0$。给定一个足够大的 λ，上述的模型中的秩约束可以被去掉，并等同地改写为：

$$\min_{S,\gamma} \sum_{v=1} \gamma^{(v)}||S-A^{(v)}||_F^2 + 2\lambda\sum_{i=1}^{c}\theta_i(L_s),$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1.$$

当 λ 足够大且对于每个i有$\theta_i(L_s)\ge 0$，上述模型的最优解S会让第二项$\sum_{i=1}^{c}\theta_i(L_s)$接近于 0，从而满足秩约束$\text{rank}(L_s)=n-c$。额外地，受文献[15]的启发，可以得到以下的等式：

$$\sum_{i=1}^{c}\theta_i(L_s)=\min_{F^TF=I}\text{Tr}(F^TL_sF).$$

因此，关于共识相似性矩阵的学习模型可以重构为：

$$\min_{S,F,\gamma} \sum_{v=1} \gamma^{(v)}||S-A^{(v)}||_F^2 + 2\lambda\text{Tr}(F^TL_sF),$$

$$\text{s.t. } 0\le s_{ij}\le 1, s_i^T\mathbf{1}=1, F^TF=I.$$

最终，考虑到多视图相似性矩阵和共识相似性矩阵的联合学习，将所提出的 LTBPL 模型构造如下：

$$\min_{\mathcal{A},A^{(v)},S,F,\gamma} \sum_{v=1}\sum_{i,j=1}^{n} ||x_i^{(v)}-x_j^{(v)}||_2^2 a_{ij}^{(v)} + \alpha||A^{(v)}||_F^2 + \beta||\mathcal{A}||_{\omega,*} + \sum_{v=1}\gamma^{(v)}||S-A^{(v)}||_F^2 + 2\lambda\text{Tr}\left(F^TL_sF\right),$$

s.t. $0 \le a_{ij}^{(v)} \le 1, \left(a_i^{(v)}\right)^T \mathbf{1} = 1$, $\mathcal{A} = G(A^{(1)}, \ldots, A^{(V)})$,

$0 \le s_{ij} \le 1, s_i^T \mathbf{1} = 1, F^T F = I.$

我们观察到，最终的共识图S和由低秩张量 $\mathcal{A}$ 约束得到的每个相似性矩阵$A^{(v)}$能够在统一框架中联合学习。

上述模型中的低秩张量正则化项是用于挖掘多个视图$A^{(v)} \in R^{n\times n}$之间的潜在全面性和高阶相关性的。一方面，张量旋转之后，旋转张量 $\mathcal{A}_r$ 的第i个正面切片 $\mathcal{A}_r^{(i)} \in R^{n\times V}$描述了在不同视图中n个样本点的关系。一个好的图$A^{(v)}$应该确保在不同视图里n个样本点之间的关系应该是一致的。考虑到不同视图通常揭示了不同的类结构，本文将张量多秩最小化约束施加于张量 $\mathcal{A}$，确保了每个 $\mathcal{A}_r^{(i)}$拥有空间低秩的结构，从而使得 $\mathcal{A}_r^{(i)}$可以很好地刻画多个视图之间的全面性信息。另一方面，和矩阵(二阶的张量)相比较，这里的三阶张量是高阶的。有了低秩的约束，高阶相关性(三阶的)可以通过张量挖掘得到，而矩阵只能捕捉到二阶的关系。因此，不同视图之间的潜在全面性和高阶相关性可以被模型中的低秩张量正则化项很好地挖掘得到。

三、实验结果

1. 数据集

表 1　九个真实世界数据集的统计数据

Datasets	Type	#Objects	View dimensions	#Classes
Yale	Image	165	4096, 3304, 6750	15
ORL	Image	400	4096, 3304, 6750	40
COIL-20	Image	1440	1024, 3304, 6750	20
UCI	Image	2000	240, 76, 6	10
Caltech-101	Image	1474	48, 40, 254, 1984, 512, 928	7
Notting-Hill	Image	4660	6750, 3304, 2000	5
Hdigit	Image	10000	784, 256	10
BBCSport	Document	544	3183, 3203	5
BBC4view	Document	685	4659, 4633, 4665, 4684	5

本文在九个真实世界数据集上对所提出的 LTBPL 模型进行了广泛的实验，以证实 LTBPL 的有效性和优越性。具体地，九个数据集的统计数据可见于表 1。所提出 LTBPL 方法的源代码可以通过以下的链接进行下载：https://github.com/ManshengChen/Code-for-LTBPL-master.

2. 对比实验

不同的聚类方法在九个真实数据集上得到的聚类结果分别报告于表 2、表 3 和表 4。"SC 1" 表示在数据集的第一个视图中执行谱聚类算法，类似地对于 "SC 2" 和 "SC 3" 等等。在不同的表中，我们用粗体强调了不同数据集上的最佳性能。

从这三个表中可以观察到，提出的 LTBPL 方法在所有基准数据集上几乎都达到了最好的聚类性能，尤其是在九个数据集中的七个上获得了聚类结构与真实标签的完全匹配(即全部 1)。例如，在 Yale 数据集上，LTBPL 明显优于第二最佳方法(UGLTL[16])，通过分别实现 ACC 和 NMI 的改进为 0.7% 和 0.8%。在 BBC4view 数据集上，LTBPL 明显优于第二最佳方法(WTNM[12])，通过分别实现 ACC 和 NMI 的改进为 0.44% 和 1.63%。

尤其是，可以观察到基于 SVD 张量核范数的方法，即 LTBPL、UGLTL[16]、ETMC[17]、tSMC[11]和 WTNM[12]，比其他的方法通常能够实现更好的聚类效果。这证实了张量核范数在捕获多视图数据高阶关系的有效性。尽管如此，本文所提出的 LTBPL 方法有更明显的优越性，其中每个视图的相似性矩阵和共识的聚类指示矩阵可以在统一的框架中联合学习得到。

此外，在 Yale 数据集上的可视化结果可见于图 2，可以看到，LTBPL 揭示了一个相当清晰的底层集群结构。

3. 消融实验

为研究低秩张量和自适应联接多个相似性矩阵和共识聚类指示矩阵策略的作用，本文进行了深入的消融实验。对于 LTBPL-t1，张量核范数项被去掉($\beta = 0$)，其他项保持不变。对于 LTBPL-t2，将学习到的多个相似性矩阵累加得到一致的相似性表征($\gamma^{(v)} = 0$)，并作为谱聚类算法的输入得到最终的聚类结果。在所有基准数据集上的对比结果可见于表 5。从表中可以看出，LTBPL 在所有的测试中都比 LTBPL-t1 和 LTBPL-t2 更优越。

四、总结

本文设计了一种新颖的基于低秩张量的多视图相似性学习方法(LTBPL)。多个相似性表征被堆叠成一个受 t-SVD 加权张量核范数约束的低秩张量，去挖掘多个视图之间的全面性和高阶相关性，其中多个视图奇异值

表 2　对比结果：在 Yale、ORL 和 COIL-20 数据集上通过不同方法得到的均值和标准差

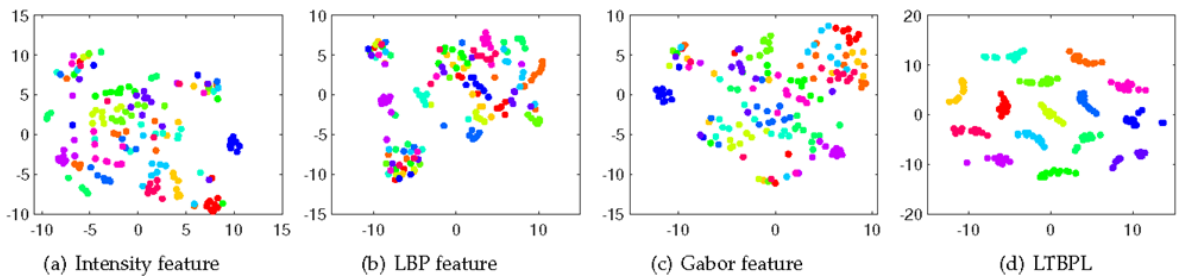
Datasets	Yale				ORL				COIL-20			
Method	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI
SC 1	0.538±0.044	0.586±0.038	0.383±0.043	0.341±0.046	0.650±0.016	0.798±0.010	0.528±0.021	0.516±0.022	0.655±0.028	0.756±0.014	0.598±0.023	0.959±0.00
SC 2	0.569±0.033	0.598±0.024	0.423±0.031	0.384±0.034	0.774±0.025	0.891±0.013	0.712±0.031	0.704±0.032	0.745±0.024	0.828±0.012	0.712±0.023	0.971±0.00
SC 3	0.640±0.039	0.657±0.031	0.489±0.037	0.454±0.040	0.704±0.027	0.842±0.013	0.611±0.030	0.602±0.031	0.691±0.022	0.792±0.009	0.654±0.016	0.965±0.00
CoTr	0.622±0.003	0.656±0.004	0.486±0.005	0.450±0.005	0.753±0.005	0.881±0.003	0.688±0.007	0.680±0.007	0.737±0.003	0.826±0.002	0.706±0.004	0.691±0.00
RMSC	0.610±0.016	0.648±0.012	0.473±0.015	0.437±0.016	0.758±0.011	0.884±0.004	0.698±0.010	0.690±0.011	0.754±0.002	0.831±0.002	0.716±0.003	0.702±0.00
CSMSC	0.766±0.036	0.782±0.020	0.645±0.032	0.621±0.035	0.816±0.026	0.917±0.010	0.774±0.025	0.768±0.026	0.732±0.035	0.832±0.016	0.694±0.029	0.677±0.03
LTMSC	0.737±0.009	0.760±0.007	0.618±0.012	0.593±0.013	0.791±0.023	0.902±0.010	0.739±0.024	0.732±0.025	0.706±0.029	0.809±0.016	0.668±0.025	0.650±0.02
LMSC	0.667±0.017	0.689±0.015	0.502±0.022	0.466±0.023	0.801±0.033	0.906±0.020	0.745±0.046	0.739±0.047	0.730±0.027	0.835±0.016	0.697±0.025	0.680±0.02
MCIAS	0.837±0.039	0.830±0.026	0.706±0.042	0.686±0.046	0.872±0.015	0.931±0.007	0.824±0.016	0.820±0.016	0.886±0.025	0.951±0.007	0.871±0.024	0.863±0.02
MLAN	0.703±0.000	0.717±0.000	0.547±0.000	0.515±0.000	0.727±0.000	0.838±0.000	0.509±0.000	0.494±0.000	0.775±0.000	0.855±0.000	0.740±0.000	0.726±0.00
GMC	0.654±0.000	0.689±0.000	0.480±0.000	0.441±0.000	0.632±0.000	0.857±0.000	0.359±0.000	0.336±0.000	0.791±0.000	0.940±0.000	0.794±0.000	0.781±0.00
WTNM	0.957±0.034	0.964±0.017	0.936±0.034	0.932±0.036	0.977±0.015	0.993±0.004	0.977±0.015	0.976±0.015	0.816±0.001	0.903±0.000	0.812±0.000	0.802±0.00
tSMC	0.913±0.044	0.919±0.027	0.857±0.050	0.848±0.053	0.974±0.013	0.992±0.003	0.974±0.013	0.973±0.013	0.825±0.012	0.902±0.003	0.817±0.009	0.808±0.01
ETMC	0.629±0.018	0.668±0.015	0.504±0.020	0.470±0.022	0.717±0.011	0.857±0.005	0.640±0.013	0.631±0.013	0.861±0.017	0.927±0.008	0.851±0.015	0.843±0.01
UGLTL	0.993±0.000	0.992±0.000	0.987±0.000	0.986±0.000	0.967±0.000	0.989±0.000	0.960±0.000	0.959±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.00
LTBPL	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.00

表 3　对比结果：在 UCI、Caltech-101 和 Notting-Hill 数据集上通过不同方法得到的均值和标准差

Datasets	UCI				Caltech-101				Notting-Hill			
Method	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI
SC 1	0.617±0.011	0.585±0.007	0.506±0.010	0.451±0.011	0.356±0.001	0.165±0.001	0.313±0.001	0.634±0.000	0.694±0.000	0.663±0.000	0.673±0.000	0.856±0.000
SC 2	0.684±0.000	0.587±0.001	0.554±0.000	0.504±0.001	0.404±0.002	0.271±0.001	0.438±0.000	0.691±0.000	0.844±0.000	0.670±0.000	0.772±0.000	0.901±0.000
SC 3	0.546±0.005	0.489±0.001	0.427±0.001	0.362±0.001	0.328±0.002	0.217±0.001	0.334±0.001	0.644±0.000	0.740±0.000	0.623±0.000	0.673±0.000	0.859±0.000
SC 4	—	—	—	—	0.379±0.014	0.385±0.010	0.435±0.010	0.699±0.005	—	—	—	—
SC 5	—	—	—	—	0.355±0.014	0.307±0.005	0.396±0.018	0.677±0.008	—	—	—	—
SC 6	—	—	—	—	0.481±0.001	0.343±0.001	0.460±0.001	0.705±0.000	—	—	—	—
CoTr	0.840±0.014	0.796±0.005	0.779±0.009	0.754±0.010	0.437±0.003	0.423±0.005	0.473±0.005	0.326±0.006	0.843±0.010	0.781±0.005	0.823±0.006	0.773±0.008
RMSC	0.859±0.018	0.822±0.009	0.800±0.014	0.777±0.016	0.460±0.000	0.394±0.000	0.483±0.000	0.332±0.000	0.827±0.000	0.772±0.000	0.822±0.000	0.772±0.000
CSMSC	0.882±0.000	0.787±0.001	0.784±0.001	0.760±0.001	0.630±0.010	0.534±0.021	0.632±0.017	0.494±0.018	0.873±0.000	0.760±0.000	0.795±0.000	0.736±0.000
LTMSC	0.800±0.006	0.768±0.007	0.748±0.009	0.720±0.010	0.602±0.000	0.547±0.000	0.613±0.000	0.475±0.000	0.868±0.000	0.779±0.000	0.825±0.000	0.777±0.000
LMSC	0.856±0.087	0.783±0.027	0.762±0.089	0.736±0.044	0.564±0.029	0.448±0.020	0.564±0.025	0.410±0.026	0.913±0.051	0.833±0.058	0.866±0.078	0.829±0.099
MCIAS	0.976±0.001	0.946±0.003	0.953±0.002	0.948±0.003	0.742±0.048	0.535±0.038	0.735±0.054	0.591±0.072	0.523±0.066	0.362±0.078	0.460±0.060	0.294±0.082
MLAN	0.968±0.000	0.925±0.000	0.937±0.000	0.930±0.000	0.627±0.004	0.544±0.003	0.618±0.000	0.428±0.003	0.365±0.000	0.114±0.000	0.376±0.000	0.044±0.000
GMC	0.735±0.000	0.815±0.000	0.713±0.000	0.677±0.000	0.692±0.000	0.659±0.000	0.721±0.000	0.594±0.000	0.312±0.000	0.092±0.000	0.369±0.000	0.022±0.000
WTNM	0.996±0.000	0.990±0.000	0.993±0.000	0.992±0.000	0.685±0.000	0.668±0.000	0.702±0.000	0.587±0.000	0.983±0.000	0.956±0.000	0.975±0.000	0.968±0.000
tSMC	0.996±0.000	0.989±0.000	0.992±0.000	0.991±0.000	0.746±0.000	0.724±0.002	0.758±0.001	0.656±0.001	0.956±0.000	0.890±0.000	0.917±0.000	0.895±0.000
ETMC	0.933±0.015	0.961±0.007	0.939±0.013	0.932±0.014	0.514±0.010	0.535±0.005	0.559±0.006	0.425±0.007	0.951±0.000	0.911±0.000	0.924±0.000	0.898±0.000
UGLTL	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.383±0.000	0.621±0.000	0.499±0.000	0.355±0.000	0.950±0.000	0.921±0.000	0.924±0.000	0.903±0.000
LTBPL	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.945±0.000	0.894±0.000	0.953±0.000	0.924±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000

表 4　对比结果：在 Hdigit、BBCSport 和 BBC4view 数据集上通过不同方法得到的均值和标准差

Datasets	Hdigit				BBCSport				BBC4view			
Method	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI	ACC	NMI	Fscore	ARI
SC 1	0.526±0.000	0.468±0.001	0.412±0.000	0.345±0.000	0.846±0.000	0.672±0.001	0.760±0.001	0.688±0.001	0.581±0.001	0.413±0.001	0.504±0.002	0.357±0.003
SC 2	0.485±0.009	0.455±0.007	0.391±0.008	0.322±0.008	0.509±0.001	0.229±0.002	0.416±0.000	0.164±0.001	0.777±0.000	0.542±0.001	0.652±0.000	0.550±0.001
SC 3	—	—	—	—	—	—	—	—	0.636±0.004	0.439±0.001	0.523±0.002	0.384±0.003
SC 4	—	—	—	—	—	—	—	—	0.725±0.000	0.500±0.000	0.580±0.000	0.460±0.000
CoTr	0.910±0.004	0.826±0.002	0.845±0.003	0.828±0.003	0.902±0.003	0.809±0.005	0.875±0.003	0.837±0.004	0.543±0.001	0.386±0.002	0.464±0.005	0.272±0.002
RMSC	0.729±0.008	0.672±0.010	0.639±0.010	0.598±0.011	0.851±0.028	0.801±0.015	0.848±0.020	0.801±0.027	0.708±0.019	0.533±0.005	0.590±0.009	0.470±0.011
CSMSC	0.834±0.000	0.738±0.000	0.731±0.000	0.701±0.000	0.955±0.000	0.861±0.000	0.914±0.000	0.888±0.000	0.919±0.000	0.775±0.000	0.861±0.000	0.819±0.000
LTMSC	0.782±0.000	0.661±0.000	0.649±0.000	0.609±0.000	0.943±0.000	0.839±0.000	0.907±0.000	0.878±0.000	0.927±0.000	0.796±0.000	0.873±0.000	0.834±0.000
LMSC	0.802±0.000	0.796±0.000	0.758±0.000	0.730±0.000	0.920±0.002	0.839±0.005	0.901±0.004	0.870±0.005	0.874±0.007	0.680±0.014	0.790±0.011	0.726±0.014
MCIAS	0.728±0.001	0.831±0.004	0.763±0.005	0.734±0.006	0.891±0.046	0.818±0.089	0.883±0.080	0.846±0.040	0.860±0.089	0.705±0.080	0.800±0.084	0.739±0.050
MLAN	0.710±0.000	0.837±0.000	0.762±0.000	0.731±0.000	0.977±0.000	0.923±0.000	0.953±0.000	0.938±0.000	0.871±0.000	0.700±0.000	0.810±0.000	0.746±0.000
GMC	0.998±0.000	0.993±0.000	0.996±0.000	0.995±0.000	0.739±0.000	0.795±0.000	0.720±0.000	0.600±0.000	0.693±0.000	0.562±0.000	0.633±0.000	0.478±0.000
WTNM	0.998±0.000	0.996±0.000	0.997±0.000	0.997±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.995±0.000	0.983±0.000	0.993±0.000	0.991±0.000
tSMC	0.997±0.000	0.991±0.000	0.994±0.000	0.993±0.000	0.998±0.000	0.992±0.000	0.997±0.000	0.996±0.000	0.994±0.000	0.977±0.000	0.990±0.000	0.987±0.000
ETMC	0.917±0.018	0.962±0.008	0.932±0.014	0.924±0.016	0.964±0.027	0.976±0.019	0.972±0.028	0.963±0.030	0.906±0.048	0.899±0.019	0.911±0.032	0.884±0.042
UGLTL	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	0.705±0.000	0.839±0.000	0.783±0.000	0.723±0.000	0.845±0.000	0.754±0.000	0.748±0.000	0.678±0.000
LTBPL	0.999±0.000	0.999±0.000	0.999±0.000	0.999±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000	1.000±0.000



(a) Intensity feature　(b) LBP feature　(c) Gabor feature　(d) LTBPL

图 2　Yale 数据集上的可视化

表 5　消融实验：LTBPL 及其变体在 NMI 方面的比较结果

Variants	Yale	ORL	COIL-20	UCI	Caltech-101	Notting-Hill	Hdigit	BBCSport	BBC4view	Average
LTBPL	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.894_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.999_{\pm 0.000}$	$1.000_{\pm 0.000}$	$1.000_{\pm 0.000}$	0.988
LTBPL-t1	$0.642_{\pm 0.000}$	$0.801_{\pm 0.000}$	$0.974_{\pm 0.000}$	$0.821_{\pm 0.000}$	$0.619_{\pm 0.000}$	$0.623_{\pm 0.000}$	$0.991_{\pm 0.000}$	$0.825_{\pm 0.000}$	$0.639_{\pm 0.000}$	0.770
LTBPL-t2	$0.916_{\pm 0.022}$	$0.987_{\pm 0.006}$	$0.967_{\pm 0.009}$	$0.998_{\pm 0.000}$	$0.664_{\pm 0.001}$	$0.970_{\pm 0.000}$	$0.995_{\pm 0.000}$	$1.000_{\pm 0.000}$	$0.988_{\pm 0.000}$	0.942

的先验知识通过赋予不同权重被显示地考虑。同时，视图对应的相似性表征和反映最终聚类的共识聚类指示矩阵能够通过不同的自适应置信度关联起来。在九个真实世界数据集的广泛实验表明了和最先进的聚类方法相比，所提出的 LTBPL 方法有明显的优越性。对于未来的工作，张量学习将被用于解决不完整多视图聚类里面具有挑战性的问题，其中视图中的样本或特征可能存在缺失。相关工作请参考中山大学团队在 IEEE TKDE 2022 等的论文[1, 19, 20]。

责任编委　崔海楠　王金甲

参考文献

[1] M.-S. Chen, C.-D. Wang, and J.-H. Lai, “Low-rank tensor based proximity learning for multi-view clustering.” IEEE Trans. Knowl. Data Eng.,2022.

[2] C. Tang, X. Zhu, X. Liu, M. Li, P. Wang, C. Zhang, and L. Wang, “Learning a joint affinity graph for multiview subspace clustering,” IEEE Trans. Multimedia, vol. 21, no. 7, pp. 1724–1736, 2019.

[3] H. Wang, Y. Yang, and B. Liu, “GMC: graph-based multi-view clustering,” IEEE Trans. Knowl. Data Eng., vol. 32, no. 6, pp. 1116–1129, 2020.

[4] A. Kumar, P. Rai, and H. Daum´e III, “Co-regularized multi-view spectral clustering,” in NIPS, 2011, pp. 1413–1421.

[5] X. Wang, X. Guo, Z. Lei, C. Zhang, and S. Z. Li, “Exclusivity-consistency regularized multi-view subspace clustering,” in CVPR, 2017, pp. 1–9.

[6] C. Zhang, H. Fu, Q. Hu, X. Cao, Y. Xie, D. Tao, and D. Xu, “Generalized latent multi-view subspace clustering,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 42, no. 1, pp. 86–99, 2020.

[7] X. He and P. Niyogi, “Locality preserving projections,” in NIPS, 2003, pp. 153–160.

[8] S. T. Roweis and L. K. Saul, “Nonlinear dimensionality reduction by locally linear embedding,” Science, vol. 290, no. 5500, pp. 2323–2326, 2000.

[9] Z. Zhao, L. Wang, H. Liu, and J. Ye, “On similarity preserving feature selection,” IEEE Trans. Knowl. Data Eng., vol. 25, no. 3, pp.619–632, 2013.

[10] R. Vidal, “Subspace clustering,” IEEE Signal Processing Magazine, vol. 28, no. 2, pp. 52–68, 2011.

[11] Y. Xie, D. Tao,W. Zhang, Y. Liu, L. Zhang, and Y. Qu, “On unifying multi-view self-representations for clustering by tensor multi-rank minimization,” Int. J. Comput. Vis., vol. 126, no. 11, pp. 1157–1179, 2018.

[12] Q. Gao, W. Xia, Z. Wan, D. Xie, and P. Zhang, “Tensor-svd basedgraph learning for multi-view subspace clustering,” in AAAI, 2020, pp. 3930–3937.

[13] F. Nie, J. Li, and X. Li, “Self-weighted multiview clustering with multiple graphs,” in IJCAI, 2017, pp. 2564–2570.

[14] F. Nie, G. Cai, and X. Li, “Multi-view clustering and semi-supervised classification with adaptive neighbours,” in AAAI, 2017, pp. 2408-2414.

[15] K. Fan, “On a theorem of weyl concerning eigenvalues of linear transformations i,” Proceedings of the National Academy of Sciences, vol. 35, no. 11, pp. 652–655, 1949.

[16] J. Wu, X. Xie, L. Nie, Z. Lin, and H. Zha, “Unified graph and low-rank tensor learning for multi-view clustering,” in AAAI, 2020, pp. 6388–6395.

[17] J.Wu, Z. Lin, and H. Zha, “Essential tensor learning for multi-view spectral clustering,” IEEE Trans. Image Process., vol. 28, no. 12, pp. 5910–5922, 2019.

[18] X.- L. Li, M.-S. Chen, and C.-D. Wang, et al. Refining graph structure for incomplete multi-view clustering. IEEE Transactions on Neural Networks and Learning Systems, 2022.

[19] M.-S. Chen, J.-Q. Lin, X.-L. Li, B.-Y. Liu, C.-D. Wang, D. H, and J.-H. L, “Representation learning in multi-view clustering: A literature review,” Data Science and Engineering, 2022:1-17.

陈曼笙

中山大学计算机学院 2022 级博士研究生，导师为王昌栋副教授。在国际期刊和会议上发表了八篇论文，包括 IEEE TKDE、IEEE TCYB、IEEE TNNLS、Information Fusion、KDD、ACM MM、AAAI 和 DASFAA。主要研究方向是多视图聚类。

Email:chenmsh27@mail2.sysu.edu.cn

王昌栋

中山大学计算机学院副教授，博士生导师，师从中山大学赖剑煌教授和美国伊利诺大学-芝加哥校区 IEEE Fellow Philip S. Yu 教授。研究方向包括数据聚类、网络分析、推荐算法和大数据信息安全。以第一作者身份或者指导学生发表了 100 余篇 CCF B 类或中科院分区表 SCI 二区以上的学术论文，其中 IEEE/ACM Trans 超过 40 篇，A 类或一区论文 50 余篇。主持了包括广东省自然科学基金-杰出青年基金、广东特支计划“科技创新青年拔尖人才”、国家重点研发计划项目-子课题、国家自然科学基金-面上项目、CCF-腾讯犀牛鸟科研基金等 13 个项目。任人工智能权威期刊 JAIR 的副编辑。

Email: changdongwang@hotmail.com

赖剑煌

中山大学计算机学院教授、博士生导师。广东省信息安全技术重点实验室主任，视频图像智能分析与应用公安部重点实验室学术委员会主任。中国图象图形学学会副理事长、会士，自动化学报副主编，中国计算机学会杰出会员，中国计算机学会计算机视觉专业组副主任(第一、二届)。广东省图像图形学会理事长(第四、五届)，广东省人工智能与机器人学会副理事长、广东省安防协会人工智能专委会主任。IEEE 高级会员。已主持承担国家自然科学基金与广东联合重点项目，科技部科技支撑课题，国家自然科学基金、广东省前沿与关键技术创新专项等多项，获得广东省自然科学一等奖(2018 年)、中国图象图形学学会自然科学一等奖(2020 年)、广东省自然科学二等奖(2020 年)、广东省科学技术奖励二等奖(2016 年)等。已发表了 200 多篇学术论文，主要发表在 IEEE TPAMI、IJCV 、IEEE TIP、IEEE TNN、IEEE TCSVT、IEEE TSMC(Part B)、Pattern Recognition 等国际权威刊物以及 ICCV、 CVPR、 ICDM 等专业重要学术会议上。拥有 30 多项国家发明专利。

Email: stsljh@mail.sysu.edu.cn

热点追踪

噪声关联学习

四川大学 杨谋星 林义杰 黄振宇 彭玺

一、引言

深度神经网络的成功依赖于大规模且高质量的标记数据。作为标签的一种重要形式，数据点间的关联/对齐关系在跨模态检索、视觉定位、图像自动描述、目标重识别、图匹配、机器阅读、对比学习等应用和学习范式中至关重要。

在实际场景中，为获得关联的成对数据，通常采用人工标注或者互联网爬取的方式。然而，由于数据繁杂和人力资源受限，关联的准确性往往难以保证，数据点间不可避免地存在噪声关联(Noisy Correspondence)[1,2]。如图 1 所示，噪声关联的样本对通常分为两类，一类是假阳性样本对(False Positive Pairs，FP)，即本来没有关联的数据对被错误当做有关联的正样本对[1]，如生活中常见的图文不符、音画不同步、答非所问现象；另一类是假阴性样本对（False Negative Pair，FN)，即描述相同或相关目标的数据点被错误当做没有关联的负样本对[2]。

需要注意的是，噪声关联可以认为是噪声标签（Noisy Label) 学习领域的一种新范式，其与注重于分类任务中错误类别标签的工作有着显著区别：i) 噪声关联指样本间的相关性/关系可能存在错误，而噪声标签主要强调样本的所属类别可能出错。从该角度出发，大部分需要成对样本作为输入的任务和应用都可能存在噪声关联，而噪声标签主要局限于传统的分类任务；ii) 噪声关联并不是非正即负，对于给定样本对，他们之间的关联其实是[0,1]之间的连续值。相比之下，在分类任务中，噪声标签的样本通常属于某个特定的类。综上，噪声关联学习丰富了噪声标签学习范式的内涵，并扩展了其外延。

图 1　噪声关联示意图，其中所有样本均选自多模态 Conceptual Captions 数据集 [3]。上图：由于该数据集是从互联网自动爬取得到的图文数据对，正样本对中不可避免地存在部分假阳性样本对。神经网络在优化过程中将拟合假阳性样本对而得到错误决策边界，最终导致性能下降。下图：给定锚点样本，除了其关联的正样本对，负样本集合中存在一些潜在相关的假阴性样本，它们与锚点可能存在完全对应、抽象对应和部分对应等情形。错误地将这些假阴性样本对当作负样本对不仅将失去关联样本对的多样性，还将导致模型被错误优化。

从 2021 年开始，一些学者认识到了噪声关联问题的重要性并开展了一系列研究。[2]最早意识到噪声关联

问题的重要性，并提出了对假阴性鲁棒的对比学习损失函数；受此启发，[1]正式揭示和定义噪声关联问题，并以跨模态匹配为验证场景，提出了对假阳性样本对鲁棒的跨模态检索方法；[4-10]进一步深入挖掘噪声关联在目标重识别、图匹配等应用场景下的特殊性，并设计了场景定制化的解决方案。在后续章节，我们将简要介绍这些工作。具体地，第二节将给出噪声关联的形式化定义，包含假阴性和假阳性关联；第三节将介绍噪声关联学习在不同场景下的研究现状；第四节将总结全文，并给出噪声关联学习研究的未来展望。

二、噪声关联学习

本节介绍噪声关联的形式化定义，主要包含对假阳性和假阴性关联的定义。

2.1 假阳性关联（False Positive）

给定正样本对集合$\{(x_i^{m_1}, x_i^{m_2}), c_i\}_{i=1}^N$，其中$(x_i^{m_1}, x_i^{m_2})$代表第$i$个正样本对，其通过以下两种方式获得：i）人工标注或互联网爬取的成对数据；ii）通过类别标签构建，即同类的样本作为正样本对。理想情况下，关联$c_i = 1$，即$x_i^{m_1}$和$x_i^{m_2}$都描述同一个或同类的目标。然而，实际应用中将不可避免得到一些假阳性样本对，它们实为无关联或弱关联的样本对，但关联c_i被错误标记为1。以常见的图文数据对为例，经常会出现字幕冗余、字幕欠完备，甚至是图文完全不符等情况，这些都将导致假阳性关联。需要注意的是，$(x_i^{m_1}, x_i^{m_2})$呈现的形式多样。例如，在跨模态检索任务中，x_i代表图像、文本等实例；在视觉定位中，x_i代表目标框、或单词等细粒度对象；在图匹配中，x_i代表图像块。同理m_1和m_2可以相等也可以不等，意味着数据可能来自同一或不同模态。例如，在跨模态任务中，$m_1 \neq m_2$；在单模态任务如行人重识别中，$m_1 = m_2$。特别的，在视频表征学习或者图像匹配等任务中，每个样本对可能包含 3 个甚至更多的样本，例如视频包含了图像、文本、音频等。综上，假阳性关联广泛存在于不同应用中，且在不同应用下可能存在不同的定制化解决方案。

2.2 假阴性关联（False Negative）

给定所构建的负样本对集合$\{(x_i^{m_1}, x_j^{m_2}), c_{ij} \mid i \neq j\}$，其中负样本对$(x_i^{m_1}, x_j^{m_2})$有以下两种来源：i）通过类别标签构建，即非同类的样本作为负样本对；ii）数据集或同批次（Batch）内随机采样获得。理想情况下，$x_i^{m_1}$和$x_j^{m_2}$将是对于不同类或不同目标的描绘，它们之间没有相关性，因而关联c_{ij}被记为 0。然而，实际应用中，上述两种构建方式均可能引入假阴性样本对。具体地，当样本的类别标签存在错误时，方式 1 不可避免地会将本属于同一类的样本错误作为负样本对。同理，方式 2 也可能将同批次内语义相似的样本对错误作为负样本对，特别是目前主流的大批次对比学习将引入更多假阴性样本对。

三、不同应用下的噪声关联学习

3.1 跨模态检索

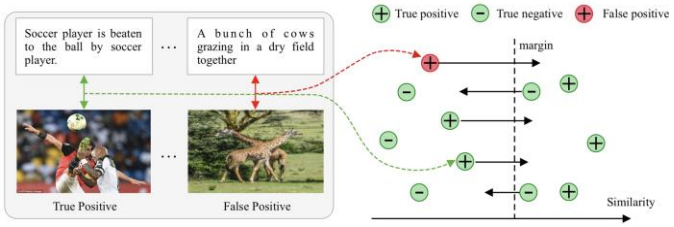


图 2　跨模态检索中的噪声关联。真阳性样本对正确地指导了跨模态匹配，但假阳性样本对则对训练的进了错误的监督。

跨模态检索大多依赖于正确匹配的跨模态数据，从而学习到一个可以衡量跨模态相似性的匹配模型。然而，在数据收集和标记过程中，常常引入噪声关联。如图 2 所示，给定的训练数据中图片和文本描述可能是错误匹配即假阳性关联的，这无疑会影响后续的跨模态匹配任务。为解决假阳性关联问题，[1]提出了基于神经网络记忆效应的噪声鉴别矫正方法，探明了神经网络对数据关联的拟合演化规律。该方法可自适应地识别噪声关联数据并对其关联进行矫正，结合所设计的鲁棒多模态匹配目标函数，最终实现假阳性关联鲁棒的多模态检索。

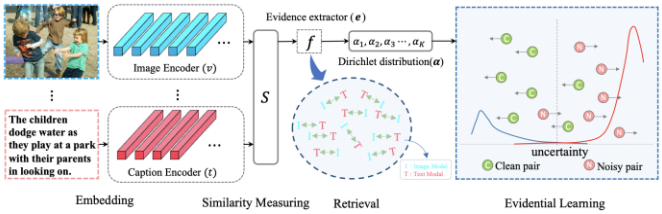


图 3　方案[4]将证据理论应用于跨模态检索中，可估计样本对的不确定度。

[4]将证据学习应用于跨模态检索学习，考虑了噪声关联在跨模态中的不确定性估计问题和难负样本选择问题，提出了一个广义的深度证据跨模态学习框架。该方案能以有效和高效的方式提供可信的检索，同时可直接应用于现有的跨模态检索方法以增强鲁棒性。

表 1　不同方法在 Conceptual Captions [3]子集上多模态检索召回率。

方法	Image → Text			Text → Image		
	R@1	R@5	R@10	R@1	R@5	R@10
SCAN[11]	30.5	55.3	65.3	26.9	53.0	64.7
SAF[12]	31.7	59.3	68.2	31.9	59.0	67.9
SGR[12]	11.3	29.7	39.6	13.1	30.1	41.6
NCR[1]	**39.5**	64.5	73.5	40.3	64.6	73.2
DECL[4]	39.0	**66.1**	**75.5**	**40.7**	**66.3**	**76.7**

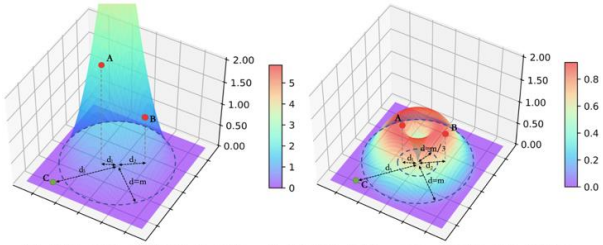
如表 1 所示，处理噪声关联的方法[1,4]在互联网爬取的噪声多模态数据集上，取得了显著的检索性能提升，充分说明解决噪声关联在实际应用下具备重要意义。

3.2 多视图表示学习

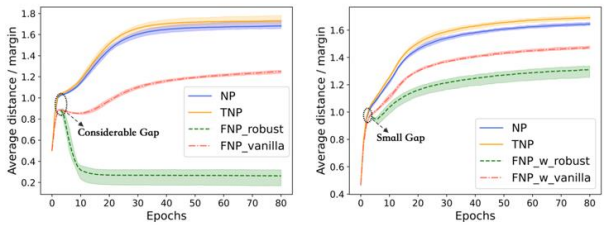
[2]观察到了对比学习范式的随机负样本选择策略将不可避免地将相同语义、属于同类的样本对错误作为负样本对，即引入假阴性，最终导致模型性能退化。为此，该工作设计了一种假阴性鲁棒的对比损失函数，该损失函数具有良好的性质及相应数学证明，能够缓解或甚至避免错误拟合假阴性样本对。为验证该损失的有效性，[2]以多视图表示学习为应用场景，验证了所提出鲁棒对比学习的有效性，如图 4(b)所示。

3.3 跨模态行人重识别

给定一张可见光/红外光模态的行人照片，跨模态



(a) 损失函数性能曲面（左：朴素对比损失，右：鲁棒对比损失）



(b) 不同数据数据集下对假阴性样本对的鲁棒性

图 4　相比朴素对比损失函数（a 图左），[2]提出的鲁棒对比损失具备非单调优化性质，能够缓解或甚至避免错误拟合假阴性样本对。如 b 图所示，使用朴素对比损失将把假阴性样本当作负样本对，错误地增加样本对距离；使用鲁棒对比损失将缓解假阴性样本对距离的错误增加，甚至把假阴性样本对正确地作为正样本对优化。

行人重识别(VI-ReID)旨在从另一模态中匹配出同一行人的其他相片。目前主流的 VI-ReID 方案，需要在各模态进行判别性学习，同时依赖身份标注构建跨模态正负样本对，进一步执行跨模态相似性学习以提升性能。因此，现有 VI-ReID 范式严重依赖于身份标注的准确性。然而，监控系统中图像可识别性差，特别是丢失行人颜色信息的红外模态，精确标注所有行人的身份是不现实的。错误的行人身份标注将产生类别级的噪声标签，并进一步导致假阳性和假阴性噪声关联。

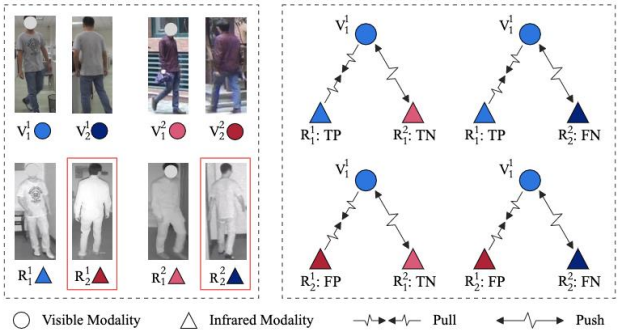


图 5　[5]指出 VI-ReID 任务会同时面临噪声标注（左）与噪声关联（右）问题。

论文[5]同时考虑了噪声标签与噪声关联问题，揭示了两种噪声之间的“孪生”关系，并针对性地提出了鲁棒的 VI-ReID 方案。其首先利用神经网络的记忆效应来估计身份标注的置信度；随后，基于置信度将跨模态正、负样本对分为不同子集并进一步校正其中的关联，最后利用所设计的双重鲁棒损失函数来实现对孪生噪声标签鲁棒的跨模态行人重识别。

3.4 图匹配

图匹配旨在寻找两个或多个存在复杂关系的集合间元素的对应关系。其最广为人知的应用场景是图像特征点的匹配，利用不同图像上匹配好的特征点，可以实现三维重建、目标追踪、运动结构理解等问题。类似分类问题中类别标签标注，图像中的关键点也是人工标注的。现实中的图像往往可见度低、图像间视角差异大，从而导致人工标注的关键点偏移、甚至错误，如图 6 所示。错误的关键点标注会导致图像间的特征点关联不正确，即噪声关联问题。不同于跨模态检索等应用中的噪声关联，图匹配任务会根据关键点构建一张图结构，并同时对齐两张或多张图中的点和边，此时噪声关联问题会同时存在于点与边的对应中。强迫图匹配网络拟合这种噪声，会显著降低其性能。

图 6　图匹配中的噪声关联问题示意，其中绿线表示正确关联，红线表示错误关联。由于可见度低、视角差异大，关键点的标注存在错误，导致噪声关联。

为同时解决噪声情况下点与边的匹配问题，论文[6]提出了一个鲁棒的图匹配损失函数，其将点的对齐与边的对齐形式化为一个二阶噪声关联问题。进一步地，该方法构造了一个动量网络以预测出点与点、边与边之间的匹配置信度，再按照置信度加权给匹配损失，实现了噪声关联下鲁棒的图匹配。

3.5 会话式机器阅读

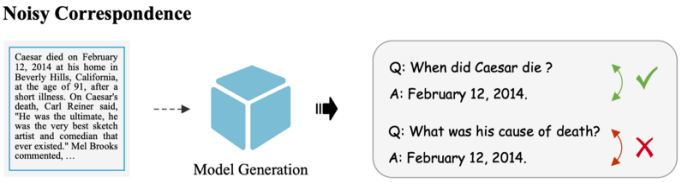


图 7　噪声关联数据可以是模型自身产生的两个数据点之间的不对齐。

会话式机器阅读理解（Machine Reading Comprehension，MRC）高度依赖于与给定文档相关的回答对，而对于无标签数据，通常会借助一个个预训练好的问题-答案生成器去自动生成伪标签数据，再去 finetune 一个预训练好的 MRC 模型。如图 7 所示，这样的方式不可避免的会遇到噪声关联问题，即构造的伪问答对是错误关联的。这显然极大损害了后续模型优化。

为此，[7]提出鲁棒的机器阅读理解域迁移方法，可以有效的缓解噪声问答对的产生。该方法包含了 QA Construction Model 和 MRC Model，前者用于在目标域构造问答对来 finetune 后者。为了解决噪声关联问题，该方案提出了一种强化自训练方法来将 MRC Model 对构造的问答对的评估作为反馈去优化 QA Construction Model，从而减少了噪声关联数据的产生。该方法被成功用于支付宝智能客服系统中，在“双十一”、“双十二”和“新春红包”等多项营销活动显著提升了智能问答的准确性。

3.6 多模态预训练

多模态预训练旨在从大规模图像-文本对中学习多模态表示，以改进下游的视觉-语言任务，例如图文检索、视觉问答、自然语言视觉推理、视觉定位等。大规模数据的预训练往往能够给下游任务提供很好的初始化和性能保障。

目前的多模态预训练所利用的数据规模从百万到十亿不等，这些海量的图像-文本数据均是通过互联网爬取得到的，包含大量的噪声关联数据。虽然预训练的数据规模足够，但是其中包含的噪声数据仍然会对模型的性能造成不利影响。为此，许多多模态预训练[8-10]尝试从噪声处理的角度来优化预训练,他们的实验也证

实了显式处理噪声关联能够提升下游任务的性能。

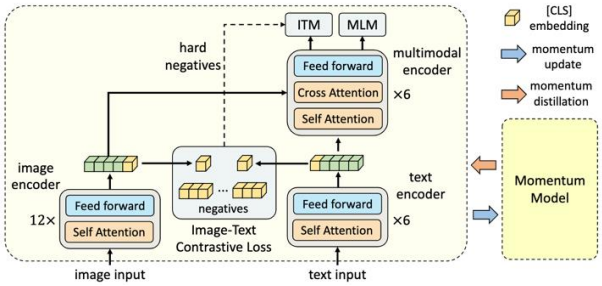


图 8　方法[8]利用动量模型作为教师，矫正对比学习（Image-text Contrastive）和掩码语言建模(Masked Language Model, ITM)中的噪声关联。

方法[8]和[10]通过蒸馏的方式，在多模态学习中同时减缓噪声关联的影响。[8]改进了对比学习和掩码语言建模技术，通过自蒸馏的方式来矫正原本的错误对应关系。[10]进一步提出渐进蒸馏方法，其在训练过程中选择部分样本（α%）按照给定的关联进行训练，同时剩余样本（1-α%）按照教师模型提供的相似度进行训练。随着训练进行，该方法逐渐依靠自身的预测来实现自我提升，不断地减少依靠给定监督信号的数据比例α，从而减缓对噪声的拟合。

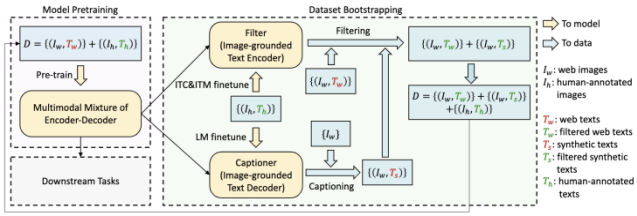


图 9　方法[9]通过对数据进行清理来去除噪声。其提出了一个标题生成器，用于为图像生成关联的标题；以及一个过滤器，用于去除噪声的图像-文本对。

相较于前两篇从鲁棒训练的角度解决噪声关联问题，[9]提出了一种新颖的数据清理方式，利用语言模型对原本的数据进行刷新从而解决噪声关联。在清洗过程中，该方法利用二分类损失判断图-文是否匹配，将不匹配的图文对通过语言模型实现关联文本的生成。最后利用清洗后的数据集在进行训练。

3.7 视听动作识别

视听动作识别旨在从视频片段中分辨不同的动作。目前视频听作识别通常使用多个模态的信息共同完成识别任务，此类方法高度依赖于数据的标签信息，要求视频信息和音频信息正确对应，才能完成动作识别。然而，在实际应用中，这样的条件常常无法满足。训练数据中常常包含噪声关联，即视频片段和音频片段不是对应的。

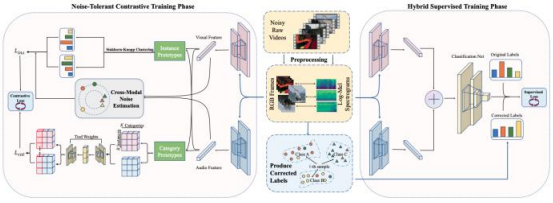


图 10　方法[13]利用噪声估计模块缓解噪声关联问题，并依赖跨模态的相似性修正噪声样本对。

为了解决这类问题，[13]提出了一种容噪学习框架，用于视听动作识别。简言之，[13]首先设计了一个跨模态的噪声估计模块去动态调整跨模态一致性，并用一个类别级的对比学习损失函数来进一步缓解噪声关联的负面影响。随后，[13]通过计算跨模态的相似性来修正噪声样本的关联信息，并将其作为一种补充信息进行模型训练。

四、总结与展望

本文介绍了噪声关联问题，调研了其在不同应用下的具象化存在以及对应的解决方案。噪声关联本质上是一种由于时空不同步所导致的数据错误关联现象，其广泛存在于不同应用和任务中。一旦使用噪声关联的数据去训练机器学习模型，即使是增大数据规模或模型容量也难以获得理想结果。

噪声关联学习赋予了传统噪声标签学习新内涵，可被视为噪声标签学习领域的一个新研究方向。目前针对该问题的研究还比较初步，后续有诸多改进点和探索点，例如：深入探索更多应用和任务下噪声关联的特殊性，设计应用定制化的解决方案；深入研究同时对假阳性和假阴性关联鲁棒的方法，避免模型过拟合假阳性样本对，同时增强正相关样本对的多样性；构建不同任务和应用下噪声关联的评估基准。

责任编委　储珺

参考文献

[1] Huang Z, Niu G, Liu X, et al. Learning with Noisy Correspondence for Cross-modal Matching, NeurIPS 2021.

[2] Yang M, Li Y, Huang Z, et al. Partially view-aligned representation learning with noise-robust contrastive loss, CVPR 2021.

[3] Sharma P, Ding N, Goodman S, et al. Conceptual captions: A cleaned, hypernymed, image alt-text dataset for automatic image captioning, ACL 2018.

[4] Qin Y, Peng D, Peng X, et al. Deep Evidential Learning with Noisy Correspondence for Cross-modal Retrieval, ACMMM 2022.

[5] Yang M, Huang Z, Hu P, et al. Learning with Twin Noisy Labels for Visible-Infrared Person Re-Identification, CVPR 2022.

[6] Lin Y, Yang M, Yu J, et al. Graph Matching with Bi-level Noisy Correspondence.

[7] Jiang L, Huang Z, Liu J, et al. Robust Domain Adaptation for Machine Reading Comprehension, AAAI 2023.

[8] Li J, Selvaraju R., Gotmare A., et al. Align before fuse: Vision and language representation learning with momentum distillation, NeurIPS 2021.

[9] Li J, Li D, Xiong C, et al. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation[J]. ICML 2022.

[10] Andonian A, Chen S, et al. Robust Cross-Modal Representation Learning with Progressive Self-Distillation, CVPR 2022.

[11] Lee K H, Chen X, Hua G, et al. Stacked cross attention for image-text matching, ECCV 2018.

[12] Diao H, Zhang Y, Ma L, et al. Similarity reasoning and filtration for image-text matching, AAAI 2021.

[13] Han H, Zheng Q, Luo M, et al. Noise-Tolerant Learning for Audio-Visual Action Recognition, arXiv:2205.

杨谋星

四川大学计算机学院博士生，研究方向：噪声关联学习，多模态学习。
Email: yangmouxing@gmail.com

林义杰

四川大学计算机学院博士生，研究方向：噪声关联学习，多模态学习。
Email: linyijie.gm@gmail.com

黄振宇

四川大学计算机学院博士生，研究方向：噪声关联学习，多模态学习。
Email: zyhuang.gm@gmail.com

彭玺

四川大学计算机学院教授，研究方向：机器学习理论及其在多媒体分析、计算机视觉、自然语言处理中的应用。
Email: pengx.gm@gmail.com

热点追踪

Distance Correlation 在深度学习中的应用

威斯康星大学麦迪逊分校 甄行践

一、引言

当我们比较两个（或多个）神经网络的时候，我们通常更在意其在某些测试集上的表现，比如准确度或AUROC。但是其实我们更值得在意的是，这个网络学到了什么信息。比如我们都知道，现在 ViT 在绝大多数情况下比 ResNet 的识别准确度更高，但是当我们真地考虑 ViT 学习到的信息量是否比 ResNet 多的时候，我们没有一种非常好的方式。换句话说，我们比较不同网络的时候，需要一种可以被理解、被解释的方式“做减法”。这个时候，我们从统计学中知道，correlation（相关性）是一种被广泛应用于比较两个随机变量的度量，甚至当随机变量增多时，我们可以用 partial correlation 或者 conditional correlation 去掉某些随机变量对于剩下的随机变量的影响[1]。

但我们该如何将 correlation 引入深度学习呢？深度网络中，什么是我们的随机变量呢？

二、方法

我们可以将深度网络抽象为特征提取器+分类器的组合，于是将原始图像输入特征提取器之后，我们可以得到对应的特征，我们将这些特征理解为随机变量。对于同一张图片，如果我们使用不同的神经网络（比如 ViT 和 ResNet），我们可以提取出两个特征向量（x 和 y）。我们想比较 x 和 y 之间的相关性。

如果使用传统的 correlation，最首要的问题就是 x 和 y 的维度需要相同。虽然我们可以用不同的方法把 x 和 y 投影到同一个空间（CCA），但是在训练过程中的 CCA 比较难收敛。这时，我们想到可以使用 distance correlation。

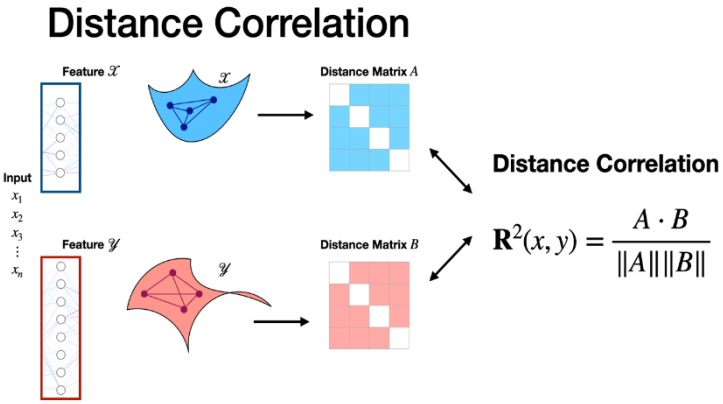


图 1　距离相关性

Distance correlation 与其说是比较 x 和 y 的相关性，不如说比较的是不同 x 之间的距离与对应的 y 之间的距离的相关性。比如我们有两张猫的图片，一张狗的图片，以及 ViT 和 ResNet 作为我们的特征提取器。直觉上，我们可以猜测两张猫的特征之间的距离在任何一种特征提取器之后都应该比较近，同时猫和狗的特征之间的距离都应该比较远。如果符合这种情况，两个特征提取器之间的相关性应当比较高；反之，如果这些特征之间的距离没有什么关系，那么相关性就应该比较低。这是 distance correlation 的基本想法，如图 1 所示，特别具体的证明请参照原文章[2,3]。

如果我们用 distance correlation，我们可以比较轻松地衡量两个神经网络之间的相关性。当我们想更进一步，比如我们想要衡量 ViT 比 ResNet 多学到了多少信息，我们可以将 ResNet 作为我们 correlation 的 condition，计算 partial distance correlation。

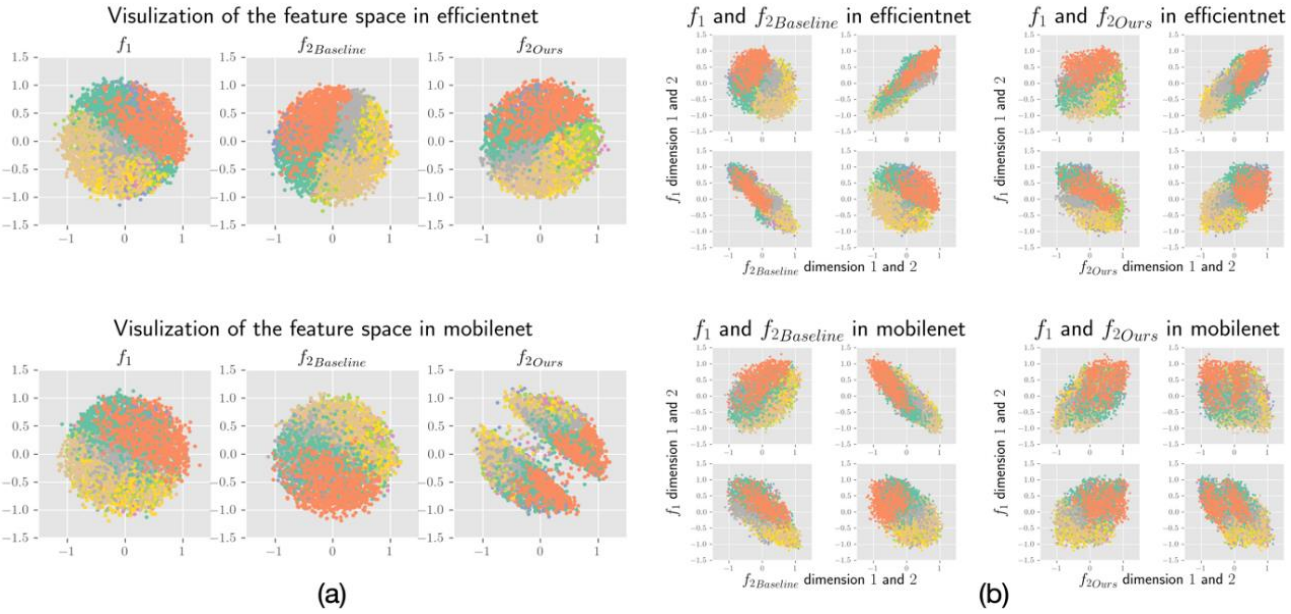


图 2　特征空间的 Picasso 可视化以及不同模型之间的相关性。（a）特征空间分布。（b）使用/不使用 DC 训练的 f_1 和 f_2 的特征空间之间的互相关。

三、应用

3.1. 提高组合网络鲁棒性（降低攻击图片的迁移能力）

现在的神经网络可能识别准确率很高，但是如果我们针对某种特定的神经网络，可以生成一些肉眼难以分辨但是神经网络无法正确识别的图像来攻击这种网络。一种非常简单直接的提高网络鲁棒性的方式是组合多个不同的神经网络。但是研究发现，在相似结构的神经网络之间，相同的攻击图片有很高的迁移概率，即相同的攻击图片针对不同但相似的神经网络，有很高的攻击成功概率。在这种情况下，即使我们组合了多个不同的神经网络，攻击图片仍然可以有较高的攻击成功几率。但是相对的，如果我们可以控制这些神经网络，使得他们学习的特征相互之间是独立的，我们可以假设攻击图片的迁移性会下降，这样组合的神经网络的鲁棒性可以得到提高。

我们比较轻松地发现，distance correlation 可以用于降低不同子网络之间的相关性，使得其彼此独立。

在文章中，我们更多地关注子网络之间的攻击迁移性。我们训练了 2 个相同结构的神经网络。第一种情况下我们不做任何操作，只让权重的初始值不同。第二种情况下我们在训练中，保持神经网络之间互相独立（最小化 distance correlation）。

如图 2 所示，我们检查了分别在使用 distance correlation 和不使用 distance correlation 训练之后，不同模型之间的相关性，可以看出使用 distance correlation 的模型随机变量之间相关度更低（更接近圆形）。

我们发现针对相同的攻击图片，独立网络之间的迁移性下降了 6%-9%不等。

3.2. Disentanglement(解开 latent vector 间的耦合)

对于生成网络，我们可以简单理解为把 latent vector 通过网络转换成图片。通常情况下我们不理解 latent vector 各维度之间的关系。但是假如我们知道某张人脸图片是亚裔少年男性，而我们希望将其转换为亚裔青年女性的图片，如果我们不理解 latent vector 或者对 latent vector 没有限制，是很难实现这种目标的。当我们想要控制 latent vector，比如前 16 维对应人种，接下来 16 维对应年龄，之后 16 维对应性别，之后的所

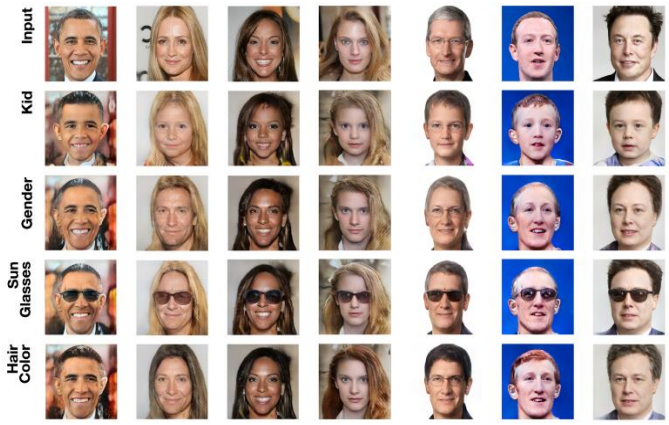


图 3　在 FFHQ 上的训练生成图像（这些结果仅使用 CLIP 的半监督数据集）

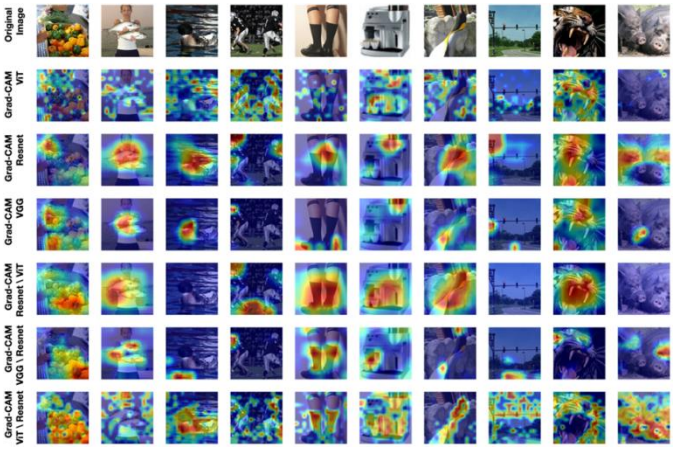


图 5　在 ImageNet 上使用 ViT、Resnet18 和 VGG16 获得 Grad-CAM 结果。

有信息都存在最后 128 维里（成为剩余信息），我们会希望剩余信息包含尽量少的人种年龄性别信息，换句话说，我们希望人种年龄性别对应的 latent vector 和剩余信息对应的 latent vector 相互独立。因此，我们可以很轻易地使用 distance correlation 来实现目标。

我们使用方法[4]，生成了一些高清的人脸图片，如图 3 所示，并且可以控制生成图片的人种年龄等信息。

3.3. 网络信息的比较

这是最重要的一部分实验，也是我们最独特的贡献点：回答“ViT 比 ResNet 的信息更多吗？”

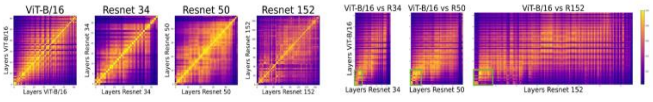


图 4　不同层之间相关性（a）左 4 是单模型中各层之间相似性。(b) 右 3 是 ViT 和 Resnets 层之间相似性。

首先我们可以使用 distance correlation 来比较同个网络不同层之间的相关性，以及不同网络不同层之间的相关性，如图 4 所示。更重要的是，我们可以使用 partial distance correlation 来回答上述问题。

我们首先使用预训练好的 BERT 来对 ImageNet 的 1000 个不同的类的名字进行 embedding，主要用于衡量不同类之间的相似程度（比如猫和老虎距离就应该比较近，而猫和飞机的距离就应该比较远。）

其次我们把 ResNet 提取出来的特征作为 condition，计算 ViT 提取的特征与文字特征之间的相关性。并且反过来将 ViT 作为 condition，计算 ResNet 提取的特征和文字特征的相关性。

实验结果如图 5 所示。结论是 ViT 在剔除 ResNet 的信息后的相关性高于 ResNet 剔除 ViT 的信息。

同时，我们也使用 Grad-CAM 来可视化检验剔除另一网络之后，当前网络更聚焦在图片的什么位置。我们发现 ViT 在剔除 ResNet 之后仍然可以“看清”图片的细微处，而 ResNet 在剔除 ViT 之后聚焦点比较散乱。

四、总结

我们将 distance correlation 和 partial distance correlation 引入深度学习，并且展示了三种完全不同的实验以证明其优势和潜力。对于 distance correlation 其他方向的挖掘，我们认为有很广阔的空间可以探索。比如最直接的 fairness，可以让网络学出的特征对于某些特定的 attribute 独立（比如收入预测独立于人种等等）；或者使用 partial distance correlation 替代 linear regression 来剔除网络信息等。

责任编委　王金甲

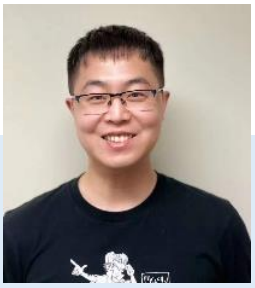
参考文献

[1] Xingjian Zhen, Zihang Meng, Rudrasis Chakraborty, and Vikas Singh: On the versatile uses of partial distance correlation in deep learning. European Conference on Computer Vision (ECCV), 2022.

[2] Gábor J. Székely, Maria L. Rizzo, and Nail K. Bakirov: Measuring and testing dependence by correlation of distances. The Annals of Statistics 35(6), 2769–2794, 2007.

[3] Gábor J. Székely, and Maria L. Rizzo: Partial distance correlation with methods for dissimilarities. The Annals of Statistics 42(6), 2382–2412, 2014.

[4] Aviv Gabbay, Niv Cohen, and Yedid Hoshen: An image is worth more than a thousand words: Towards disentanglement in the wild. Advances in Neural Information Processing Systems 34, 9216–9228, 2021.



甄行践

UW-Madison 计算机科学博士生，师从 Vikas Singh。在 NeurIPS，ICCV，CVPR，AAAI，ECCV 等高水平会议上发表学术论文 7 篇，其中 2 篇获的 CVPR oral，1 篇获的 ECCV 最佳论文奖。
Email: xzhen3@wisc.edu

顶会观察

ECCV 2022

东南大学 祁磊

欧洲计算机视觉国际会议（European Conference on Computer Vision, ECCV）是计算机视觉顶级会议之一，与 CVPR 和 ICCV 并称为计算机视觉领域三大顶会。ECCV 会议两年召开一次，与 ICCV 会议正好错开。今年大会的主席成员包括：来自摩德纳大学的 Rita Cucchiara、来自捷克理工大学的 Jiri Matas、来自 Mobileye 的 Amnon Shashua 和来自以色列理工学院的 Lihi Zelnik-Manor。值得关注的是，本届大会获得最佳论文奖的论文为 On the Versatile Uses of Partial Distance Correlation in Deep Learning。该论文的作者来自威斯康星大学麦迪逊分校和 Butlr，其中第一、二作者皆为华人，他们是 Xingjian Zhen 和 Zihang Meng，本科分别毕业于清华大学和浙江大学。本届 ECCV 大会于 2022 年 10 月 23 日至 10 月 27 日在以色列特拉维夫举办，包括 3 天的正会和 2 天的 Workshops & Tutorials。

一、 会议概况

自新冠疫情流行以来，ECCV 2022 首次线下举办，无法线下参会人员仍可选择线上参会。据主办方统计，截至大会开幕，有来自 76 个国家约 5000 人注册参会，其中约 3200 人现场参会，约 1800 人以线上方式参会。本届会议的所有与会者，无论是线上还是线下，都可以在会议平台上观看每篇论文时长为五分钟的演示视频。此外，亲临现场的作者们可以现场做 presentation，线上参会者只能观看，不做在线工作汇报。

大会的主席团成员（General Chairs）介绍了会议的具体安排：60 场 Workshops，13 场 Tutorials、2 场主题报告、2 场现场演示、2 场 Mentoring Events 和 1 场 Industry Track。大会的程序主席们(Program Chairs)则对 ECCV 2022 论文的审核情况作了详细介绍：在审稿过程中，有 846 篇投稿因各种原因在第一阶段被拒稿（Desk Rejected），其中很多是因为暴露了作者的身份，违反了双盲政策。在排除不同阶段被撤回的论文后，大会最后收稿数量为 5804 份。所有的这些投稿都至少收到了三篇评审意见(只有 6 篇文章只收到 2 条)，评审意见总数超过 15000 条。论文作者有机会提交 rebuttal，然后由领域主席（Area Chairs）和分配给每篇论文的审稿人进行讨论。为了确保尽可能公平公正，每篇论文的最终录用决定是由领域主席与一位协作领域主席讨论后作出的。此外，程序委员会主席监督整个审稿过程，尤其是在领域主席的决定与所有审稿人的意见存在明显分歧时，程序委员会主席会增加关注。

本届 ECCV 较为特别的是首次举办 Industry Track，其旨在通过组织一系列的特邀报告和 panel，来突出计算机视觉技术应用性质和所产生的广泛影响，以丰富本次大会的会议内容。Industry Track 于 10 月 26 日下午进行，包括一系列关于创业的受邀演讲和小组讨论，尤其是探讨如何跨越计算机视觉研究与其现实世界应用之间的界限。讲者包括以科研起家的公司创始人和 CEO，以及相关领域的投资人和其他专家。

二、 录用情况

ECCV 2022 收到的有效投稿和录用数量相比上一届都有大幅提高。大会共收到来自 18310 位作者的 6773 篇投稿，总共有 5804 份有效投稿，由 276 位 Area Chairs（AC）处理，并征求了 4719 位审稿人的意见。整个过程由 Program Chairs（PC）监督，并得到

General Chairs（GC）的持续支持。最终有 1645 篇（28%）论文被接收，包括 157 篇 Orals（2.7%）。相较于上一届，今年 ECCV 的投稿量提升 15.5%(779 篇)，同时录用数量也有所上升，录用论文数量提升为 20.9%（284 篇）。ECCV 2022 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、表征学习和深度学习、视频分析与理解、视觉与语言、迁移学习、计算摄影、姿态估计与跟踪、多任务学习、无监督学习、行为识别、自动驾驶等方向。值得关注的是，国内科研机构和企业在 ECCV 2022 上斩获颇丰。商汤共 70 篇论文入选，包含 6 篇 Oral，主要方向为自动驾驶、计算摄影、视频理解与分析、迁移学习、多模态等前沿研究和应用。旷视共有 20 篇论文入选，其中 3 篇 Oral，内容涵盖目标检测、3D 重建、图像复原等多个研究方向。腾讯则有 29 篇论文入选，内容涵盖人脸安全、图像分割、目标检测等研究方向。国外机构方面，谷歌在本次会议中有很不错的表现，共有 60 多篇论文入选。苹果和亚马逊则分别有 7 篇和 12 篇论文入选。

三、 热点论文

2022 年度 ECCV 最佳论文奖评审委员会由 7 名国际权威学者组成，其中包括两名华人学者：宾夕法尼亚州立大学 Yanxi Liu 教授和宾夕法尼亚大学 Jianbo Shi 教授。本年度大会共评选出了 1 篇最佳论文，2 篇最佳论文提名，论文具体介绍如下。

最佳论文：On the Versatile Uses of Partial Distance Correlation in Deep Learning[1]，来自威斯康星大学麦迪逊分校和 Butlr。目前神经网络模型的功能行为比较虽然取得了一些进展，但系统的功能比较，特别是不同网络之间的功能比较，仍然是很困难的，且往往是逐层进行的。本文重新审视了一种并不为人熟知的统计学的方法，称为距离关联方法，旨在评估不同维度特征空间之间的相关性。文章还描述了该方法适用于大规模模型的部署必要步骤，可以为一系列新应用的开发打开大门，比如可以调节一个深度模型与另一个模型的关系、学习不相干的表征以及优化不同的模型等。实验表明，一个多功能的正则器 (或约束) 具有许多优点，可以规避在这类分析中出现的一些常见困难。值得一提的是，该文章第一作者和第二作者皆为华人，第一作者 Xingjian Zhen 为威斯康星大学麦迪逊分校计算机科学博士生，曾于 2017 年获得清华大学电子工程系的学士学位；第二作者 Zihang Meng 现为 Meta AI（原 Facebook AI）研究科学家，曾于 2017 年获得浙江大学电子信息工程学士学位。

最佳论文提名 1：Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields[2]，来自图宾根大学、马克斯·普朗克计算机科学研究所和 Meta 现实实验室。众所周知，姿态或动作先验对于生成逼真的新姿态非常重要，对从有噪声或局部观察数据重建精确的姿态也非常重要。本文提出了一种基于神经距离场(NDFs)的人体姿态连续模型：Pose-NDF。该模型学习一系列各种可能的姿态作为神经隐式函数的零水平集（zero level set），将 3D 建模隐式曲面的思想扩展到高维域 $SO(3)^K$。与之前基于 VAE 的人体姿态先验（将姿态空间转换为高斯分布）相比，该研究对真实姿态流形进行建模，以保留姿态之间的距离。该方法在各种下游任务中优于 SOTA 方法，包括对真实世界的人体动作数据进行去噪、从遮挡数据中恢复姿态以及从图像中重建 3D 姿态。此外，与基于 VAE 的方法相比，Pose-NDF 可生成更多样化的姿态。

最佳论文提名 2：A Level Set Theory for Neural Implicit Evolution under Explicit Flows[3]，来自加利福尼亚大学圣迭戈分校。基于坐标的神经网络参数化隐式表面已经成为几何的有效表示，它们高效充当了参数水平集，其中零水平集定义了感兴趣的表面。本文提出了一个新框架，允许将三角形网格定义的变换操作应用于这类表面。这其中的部分操作中可以被视为在显式表面引起瞬时流场的能量最小化问题。该方法通过扩展水平集的经典理论，利用流场来实现参数化隐式表面变形。此外，通过形式化与水平集理论的关联，研究者分析认为现有的可微表面提取和渲染方法偏离了理论，并利用本文方法对表面平滑、平均曲率流、逆渲染和用户定义的隐式几何编辑等应用加以改进。

此外，本次大会有 157 篇论文入选 oral 环节，其中华人学者为第一作者的论文数量超过六成，多篇文章

也引起广泛关注。华中科技大学、约翰霍普金斯大学、字节跳动和牛津大学的论文 In Defense of Online Models for Video Instance Segmentation[4]提出了一个基于对比学习的视频实例分割 online 算法：IDOL，可学习更具有区分度的 instance embedding，并充分利用视频的历史信息来保证算法稳定性，将 online 模型表现提高到与 offline 模型相当甚至更高的水平。华为诺亚实验室的 CLIFF: Carrying Location Information in Full Frames into Human Pose and Shape Estimation[5]基于 HMR 网络结构提出新的动作捕捉算法，在网络输入和监督信号中引入裁剪框的全局位置，同时使用新的方法构造人体动作捕捉伪标注，降低估计误差 40%以上。谷歌、康奈尔大学和加州大学伯克利分校的 InfiniteNature-Zero: Learning Perpetual View Generation of Natural Scenes from Single Images[6]提出一种新型混合方法，在迭代渲染、优化和重复框架中集成了几何和图像合成功能，从而可采用单一图像，生成由数百个具有逼真和多样化内容的新视图组成的长相机轨迹，对永久视图生成（prepetual view generation）问题进行了进一步探索。

四、 大会获奖

Koenderink Prize. 也称时间检验奖（test of time），该奖旨在表彰计算机视觉领域的基础性贡献研究，获奖论文均为发表时间超过十年并经受住时间检验的研究。今年获奖论文共有两篇，第一篇是来自纽约大学、伊利诺伊大学厄巴纳-香槟分校和微软剑桥研究院 2012 年发表的论文 Indoor Segmentation and Support Inference from RGBD Images[7]。该文章提出了一种从 RGBD 图像中将典型、通常杂乱无章的室内场景解析为地板、墙壁、支撑面和对象区域，并恢复其支撑关系的方法，还创建了一个新颖的整数规划公式来推断物理支撑关系，并提供了一个包含 1449 张 RGBD 图像，捕获了 464 个具有详细注释的不同场景的新数据集。实验证明了该方法在复杂场景中推断支撑关系的能力，并验证了 3D 场景提示和推断支持能够实现更好的对象分割效果；第二篇是来自华盛顿大学、马普研究所、佐治亚理工学院 2012 年发表的论文 A Naturalistic Open Source Movie for Optical Flow Evaluation[8]。该文章基于虚拟 3D 动画短片生成一种新的光流数据集，以克服光流算法难以在现实数据上进行训练和测试的难题，并基于此数据集对主流光流算法进行有效评估，以促进光流算法的改进。文章开展的有效实验也证明了该数据集与现实场景数据集的相似性。

PAMI Everingham Prize. 该奖项旨在纪念英国计算机视觉领域专家、The PASCAL Visual Object Classes （VOC）数据集的主要贡献者以及该比赛项目的发起人 Mark Everingham，由 IEEE 计算机协会模式分析与机器智能（PAMI）技术委员会颁发，以表彰对计算机领域社区做出无私贡献的研究者或研究团队。本届获奖者是美国圣母大学计算机科学与工程副教授 Walter J. Scheirer，其主要研究领域包括开集识别、视觉识别的极值理论模型以及受生物启发的学习算法。此外，UCF101（2012 年）和 HMD51（2011 年）动作识别数据集团队的全体成员也获得该奖项。

Young Researcher Award. 该奖项由欧洲计算机视觉联盟（ECVA）设立，旨在鼓励年轻研究人员在计算机视觉方面的杰出研究成就。今年该奖项颁发给欧洲联邦理工学院(EPFL)计算机科学助理教授 Amir Zami。此外本次会议还补发了去年（2021 年）青年学者奖，图宾根大学计算机科学教授 Zeynep Akata 获此殊荣。

ECVA PhD Award. 该奖项旨在鼓励和表彰近两年完成博士学位论文的杰出研究者，通常来说每年会有两个获奖名额，并在接下来的 ECCV 会议上颁发。在本次大会中牛津大学的 Triantafyllos Afouras、摩德纳雷焦艾米利亚大学的 Marcella Cornia、慕尼黑工业大学的 Iro Laina 和马克斯·普朗克计算机科学研究所的 Yongqin Xiang 共四位研究者凭借 2020 年和 2021 年的博士学位论文获得该奖项。

五、 总结展望

本年度 ECCV 大会中 3D 视觉分析、目标检测（跟踪）、图像修复、图像分割、多模态分析、视觉 Transformer、无监督方法、对比学习、小样本学习等领域保持较高热度。综合近年来 ECCV 收录论文情况来看，

会议越发重视实际应用问题的解决，如针对数据量缺乏现实，研究小样本学习、迁移学习方法；针对数据标注成本高昂问题，发展半监督、无监督领域自适应研究；用多模态方法，尽可能使用各种模态（视角）数据来优化提高模型性能；3D 视觉的检测与分割、视频理解等技术的发展，把模型应用从二维的、静止的场景拓展到更贴合实际的三维的、动态的场景；开展深度模型泛化性研究，以提升模型的环境适应性。笔者认为，计算机视觉技术发展的目的是能解决工业应用和人们生活中的实际需求，因此如何推动模型落地应用，如何让模型从识别到理解、从大数据驱动到小样本学习、从单一静止场景到多模态多风格场景，是计算机视觉任务未来的发展方向，也是开展科研工作的重要切入点。

责任编委 魏秀参

参考文献

[1] Xingjian Zhen, Zihang Meng, Rudrasis Chakraborty, Vikas Singh. On the Versatile Uses of Partial Distance Correlation in Deep Learning. ECCV 2022.

[2] Garvita Tiwari, Dimitrije Antic, Jan E. Lenssen, Nikolaos Sarafianos, Tony Tung, Gerard Pons-Moll. Pose-NDF: Modeling Human Pose Manifolds with Neural Distance Fields, ECCV2022.

[3] Ishit Mehta, Manmohan Chandraker, Ravi Ramamoorthi. A Level Set Theory for Neural Implicit Evolution under Explicit Flows. ECCV2022.

[4] Junfeng Wu, Qihao Liu, Yi Jiang, Song Bai, Alan Yuille, Xiang Bai. In Defense of Online Models for Video Instance Segmentation. ECCV 2022.

[5] Zhihao Li, Jianzhuang Liu, Zhensong Zhang, Songcen Xu, Youliang Yan. CLIFF: Carrying Location Information in Full Frames into Human Pose and Shape Estimation. ECCV2022.

[6] Zhengqi Li, Qianqian Wang, Noah Snavely, Angjoo Kanazawa. InfiniteNature-Zero: Learning Perpetual View Generation of Natural Scenes from Single Images. ECCV2022.

[7] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, Rob Fergus. Indoor Segmentation and Support Inference from RGBD Images. ECCV 2012.

[8] Daniel J. Butler, Jonas Wulff, Garrett B. Stanley, Michael J. Black. A Naturalistic Open Source Movie for Optical Flow Evaluation. ECCV 2012.

祁磊

东南大学计算机科学与工程学院 PLAM 实验室助理研究员（至善博士后），2020 年博士毕业于南京大学计算机科学与技术系。主要研究方向为计算机视觉、模式识别。

Email:qilei@seu.edu.cn

北京航空航天大学徐迈教授访谈

2022 年 11 月 9 日，《CCF-CV 专委简报》在线采访了北京航空航天大学博士生导师徐迈教授。下面是采访实录。

徐老师，您好！首先，请您分享一下您的个人学习和研究经历。

我本科毕业于北京航空航天大学电子信息工程学院，硕士毕业于清华大学电子工程系，博士毕业于伦敦帝国理工学院电气与电子工程系，师从英国皇家工程院院士 Maria Petrou 教授。博士毕业后，在清华大学从事博士后研究工作，合作导师是中国科学院院士陆建华教授。首先，非常感谢在国外的求学经历，在 Maria Petrou 教授指导下，使我逐步了解图像处理领域的国际学术前沿，也使我热爱上了图像处理这门学科；同时，博士阶段的求学经历也使得我掌握了独立从事科研的能力，为后面的理论研究工作打下了一定的基础；更要感谢在清华大学期间的博士后研究经历以及陆老师对我的指导，让我深刻认识到要将基础理论研究与国家重大需求相结合，使我们的科研成果能够更好地服务国家的需求。

2013 年加入北航后，我已经从事了十年的图像处理、视频压缩与质量增强等领域的研究工作，这十年我始终坚持开展基础理论研究以及面向国家需求的关键技术攻关与成果应用转化。在基础理论方面，获得了国家自然科学基金委的首批原创探索项目，并且在今年获得滚动支持；在成果应用方面，获得中国科协“求是”杰出青年成果转化奖。

您在图像处理、视频压缩与增强等领域内颇有建树，能否介绍一下您在这些领域中最突出的几项研究成果？针对这些领域的研究者，您有什么建议？

在视频压缩领域，我们团队的研究特色是契合人类视觉感知的视频压缩方法。尤其是，每一代视频编码都是以率失真为目标优化演进的；但是，这些标准并没有考虑到用户到底需要什么。因此，我们在视觉感知领域，研究人类视觉关注模型；在视觉关注模型基础上，发明了新的码率控制技术，提出了率-感知失真的优化方法，为用户提供更加契合视觉体验的视频业务，同时可以显著降低视频数据量。此外，针对视频编码复杂度高的实际问题，我们团队率先将深度学习方法应用于视频编码的块划分，显著降低了视频编码的复杂度，确保了视频编码器的实时性。为进一步提升视频质量，我们也是国内较早开展视频质量增强团队之一，其中代表性成果有多帧联合优化的 MFQE 方法，率先利用多帧信息互补的方法来增强视频编码每一帧的质量，同时也获得了视频质量增强领域最权威的竞赛 2022 年 CVPR NTIRE 两个赛道的冠军。此外，我们团队面向全景视频这一类新的媒体业务，也从用户感知模型、质量评价方法、视频压缩技术等多个方面开展前瞻性研究工作，代表性工作发表在 IEEE TPAMI 等顶级期刊上。

视频压缩与质量增强一直是刚性需求。据我了解，虽然今年互联网企业就业情况不容乐观，但是在视频压缩、质量增强领域还是有大量的岗位需求，可见其具有重要的实用价值，已经成为人类生活数字化的基本需求。因此，我建议能够有更多的研究者参与到这些领域的研究中来，并且以实际应用为导向，开展能够解决实际问题的研究工作。

作为知名青年学者，您获得了多个国家级头衔，您认为这些头衔对您的学术生涯有什么样的影响？

我觉得学术头衔本身对我的研究能够起到较好的促进工作。比如说，可以使我的团队有更多的科研经费，支持我们做更加深入的基础理论研究；也可以吸引一些志同道合的青年人才加入我的团队，一起开展学术研究工作。

最后，我觉得学术头衔只是对个人过去成绩的一些肯定；它对我们只是一种鞭策作用，后面还需要更加努力，培养人才、回馈社会。

您很注重教学工作，并获得了多项教师奖/教学奖，您认为教学与科研是什么关系呢?可否分享一下您的教育理念和教学方法？

我始终觉得科研和教学是相辅相成、相互促进的。比如说，我从 2013 年春季学期开始，主要从事《图像处理》的本科课程教学工作，至今已有 10 年。这 10 年，图像处理发生了翻天覆地的变化，在课程教学中，我喜欢和学生讲一些图像处理领域的新技术、新方向；一般先从效果讲起，再讲背后的原理。这样可以保证课程的先进性，也能吸引学生的兴趣；学生一旦感兴趣，就愿意加入我的团队，作为硕士或博士研究生，从事相关方向的研究工作，最终做到科研促进教学。很庆幸，在这方面，我也得到了学生的认可，获得由学生选出来的北航“我爱我师”十佳教师，也获得了 2021 年高校计算机专业优秀教师奖励计划的个人荣誉以及北航研究生卓越课程奖。

另一方面，由于这些学生的加入，我们又能保证科研工作的先进性，进而促进我们的本科课程教学工作，由此做到“科研反哺教学”。

可否请您谈一谈在第三代人工智能时代，计算机视觉将如何发展？面临哪些挑战？哪些研究方向会特别有价值呢？

近年来，由于深度学习的快速发展，人工智能进入了新的黄金期，也极大推动了计算机视觉的发展。然而，我们也发现，深度学习是数据驱动下的黑盒子，其可解释性差，导致将深度学习应用到实际的计算机视觉任务的工作中，存在“不可信、不可靠、不安全”的实际问题，为计算机视觉进一步发展带来了巨大挑战。比如说，我们在和 301 医院的合作过程中，将深度学习应用于医学图像智能诊断，但是仅仅诊断出结果是不行的，“知其然”的同时还要“知其所以然”；因此，亟需开展面向医学图像智能诊断的深度学习可解释性研究。由此可见，“可解释的深度学习”这一研究方向比较有价值；同时，这一方向也入选了今年中国科协的十大科学问题。

您有多个社会学术兼职，能否跟大家分享一下您是如何兼顾本职工作和兼职工作的？能分享一下您的经验吗？

首先，我个人非常认同青年学者还是应该有些社会学术兼职的，这样有助于开拓学术视野、开展学术交流，对个人的学术生涯也有正向的促进作用。很多时候，学术灵感是在讨论甚至争论中获得的。但是，我也非常认同一位前辈的观点，在 35 岁以前，尽可能少承担学术兼职，主要因为这个时候正是学术思想的活跃期，更需要沉下心来，开展研究工作。在 35 岁以后，可以适当承担学术兼职，服务学术社区。所以，我也是在 35 岁才逐渐承担社会兼职活动，包括 IEEE TIP、TMM 的编委，并且也组织了一些特刊，希望推动我们领域的进一步发展。

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

我个人认为人才培养是高校教师的首要任务，所以从一开始我就非常重视人才培养工作。在研究生选拔过程中，我比较注重优中选优，尤其是前几届研究生，我对他们的要求比较高、指导也非常多。一旦前几届研究生养成了好的习惯、取得了好的成果后，能起到带头作用，整个实验室的学习与研究氛围就会很好，后面研究生也会得到传承、发扬好的传统，出更加优秀的成果。因此，从 2015 年至今，我的第一个硕士生获得北航研究生十佳(北航研究生最高荣誉)开始，已有 6 位硕士、博士同学获此殊荣。

我在研究生管理过程中，还是比较柔性的，我坚信学生本身都很优秀，只要他们有目标，一定能够出成果。因此，先了解大家的想法是什么，未来的规划，然后再因材施教；此外，我也愿意与研究生一起投入到一线科研工作中，做到示范带头作用，这样学生也会具有更高的积极性。

如果吐露研究工作者的心声，您最想说的是什么？

我最想说的是科研工作一定要“锲而不舍,金石可镂”，也希望每个科研工作者都能实现心中的梦想。

责任编委 赵振兵 余烨

徐迈

徐迈，北京航空航天大学电子信息工程学院教授、博导，教育部长江学者特聘教授，中国图象图形学学会青工委副主任，CCF-CV 专委委员。作为负责人承担了国家自然科学基金首批原创探索、重点、优青以及北京市杰青等项目。获教育部技术发明一等奖（第一完成人）、中国人工智能学会技术发明一等奖（第二完成人）、中国科协求是杰出青年成果转化奖。研究兴趣为图像处理、视频压缩与增强。近五年，在 IJCV、IEEE TPAMI、TIP、JSAC、TMM 等权威期刊以及 IEEE CVPR、ICCV、ECCV、ACM MM、AAAI、DCC 等重要会议上发表论文 100 余篇，多篇论文入选 ESI 高被引论文/热点论文。连续 10 年从事《图像处理》本科课程教学工作，获高校计算机专业优秀教师奖；连续 6 年从事《机器学习》研究生课程教学工作，获北航研究生课程卓越教学奖；入选北航“我爱我师”十佳教师。

委员好消息

✪ 2022 年 9 月 28 日，CCF 公布了 2022 下半年新晋升高级会员名单，共 358 人，CCF-CV 专委会 13 位执行委员从专业会员晋升为高级会员，他们是：**邓红霞**（太原理工大学）、**范登平**（阿联酋起源人工智能研究院）、**高广谓**（南京邮电大学）、**黄栋**（华南农业大学）、**雷柏英**（深圳大学）、**刘静**（中国科学院自动化研究所）、**刘天歌**（燕山大学）、**桑农**（华中科技大学）、**施柏鑫**（北京大学）、**王利民**（深圳大学）、**王鹏**（西北工业大学）、**张世辉**（燕山大学）、**朱磊**（山东师范大学）。

✪ 2022 年 9 月 29 日，CCF 公布了 2022 下半年新晋升杰出会员名单，共 173 人，CCF-CV 专委会共 23 位执行委员晋升为杰出会员，他们是：**陈俊颖**（华南理工大学）、**程健**（中国科学院自动化研究所）、**方玉明**（江西财经大学），**何晖光**（中国科学院自动化研究所）、**何震宇**（哈尔滨工业大学（深圳））、**胡建国**（广州智慧城市发展研究院）、**雷震**（中国科学院自动化研究所）、**李永杰**（电子科技大学）、**李振波**（中国农业大学）、**刘家瑛**（北京大学）、**鲁继文**（清华大学）、**曲延云**（厦门大学）、**王昌栋**（中山大学）、**王胜科**（中国海洋大学）、**邬向前**（哈尔滨工业大学）、**许信顺**（山东大学）、**严骏驰**（上海交通大学）、**张淳杰**（北京交通大学）、**张文强**（复旦大学）、**张晓宇**（中国科学院信息工程研究所）、**赵才荣**（同济大学）、**赵涓涓**（太原理工大学）、**赵振兵**（华北电力大学）。

✪ 2022 年 10 月 12 日,CCF-CV 专委会执行委员、哈尔滨工业大学(深圳)**聂礼强**等的论文 Search-oriented Micro-video Captioning 获 ACM Multimedia 大会最佳论文奖。

✪ 2022 年 10 月 14 日，国家市场监督管理总局（国家标准化管理委员会）批准了 708 项推荐性国家标准和 3 项国家标准修改单，其中 GB/T 41864-2022《信息技术 计算机视觉 术语》有 CCF-CV 专委会 39 位执行委员参与制定，他们是：中国科学院计算技术研究所**陈熙霖**、**王瑞平**，北京工业大学**毋立芳**、**马伟**、**简萌**、**段立娟**、**贾熹滨**，中国科学院自动化研究所**王亮**，爱奇艺**王涛**，北京格灵深瞳信息技术有限公司**邓亚峰**，上海科技大学**李实英**、**虞晶怡**，南开大学**杨巨峰**、**程明明**，南京邮电大学**周全**，福州大学**牛玉贞**，北京科技大学**殷绪成**，北京邮电大学**马占宇**，中国科学院深圳先进技术研究院**乔宇**，华北电力大学**赵振兵**、**翟永杰**，北京交通大学**韦世奎**，北京电子科技学院**金鑫**，中国科学技术大学**王上飞**，电子科技大学**姬艳丽**，湘潭大学**欧阳建权**，西北工业大学**韩军伟**，哈尔滨工程大学**刘海波**，西安电子科技大学**苗启广**、**邓成**，中国科学院信息工程研究所**葛仕明**，中国石油大学（华东）**刘伟锋**，郑州大学**徐明亮**，南京大学**任桐炜**，兰州理工大学**李策**，中国科学院大学**马丙鹏**，中国科学院自动化研究所**何晖光**，重庆大学**张磊**和字节跳动**王长虎**（现单位龙湖集团）。

✪ 2022 年 10 月 31 日，2022 年度中国人工智能学会-华为 MindSpore 学术奖励基金入选名单公示，本年度有 60 个项目入选 A 类奖励基金名单，10 个项目入选 B 类奖励基金名单。北京航空航天大学**刘偲**、香港中文大学（深圳）**吴保元**、郑州大学**徐明亮**、南京理工大学**张姗姗**主持的项目入选 A 类奖励基金，北京交通大学**丛润民**、**张淳杰**、浙江大学**李玺**、南京理工大学**魏秀参**主持的项目入选 B 类奖励基金。

✪ 2022 年 11 月 14 日，2022 年《麻省理工科技评论》"35 岁以下科技创新 35 人" 亚太区入选名单揭晓，CCF-CV 专委会执行委员、京东集体**刘武**因 "参与建设了中国国家级智能供应链人工智能开放创新平台，支持了智能生产、流通和服务场景的示范应用，在疫情防控、

科技冬奥中发挥了重要作用”而入选。

✪ 2022 年 11 月 23 日，2022 年度中国图象图形学学会科学技术奖评选结果揭晓，本年度共评选出自然科学奖 6 项，技术发明奖 3 项，科技进步奖 6 项，高等教育教学成果奖 9 项，青年科学家奖 5 人，石青云女科学家奖 4 人，优秀博士学位论文奖 10 篇、优秀博士学位论文提名奖 7 篇。CCF-CV 专委会 22 位执行委员获奖：四川大学**彭玺**等完成的“高维数据的潜在结构发现”、中科院计算所**王瑞平**、**山世光**、**陈熙霖**等完成的“开放场景中视觉数据的关联建模与学习”获自然科学一等奖；中山大学**任传贤**等完成的“图像的多尺度特征融合与自适应判别分析”、北京交通大学**阮秋琦**、中科院自动化所**雷震**、北京航空航天大学**黄迪**等完成的“以人为中心的视频时空特征表示与学习”获自然科学二等奖；复旦大学**张军平**等完成的“端云结合视觉细粒度分析关键技术和应用”、北京邮电大学**何召锋**等完成的“视听觉个体感知与交互关键技术与应用”获技术发明二等奖；北京理工大学杨**健**等完成的“对抗环境下多模态虚实交互技术及应用”、新疆大学**库尔班•吾布力**等完成的“多文种离线手写签名识别与鉴别关键技术及应用”获科技进步二等奖；华中科技大学**白翔**等完成的“智能机器人复合人才培养创新与实践”获高等教育教学成果一等奖；江西财经大学**方玉明**等完成的“财经高校信息类专业创新人才培养探索与实践”、中南大学**陈再良**等完成的“计算思维引领的立体式实践平台，赋能计算视觉领域大学生创新能力培养”获高等教育教学成果二等奖；西北工业大学**戴玉超**和中科院自动化所**黄岩**获青年科学家奖；深圳大学**雷柏英**获石青云女科学家奖（青年组）；北京航空航天大学**徐迈**指导的“实数域与复数域下显著性检测模型关键技术研究”、华中科技大学**白翔**指导的“自然场景端到端文字识别方法研究”、西北工业大学**韩军伟**指导的“多视图数据机器学习算法研究”、武汉大学**夏桂松**指导的“图像几何结构的矢量化感知”、华中科技大学**桑农**“面向场景分割的判别特征感知方法研究”获优秀博士学位论文奖；南开大学**程明明**指导的“知识引导的自适应图像理解”、中科院计算所**陈熙霖**指导的“面向自动驾驶场景下图像语义分割的表示学习”获优秀博士学位论文提名奖。

✪ 2022 年 11 月 26 年，2022 年度教育部-华为“智能基座”优秀教师奖励计划获奖名单揭晓，CCF-CV 专委会执行委员、哈尔滨工程大学**刘海波**入选，本年度共 20 人入选。

✪ 2022 年 12 月 2 日，2022 年度上海市科学技术奖复评结果公布，CCF-CV 专委会执行委员、同济大学**何良华**、**赵才荣**等完成的“复杂条件下视频理解与传输关键技术及在智慧城市中的应用研究”、上海交通大学**林巍峣**等完成的“异构混合网络的内容适配化多媒体实时传输关键技术及系统”获科技进步一等奖。2022 年度共评出 16 名青年科技杰出贡献奖、60 项自然科学奖、50 项技术发明奖、186 项科技进步奖、18 项科学技术普及奖、3 名科技功臣奖。

✪ 2022 年 12 月 2 日，2022 年度中国人工智能学会-华为 MindSpore 学术奖励基金 C 类项目入选名单正式揭晓，CCF-CV 专委会执行委员、中国科学院空天信息创新研究院**孙显**主持的“遥感数据跨模态预训练基础模型研究及应用”项目入选，本年度有 16 个项目入选。

✪ 2022 年 12 月 15 日，山东省人民政府发布了《山东省人民政府关于 2022 年度山东省科学技术奖励的决定》，CCF-CV 专委会执行委员、上海交通大学**杨杰**等参与完成的“医学影像智能分析关键技术及应用 ”获科技进步一等奖，北京工业大学**胡永利**等参与完成的“复杂场景下特需任务道路交通保障关键技术及应用”获科技进步二等奖。本年度共授予科学技术最高奖 2 项、科学技术青年奖 10 项、自然科学奖 36 项、技术发明奖 19 项、科技进步奖 142 项、国际科学技术合作奖 4 项。

✪ 2022 年 12 月 23 日，CCF-CV 专委会执行委员、复旦大学**张军平**、南方科技大学**于仕琪**、兰州理工大学**李策**、西安电子科技大学**沈沛意**、南京理工大学**魏秀参**获 CCF-CV 专委会中科视拓 Seeta 服务贡献学者称号。

责任编委 刘海波

安检 X 线图像自动检测开源代码

东北大学 贾同 马博文

基于 X 射线图像进行包裹检查是目前公共场所（航空、运输等）使用最广泛的维护安全的手段，可以有效减少犯罪和降低恐怖袭击的风险。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、自动的包裹检查，以分担安检人员重复性工作，解决漏检率高、检测效率低等问题。其难点在于 X 线图像经常存在重叠、遮挡、类内差异、类别不平衡等，解决这些问题将促进其在现实中的潜在应用。安检 X 线图像自动检测由于其重要的应用价值，已成为计算机视觉领域的重要研究方向之一。本文主要从三个具体任务（分类、检测、分割）介绍该领域的研究成果，如图 1 所示。

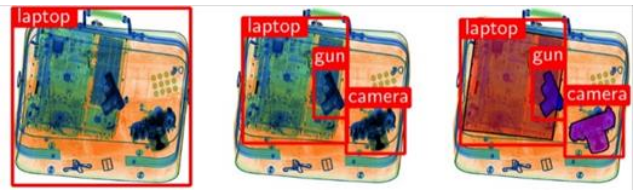


图 1 以不同任务实现安检 X 线图像自动检测

1、Class-balanced Hierarchical Refinement (CHR)

工作： CHR 将 X 线图像检查视为多目标分类任务，并将每张安检 X 线图像近似视为是一系列透明子图像的集合，因此可以使用一个混合模型来建模图像：

$$\mathrm{x}_n \approx \sum_{c=1}^{C} y_{n,c} \cdot \mathrm{x}_{n,c} \qquad (1)$$

其中 x_n 表示一张安检 X 线图像；$\mathrm{x}_{n,c}$ 表示对应于特定类别的透明子图，类别来自于一个开放世界集；$y_{n,c}$ 表示属于该类别的子图是否存在，1 表示存在，0 表示不存在。上述公式导致安检 X 线图像在基于 CNN 的特征学习过程中产生严重的特征混叠现象。为了解决这个问题，CHR 从三个连续层中提取图像特征，然后将高层特征上采样后与低层特征连接，细化函数 $g(\cdot)$ 从连接的特征图中删除无关的冗余信息，这种利用高级监督信号通过反向连接指导低级信息的方式实现了各层特征的优化，最后最小化三个连续层优化后的特征图的分类加权和损失以解决正负样本不平衡，整个网络结构如图 2 所示。

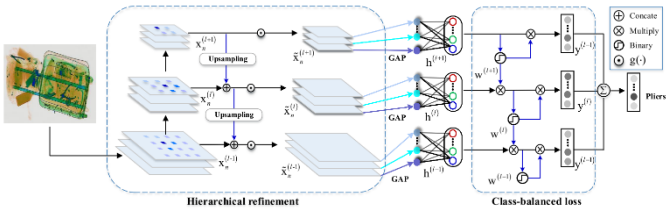


图 2 CHR 结构图

实验结果表明 CHR 可以适用于任意经典的骨干网络，并取得一致的性能增益（例如，在 ResNet-101 上取得了 2.13%的 mAP 提升）。

更多有关 CHR 的详细内容可参考发布该方法的论文“SIXray: A Large-scale Security Inspection X-ray Benchmark for Prohibited Item Discovery in Overlapping Images”。

论文地址： https://arxiv.org/pdf/1901.00303.pdf
代码地址： https://github.com/MeioJane/CHR

2、De-occlusion Attention Module (DOAM)

工作： DOAM 将安检 X 线图像自动检测视为经典的目

标检测任务，旨在解决图像中存在的遮挡问题。该模块可以即插即用到大多数经典检测器中，例如 SSD、YOLO、FCOS 等。DOAM 的核心部件是三个子模块：边缘信息引导模块、材料信息感知模块和注意力图生成模块。该文认为安检 X 线图像虽然存在严重的重叠、遮挡现象，但违禁品的边缘轮廓和由材料不同定义的颜色信息却十分显著。因此 DOAM 利用边缘信息引导模块、材料信息感知模块分别捕获违禁品的边缘信息和材质信息。注意力图生成模块利用上述两种信息来生成注意力图，以增强违禁品的识别能力，其结构图如图 3 所示。

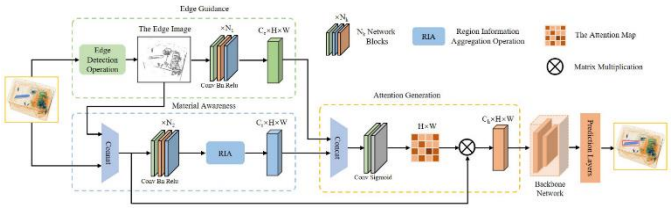


图 3 DOAM 结构图

边缘信息引导模块使用 sobel 算子的水平核和垂直核分别计算输入图像水平和垂直方向的边缘图像，合并后再与输入的图像在通道维度拼接以整合边缘信息。材料信息感知模块使用不同尺度核的平均池化层来聚合各个区域的信息，并对每个生成区域聚合特征图，从而进一步定义不同聚合区域尺度下的特征集，区域聚合操作如图 4 所示。

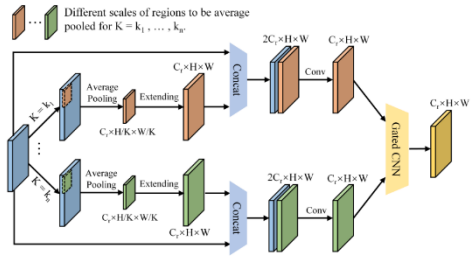


图 4 区域聚合操作示意图

最后，注意力图生成模块拼接上述两个子模块生成的特征信息，随后使用 1×1 的卷积核对通道维度进行压缩，生成适用于不同检测器的优化的注意力分数特征图。

该文在自己提出的 OPIXray 数据集上进行了大量实验，证明了 DOAM 在安检 X 线图像自动检测任务中的有效性和潜力。

更多有关 DOAM 的详细内容可参考发布该方法的论文“Occluded Prohibited Items Detection: An X-ray Security Inspection Benchmark and De-Occlusion Attention Module”。

论文地址： https://arxiv.org/pdf/2004.08656.pdf
代码地址： https://github.com/OPIXray-author/OPIXray

3、Dense De-overlap Attention Snake (DDoAS)

工作： DDoAS 将安检 X 线图像自动检测视为实例级分割任务，不仅可以定位违禁品在 X 线图像中的位置，同时可以分割出违禁品的轮廓，方便安检人员进一步判断是否为违禁品。除此之外，分割后的结果也可以为液体检测、图像注入等任务服务。方法上该文深入分析了 CHR，并在优化方法，优化目标和优化函数上提出改进以解决特征混叠问题。DDoAS 主要包括三个模块：密集去重叠模块，一对一融合模块和自适应形变模块，总体结构如图 5 所示。

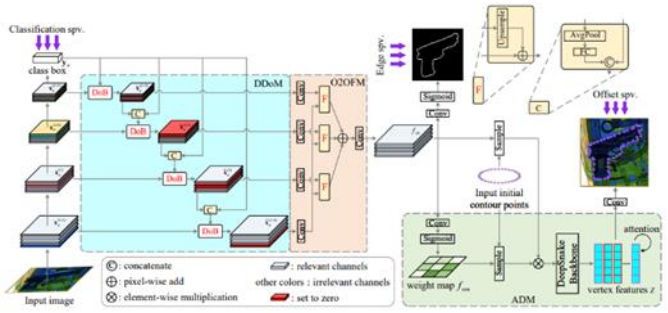


图 5 DDoAS 结构图

密集去重叠模块认为由公式 (1) 造成的特征混叠现象可以在通道维度上得到解耦，因为不同的通道只提取图像的一种特征，因此利用高层特征生成指导信号，用于评估低层特征中每个通道提取的特征是否属于违禁品，以过滤掉无用的冗余信息，结构图如图 6 所示。

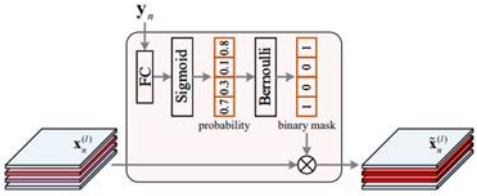


图 6 密集去重叠模块结构图

一对一融合模块将最底层特征图作为公共块，将其分别与高层特征融合，以实现底层空间信息的重复、高效利用，生成准确、细致的边缘轮廓。

自适应形变模块引入自注意力思想，根据初始轮廓中不同顶点位置的重要程度为不同的初始顶点分配不同的权值，以提高由于不同违禁品之间形状差异大导致的低学习效率。

广泛的实验证明了 DDoAS 在安检 X 线图像和自然图像上的有效性和优越性，暗示了其在安检领域和其他领域的潜在应用。

更多有关 DDoAS 的详细内容可参考发布该方法的论文“Automated Segmentation of Prohibited Items in X-ray Baggage Images Using Dense De-overlap Attention Snake”。

论文地址： https://ieeexplore.ieee.org/document/9772992

代码地址： https://github.com/Mbwslib/DDoAS

责任编委 李策 樊鑫

贾同

东北大学信息科学与工程学院教授、博士生导师，智能感知与机器人研究所所长。研究方向为计算机视觉、模式识别、图像处理和深度学习等领域。电子邮箱：jiatong@ise.neu.edu.cn

马博文

博士研究生，东北大学信息科学与工程学院，研究方向为图像处理、计算机视觉和深度学习
电子邮箱：2010285@stu.neu.edu.cn

水下目标检测数据集

大连理工大学 付陈平 樊鑫

近年来，水下目标检测（Underwater Object Detection，UOD）得到广泛关注。水下目标检测是指识别并定位出水下图像中感兴趣的目标。相较于一般检测，水下检测面临各种检测挑战。一方面，水下目标多遮挡、形变、伪装等问题，且小目标居多，不易检测；另一方面，水下成像环境复杂，往往存在色偏、雾效应、光干扰等现象，这些环境问题进一步增加了水下检测的难度。为了提升 UOD 的检测性能，研究人员一方面设计专门的水下检测器，另一方面提出大量水下检测数据集，以挖掘现有深度检测器在水下环境的检测性能。下面，本文将详细介绍三种代表性的 UOD 数据集，分别为 RUOD、UODD 和 UDD。

1、RUOD 数据集

介绍： RUOD (Real-world Underwater Object Detection) 数据集发布于 2022 年，由大连理工大学樊鑫团队收集发布。RUOD 数据集是第一个面向一般真实水下场景的 UOD 数据集，其图片来源于互联网以及现有的水下数据集 URPC 系列。该数据集共包括 14，000 张图像，10 类水下目标，以及 74，903 个标记实例。其中，10 类目标包括：鱼、潜水员、海星、珊瑚、海龟、海胆、海参、扇贝、鱿鱼、水母。RUOD 除了常规的训练集和测试集之外，该数据集还包括 3 个环境挑战测试集，分别为光干扰、色偏和雾效应测试。其中，训练集包括 9，800 张图像，测试集包括 4，200 张图像。3 个环境测试集分别包括 100 张图像。图 1 展示了 RUOD 数据集中图像的例子，接下来，将详细介绍该数据集。

图 1 RUOD 数据集示例

RUOD 数据集的类别集确定坚持两个主要原则：（a）类别具有代表性，与日常生活密切相关，且出现频率高，可满足大规模数据集建立所需的实例数量。（b）类内目标形态可以变化显著，而类间形态变化可以仅有细微差异，这样可训练探测器获得更为鲁棒的辨别能力。除了制订目标类别选取的原则，研究人员还制订了图像的筛选原则。具体来说，RUOD 旨在建立一个用于一般水下场景中检测的数据集。鉴于该目标，研究人员制订了筛选图片的两个原则。首先，图片要从照片共享网站以及现有的 UOD 数据集而不是从个人相册中进行收集。该过程可实现图片的“无偏见”收集。其次，图像中应包含多样的背景以及观察视角，应尽量避免图像中仅有一个目标的情况。

如图 1 所示，RUOD 数据集具有跨时空、多场景、多目标的特点，且涵盖多样水下检测挑战。具体而言，RUOD 包含的水下检测挑战主要有，（a）雾效应：在水下摄影中，水介质和悬浮粒子对光的吸收和散射会导致图像出现低对比度和模糊现象。此外，图像采集设备的

移动同样会导致图像出现雾状的模糊效果。（b）色偏：由于水介质对光的吸收，导致水下图像往往出现偏蓝或偏绿的问题。（c）光干扰：人工光源的使用以及浅水对光的反射等原因，水下图像会存在大光斑、光波纹等干扰。（d）复杂目标：群居海洋目标通常大规模出现，造成严重的相互遮挡等问题。水生生物还存在体型小、姿态变形多样，伪装外表等问题。此外，海洋目标还存在类内差异较大，类间差异较小的问题，例如，热带鱼和鲨鱼有完全不同的外观，但属于同一种类目标。而水母和塑料袋的外观非常相似，却属于不同目标类型。

更多的数据集建立及内容细节可参考论文“Rethinking general underwater object detection: Datasets, challenges, and solutions”。

论文地址：https://linkinghub.elsevier.com/retrieve/pii/S0925231222013169

数据集下载地址：https://github.com/dlut-dimt/RUOD

2、UODD 数据集

介绍： 2021 年，UODD（Underwater Object Detection Dataset）数据集发布。该数据集 2，500 张图像用于训练，128 张用于验证，506 张用于测试，共计 3，134 张图像。UODD 包括海参、海胆、扇贝三类目标，共计 19，212 个标记实例。其中，海参 4，714 个，海胆 13，083 个，扇贝 1，415 个。图 2 展示了 UODD 数据集的目标统计数据详情。

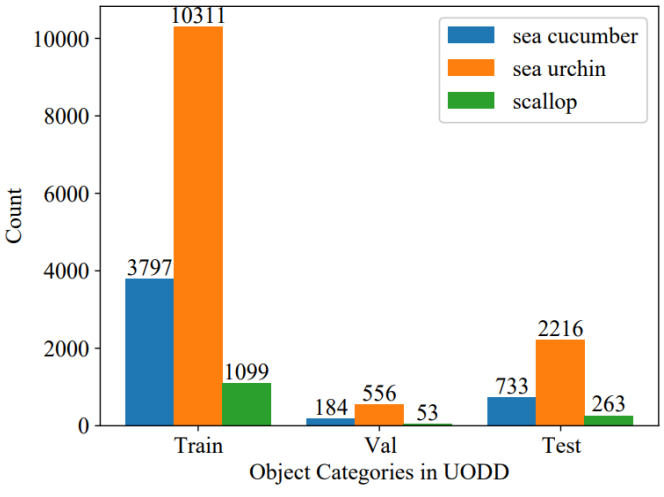


图 2 UODD 目标统计数据

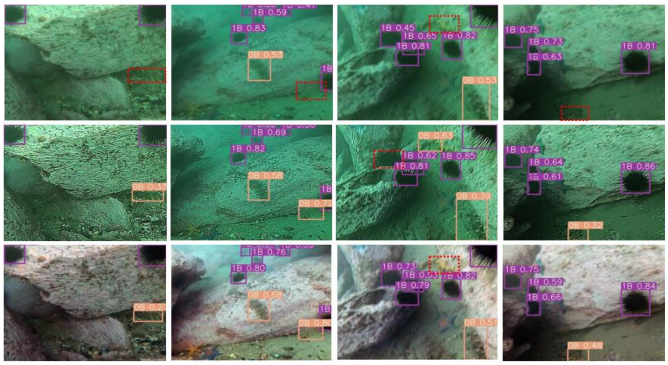


图 3 UODD 数据集示例

UODD 对现存水下增强数据集 RUIE 的图片进行了筛选与标记。RUIE 包含丰富的水下目标以及多样的光散射场景。然而，RUIE 数据集不提供目标检测任务所需的类别和位置标签。因此，UODD 对 RUIE 数据集进行了处理，使之可用于 UOD 任务。UODD 在标注时遇到一些棘手情况，例如：图像中有很多模糊的小目标，或者一些图像中的某些目标被其他物体部分遮挡。对此，数据集标记人员忽略了这些没有明显特征的目标。如图 3 所示，该数据集涵盖了多种水下检测挑战，例如：多种目标视角，单幅图像中多个目标数量，低对比度场景，各种尺度目标等。

更多的数据集建立及内容细节可参考论文“Underwater Species Detection using Channel Sharpening Attention”。

论文地址：https://dl.acm.org/doi/10.1145/3474085.3475563

数据集下载地址：https://github.com/LehiChiang/Underwater-object-detection-dataset

介绍：2022 年，UDD (Underwater Object Detection)

3、UDD 数据集

数据集发布。UDD 是一个 4K 高清数据集，包括 2，227 张图像和 3 类目标。 其中，海参 1，148 个，海胆 13，592 个，扇贝 282 个，共计 15，022 个实例目标。由于该数据集面向水下机器人海产品抓捕项目，因此对于海参、海胆等经济效益高的目标进行了大规模的标记，而对于扇贝等经济效益较低的目标标记较少，从而存在较为严重的类别不平衡问题。UDD 数据集包括常规的

训练集和测试集，其中训练集 1，827 张图像，测试集 400 张图像。图 4 展示了 UDD 数据集中图像的例子，接下来，将详细介绍该数据集。

2018 年，为了获取 UDD 图像，潜水员和机器人操作高清摄像机 Yi 4K 在中国大连獐子岛录制水下视频。潜水员和机器人在距离獐子岛约 500 米的两个水下地点进行了视频录制，以此提高多样化图像数据。对于同一地点，机器人在海底行走，录制 720P、1080P 和 4K 视频，同时，潜水员从俯视角录制相同分辨率的视频。两者遵循特定的循环路线。录制结束，研究人员根据不同地形，截取了四个视频。最后，研究人员从四个视频中分别截取了 1000 帧采样图像来构建 UDD 数据集。

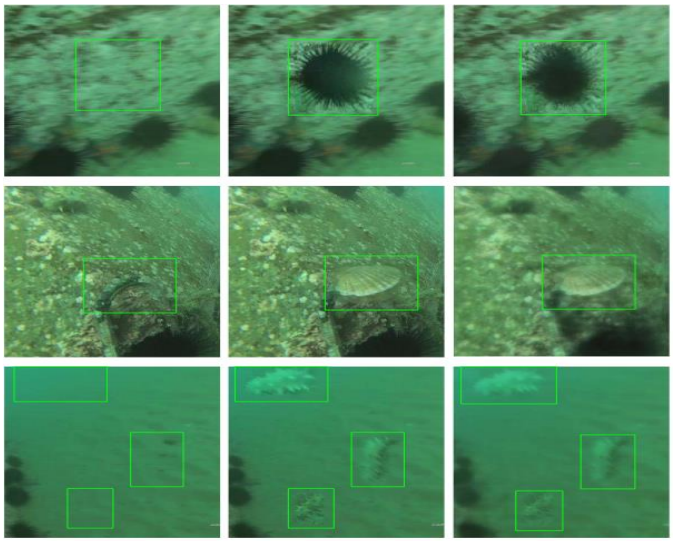


图 4 UDD 数据集示例

UDD 数据集中的图像具有 4K 的分辨率，每张图像平均有 10 个实例目标。相较于现存的检测数据（例如 URPC 系列，PASCAL VOC，MS COCO 等），UDD 数据集具有高清分辨率与单幅图像具有多目标数量的优势。此外，相较于先前的目标检测数据集，UDD 数据集包含大量的小目标，目标平均尺寸大小仅有 44×28 像素。对于检测器而言，小目标是难以检测的，因此 UDD 是一个极具挑战的水下检测数据集。同样，水下机器人拾取海洋目标时，如何设计一个既能检测小尺度目标又能保证拾取效率的检测器成为目前亟需解决的问题。

更多的数据集建立及内容细节可参考论文“A New Dataset, Poisson GAN and AquaNet for Underwater Object Grabbing”。

论文地址： https://ieeexplore.ieee.org/document/9496608
数据集下载地址： https://github.com/chongweiliu

责任编委 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，特殊场景下的目标检测。

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

好 文 推 荐

中国科学院自动化研究所“DVG-Face: Dual Variational Generation for Heterogeneous Face Recognition”的成果发表在 IEEE TPAMI 2022。

论文：Chaoyou Fu, Xiang Wu, Yibo Hu, Huaibo Huang, and Ran He. T DVG-Face: Dual Variational Generation for Heterogeneous Face Recognition, IEEE TPAMI, 44: 2938-2952 (2022)

异构人脸识别（Heterogeneous Face Recognition，HFR）是指跨域人脸的匹配，在公共安全中起着至关重要的作用，然而 HFR 面临着领域差异较大和异构数据不足的挑战。

文章将 HFR 表示为一个双生成问题，并通过一个新的对偶变分生成（Dual Variational Generation，DVG-Face）框架来解决这个双生成问题。作者提出了一个双变分发生器学习成对异构图像的联合分布，为避免小规模的配对异构训练数据可能会限制采样的身份多样性，将大规模视觉图像的丰富的身份信息整合到联合分布中。文章中对生成的成对异构图像采用了成对恒等损失，以确保其身份的恒等一致性，借此可以从噪声中生成大量具有相同身份的不同配对的异构图像，缓解了异构数据不足的问题。DVG-Face 框架如图 1 所示。

文章提出的方法在 HFR 领域的 5 个相关任务,包括红外-视觉图像，草图-照片，正面照，热感-视觉图像，身份证摄像，7 个数据集上取得的 state-of-the-art 的效果。

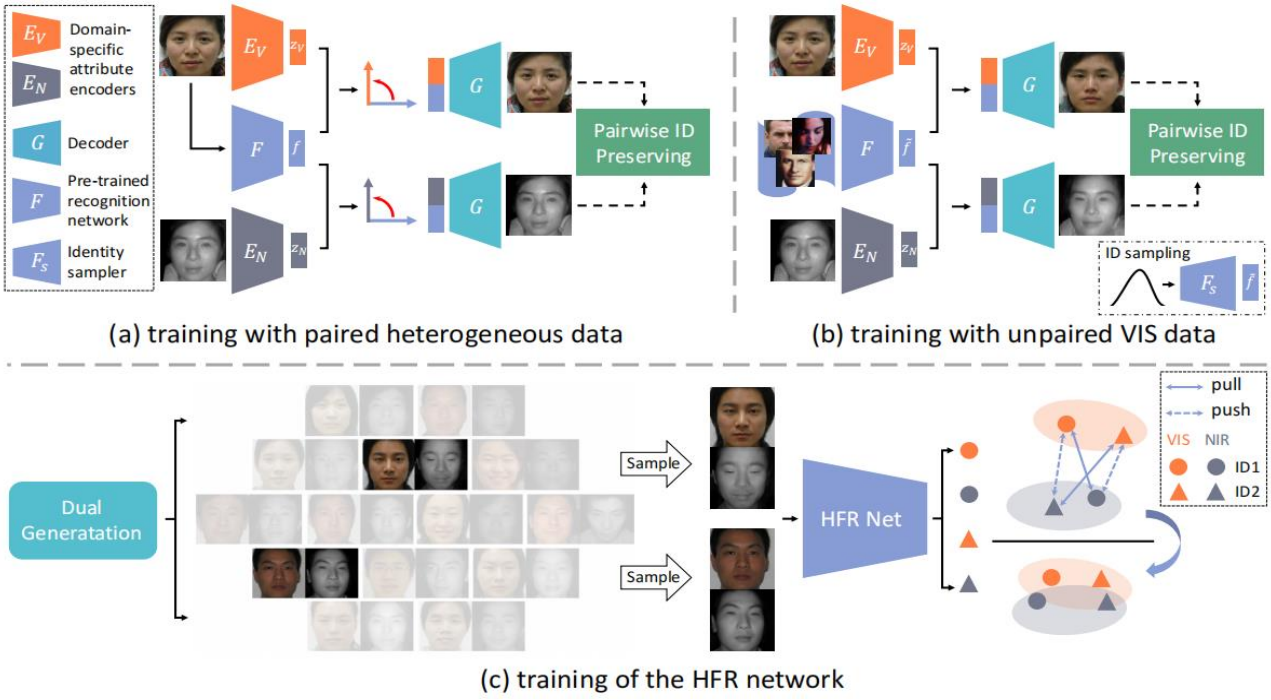


图 1 DVG-Face 框架

责任编委 李策 樊鑫

好 文 推 荐

西安电子科技大学“Transferable Coupled Network for Zero-Shot Sketch-Based Image Retrieval”的最新成果发表在 IEEE TPAMI 2022。

论文：Hao Wang , Cheng Deng , Tongliang Liu , and Dacheng Tao. Transferable Coupled Network for Zero-Shot Sketch-Based Image Retrieval, IEEE TPAMI, 44(12): 9181-9187 (2022)

手绘草图图像检索（Sketch-Based Image Retrieval，SBIR）主要目的是根据手绘草图的内容对图像进行检索，得到符合手绘草图内容的图像。传统手绘草图图像检索在实际应用时，需要满足对未知类检索的适配，要达到这一条件，需要使用零次学习(Zero-Shot Learning)。随着智能手机，平板点点等可触屏设备的不断发展，手绘草图图像检索任务和与之相关的零次学习方法也在不断取得新的进展。

文章探讨了之前零次学习手绘草图图像检索任务工作中的问题，对此提出了可迁移耦合网络框架，包含耦合网络，特征嵌入网络和语义度量学习模块。文章首先提出使用软权重共享的方法，减少硬权重共享中网络优化不均衡的问题，实现网络耦合；为了让网络学到的特征更具有判别性，使用知识蒸馏的方式，在 ImageNet 数据集中预训练的网络作为 Teacher Network，对耦合网络的特征向量进行约束。最后使用语义度量学习模块，提高网络整体的可迁移性。整个网络的框架图如图 1 所示。

文章在 Sketchy、TU-Berlin 和 QuickDraw 数据集上广泛评估了所提方法。实验表明，在上述数据集中均取得了最好的效果，验证了算法的效果。

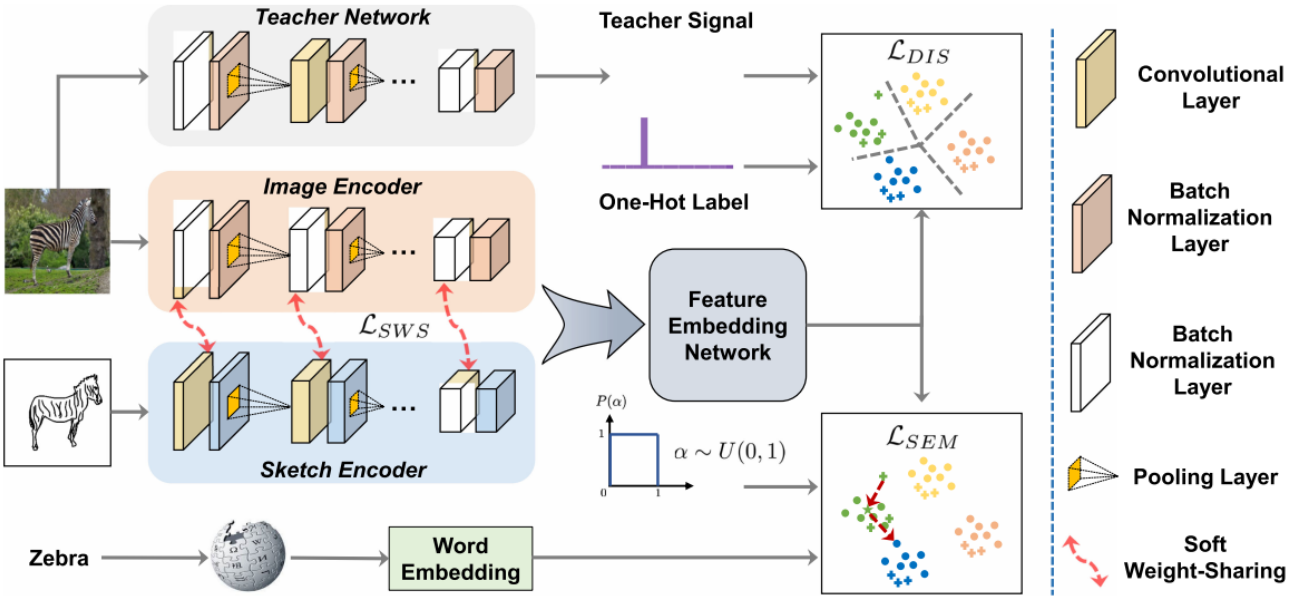


图 1 可迁移耦合网络框架

责任编委 贾同 樊鑫

好 文 推 荐

兰州理工大学、西安交通大学和腾讯科技 PCG ARC 实验室最新联合成果“DAGCN: Dynamic and Adaptive Graph Convolutional Network for Salient Object Detection”发表在 IEEE TNNLS 2022。

论文：Ce Li , Fenghua Liu, Zhiqiang Tian , Shaoyi Du, and Yang Wu. DAGCN: Dynamic and Adaptive Graph Convolutional Network for Salient Object Detection, IEEE TNNLS , 2022, DOI: 10.1109/TNNLS.2022.3219245

显著性目标检测在图像/视频等诸多领域有重要研究意义。显著性目标检测是将人类视觉选择性注意机制具体应用于重要目标检测的一类工作，如何有效地对场景中的上下文关系进行建模是显著性目标检测任务的关键。图卷积网络(GCN)因其在许多任务中的出色性能而广为人知。然而，现有的图卷积模型仍然存在邻域共性依赖、拓扑适应性差等问题。这些问题使得 GCN 难以直接应用于显著目标检测等视觉任务中。

文章提出了一种有效的 DAGCN 模型，如图 1 所示。以图模型为基础，设计了一组新颖的图卷积组件 DAGCN。组件包括一种非欧空间的图构建模块 SRKNN 以及一种新颖的自适应图卷积 AnwGCN。模型首先通过主干网络构建包含显著性信息的初级节点表示，并根据节点特征使用 SRKNN 征构建初始图。所构建的图将通过一组由 AnwGCN 以及 SRKNN 组成的动态图处理单元。模型可以有效适应拓扑变化带来的邻域信息差异。通过优化特征之间的信息交互网络来增强有用信息的上下文关系，突出显著区域的特征。该过程有效地减少了显著性目标二义性的问题，提高了显着目标检测的性能。

综合 6 个公开数据集的主、客观实验结果表明，DAGCN 优于 19 个流行的显著性目标检测方法。此外，DAGCN 还能有效地检测出具有弱边界的伪装目标，成为显著性目标检测任务的有效解决方案。

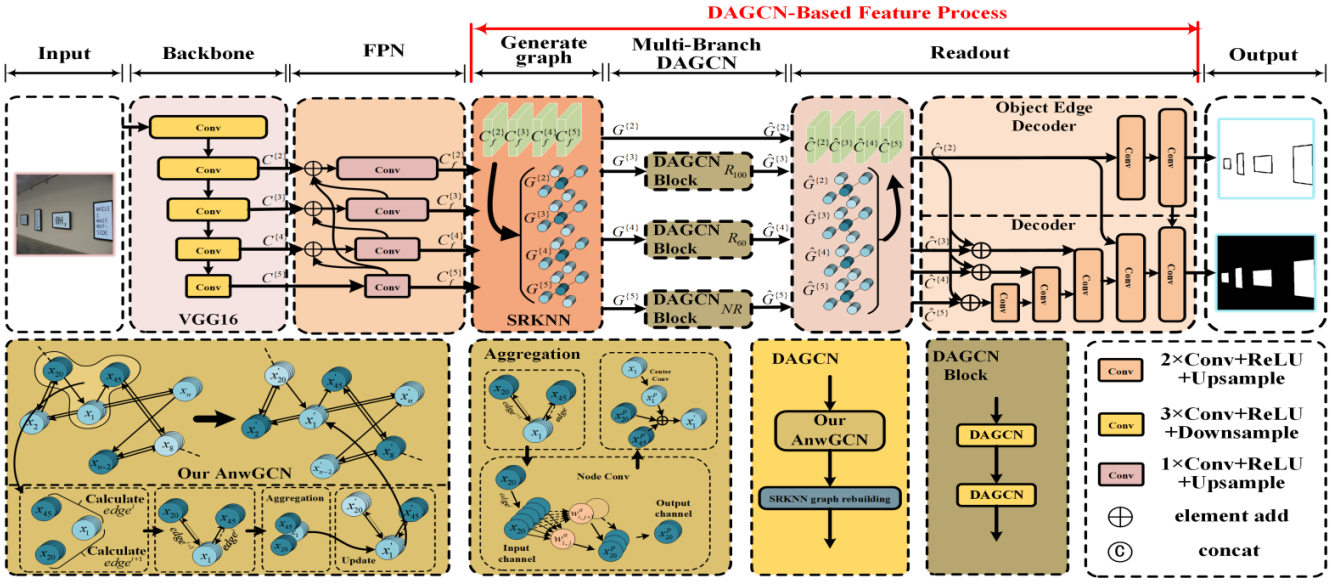


图 1 DAGCN 模型结构流程图

责任编委 沈沛意 樊鑫

征文通知

1 会议征文

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

2 期刊征文

计算机视觉领域近期相关期刊专刊的征文通知如表 2 所示，包括 IEEE Transactions on Multimedia, Visual Intelligence 和 IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing。

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)，由中国人工智能学会（CAAI）、中国计算机学会（CCF）、中国自动化学会（CAA）和中国图象图形学学会（CSIG）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。

第五届 PRCV 将于 2022 年 12 月 24 日至 26 日在深圳举行，由南方科技大学和深圳职业技术学院共同承办，香港浸会大学、香港中文大学（深圳）、哈尔滨工业大学（深圳）和中国科学院深圳先进技术研究所联合承办。本届会议旨为促进 PRCV 和湾区学者交流融合、聚焦前沿理论，提高学术交流氛围和质量、技术赋能产业，吸引科技类企业和投资类企业参与、联合粤港澳，提供一个全国科研团队和粤港澳企业近距离交流科研平台和机会。会议论文集将由 Springer 出版社出版，并被 EI 和 ISTP 检索。

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
IJCAI 2023	2023.08.19-25	Cape Town, South Africa	2023.01.19	https://ijcai-23.org/
ICML 2023	2023.07.23-29	Hawaii, USA	2023.01.27	https://icml.cc/
CHIL 2023	2023.06.22-23	Cambridge, USA	2023.02.03	https://www.chilconference.org/
ICCV 2023	2023.10.01-07	Paris, France	2023.05.09	http://iccv2023.thecvf.com/

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
TMM	Pre-trained Models for Multi-modality Understanding	https://signalprocessingsociety.org/sites/default/files/uploads/special_issues_deadlines/TMM_SI_pre_trained.pdf	2023.01.15
JSTAR	Remote sensing land surface temperature (LST) for generation, analysis, and application	https://grssiee.wpenginepowered.com/wp-content/uploads/2022/10/cfp_LST_extended.pdf	2022.02.28
Visual Intelligence	多模态视觉理解与生成	https://mp.weixin.qq.com/s/vxW2ASwUFXtdNKRAuICqjA	2023.03.01
JSTAR	Remote Sensing of Land Surface Variables over the Tibetan Plateau	https://grssiee.wpenginepowered.com/wp-content/uploads/2022/06/Call_for_Paper_Remote-Sensing-of-Land-Surface-Variables-over-theTibetan-Plateau.pdf	2023.03.31

心底无私视界宽∞肖自美教授专访

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。在这一过程中，有一批见证了视觉领域的发展，为我国计算机视觉领域的奠基做出了重大贡献的先驱。

至今，《视界专访》已经采访了 8 位在计算机视觉领域耕耘的资深教授。我们发现这些教授主要分布在上海、北京，但从地域看，广东还没有。因此，我们联系了中山大学赖剑煌教授，请他分享几位计算机视觉领域的南派资深教授，让计算机视觉学会及相关研究领域的人员也听听南派在计算机视觉和图像处理领域的故事。赖教授推荐了肖自美教授，并帮忙联系上了肖教授的学生梁凡副教授。由于疫情原因，我们对肖教授的采访是通过微信提问，由梁凡副教授录音整理后，再加工完成的。

图 1 肖自美教授

为能更好地帮助我们回顾本次采访，**我们采用问答的形式表述。**以下是肖自美教授的简介和专访内容。

问：现在短视频非常流行，您觉得视频编码的改进在其中是否起了非常大的作用?

肖教授：视频编码（图像编码）在音视频的数字化应用中是一项关键技术。不经过压缩，视频数据的传输、存储等问题都解决不了。

问：您是国内多功能视像电话系统的提出人，并因此获得科技进步奖二等奖。现在线上视频会议用得很广泛，而您是在 1985 年就开始考虑这一问题的。您能谈谈您当时做这个系统的初衷，以及存在的困难吗?

肖自美教授 1938 年出生于广西桂林市，1961 年毕业于武汉大学物理系无线电技术专业，1963 年东南大学无线电工程系研究生毕业后在武汉大学物理系任教。1981-1984 年公派赴德国留学，在德国亚琛技术大学(TH-Aachen)信息处理研究所担任访问学者。1987 年到中山大学任教。

1968 年参加了重大国防工程的外围研究工作。1970-1972 年参加中国彩色电视制式攻关，参与了我国电视制式（PAL）的制订工作。1974-1979 年参加了多部门联合下达的国防重大项目，并担任该项目技术负责人之一，该项目 80 年代初在首届全国科学大会上荣获国家科学技术进步奖。

肖自美教授指导培养了硕士、博士、博士后等 60 余名。其中绝大部分成为中外企业高管、企业家、国内外大学教授、高级工程师。肖老师培养的学生参与制订了我国高清、超高清数字电视使用的 AVS 系列视频编码标准，研发了支持 AVS 系列标准的编解码器等产品。

肖自美教授曾任中山大学图像研究室主任、中山大学信息与通信技术研究中心主任、中国图象图形学会常务理事、中国图象图形学报编委、广东省图象图形学会副理事长； 广东省通信学会副理事长、中国计算机学会多媒体技术专业委员会常务委员、中国通信学会通信信息与信息处理专业委员会委员、广东省“九五”科技规划“电子与通信”专家组组长，被聘为广州市政府科技顾问、电子与通信专家组组长。曾任广东省信息化专家咨询委员会常务副主任委员等社会学术职务。肖自美教授 1992 年获享受国务院颁发的政府特殊津贴突出贡献专家，1995 年入选《中国科技名人录》，1996 年入选《英国剑桥国际传记名人录》，2011 年 9 月荣获中国计算机学会与中国图象图形学会联合颁发的“多媒体技术终身成就奖”。

肖自美教授主持研制开发的“多功能可视通信系统”、“远程可视遥控监视系统”、“IP 网上多点视频会议系统”“数字视频监控/录象系统”，“广播级视频点播系统”，“实时视频广播/点播系统”，“远程可视医疗会诊系统”、“数字移动多媒体通信系统”、“数字语音/数据同步器”、“多媒体信息服务系统”，“指纹图像识别”等一系列成果，均为达到国内领先、国际先进水平的高科技成果，在国民经济众多部门和领域获得了广泛应用。

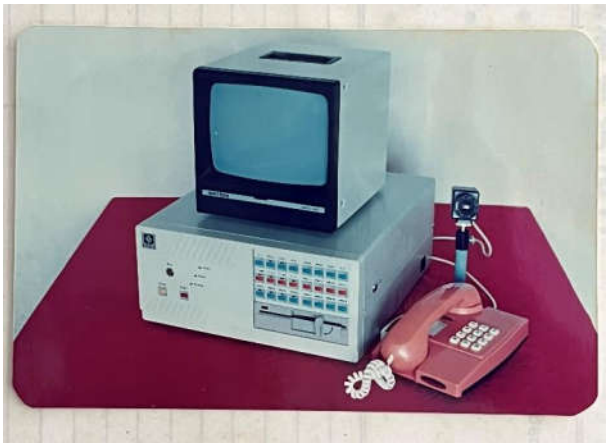


图 2 视像电话系统（1989）-01

图 3 视像电话系统 （1989）-02

图 4 视像电话系统 （1989）-03

肖教授：国内从 1980 年代初期开始发展程控交换，安装电话是一件非常困难的事情，我在德国 3 年都无法给家里打过一个电话。信道条件很差，除了电话线，根本就没有宽带的信道。从德国回来以后，我就考虑能不能利用电话线传输图像和传真。当时国内外尚未见到有类似成果，这项工作起步的时候得到了湖北省科委的支持，觉得我这项工作有很大的应用价值。当时电话线能提供的带宽只有 300-3400Hz，除了应用 DPCM 技术，还应用了像素抽取/填补等技术。设计开发了硬件电路板和 PC 机一起配合完成图像的压缩、传输。“数字式静态图像窄带传输系统”1986 年通过了湖北省组织的技术鉴定。每秒最高能传输 6-8 幅图像，也可以称为准动态的。

1989 年“多功能可视电话”获得首届“火炬杯”优秀产品奖，包括图像、语音、传真等功能。能获得“火炬杯”，当时取得了一定的经济效益。这套系统当时的主要应用是铁道部，用在铁路和公路交叉道口的监控上，以及做为视频通话用。

图 5 火炬杯优秀产品奖（1989 年）

问：您觉得目前的计算机视觉领域，还有哪些值得做的研究方向。

肖教授：我从 1988 年开始研究指纹识别，我觉得计算机视觉的核心是识别问题。特征点提取的算法还有待更深入的研究。我觉得图像压缩继续发展下去一定要与计算机视觉相结合。

对人的识别，要和医学、生物学等学科结合起来。不仅仅是身份验证，也可以用于其他领域。应该重视计算机视觉在国家安全、农业、医疗等领域的应用。

图 6 窄带传输系统鉴定会（1984）

问：能否分享一下您当年切入到图形图像领域的经历？能否分享一些南派图形图像、计算机视觉领域的早期研究趣事和有趣的研究者？

肖教授：1979 年国家派出了一大批留学生，有 2 万多人，主要是去美国、英国、加拿大。除了英语国家，后来又提出也要往非英语国家如德国、法国派留学生。我先是利用业余时间自学了一点德语，然后参加武汉大学的一个培训班学习了半年，1980 年初参加了资格考试，取得了比较好的成绩。

我为什么选择搞图像技术呢？在 1970 年代初期（大致是 1971-1973 年），国家组织彩色电视的技术攻关，全国有 10 几 20 万人参加。我是 1961 年大学毕业的，毕业以后先搞微波技术。后来武汉电视机厂生产电子管黑白电视机，我参加了研制工作，主要做高频头部分。因为参与过黑白电视机的研制，对电视的知识有所了解，因此我也参加了彩色电视攻关，算是武汉大学方面参加彩色电视攻关的主要人员之一。彩色电视攻关的工作前后做了有 2 年时间，由此对电视、视频、图像产生了兴趣。

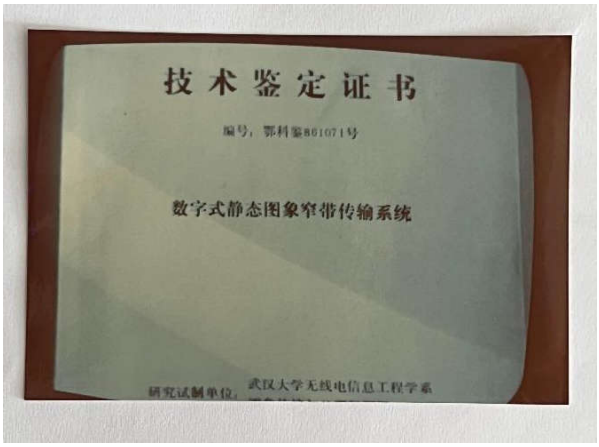


图 7 窄带传输系统鉴定证书（1984）

我考取公派留学资格以后，当时我们的系主任是龙咸灵教授，他是搞地球物理，特别是研究短波传播的。武汉大学的特点之一就是电波传播，他就要求我去学习日地空间，特别是研究太阳风对地球的影响。对此我不同意，我是无线电技术专业的，不是这个专业的，不想去搞这一行。我当时坚持要搞图像，通过联系到德国亚琛工业大学去做图像方面的研究。考取留学资格以后，我还到西安去学习外语，按期是学习一年，但是我半年

就毕业了，在 1980 年 10 月出发去德国学习。

我是 1984 年冬天回国的，在德国待了 3 年。我去德国接替了柴振民，他回国空出个位置，我正好去接替他，师从德国的 Taffe 教授。教授在西门子工作了很多年，下面有几个研究组，其中一个研究组做图像处理方面的工作，大概有 5-6 个人，我就在这个组。我去的时候，德国方面研究图像的主要目标是压缩，研究图像编码，当时都是研究静态图像的编码。当时在德国研究图像编码比较有名的有三家，一是汉诺威大学（Universität Hannover），二是亚琛工业大学（RWTH Aachen University），三是布伦瑞克工业大学（Technische Universität Braunschweig）。

图 8 多媒体技术终身成就奖（2011 年）

我到了德国以后，和研究组的几名博士生一起研究图像压缩。图像压缩是从 PCM 数据开始的，原始图像每像素 16bit，当时最好的水平能压缩到每像素 4bit，大多数人看不出原始图像和解压缩重建图像之间区别。当时有两条技术路径，一是做 DPCM，这个是在空间域做的；二是到变换域去做。我尝试了一下变换域的方法，使用 M 变换，效果不太好，因此把研究的重点放在了 DPCM 上。我在德国 3 年时间，大概有 2 年的时间用在 DPCM 上。那时整个大学没有一台个人电脑，学校有计算中心，整个图像组只有 1 台计算机终端可供使用。大家分时间使用，一周大概有 4-6 个小时的使用时间。我在德国研究的成果回国后发表在《通信学报》上（《压扩量化器》），做到了 4bit/pixel（即每像素）的水平，在德国也是最好的。当时天津大学的张春田教授，使用预测技术也做到了 4bit/pixel 的水平。

在《通信学报》发表文章时，还有一件趣事，图像压缩使用的一幅戴草帽女郎的测试图像(即著名的 Lena 图），学报编辑部说这幅图像内容不合适，是不是可以换一幅。但是，这幅测试图像是当时压缩难度最大的图象之一。所以，最终没换。

图 9 莲娜图（Lena），图像编码经典的测试图片

问：您认为培养研究生，导师需要注意哪些方面呢？如何提高学生的创新能力呢？

肖教授：作为老师，自己经历过的才容易讲得生动活泼，学生也容易接受。如果没有实际经历，照本宣科效果就不好。像图像编码领域，90%以上都是面向应用的。另外，导师要给研究生提供前沿的课题、经费、设备，研究生也必须从事实际的操作和研究。所以，如果没有研究课题，就谈不上培养。关于研究生的培养，我认为导师只能指导选题的方向、研究的方法，具体问题需要学生独立完成。导师不能代替学生做具体工作。

还需要注意的是，研究生不能都按同一个模式培养，应该加强导师的自主权和决定权。除此以外，研究生培养不能只唯论文，导师是否认可才是关键。导师既然有资格指导学生，他就有能力判断学生是否满足要求。

责任编委 张军平 明悦 贾熹滨

COMPUTER VISION NEWSLETTER

04 2022

总第 34 期

计算机视觉专委会简报

CCF 计算机视觉专委会