

主办 CCF 计算机视觉专业委员会

COMPUTER
VISION
NEWSLETTER

CCCF 计算机视觉 专委会简报

03 2023

总第 37 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 03 期

总第 37 期

主 办 编委会

CCF 计算机视觉专业委员会



CCF 计算机视觉
专 委 会

/专委动态/

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

/科技前沿/

潘金山 南京理工大学
任传贤 中山大学
杨巨峰 南开大学
朱安娜 武汉理工大学
主 编 **王金甲** 燕山大学
编 委 **储 珺** 南昌航空大学
崔海楠 中国科学院自动化研究所
魏秀参 南京理工大学

/委员风采/

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

/学术资源/

主 编 **李 策** 兰州理工大学
编 委 **樊 鑫** 大连理工大学
贾 同 东北大学

/海外学者/

王 田 北京航空航天大学
主 编 **金 鑫** 北京电子科技学院
编 委 **刘帅奇** 河北大学
张汗灵 湖南大学

/视界专访/

主 编 **张军平** 复旦大学
编 委 **贾熹滨** 北京工业大学
明 悦 北京邮电大学

CONTENTS

简报目录

| 专委动态

- 04 CCF-CV 走进高校系列报告会
- 05 CCF-CV 视界无限系列研讨会
- 09 CCF-CV 计算机视觉前沿讲习班
- 12 CCF-CV 计算机视觉前沿进展研讨会 (RACV 2023)

| 科技前沿

- 15 基于极大团的三维配准方法
- 22 基于超像素交互的图像超分辨率
- 27 CVPR 2023

| 委员风采

- 32 中国农业大学李振波教授访谈
- 36 委员好消息

| 学术资源

- 37 点云域自适应领域开源代码
- 40 裂痕缺陷分割数据集
- 44 好文推荐

| 海外学者

- 47 征文通知

| 视界专访

- 48 安徽大学韦穗教授专访

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 126 期 鲁东大学



2023年6月25日，由中国计算机学会计算机视觉专委会（CCF-CV）主办、鲁东大学信息与电气工程学院承办的走进高校系列报告会第126期活动在鲁东大学逸夫楼B座成功举办。本次活动邀请了复旦大学姜育刚教授、中国科学院计算技术研究所蒋树强研究员、山东师范大学朱磊教授做特邀报告。鲁东大学信息与电气工程学院院长王丽丽教授、副院长姚涛副教授担任本次活动的执行主席。

鲁东大学党委常委、副校长毕研强教授出席活动并致欢迎辞，随后，复旦大学姜育刚教授、山东师范大学朱磊教授、中国科学院计算技术研究所蒋树强研究员做主题报告。最后，鲁东大学信息与电气工程学院院长王丽丽教授进行了活动总结，首先感谢了三位专家的精彩报告与学术交流分享，同时感谢了师生们的热情参与，最后再次感谢 CCF-CV 专委会和学校对信息与电气工程学院承办本次活动的大力支持，并祝贺本次活动取得了圆满成功！

第 127 期 石河子大学



2023年8月12日，由中国计算机学会计算机视觉专委会（CCF-CV）主办、石河子大学信息科学与技术学院承办的走进高校系列报告会第127期活动在石河子大学中区科技楼报告厅成功举办。本次活动邀请了中国科学院自动化所刘成林研究员、武汉大学夏桂松教授、华南理工大学陈俊颖副教授做特邀报告。石河子大学信息科学与技术学院党委书记赵庆展教授、计算机科学系副主任郑煜辰副教授担任本次活动的执行主席。

石河子大学信息科学与技术学院党委书记赵庆展教授致欢迎辞。随后，刘成林研究员、夏桂松教授、陈俊颖副教授做主题报告。最后，石河子大学信息科学与技术学院郑煜辰副教授进行了活动总结，感谢三位专家的精彩报告和师生们的热情参与，对 CCF-CV 专委会和学校为本次活动提供支持表示感谢，并祝贺本次活动取得了圆满成功！

第 128 期 桂林电子科技大学



2023 年 9 月 16 日上午，由中国计算机学会计算机视觉专委会主办，桂林电子科技大学数学与计算科学学院承办的“CCF-CV 走进高校”学术交流活动在桂林电子科技大学花江校区第 6 教学楼 6306 报告厅隆重举行。本次活动由桂林电子科技大学人工智能学院副院长**刘振丙**教授、数学与计算科学学院**徐增敏**副教授担任执行主席。

活动首先由武汉工程大学计算机科学与工程学院/人工智能学院院长**卢涛**教授、中电科新型智慧城市研究院有限公司副总经理**胡金晖**研究员分别做主题报告。随后，广西医科大学**张茂胜**教授和各位讲者参与了 panel 讨论环节，主要以“通用大模型与垂直领域大模型发展”为议题展开讨论。报告结束后，刘振丙教授首先感谢了专家们的精彩报告与学术交流分享，然后感谢 CCF-CV 专委会和参加本次活动的校内外师生大力支持！并欢迎各位专家学者能再次来桂林电子科技大学进行学术指导和交流！祝贺本次活动取得了圆满成功！

责任编辑 朱安娜

第 18 期 视觉大模型的前沿进展与未来趋势

CCF-CV 视界无限系列研讨会

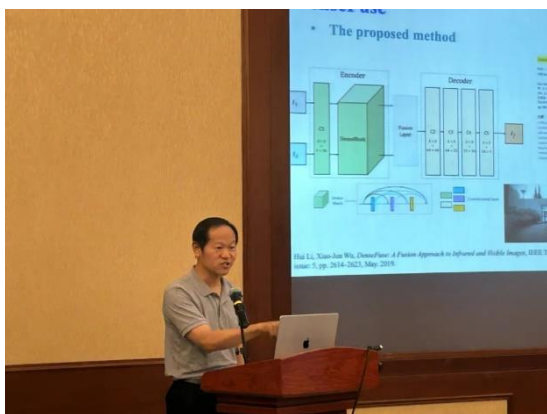


2023 年 7 月 8 日，由中国计算机学会计算机视觉专委会主办、复旦大学大数据学院承办的第 18 期 CCF-CV “视界无限”系列活动——“视觉大模型的前沿进展与未来趋势”研讨会在复旦大学生子彬院 205 会议室举行。研讨会在复旦大学**张力**研究员和**付彦伟**研究员的

主持之下，由复旦大学**姜育刚**教授致辞。随后浙江大学**杨易**教授，江南大学**吴小俊**教授，上海人工智能实验室**乔宇**教授，香港大学**韩楷**助理教授，复旦大学**张力**研究员，北京大学**田永鸿**教授，浙江大学**李玺**教授做主题报告。



杨易教授以“垂直领域应用中的大小模型协同技术”为题，指出如今大模型的迸发是 AI 领域不可避免的趋势之一。由于其海量训练数据，大模型对图片和文本具有良好的归纳能力，但是现在的通用多模态大模型还没有达到人们想象中的优越性能；而训练或者微调大模型所需要的电力和硬件以及时间成本是相对巨大的，这就导致大模型的灵活性较弱。对于视觉任务而言，由于其多样性和长尾效应，在一些视觉子任务中大模型难以被直接用于推理，而小模型以其灵活性和相对较好的可解释性则恰好弥补这一缺点，通过大模型蒸馏得到的知识往往能够提高小模型的能力和扩展其应用场景。但对于不同的任务，我们需要构造不同的大小模型协同方式。因此，探究如何做到充分利用大模型和小模型的协同作用是一个具有良好发展前景和极高价值的方向。



吴小俊教授的报告主题为“小模型 vs.大模型：以多模态视觉为例”。此次报告首先介绍了智慧城市中信息模态的多样性，例如红外光、可见光、深度信息等等，讲解了图像融合任务对于处理数据冗余和抽取信息的重要性及必要性。并且简单介绍了传统模式识别方法在图像融合领域的应用，进而引出近些年来大模型和深度学习在图像融合领域应用的发展，深入浅出的讲解了传统模式识别和深度学习在该领域的关系。吴教授介绍了他们团队发表在 TPAMI 的工作 LRRnet。不同于传统的 CNN 和 Residual Block, LRRnet 一种是针对图像融合任务设计的基于深度学习框架的新型网络架构。最后，吴教授总结到大模型和小模型的思想和方法论有重要的区别，要根据具体的问题选择和设计模型，对视觉融合问题亦然；即使 AI 进入大模型时代，也不要丢到传统的 CV、PR 和 ML。



乔宇教授以“书生通用智能大模型--我的大模型之路--”为题，生动地讲述通用大模型领域取得的突破性进展，特别是上海人工智能实验室在该领域的相关工作。他的报告从专用模型说起，梳理了从 2011 年以来，从针对特定任务的专用模型到应对多种任务多种模态的通用大模型的发展历程。他先介绍了在人脸识别领域视觉大模型的训练突破，以及近些年来在手机，安防，以及闸机等方面实现人脸识别模型的跨场景通用。他重点介绍了通用多模态大模型“书生”。这是一个整合了语言、图像、视频等多种模态的通用大模型，实现通过自然语言定义视觉任务，并具有多模态交互和跨模态生成能力，能够完成复杂场景下的各种 AI 任务。然后他对比了语言和视觉功能在人类进化中的历史，并从脑科学的角度展示了两类神经元之间的比例，提出了一个开放式的问题：“视觉比语言更加复杂，甚至更加重要？”最后他总结了通用视觉模型的主干网络的设计和训练的挑战，并展示了书生·浦语和书生·天际等基础模型在语言推理和实景三维重建中的应用。



韩锴助理教授的报告主题为“开放世界学习：类别发现，人体数字化和内容生成”，他首先介绍了如何利用先验知识减少聚类的模糊性，从无标签数据中发掘新类别并提高新发现类的质量的方法。随后针对以往人体重建算法中存在的难以重建衣物以及头发等问题，他介绍了他们组最近的工作 SeSDF，提出了利用自我进化的符号距离场从单个或多个未校准的图像中实现高质量人体三维重建。最后他介绍了近期结合生成式大模型 Stable Diffusion 在基于文本和形状引导的 3D Human Avatar 生成算法上的工作 DreamAvatar，该算法能进行个性化的文本到图像生成和文本到三维生成的方法，满足更多开放世界应用。



张力研究员报告的主题为“视觉基础模型与自动驾驶算法研究”。他的报告从视觉基础模型领域的工作出发，首先介绍了能够灵活建模视觉数据中远距离依赖关系的动态图消息传递网络。针对视觉 Transformer 架构的效率问题，他提出了具有线性复杂度的新型注意力机制，以及具有低延迟的移动端变体。此外，他还介绍了将语义分割任务建模为序列到序列预测的 SETR，该工作在国际上首次将 Transformer 架构应用于像素级理解任务中。随后，他重点介绍了今年发表在 CVPR 上的工作 GSS，这一工作创造性地将经典的语义分割任务纳入生成式框架，为统一各种视觉任务做出了初步探索。张力研究员指出，生成模型能够更完整地把握数据分布，因而有望突破当下视觉模型的规模化瓶颈，实现真正的大规模视觉基础模型。最后，张力研究员还介绍了他们团队近年来在自动驾驶领域取得的一系列重要成果，包括基于动态信息传递的融合深度的单目 3D 目标检测网

络 DDMP、鸟瞰视角下的多任务自动驾驶感知模型 Ego3RT，基于 Transformer 将相机视角下的图片特征转换到鸟瞰图极坐标网格中执行下游感知任务的 PolarFormer，模型间保留独立的多模态场景表征的 DeepInteraction，以及基于 NeRF 技术能够生成具有照片级真实感且可编辑的大规模无边街景视图以助力下游感知模型训练的自动驾驶场景仿真系统 S-NeRF 和 LiDAR-NeRF。



田永鸿教授报告的题目是“神经形态视觉：计算视觉的新范式”。田教授首先立足科技发展史，回顾了对大脑功能结构的借鉴为智能技术发展的带来的一系列启发。随后，他基于“摄像机的生物学原型是视网膜”这一事实，通过学习生物视网膜的结构和机制，突破了经典的视觉成像和表示理论，提出了具有高时域分辨率、高动态范围、低数据冗余和低功耗的神经形态传感器，并在此基础上发展了与之相适应的神经形态数据表征和处理方法。



李玺教授报告主题是“多模态视觉结构建模和特征学习”。他主要介绍了围绕数据驱动的人工智能学习方法，如 SAM 等视觉大模型算法，并从几何结构建模、动态推理结构、结构层面的图像生成等方面介绍了团队的相关工作。在车道线检测任务中，他们放弃了以往的密集预测方式，直接采用基于 anchor 的离散预测，从而大大提高了准确率和实时性。此外，他还介绍了一种利用动态路由实现样本自适应推断的新方法，以及基于自然语言文本引导的视觉模型权重生成框架。最后，他还介绍了布局图引导的图像生成算法 Layout Diffusion。

Panel 环节由复旦大学付彦伟研究员主持，参与嘉宾包括北京大学田永鸿教授、江南大学吴小俊教授、浙江大学李玺教授、复旦大学姜育刚教授、香港大学韩锴助理教授、复旦大学付彦伟研究员、复旦大学张力研究员。

大家就“在当下高校计算资源有限的情况下，要不要追当下生成式和大模型的发展热潮？”、“端到端的方法在自动驾驶领域的发展前景如何？”、“神经形态学的相关任务能否用当前数据驱动的范式来进行训练？能不能在比较成熟的模型上微调，从而完成神经形态学中的任务？”三个问题展开了全方位、多角度的深入讨论，与会专家与老师同学们进行了深入交流与探讨，最后论坛在大家热烈的思想碰撞中落下帷幕。



责任编辑 杨巨峰

CCF-CV 计算机视觉前沿讲习班

第二届CCF计算机视觉前沿讲习班 中国 重庆 2023.07



2023年7月26-27日，第二届CCF计算机视觉前沿讲习班在雾都重庆成功举办，活动共吸引了计算机视觉领域的研究人员、学生、工程师等200余人报名参加。本次活动由中国计算机学会（CCF）主办，中国计算机学会计算机视觉专委会（CCF-CV）、重庆邮电大学与西安电子科技大学联合承办，重庆邮电大学校长**高新波**教授与西安电子科技大学**王楠楠**教授担任执行主席。讲习班旨在促进计算机视觉领域的学术交流与高级人才培养，帮助该领域青年从业者提升技术水平，开拓实践视野，掌握最前沿的理论成果和创新应用。本届讲习班邀请八位知名专家报告前沿学术进展，帮助学员全面学习并系统掌握计算机视觉前沿相关技术。



讲习班开幕式首先由高新波校长致辞，高校长介绍了本次活动的目的、重庆邮电大学的基本情况、本次活动组织和报名情况，对讲者和学员表示热烈欢迎，执行主席**王楠楠**教授介绍了专委会的宗旨和讲习班的意义。中科院计算所**山世光**研究员发表感谢致辞，并为本次活动的承办方重庆邮电大学、以及组织活动的主席和组织委员颁发了感谢证书。

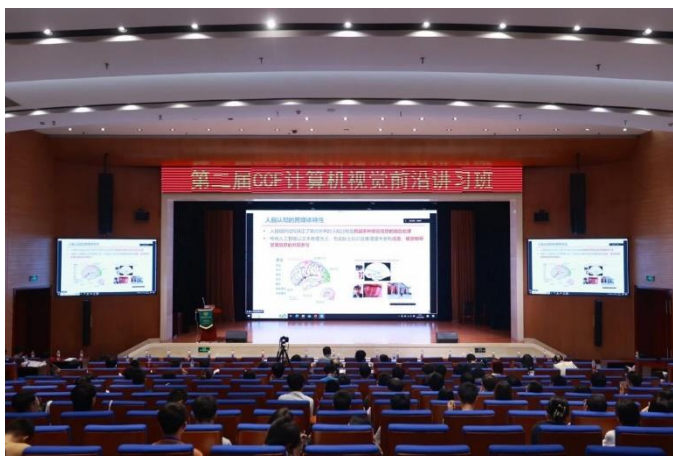


中国科学院计算所**山世光**研究员为学员讲授第一课。山老师以“基于视觉的心理感知技术”为题，首先阐述了在人脸识别逐步成熟的背景下，视觉感知技术对于情感理解的需求越来越高，模型不仅要会“看脸”，更需要会“读心”。然后，山老师从精神状况（长期）、心理状态（短期）、生理指标（瞬态）三个层次系统深刻地介绍了人的视觉感知技术之内涵和外延。随后，山老师重点从表情与面部动作检测和视线估计研究两个方面，介绍了课题组在相关领域的研究进展，并特别展示了相关数据集的构建工作及特殊儿童视线模式分析等相关应用。最后，山老师总结了基于视觉的情感感知现状与面临的挑战，并从知识驱动、大规模自监督学习、多模

态和上下文情感感知与理解等方面对未来发展趋势进行了展望。



南开大学程明明教授以“图像自适应感知”为题，从“粒度、算力、数据”三个方面深入探究了图像内容感知这一人工智能的基本任务。程教授首先阐述了现有视觉感知技术依然面临着特征表达粒度自适应性差所带来的算不准问题、算力自适应性差所带来的算不动问题、以及对有限标注的数据自适应性所带来的算不了的问题。然后，程教授从粒度自适应表征、算力自适应的高效计算、和数据自适应的无监督学习等角度出发，介绍针对上述问题的图像自适应感知技术的最新研究进展。最后，程教授特别提出了关于开源生态建设的DOCX行动倡议，包括Demo（方便科普和教学）、Open source（方便同行复现与验证）、Chinese（提供中文版本论文）和eXplain（及时回应项目主页上读者提问）。



北京大学彭宇新教授以“细粒度多模态协同感知、认知与生成”为题，首先介绍了互联网数据具有图像、

文本、视频、音频等多模态并存的特点，并阐述了现有多模态分析技术通常聚焦于粗粒度的大类，难以满足医疗、交通等诸多领域精细化需求的问题，从而引出细粒度多模态协同感知、认知与生成旨在使计算机能够对多模态内容进行精细化分析。然后，彭教授强调了借鉴人脑的跨模态特性，通过挖掘并协同多源、互补、关联的细粒度和多模态信息，使得计算机从能看会认的“感知智能”迈向能理解会思考的“认知智能”的研究目标，并从细粒度辨识增强、多模态关联、多模态协同、跨模态生成四个方面介绍课题组在细粒度图像分类、行人再识别、细粒度视频检索、细粒度跨模态检索、跨模态推理、文本到视觉生成上的最新研究进展。最后，彭教授对细粒度多模态分析的未来挑战进行了分析与展望。



中国科学院自动化所张兆翔研究员以“全域视觉场景感知”为题，首先针对视觉场景感知这一对于探索场景表征与认知计算具有重要意义的经典问题，系统介绍并分析了其广泛的应用领域与相关人工智能算法的研究进展。然后，张老师着重介绍了视觉场景感知的基本概念，具体阐述联合多数据类型、多监督类型，多汇聚类型的全域场景感知方法；并提出致力于探索面向高性能、少标注、自适应为主要特点的全域视觉感知与学习新理论，在2D、3D、4D视觉感知的范式下从模型、数据、知识等层面开展研究，探索全域视觉场景中的自主整合机制，汇聚一体化全域自主学习理论范式。最后，张老师从联合数据的视觉场景感知、4D融合的感知模型、终身学习与多任务联合机制等方面对未来研究趋势进行了展望。

第二届CCF计算机视觉前沿讲习班 @重庆邮电大学

书生·通用智能大模型 INTERN: A System of Foundation Models

—我的大模型之路—

乔宇
上海人工智能实验室
2023年7月27日



上海人工智能实验室**乔宇**教授以“书生·通用智能大模型——我的大模型之路”为题，首先介绍通用视觉大模型领域的最新进展，特别是详细介绍通用视觉大模型书生取得的进展。然后深入探究视觉与语言之间的区别与联系，并介绍了视觉大模型与语言大模型的结合以及未来的发展趋势。在本次分享中，乔教授在大模型引领人工智能进入跨越式发展的背景下，介绍了对于“规则-技术-场景-评测”一体协同治理创新体系的探索。

浙江大学**杨易**教授以“多重知识驱动的视觉理解与生成技术”为题，介绍了融合多种知识实现图像和视频高效理解和生成的技术。通过将结构化特征与深度学习方法相结合，可以提高图像和视频的理解和生成效果。针对当前单一深度学习表达单一、推理能力缺乏的问题，多重知识表达理论旨在建立多源互补多重互联、异构贯通的新型跨媒体表达。在本报告中，杨教授首先分析多重知识表达理论的优势与特点，然后面向实际应用具体分析视觉关系理解方法，结合案例讨论视觉知识的分解、类比、推理等高阶操作。最后，杨教授强调大小模型协同是多重知识表达在大模型时代的一种实践，并讨论通过使用深度生成模型，实现高质量图像和视频生成的方法，这类方法为游戏、虚拟现实和动画等领域带来了新的应用。



华中科技大学**白翔**教授以“Deep Neural Networks for Scene Text Reading”为题，介绍了场景文字检测与识别这一 OCR (Optical Character Recognition, 光学字符识别) 有关应用中的核心技术。白教授回顾了近十年来场景文字检测与识别方向的里程碑式工作及其发展脉络，很好地帮助同学们全面快速地了解场景文字检测与识别这一关键技术。在本次报告中，白教授从视觉与语言预训练模型出发，结合开放集合图像语义分割任务，以文本检测技术为主干，扩展介绍了人群计数领域的研究进展，最后进行了总结与讨论。



华为**谢凌曦**高级研究员以“Towards AGI in Computer Vision: Lessons Learned from GPT and Large Language Models”为题，深入探讨了计算机视觉迈向通用人工智能 (AGI) 所遇到的瓶颈问题，希望通

通过分析 NLP 领域中大语言模型驱动的对话系统具有强大能力的本质原因,借鉴 NLP 的经验为 CV 领域寻求一条发展 AGI 的有效路径。在这次报告中,谢老师从 AGI 的定义和 NLP 所达到的成就为引子,揭示 AGI 的关键在于设计大一统模型。随后,谢老师简要回顾 CV 领域在大一统方面的努力,并指出 CV 的本质困难在于无法建立起有效的交互环境,长期以来只能依靠代理任务来推进研究,而这种范式已经走到尽头。在此基础上,谢老师设想了未来 CV 的新范式,即从建立大规模交互环境,通过生成式预训练和指令微调的方式来学习,并展望几个重要的研究方向。

第二届 CCF 计算机视觉前沿讲习班为计算机视觉

领域的广大师生和工程师提供了一个与专家学者近距离交流学习的宝贵机会,大家对专家所讲的课程内容产生了极大兴趣,现场互动非常活跃,学术氛围浓厚。

在结业典礼上,执行主席王楠楠教授对本次活动进行了总结。首先祝贺本次活动在重庆邮电大学举办取得了圆满成功,然后征集学员意见,请大家为下一届讲习班的组织工作提供建议。最后,为学员颁发学员证书。全体学员合影留念,记录下了自己在第二届 CCF 计算机视觉讲习班顺利结业的瞬间。

责任编辑 潘金山

CCF-CV 计算机视觉前沿进展研讨会



2023年7月24日,中国计算机学会计算机视觉专委会(CCF-CV)年度学术研讨会 RACV (Recent Advances on Computer Vision) 在重庆成功召开。RACV 定位为国内计算机视觉领域的小规模精品研讨会,

通过定向邀请方式汇集领域专家,深度研讨计算机视觉领域中的若干核心问题并形成进展报告。研讨会试图通过务实、开放与平等的对话与讨论,深入发掘相关研究领域潜在的问题,为广大的科研人员提供观察问题的新视角与新观点。



本次会议开幕式由专委会副主任、中国科学院自动化研究所**王亮**研究员主持，中国计算机学会副理事长、清华大学**胡事民**教授，专委会主任、北京大学**查红彬**教授，重庆邮电大学**王国胤**副校长致开幕辞。根据专委会前期的讨论票选，本次会议设置了 3 项研讨主题。每项主题首先由特邀嘉宾进行主题发言，之后所有与会人员进行自由讨论。



24 日上午首先进行了专题一“可信人工智能在计算机视觉中的前沿进展”的研讨。该专题由专委会常委、华中科技大学**白翔**教授，专委会执行委员、西安电子科技大学**王楠楠**教授，北京邮电大学**梁孔明**副研究员组织，邀请了重庆邮电大学**高新波**教授、中山大学**操晓春**教授、北京航空航天大学**徐迈**教授、香港中文大学（深圳）**吴保元**副教授 4 位嘉宾进行主题发言。近几年，可信人工智能吸引了大量学者的关注，与会嘉宾围绕可信人工智能的内涵和外延、实现路径、挑战与机遇等问题进行了精彩的讨论和观点分享。



24 日下午进行了专题二“大模型&ChatGPT 对计算机视觉的影响”的研讨。该专题由专委会常委、百度

王井东博士，专委会执行委员、华中科技大学**王兴刚**教授，专委会执行委员、上海人工智能实验室**李弘扬**博士组织，邀请了中国科学院计算技术研究所**陈熙霖**研究员、浙江大学**沈春华**教授、清华大学**代季峰**副教授、华为云**谢凌曦**博士 4 位嘉宾进行主题发言。嘉宾们围绕大模型对传统计算机视觉任务的冲击与影响、大模型如何定义与应用、大模型与垂直行业模型的对比、大模型的安全性可解释性、新形势下学术界与工业界如何分工等议题展开了深入探讨。



24 日下午继续进行了专题三“人工智能内容生成”的研讨。该专题由专委会执行委员、中国科学院自动化研究所**赫然**研究员，专委会常委、南开大学**程明明**教授、大连理工大学**李祎**副教授组织，邀请了浙江大学**周昆**教授、北京大学**陈宝权**教授、清华大学**刘焯斌**教授、国防科大**徐凯**教授 4 位嘉宾进行主题发言。嘉宾们围绕人工智能内容生成的核心科学问题及在计算机视觉领域的角色、技术路线、评价标准、发展前景等议题展开了深入探讨。



最后，专委会主任、北京大学**查红彬**教授代表专委会感谢承办方、赞助方和参会人员。专委会副主任、南京邮电大学**刘青山**教授对活动做了总结并宣布研讨会闭幕。本次研讨会深入探讨了本领域最前沿研究问题，主题发言视角广阔，自由讨论热情激烈，参会嘉宾们纷纷表示本次会议内容丰富，收获良多。按照计划，组委会后续将整理相关主题的发言与讨论文稿，形成观点性

文档进行发布，把讨论从线下延伸到线上，欢迎更多专家学者积极参与。本次研讨会由重庆邮电大学高新波教授和西安电子科技大学王楠楠教授团队承办，百度提供独家赞助。

责任编辑 任传贤 黄岩

热点追踪

基于极大团的三维配准方法

西北工业大学 张曦予 杨佳琪* 张世坤 张艳宁

本文是西北工业大学团队解读其在 CVPR 2023 获得最佳学生论文奖的工作 MAC。论文研究的问题是三维点云配准 (PCR)，旨在寻找最佳姿态以对齐点云对。由于现有的 3D 关键点检测器和描述符的局限性，点云之间的有限重叠以及数据噪声，特征匹配通常包含外点，对精确的点云配准产生极大挑战。我们提出了一种基于极大团 (MAC) 的三维配准方法。首先，我们提出了一种名为 MAC 的假设生成方法。与以前的最大团约束相比，我们的 MAC 方法能够在图中挖掘更多的局部信息。即使存在严重的异常值，由 MAC 生成的假设也具有非常高的准确性。其次，基于 MAC 我们提出了一种新的点云配准方法，该方法在 U3M、3DMatch、3DLoMatch、KITTI 和 ETH 数据集上实现了最先进的性能。值得注意的是，我们的仅凭几何的 MAC 方法优于几种最先进的深度学习框架^[1-4]。MAC 还可以作为模块插入多个深度学习框架^[5-9]中，以提高它们的性能。

一、研究背景

我们研究的是计算机视觉中的一个经典问题——三维点云配准：寻找最佳姿态以对齐点云对。为了解决三维点云配准问题，有传统几何与深度学习两类方法。

传统几何方法^[10-14] RANSAC 从初始匹配集中迭代地采样匹配，为每个子集生成和评估几何估计，直到得到最佳的变换。尽管基于 RANSAC 的方法简单高效，但在高外点率情况下，它们的性能非常脆弱，并且需要大量迭代才能获得可接受的结果。为了解决当前评估指标耗时长以及对噪声敏感的问题，Yang 等人^[15]在计算过程中分析了内点和外点的贡献，并提出了几种可以有效提高 RANSAC 配准性能的指标。最近也出现了一系

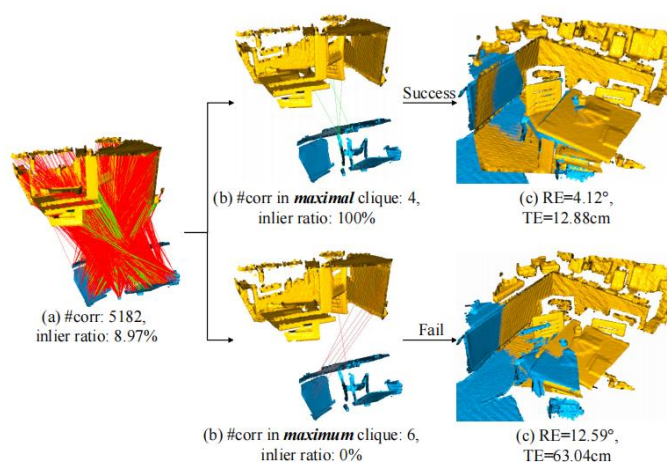


图 1 极大团与最大团在低重叠率点云数据上的区别

列基于 BnB 的全局配准方法。Bustos 和 Chin^[16]提出了保证的异常值去除 (GORE)，它计算每个匹配的紧密下界和紧密上界，并通过拒绝真正的外点来减少匹配集的大小。但这些方法的计算复杂度较高，尤其是当匹配集很大并且外点率极高时。

深度学习方法 除了仅使用几何方法外，最近的工作还采用了深度学习技术来进行三维点云配准。一些方法旨在检测更多可重复的关键点^[17,18]，并提取更多的描述性特征^[5,6]。Choy 等人^[6]提出的 FCGF 在没有关键点检测的情况下，通过全卷积神经网络在单次通过中计算特征。Bai 等人^[17]提出的 D3Feat 使用全卷积网络来获得点云的局部信息，并使用联合学习框架来实现 3D 局部特征检测和描述。为了对点云提取描述能力更强的特征，Ao 等人^[5]提出 SpinNet 提取局部特征，这些局部特征是旋转不变的并且具有足够的信息量以实现准确的配准；为了有效区分匹配中的内点和外点，Bai 等人^[1]提出 PointDSC 明确探索空间一致性，以消除异常匹配并使点云配准成功；为了遵循无检测方法，并实现端

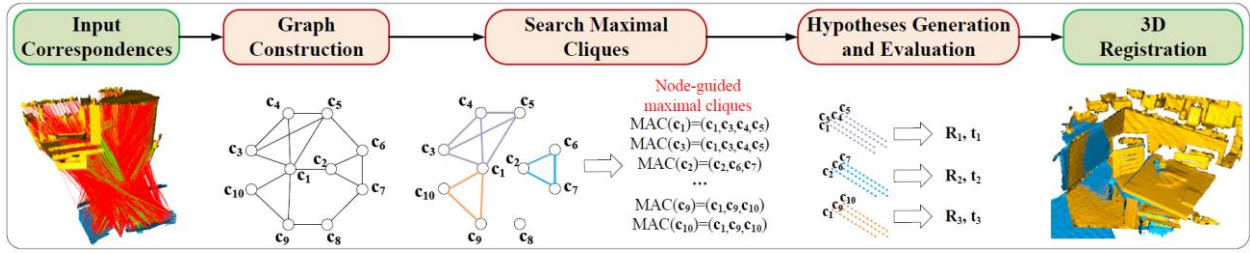


图 2 MAC 的流程图。a) 将初始匹配建模成兼容性图；b) 从兼容性图中搜索极大团作为兼容匹配集合；c) 根据兼容匹配集合生成并评估假设；d) 最佳假设作为配准结果。

到端方式的估计变换，Qin 等人^[8]提出 GeoTransformer 学习几何特征以进行强大的超点匹配，在低重叠情况下具有鲁棒性并且对刚性变换具有不变性。但是基于深度学习的方法需要大量的数据进行训练，并且通常在不同的数据集上缺乏泛化能力。

二、MAC方法介绍

2.1. 方法框架介绍

为了使两个点云 P^s 和 P^t 对齐，我们首先使用几何或深度学习描述子为它们提取局部特征。让 p^s 和 p^t 分别表示 P^s 和 P^t 中的点。通过匹配特征描述符，形成初始匹配集 $C_{initial} = \{c\}$ ，其中 $c = (p^s, p^t)$ 。MAC 从输入的 $C_{initial}$ 估计 P^s 和 P^t 之间的 6-DoF 姿态变换。方法的框架如图 2 所示。

2.2. 构建兼容性图

由于图空间 $G = (V, E)$ 相较于欧几里得空间可以更准确地描述匹配之间的亲和关系，因此我们将初始匹配建模为兼容性图，其中匹配由节点表示，几何上兼容的点由边连接。兼容性图的构建方式有以下两种：

1) 一阶图 (FOG)：FOG 是基于匹配对之间的刚性距离约束构架的，这种约束可以定量测量为：

$$S_{dist}(c_i, c_j) = ||p_i^s - p_j^s| - |p_i^t - p_j^t||. \quad (1)$$

c_i 和 c_j 兼容性分数的计算方式为：

$$S_{cmp}(c_i, c_j) = \exp\left(-\frac{S_{dist}^2(c_i, c_j)}{2d_{cmp}^2}\right), \quad (2)$$

其中 d_{cmp} 是一个距离参数。若 $S_{cmp}(c_i, c_j)$ 大于阈值， c_i 和 c_j 之间形成边 e_{ij} 并且 $S_{cmp}(c_i, c_j)$ 是 e_{ij} 的权值；否则， $S_{cmp}(c_i, c_j)$ 被设为 0。由于兼容性图是一个无向图，图的权重矩阵是一个对称矩阵。

2) 二阶图 (SOG)：之前的研究^[11]提出了二阶兼容性度量，它与全局集当中的共同兼容匹配的数量有关。SOG 由 SOG 演变而来，其权重矩阵 W_{SOG} 的计算方式为：

$$W_{SOG} = W_{FOG} \odot (W_{FOG} \times W_{FOG}), \quad (3)$$

其中 \odot 表示矩阵按元素相乘。

如图 3 所示：SOG 考虑了匹配对全局集中的共同兼容性，而不仅仅是几何一致性，使其在高外点率的情况下更一致和鲁棒；SOG 比 FOG 更稀疏，因此有助于团搜索更加快速。基于以上两点，MAC 构建 SOG 来表示匹配的兼容性关系。

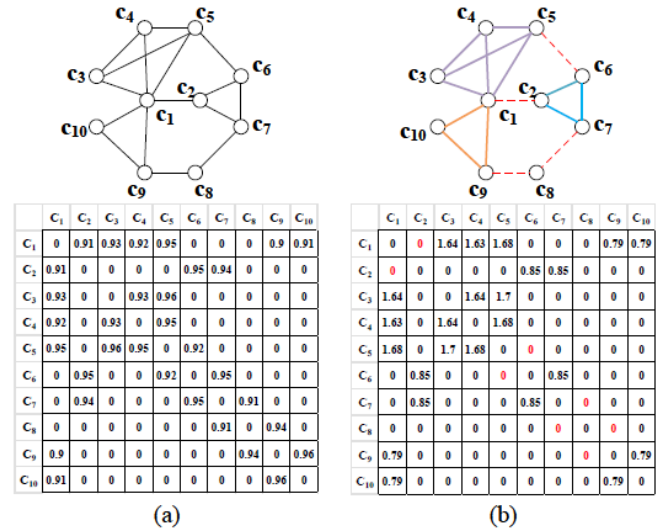


图 3 阐述一阶图 FOG 与二阶图 SOG 之间联系的示例。

(a) FOG 及其邻接矩阵； (b) SOG 及其邻接矩阵。

2.3. 搜索极大团

给定一个无向图 G ，团 $C = (V', E')$ ， $V' \subseteq V, E' \subseteq E$ 是 G 的子集，其中任意两个节点都由边连接。极大团是不能通过添加任何节点来扩展的团。特别地，具有最多节点数的极大团是图的最大团。理论上，内点会在图中形成团，因为内点通常在几何上彼此兼容。之前的研究^[23]专

注于在图中搜索最大团，但是最大团是非常严格的约束，只关注图中的全局共识信息。相反，我们放宽了约束，并利用极大团来挖掘更多的局部图信息。极大团的搜索使用 igraph 完成。

执行极大团搜索后，我们获得了极大团集合 $MAC_{initial}$ 。在实践中， $MAC_{initial}$ 通常包含数万个极大团，如果我们考虑所有极大团将会非常耗时。本节中，我们引入了一种节点引导的团筛选方法来减少 $|MAC_{initial}|$ 。首先，我们计算集合中每个团的权重。给定一个团 $C_i = (V_i, E_i)$ ，权重 $w(C_i)$ 计算如下：

$$w(C_i) = \sum_{e_j \in E_i} w(e_j), \quad (4)$$

其中 $w(e_j)$ 表示 W_{SOG} 中边 e_j 的权重。一个节点可能被周围的多个极大团包含，我们只保留局部权重最大的团与节点对应。然后，对剩余的团去重，获得极大团集合 $MAC_{selected}$ 。此步骤的动机是使用图节点周围局部几何结构的信息来找到最佳一致性的对应节点集。很明显， $|MAC_{selected}|$ 不会超过 $|V|$ 。我们可以将这些极大团直接送到下一阶段进行三维点云配准。

2.4. 假设生成与假设

从上一步筛选出的每个极大团都代表了一组一致的匹配。通过对每个一致性集应用 SVD 算法，我们可以获得一组 6-DoF 姿态假设。MAC 的最终目标是估计最大化如下目标函数的最优 6-DoF 刚性变换（由旋转姿态 $R^* \in SO(3)$ 和平移姿态 $t^* \in \mathbb{R}^3$ 组成）：

$$(R^*, t^*) = \arg \max_{R, t} \sum_{i=1}^N s(c_i), \quad (5)$$

其中 $c_i \in C_{initial}$ ， $N = |C_{initial}|$ 。我们在这里考虑了几个 RANSAC 假设评估指标^[15]，包括 MAE，MSE 和内点数。 $s(c_i)$ 表示一致性集中每个 c_i 的指标得分。拥有最高评估分数的假设被视为最佳假设，并作为最终变换以执行点云配准。

三、实验结果

3.1. 在模型数据集 U3M 上的结果

从图 4 中广泛比较的结果可以看出：MAC 性能最好，显著优于所有测试的 RANSAC 方式估计器，如 SAC-COT、OSAC、SAC-IA 和 RANSAC。基于 MAE 评估度

量的 MAC 在 U3M 上的配准性能最好。

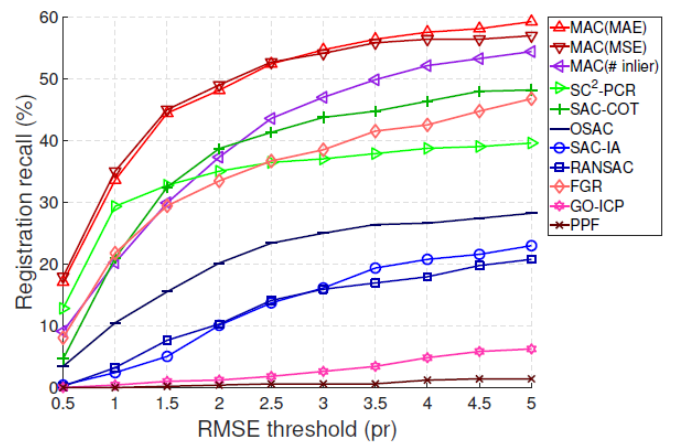


图 4 各方法在 U3M 数据集上的配准结果

3.2. 在室内场景数据集 3DMatch 和 3DLoMatch 上的结果

表 1 3DMatch 数据集上的配准结果

	FPFH			FCGF		
	RR(%)	RE(°)	TE(cm)	RR(%)	RE(°)	TE(cm)
<i>i) Traditional</i>						
SM [20]	55.88	2.94	8.15	86.57	2.29	7.07
FGR [45]	40.91	4.96	10.25	78.93	2.90	8.41
RANSAC-1M [13]	64.20	4.05	11.35	88.42	3.05	9.42
RANSAC-4M [13]	66.10	3.95	11.03	91.44	2.69	8.38
GC-RANSAC [5]	67.65	2.33	6.87	92.05	2.33	7.11
TEASER++ [36]	75.48	2.48	7.31	85.77	2.73	8.66
CG-SAC [30]	78.00	2.40	6.89	87.52	2.42	7.66
SC ² -PCR [8]	83.73	2.18	6.70	93.16	2.09	6.51
<i>ii) Deep learned</i>						
3DRegNet [27]	26.31	3.75	9.60	77.76	2.74	8.13
DGR [9]	32.84	2.45	7.53	88.85	2.28	7.02
DHVR [19]	67.10	2.78	7.84	91.93	2.25	7.08
PointDSC [3]	72.95	2.18	6.45	91.87	2.10	6.54
MAC	84.10	1.96	6.18	93.72	1.89	6.03

表 2 3DLoMatch 数据集上的配准结果

	FPFH			FCGF		
	RR(%)	RE(°)	TE(cm)	RR(%)	RE(°)	TE(cm)
<i>i) Traditional</i>						
RANSAC-1M [13]	0.67	10.27	15.06	9.77	7.01	14.87
RANSAC-4M [13]	0.45	10.39	20.03	10.44	6.91	15.14
TEASER++ [36]	35.15	4.38	10.96	46.76	4.12	12.89
SC ² -PCR [8]	38.57	4.03	10.31	58.73	3.80	10.44
<i>ii) Deep learned</i>						
DGR [9]	19.88	5.07	13.53	43.80	4.17	10.82
PointDSC [3]	20.38	4.04	10.25	56.20	3.87	10.48
MAC	40.88	3.66	9.45	59.85	3.50	9.75

配准方法对比 从表 1 和表 2 可以得出以下结论：

1) 无论使用哪种描述符，MAC 在 3DMatch 和 3DLoMatch 数据集上都优于所有比较方法，表明其强大的室内场景点云配准能力；2) 即使与深度学习方法进行比较，MAC 仍然在没有任何数据训练的情况下实现了更好的性能；3) 除了配准召回率 (RR) 指标外，MAC 实现了最佳的 RE 和 TE 指标。这表明 MAC 的配准非常准确，并且 MAC 能够对齐重叠度低的数据。

使用 MAC 增强深度学习性能 如表 3 所示, MAC 在 3DMatch 和 3DLoMatch 数据集上的所有测试方法中都显著提高了配准召回率。值得注意的是, SpinNet, Predator 和 CoFiNet 在 MAC 的增强下的表现超过了 GeoTransformer。MAC 与 GeoTransformer 配合使用, 在 3DMatch / 3DLoMatch 上实现了 95.7%/ 78.9% 的最新配准召回率。结果表明: 1) MAC 可以极大地提高现有的深度学习性能; 2) MAC 对样本数量不敏感。

表 3 3DMatch 与 3DLoMatch 数据集上深度学习方法与 MAC 结合后的性能提升

# Samples	3DMatch RR(%)					3DLoMatch RR(%)				
	5000	2500	1000	500	250	5000	2500	1000	500	250
FCGF [10]	85.1	84.7	83.3	81.6	71.4	40.1	41.7	38.2	35.4	26.8
SpinNet [11]	88.6	86.6	85.5	83.5	70.2	59.8	54.9	48.3	39.8	26.8
Predator [18]	89.0	89.9	90.6	88.5	86.6	59.8	61.2	62.4	60.8	58.1
CoFiNet [43]	89.3	88.9	88.4	87.4	87.0	67.5	66.2	64.2	63.1	61.0
GeoTransformer [29]	92.0	91.8	91.8	91.4	91.2	75.0	74.8	74.2	74.1	73.5
FCGF+MAC	91.3	92.2	91.6	90.4	85.6	57.2	56.0	52.6	42.4	32.1
	6.2↑	7.5↑	8.3↑	8.8↑	14.2↑	17.1↑	14.3↑	14.4↑	7.0↑	5.3↑
SpinNet+MAC	95.3	95.1	93.3	91.4	81.2	72.8	69.9	59.2	54.8	32.1
	6.7↑	8.5↑	7.8↑	7.9↑	11.0↑	13.0↑	15.0↑	10.9↑	15.0↑	5.3↑
Predator+MAC	94.6	94.4	94.0	93.5	92.3	70.9	70.4	69.8	67.2	64.1
	5.6↑	4.5↑	3.4↑	5.0↑	5.7↑	11.1↑	9.2↑	7.4↑	6.4↑	6.0↑
CoFiNet+MAC	94.1	94.4	94.5	93.8	92.7	71.6	71.5	70.6	69.2	68.1
	4.8↑	5.5↑	6.1↑	6.4↑	5.7↑	4.1↑	5.3↑	6.4↑	6.1↑	7.1↑
GeoTransformer+MAC	95.7	95.7	95.2	95.3	94.6	78.9	78.7	78.2	77.7	76.6
	3.7↑	3.9↑	3.4↑	3.9↑	3.4↑	3.9↑	3.9↑	4.0↑	3.6↑	3.1↑

3.3. 在室外场景数据集 KITTI 上的结果

表 4 KITTI 数据集上的配准结果

	FPFH			FCGF		
	RR(%)	RE(°)	TE(cm)	RR(%)	RE(°)	TE(cm)
<i>i) Traditional</i>						
FGR [45]	5.23	0.86	43.84	89.54	0.46	25.72
TEASER++ [36]	91.17	1.03	17.98	94.96	0.38	13.69
RANSAC [13]	74.41	1.55	30.20	80.36	0.73	26.79
CG-SAC [30]	74.23	0.73	14.02	83.24	0.56	22.96
SC ² -PCR [8]	99.28	0.39	8.68	97.84	0.33	20.58
<i>ii) Deep learned</i>						
DGR [9]	77.12	1.64	33.10	96.90	0.34	21.70
PointDSC [3]	<u>98.92</u>	0.38	8.35	97.84	0.33	20.32
MAC	99.46	0.40	<u>8.46</u>	97.84	<u>0.34</u>	<u>19.34</u>

从表 4 可以看出: 就配准召回性能而言, MAC 表现最好。MAC 也具有比现有技术的仅几何方法 SC²-PCR 更低的 TE。值得注意的是, 室外点云非常稀疏且分布不均匀。在物体、室内场景和室外场景数据集上的配准实验一致验证了 MAC 在不同应用场景下具有良好的泛化能力。

3.4. 在室外场景数据集 ETH 上的结果

表 5 不同方法在 ETH 数据集上生成匹配的内点率

	Gazebo		Wood		Avg.
	Autumn	Summer	Autumn	Summer	
FPFH [7]	0.42	0.24	0.21	0.26	0.29
FCGF [2]	2.34	1.25	1.35	1.68	1.62
Spinnet [1]	16.67	13.73	12.20	14.67	14.40

表 6 ETH 数据集各方法与 MAC 结合后的性能提升

	Gazebo		Wood		Avg.
	Autumn	Summer	Autumn	Summer	
FPFH [7]	16.85	10.03	10.43	10.40	11.92
FCGF [2]	54.35	28.03	52.17	51.20	42.78
Spinnet [1]	98.37	83.05	100.00	99.20	92.57
FPFH+MAC	46.74	27.68	33.04	43.20	36.12
	29.89↑	17.65↑	22.61↑	32.80↑	24.20↑
FCGF+MAC	75.54	42.91	71.30	73.60	61.29
	21.19↑	14.88↑	19.13↑	22.40↑	18.51↑
Spinnet+MAC	98.91	87.54	100.00	100.00	94.67
	0.54↑	4.49↑	-	0.80↑	2.10↑

与 3DMatch 相比, ETH^[19]包含更复杂的几何形状。FPFH^[20]、FCGF 和 SpinNet 用于生成匹配, 然后 RANSAC-50K 和 MAC 将根据这些匹配执行配准。采样点或匹配的数量统一设置为 5000。当 $RE \leq 15^\circ$ 和 $TE \leq 30$ cm 时, 配准被视为成功。表 5 和表 6 分别报告了生成匹配的质量和配准结果。

结果表明, 当描述符中的缺陷导致生成匹配集的内点率非常低时, MAC 仍然可以有效地从中找到准确的一致子集, 从而大大提高配准召回率。当与 FPFH 组合时, 使用 MAC 获得的配准召回率比 RANSAC 高 24.2%, 与 FC 组合时高 18.51%。MAC 与 SpinNet 结合在 ETH 上实现了 94.67% 的配准召回率。

3.5. MAC 生成假设的质量

表 7 RANSAC 和 MAC 生成正确假设数量的比较

# hypotheses	3DMatch				3DLoMatch			
	RANSAC		MAC		RANSAC		MAC	
	FCGF	FPFH	FCGF	FPFH	FCGF	FPFH	FCGF	FPFH
100	10.45	0.76	61.94	50.67	1.25	0.05	30.47	12.22
200	20.76	1.50	119.20	89.27	2.52	0.09	55.57	17.59
500	51.74	3.68	269.06	162.41	6.21	0.21	109.32	23.32
1000	103.65	7.39	456.18	217.32	12.43	0.41	156.11	26.02
2000	208.24	14.90	669.32	254.13	24.80	0.81	202.12	29.31

我们通过将 RANSAC 和 MAC 的假设与真值变换进行比较来评估生成的假设的质量。结果如表 7 所示。

与 RANSAC 在没有几何约束的情况下随机选择匹配并从中生成假设相比, MAC 有效地从兼容性图中的极大团中生成了更令人信服的假设, 充分利用了图中的一致性信息。

3.6. MAC 的性能上限

我们通过改变生成的正确假设数量的判断阈值来测试 MAC 的性能上限, 结果如表 8 所示。令人印象深刻的是, MAC-1 在 3DMatch / 3DLoMatch 上实现了 98.46% / 91.24% 的配准召回率。这表明, 即使在重叠

度较低的数据集上，MAC 也能够为大多数点云对生成正确的假设。此外，我们可以推断出，使用更好的假设评估度量可以进一步提高 MAC 的性能。

表 8 性能上限探究。MAC-n 表示至少生成 n 个假设即认为可成功配准。

	3DMatch RR(%)	3DLoMatch RR(%)
MAC-1	98.46	91.24
MAC-5	97.10	83.32
MAC-10	96.43	77.93
MAC-20	94.70	70.47
MAC-50	91.13	56.37
MAC-origin	93.72	59.85

3.7. 时间和空间效率分析

表 9 平均耗时对比

# Corr.	250	500	1000	2500	5000
PointDSC	32.24±0.81	78.38±0.89	240.46±2.18	1401.97±12.24	5504.11±10.32
TEASER++	6.40±1.88	6.68±0.66	16.74±1.21	104.24±0.53	484.93±1.87
SC ² -PCR	19.34±0.63	63.23±0.55	215.98±1.24	1282.73±4.05	5210.17±8.30
MAC	7.32±0.55	23.32±0.38	56.45±1.41	282.67±7.83	3259.38±12.66

表 10 平均内存消耗对比

# Corr.	250	500	1000	2500	5000
PointDSC	3531.46	3538.26	3582.57	3634.22	3736.10
TEASER++	1631.92	1634.77	2029.22	2266.84	2484.83
SC ² -PCR	448.01	453.18	508.40	621.27	690.22
MAC	15.59	17.43	23.49	52.79	150.86

表 9 和表 10 分别显示了几种执行良好的方法的时间效率和内存消耗结果。关于时间效率实验，所有方法都进行了十轮测试，并报告了平均值和标准差结果。

所有方法都仅使用 CPU 进行测试。结果表明，当输入匹配数小于 2.5k 时，MAC 是非常轻量和高效率的。

3.8. 可视化结果

在图 5 和图 6 中我们展示了更多可视化结果。图 5 表明在低内点率数据上，MAC 能够进行高精度配准；图 6 表明 MAC 能够处理其他方法配准失败的低重叠率数据。

四、总结

我们提出了 MAC 来解决点云配准问题，通过使用极大团约束从匹配中生成精确的姿态假设。我们的方法在所有测试的数据集上都实现了最先进的性能，并且可以适应深度学习方法来进一步提高其性能。

责任编辑 金鑫 王金甲

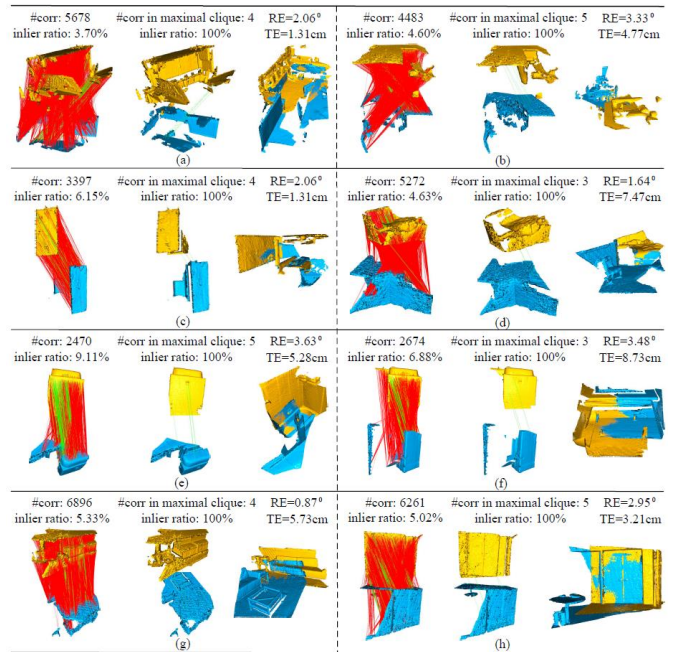


图 5 MAC 在 3DMatch 数据集上的配准过程可视化。红色和绿色的线分别代表外点和内点。

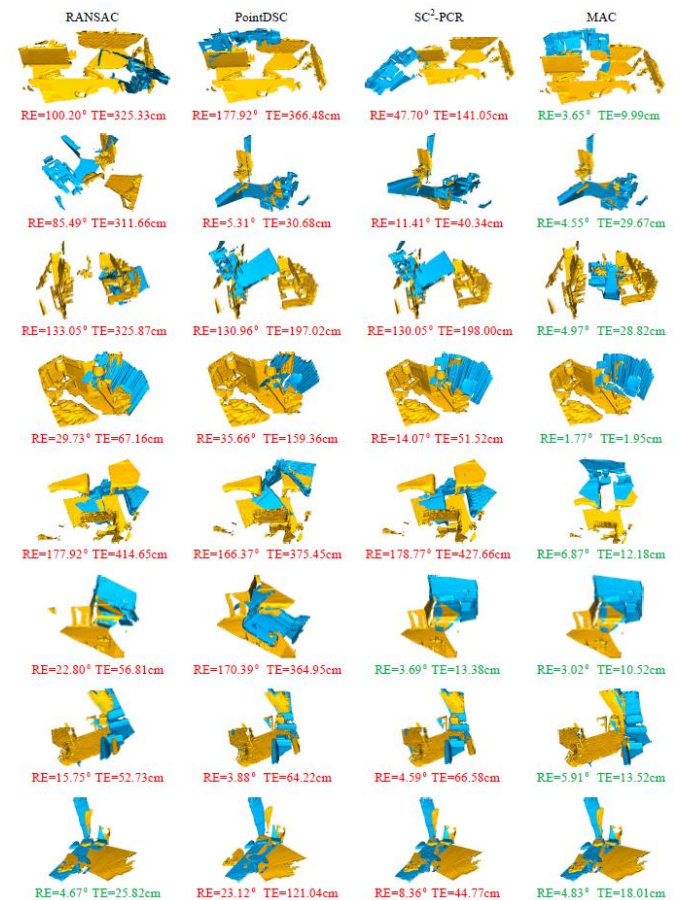
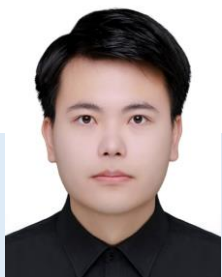


图 6 不同方法在 3DLoMatch 数据集上的配准结果对比。红色和绿色分别代表配准失败与配准成功。

参考文献

- [1] Xuyang Bai, Zixin Luo, Lei Zhou, Hongkai Chen, Lei Li, Zeyu Hu, Hongbo Fu, and Chiew-Lan Tai. Pointdsc: Robust point cloud registration using deep spatial consistency. In CVPR 2021.
- [2] Christopher Choy, Wei Dong, and Vladlen Koltun. Deep global registration. In CVPR 2020.
- [3] Junha Lee, Seungwook Kim, Minsu Cho, and Jaesik Park. Deep hough voting for robust global registration. In CVPR 2021.
- [4] G Dias Pais, Srikumar Ramalingam, VenuMadhav Govindu, Jacinto C Nascimento, Rama Chellappa, and Pedro Miraldo. 3Dregnet: A deep neural network for 3d point registration. In CVPR 2020.
- [5] Sheng Ao, Qingyong Hu, Bo Yang, Andrew Markham, and Yulan Guo. Spinnet: Learning a general surface descriptor for 3d point cloud registration. In CVPR 2021.
- [6] Christopher Choy, Jaesik Park, and Vladlen Koltun. Fully convolutional geometric features. In CVPR 2019.
- [7] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3D point clouds with low overlap. In CVPR 2021.
- [8] Zheng Qin, Hao Yu, Changjian Wang, Yulan Guo, Yuxing Peng, and Kai Xu. Geometric transformer for fast and robust point cloud registration. In CVPR 2022.
- [9] Hao Yu, Fu Li, Mahdi Saleh, Benjamin Busam, and Slobodan Ilic. Cofinet: Reliable coarse-to-fine correspondences for robust point cloud registration. In CVPR 2021.
- [10] XAlvaro Parra Bustos and Tat-Jun Chin. Guaranteed outlier removal for point cloud registration with correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(12):2868–2882, 2017.
- [11] Zhi Chen, Kun Sun, Fan Yang, and Wenbing Tao. Sc²-pcr: A second order spatial compatibility for efficient and robust point cloud registration. In CVPR 2022.
- [12] Marius Leordeanu and Martial Hebert. A spectral technique for correspondence problems using pairwise constraints. 2005.
- [13] Heng Yang, Jingnan Shi, and Luca Carlone. Teaser: Fast and certifiable point cloud registration. IEEE Transactions on Robotics, 37(2):314–333, 2020.
- [14] Qian-Yi Zhou, Jaesik Park, and Vladlen Koltun. Fast global registration. In ECCV 2016.
- [15] Jiaqi Yang, Zhiqiang Huang, Siwen Qian, Qian Zhang, Yanning Zhang, and Zhiguo Cao. Toward efficient and robust metrics for ransac hypotheses and 3D rigid registration. IEEE Transactions on Circuits and Systems for Video Technology, 893–906, 2021.
- [16] Alvaro Parra Bustos and Tat-Jun Chin. Guaranteed outlier removal for point cloud registration with correspondences. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2868–2882, 2017.
- [17] Xuyang Bai, Zixin Luo, Lei Zhou, Hongbo Fu, Long Quan, and Chiew-Lan Tai. D3feat: Joint learning of dense detection and description of 3d local features. In CVPR 2020.
- [18] Shengyu Huang, Zan Gojcic, Mikhail Usvyatsov, Andreas Wieser, and Konrad Schindler. Predator: Registration of 3d point clouds with low overlap. In CVPR 2021.
- [19] François Pomerleau, Ming Liu, Francis Colas, and Roland Siegwart. Challenging data sets for point cloud registration algorithms. The International Journal of Robotics Research, 31(14):1705–1711, 2012.
- [20] Radu Bogdan Rusu, Nico Blodow, and Michael Beetz. Fast point feature histograms (fpfh) for 3D registration. In IEEE International Conference on Robotics and Automation, 3212–3217, 2009.



张曦予

西北工业大学计算机学院 2022 级硕士研究生，导师为杨佳琪副教授，主要研究方向为三维特征匹配与点云配准。

Email: 2426988253@mail.nwpu.edu.cn



杨佳琪

西北工业大学计算机学院的副教授，研究方向包括局部几何描述、三维配准、三维特征匹配和三维物体识别。他分别于 2014 年和 2019 年在华中科技大学获得学士学位和博士学位。在中国国家留学基金委员会的资助下，他于 2017 年至 2018 年期间访问了宾夕法尼亚大学的 GRASP 实验室。

Email: jqyang@nwpu.edu.cn



张世坤

西北工业大学计算机学院 2021 级硕士研究生，导师为杨佳琪副教授，主要研究方向为多视图点云配准。

Email: zhangshikun@mail.nwpu.edu.cn



张艳宁

西北工业大学副校长，教授，国家级人才，国防 973 项目首席。长期致力于图像处理、模式识别、计算机视觉与智能信息处理等的研究，并与航天、航空等方面的国家重大需求相结合。获国家教学成果二等奖 1 项、省部级科技进步奖 3 项，曾获全国三八红旗手称号和总装 863 科技攻关先进个人。先后承担国防 973 项目、国家自然科学基金重点项目、国家/国防 863、总装预研等国家级项目 40 余项。在 IEEE TPAMI、IEEE TIP、PR、IEEE TSMC-B、Info. Fusion、CVPR、ICCV 等国内外本领域权威期刊和重要国际会议上发表论文百余篇。研究成果被多个国家级重大工程攻关项目采用，成功应用于航天、航空、能源、水利等行业的 20 余家单位。获国家/国防授权发明专利 50 余项，陕西省科技进步一等奖一项，国防技术发明一等奖一项，出版专著 3 部。

Email: ynzhang@nwpu.edu.cn

专题综述

基于超像素交互的图像超分辨率

中山大学 张爱平 任文琦 操晓春

本文是中山大学团队解读其在ICCV2023录用的工作SPIN^[1]。论文研究的问题是**基于超像素交互的图像超分辨率 (Super-Resolution, SR)**。现有方法中, 基于Transformer的方法在单幅图像SR任务上取得了高质量结果, 将低分辨率输入图像划分为矩形窗口并行处理, 然后融合以生成高分辨率图像。然而, 这种划分过于粗糙且缺乏可解释性, 影响相似结构间的信息交互从而降低超分性能, 并容易出现伪影。为了应对这些挑战, 我们提出了一种新的超像素信息交互网络 (SPIN), 利用超像素对局部相似像素进行聚类, 形成可解释的连通区域, 并利用超像素内的注意力实现局部信息交互。它是可解释的, 因为只有相似的区域相互补充, 而不同的区域被排除在外。此外, 我们设计了一个超像素交叉注意模块, 利用超像素作为代理来促进全局信息的传播。实验表明, 所提出的SPIN模型在准确性和轻量级方面优于最先进的SR方法。

一、研究背景

单幅图像超分辨率 (Single Image Super-Resolution, SISR) 是底层计算机视觉中的一项关键任务, 旨在提高低分辨率 (LR) 图像的分辨率和视觉质量, 生成高分辨率 (HR) 图像。

自从 SRCNN^[2] 的开创性工作以来, 已经有许多神经网络来应对从低分辨率输入重建高质量图像的挑战。近年来, 注意力机制^[3], 尤其是 Transformer, 已被证明对高层视觉任务和底层视觉领域都有显著效果, 包括超分辨率 (SR) 任务。注意力机制有很强的上下文特征提取能力, 允许网络选择性地关注输入的相关区域,

这可以提高 SR 输出的质量。然而, 注意力机制具有较高的计算复杂度和内存消耗, 因此通常需要将大图像分割成小窗口, 分别在窗口内做注意力操作, 如 SwinIR^[4] 和 ESRT^[5] 和 ELAN^[6]。然而, 基于固定矩形形状划分窗口会导致本身连续的相似结构的分割在不同的窗口, 影响信息直接交互。同时, 每个窗口中应用的局部注意力机制涉及语义不相关区域, 从而导致不理想的模型预测。为了解决这些问题, 我们提出了一种新的方法, 将局部和全局注意力机制与精细的超像素划分相结合。

二、基于超像素交互的超分辨模型

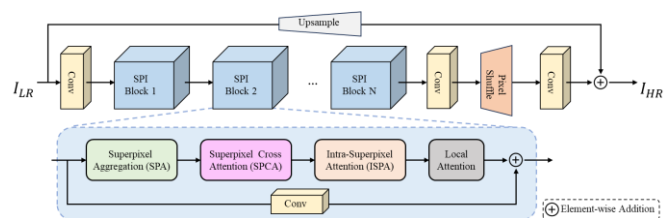


图 1 基于超像素交互的图像超分辨率模型架构

我们提出的模型的架构如图 1 所示。首先, 我们使用编码器, 即 3×3 卷积, 将低分辨率图像 I_{LR} 嵌入到高维特征空间中

$$x_{emb} = f_{encoder}(I_{LR})$$

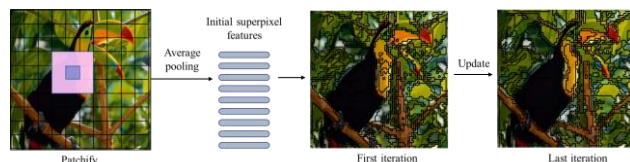


图 2 超像素迭代更新示意图

然后, 通过多个超像素交互 (SPI) 模块进行深层特征提取。该模块由四个组件组成: 超像素聚合 (SPA)、超像素交叉注意力 (SPCA)、超像素内注意力 (ISPA)

和局部注意力。如图 2 所示, SPI 模块利用 SSN^[7] 的基于 soft k-means 的可微超像素算法, 迭代更新超像素特征和像素到超像素的关联映射图 A 。SPA 模块负责聚合来自输入图像中的超像素的信息, SPCA 模块通过超像素的替代来捕捉像素之间的交互, 而 ISPA 模块捕捉每个超像素内像素之间的相互作用, 局部注意力模块用来加强局部内部的信息交互。对于第 i 个 SPI 模块, 我们有

$$x_i = x_{i-1} + f_{local}(f_{ISPA}(f_{SPCA}(x_{i-1}, s_i)))$$

其中, s_i 为超像素特征。最后, 我们利用 3×3 卷积和 pixel-shuffle [34] 来获得全局残差信息, 该信息被添加到 I_{LR} 的上采样图像中, 用于获取高分辨率图像 I_{SR} 。

2.1. 超像素交叉注意力

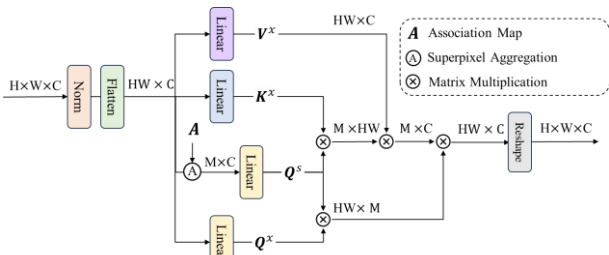


图 3 超像素交叉注意力

如图 3 所示, 超像素交叉注意力 (SPCA) 旨在利用超像素进行长距离的信息传递。由于超像素有着单一的语义信息, 且与像素特征的相似性高, 很适合作为代理 (proxy) 特征来进行信息传递。具体来说, 我们使用注意力机制^[3] 首先将像素信息传播到超像素。我们使用线性映射来计算 query, key 和 value:

$$Q^s = sW_q^s, \quad K^x = xW_k^x, \quad V^x = xW_v^x$$

其中, $W_q^s \in \mathbf{R}^{C \times D}$, $W_k^x \in \mathbf{R}^{C \times D}$, $W_v^x \in \mathbf{R}^{C \times C}$ 分别为 query, key 和 value 的权重矩阵, x 代表像素, s 代表超像素。首先计算 query 和 key 之间的相似性, 并将其用作权重来聚合 key 来获得输出

$$s_u = \text{softmax}(Q^s(K^x)^T / \sqrt{D})V^x$$

s_u 相当于新的超像素特征。这个过程不考虑邻居的限制, 确保了长距离信息的传播。

由于信息从像素传播到了超像素, 有必要将聚合的信息分发回像素, 以实现像素之间的信息传播。在这里,

我们进一步使用注意力机制。类似地, 我们利用另一个权重矩阵 W_q^s 来从像素特征获得 query。为了减少参数的数量, 我们直接使用 Q^s 作为 key, 并使用更新的超像素特征 s_u 作为 value, 并利用交叉注意力将更新的超像素特征映射回像素级别。与 Transformer^[3] 类似, 在注意力机制后使用 FFN, 我们的 FFN 包含一个层归一化^[8] 层, 之后我们使用特征门控^[9] 来调制输入特征, 并使用通道注意力^[10] 来提取全局信息。之后, 使用两个完全连接层和 GELU^[11] 激活函数。

2.2. 超像素内部注意力

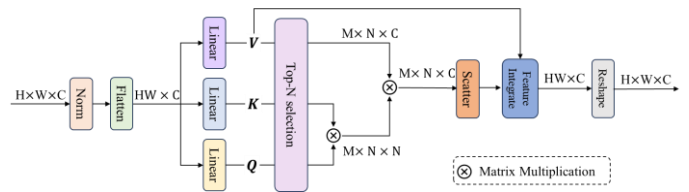


图 4 超像素内部注意力

如图 4 所示, 超像素内部注意力 (ISPA) 旨在利用超像素内部特征的相似性进行信息互补。提高超分辨率图像质量的直观方法是利用同一超像素内相似像素的互补性。为了实现这一目的, 我们需要获得每个超像素的相应像素。然而, 不同的超像素可能包含不同数量的像素, 这使得难以进行并行处理, 也会导致额外的内存消耗, 因为总有一些超像素包含大量像素。为了解决这个问题, 如图 5 所示, 我们求助于关联图 A , 并选择与每个超像素最相似的前 N 个像素。假设一个超像素的隶属像素为

$$f = \{x(i)\}_N \in \mathbf{R}^{N \times C}$$

其中 N 表示所选像素的数量。我们遵循标准的自注意机制^[3], 进行超像素内注意, 其中包括用于 query、key 和 value 投影的权重矩阵 W_q^f , W_k^f , W_v^f 。在超像素内交互之后, 我们利用前 N 个选择过程中生成的索引, 将细化的像素特征分散回图像中它们各自的位置。前 N 个选择可能导致一些“被忽略”的像素, 即这些像素不包括在任何超级像素中。对于那些“被忽略”的像素, 我们利用值投影 W_v^f 来投影它们以获得更新的特征, 然后将其与通过超像素内交互更新的像素进行集成。与 SPCA 模块类似, 我们在 ISPA 模块之后采用了相同的 FFN。

表 1 超分辨率实验结果

Methods	Scale	Params	Set5		Set14		BSDS100		Urban100		Manga109	
			PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
CARN	×2	1592K	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	38.36/0.9765					
IMDN		694K	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774					
LatticeNet+		756K	38.15/0.9610	33.78/0.9193	32.25/0.9004	32.29/0.9291	-					
SwinIR-light		878K	38.14/0.9611	33.86/0.9206	32.31/0.9012	32.76/0.9340	39.12/0.9783					
ESRT		677K	38.03/0.9600	33.75/0.9184	32.25/0.9001	32.58/0.9318	39.12/0.9774					
ELAN-light		582K	38.17/0.9611	33.94/0.9207	32.30/0.9012	32.76/0.9340	39.11/0.9782					
SPIN (Ours)		497K	38.20/0.9615	33.90/0.9215	32.31/0.9015	32.79/0.9340	39.18/0.9784					
CARN	×3	1592K	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	33.43/0.9427					
IMDN		703K	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445					
LatticeNet+		765K	34.53/0.9281	30.39/0.8424	29.15/0.8059	28.33/0.8538	-					
SwinIR-light		886K	34.62/0.9289	30.54/0.8463	29.20/0.8082	28.66/0.8624	33.98/0.9478					
ESRT		770K	34.42/0.9268	30.43/0.8433	29.15/0.8063	28.46/0.8574	33.95/0.9455					
ELAN-light		590K	34.64/0.9288	30.55/0.8463	29.21/0.8081	28.69/0.8624	34.00/0.9478					
SPIN (Ours)		569K	34.65/0.9293	30.57/0.8464	29.23/0.8089	28.71/0.8627	34.24/0.9489					
CARN	×4	1592K	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.42/0.9070					
IMDN		715K	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075					
LatticeNet+		777K	32.30/0.8962	28.68/0.7830	27.62/0.7367	26.25/0.7873	-					
SwinIR-light		897K	32.44/0.8976	28.77/0.7858	27.69/0.7406	26.47/0.7980	30.92/0.9151					
ESRT		751K	32.19/0.8947	28.69/0.7833	27.69/0.7379	26.39/0.7962	30.75/0.9100					
ELAN-light		601K	32.43/0.8975	28.78/0.7858	27.69/0.7406	26.54/0.7982	30.92/0.9150					
SPIN (Ours)		555K	32.48/0.8983	28.80/0.7862	27.70/0.7415	26.55/0.7998	30.98/0.9156					

三、实验结果

3.1. 定量结果

将我们的模型与最先进的轻量级 SR 模型进行了比较，包括基于 CNN 的 CARN^[12]、IMDN^[13]、LatticeNet^[14]等模型，以及基于 Transformer 的 ESRT^[5]、SwinIR^[4]和 ELAN^[6]模型。如表 1 所示，通过利用图像块之间的长距离相似性，基于 Transformer 的模型^[4,5,6]在 PSNR 和 SSIM 方面始终优于基于 CNN 的方法^[12,13,15,16,17,18,19]。然而，他们总是将图像划分为规则的窗口，这可能会破坏输入图像中的对象、边界等。相反，我们的方法利用超像素为 Transformer 进行可解释和连续的区域划分。我们在所有五个基准数据集和所有三个量表上获得了最好或第二好的 PSNR/SSIM 分数。此外，参数的数量小于现有的基于 Transformer 的方法。

3.2. 定性结果

图 5 显示了 Urban100、BSDS100 和 Set14 数据集上比例×4 的可视化比较。结果表明，只要 LR 图像中存在相应的非局部信息，所提出的 SPIN 就可以有效地恢复被严重破坏的纹理。而缺乏非局部注意力的模型无法准确重建严重低质的结构，例如，比较图像“B100/148026”的重建结果时，很明显我们的模型产生的结果非常接近 HR，而其他没有非局部关注的 SISR

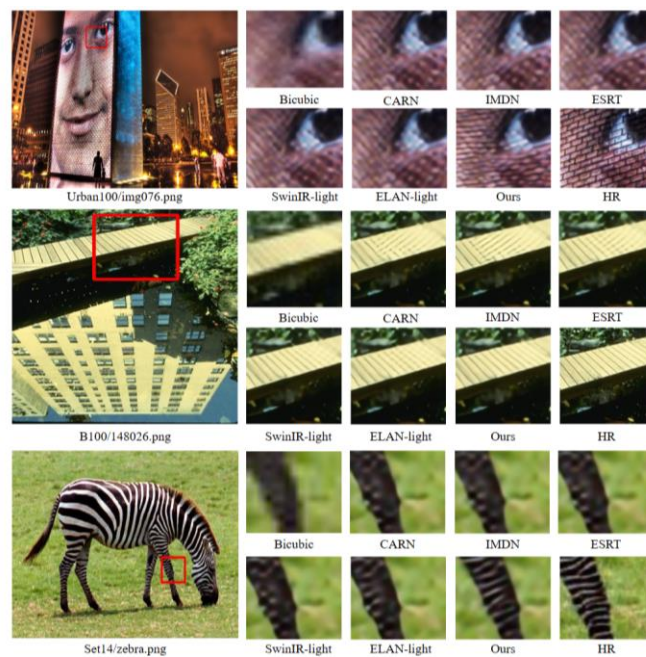


图 5 视觉定性结果

模型，如 CARN^[12]和 IMDN^[13]，不适合恢复这种严重受损的区域此外。与其他基于注意力的深度 SISR 方法（如 ESRT^[5]、SwinIR-light^[4]和 ELAN-light^[6]）相比，我们的 SPIN 模型仍然保持着卓越的重建质量。此外，对于图像“Urban100/img020”，即使没有太多的纹理信息，我们的方法也可以准确地恢复低质图像。

四、总结

我们提出了一种基于超像素交互网络（SPIN）的图像超分辨率新方法，该方法利用超像素将局部相似像素分割为可解释的连通区域。我们的方法采用超像素内注意力来促进不规则局部超像素区域内的局部信息交互，而超像素交叉注意力模块通过超像素代理来促进长距离信息交互。实验表明，SPIN在精度和轻量级方面优于最先进的超分辨率方法。此外，所提出的方法为利用可解释的区域划分处理整个图像的挑战提供了一个新的思路。

责任编辑 崔海楠

参考文献

- [1] Aiping Zhang, Wenqi Ren, Yi Liu, Xiaochun Cao. Lightweight Image Super-Resolution with Superpixel Token Interaction. In ICCV 2023.
- [2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In ECCV 2014.
- [3] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In Advances in Neural Information Processing Systems (NeurIPS), 2017.
- [4] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In ICCV 2021.
- [5] Zhisheng Lu, Juncheng Li, Hong Liu, Chaoyan Huang, Lili Zhang, and Tiejiong Zeng. Transformer for single image super-resolution. In CVPR 2022.
- [6] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In ECCV 2022.
- [7] Varun Jampani, Deqing Sun, Ming-Yu Liu, Ming-Hsuan Yang, and Jan Kautz. Superpixel sampling networks. In ECCV 2018.
- [8] Jimmy Lei Ba, Jamie Ryan Kiros, and Geoffrey E Hinton. Layer normalization. arXiv preprint arXiv:1607.06450, 2016.
- [9] Xu Qin, Zhilin Wang, Yuanchao Bai, Xiaodong Xie, and Huizhu Jia. Ffa-net: Feature fusion attention network for single image dehazing. In AAAI 2020.
- [10] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In CVPR 2018.
- [11] Dan Hendrycks and Kevin Gimpel. Gaussian error linear units. arXiv preprint arXiv:1606.08415, 2016.
- [12] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In ECCV 2018.
- [13] Zheng Hui, Xinbo Gao, Yunchu Yang, and Xiumei Wang. Lightweight image super-resolution with information multi-distillation network. In Proceedings of the ACM International Conference on Multimedia (ACM MM), 2019.
- [14] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In ECCV 2020.
- [15] Chaofeng Wang, Zheng Li, and Jun Shi. Lightweight image super-resolution with adaptive weighted learning network. arXiv preprint arXiv:1904.02358, 2019.
- [16] Rushi Lan, Long Sun, Zhenbing Liu, Huimin Lu, Cheng Pang, and Xiaonan Luo. Madnet: a fast and lightweight network for single-image super resolution. IEEE Transactions on Cybernetics, 51(3):1443–1453, 2020.
- [17] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. In Advances in Neural Information Processing Systems (NeurIPS), 33:20343–20355, 2020.
- [18] Jie Liu, Jie Tang, and Gangshan Wu. Residual feature distillation network for lightweight image super-resolution. In ECCV 2020.
- [19] Xinyan Zhang, Peng Gao, Sunxiangyu Liu, Kongya Zhao, Guitao Li, Liuguo Yin, and Chang Wen Chen. Accurate and efficient image super-resolution via global-local adjusting dense network. IEEE Transactions on Multimedia, 23:1924–1937, 2021.



张爱平

中山大学网络空间安全学院 2022 级博士研究生，主要研究方向为计算机视觉。

Email: zhangaip7@mail2.sysu.edu.cn



任文琦

副教授，硕士生导师，中山大学“百人计划”引进人才。国家自然科学基金优秀青年基金获得者。天津大学与美国加州大学 Merced 分校联合培养博士，从事计算机视觉与多媒体内容安全领域的研究。近年在 CCF-A 类期刊和会议长文发表学术论文 60 余篇，Google 学术引用 9000 余次，入选爱思唯尔中国高被引学者、百度全球高潜力 AI 华人青年榜。担任 AI 和 CV 领域多个顶级学术会议的领域主席和程序委员会委员。获中国计算机学会优博奖、吴文俊人工智能优秀青年奖。

Email: renwq3@mail.sysu.edu.cn



操晓春

中山大学信息学部副主任、网络空间安全学院院长，国家杰出青年/优秀青年基金获得者。主要从事人工智能基础研究和网络空间内容安全应用研究；发表 ACM/IEEE 汇刊 100 余篇，CCF-A 类期刊及会议长文文章 160 余篇；Google 引用 18000 余次，H-index 64；获得省部级一等奖和二等奖各 1 项。现兼任 TPAMI 的 Associate Editor、TIP 的 Senior Area Editor、电子学报的编委，曾兼任 TMM 和 TCSVT 的 Associate Editor，10 余次兼任 NeurIPS/ICCV/CVPR/IJCAI/ACMMM 的 Area/Track Chairs。指导博士生获得中国电子学会优博、CCF 优博、中科院优博论文；指导的研究生有 3 人入选国家级人才计划。。

Email: caoxiaochun@mail.sysu.edu.cn

顶会观察

CVPR 2023

南京理工大学 周涛

国际计算机视觉与模式识别会议 (IEEE Conference on Computer Vision and Pattern Recognition, CVPR) 是世界顶级的计算机视觉会议 (三大顶会之一, 另外两个是 ICCV 和 ECCV)。CVPR 会议每年举办, 旨在推动计算机视觉、模式识别及相关领域的发展, 展示这些领域最新的研究成果。会议涵盖的主题包括图像处理、模式识别、计算几何、机器学习、数据挖掘以及其他相关领域。目前, 根据 Google Scholar, CVPR 的 h5 指数所有会议或出版物中位列第四, 仅次于《自然》、《科学》和《新英格兰医学杂志》。过去 5 年被引用最多的出版物中, CVPR 位列第 4。今年大会组委会仍有不少华人学者, 美国国家工程院院士、斯坦福大学李飞飞教授担任 General Chair, 南京大学吴建鑫教授和 ETH Zurich 助理教授汤思雨担任 Tutorials Chairs 等。本届 CVPR 大会于 2023 年 6 月 18 日至 22 日在加拿大温哥华国际会议中心举办, 包括 2 天的研讨会 (workshops) & 讲习班 (tutorials) 和 3 天的主会 & 博览会 (Expo)。

一、投稿与录用情况

CVPR 的论文投稿量持续增加, 从 CVPR2019 的 5160 篇有效投稿增长到 CVPR2020 年 6656 篇, 到 CVPR2021 的 7500 篇, 再到 CVPR2022 的 8161 篇。而今年又再创新高, 达到了 9155 篇, 收录 2360 篇, 接收率为 25.78%。相比 2022 年, 今年 CVPR 的投稿量提升 12.12% (994 篇); 相比 2021 年, 投稿量提升 30.06% (2116 篇)。今年投稿的论文主要研究主题包括多视图和传感器 (隐式表示, 大量 NeRF), 图像

和视频生成 (主要是扩散模型), 面部/身体/姿势估计, 持续学习, 多模态 (视觉 + 语言) 建模等。投稿量排在前十的研究方向包括: 1) 多视角和传感器下的三维重建; 2) 图像和视频的合成与生成; 3) 面部、身体姿势和手势运动; 4) 迁移学习、元学习、小样本学习、持续学习或长尾学习; 5) 识别: 分类、检测、检索; 6) 视觉、语言和推理; 7) 低层次视觉; 8) 分割、分组和形状分析; 9) 深度学习架构和技术; 10) 多模态学习。此外, 在医学和生物视觉、细胞显微镜技术领域有 420 篇投稿 (53 篇录用), 反映出 AI 医疗相关的研究人员也在关注偏向计算机视觉和人工智能的顶级会议 CVPR。

在录用的论文中, 有 12 篇入围本届最佳论文候选名单 (award candidate), 占提交论文的 0.13%, 录用论文的 0.51%; 以及 235 篇录用为 highlights 论文, 占提交论文的 2.6%, 录用论文的 9.96%。值得注意的是, 在 12 篇最佳候选论文中, 有部分来自于武汉大学、香港中文大学、香港科技大学、上海人工智能实验室、商汤科技、华为等国内大学及研究机构。

二、大会概览

会议形式: 自新冠疫情流行以来, CVPR2023 是第二次进行线下举办, 大约 8300 多人注册参会, 其中线下人数近 7000。可以看出, 今年参会人数也是自疫情以来创历史最高。与以往不同的是, 今年的 CVPR 采用单一轨道模式, 让所有参会者都能参加所有活动。每篇论文都以海报形式展示。此外, 所有最佳候选论文额外进行全体大会口头报告。每个参会者都获得个性化的数字计划, 以便轻松浏览每个海报展示单元中的约 400 个

海报。为了方便交流，虚拟平台为每篇论文提供了海报展示、视频和聊天交流等功能。另外，无法线下的参会人员可以通过在线直播观看大会活动。

评审机制：CVPR2023 的审稿过程由 400 多名领域主席 (area chair) 负责管理，而今年新增了 30 位高级领域主席。高级领域主席在许多方面提供了帮助，主要负责对于处理困难案例的裁决、应急情况的处理、选择重点论文和评选奖项候选人等。今年共有 6625 名审稿人，每篇论文至少收到 3 份评审意见。与往年一样，在收到初步评审后，作者对意见进行回复，然后 AC 负责组织线上讨论并给出初步推荐意见。最后，由领域主席三人小组共同作出每篇论文的接收/拒绝决定。

主题报告：本届 CVPR 2023 的主题演讲主要围绕人工智能领域的下一步发展方向展开讨论。麻省理工学院 Rodney Brooks 教授带来了“Recycling old vision ideas in a modern computational world”的报告。许多计算机视觉领域的想法一次又一次地被重新审视，包括基于神经计算的现代计算机视觉。特别是当前计算硬件取得了令人难以置信的发展，报告中主要探讨这样强大的计算能力是否能够被忽视的旧想法注入新的动力。华盛顿大学 Yejin Choi 教授带来了“An AI Odyssey: the Dark Matter of Intelligence”主旨报告。在追求更强大的智能系统的过程中，我们发现了一种被称为“智能的暗物质”的现象，这是指 AI 中存在的巨大而复杂的知识和能力，但我们仍然对其运作方式和内部机制知之甚少。报告主要探讨这个令人困惑且引人瞩目的领域，并探索解决这个问题的途径。此外，Meta AI 的研究科学家 Larry Zitnick 带来的“Modeling Atoms to Address Our Climate Crisis”报告。气候危机是当今世界面临的巨大挑战之一，而模拟和理解原子水平的过程对于解决这一问题至关重要。该报告主要探讨使用原子模型并结合人工智能技术来应对当前面临的气候危机的重要性。

三、最佳论文

2023 年度最佳论文奖评审委员会由计算机视觉相关领域的 7 名国际权威学者组成。本届 CVPR 大会共评选出 5 篇论文获奖，其中 2 篇最佳论文，1 篇最佳学生

论文，以及最佳学生论文提名和最佳论文提名各 1 篇。

最佳论文：Planning-oriented Autonomous Driving^[1]，来自上海人工智能实验室、武汉大学、商汤科技。论文作者首次将感知、预测和规划等 3 大类主任务、6 小类子任务（目标检测、目标跟踪、场景建图、轨迹预测、栅格预测和路径规划）整合到统一的基于 Transformer 的端到端网络框架下，实现了全栈关键任务驾驶通用模型。该论文也是首次提出感知决策一体化的自动驾驶通用大模型 (UniAD)，开创了以全局任务为目标的自动驾驶大模型架构先河，为自动驾驶技术与产业的发展提出了新的方向。值得注意的是，这是近 10 年来计算机视觉三大顶级会议中 (CVPR、ICCV、ECCV)，第一篇以中国学术机构作为第一单位的最佳论文。

最佳论文：Visual Programming: Compositional visual reasoning without training^[2]，来自艾伦人工智能研究所 (Allen Institute for AI, AI2)。通用人工智能系统旨在为用户提供简单的自然语言界面与模型进行交互。构建这些系统的主要方法是采用大规模无监督预训练和监督多任务训练。然而，这种方法需要为每个任务准备精心策划的数据集，难以扩展到复杂任务的无限长尾部分。在本文研究中，研究人员探索使用大型语言模型来处理复杂任务的长尾部分，通过将这些任务用自然语言描述并分解为较简单的步骤，可以由专门的端到端训练模型或其它程序来处理。基于此，论文提出了 VISPROG，一种根据自然语言指令解决复杂和组合视觉任务的方法。只需要给出几个自然语言指令的示例和所需的高级程序，VISPROG 使用 GPT-3 中的上下文学习为任何新指令生成一个程序，然后在输入图像上执行该程序以获得预测结果。

最佳学生论文：3D Registration with Maximal Cliques^[3]：来自西北工业大学。三维点云配准是计算机视觉的关键研究问题之一，其目的是搜索最佳位姿以对齐一对点云。本文作者提出了一种基于最大种群的 3D 配准方法，核心是放宽先前的最大种群约束，在图中挖掘更多局部一致性信息来生成姿态假设。首先构建了一个兼容性图来表示初步对应关系之间的亲和矩阵，然后

搜索最大种群，每个种群代表一个一致集。然后执行基于节点的种群选择，每个节点对应图权重最大的最大种群。对于选择的种群，通过奇异值分解算法（SVD）计算变换假设，并选择最佳假设进行配准。在多个数据集（U3M、3DMatch、3DLoMatch 和 KITTI）上进行了大量实验，结果表明该方法能够有效提高配准性能，且优于各种最先进的方法。

最佳论文提名：DynIBaR: Neural Dynamic Image-Based Rendering^[4]，来自谷歌和康奈尔大学。

从动态场景的单目视频中合成新视图是一个更具挑战性的动态场景重建问题。基于时间变化的神经辐射场（也称为动态 NeRF）的最先进方法在这个任务上展示了令人印象深刻的结果。然而，对于具有复杂物体运动和不受控制的相机轨迹的长视频，这些方法可能会产生模糊或不准确的渲染图像，限制了它们在实际应用中的使用。本文提出了一种新方法（DynIBaR），通过采用基于体素的图像渲染框架来处理这些局限性，以一种感知场景运动的方式聚合来自附近视角的特征来合成新视角。该方法在能够建模复杂场景和视角依赖效果方面保留了之前方法的优势，同时还能够从具有复杂场景动态和不受约束的相机轨迹的长视频中合成逼真的新视角。在动态场景数据集上的结果展示了该方法的先进性，并将方法应用于具有挑战性的相机和物体运动的野外视频中，在这些情况下，之前的方法无法生成高质量的渲染图像。

最佳学生论文提名：Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation^[5]，来自谷歌和波士顿大学。大型文本到图像生成模型在人工智能的发展中取得了显著的突破，能够根据给定的文本提示（Prompt）生成高质量且多样化的图像。然而，这些模型缺乏模仿或再现给定参考集中主体外观的能力，特别是对个人定制的物体，很难实现个人物体在风格、造型等方面的实现能力。在本研究中，研究人员提出了一种用于“个性化”的文本到图像的扩散模型。仅提供主体的几张图像作为输入，对预训练的文本到图像模型进行微调，使其学习将唯一标识符与特定主体绑定。一旦主体嵌入到模型的输出领域

中，该唯一标识符可以用于合成在不同场景中具有上下文的主体的新的逼真图像。通过利用模型中嵌入的语义先验和新的自生类别特定先验保持损失，该模型可以合成出在参考图像中不存在的多样化场景、姿势、视角和光照条件下的图像。最后，作者将该模型应用到多个挑战的任务上，包括主题重构、文本引导视图合成和艺术渲染。此外，本文作者为这个主题驱动生成的新任务提供了一个新的数据集和评估协议。

四、大会奖项

Longuet-Higgins Prize: Longuet-Higgins 奖由 IEEE 模式分析和机器智能(PAMI)技术委员会在每年的 CVPR 上颁发，以表彰在计算机视觉领域所作出的基础贡献。该奖项旨在表彰十年前对计算机视觉研究产生重大影响的 CVPR 论文，它是以理论化学家和认知科学家 H. Christopher Longuet-Higgins 的名字来命名的。本届 Longuet-Higgins 奖颁发给了 2013 年 CVPR 论文“Online Object Tracking: A Benchmark”。

Young Researcher Awards: 青年研究员奖是颁发给在获得博士学位后 7 年内，对计算机视觉领域做出杰出研究贡献的一位或两位青年研究人员。本届的青年研究员奖颁发给了 Facebook 人工智能研究院研究科学家 Christoph Feichtenhofer 和佐治亚理工学院助理教授 Judy Hoffman。

Thomas S. Huang Memorial Prize: 这个奖项从 2020 年开始设立，是为了纪念已故的黄煦涛教授。黄教授是在 CV 和图像处理领域做出大量贡献的先驱学者。本届该奖项颁给了加利福尼亚大学伯克利分校教授 Alyosha Efros，他的研究兴趣包括计算机视觉、计算机图形、计算摄影、机器学习及人工智能等。

五、Tutorials和Workshops

Tutorials: 本届 CVPR 包括 33 个 tutorials，例如“Large-Scale Visual Localization”，“Vision Transformer: More is different”，“Object localization for free: Going beyond self-supervised learning”，“Denoising Diffusion Models: A Generative Learning Big Bang”，“Recent advances

in anomaly detection”，“Recent Advances in Vision Foundation Models”，“ML Systems for Large Models and Federated Learning”，“All Things ViTs: Understanding and Interpreting Attention in Vision”等。通过这些也不难发现，当前的研究热点仍然围绕视觉定位、Transformer、自监督学习、扩散模型、视觉基础模型、联邦学习、注意力机制等。因此，这些 tutorials 资料，系统地为我们梳理了相关领域的发展历史脉络、知识结构、研究进展以及未来的发展方向。

Workshops: 本届 CVPR 共包含 100 个 workshop，其内容涵盖非常丰富，既有关注计算机视觉基础任务，也有学习范式，还有特定应用。粗略选取收录论文数量相对较多的一些 workshop，例如 1) 8th New Trends in Image Restoration and Enhancement Workshop and Challenges (77 篇)，主要聚焦图像与视频恢复与增强研究；2) 7th AI City Challenge Workshop (35 篇)，主要聚焦人工智能城市中的实体零售业务和智能交通系统；3) 4th International Workshop on Event-Based Vision (33 篇)，聚焦于智能相机和处理这些传感器数据的算法研究；4) 6th Efficient Deep Learning for Computer Vision (29 篇)，聚焦于高效深度学习算法、以及低资源学习等方面的研究；5) 2nd Workshop on Learning With Limited Labelled Data for Image and Video Understanding (27 篇)，聚焦于面向标记数据受限情况下的图像与视频理解算法研究；6) Generative Models for Computer Vision (24 篇)，聚焦生成模型在计算机视觉领域的应用和发展；7) 19th CVPR Workshop on Perception Beyond the Visible Spectrum (22 篇)，聚焦复杂场景下的目标检测、识别和跟踪等算法研究；8) Workshop on Autonomous Driving (16 篇)，聚焦自动驾驶的运动预测研究；9) EarthVision Large Scale Computer Vision for Remote Sensing Imagery (16 篇)，聚焦计算机视觉和机器学习技术在遥感数据上的研究；10) 4th Workshop on Continual Learning in Computer Vision (14 篇)，聚焦持续学习在计算机视觉中的应用。可以看出，部分 workshop

已经连续举办多届，而有些是第一届举办。此外，也不难发现这些 workshop 更关注计算机视觉中的许多应用场景，例如图像恢复与增强、智慧城市、目标检测识别以及自动驾驶中轨迹预测等问题。

此外，在 CVPR2023 研讨会上举办的各项挑战赛中，国内学术界和工业界在不同任务上取得了备受关注的业绩。西安电子科技大学在在多个赛道取得冠军（CVPR 2023 FMDC 挑战赛中零样本/少样本图像分类、VizWiz 视觉问答、跨模态视频检索等）。合肥工业大学 CVPR 2023 国际情感行为分析挑战赛冠军。南京理工大学在细粒度视觉分类研讨会挑战赛的多个赛道多个赛道取得冠军。联想研究院 PC 创新与生态系统团队在 CVPR 2023 WAD BDD100K MOTs (多目标追踪与分割) 挑战赛中获得冠军，在 CVPR 2023 WAD ARGOVERSE 挑战赛中，斩获端到端运动预测赛道中的 3D 运动预测，3D 多目标跟踪，3D 目标检测三项子赛道冠军等。旷视研究院在 CVPR 2023 自动驾驶国际挑战赛中 OpenLane Topology 赛道夺得冠军。中国电信天翼云 AI 团队在 CVPR 2023 Workshop on Foundation Model 中多任务大模型赛道获得冠军。小视科技在第一届 CVPR 2023 WorkShop 大模型挑战赛中，取得了 A 榜第一名。腾讯优图荣获 CVPR 2023 视觉异常检测挑战赛冠军。

六、总结展望

本届 CVPR 的研究热点和关键词包括 3D 视觉、图像和视频的合成与生成（扩散模型）、识别、迁移学习/元学习/小样本学习、视觉/语言推理、深度学习、Transformer 等。值得关注的是，神经-符号方法重新回到人们视野，它结合了神经网络和符号推理，在扩展人工智能能力方面展示了潜力。此外，数据驱动方法和迁移学习对于提高模型性能和泛化能力起着重要作用。更值得关注的是，将多个任务整合到一个框架中可以提高性能和协调能力。这种趋势强调开发统一的框架来优化多目标，实现复杂场景下不同任务之间更好的协调和性能。此外，大模型技术引起了广泛关注，数据集和基准对于推动计算机视觉研究和评估算法性能仍然至关重要。由于大模型在自然语言处理和计算机视觉等领域取

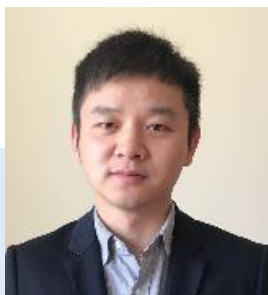
得成功，研究人员开始探索将其引入计算机视觉任务中，这也是一个令人期待的趋势。此外，通用人工智能 (AGI) 旨在模仿人类的认知能力，包括感知、理解、学习和推理，跨越更广泛的领域解决复杂问题并独立做出决策。

因此，我们期待大模型与 AGI 将推动计算机视觉技术的发展，并为解决复杂视觉任务和实际应用提供新的思路和方法。

责编委 魏秀参

参考文献

- [1] Hu, Yihan, Jiazhi Yang, Li Chen, Keyu Li, Chonghao Sima, Xizhou Zhu, Siqi Chai et al. Planning-oriented autonomous driving. CVPR 2023.
- [2] Gupta, Tanmay, and Aniruddha Kembhavi. Visual programming: Compositional visual reasoning without training. CVPR 2023.
- [3] Zhang, Xiyu, Jiaqi Yang, Shikun Zhang, and Yann. ing Zhang. 3D Registration with Maximal Cliques. CVPR 2023.
- [4] Li, Zhengqi, Qianqian Wang, Forrester Cole, Richard Tucker, and Noah Snavely. Dynibar: Neural dynamic image-based rendering. CVPR 2023.
- [5] Ruiz, Nataniel, Yuanzhen Li, Varun Jampani, Yael Pritch, Michael Rubinstein, and Kfir Aberman. Dreambooth: Fine tuning text-to-image diffusion models for subject-driven generation. CVPR 2023.



周涛

南京理工大学计算机科学与工程学院教授、博导
主要研究方向为医学图像分析、计算机视觉、机器学习等。
Email:taozhou@njust.edu.cn

中国农业大学李振波教授访谈

2023年8月26日,《CCF-CV专委简报》在线采访了中国农业大学博士生导师李振波教授。下面是采访实录。

问题 1: 李老师,您好!首先,请您分享一下您的个人学习和研究经历。

我2000年和2003年分别于山东大学计算机学院获得学士和硕士学位,硕士期间主要从事的研究方向偏重于计算机图形学、虚拟现实与人机交互方面。为了能够在计算机研究领域有一些更大的突破,我于2007年在中国科学院计算技术研究所获得了博士学位,在攻读博士期间,我的研究领域转向了偏计算机视觉方向。之后我赴法国国立巴黎电信学院从事博士后研究,继续开展计算机视觉与虚拟现实方面的研究工作。2010年我被以优秀人才引进到中国农业大学工作,之所以做出这个选择,是因为我感觉计算机视觉在农业场景中有很多挑战需要去解决,因此,我选择来到中国农业大学任教。

问题 2: 您主攻计算机视觉在农业方面的研究和应用,请问这方面的研究和应用与计算机视觉在其它领域的研究与应用,最大的区别和难点分别在哪里呢?

我来到中国农业大学以后,一方面继续从事计算机视觉基础理论研究,另一方面开展计算机视觉在农业领域的应用研究,为保障国家粮食安全做贡献。在农业领域,主要有种植业和养殖业,种植业中的大田种植、设施园艺及养殖业中的水产养殖和畜禽养殖都是比较典型的农业应用场景。对于大田种植,它的环境是非常开

放的,例如雨天或者阴天等,很难去管理和控制。但是在养殖业方面,以设施园艺、畜禽养殖和工厂化水产养殖为代表的设施养殖环境是相对可控的。以我们目前的研究来看,在相对可控环境下,计算机视觉在农业方面的应用研究还是相对容易的。在开放和半开放环境下,我们做了一些计算机视觉方法的研究,例如典型作物表型信息的提取、作物病害的自动诊断识别、作物生长状态监测、作物叶片的实例分割等;对特征提取和表达方法、低质量图像和视频的复原增强等基础任务进行了研究,相关工作发表在ICCV、ECCV、TIP等视觉领域知名国际会议和期刊上。

我在农业农村部国家数字渔业创新中心、农业农村部农业信息获取技术重点实验室和农业农村部智慧养殖技术重点实验室等省部级研究机构的支持下,也开展了智慧养殖方面的应用研究,以数字渔业为例,我们通过传感器来获取养殖环境数据,通过运动跟踪方法跟踪分析不同鱼的行为数据,为智慧养殖提供依据。

我觉得最大的区别和难点,一是由于农业外部环境的开放性,导致光照、背景等很多因素都是不可控的,对于高质量数据的采集有很大影响,因此在农业领域做研究,大规模的开放数据集还是很少的。另一个难点是农业领域很难有一个统一的模型或方法,针对不同的农业对象有不同的研究方法,例如我们也做过针对鸡蛋蛋形的专门检测,奶牛体尺的自动测量等研究,旨在基于实际应用需求和不同对象的外在形态开展一系列有意思的研究。

问题 3: 您主要聚焦于智慧农业、数字渔业方面的研究, 能否谈谈我们国家目前在智慧农业和数字渔业方面的一些研究进展? 与其他一些国家相比, 我们在这些领域还存在哪些不足? 目前我们国家在这两个领域都是如何布局的, 后续主要计划发展哪些方面?

我们国家农业发展可分为四个阶段, 第一阶段 1.0 主要是传统农业, 主要依靠人力和畜力来劳作; 第二阶段 2.0 是机械化农业, 通过使用农业机械进行作业; 第三阶段 3.0 是智慧化生产阶段, 通过物联网等技术进行数据获取, 通过计算机软件进行生产管理; 第四阶段 4.0 是智能化生产阶段, 各类农业装备能够更加智能地自主作业。我们目前的农业是从 2.0 向 3.0 逐渐过渡的阶段。

我国对智慧农业的发展非常重视, 近年来每年中央一号文件, 都会部署智慧农业、互联网+农业、数字农业、农业农村大数据中心、数字乡村等工作, 旨在通过信息技术加快实现农业转型升级, 实现乡村振兴的目标。国家部委和许多省市也开展了许多智慧农业相关的研究和示范, 如在温室为代表的农业物联网的应用示范中, 通过安装环境传感器来实时监测温室的环境参数、监测作物长势、通过调控模型控制温室温度、湿度、二氧化碳浓度等, 可以较好地节约人工, 降低劳动强度、提高温室的生产效率。

在渔业方面, 例如在池塘养殖中, 我们通过无人机空地一体化装备, 实时采集养殖溶解氧等水质数据, 通过视觉相机观测鱼群行为, 实现自动生物量计数, 为养殖水质管理和变量投喂决策提供依据。在工厂化水产养殖、海洋牧场养殖、鱼菜共生养殖等模式方面我们也采用人工智能、数字孪生等技术开展了一些探索。总体而言, 这些探索还处于一个发展中的阶段。

与国外相比, 我们在智慧农业领域的研究还是有一些短板, 例如之前提到的传感器技术, 我们适用于农业领域的传感器测量的精度、稳定性和成本等问题都还有提升空间, 农业模型的研究也有一些差距, 需要我们不断努力。

未来农业将会是无人或者少人农业, 我国农业人口面临老龄化问题, 未来谁来种地, 谁来养猪、谁来养鱼, 谁来保障国家粮食安全是摆在我们面前的课题。无人农业就是希望少数人或者没有人来直接从事农业劳作, 更多使用智能化的农业装备进行农业生产。比如在一块地里面, 完全靠装备来进行耕地播种、施肥施药, 包括最后的收获, 而人们只需要在外面去操纵这些机械就可实现农业的整个生产过程。我们团队也专门编写过关于无人农场的书籍。现在比较有意思的现象是以智能手机为代表的移动终端正逐渐成为像拖拉机、联合收割机一样的新型农具, 教会农民更好地使用智能手机来获取农业信息、操控农业装备、通过手机来做农产品电商, 会吸引更多年轻人从事农业, 这样就使农业成为一种全新的形态, 农民就不再是传统意义上的农民了, 也是一种有身份地位的职业了。

问题 4: 能否分享几个您认为在智慧农业、数字渔业方面做得非常成功的研究或应用? 您认为这些研究和应用具体对我们国家的农业、渔民有哪些帮助呢?

我们团队从 2009 年开始做一些农业物联网相关的研究, 开始做河蟹养植物联网系统, 通过水质传感器实时获取养殖水质数据, 通过预测模型预测水质变化, 通过视觉技术分析河蟹行为, 通过养殖模型实现变量投喂和增氧机等设备的科学管控, 改变了传统养殖主要依靠渔民经验的问题, 为养殖系统提供一个相对可控的外部环境。对于自然灾害和疫病的诊断和预防也是很重要的一个难点, 传统农业农民常说的一句俗语是“看天吃饭”, 因为很多时候, 对于防不胜防的自然灾害我们办法不多。通过尽量调整环境参数, 提供一个相对稳定的外部环境, 以降低养殖的损失。而病害对于养殖危害很大, 养殖疫病的自动诊断和主动防控也是我们研究的课题, 在智慧养殖方面我们团队的工作也获得了国家和省部级的一些奖励, 也是对我们工作的肯定。

问题 5: 您曾主持科技部国家重点研发计划课题和子课题, 能否分享您申请和承担这类课题时的一些经验?

我申请和承担的课题大部分都是和农业领域应用非常紧密的，需求非常明确，比如在国家重点研发计划蓝色粮仓课题中，为实现学科交叉研究，我们与养殖等不同领域的行业专家进行充分的交流，因为如果你不了解行业真正的需求，就很难实现以视觉为代表的人工智能技术和农业产业的深度融合。这里就涉及到了学科交叉研究的问题，交叉学科已经成为国家第 14 个学科门类，国家自然科学基金委设有交叉学部，在信息学部里也有专门的人工智能和不同领域学科交叉的方向，交叉学科研究是一个非常重要的研究发展趋势，和不同行业、不同领域的技术人员或研究人员进行交流，了解真正的需求，然后再有针对性地去研究解决实际问题，实现视觉技术的应用落地。我们农业领域有一名言“把论文写在祖国大地上面”，可以写在农田，写在温室，写在鱼塘里，所以我们需要了解产业发展的需求和痛点，了解行业里面存在的问题，再将这些问题进行凝练，提炼成科学问题或关键技术，然后通过进一步的研究，再去攻克这些课题，在产业应用中进行验证。

问题 6：您在教学方面曾获得 2022 宝钢优秀教师奖、北京市教学成果二等奖和中国农业大学教学成果特等奖等荣誉，培养多名学生获校优秀研究生学位论文和北京市优秀毕业设计，能否介绍一下您的教学理念，以及您在培养学生时的一些优秀做法？

我投入教学的时间还是很多的，我在学校也承担了多门课程的教学任务。在教学方面，我一直秉持着一个很重要的理念，就是教学模式的改革，我将它称之为“课堂革命”。就是让学生成为运动员，我们老师作为教练的角色，比如，我们把学生当成刘翔，那我们就要把自己变成教练孙海平。如何去衡量一门课学生学得怎么样呢？就是要让学生学完这门课以后，既有满头大汗的疲惫感，又有兴高采烈的兴奋感。所以在课程设计过程中，我就会多设置一些随堂测验，阶段测验还有翻转课堂，让学生来当课堂的主人，老师进行课程点评和总结，这样的话可以调动学生学习的积极性，不至于让学生觉得课堂很枯燥。对于一些需要实验操作的课程，我在教学

过程中也会很重视学生的动手操作能力，让学生自己去多去实验。教学不仅仅是让学生学习课本上面的理论知识，更多的是要教会学生如何将学习到的知识加以应用，从而将知识记得更深刻。这种教学模式的改革也是一个不断探索的过程，所以我们老师之间也需要不断地交流学习，这也是一个不断提高自身能力的过程。

对于研究生的培养，我是实行分类培养，主要分为两种类别，一种是学术型研究生，培养方案是科教融合，我们的要求就是以研究为主，做高水平 and 顶天的研究，他们的课程是偏理论层面，对于他们的未来发展方向是成为一个具体领域的拔尖创新领军人才。对于专业型研究生，培养方案是产教融合，在培养过程中，会设置初级实践环节，让学生先去我们的一些专业实践基地，先去接触产业，了解产业的实际问题 and 需求，然后让学生带着问题再去学习，去解决产业中落地的问题。

对于我的研究生，我们坚持例会制度，每周进行组会交流。组会上，会有主讲同学来介绍最近所阅读的一些前沿文献和自己的一些研究工作，然后课题组的其他同学也可以提出自己的见解或者自己研究过程中遇到的困难，大家一起讨论解决。这是我们的组会模式，我们还有一对一研究讨论，可能学生在实验方面、撰写论文方面有什么问题，我们交流解决；或者论文投稿之后的一些修改意见，我们也会一起来进行讨论。

问题 7：您曾出版专著 1 部，参编《数智驱动乡村振兴》《农业物联网导论》等多部著作，可否分享一下您在出版专著方面的一些经验？

专著就是一个成果的整理和积累，就是把你在一个领域里面多年的积累总结归纳成册，从而成为一种信息传播渠道，为他人提供了解和学习的便利。比如说我参编的《农业物联网导论》，现在就被很多学校尤其是农业类学校选为课程教材用来学习农业物联网的知识。对于《数智驱动乡村振兴》这部著作，当时是为了调研乡村振兴中数字乡村的情况，经过长期积累从而将我们的调研工作整理出版的。出版专著方面的一些经验，我觉

着很多时候都是需要积累，特别是一些数据的积累以及平时的调研工作。很多时候调研工作是需要花费大量时间的，不仅仅要调研我们国家不同省份、不同地区的情况，还需要调研该领域在国际上的一些发展，调研需要全面。

问题 8：在科研之余，您还担任中国农业大学信息与电气工程学院副院长，社会兼职工作也很多，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人？

学会平衡和分配时间是一件很重要的事情，在今年的新生开学典礼上，我就说到了这个问题，让同学们要学会平衡学习、科研和生活。文武之道，一张一弛，虽然科研是研究生生活很重要的一部分，但是也不能没有生活，所以我平时也会多倡导学生走出实验室，积极组织实验室运动等。我个人对于时间是有一定的分配的，我大概会把三分之一的时间用在学院的行政管理上面，

三分之一的时间用在教学工作上面，剩下三分之一的的时间用于科研和社会服务工作。

虽然每天都有做不完的事情，但是只要有时间我很乐意去健步走，多运动运动，锻炼身体。

如果不谈工作，私底下的我可能相对安静一些，喜欢阅读一些书籍。在家庭中，因为我和我爱人都是大学老师，所以我爱人也比较理解并支持我的工作。

问题 9：如果吐露研究工作者的的心声，您最想说的是什么？

我觉得我们所有的研究工作者都有一个共同的心声，就是希望能够为我们国家的科技发展做一些贡献。同时，我们作为老师，能够把人才培养好，把学生培养好，是我们最大的荣誉。学生能够取得好的成绩，获得更好的发展，我作为老师，比学生还要高兴，这是身为老师的一种幸福感。

责任编辑 余焯 赵振兵



李振波

教授，博士生导师，中国农业大学信息与电气工程学院副院长。2000年获山东大学计算机学院学士学位、2003年获山东大学计算机学院硕士学位，2007年获中国科学院计算技术研究所博士学位，后赴法国国立巴黎电信学院从事博士后，2010年以优秀人才引进进入中国农业大学工作。中国计算机学会（CCF）杰出会员、CCF数字农业分会常务委员、CCF计算机视觉专委会执行委员、中国农业工程学会高级会员、中国图学学会高级会员、中国人工智能学会会员、IEEE会员、国际信息处理联合会IFIP农业信息处理分会副主席，全国高等院校计算机基础教育研究会理事，全国高等院校计算机基础教育研究会农林专委会主任，教育部计算机类专业教指委物联网工程专业教学研究专家组成员。研究领域聚焦在计算机视觉和图像处理、智慧农业等。农业农村部国家数字渔业创新中心委员，农业农村部农业信息获取技术重点实验室委员，农业农村部智慧养殖技术重点实验室委员，近年来围绕智慧农业中以计算机视觉技术为代表的人工智能关键技术基础理论、方法和系统开展研究和技术应用研究，主持国家自然科学基金、国家重点研发计划、国际科技合作专项等科研课题和任务12项。在国际人工智能顶级期刊TIP和ICCV、ECCV等人工智能领域顶级学术会议、智慧农业领域顶级期刊COMPAG上发表高水平论文90多篇，获授权国家发明专利23项，公开国家发明专利25项，获得软件著作权62项，出版专著1部，参编《数智驱动乡村振兴》《农业物联网导论》等多部著作，获省部级一等奖2项，二等奖3项，2022宝钢优秀教师奖，参与制定国家标准4项，行业标准4项，地方标准1项。获北京市教学成果二等奖、中国农业大学教学成果特等奖等，培养多名学生获校优秀研究生学位论文和北京市优秀毕业设计。

委员好消息

✪ 2023年6月30日，北京高校庆祝中国共产党成立102周年表彰大会在中国人民大学举行，CCF-CV专委会委员、北京航空航天大学**王蕴红**和北京科技大学**殷绪成**获北京高校优秀共产党员荣誉称号，此次共有100名“北京高校优秀共产党员”、100名“北京高校优秀党务工作者”和100个“北京高校先进基层党组织”受到表彰、获得相应称号。

✪ 2023年6月21日，CVPR 2023最佳论文奖公布，CCF-CV专委会执行委员、上海AI实验室**李弘扬**等的论文 Planning-oriented Autonomous Driving 获 CVPR 2023 最佳论文，该文提出一个端到端的感知决策一体框架，融合了多任务联合学习的新范式，使其可进行更有效的信息交换，协调感知预测决策，以进一步提升路径规划能力。CCF-CV专委会执行委员、西北工业大学**张艳宁**指导的论文 3D Registration with Maximal Cliques 获最佳学生论文，该文利用来自噪声点对应的兼容性图中的最大团约束，提出了一种解决点云配准基本问题的方法。

✪ 2023年7月17日，第5届科学探索奖获奖名单揭晓，CCF-CV专委会2位执行委员、东南大学**耿新**和上海交通大学**卢策吾**上榜，共48位青年科学家入选。

✪ 2023年7月28日，2022年度北京市科学技术奖评审结果公示，CCF-CV专委会4位执行委员完成的3个项目获奖，其中，中国科学院自动化研究所**张兆翔**和**谭铁牛**等完成的“面向动态复杂环境的多源融合智能感知技术及产业化”获科技进步一等奖，北京工业大学**胡永利**等完成的“超大城市交通视觉大数据高效表达与运行决策关键技术与应用”、北京交通大学**白慧慧**等完成的“空天地协同的视觉智能关键技术及应用”获科技进步二等奖。

✪ 2023年8月22日，由教育部高等教育司指导、中国高等教育学会主办的第三届全国高校教师教学创新大赛在浙江大学圆满结束，CCF-CV专委会副主任、南京信息工程大学**刘青山**所在的参赛团队获二等奖，CCF-CV专委会执行委员、哈尔滨工程大学**刘海波**带领的团队获三等奖，此前，两支队伍分别获江苏省和黑龙江省高校教师教学创新大赛特等奖。

✪ 2023年8月29日，2023年度中国人工智能学会会士增选名单公示，CCF-CV专委会执行委员、北京大学**彭宇新**入选，本年度共8人入选。

责任编辑 刘海波

点云域自适应领域开源代码

北京航空航天大学 李俊俏 王田

近年来，基于深度神经网络的点云表示学习在 3D 视觉领域备受关注。然而，由于数据采集过程、传感器视角、现实噪声等存在巨大的域差异，在源域上训练的模型无法适应于目标域，为此点云域自适应技术被提出。本文关注以室内点云 PointDA 数据集为主要实验数据集的论文，介绍相关点云域自适应（Domain Adaptation, DA）领域的研究成果。

1、DefRec_PCM

工作：该文首次将自监督学习（Self-supervised Learning）应用于点云 DA 领域，受到模拟到现实转换中遇到的形变的启发，设计变形重建自监督任务来帮助网络捕捉点云数据中的常见变形，学习点云的潜在几何结构特征。此外，该文受到 MixUp 方法的启发，设计针对有标签点云数据的新的训练方法 PCM（Point Cloud Mixup）。实验表明在点云分类和分割任务中，该方法相对存在的方法性能有很大提升。

DefRec_PCM 结构如图 1 所示，其是一个基于多任务架构的多头网络。有标签的源域数据和没有标签的目标域数据被送入一个共享的特征编码器中学习；源域被编码的特征送入分类头使用 Mixup 方法去做预测；源域和目标域被编码的特征共同被送入自监督任务头，使用 Chamfer 距离设计 SSL 损失来训练。具体而言，在自监督任务头中，它通过移动错位一些点来变形一个区域，之后网络预测这些点的原始位置来重建形状变化的区域。如图 2 所示，文中提出三种类型的形变：基于

体积、基于特征和基于采样，并对三种类型的形变进行实验和性能分析，最终采用基于体积的形变。在实践中，作者发现只将 DefRec 应用在目标域样本上会得到更好的结果，因此该文最终选择采用这种方式进行训练。

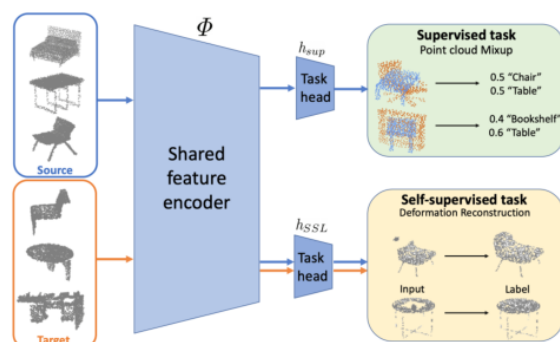


图 1 DefRec_PCM 结构图

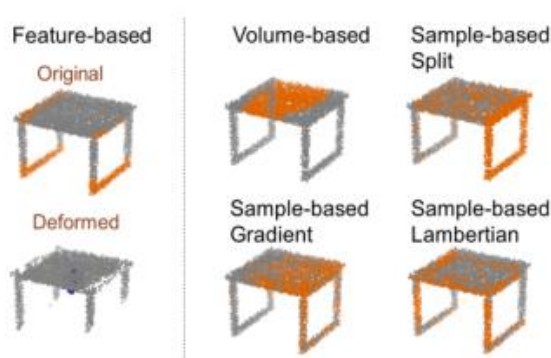


图 2 不同形变类型示意图

论文地址: <https://arxiv.org/pdf/2003.12641.pdf>

代码地址: <https://github.com/ldanachituve/DefRec> and PCM

2、GAST

工作: 本文提出几何感知自训练(Geometry-Aware Self-Training, GAST)的方法, 用于点云自适应领域分类任务。该方法通过结合两个新颖的自监督几何学习任务, 学习语义类别的域共享表示。一方面, 构建旋转角预测任务, 通过点云样本和其自身生成的旋转标签的线性混合来增强表示学习, 以捕获局部几何的全局拓扑特征。一方面, 构造失真位置曲率预测任务, 使用曲率感知失真定位来归一化不同数据集之间的分布。PointDA 数据集的实验表明, GAST 方法优于当时最先进的方法, 特别是在合成到真实任务场景中表现最佳。

为了更好地以自监督学习的方式训练无标签目标域数据, 该文以自步调 (Self-paced) 学习的方式采用自训练方案来逐步优化选择置信度高的目标样本, 将其和伪标签一起送入分类模型中学习, 以细化具有目标判别能力的语义特征。

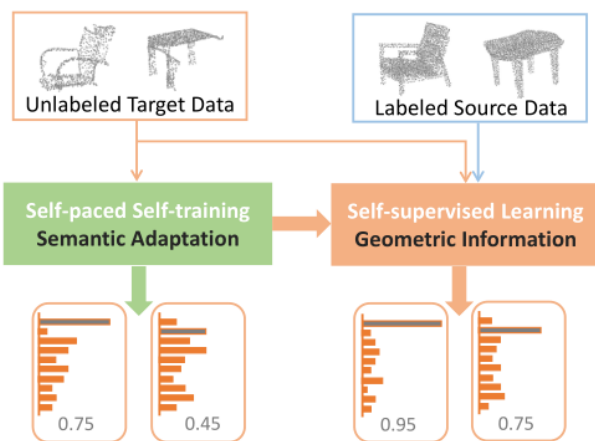


图 3 几何感知自训练

图 3 所示比较了传统的 Self-paced 自训练 (绿色

块) 和该文提出的几何感知自训练 (使用两个块), 从底部的预测分类概率图上可以对比看出 GAST 方法在语义表示学习上的有效性。图 4 为 GAST 模型网络结构图。

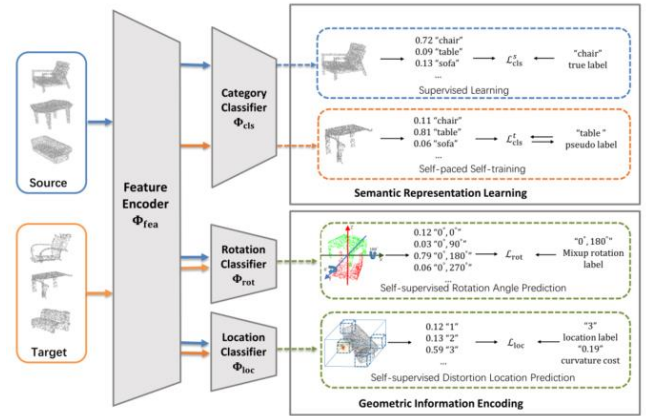


图 4 GAST 结构图

论文地址: <https://arxiv.org/pdf/2108.09169.pdf>

代码地址: <https://github.com/zou-longkun/GAST>

3、MLSP

介绍: 该文提出掩蔽局部结构预测 MLSP (Masked Local Structure Prediction) 方法。除了在源域上进行监督学习外, 该方法同 DefRec 和 GAST, 通过构建自监督任务学习共享的特征编码器将源域数据和目标域数据嵌入到共享的特征空间中。具体来说, MLSP 通过估计点基数、位置和法向量来预测被掩蔽的局部结构。

MLSP 的设计理念在于: (1) 点基数反映了对于特定域不变的基本结构特征 (点、线、面); (2) 预测被掩蔽区域中的点位置信息将使模型能够从部分观测中推断结构, 从而学习不完全引起的域偏差的鲁棒表示; (3) 点法向量由其周围邻域点计算得来, 因此其具备对噪声的鲁棒性。

MLSP 还发展了一种自步调学习的变体, 其利用预测概率熵来选择可靠的伪标签样本。其动机是具有较小

熵的目标样本更有判别性，其伪标签大概率是正确的。为此，模型计算各个目标样本的熵，对熵小于阈值 r 的样本选择生成伪标签，和源域一起送入分类头进行学习。虽然阈值 r 固定，但是随着模型训练过程逐渐变强，模型所选的目标样本数量越来越多。

图 5 所示为 MLSP 的结构。有监督支路将源域点云作为输入，并使用真实标签计算交叉熵损失。自监督支路将目标域点云作为输入，并使用提出的掩码局部结构预测任务来计算自监督损失，包括预测基数、位置坐标和法向量。

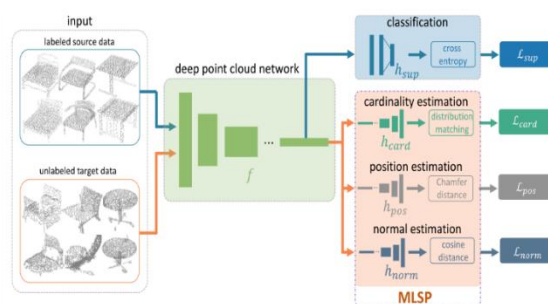


图 5 MLSP 结构图

论文地址: https://www.ecva.net/papers/eccv_2022/papers_ECCV/papers/136630159.pdf

代码地址: <https://github.com/VITA-Group/MLSP>

责任编辑 李策 贾同



李俊俏

硕士研究生，北京航空航天大学人工智能研究院，研究方向为计算机视觉。



王田

北京航空航天大学人工智能研究院从事教学与科研工作。研究方向为计算机视觉，模式识别，图像处理，机器学习，类脑计算和数据挖掘。

裂痕缺陷分割数据集

深圳大学 钟小品 刘维湘 易健业

语义分割是指将一幅图像中的像素逐一进行分类的方法。在表面缺陷分割中，我们对每个像素进行密集的预测推断，将表面缺陷进行像素级分类，从而可以精确地定位缺陷并细致地描述缺陷的轮廓。基于机器视觉的表面缺陷检测装备已经在各个工业领域广泛应用，包括汽车、家电、机械制造、半导体及电子、化工、医药、食品包装、公路检修、航空航天等行业。这些装备在表面缺陷检测方面取代了传统的人工肉眼检测，实现了自动化和高效。

目前，针对表面缺陷检测的数据集众多，这些数据集为评估表面缺陷检测算法的性能提供了相对客观的衡量标准，推动了该领域的标准化发展，成为科学研究的重要指南。其中，裂痕缺陷作为表面缺陷的典型代表，一直受到广泛关注。本文将依次介绍四个精标注的数据集：道路裂缝公共数据集 Crack Forest Dataset (CFD)、瓷瓦裂痕缺陷公共数据集 Magnetic Tile Surface Defects (MTDD) 的子集 Crack Defects、金属裂痕缺陷公共数据集 Kolektor Surface-Defect Dataset (KolektorSDD) 以及我们公开的私有数据集太阳能隐裂缺陷数据集 Tiny Hidden Crack (THC)。这些数据集中的裂痕像素占比依次减小，通过对它们的研究和应用，我们不仅能够更好地理解 and 解决表面缺陷问题，推动相关领域的发展和创新，还可以探究前背景像素不平衡的语义分割问题。

1、CFD 数据集

介绍：公共数据集 CFD 是一个道路裂缝数据集，裂缝像素与背景像素的比值较高，达到 1:25。该数据集由 118 幅图像组成，能够大致反映北京城市路面状况。每张图片都有手工标记的真实标签，这些图像中包含了阴影、油污、水渍等噪声。该数据集的分辨率为 480x320。由于道路裂缝对交通安全和路面质量有重要影响，准确地检测和定位道路裂缝缺陷对于道路维护和管理至关重要。

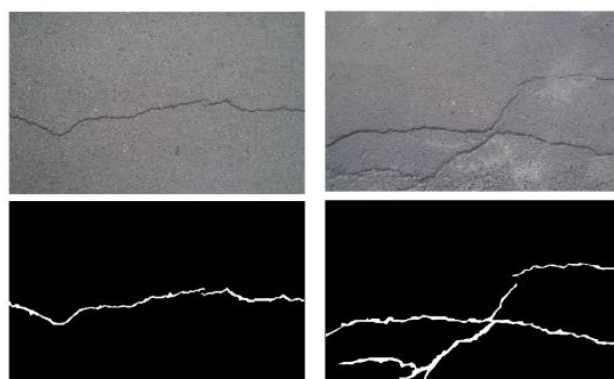


图 1 CFD 数据集示例

这个数据集提供了一个宝贵的资源，可以用于研究和分析城市道路的裂缝问题。通过对这些图像进行分析，研究人员可以更好地了解裂缝的形成和发展过程，从而为道路维护和安全提供更有效的解决方案。此外，该数据集还可以用于开发自动化的裂缝检测算法和工具，从

而提高道路维护的效率和准确性。

数据集下载地址: <https://github.com/cuilimeng/CrackForest-dataset>

相关论文链接: <https://ieeexplore.ieee.org/document/7471507>

2、MTDD 数据集

介绍: 公共数据集 MTDD (Multi-Tile Defect Dataset) 是一个提供磁瓦表面缺陷数据的开放数据集。为了模拟实际装配线上的生产过程, 在多光照条件下对给定的磁瓦进行了图像采集。这样可以更好地模拟不同环境下的光照变化对裂痕缺陷的影响。该数据集根据缺陷类型将数据划分为 6 个子数据集, 分别是 Blowhole (气泡)、Crack (裂痕)、Fray (磨损)、Break (破裂)、Inconsistent (不一致) 和 Free (无缺陷)。每个子数据集都提供了像素级标签。



图 2 MTDD 数据集示例

我们特别关注裂痕缺陷 Crack Defects, 并将其收集整理为一个独立的子数据集。实验计算后发现裂痕像素与背景像素的比例相对较低, 大约为 1:110。这意味着在图像中, 裂痕所占的像素数量相对较少, 相对于整个图像来说是一个较小的比例。经过数据增广处理后, 该数据集的裂痕缺陷共有 306 对像素级标注的样本, 每个样本的分辨率为 480x480。

通过使用公共数据集 MTDD 中的 Crack 子数据集, 可以进行各种图像处理和机器学习算法的研究和实验, 以提高对磁瓦表面裂痕缺陷的检测准确性和效率。

数据集下载地址: <https://github.com/KLIVIS/DIBE/tree/master/Crack>

相关论文链接: <https://link.springer.com/article/10.1007/s00371-018-1588-5>

3、KolektorSDD 数据集

介绍: 公共数据集 KolektorSDD 提供了一系列金属表面裂痕缺陷的图像样本。这个数据集中, 裂痕像素与背景像素的比值相对较低, 达到 1:300。金属裂痕缺陷的检测对于保证金属制品的质量和安全性具有重要意义, 因此引起了广泛的关注和研究。

该数据集是由 Kolektor Group 提供并注释的一组电子换向器图像。这些图像展示了电子换向器表面细小碎裂或裂痕的情况。数据集共包含 399 张图片, 其中 52 张为可见裂缝缺陷的图像, 而 347 张则是没有任何缺陷的样本。这个数据集的原始图像宽度为 500 像素, 高度在 1240 到 1270 像素之间。为了方便后续的训练和评估, 图像的分辨率已经被调整为 512x1408。

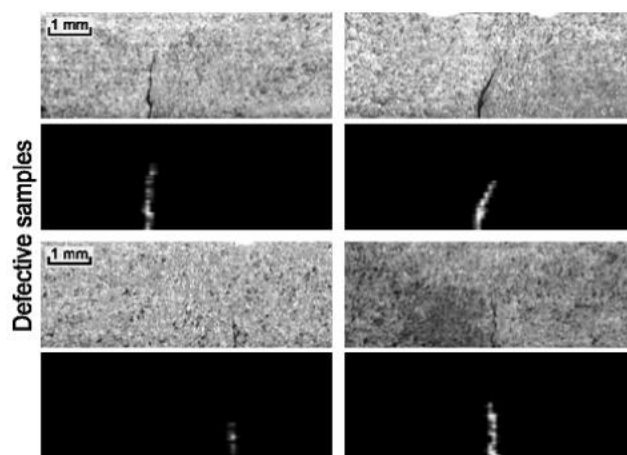


图 3 KolektorSDD 数据集示例

通过这个数据集, 研究人员和开发者可以深入了解电子换向器的缺陷问题。裂缝的出现和发展是电子换向器质量控制和性能改进的重要因素。通过分析这些图像, 可以研究裂缝形成的原因、发展的规律以及与之相关的因素, 为改进电子换向器的设计和制造提供有力支持。

数据集下载地址: <https://www.vicos.si/resources/kolektorsdd/>

相关论文链接: <https://link.springer.com/article/10.1007/s10845-019-01476-x>

4、THC 数据集

介绍: 原片上料是太阳能硅晶片制作的第一道工序, 它对整个生产流程的质量和经济效益影响至关重要。如果在这个阶段能够及时发现具有隐裂缺陷的硅晶片, 并将其退回给硅片厂家供应商, 将能够最大限度地减少经济损失。隐裂缺陷通常是肉眼不可见的微小裂痕, 因此, 需要采用特殊的成像技术来检测隐裂缺陷。目前, 激光透射成像方案被广泛应用, 因为它具有价格低廉且符合工业生产需求的优势。

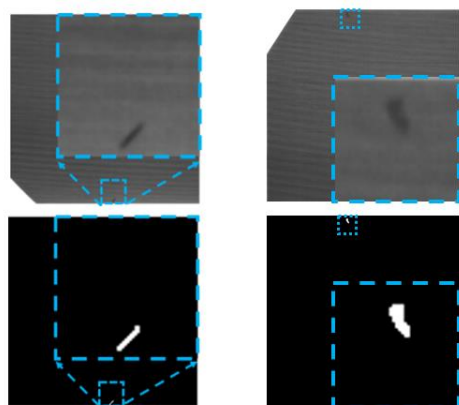


图 4 CFD 数据集示例

通过激光透射成像获取的原始图像分辨率为 2800x2600, 但为了方便后续的处理和训练, 经过预处理后, 得到的太阳能硅晶片图像分辨率为 1344x1344。然而, 由于 1344x1344 的图像尺寸较大, 直接将其输入网络进行训练会消耗大量的计算资源。同时, 隐裂缺陷非常微小, 如果进行下采样处理, 将会丢失大量缺陷信息, 使得检测变得更加困难。因此, 预处理后的图像被切割成 3x3 的小块, 分辨率为 448x448。这样做的好处是既减小了图像的尺寸, 减少了计算资源的消耗, 又能够保留更多的缺陷信息, 有助于更准确地进行检测。对该数据集进行精细的像素级标注, 进而构建出 1874 对精标注的太阳能硅晶片隐裂缺陷数据集。值得一提的是, 该数据集中隐裂缺陷与背景像素的比值极小, 达到 1:2200, 这进一步突显了隐裂缺陷的微小和难以察觉的特性, 准确地检测和定位隐裂缺陷对于太阳能硅晶片的质量控制和性能提升至关重要。

通过以上的技术处理和数据集构建, 我们得到了太阳能硅晶片隐裂缺陷的缺陷分割数据集。这将有助于提高生产效率, 降低质量风险, 并为太阳能硅晶片行业的发展做出积极贡献。

数据集下载地址: <https://github.com/KLIVIS/DIBE/tree/master/THC448>

相关论文链接: <https://arxiv.org/abs/2211.05295>

责编委 王田 樊鑫



钟小品

副教授，硕导，深圳大学机电与控制工程学院，研究方向为计算机视觉、智能制造、工业缺陷检测。



刘维湘

副教授，硕导，深圳大学机电与控制工程学院，研究方向为模式识别和机器学习、智能制造、医学图像处理。



易健业

硕士研究生，深圳大学机电与控制工程学院，研究方向为语义分割、工业缺陷检测。

好文推荐

南开大学的“A Highly Efficient Model to Study the Semantics of Salient Object Detection”最新成果发表在 IEEE TPAMI 2022。

论文: Ming-Ming Cheng, Shang-Hua Gao, Ali Borji, Yong-Qiang Tan, Zheng Lin, Meng Wang. A Highly Efficient Model to Study the Semantics of Salient Object Detection, IEEE TPAMI, 44(11): 8006-8021, 2022

基于卷积神经网络的显著性检测 (SOD) 方法已经实现了出色的检测效果。然而, 语义信息的编码方式以及这些方法是否与类别无关仍然缺少探索。显著性检测模型是基于 ImageNet 预训练的骨干网络建立的, 这导致了信息泄漏和特征冗余, 进而成为了探索上述这些问题的一个主要障碍。为解决这一问题, 本文首先提出一个极其轻量化的整体网络用于显著性检测, 这一网络可以摆脱分类任务的骨干网络并且从头进行训练。本文之后使用这一网络探索显著性检测模型的语义信息问题。基于本文提出的整体网络, 以及通过一种新颖的动态权

重衰减机制实现的表征冗余缩减, 所提模型的参数数量仅仅 100K (约大模型的 0.2%), 并且在主流的 SOD 基准测试上的表现与先进模型相当。使用 CSNet, 本文发现 a) 显著性检测方法和分类方法有着不同的机制, b) 显著性检测模型对类别不敏感, c) 基于 ImageNet 的预训练对于显著性检测模型是不必要的, 并且 d) 显著性检测模型需要的参数远少于分类模型。

具体而言, 如图 1 所示, 本文提出了一个极其轻量化的模型, 其整体地考虑特征提取器和显著性检测模块。首先, 本文将 OctConv 一般化, 得到 gOctConv, 其有着更高的灵活性以及额外的自适应能力。其次, 本文提出一种动态权重衰减机制, 基于此 gOctConv 可以实现自适应可学习的通道数。这种机制用于帮助在 SOD 模型中分析语义信息, 同时有助于模型在损失微不足道的性能的情况下减少约 80% 的参数。接下来, 利用 gOctConv, 本文提出了一种极其轻量化的整体的跨阶段跨尺度的网络, 即 CSNet。受益于这种整体的设计以及动态权重衰减机制, CSNet 在仅仅有 100K 参数(约 SOTA 模型参数的 0.2%)的情况下实现了与先进方法相当的效果。

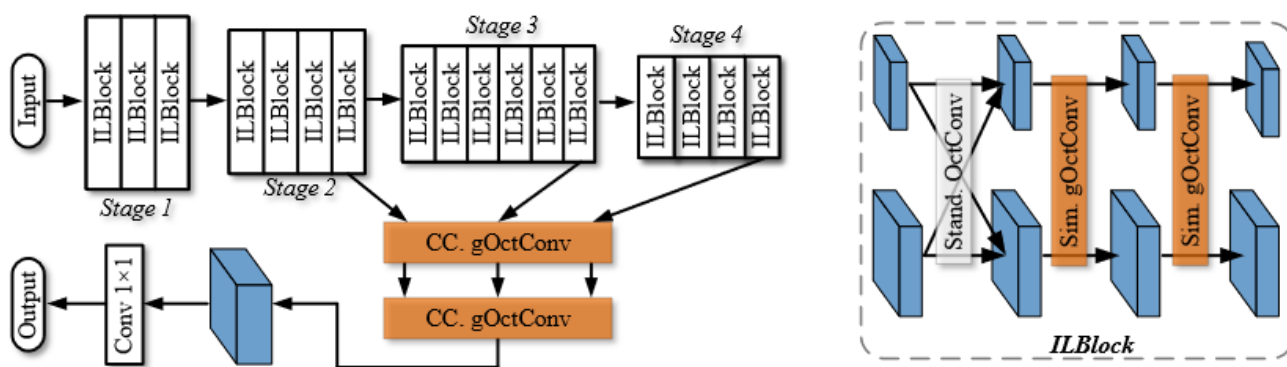


图 1 CSNet 的显著性检测模型的流程

责任编辑 贾同 樊鑫

好文推荐

哥伦比亚南卡罗莱纳大学的最新成果 “A One-Stage Domain Adaptation Network With Image Alignment for Unsupervised Nighttime Semantic Segmentation” 发表在 IEEE TPAMI 2023。

论文: Xinyi Wu, Zhenyao Wu, Lili Ju, Song Wang.
A One-Stage Domain Adaptation Network With Image Alignment for Unsupervised Nighttime Semantic Segmentation, IEEE TPAMI, 45(1): 58-72, 2023

本文解决了夜间图像的语义分割问题。在自动驾驶中夜间图像的语义分割与白天图像同等重要，但由于光照非常差且缺乏标注数据集，因此更具挑战性。夜间图像的语义分割问题可以被视为一个无监督的域适应 (Unsupervised Domain Adaptation, UDA) 问题，即应用白天拍摄的其他标记数据集来指导分割网络的训练，同时减少域偏移，使训练后的模型可以很好地泛化

到所需的夜间图像域。然而，目前的通用无监督的域适应方法不足以解决白天和夜间域之间的显著外观差异。为克服如此大的域差距，本文提出一种新的域适应网络，名为 DANIA。DANIA 用于夜间语义图像分割，利用有标记的日间数据集(源域)和包含粗对齐的昼夜图像对的无标记数据集(日间和夜间的目标域)。这三个域通过对抗性训练在网络中进行多目标自适应。对于未标记的昼夜图像对，使用白天图像上静态物体类别的像素级预测作为伪监督，来分割对应的夜间图像。此外，还有一个图像对齐步骤，通过估计一个流来细化白天图像产生的伪监督，以缓解昼夜图像对间的不对齐造成的不准确性。最后，本文采用重加权策略 (Are-weighting Strategy) 进一步提高预测精度，特别是提高小目标的预测精度。所提出的 DANIA 是一个用于夜间语义分割的单阶段自适应框架，不作为单独的预处理阶段训练额外的昼夜图像迁移模型。在 Dark Zurich 和夜间驾驶数据集上的广泛实验表明，DANIA 在夜间语义分割方面取得了最先进的性能。

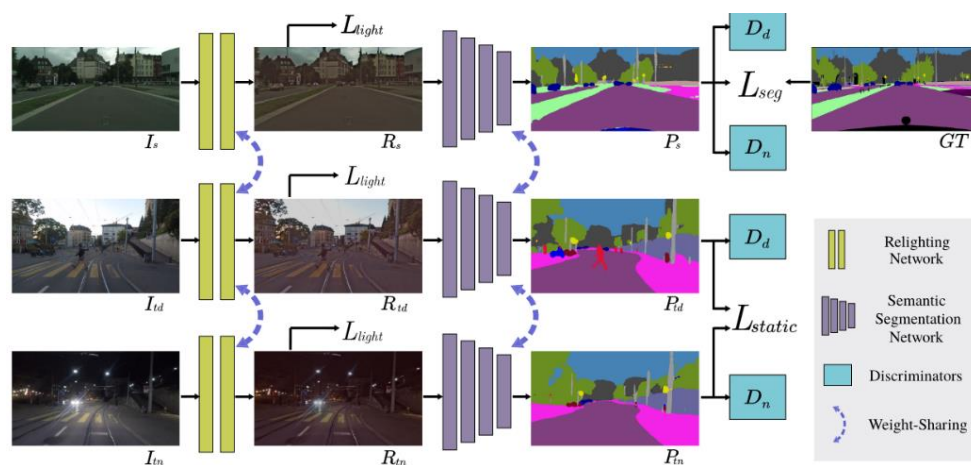


图 1 DANIA 结构图

责任编辑 李策 樊鑫

好文推荐

西安交通大学、苏黎世联邦理工学院、西北工业大学深圳研究院、西北工业大学、哈佛大学和维尔茨堡大学 “CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion” 最新成果发表在 CVPR-2023。

论文: Zhao Z, Bai H, Zhang J, et al. CDDFuse: Correlation-driven dual-branch feature decomposition for multi-modality image fusion[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2023: 5906-5916.

图像融合是计算机视觉领域的重要研究方向，其目的是将源图像中的重要信息结合起来，生成信息融合后的图像。可见光-红外图像融合(IVF)与医学图像融合(MIF)是多模态图像融合(MMIF)中两个具有挑战性的子方向，其重点是对来自所有传感器的跨模态特征进行建模，并将其聚合到输出图像中。近年来，学者研究了多种方法以应对 MMIF 的挑战。然而，现有方法有三个主要缺点：首先，CNN 的内部工作机制难以控制和解释，导致跨模态特征提取不足；其次，与上下文无关的

CNN 只在相对较小的接受域内提取局部信息，很难提取全局信息，从而难以生成高质量的融合图像；最后，融合网络的前向传播过程经常导致高频信息丢失。

为了解决跨模态特征建模和模态共享特征分解难题，文章提出了一种新的基于相关性驱动的特征分解融合网络(Correlation-Driven Feature Decomposition Fusion, CDDFuse)。首先，CDDFuse 使用 Restormer 块提取跨通道浅层特征。然后，引入了一种双分支 Transformer-CNN 特征提取器，该特征提取器使用 Lite Transformer(LT)来处理低频全局特征，而可逆神经网络 Invertible Neural Networks(INN)用于提取高频局部信息。在此基础上，进一步提出了一种基于嵌入信息的相关性驱动损失，使低频特征相关而高频特征不相关。最后，基于 LT 的全局融合层和基于 INN 的局部融合层输出融合图像。CDDFuse 方法体系结构如图 1 所示。

文章在 MSRS 数据集上训练网络，并采用 RoadScene 数据集进行验证；采用 MSRS、RoadScene 和 TNO 作为测试数据集，对融合性能进行全面验证。实验结果表明 CDDFuse 能够有效地提高图像融合的性能，在可见光-红外图像融合和医学图像融合等多种融合任务中都取得了很好的效果。

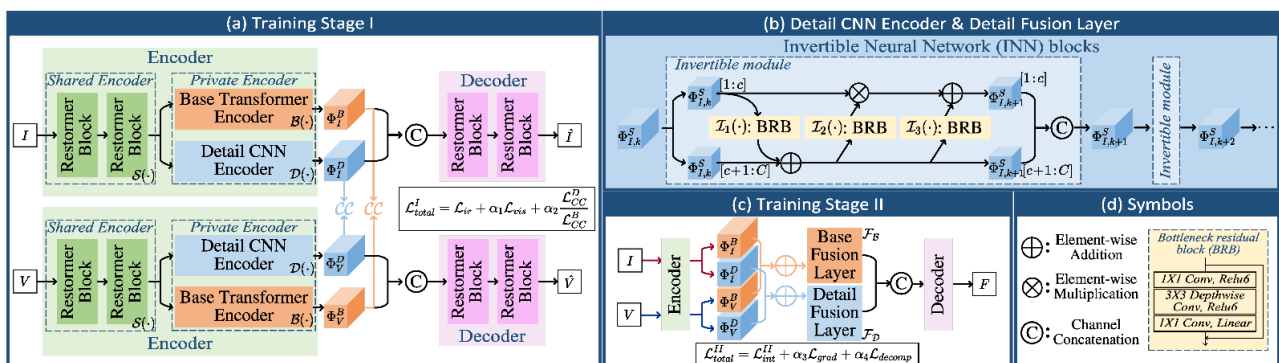


图 1 CDDFuse 方法体系结构 (以 IVF 为例)

责任编辑 王田 贾同

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 Journal of Biomedical and Health Informatics, Pattern Recognition Letters, Image and Vision Computing 和 Neural Networks。

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
AISTATS 2024	2024.05.02-04	Valencia, Spain	2023.10.14	https://www.aistats.org/aistats2024/
CVPR 2024	2024.06.17-21	Seattle, USA	2023.11.11	https://cvpr.thecvf.com/Conferences/2024
ISBI 2024	2024.05.27-30	Athens, Greece	2023.11.11	https://biomedicalimaging.org/2024/
ICPRAI 2024	2024.06.18-21	Jeju, South Korea	2023.12.15	https://brain.korea.ac.kr/icprai2024/

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
JBHI	Multi-modal joint learning in healthcare imaging	https://www.embs.org/jbhi/wp-content/uploads/sites/18/2023/04/IEEE-JBHI_call-multi-modal.pdf	2023.10.25
PRL	11th Iberian Conference on Pattern Recognition and Image Analysis	https://www.sciencedirect.com/journal/pattern-recognition-letters/about/call-for-papers#special-section-for-awarded-papers-from-11th-iberian-conference-on-pattern-recognition-and-image-analysis-ibpria-2023-ibpria-2023	2023.10.20
IVC	AI on Digital Health: Computer vision applications in medical imaging	https://www.sciencedirect.com/journal/image-and-vision-computing/about/call-for-papers#ai-on-digital-health-computer-vision-applications-in-medical-imaging	2023.12.01
NN	Graph Representation Learning	https://www.sciencedirect.com/journal/neural-networks/about/call-for-papers#graph-representation-learning	2023.11.01

心底无私视界宽 ∞ 韦穗教授专访

自 50 年代以来，我国就已经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。

《视界专访》栏目希望通过对这些见证者作一个系列专访，帮助从事计算机视觉及相关领域的科研工作者或爱好者，全方位地了解 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。

2023 年 7 月 11 日，张军平教授代表中国计算机学会计算机视觉专委会 (CCF-CV)，在安徽大学采访了韦穗教授，向她请教了其从事计算机视觉研究，尤其是三维成像的经历和经验。韦穗教授虽然年近 80，仍然坚持每天来办公室继续从事科研工作，且思维敏捷，说话铿锵有力，非常健谈。不知不觉，韦老师、她的学生们和我就在其办公室聊了一个多小时。通过沈川老师、徐佳博士后和我对语音稿的整理，最终形成了问答形式的采访纪录。以下是韦穗教授的简介和专访内容。

张军平 (后缩写为张): 您是如何选择计算机视觉和人工智能这个研究方向的？什么样的家庭环境或成长机遇造就了您和姐姐韦钰同时成为男性占据主流的工科领域的女性科学家，还成为了被人赞誉的中国“校长姐妹花”。

韦穗 (后缩写为韦): 对于这个问题，要讲我和我姐姐的事，姐姐她是比较反对“校长姐妹花”的描述。她觉得自己就是一名普通的教育工作者。曾经有一个出版社要给每一位工程院院士写一本书，结果我的姐姐韦钰带头反对，她不喜欢这样的宣传方式。不过，我可以说说家庭和教育对于我们成长的影响。尽管那时候家里姊妹特别多，但是父亲有一句话让我影响深刻，就是“你们能读书、我尽量供你们读”。

我特别感谢中小学校给了我优良的教育基础。小的时候，我读的是南师附小，当时南师附小正好实行师霞教育法—南师附小现在名叫师霞学校。师霞老师教得特别好，那里的学生也特别好。不过，我不是师霞老师教的，而是吴家翼老师教的，她们都非常好，对我们都非常好。吴家翼老师特别喜欢我，她说她没有女儿就把我当女儿。当时江苏省 1956 年学苏联，举办江苏省暨南



图 1 张军平 (左) 采访韦穗教授 (右)



图2 韦穗在南师附中第四小队的合影

京市少先队夏令营。南师附小送一男一女，就把我送去了，所以我有机会接触到很多不同的小伙伴，这对我的性格方面有很大的影响。

后来我又读了部队的一个小学，国家提供了很多好的教学环境，印象最深的就是我们校长叫林奔，还是长征过的。我们当时住的条件非常好了，都是打蜡地板，但是我们脱鞋在门口甩了一地。她第二天把我们召集起来，她就说我一晚上没睡好，我不希望我们的孩子是这样没有纪律的，不知道珍惜的孩子，所以我觉得这个印象很深。后来我进了南师附中，南师附中的教育也是非常好的。再后来我能有机会出国访学，实际上依赖于南师附中打了很好的底子。南师附中的师资非常好，当时老师们的外语都教得特别好，我觉得老师那种鼓励的教学方法对我们非常有用。

张：您认为作为一名教育工作者如何培养学生的社会责任感，您曾任过校长，也担任过全国政协委员，同时也是一线的教育科研人员，能否分享一些你和学生曾经的故事或经历？

韦：尽管我担任过一段时间的高校副校长工作，但是我觉得说清楚教育这个东西是特别难的，可以说每个人都是教育家，但是我们怎么去理解教育却很困难。记得我在做全国委员政协委员的时候，有一件事情印象深刻。曾经有一段时间，社会风气比较功利，我们的公办学校被贬得很低，有人说国家投入你们那么多钱，你们还不

如民办学校，你看民办学校还不需要国家投入很多钱，但是效益挺不错。我想，这里就是有一个社会责任感的问題。当时我就反驳说，大家要想清楚了，我们也不都是无能之辈，也不都是贪官污吏，如果一切都要看效益，那我们这些学校可以不要数学，不要物理，不要化学，更不要哲学、历史。比如这些民办高校，条件有限时，他肯定不开办这些专业，会直接只上一些应用学科。但是，不能大家都去这么干，我们有一些大学，它必须不计成本地开办一些基础学科，把文化、科学传承下去。如果大家说这些都可以不要，我们可以集中精力与民办高校一起在应用学科比拼，看谁做得更好。后来经过我这么一说，质疑的声音小了很多。

我对大学的理解首先是要有文化和科学的传承，这个是肯定要做到的。实际上怎么办好大学确实有各种各样的说法，比如，之前，有部长就希望大学里办科学院。但我自己是从科学院调过来的，我觉得大学应该有大学的特色。孩子们把人生中最好的时间交给了我们，我们首先应考虑的是育人！通过他们将文化和科学传承了，能使他们出去有个好平台发展自己，我认为这是我们最主要考虑的问题。

具体如何实现，我觉得可能有各种方法。但是无论怎么教，至少要先教一些科学的ABC内容。比如我们开玩笑讲，要传承科学，首先得有共同的语言。就像那个时候的工农兵学员，考试的时候说1/2大还是1/4大，答1/4大，他没有这个语言，1/2和1/4它表示的什么东西他都不清楚，所以就没有办法进一步科学教育。我觉得基础的东西要教，但是更重要的，应该是教会孩子们怎么能够面对一件新的事物。我觉得这应该是研究生的事，但有的时候有些本科生你带得好，他们也非常地出色。就像我之前带本科生做创新训练计划，因为要做一个课题的话，需要综合的能力，有时候碰到的问题是书本上找不到的，这对孩子们来讲就是有挑战性的。我们可以教学生如何面对陌生的问题。我跟学生们讲我们是工科专业，是学工程的，我们造一座大桥不可能把每个螺丝钉都搞清楚。但是，你必须把你这里的主要问题



图 3 韦穗教授与指导的本科生



图 4 韦穗教授与指导的研究生

搞清楚，把主要问题啃下来就好，这是你的能力。我觉得要把孩子们教到一定的能力，如果你没有一定的能力，你没有一定的数学、物理和外语的能力，你就没办法把先进的科学传承。

所以我主张大学里头还是要重视传承和人的培养。我对现在不满意的是，我们的教学能不能清楚什么是基本的东西，什么是先进的东西。我们又怎么能把这两个兼容的东西教给孩子，有些基本的东西还要坚持。

张：能否分享一些您的科研经历中印象深刻的故事。

韦：我曾经在中国科学院工作，75 攻关做通用实时图像处理系统，在上海气象局和无锡 702 所得到应用。在我过来安大之前，我到白纸坊 23 号北京印钞厂参观过，他们就给我提出过一个问题。就是套印毛主席像，即毛主席像要和钞票套印起来，但精度要求特别高。他们套印的时候经常需要请对应公司的工程师到现场。后来我说，你停机给我看一看，一看就发现工程上的东西有的

时候并不是那么难。实际上他们就对准了几个标记点，因为两个板子本身的印刷精度是很好的，就是对准问题。毛主席的两个眼睛、嘴角，毛主席的痣，几个标记点对准了，就可放心套印了。所以工程师们解决工程问题的方法是很聪明的。后来我说你以后套板可以不请对应的工程师到现场了，只要把几个标记点对准了就可以了。他们开玩笑说你是真专家，你能不能给我们做钞票检测。然后我看了一下现场，因为钞票检测要求还是特别高的。我说可以，能做一个样机。这个样机可以做到，计算机识别是绝对正确的，那肯定是绝对正确的；计算机识别是绝对错的，那肯定是绝对错的；但是，我会有一大堆的拒识。这些我是不知道好坏的，还得人工比较，然后我再逐步提高。对方说我说的比较清楚，没有忽悠他们。我也希望我们的系统能继续在这里得到应用和改进。可惜，后来因为很快调到安徽大学，因为工作单位变更等诸多原因，就没有再继续与对方合作了。但是作为计算机视觉技术在应用领域的实例，让我印象深刻。

张：互联网学院院长梁栋是您的学生，另外，您还有哪些得意门生？他们有没有一些共性的优点呢？关于大学生研究生博士生的学生，你有什么需要强调的呢？

韦：梁栋是我到安徽大学后早期带出的博士生。我到安徽大学来的时候，安徽大学是处于全国 211 学校比较靠后的位置。因为当时，全国的 211 学校只有我们学校和广西大学还没有博士授予单位，不是说没有博士点，而是没有博士授予单位。这个与 211 工程建设背景有一定渊源。当时李铁映做教育部部长的时候，国家提出面向 21 世纪重点建设培养 100 多所大学，规定教育部所属学校建设一部分，然后各个机部建设一部分，各省市建设一部分。安徽大学是安徽省省属学校，所以安徽省重点支持安徽大学进入 211。合肥工业大学（工大）当时吃了这个亏，工大当时属于一机部，但是一机部当时选一所学校的时候，他选的是吉林工业大学。当时，北航是属于航天，南航就属于航空，科学院只有科大。就是每个基部建设一所，非要找到一个基部。像浙大就属于教育部所属学校，所以他们教育部所属学校挑 1/3 就 30

多所。后来再给一机部一个新名额的时候，还是没有给工大，给了湖南大学。这时工大才回过神来，要到安徽省申报，但安徽省党代会已经确定支持安徽大学了，所以工大开始不是 211，但后来学校合并了，工大就有了进入 211 的名额，另外当时他们进入教育部了也很不错。

对于我们学科而言，优秀的学生都应具有好的编程能力。之前给本科生上课时，我们的孩子说自己不会编程，我说不会编程就等于不会说话，在这个社会上，对于理工科而言，不管你具体学什么专业，必须会编程。不管是学什么电子的，学什么自动化的，学什么的，你必须会编程。现在计算无处不在，不会编程不行。我以前的学生阮宗才的编程是非常好的，是合肥科学岛上过来的。后来的学生跟着他学得非常好。因为有这个基础，他们这一代培养比梁栋那代人培养要好一些。而梁栋那代人的培养应该说还是年纪大一点，但比我的培养好多了。我是 64 年入大学的，66 年文化大革命，实际上我就读了两年。因为两年后就开始文化大革命了，所以最早期的一些学生应该说在基础教育方面还是有缺陷的。如果要站在比较高的视角来做研究，还是需要好的基础。像我在美国访学期间的导师是数学的本科，统计的硕士，工程的博士。给我感受最深的是他送我的证书，提到我在“小面积拟合”算法上的工作。事实上这是他升院士的代表作中的相关工作。我的工作主要证明算法可用的可能和局限。他却非常认可。这也证明他说到一句话：要有勇气不仅追求知识的表面，而且要追求它的深度。

当年我大学毕业后，是被分配当工人的。当时我们工科学的高等数学是樊映川老师编的，学好他的，当个工人是够的。后来我到了美国以后，发现樊老师讲的太浅了。要想做科研的话，只有这个台阶或基础是远远不够的。这其实也是大学生普遍有的问题。在大学里，学生们读大学的时候都认为自己什么都懂了。我说什么都懂，就叫大学生；只懂一点是硕士生；什么都不懂，就是博士生。我们的目标是要能把一个博士生培养出来，知道什么都不懂了，这就是博士生了。当然我们大学也不能把学生们的一生都包了。我希望教会他们将来碰到



图 5 韦穗教授在访学期间，导师针对其在图像分析上的贡献颁发的证书

问题，知道该如何解决问题。举例来说，我以前没做过校长，但是我来了以后，我按照我做科研那套习惯走。首先看看周围人是怎么做的，然后看看别的学校跟我一样的校长是什么情况，然后再去查查书再看看，然后再根据我自己情况我自己再分析，再不断地实践，在失败中总结。我觉得这个能力如果你会了，不管你现在学什么，将来做什么，都是有利的。所以我说，科研能力可以触类旁通。只要拥有这种能力，你将来不管做什么事情都是有好处的。

张：您现在还经常来学校实验室从事科研工作，能否聊聊，驱使您继续研究工作的动力是什么？

韦：我现在还要到实验室来，这就跟科研有关系了。就我所知，之前有一段时间，国内科研的风气不是太好。当然，有一些也是被逼的，因为需要科研人员出好论文、拿项目才能升职称。所以，实际上有的时候很多研究都是从 paper 到 paper。很多工作是因为人家已经有基础了，大家在上面再找一点问题，后来再往前走一点，很多东西都沿着一条线或同样的线索在走。这样下去的话，我觉得如果拿打仗来打比方的话，是要死人的。但有些东西也不能完全怪我们，就是说有的时候我们也不



图 6 韦穗参加 863-512 机器人传感技术实验室学术委员会三届四次工作会议



图 7 物联网中的图像传感器前沿技术发展预测和对策研究合影

能一步到位，立马出产品，因为产品化的事有时不一定是大学能做的。我举一个非常简单的例子，当时我做 863 专家，邓小平同志给的指示就是发展高技术，实现产业化，然后我们做机器人研究的专家就跑到一汽去推广机器人技术了。但是给我们的感觉是：一汽不是很欢迎，为什么呢？原因在于，一汽当时做的汽车采用的是刚性生产线，但如果用机器人必须上柔性生产线，全是套膜套好了，咔哒一下就完成了。然而那时的生产力并没有达到这个程度。反观现在，汽车已经有各种各样的，这是因为生产力到了。换句话说，这个产业化不是一所大学能承担下来的，而是需要国家的科技发展一起跟上。后来，我们定的是技术驱动，工科做一些核心技术的前期研究是比较合适的。所以，我们必须努力找准自己的位置，不要为了争帽子去做那些，这个帽子有一顶就可以了。有的人开玩笑讲，除了绿帽子不要什么帽子都要，其实不一定是好事。当然有竞争的压力是好，但有的时候如果钻歪了就不一定是好事。

我们继续往前做研究还是值得的。之前我是从事计算机视觉方面的研究，是二维的，后来我想扩展一点，就选了三维显示技术。三维显示最简单的一种实现方法可以是戴眼镜的，就是左眼看一幅图像，右眼看一幅图像，就会有立体感，当然也有柱透镜的方案，还有集成成像微透镜阵列的方案。裸眼三维技术上最好的是全息显示技术。光全息显示技术可以实现很好的静态三维显

示效果，但是我们现代社会期望实现的是动态的视频显示效果。所以，我们要做视频全息。我编写了“全息成像概论”双语教材，并为本科生授课多年。视频全息技术面临两个主要问题，一个就是三维获取，另一个是三维显示。目前来看都是比较困难的，我主要专注于三维显示的问题，期望得到很高的空间带宽积，根据衍射的原理，通过微纳结构来实现。但是，要解决此问题还是很困难。

习主席提到实现中国梦，之前我跟学生们说我有一个梦，弥补我的一个遗憾。当时我从美国回来的时候，知道日本人在做数字摄像机，但是我们没有认识到此技术是成像的关键技术，只是用模拟摄像机做算法和应用。就是说日本人在研究数字摄像机的时候，我们当时其实错过了这个机会。如果开始跟日本一道研究，也许我们在数字摄像机领域不会有这么大的差距，不会导致 CCD 什么是人家的。也就是说，开始的时候没有抓住这个机会，没有开始研究一个核心的东西，核心器件应该是最重要的东西。后来选择另外一条路，看来选的不是那么成功，就是我去做系统，做出通用实时图像处理系统。当时做系统我对自己的工作比较满意，还得了中国科学院二等奖，而且跟加拿大的运行系统比，我们也达到世界先进水平的。现在回过头来想，其实并没有人家那么好，我们做的是样机，人家已经是产品。

所以我觉得对三维成像，我们可以做一点涉及核心

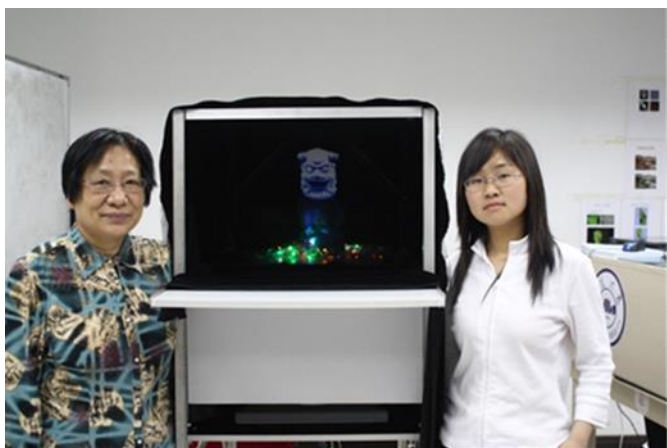


图 8 三维成像创新项目，韦穗教授与指导的大学生合影

技术突破的器件，而且对我目前的状态来讲，我不需要评职称了，也不需要“帽子”了，反而可以做一些自己喜欢的研究工作。尽管这个事很难，也有可能失败，但是至少自己不后悔。就是想试一试，我想如果不能成功，我做先烈，也可以给大家提供失败的经验，告诉后来人什么路走不通，也是有贡献的。

张：您认为女性科学家从事科研工作的难点是什么？

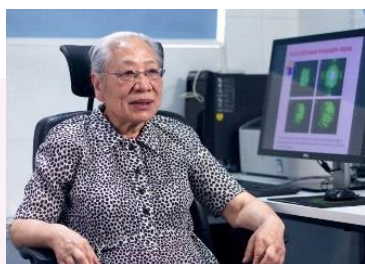
韦：我个人觉得还好。当时时代的原因，一开始我毕业后被分配当工人，后来国家提出“归口”政策就回来了。回来了以后我孩子也大了，不像现在年轻人，我从事科

研工作时面临的带小孩的压力小了很多。我觉得负担不大，因为我自己家庭的情况是我爱人付出了很多，我自己比较专注，家里的事基本不操心。所以我想我们还是要抓主要矛盾，必须有取舍，不可能全部都做。

张：能否给计算机视觉领域的青年工作者一个寄语？

韦：期望我们从事计算机视觉领域的青年工作者，在关注本领域重难点问题的同时，也多跟物理、多跟光学相结合。如果不把这些问题搞清楚，有些问题光是靠计算机视觉技术自身是很难解决的，有些非常经典的东西还很值得去看。例如我看到最近在北京举办的“首届国际基础科学大会”中请了英国学者 Michael Berry，他的一篇经典文章值得一读：Beyond Rainbows。他说光学是 Ramam 一生都在研究的东西，焦散是数学上的突变，这对我们研究成像包括计算摄像学都很有意义。尽量不要满足在现有的技术条件下做基本同一水平的工作，将电子和光子结合起来。我们的年轻人有朝气，是好事，就像我年轻时也一样年轻气盛，觉得自己就已经在世界水平上，但是在经历这么多年以后，就会发现我自己当时看的太浅了。希望大家静下心来，埋头苦干，认真把一些问题再理一理，再捋一捋，再次出发，再创新高。

责任编辑 张军平 贾熹滨 明悦



韦穗

安徽大学原副校长、教授、博士生导师，曾任全国政协委员、中国图象图形学学会副理事长、教育部科学技术委员会委员。她毕业于南京工学院（现东南大学），1983年4月至1985年9月在美国密执安大学及弗吉尼亚理工学院作访问学者，2007年受聘为安徽省人民政府参事。长期从事计算机视觉、图像图形学、模式识别、数学形态学和全息成像等领域的研究。承担了多项国家自然科学基金项目和“863项目”，在国内外期刊上公开发表学术论文80余篇。荣获国家科技部授予的“国家863计划先进工作者”称号及首批国家政府津贴，负责的大容量快速图像分析系统获中科院科技进步二等奖，是国家“863计划”机器人主题专家组成员。主持了2002年第二届国际图象图形学学术会议（ICIG），编辑了两本会议论文集，翻译出版了《计算机视觉中的多视图几何》。

COMPUTER VISION NEWSLETTER

03 2023
总第 37 期



计算机视觉专委会简报



CCF 计算机视觉
专委会