

主办 CCF 计算机视觉专业委员会

COMPUTER
VISION
NEWSLETTER

CCCF 计算机视觉 专委会简报

02 2025

总第 44 期



CCF 计算机视觉
专委会

COMPUTER VISION NEWSLETTER



计算机视觉专委会 简报

2025 年第 02 期

总第 44 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 | 王瑞平 | 中国科学院计算技术研究所 |
| 执行主编 | 朱安娜 | 武汉理工大学 |
| | 潘金山 | 南京理工大学 |
| 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/

| | | |
|-----|-----|--------------|
| | 任传贤 | 中山大学 |
| | 杨巨峰 | 南开大学 |
| 主 编 | 王金甲 | 燕山大学 |
| 编 委 | 崔海楠 | 中国科学院自动化研究所 |
| | 魏秀参 | 东南大学 |
| | 张 杰 | 中国科学院计算技术研究所 |

/委员风采/

| | | |
|-----|-----|---------|
| | 张 青 | 中山大学 |
| 主 编 | 余 烨 | 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 |
| | 王 田 | 北京航空航天大学 |

/海外学者/

| | | |
|-----|-----|----------|
| 主 编 | 金 鑫 | 北京电子科技学院 |
| 编 委 | 刘帅奇 | 河北大学 |
| | 于 茜 | 北京航空航天大学 |

/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 走进高校系列报告会
- 05 CCF CV 视界无限系列研讨会
- 12 CCF CV2025 年执行委员增选申请开始

| 科技前沿

- 13 基于反思的反绎学习
- 18 生成式遥感图像处理
- 27 ICLR 2025

| 委员风采

- 33 新疆大学库尔班·吾布力教授访谈
- 38 委员好消息

| 学术资源

- 39 多模态医疗图像辅助诊断开源代码
- 42 多模态三维目标检测数据集
- 45 好文推荐

| 海外学者

- 48 征文通知

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 144 期 北京工业大学



2025年5月15日，由中国计算机学会主办，中国计算机学会计算机视觉专委会、北京工业大学联合承办的第144期CCF-CV走进高校系列报告会在北京工业大学理科楼844学术报告厅成功举行。本次报告会邀请了北京大学蒋婷婷副教授、北京理工大学宋红教授、中国科学院计算技术研究所陈益强研究员三位专家学者作特邀报告。出席本次报告会的领导有北京工业大学计算机学院院长韩红桂教授，专委会常务委员毋立芳教授，本期报告会执行主席和主持人是北京工业大学的贾熹滨教授和马伟教授。

本次报告会聚焦“智慧医疗技术与未来发展”主题，吸引了众多的青年学者与在校学生。在开场致辞中，北京工业大学计算机学院院长韩红桂教授代表学院向与会专家和嘉宾表示热烈欢迎，并表示希望借助此次学术交流，进一步深化与各高等院校、科研机构的产学研合作，融入数智医疗领域国家及首都卫生健康发展战略布局，共同推动计算机视觉在医疗健康领域应用的技术发展和协同创新。随后，CCF-CV 常委毋立芳教授代表专

委会致辞，感谢讲者、执行主席及承办单位的大力支持，并简要介绍了专委会的主要工作与宣传平台，鼓励更多师生积极参与和关注相关活动。最后预祝本次报告会取得圆满成功！

本次报告会由北京大学蒋婷婷副教授、北京理工大学宋红教授、中国科学院计算技术研究所陈益强研究员做主题报告，内容涵盖了生物医学影像分析、微创手术导航以及模型联邦网络构建及应用等智能医疗技术。随后，三位专家与师生互动，共同探讨、交流，并对师生提出的问题做出详尽的回答，提出了许多有价值的学术见解，论坛现场气氛热烈。

本期 CCF-CV 走进北京工业大学系列报告会围绕“智慧医疗技术与未来发展”等前沿技术展开，涵盖了计算机视觉与医疗健康的交叉研究。三位专家分别从理论创新、技术突破和系统架构三个维度，全面展现了计算机视觉赋能智慧医疗的最新研究成果和实际应用。为与会师生呈现了一场精彩纷呈的学术盛宴。在报告环节中，专家们深入浅出的讲解不仅拓展了与会者的学术视野，更激发了大家对学科交叉研究的浓厚兴趣。随后的互动交流环节气氛热烈，参会师生积极提问，专家们耐心解答，思想的碰撞迸发出智慧的火花。此次报告会极大地鼓舞了师生们的科研动力，为师生搭建了一个难得的学习与交流平台，促进了知识与思想的碰撞与融合。

责任编辑 朱安娜

第 22 期 动态场景重建、理解与生成

CCF-CV 视界无限系列研讨会



圆满成功，并期望 CCF-CV 视界无限系列研讨会能够不断发展、越办越好！

上午的研讨会由北京大学**查红彬**教授，中国科学技术大学**张举勇**教授、复旦大学**张力**教授和南京大学**龙霄潇**副教授做主题报告，下午的研讨会由南京理工大学**李泽超**教授、华中科技大学**陶文兵**教授、南京大学**王贝贝**教授、香港大学**赵恒爽**助理教授做主题报告。

2025年3月29日，由中国计算机学会主办，中国计算机学会计算机视觉专委会（CCF-CV）、西北工业大学电子信息学院和陕西省信息获取与处理重点实验室承办的“CCF-CV 视界无限——动态场景重建、理解与生成”论坛于2025年3月29日在西北工业大学友谊校区正禾宾馆西四会议室举行。

研讨会由西北工业大学**戴玉超**教授、**惠乐**副教授、**刘奇**副研究员主持。研讨会开幕式由西北工业大学电子信息学院党委书记**侯俊**致辞。

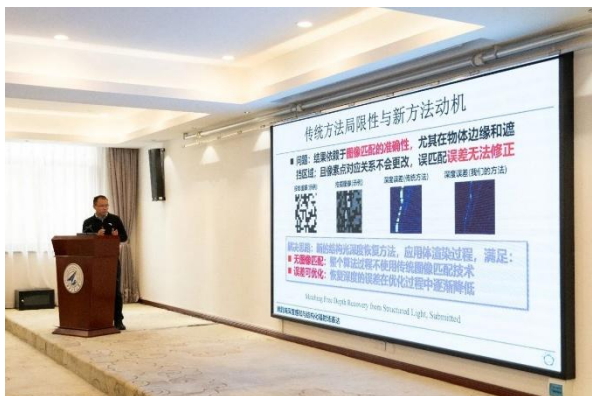


查红彬教授以“具身视觉与 SLAM：在线学习的途径”为题，指出当前具身智能的发展趋势不可避免地需要深入探讨环境感知与自适应能力的结合。SLAM 作为同步实现传感器定位与环境地图构建的核心技术，对于具身智能的实现至关重要。尽管现今的视觉系统在处理复杂环境时面临挑战，尤其是在泛化能力和环境自适应性方面，但在线学习方法为我们提供了有效的解决方案。在这一背景下，查教授分享了他们的近期研究成果，着重介绍了几项关键技术：首先是利用在线学习系统的忘却机制来优化动态 SLAM 算法，其次是应用于单目视频处理的全景三维几何与实例语义重建技术。此外，他们提出了基于在线持续学习的自适应 VIO 系统，并探讨了利用隐式地图提升全局性与不确定性表达能力的自主



侯俊书记简要介绍了学校和学院的基本情况，并对 CCF 计算机视觉专委会及各兄弟院校长期以来对电子信息学科发展的支持和帮助表示感谢，预祝此次研讨会

建图方法。这一系列研究显示，如何将在线学习与 SLAM 技术有效结合，将极大推动具身智能在真实复杂场景中的应用，从而实现智能体与环境的更为紧密的耦合。这为未来的智能系统发展开辟了新的方向，蕴藏着广阔的研究价值和实际应用潜力。



张举勇教授在“端到端深度感知与结构化辐射场表达”报告中，探讨了近年来以 NeRF 和 3DGS 为代表的辐射场表达技术在物体、场景与人物的重建与生成方面的显著进展，显示出高保真度和端到端可微的优势。然而，NeRF 等隐式辐射场依赖光线追踪的密集采样，导致计算资源消耗庞大，实时渲染难以实现；而 3DGS 等显式辐射场则由于无序点云特性，存在几何拓扑缺失的问题，无法有效支持物理属性嵌入与真实物理仿真。在本次报告中，张教授将首先介绍他们团队在无需匹配的结构光深度感知方面的研究进展。接着，将详细阐述针对 NeRF 与 3DGS 表达存在的不足，他们团队提出的 StructuredField 表达技术。这一结构化的可微辐射场表达方法能够同时实现高精度几何、高保真渲染及物理仿真，充分克服了现有技术的局限性，展示了其在多领域应用中的广阔前景。

张力教授以“基于生成式物理智能的自动驾驶闭环仿真研究”为题，强调了任意轨迹的街景合成技术使得端到端驾驶策略的闭环评估成为可能。尽管现有的方法在录制轨迹上的新视角合成表现出色，但由于驾驶视频通常以有限视角记录无边界的广阔环境，这些基于重建的方法在处理新轨迹时仍面临不少挑战。为了解决这一问题，张教授在本报告中介绍了一种新颖的任意轨迹驾驶视图合成方法。该方法通过利用视频生成先验来优化

自由轨迹下的三维模型。基于此，他们团队构建了一个全新的自动驾驶仿真引擎，具备以下功能：首先，支持多模态（摄像头和激光雷达）逼真场景渲染；其次，支持闭环评估，以适应各种自由形式的轨迹行为；同时，提供高度多样化的交通场景，以进行全面评估；并且支持多智能体协作，以考虑交互动态；最后，具备高计算效率，以确保经济性和可扩展性。在这一仿真环境下，团队模拟了三种驾驶类别：非交互性驾驶、安全测试以及多智能体交互模拟，以提供一个可靠且全面的基准，用于评估驾驶代理在现实世界中的表现。这些研究为自动驾驶技术的进一步发展提供了新的思路和工具。



龙霄潇副教授以“基于生成式 AI 的三维建模方法”为题，指出随着生成式人工智能(AIGC)的突破性进展，以 ChatGPT、Stable Diffusion 和 Sora 为代表的跨模态生成系统正在重构数字内容生产的范式，成为新质生产力的核心驱动力。作为数字内容生态的关键载体，三维模型的高效生成技术长期面临几何拓扑优化、材质物理属性建模等多维复杂性挑战，因此亟需通过 AIGC 技术实现范式的革新。在本次报告中，龙教授系统梳理了 3D AIGC 技术的演进脉络。首先，解析了神经辐射场

(NeRF) 与高斯泼溅 (Gaussian Splatting) 这两大革命性表征技术的原理与应用。此外，还将分享 3D AIGC 的背景与现状，以及面临的技术挑战。报告将探索如何重建和生成高质量的三维模型，包括从单张图像或文本描述中提取信息以重建高质量三维几何和纹理材质。这一系列的研究和探讨将为后续的三维内容生成技术提供重要的理论基础与应用指导，为实现高效和精确的数字内容生产奠定坚实基础。



李泽超教授在线上做了报告。他以“面向视觉内容异常检测的预训练模型知识增强”为题，强调图像视频大数据智能分析与识别在工业生产、瑕疵检测等多种实际应用中具有至关重要的作用。为此，课题组深入研究了开放环境下的视觉异常检测问题，特别关注如何通过预训练模型知识增强，挖掘模型知识实现预训练模型与下游异常检测任务的有效适配。在本次报告中，李教授介绍了通道-空间串联分析的语义分割模型，并阐述基于特征蒸馏学习、分区记忆、自监督分割引导和非对称蒸馏分割的异常检测方法。这些研究成果旨在提高视觉异常检测的准确性和鲁棒性。过这一系列研究，李教授期待为视觉异常检测领域的进步提供新的思路和实践指导。

陶文兵教授以“自适应表面网格重建”为题，介绍利用三维点云信息进行三维表面重建的相关技术。在这一领域，如何在充分表达物体三维细节信息的同时，尽可能减少三角面片的数量，一直是一个备受关注的矛盾。目前，典型的处理方法通常是先采用表面重建算法生成具有足够精度的三角网格表面，然后再通过表面简化算法对生成的表面进行简化，以降低面片数量、减少数据

存储开销并提高显示效率。然而，这种简化过程往往会导致大量表面细节信息的丢失。



在本次报告中，陶教授提出了一种创新的方法，将表面重建和网格简化统一到一个网络架构中。该方法在不降低表面重建质量的前提下，实现了三角网格面片数量减少超过 10 倍的显著效果。大量实验结果充分验证了该方法的性能和有效性。通过这一研究，陶教授希望为三维表面重建技术的发展提供新的思路，并推动相关领域的应用进展。



王贝贝教授以“三维几何材质重建：从宏观到微观”为题进行报告，强调现实世界丰富的几何结构、多样的外观和细腻的光影变化在自动驾驶和工业设计等应用中的重要性。构建真实感的三维数字世界需要从物理世界获取几何和材质信息，核心问题在于如何有效表达和约束这一信息。教授首先回顾了传统几何重建方法和材质表达模型的局限，指出自动化和高效性的重要性。她介绍了团队提出的创新方法，通过可微分模型联合优化几何与材质，以提高重建精度和效率。特别强调了结合不同材料表达形式与渲染技术，在多种材质类型的恢复

上取得的进展，包括高光材质和毛茸茸表面的处理。王教授还探讨了微观层面的研究，以布料恢复为例，说明了利用显微镜进行细微纤维重建的挑战和解决方案。通过这项研究，团队实现了对布料材料的更精细表达和恢复。在总结时，王教授强调尽管已有成果，但在三维重建领域仍面临诸多挑战，未来需继续探索更统一和有效的模型以应对复杂的物理和视觉现象。她鼓励与会者积极参与到该领域的研究与探索中，为实现更真实的三维数字世界贡献力量。



赵恒爽助理教授以“智能视觉空间理解与推理”为题进行报告，指出随着深度学习模型的增强以及海量数据的高效获取与利用，大规模视觉基础模型的构建备受关注。这些模型在处理跨领域复杂视觉场景任务时展示了强大的泛化能力。然而，它们通常集中在图像和视频理解上，忽视了高维视觉场景的空间理解与推理，这对具身智能和自动驾驶等下游应用至关重要。报告首先介绍了空间理解的基础模型，Point Transformer V3 和 Depth Anything，分别作为高效且性能出色的点云理解架构和数据驱动的单目深度估计框架。接着，赵教授阐述了 Sonata，一个通过大规模自监督学习实现可信空间表征的框架。随后，介绍了空间推理架构 GPT4Point 和 GPT4Scene，这两者分别针对点云和视频场景实现空间推理的能力。通过这两种架构，视觉系统能够在不同的场景中进行高效的空间推理，从而提升理解与决策能力。最后，报告讨论了智能视觉理解与推理面临的现有挑战以及未来的研究前沿，包括如何进一

步提升模型的泛化能力和推理准确性。赵教授期待通过这些研究推动智能视觉系统在多领域的应用，助力具身智能与自动驾驶等技术的发展。

在会场交流环节，与会专家与在场老师和同学们针对各位专家的报告内容进行了深入的交流与讨论。

在主题研讨部分，研讨嘉宾樊斌、刘奇、惠乐围绕“动态场景技术的挑战与应用边界”展开深入讨论。

樊斌探讨了“脉冲视觉相机动态重建影像重构”。他介绍了脉冲相机的工作原理及其在快速运动场景中的优势，通过深度学习方法恢复高质量图像，提高动态场景的感知精度，并展示了其在工业检测和气象监测等领域的应用前景。

刘奇在报告中探讨了“卫星影像城市三维建模”的重要性，强调其在经济和国防中的作用，指出基于卫星影像的三维重建相较于其他方法有独特优势，但面临建筑密集、视角限制及信息缺失等挑战。为提升重建质量，提出一种引入语义信息的新方法，并展示了相关实验结果。

惠乐的报告聚焦于“三维场景理解与感知”，讨论了三维目标检测、场景识别和语义分割等研究，特别是点云数据处理的挑战。他介绍了基于深度学习的方法，旨在提高处理效率和识别精度，并展示了语言驱动模型在提升三维数据理解上的应用潜力。

最后，戴玉超教授在闭幕致辞中强调，本次论坛邀请到动态场景分析领域专家带来最前沿的高水平报告，全面展现了动态场景重建、理解与生成方向的创新活力，为计算机视觉与产业应用的深度结合指明了路径。论坛在热烈的学术讨论氛围中圆满落幕。

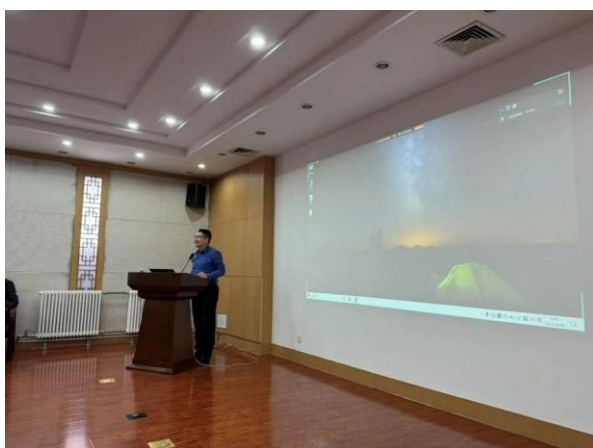
第 23 期 具身视觉：基础学习理论与关键感知技术

CCF-CV 视界无限系列研讨会



2025年4月20日，由中国计算机学会计算机视觉专委会主办、大连理工大学软件学院承办的第23期CCF-CV“视界无限”系列活动——“具身视觉：基础学习理论与关键感知技术”研讨会在大连理工大学开发区校区举行。

研讨会由大连理工大学刘日升教授的主持，大连理工大学软件学院院长樊鑫教授致辞。



大连理工大学软件学院院长樊鑫教授致欢迎词，并介绍软件学院基本情况。



大连理工大学软件学院党委书记雷娜教授对各位专家到来表示感谢。



CCF-CV 专委会秘书长王瑞平研究员对研讨会的活动背景进行介绍并对大连理工大学软件学院承办此次会议表示感谢。

随后北京交通大学于剑教授、吉林大学王世刚教授、上海大学张新鹏教授、湖南大学刘敏教授、中科院计算

所王瑞平研究员、中山大学任文琦教授、山东大学丛润民教授做主题报告。



于剑教授深入分析了具身智能研究中的理论约束，特别是在知识获取、感知-行动闭环和自主决策等领域的挑战。报告结合“机器学习公理化”思想，探讨了具身智能在与物理世界交互中的理论瓶颈，并提出了构建统一认知框架和推理机制的可能路径。于剑教授强调，具身智能的发展不仅依赖数据和算法，更需要坚实的理论基础，呼吁加强基础理论的研究，以推动具身智能的进一步发展。



王世刚教授首先阐述了光场成像技术、光场环境下虚拟视点生成技术和光场环境下的儿童孤独症智能行为分析和多模态辅助诊断技术，他首先解释了光场显示的基本原理，描述光场全光函数与光场成像技术及未来发展，从复眼仿生技术到光场内容生成算法、显示系统构建、透镜阵列光学参数匹配及优化，元宇宙诞生意义，光场稀疏采集环境下任意虚拟视点生成技术及应用，光场环境下的儿童孤独症表行组学智能多模态辅助诊断技术的研究成果。



张新鹏教授分享了数字水印技术在 AIGC (人工智能生成内容) 背景下的新发展与新挑战。报告首先分析了模型确权对数据安全与知识产权的重要性，提出基于深度神经网络的水印嵌入与验证机制。张新鹏教授重点介绍了数字水印算法上的研究成果，特别是在多模态内容生成普及的今天，如何利用水印技术实现内容生成的溯源、确权与防伪。最后展示了水印在抗篡改、防伪造与责任追踪中的实际应用，并展望了其在 AIGC 监管与版权治理中的广阔前景。



刘敏教授从智能手术机器人在复杂医疗环境中的感知挑战入手，阐述了多模态感知技术的重要性，特别是如何将视觉、触觉、力觉等多种感知信息融合，提升机器人的环境感知与动态决策能力。深入讲解了多模态传感器协同、实时数据处理以及人机协同控制的技术突破，并展示了在视觉导航、组织识别、手术监控等领域的研究成果。刘敏教授指出，多模态感知能力的提升是推动手术机器人从工具型向智能型转变的关键，具有重大的临床应用潜力。



王瑞平研究员聚焦于开放世界中具身智能的持续学习问题，指出当前主流视觉模型在动态任务和环境变化下容易发生灾难性遗忘和泛化困难。报告回顾了连续学习与开放世界学习的发展历程，结合在视频理解、增量学习等领域的研究成果，介绍了在复杂场景中实现感知、认知与行为协同的技术路径。特别强调了跨时空建模、多模态协同以及自适应学习机制在具身智能中的应用，展望了该研究方向的未来研究趋势。



丛润民教授深入分析了视觉内容理解中的持续学习问题，介绍了持续学习领域的最新研究进展。他首先阐述了持续学习的研究背景及其重要意义，并通过分析当前存在的主要问题，引出团队最新研究成果。针对持续学习中的灾难性遗忘问题以及在不断“复习旧知识”时所需的巨大存储空间问题，丛教授介绍了团队提出的最新算法，这些算法能够显著提升模型性能，实现存储空间节约百倍的技术突破。最后，丛教授展望了团队未来的研究方向。

研讨会的最后是 Panel 与交流环节，与会专家与老师、同学们进行了深入交流与探讨。Panel 环节由大连理工大学助理研究员刘晋源主持，参与嘉宾包括吉林大学王世刚教授，上海大学张新鹏教授，中山大学任文琦教授，山东大学丛润民教授。

围绕“如何才能从理论基础、算法创新、硬件平台、实际应用与监管合规等多维度协同构建并推动具身感知系统的发展？”“针对硬件平台和算法创新之间的关系，是不是要对硬件平台进行定制化设计来支持算法的发展呢？是否有相关的示范或实际案例？”“在任务驱动的场景下，认知模型如何与深度视觉感知模块高效对接？”等问题各位老师结合自己的研究领域进行了深入探讨。与会的老师和同学积极提问，同各位老师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



任文琦教授探讨了基于生成先验的图像超分辨技术，指出传统的超分辨方法在细节恢复方面存在瓶颈，单纯依赖像素级重建难以实现高质量的图像复原。他提出，通过利用生成模型中的图像先验，能够显著提升超分辨模型的细节还原能力与视觉质量。报告介绍了生成对抗网络 (GAN)、扩散模型等先进技术在超分辨中的应用。介绍了几种可用于实时处理 4K 超高清视频的轻量级图像复原网络结构。分别从双边网格映射、分片并行网络、网络架构搜索等模型出发介绍了几种 4K 超高清图像视频实时复原网络结构。

中国计算机学会计算机视觉专委会（CCF-CV）

2025 年执行委员增选申请开始啦！

自 2013 年 10 月成立以来，中国计算机学会（CCF）计算机视觉专业委员会（ccfcv.ccf.org.cn）发展迅速，举办了很多有影响力的活动，如计算机视觉前沿进展研讨会（RACV）、CCF-CV 走进高校系列报告会、CCF-CV 走进企业系列交流会、CCF-CV 视界无限系列研讨会、计算机视觉前沿讲习班，与中国自动化学会模式识别与机器智能专委会、中国图象图形学学会视觉大数据专委会、中国人工智能学会模式识别专委会共同举办中国模式识别与计算机视觉大会（PRCV），定期出版专委简报，建设专委中英文网站，专委微信公众号文章平均阅读上千次，专委活动视频在专委 Bilibili 账号（<https://space.bilibili.com/611909696>）发布。搭建了全方位、高水平、大规模的计算机视觉领域交流平台。专委会成立 11 年以来，在 CCF 专委评估中获得“特色活动奖”、“综合进步奖”、“优秀专委奖”、“年度特别奖”等 7 个奖项。为了保持专委会的活力、促进国内外视觉领域人员的交流和合作，专委会现开放 2025 年计算机视觉专委会的执行委员增选工作。

一、申请时间

2025 年 6 月 1 日— 2025 年 10 月 1 日。

二、申请流程

填写申请表（点击最下方阅读原文可直接下载），发送给秘书处（ccfcv@139.com），主题“2025 新执行委员申请-姓名-单位”。（注：推荐人必须是现任专委执行

委员，名单可以从专委网站查询。电子版申请表中需填写推荐人姓名和意见，执行委员增选成功后可以补签签名）。

三、申请资格

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

四、申请需求

现任专委执行委员每人可推荐最多 3 名候选人。本次申请结果将在“2025 年中国模式识别与计算机视觉大会”（<http://www.prcv.cn>）期间（2025 年 10 月 16 日-19 日）举行的专委工作年会上投票确定（申请者届时必须“注册参会”）。

五、特别说明

按照 CCF 的新规定，CCF 专业会员通过 CCF 会员系统关注相关专委后即加入专委并成为其委员，其后每年可以更改一次关注的专委。委员在专委中无选举权和被选举权，但具有对专委的评价权。原来的专委委员自动升级为专委执行委员，享有选举权、被选举权以及对专委的评价权。

欢迎计算机视觉及相关领域的同仁加入！

责任编辑 毋立芳 任传贤

热点追踪

基于反思的反绎学习

南京大学 胡文超 戴望州

本文是南京大学LAMDA研究所的工作成果,提出的基于反思的反绎学习 (ABL-Refl) 方法已发表在AAAI 2025并荣获大会杰出论文奖 (Outstanding Paper Award) [1]。该方法旨在缓解反绎学习中的效率瓶颈。通过引入反思机制, ABL-Refl 利用领域知识快速识别并修正神经网络输出中的潜在错误, 从而提升符号推理部分的效率。传统的反绎学习方法通常依赖外部一致性优化模块进行错误修正, 这一过程涉及基于离散搜索的逻辑推理, 计算复杂且时间消耗大。ABL-Refl创新性地将反思机制内嵌至模型结构中, 通过生成反思向量来迅速识别并调用反绎来修正错误, 从而有效减少推理所需的计算负担和时间。该方法能够在多种任务中实现高效的推理修正, 尤其在处理具有复杂符号与数据混合输入的任务时, 展现出明显的优势。

本工作基于反绎学习框架展开, 在其基础上进行了具体的效率优化探索。反绎学习框架由南京大学周志华教授原创提出, 致力于实现数据与知识的深度融合, 目前已在多个方向得到广泛拓展和成功应用。此次研究成果获得国际顶级人工智能会议的高度认可, 充分体现了我国学者在数据知识融合智能学习方法构建领域的前沿探索和重要理论贡献, 相关研究已步入国际领先行列。

一、研究背景

近年来, 融合数据与知识的人工智能方法受到广泛关注, 其核心目标是将神经网络等机器学习模型与符号推理系统有机结合。具体而言, 其中机器学习致力于研究如何通过计算手段, 利用经验来改善系统自身的性能。符号推理的目标是通过一阶逻辑等人类可理解的形式语言 (formal language) 构建一种基于离散搜索的通

用推理机制。两者分别拥有各自的长处: 机器学习技术比较善于利用数据事实, 从大量有噪声的数据中学习; 符号推理技术比较善于利用规则知识, 具有高度可解释性且能够进行复杂、可靠的逻辑推理。二者的协同, 有望有效缓解纯数据驱动的人工智能方法在处理复杂任务时易出现的可靠性差等问题, 从而提升智能系统的稳定性与可信性。

虽然融合数据与知识的目标意义重大, 但以往方法往往难以实现真正的均衡融合, 它们或是在逻辑推理中引入少量机器学习的概率元素, 呈现为“推理重、学习轻” [2]; 或是在机器学习引入部分逻辑知识以辅助模型训练, 呈现为“学习重、推理轻” [3]。针对这一问题, 南京大学 LAMDA 研究所的周志华教授于 2019 年首次提出了能够均衡利用机器学习与逻辑推理的反绎学习 (Abductive Learning, ABL) 框架 [4,5]。与以往方法不同, ABL 构建了数据与知识双驱动人工智能的新框架, 借助反绎推理实现了机器学习模型与逻辑推理循环互促, 突破了传统“一头重一头轻”的局限性。ABL 已被成功应用于图像理解 [6]、法律判案分析 [7]、竞技体育指导 [8] 等多个实际任务, 并配套发布了开源工具包 [9], 展现出良好的通用性与实用价值。然而, ABL 框架的关键问题之一是如何高效地实现离散与连续变量的联合优化, 以最大化机器学习模型与逻辑推理系统之间的输出一致性 [10]。以往的方法一般需要在完整的离散变量空间中进行搜索, 因此优化过程的计算复杂度较高, 并会随着问题规模的增加而显著增长, 限制了 ABL 方法在大规模任务中的应用潜力。

针对上述挑战, 本文在 ABL 框架的基础上, 进一步提出了一种基于反思的反绎学习方法 (Abductive

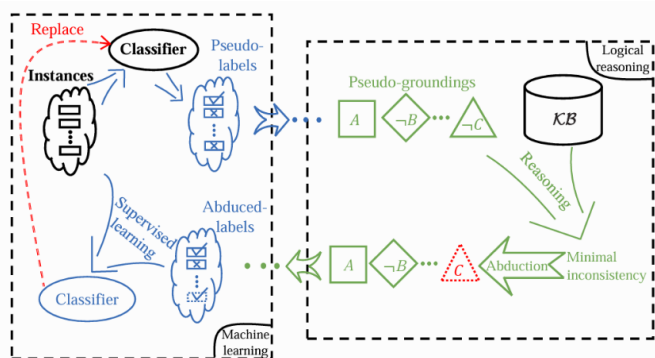


图1 反绎学习示意图

Reflection, ABL-Refl) 。ABL-Refl 的设计灵感来源于人类推理过程中常见的“反思”现象：当直觉判断与已有知识产生冲突时，人类往往能够迅速察觉到潜在的错误位置，并进一步调用深入的推理进行针对性的修正[9]。ABL-Refl 模拟了这一人类认知机制，通过引入“反思向量”，快速标记出模型预测中可能存在逻辑错误的位置，避免了传统 ABL 中计算成本高昂的离散优化过程。这一机制显著提高了模型的整体推理效率与逻辑一致性，同时进一步拓展了 ABL 方法的应用场景与适用范围。

二、反绎学习 (ABL) 框架介绍

本节简要介绍 ABL 框架的任务设定与基本流程。

反绎学习的目标是根据输入数据生成一组符号化的输出结果，使其满足领域知识库预定义的逻辑约束。ABL 系统通常包含两个核心组件：用于生成初步预测的神经网络等机器学习模型，以及提供逻辑规则与推理机制的知识库模块。知识库可以以命题逻辑、一阶逻辑、数学公式等多种形式呈现，具备可执行的符号推理能力。

反绎学习示意图如图 1 所示。ABL 的基本流程如下：学习模型首先对输入进行初步预测。由于模型可能尚未充分训练或任务本身具有复杂性，初步预测的结果往往存在一些与知识库约束相冲突的地方。ABL 框架的关键便在于引入反绎 (abduction)，即把演绎反过来嵌入到归纳中的推理过程。通过反绎，基于符号规则表征的知识库可以实现对这些潜在错误位置进行逻辑上的修正。具体而言，ABL 系统首先通过一致性优化找到预测结果中可能违反逻辑约束的部分，然后在这些位置上调用知识库进行反绎推理，通过逻辑推导与修补来确保输出满足知识约束。

然而，在实际应用中，识别不一致位置的一致性优化阶段往往成为性能瓶颈。传统 ABL 方法通常使用基于零阶优化的离散搜索，每次需遍历大量候选组合并与知识库逐一交互，故而计算成本随问题规模的增长呈指数级上升。

三、ABL-Refl 方法介绍

为解决上述问题，本文提出了一种高效的反绎学习方法——基于反思的反绎学习 (ABL-Refl) 。ABL-Refl 的核心创新在于引入反思机制，通过一个动态生成的反思向量，取代传统方法中耗时的离散一致性优化过程。

3.1 架构设计

ABL-Refl 的关键设计思想是，在神经网络模型给出初步预测结果的同时，额外引入一个反思向量的辅助机制，以快速标记出预测中可能存在逻辑错误的位置，从而有针对性地指导后续的逻辑修正过程。ABL-Refl 架构示意图如图 2 所示。具体而言，模型首先利用一个主干网络对输入数据进行处理，提取出高度抽象的语义特征表示。这些语义特征随后被送入两个并行的分支网络中进行进一步处理：其中一个分支专注于根据数据生成初步的符号预测结果 (记为 \hat{y})，而另一个分支则专门用于生成对应的反思向量 (记为 r)。反思向量与初步预测在维度上完全一致，每一个元素都可看作一个独立的二分类判断器，取值为“0”或“1”。取值为“1”表示该位置对应的初步预测可能存在与领域知识约束不符的问题，亟需进行进一步的逻辑校正；取值为“0”则意味着相应预测项相对可靠，可直接保留。接下来，系统会对反思向量中标记为“1”的位置进行置空处理，然后调用符号推理模块，对这些被置空的位置进行反绎推理，从而获得修正后的、与逻辑知识库更为一致的最终输出。

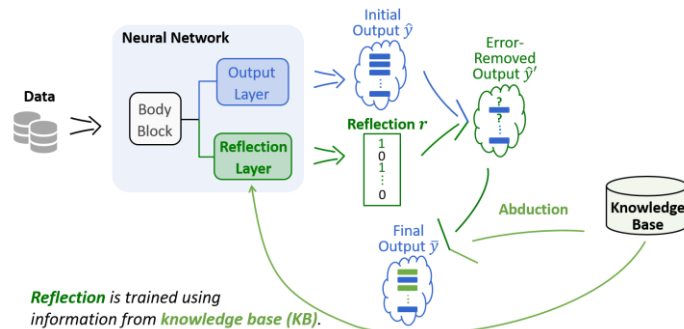


图2 ABL-Refl 架构示意图

以经典的数独任务为例示意图如图 3 所示,在 ABL-Refl 方法中,神经网络首先给出一个初步的数独预测解,其中可能存在某一行或某一宫格出现数字重复的问题,违反了数独的唯一性规则。此时,反思向量能够自动且迅速地标记出这些可能错误的位置,并将其暂时置空,然后直接调用符号逻辑模块进行局部的反绎推理修复,从而无需穷举大量位置组合,即可快速获得满足所有规则的数独解。这种方式大幅缩小了逻辑推理的搜索空间,有效提高了推理效率和最终输出的逻辑一致性。

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 8 | 0 | 4 | 0 | 0 | 1 | 2 | 0 |
| 6 | 0 | 0 | 0 | 7 | 5 | 0 | 0 | 9 |
| 0 | 0 | 0 | 6 | 0 | 1 | 0 | 7 | 8 |
| 0 | 0 | 7 | 0 | 4 | 0 | 2 | 6 | 0 |
| 0 | 0 | 1 | 0 | 5 | 0 | 9 | 3 | 0 |
| 9 | 0 | 4 | 0 | 6 | 0 | 0 | 0 | 5 |
| 0 | 7 | 0 | 3 | 0 | 0 | 0 | 1 | 2 |
| 1 | 2 | 0 | 0 | 0 | 7 | 4 | 0 | 0 |
| 0 | 4 | 9 | 2 | 0 | 6 | 0 | 0 | 7 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 8 | 5 | 4 | 3 | 9 | 1 | 2 | 6 |
| 6 | 1 | 2 | 8 | 7 | 5 | 3 | 4 | 9 |
| 4 | 9 | 4 | 6 | 2 | 1 | 5 | 7 | 8 |
| 8 | 5 | 7 | 9 | 4 | 3 | 2 | 6 | 1 |
| 2 | 6 | 1 | 7 | 5 | 8 | 9 | 3 | 4 |
| 9 | 3 | 4 | 1 | 5 | 2 | 7 | 8 | 5 |
| 5 | 7 | 8 | 3 | 9 | 4 | 6 | 1 | 2 |
| 1 | 5 | 6 | 5 | 8 | 7 | 4 | 6 | 3 |
| 3 | 4 | 9 | 2 | 1 | 6 | 8 | 5 | 7 |

a. 数独问题

b. 神经网络初步预测

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 7 | 8 | 5 | 4 | 3 | 9 | 1 | 2 | 6 |
| 6 | 1 | 2 | 8 | 7 | 5 | 3 | 4 | 9 |
| 9 | 4 | 6 | 2 | 1 | 5 | 7 | 8 | |
| 8 | 5 | 7 | 9 | 4 | 3 | 2 | 6 | 1 |
| 2 | 6 | 1 | 7 | 5 | 8 | 9 | 3 | 4 |
| 9 | 3 | 4 | 1 | 5 | 2 | 7 | 8 | 5 |
| 5 | 7 | 8 | 3 | 9 | 4 | 6 | 1 | 2 |
| 1 | 5 | 6 | 5 | 8 | 7 | 4 | 6 | 3 |
| 3 | 4 | 9 | 2 | 1 | 6 | 8 | 5 | 7 |

c. 反思向量识别的错误项

图 3 数独求解任务举例

与以往的 ABL 方法相比,ABL-Refl 将传统计算复杂度较高的外部一致性优化模块替换成了上述高效的反思机制,避免了大量的搜索与交互过程。基于此,ABL-Refl 在接收到数据时仅需一次性标记出可能存在的问题的位置并进行局部修复,从而大幅降低了符号推理的计算复杂性。得益于这一创新设计,ABL-Refl 在训练和推理效率上均实现了显著提升,且逻辑一致性与稳定性也得到了有效保证,更适合处理大规模、复杂的现实任务。

3.2 训练机制

ABL-Refl 的训练目标是直接利用知识库反馈,指导模型生成反思向量,以提升输出的一致性。每次模型前

向传播时,神经网络同步给出初步预测结果 \hat{y} 和反思向量 r 。根据反思向量,模型将预测结果中标记为可能错误的位置置空,从而生成中间结果 \hat{y}' 。然后,系统调用知识库对置空位置进行逻辑修复,得到最终输出 \bar{y} 。

在训练时,有了 \hat{y} 和 \hat{y}' ,我们可以分别比较他们与领域知识库 KB 的一致性,从而定义两个一致性指标 $\text{Con}(\hat{y}', \text{KB})$ 和 $\text{Con}(\hat{y}, \text{KB})$ 。通过衡量置空修复前后的一致性差异,我们得到一致性指标提升:

$$\Delta\text{Con}_r(\hat{y}) = \text{Con}(\hat{y}', \text{KB}) - \text{Con}(\hat{y}, \text{KB})$$

这一提升便可以用于度量置空操作是否有效地引导了更一致的逻辑输出:越高的提升值,可以代表反思向量能够更好地识别出神经网络初始预测中的错误。我们将该提升作为奖励信号,通过 REINFORCE 算法^[12]优化模型参数,使其逐步学会生成高质量的反思向量。优化目标如下:

$$L_{\text{con}}(x) = -\Delta\text{Con}_r(\hat{y}) \cdot \nabla_{\theta} \log f_{\theta}(\hat{y}, r | x)$$

其中, θ 为神经网络的参数。这一训练过程无需依赖标签信息,仅通过知识库反馈引导模型学习如何更高效地协同感知与逻辑推理。

此外,为了控制推理成本,ABL-Refl 还引入了一个反思规模的正则项,鼓励模型仅在必要时触发反绎推理。该部分损失定义如下:

$$L_{\text{size}}(x) = \Phi \left(c - \frac{1}{n} \sum_{i=1}^n (1 - r_i) \right),$$

其中, $\Phi(a) = \max(0, a)^2$ 。

此外,在有标签数据的情况下,也可结合交叉熵等监督损失 $L_{\text{labeled}}(x, y)$ 。ABL-Refl 的最终损失函数由监督损失、一致性损失与反思规模控制损失构成,注意到后两项损失函数与反思向量的训练直接相关,而它们无需使用数据标签的信息,可适用于半监督训练环境。

四、实验

为验证 ABL-Refl 方法的通用性与有效性,本文在三个具有代表性的任务场景中进行了系统性的评估。这些任务涵盖了符号推理、感知推理以及结构化推理场景,分别为符号数独 (Symbolic Sudoku)、图像数独 (Visual Sudoku) 和最大团识别 (Maximum Clique),

广泛用于数据与知识双驱动人工智能方法的能力测试。

在符号数独任务中，模型需根据数独规则（行、列与九宫格中的唯一性约束）填充一个部分缺失的 9×9 数字矩阵。我们选择了以下代表性方法作为基线进行比较：

- **Recurrent Relational Network (RRN)**^[13]: 纯数据驱动的神经网络模型，不显式引入规则知识。
- **CL-STE**^[14]: 通过将逻辑规则转换为损失项，以正则化的方式实现符号知识的约束。
- **SATNet**^[15]: 使用可微 SAT 求解器，使得逻辑推理过程能够嵌入到神经网络中。

实验结果表明，ABL-Refl 在推理准确性、训练效率和收敛速度上均明显优于这些基线方法。这主要归因于反思机制有效地缩小了符号推理的搜索空间，显著提升了模型的逻辑一致性与稳定性。

在图像数独任务中，输入数据为 81 张手写数字图像，模型需首先识别图像中对应的数字，再进行逻辑补全。这一任务融合了感知与符号推理，体现了数据与知识双驱动方法的典型应用场景。实验中，ABL-Refl 表现出更高的推理质量与一致性修复能力，与 SATNet、CNN+Solver 等代表方法相比，ABL-Refl 显示出了更强的推理准确率与泛化能力。

此外，我们还评估了 ABL-Refl 在最大团识别任务中的表现。该任务是一个组合优化问题，输入为图的邻接矩阵，目标是找到图中规模最大的完全子图。此类问题具有 NP 复杂性，对模型的图结构理解与组合推理能力提出了较高的要求。实验结果表明，ABL-Refl 能够有效适应不同图结构与密度的变化，无需专门的图结构假设即可稳定地获得高质量的推理结果，体现出反思机制在复杂高维任务中的优势与通用性。

参考文献

- [1] Wen-Chao Hu, Wang-Zhou Dai, Yuan Jiang, and Zhi-Hua Zhou. "Efficient rectification of neuro-symbolic reasoning inconsistencies by abductive reflection." In Proceedings of the AAAI Conference on Artificial Intelligence, pp. 17333-17341, 2025.
- [2] Luc De Raedt, and Angelika Kimmig. "Probabilistic (logic) programming concepts." Machine Learning 100: 5-47, 2015.
- [3] Jingyi Xu, Zilu Zhang, Tal Friedman, Yitao Liang, and Guy Broeck. "A semantic loss function for deep learning with

五、总结与未来展望

本文基于南京大学周志华教授提出的反绎学习框架，实现了一种高效的数据与知识融合方法——基于反思的反绎学习 (ABL-Refl)，有效缓解了 ABL 框架中连续离散变量联合优化的效率瓶颈。该方法通过引入反思机制，自动利用领域知识生成反思向量，实时识别数据驱动模型输出中的潜在逻辑错误，并通过逻辑反绎推理进行针对性修正。ABL-Refl 的反思机制实质上构成了一种逻辑意义上的注意力机制，有效地缩小了逻辑推理的搜索空间，使推理过程更加高效和准确。在多个代表性的推理任务中，ABL-Refl 均表现出明显的性能提升，证实了该方法在提高逻辑一致性、推理效率与模型泛化能力方面的显著优势。

ABL 框架通过有效融合数据驱动的机器学习与知识驱动的逻辑推理，可显著提升人工智能系统的推理能力和决策质量，并有望在多种高复杂性场景中实现关键突破。例如在医疗、法律、教育等同时具有丰富数据与明确领域知识的场景中，数据驱动与知识驱动的融合能够有效弥补单一方法的不足，提高系统的性能、稳定性与可解释性；在工业控制、国防等安全性要求极高的场景中，传统的数据驱动方法难以满足严格的规则约束与安全性保障，而 ABL 通过引入符号推理与知识约束，能够在保证系统灵活性的同时确保系统的可靠性与可信性。此外，随着大型语言模型的快速发展，如何有效识别与修正其输出中的知识错误与逻辑冲突问题日益受到关注。ABL-Refl 提出的反思式推理框架，将为构建更加稳健、可信的通用人工智能系统提供重要的技术路径，有望推动下一代人工智能向更具通用性与可靠性的方向进一步迈进。

责任编辑 魏秀参

- [4] Zhi-Hua Zhou. "Abductive learning: towards bridging machine learning and logical reasoning." *Science China. Information Sciences*, 62(7): 76101, 2019.
- [5] Zhi-Hua Zhou, and Yu-Xuan Huang. "Abductive learning." In *Neuro-Symbolic Artificial Intelligence: The State of the Art*, pp. 353-369. IOS Press, 2021.
- [6] En-Hao Gao, Yu-Xuan Huang, Wen-Chao Hu, Xin-Hao Zhu, and Wang-Zhou Dai. "Knowledge-enhanced historical document segmentation and recognition." In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 8409-8416, 2024.
- [7] Yu-Xuan Huang, Wang-Zhou Dai, Jian Yang, Le-Wen Cai, Shaofen Cheng, Ruizhang Huang, Yu-Feng Li, and Zhi-Hua Zhou. "Semi-supervised abductive learning and its application to theft judicial sentencing." In *2020 IEEE international conference on data mining*, pp. 1070-1075, 2020.
- [8] Jiachen Wang, Dazhen Deng, Xiao Xie, Xinhuan Shu, Yu-Xuan Huang, Le-Wen Cai, Hui Zhang, Min-Ling Zhang, Zhi-Hua Zhou, and Yingcai Wu. "Tac-valuer: Knowledge-based stroke evaluation in table tennis." In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pp. 3688-3696, 2021.
- [9] Yu-Xuan Hu, Wen-Chao Hu, En-Hao Gao, and Yuan Jiang. "ABLkit: a Python toolkit for abductive learning." *Frontiers of Computer Science*, 18(6):186354. 2024.
- [10] Wang-Zhou Dai, Qiuling Xu, Yang Yu, and Zhi-Hua Zhou. "Bridging machine learning and logical reasoning by abductive learning." *Advances in Neural Information Processing Systems* 32, 2019.
- [11] Shane Frederick. "Cognitive reflection and decision making." *Journal of Economic perspectives* 19, no. 4: 25-42, 2005.
- [12] Ronald J Williams. "Simple statistical gradient-following algorithms for connectionist reinforcement learning." *Machine learning* 8: 229-256, 1992.
- [13] Rasmus Palm, Ulrich Paquet, and Ole Winther. "Recurrent relational networks." *Advances in neural information processing systems* 31, 2018.
- [14] Zhun Yang, Joohyung Lee, and Chiyoun Park. "Injecting logical constraints into neural networks via straight-through estimators." In *International Conference on Machine Learning*, pp. 25096-25122, 2022.



胡文超

南京大学人工智能学院 2021 级博士研究生，导师为姜远教授，主要研究方向为反绎学习。
Email: huwc@lamda.nju.edu.cn



戴望州

戴望州，南京大学苏州校区智能科学与技术学院准聘副教授，博导，获国家级人才计划（青年）项目资助。2019 年在南京大学取得博士学位。2019 年 4 月-2022 年 7 月在伦敦帝国理工学院从事博士后研究工作，2022 年 8 月加入南京大学苏州校区。研究方向为人工智能与机器学习，目前研究兴趣是数据与知识双驱动的机器学习，代表性工作为融合逻辑推理与机器学习的反绎学习框架及算法。在人工智能与机器学习领域共发表多篇学术论文，获 AAAI-2025 杰出论文奖，担任第 4 届国际学习与推理联合会（IJCLR）程序主席。
Email: daiwz@nju.edu.cn

专题综述

生成式遥感图像处理

西安交通大学 庞立 唐大涛 曹相湧

本文主要介绍西安交通大学计算机科学与技术学院曹相湧副教授课题组在生成式遥感图像处理领域的系列研究成果，包括发表在Information Fusion 2024的PLRDiff^[1]，CVPR 2024的HIR-Diff^[2]，Information Fusion 2025的LatentHSI^[3]，TGRS 2024的CRS-Diff^[4]，ArXiv 2024的HSIGene^[5]和CVPR 2025的AeroGen^[6]，下面将围绕以上工作的技术细节和创新点展开讨论。

一、研究背景

遥感图像处理作为现代地球观测系统的核心技术，已成为推动国民经济发展和保障社会民生的重要技术支撑，被广泛应用于农业监测、城市规划、灾害应急、生态保护等关键领域。近年来，基于深度学习的遥感图像处理方法利用数据驱动的方式，能够从给定图像训练对中学会处理特定遥感任务的规律。然而，基于深度学习的遥感图像处理方法在实际应用中仍然面临巨大挑战，比如有标签遥感影像数据缺乏、模型泛化能力不足等关键难题。为了解决这些难题，本文探讨如何利用生成式模型来赋能遥感图像处理方法，分别从无监督模型设计和稀缺数据扩充两个角度来进行研究，并在遥感底层图像处理以及遥感高层图像处理两方面来展开探索，旨在抛砖引玉，助力遥感图像处理方法的实际应用。

二、无监督生成式遥感图像复原

2.1 背景介绍

高光谱成像在图像采集过程中常会遭受各种类型的退化和压缩，严重影响了下游任务的性能^[7]。因此，高光谱图像复原技术对于高光谱应用至关重要，在过去数十年中吸引了大量研究。现有的高光谱图像复原方法

主要分为基于模型的方法^[8]和基于深度学习的方法^[9]。基于模型的方法将复原问题转化为优化问题，依赖手工先验约束解空间，虽泛化能力强但先验设计主观，难以精准建模复杂图像特征，性能受限。深度学习方法通过大量清晰-退化图像对进行模型训练，性能优越但存在泛化能力弱的问题。因此本文引入基于预训练扩散模型的无监督高光谱图像复原框架，旨在融合传统模型的泛化能力与扩散模型的强大特征表达能力，实现高效、高质量的图像复原。

2.2 研究内容

研究内容一：无监督高光谱全色锐化

本研究提出了一种无监督的高光谱全色锐化方法PLRDiff^[1]，该方法创新性地将图像的低秩分解思想与无条件扩散生成模型相结合。此方法旨在规避监督学习对大规模成对数据的依赖，同时克服传统方法中手工先验表达能力不足的缺陷。其核心思想是将待融合的高分辨率多光谱图像（HRMS）分解为一个低秩的基底张量和一个系数矩阵的乘积，再通过分别估计这两个分量来完成融合任务。该方法首次实现贝叶斯方法和扩散先验相结合的方式来进行全色锐化，整体流程如图1所示。

具体而言，首先对于基底张量的定义，传统方法通常将奇异值分解方法SVD或主成分分解方法PCA等低秩分解后得到的基向量作为基底张量进行求解，然而该定义下的基底张量不具备真实图像的视觉特征和空间结构，这使得利用强大的图像生成模型对其进行建模得十分困难。为了解决这一根本性问题，不同于传统的主成分分析等方法，本研究提出从HRMS影像自身选取若

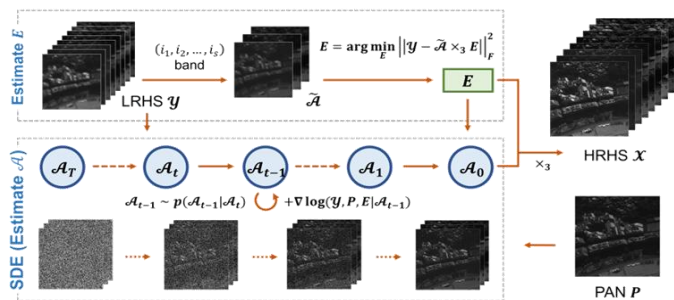


图1 无监督高光谱全色锐化方法 PLRDiff 的总体框架

干线性无关的波段来直接构成基底张量，如此定义的基底张量本身就处于图像域，其空间特征与真实遥感影像高度相似，这为利用预训练的遥感图像扩散模型进行结构生成和先验建模提供了理论基础。此外，研究发现，在常见的线性退化模型下，低分辨率高光谱/多光谱影像（LRMS）与其对应的高分辨率高光谱/多光谱影像共享同一个系数矩阵。基于这一关键观察，本研究提出了一种简单而有效的系数矩阵预估方法：首先从已知的LRMS影像中选取部分波段构成其自身的基底张量，进而通过求解一个最小二乘问题来反解出能够保留光谱信息的系数矩阵。在系数矩阵被固定后，基底张量的复原问题便转化为一个条件生成任务。本研究将该任务构建于一个逆向随机微分方程（SDE）框架内，利用一个预训练的、无需任何监督信息的遥感图像扩散模型，以观测到的LRMS、全色影像（PAN）以及预估的系数矩阵为条件，逐步从随机噪声中采样生成高真实度的基底张量。最终，通过将生成出的基底张量与预估的系数矩阵相乘，即可重构出最终的高分辨率高光谱/多光谱。该无监督框架不仅利用了扩散模型强大的图像分布学习能力，还通过低秩分解有效降低了生成模型的处理维度，使其能够灵活地应用于不同波段数量的数据集。

研究内容二：无监督高光谱图像复原

本研究核心研究内容在于构建一个名为HIR-Diff^[2]的无监督高光谱图像重建统一框架，它巧妙地将高光谱图像固有的低秩特性与预训练扩散模型的强大先验相结合。该框架的核心思想是将复杂的高维高光谱图像复原问题，分解为两个关联但更易于求解的子问题：一个低维“降维图像”（reduced image）的生成和一个“系数矩阵”（coefficient matrix）的估计。最终，清

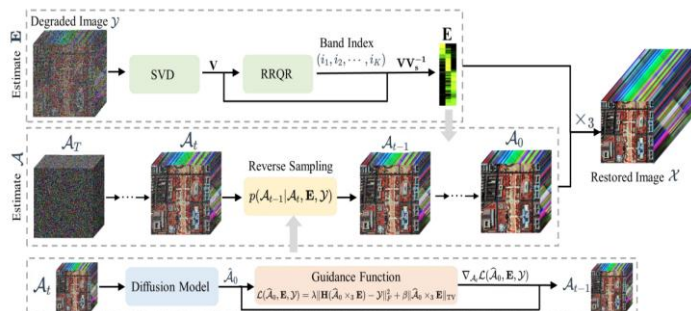


图2 无监督高光谱图像复原方法 HIR-Diff 整体框架

晰的高光谱图像通过这两个分量的乘积重构得到，其整体框架如图2所示。

首先，本研究深入研究了如何实现对高光谱图像的有效低秩分解与系数矩阵的鲁棒估计。鉴于高光谱图像在采集过程中不可避免地会引入噪声，直接对观测图像进行分解容易导致误差累积。为此，本文提出了一种联合利用奇异值分解（SVD）和显秩QR分解（Rank-Revealing QR, RRQR）的策略。SVD作为一种经典的降维技术，其分解出的右奇异向量对于噪声具有天然的鲁棒性，能够稳定地表征高光谱数据的主要光谱基底。然而，如何从这些基底中选择一个最优子集来重构系数矩阵是一个关键问题。本研究创新性地引入RRQR算法来解决此问题。RRQR能够从SVD分解出的右奇异向量矩阵中，自动筛选出一组线性无关性最强的基向量，这确保了后续系数矩阵求解过程的数值稳定性，有效避免了因波段间高度相关而导致的病态问题，从而实现对系数矩阵的精确预估。

其次，本研究提出了利用扩散模型生成高质量的降维图像的具体方法。通过上述分解，高光谱图像的主要信息被压缩到一个光谱维度极低的降维图像中。这个降维图像在空间结构上与普通RGB遥感图像非常相似，因此可以直接利用在海量遥感图像上预训练好的无条件扩散模型进行生成，从而将大规模数据集学到的丰富结构先验迁移至复原任务中。为了将观测信息融入生成过程，本文设计了一种新的引导函数。该函数不仅包含一个数据保真项，用于确保复原结果与观测到的退化图像保持一致，还引入了一个总变分（TV）正则项，约束图像空间的局部平滑性，在抑制噪声的同时，精细地复原地物的边缘和纹理细节。

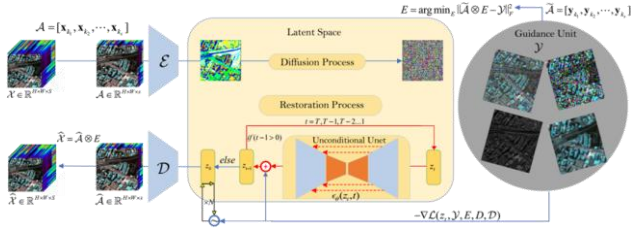


图3 基于隐空间的无监督高光谱图像复原方法 LatentHSI 整体框架

总之，本研究提出了一个通用的高光谱无监督复原框架，该方法成功地融合了低秩先验与扩散模型的强大建模能力，为高光谱图像复原提供了新的有效解决方案。

研究内容三：隐空间高光谱图像复原

本研究构建了一个名为 LatentHSI^[3]的基于隐空间的无监督高光谱图像复原框架，该框架旨在通过在一个全新构建的、数据驱动的隐空间 (Latent Space) 中实现特征复原，以克服现有复原方法在高维数据处理上的局限性，其整体框架如图3所示。

该方法首先利用了高光谱图像在光谱维度上固有的低秩特性。具体而言，该方法将高光谱图像分解为一个降维表征 (reduced representation) 和一个系数矩阵的乘积。其中，降维表征通过从原始高光谱数据中选取s个等间隔的波段构成，而系数矩阵则通过求解最小二乘问题从退化图像中预先估计得出。这一步骤将不同波段数的原始高光谱图像统一转换成了具有固定通道数的标准化数据形式。

与以往方法直接在像素空间或手工设计的子空间中进行复原不同，LatentHSI引入了变分自编码器 (VAE)，为上述的降维表征构建了一个全新的、维度更低、结构更优的潜在空间。这个通过VAE学习到的潜在空间具有良好的连续性与平滑性，其先验分布被约束为标准高斯分布，这意味着该空间结构简单、清晰，能够有效避免像素空间中因数据分布离散、复杂而导致的优化难题。

在此基础上，该研究在潜在空间中训练了一个无条件的扩散模型，以精确学习数据分布。在执行复原任务时，该过程被构建为一个条件采样问题。具体地，模型使用 DDIM^[12] 采样器在潜在空间中进行迭代去噪，并通过一个为复原任务定制的引导函数 (guidance

function) 对每一步的采样进行修正。该引导函数包含一个数据保真项，以确保复原结果与观测到的退化图像一致；同时包含一个总变分 (TV) 正则项，用于抑制噪声和保护图像的边缘细节。最后，为了修正引导采样过程中可能引入的分布偏移和误差累积，本文在扩散采样结束后，额外增加了一个校正步骤。该步骤使用 Adam优化器对最终得到的潜在向量进行精细微调，使其更精确地满足引导函数所定义的约束。经过优化的潜在向量通过VAE的解码器重构为清晰的降维表征，再与预先估计的系数矩阵相乘，最终得到高质量的复原高光谱图像。

2.3 实验分析与结果

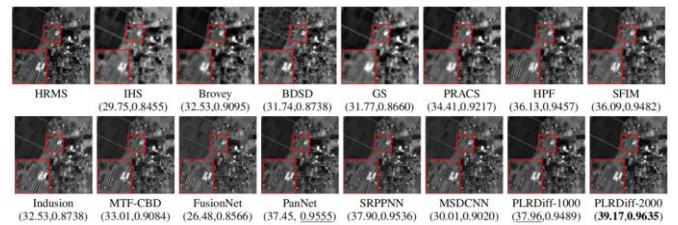


图4 Chikusei 数据集第 43 个波段可视化结果

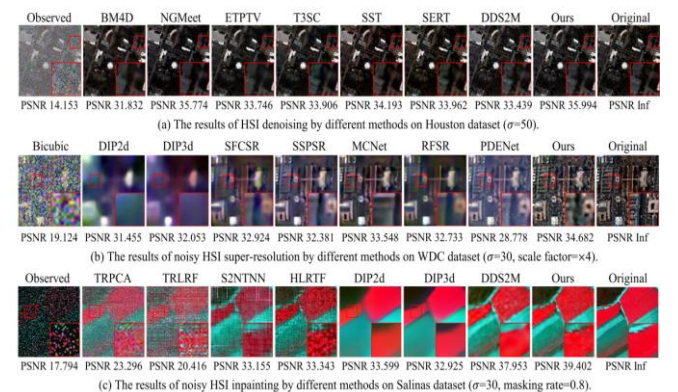


图5 多种复原任务 HIR-Diff 与其他方法可视化结果

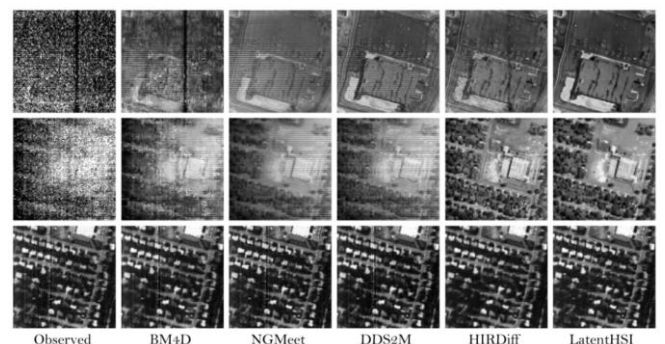


图6 Urban 数据集复原结果可视化

详细的实验设置、定量或定性实验可参考相关方法原文，下面仅对关键实验结果进行展示。

对比实验：如图4所示，相比于现有的SOTA方法，PLRDiff可以实现更为准确自然的高/多光谱图像全色锐化。从图5中可以看出HIR-Diff能够在包括去噪、超分、填充等多个任务上性能超过现有方法。图6则展示了LatentHSI能够在真实场景高光谱复原获得最好的复原效果。

三、生成式遥感大模型

3.1 研究背景

遥感图像与自然图像相比具有独特的特征，包括更高的分辨率、更广的覆盖范围以及更丰富的信息内容。然而，获取高质量的标注遥感数据非常困难且成本高昂，这严重限制了遥感图像领域深度学习模型的训练和应用。近年来，随着扩散模型 (Diffusion Models) [10]等生成模型的发展，为解决遥感数据稀缺问题提供了新的可能。

传统的遥感图像生成方法主要依赖于基本的数据增强技术 (如旋转、缩放等)，但这些方法难以产生足够多样性的样本。而现有的生成模型虽然能够生成高质量图像，但大多专注于通用图像生成，对遥感图像的特殊需求考虑不足，尤其是在精确控制生成内容方面存在局限性。针对这些挑战，曹相湧副教授团队开发了三个创新的生成模型框架：CRS-Diff, AeroGen和HSIGene，分别解决光学遥感图像和高光谱图像的多条件可控生成、目标检测数据增强问题。

3.2 研究内容

研究内容一：多条件可控扩散模型 CRS-Diff

CRS-Diff^[4]是一个专为遥感图像生成设计的多条件可控扩散模型框架。与仅依赖文本条件的现有方法不同，CRS-Diff同时支持三种控制条件：文本条件、元数据条件和图像条件，从而实现更精确的生成控制。这些条件及其对应生成图像如图7所示。

CRS-Diff的核心创新在于其多条件控制机制，这是首次在遥感图像生成领域同时支持文本、元数据和图像三种控制条件的框架。该机制通过设计特定的条件编码器，将不同类型的控制信息转换为模型可理解的特征表示，然后通过条件注入模块将这些特征引入扩散过程，实现对生成图像的精确控制。

在模型架构上，CRS-Diff采用了多尺度特征融合策略，有效整合了不同层级的特征信息，增强了控制条件的引导效果。具体来说，如图8所示，模型在U-Net backbone的不同层次设置了特征融合模块，将文本、元数据和图像条件的特征与扩散模型的中间特征进行融合，确保控制信息能够在不同尺度上影响生成过程。CRS-Diff还具备强大的数据引擎功能，可以为各种下游任务生成高质量的训练数据。例如，在道路提取任务中，CRS-Diff能够根据指定条件生成包含复杂道路结构的遥感图像，显著提升道路提取模型的性能，特别是在原始数据集中罕见或缺失的道路场景上。

通过严格控制生成过程中的条件参数，CRS-Diff能

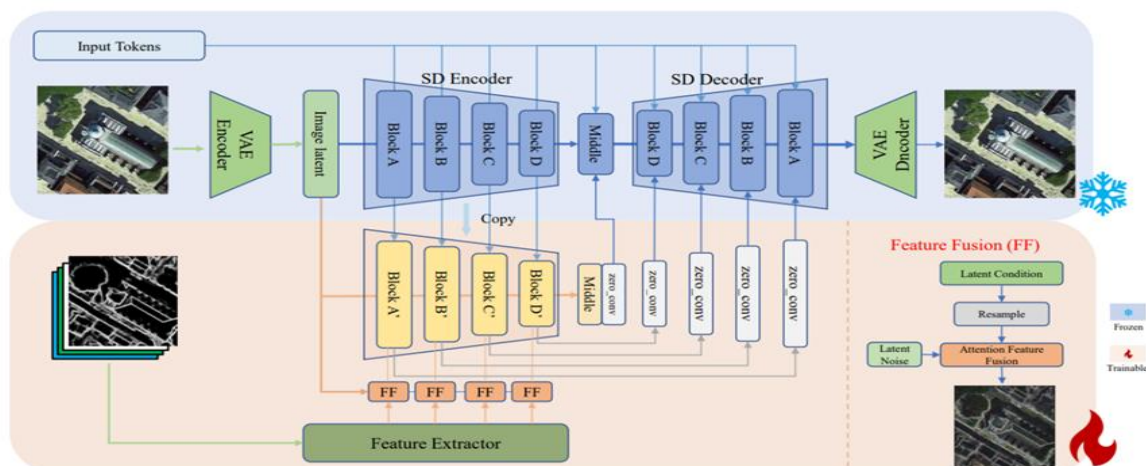


图7 CRS-Diff 模型结构

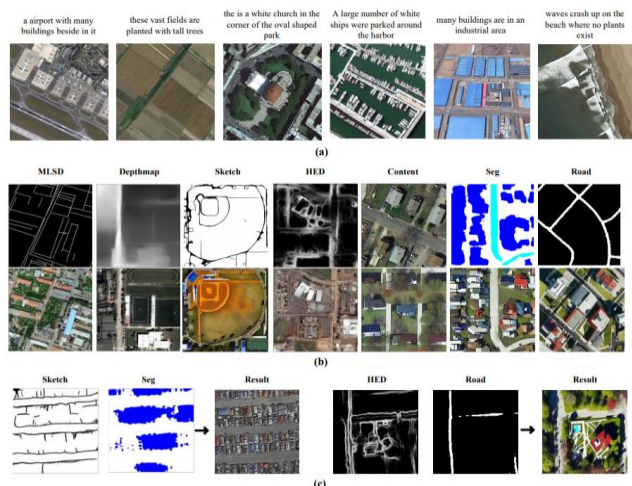


图 8 CRS-Diff 生成结果可视化

够生成符合特定要求的遥感图像，如指定地理位置、特定地物分布或特定时间的景观特征，这为遥感应用领域提供了前所未有的灵活性和可能性。

研究内容二：高光谱生成式大模型HSIGene

本研究的核心研究内容在于构建了一个名为HSIGene的高光谱图像（HSI）生成式基础模型。该模型旨在解决高光谱图像因采集成本高昂而导致的数据稀缺问题，进而提升各类下游任务的性能。HSIGene基于潜在扩散模型（Latent Diffusion Models, LDMs）^[11]构建，并具备强大的多条件可控生成能力，能够根据用户提供的多种控制信号（如草图、分割图、边缘图等）精确、可靠地合成高光谱图像。

为有效缓解训练数据不足对模型生成能力和多样性的限制，本研究提出了一种新颖的数据增强策略。该策略的核心思想是对现有的真实高光谱图像进行空间超分辨率处理，然后从放大后的高分辨率HSI中裁剪出大量的训练样本块。这种方法能够在保持原始光谱保真度的同时，显著增加训练数据的空间内容多样性。为了进一步提升增强数据的感知质量，本文设计了一个两阶段高光谱图像超分辨率框架。第一阶段，利用大量与真实HSI地理位置和内容相似的高分辨率RGB遥感影像，训练一个基于扩散模型的RGB超分辨率网络（DSRNet），用于提升HSI中RGB波段的空间分辨率。第二阶段，基于已获得的高分辨率RGB波段作为精确的引导信息，本文提出了一种新颖的矩形引导注意力网络（Rectangular

Guided Attention Network, RGAN)，用于实现高质量的引导式高光谱图像超分辨率重建。RGAN网络通过自注意力层（SAL）、矩形交叉注意力模块（RCA）、光谱注意力层（SpecAL）和前馈网络（FFD）等组件，能够高效地将RGB模态的精细细节迁移至高光谱模态。

HSIGene的另一项关键研究内容是其多条件可控生成机制。该模型支持包括整体嵌套边缘检测（HED）图、图像分割掩码（Segmentation）、手绘草图（Sketch）、多尺度线段检测（MLSD）结果、图像内容特征（Content）以及文本描述（Text）在内的多达六种控制条件输入。这种多模态、多角度的控制能力，使得模型能够根据具体需求生成具有特定结构、内容或语义的高光谱图像，并支持多种条件的组合使用以实现更精准的控制。HSIGene不仅能够合成视觉逼真且光谱可靠的高光谱图像，并且这些合成数据能够显著提升下游高光谱图像处理任务的性能和泛化能力，充分证明了HSIGene作为数据增强工具的巨大潜力。

研究内容三：遥感目标检测任务导向的生成式模型

AeroGen是一个专为遥感图像目标检测（RSIOD）任务设计的布局可控扩散生成模型。针对RSIOD数据集中标记数据稀缺和类别不平衡的问题，AeroGen提供了一种创新的数据增强解决方案。

AeroGen的最大亮点在于它是首个同时支持水平和旋转边界框条件生成的模型。遥感图像中的目标通常呈现多角度分布，传统的仅支持水平边界框的生成模型难以满足实际需求。AeroGen通过设计专门的布局条件编码器，能够处理任意方向的目标布局信息，生成符合预设布局要求的高质量图像，生成结果如图9所示。

在框架设计上，AeroGen提出了一个端到端的数据

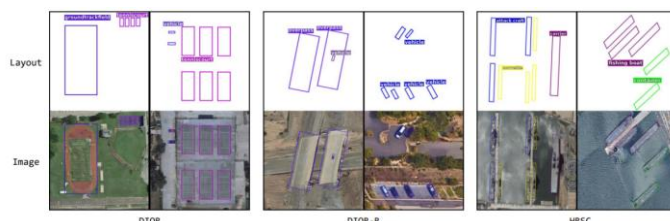


图 9 AeroGen 布局合成效果

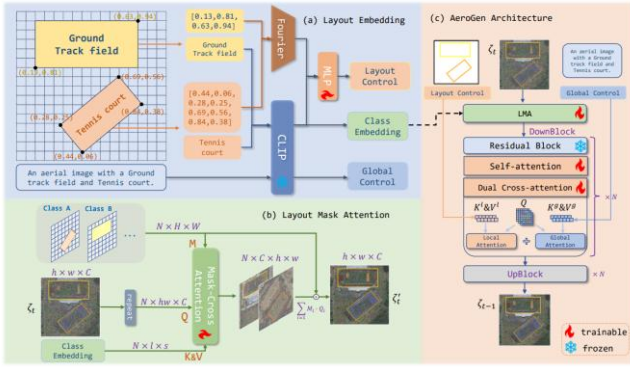


图 10 AeroGen 模型架构

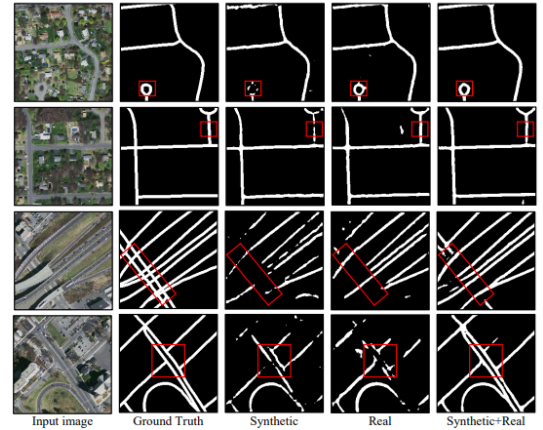


图 12 CRS-Diff 数据增强示例

增强架构,如图10和图11所示,包括多样性条件生成器和质量过滤机制。多样性条件生成器通过随机组合不同类别和数量的目标布局,创建多样化的生成条件,确保合成数据的丰富性。质量过滤机制则通过设计的评分函数,评估生成图像的质量和与条件的一致性,筛选出高质量的合成样本用于下游任务训练。AeroGen特别关注稀有目标类别的生成问题。针对数据集中样本数量稀少的类别,AeroGen采用类别平衡采样策略,在生成过程中提高这些类别出现的频率,有效缓解了类别不平衡问题。实验表明,这种策略显著提升了目标检测模型在罕见类别上的性能。

在实际应用中,AeroGen不仅能生成单一目标的图像,还能处理复杂的多目标组合场景,模拟真实世界中目标的复杂分布和相互关系,进一步提高了生成数据的实用性和价值。

3.3 实验分析与结果

详细的实验设置、定量或定性实验可参考相关方法原论文,下面仅对关键实验结果进行展示。

3.3.1 CRS-Diff

CRS-Diff在多个实验场景下进行了全面评估,包括单条件控制、多条件协同控制以及下游任务应用等方面。

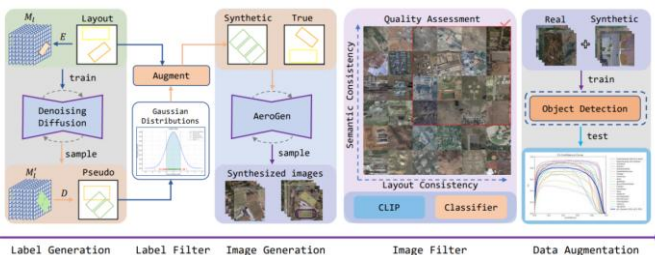


图 11 AeroGen 的数据增强管道

在生成质量评估中,CRS-Diff与现有最先进的方法进行了对比。在RSICD数据集上,CRS-Diff的FID指标达到了50.72,比基线模型提升了12.7%,表明CRS-Diff生成的遥感图像具有更高的真实性和多样性。在控制精度方面,实验结果显示CRS-Diff能够准确响应不同类型的控制条件。在文本条件控制下,模型能够根据复杂的文本描述生成符合要求的遥感场景,如"一个被森林环绕的小型机场"或"城市边缘的工业区与农田交界处"。元数据条件控制实验表明,模型能够根据地理坐标、高度、季节等元数据信息生成对应特征的遥感图像。在图像条件控制下,CRS-Diff能够保持输入图像的结构和布局特征,同时根据其他条件要求调整内容细节,结构相似性指标SSIM达到0.87。

在下游任务应用实验中,研究团队利用CRS-Diff生成的数据作为额外训练集,对道路提取模型进行训练。结果如图12所示,与仅使用原始数据集相比,加入CRS-Diff生成数据后,道路提取模型的IoU指标从52.84提升至55.27,尤其在复杂道路结构和罕见道路类型上的提升更为显著。这一结果有力证明了CRS-Diff作为数据引擎的实用价值。

3.3.2 HSIgene

可视化结果:HSIgene生成结果如图13所示,能够在多种引导条件下生成可靠的高光谱影像。

对下游任务性能提升:如图14所示,HSIgene的光谱数据可以提升多个下游任务的性能。

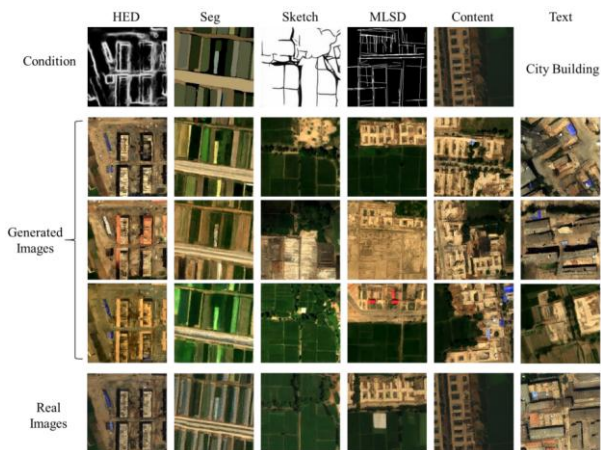


图 13 HSiGene 生成图像可视化

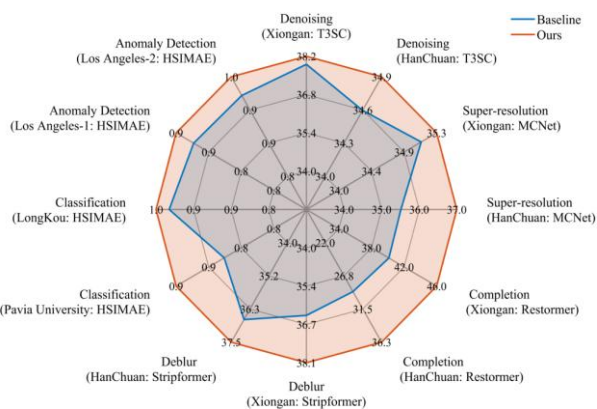


图 14 HSiGene 对多个下游任务提升性能比较

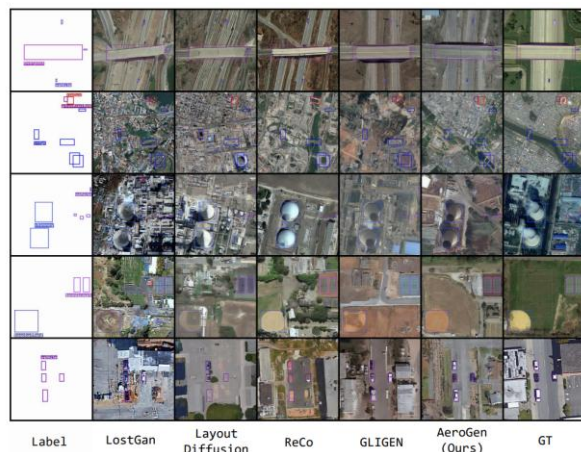


图 15 AeroGen 数据合成效果展示

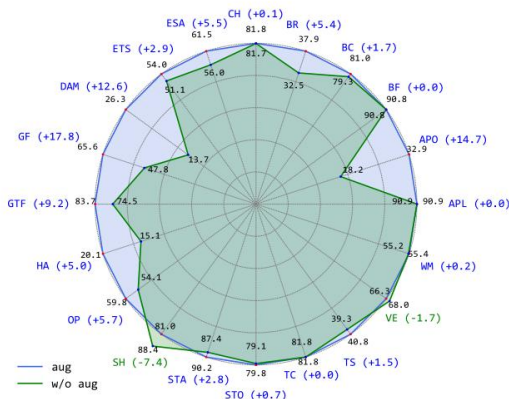


图 16 AeroGen 数据增强效果展示

明 AeroGen 在解决遥感检测中类别不平衡问题方面具有独特优势。如图 16 所示，论文通过在多个结构不同的数据集 (DIOR、DIOR-R、HRSC) 上验证合成数据提升下游性能，间接证明其良好的迁移适应能力，为遥感图像检测中的小样本和稀缺数据问题提供了通用性解决方案。

四、总结

本文围绕生成式遥感图像处理展开研究，在遥感图像底层视觉研究领域，提出了首个基于扩散模型的无监督遥感图像复原系列方法 (PLRDiff, HIR-Diff, LatentHSI)，能够高效快速的重建各种退化的遥感图像且不依赖于任何训练数据对。在遥感图像高层视觉任务领域围绕“图像生成-面向任务生成”的遥感图像生成思路，提出了遥感图像生成的系列大模型 (CRS-Diff, HSiGene, AeroGen)，能够快速有效地生成各种类型的遥感图像，用以扩充道路检测、语义分割、目标检测等

3.3.3 AeroGen

AeroGen 的实验评估主要集中在生成质量和目标检测性能提升两个方面，在 DIOR、DIOR-R 和 HRSC 三个主流遥感目标检测数据集上进行了全面验证。AeroGen 在多个指标上显著优于现有方法。在 DIOR 数据集上，AeroGen 在 FID、分类准确性得分 (CAS) 和 YOLO 布局一致性评分方面均为最佳。

AeroGen 生成的合成数据被用于目标检测训练。结果表明，在 DIOR 数据集上加入 50k 合成样本后，检测精度不断提升。生成结果如图 15 所示，这些增幅表明 AeroGen 生成的样本具备极高的实用价值，在遥感检测中能有效替代或补充真实数据。

针对罕见类别的性能提升尤为显著。在 DIOR-R 数据集中，样本相对较少的类别如 GF (地面跑道)、DAM (水坝) 和 APO (机场) 分别提升了 17.8%、12.6%、14.7% 的 mAP50，远高于大多数常见类别的增幅。这说

遥感高层视觉任务地训练数据集，旨在缓解特殊场景下小样本的难题。大量充分的实验证明了上述所提系列方法的有效性。该系列成果已发表于中科院 1 区顶级期刊

Information Fusion 和 TGRS，以及计算机视觉国际顶级会议 CVPR 2024 和 CVPR 2025。

责任编辑 张青

参考文献

- [1] X. Rui, X. Cao, Z. Zhu, Z. Yue, and D. Meng. Unsupervised pansharpening via low-rank diffusion model. Information Fusion, 2024.
- [2] L. Pang, X. Rui, L. Cui, H. Wang, D. Meng, and X. Cao. HIR-Diff: Unsupervised hyperspectral image restoration via improved diffusion models. In CVPR 2024.
- [3] J. Cao, X. Rui, L. Pang, D. Meng, and X. Cao. LatentHSI: Restore hyperspectral images in a latent space. Information Fusion, 2025.
- [4] D. Tang, X. Cao, X. Hou, Z. Jiang, J. Liu, and D. Meng. Crs-diff: Controllable remote sensing image generation with diffusion model. TGRS, 2024.
- [5] L. Pang, X. Cao, D. Tang, S. Xu, X. Bai, F. Zhou, and D. Meng. Hsigene: A foundation model for hyperspectral image generation. arXiv, 2024.
- [6] D. Tang, X. Cao, X. Wu, J. Li, J. Yao, X. Bai, D. Jiang, Y. Li, and D. Meng. AeroGen: Enhancing remote sensing object detection with diffusion-driven data generation. In CVPR 2025.
- [7] W. He, Q. Yao, C. Li, N. Yokoya, and Q. Zhao. Non-local meets global: An integrated paradigm for hyperspectral denoising. In CVPR 2019.
- [8] Q. Yuan, L. Zhang, and H. Shen. Hyperspectral image denoising employing a spectral-spatial adaptive total variation model. TGRS, 2012.
- [9] Q. Shi, X. Tang, T. Yang, R. Liu, and L. Zhang. Hyperspectral image denoising using a 3-D attention denoising network. TGRS, 2021.
- [10] J. Ho, A. Jain, and P. Abbeel. Denoising diffusion probabilistic models. In NeurIPS, 2020.
- [11] R. Rombach, A. Blattmann, D. Lorenz, P. Esser, and B. Ommer. High-resolution image synthesis with latent diffusion models. In CVPR 2022.
- [12] J. Song, C. Meng, and S. Ermon. Denoising diffusion implicit models. arXiv, 2020.



庞立

西安交通大学计算机学院硕士研究生，导师为曹相湧副教授，主要研究方向为图像复原与图像生成。

Email: pp2373886592@gmail.com



唐大涛

西安交通大学计算机学院硕士研究生，导师为曹相湧副教授，主要研究方向为遥感图像解译与遥感图像可控生成。

Email: tdt200107@stu.xjtu.edu.cn



曹相湧

西安交通大学计算机学院副教授，博士生导师。主要研究方向为遥感图像解译、底层图像处理、生成式模型，发表包括 IJCV、TIP、TGRS、CVPR、ICCV 等在内的论文 60 余篇，ESI 高被引论文 7 篇，先后主持国家自然科学基金面上项目、青年项目、华为云、中国电信、教育部考试院等项目 5 项，获国际会议竞赛前三名 5 项。

Email: caoxiangyong@mail.xjtu.edu.cn

顶会观察

ICLR 2025

中国科学院计算技术研究所 徐逸峰 王中琦 何振梁

国际表征学习大会 (International Conference on Learning Representations, ICLR) 是机器学习领域的顶级会议之一，每年举办一次。ICLR 2025于4月24日至4月28日在新加坡举办，这也是该会议首次在亚洲举办。

一、会议概况

ICLR 2025 共收到 11,500 份投稿，最终接收论文 3,705 篇，整体录用率为 32%。其中，213 篇论文被评为 Oral，380 篇被评为 Spotlight。本届会议的论文主题高度集中于大语言模型 (LLMs)。以论文的第一个关键词进行统计，在排名前十的关键词中，有一半与大模型相关，包括“Large Language Models”及其各种大小写、缩写变体。此外，强化学习主题在前十中占据两个位置。前十关键词中还包括联邦学习、图神经网络以及扩散模型等热门方向。ICLR 2025 共组织了 40 个 Workshops，涵盖人工智能多个前沿领域。主会议的所有入选论文，包括获得口头报告机会的论文，均以墙报的形式展示，以便与会者深入交流讨论。



图1 ICLR 2025 主会场一隅

二、获奖论文

本次会议共选出了 3 篇杰出论文 (Outstanding Papers) [1][2][3]、3 篇荣誉提名 (Honorable Mentions) [4][5][6]、1 篇时间检验奖 (Test of Time Awards) [7]和 1 篇时间检验候选 (Runner Up) [8]。下面简要介绍 3 篇杰出论文和 1 篇时间检验奖。

2.1 杰出论文 1: Safety Alignment Should be Made More Than Just a Few Tokens Deep^[1]

此文作者来自于普林斯顿大学和谷歌，主要内容为研究大语言模型 (LLM) 的对齐缺陷发现与修复。近年来，大语言模型的安全性主要依赖于安全对齐方法，涉及有监督微调 (SFT) 与基于偏好的优化方法 (如 RLHF) 等。然而，最近的研究表明这类对齐方法的脆弱性。对齐模型在对抗性输入或有限步数微调后仍会输出不安全内容。在本文中，作者深入分析了造成当前对齐脆弱性的原因，发现当前安全对齐仅作用于回答中的前几个 token，这种“安全对齐捷径”被作者称为大语言模型的“浅层安全对齐 (Shallow Safety Alignment)”。具体的，作者计算了对齐模型和未对齐模型中，不同 token 长度处二者的特征 KL 散度，作者发现 KL 散度在前几个 token 中显著高于其他 token，证明安全对齐主要作用于前几个 token (如拒绝前缀 “I cannot...”)。此外，作者还通过实验证明了这种浅层对齐是许多脆弱性的来源：在推理阶段中，对齐模型可能会遭受前缀攻击、基于优化的后缀攻击和纯随机采样的越狱攻击。作者表明如果模型仅通过引导一个简短的“拒绝响应”的前缀 token 来屏蔽有害内容模型，攻击者可以轻易替换或优化肯定性前缀 (如 “Sure, here is...”) 完成

越狱攻击。在下游微调阶段中，作者也通过实验证明，基于微调的攻击主要扰动生成分布的前几个 token。为了实现“深层安全对齐 (Deep Safety Alignment)”，作者设计了一种数据增强方法，其核心思想是训练模型在“开始被恶意诱导”的情况下依旧能做出拒绝回应。实验表明，数据增强的方法使模型得到了更深的对齐结果，并且原始性能也得到保留。此外，作者为了增强模型抗微调攻击，提出了一种约束性微调优化目标，通过限制初始 token 的分布变化，使安全对齐在下游微调时保持得更为持久。总结而言，该论文系统性验证了 LLM 对齐中“浅层安全对齐”现象，从根本上分析了 LLM 安全对齐的薄弱之处，并提出加强对齐深度的策略，对防御大模型越狱具有重要意义。

2.2 杰出论文 2: Learning Dynamics of LLM Finetuning^[2]

此文作者来自不列颠哥伦比亚大学，主要内容为研究微调大语言模型 (LLM) 时模型的学习动态。近年来，大语言模型的强大表现使其成为自然语言处理研究的焦点。为了使这些模型更好地遵循人类指令、符合人类偏好，微调成为关键步骤，通常包括两个阶段：指令微调和偏好微调。尽管已有大量研究尝试从训练目标、最终模型表现或与强化学习的关系角度来理解微调机制，但此论文提出了一种全新的视角——从学习动态出发，对 LLM 微调过程进行深入解析。学习动态描述的是：在通过梯度下降更新模型参数的过程中，训练样本如何影响模型对其他样本的预测。这种动态视角不仅揭示了训练过程中一些有趣现象，如“Z 字形”学习路径和组合性概念空间的形成，也为设计更高效的训练算法提供了理论基础。在此文中，作者通过将模型预测变化分解为三个不同的组成部分，建立了一个通用的学习动态框架。该框架可适用于多种主流微调算法，包括：有监督微调、直接偏好优化、强化学习方法（如 PPO）。这一框架不仅统一了对不同微调算法训练过程的理解，还解释了若干在实践中观测到的、看似违反直觉的现象。例如：复读机现象，即偏好微调后，模型倾向于重复简单、模板化的短语；幻觉增强，即模型可能在回答一个问题时引用另一个问题的错误信息或无关事实；输出置信度下降，即在离线策略 DPO 训练过程中，几乎所有候选回答的

置信度都降低。为了进一步解释这些现象，作者提出了一个关键机制——压缩效应。该效应源于交叉熵损失函数与 softmax 层结合下的梯度上升。具体来说，当模型对一个不太可能的标签进行负向梯度更新时，这种更新会压缩其对所有标签的预测概率，仅将概率推向少数最可能的标签上。这种压缩可能导致原本合理的输出也变得置信度较低，从而损害模型与人类偏好的一致性。该发现也解释了为何在线策略 DPO 通常优于离线策略方法。作者通过深入的梯度分析，揭示了不同优化策略在动态层面上的根本差异。在理论分析的基础上，此文还提出了一种简单而有效的对齐提升方法，尽管这种方法在直觉上可能与常规策略相反，却在实验证明中显示出显著提升的效果。总结而言，此论文通过构建统一的学习动态框架，为理解大语言模型的微调过程提供了全新视角。该框架不仅帮助解释多个实证现象，还揭示了不同微调方法的本质差异，并进一步启发出实用的优化策略。对研究者和工程师而言，这项工作不仅拓展了理论边界，也为提升模型性能和对齐质量指明了方向。

2.3 杰出论文 3: AlphaEdit: Null-Space Constrained Knowledge Editing for Language Models^[3]

此文作者来自于中国科学技术大学和新加坡国立大学，主要研究内容为面向大型语言模型 (LLM) 的知识编辑问题。近年来，为解决 LLM 中出现的幻觉问题，许多工作从知识编辑角度进行实验。目前主流的编辑方式是“先定位-再编辑”，然而有研究表明，现有编辑方法会不可避免地会破坏模型中原有知识。作者认为这是由于编辑方法导致模型隐空间表示发生分布偏移。为了解决这个问题，作者引入了零空间 (Null Space) 的概念。零空间指的是那些不会改变原始知识表现的模型参数扰动方向所构成的空间。具体来说，作者先识别出用于保持原有知识的重要表示，然后将编辑过程中的参数扰动严格投影到这一零空间中，从而实现在编辑目标知识的同时，保护原有知识不被破坏。特别地，由于这一思想十分简洁，仅需在现有的大多数模型编辑方法中增加一行代码，即可将参数扰动投影到零空间，从而显著提升编辑后的性能与原有知识的保留效果。在实验中，作者表明该工作在编辑有效性和泛化性能上要优于基线 12.54% 以上。此外，在连续编辑任务中，当编辑 3000

个样本规模时，编辑模型仍能保持原始性能。总体而言，该工作开创性地引入以零空间投影为核心的语言模型编辑新范式，为实现可持续、可靠的大模型知识更新提供了强有力的方法论支持。

2.4 时间检验奖: Adam: A Method for Stochastic Optimization^[7]

Adam 是由 Kingma 和 Ba 提出的一种用于优化随机目标函数的一阶梯度优化算法。它能够自适应地估计梯度的一阶和二阶矩来调整每个参数的学习率，从而提升优化效率与稳定性。该方法实现简单、计算高效，能够很好地处理目标函数非平稳、梯度稀疏或噪声较大的情况，而且通常无需复杂的调参，因此特别适合用于大规模数据或参数量庞大的问题。如今，Adam 已成为深度学习领域应用最广泛的优化算法之一，彻底改变了神经网络训练方式，其在实际应用中的卓越表现使之成为从计算机视觉、自然语言处理到强化学习等众多领域最先进模型的首选优化器，展现了卓越的通用性。在颁奖仪式上，两位作者用生动有趣的语言分享了开发 Adam 优化器背后不为人知的故事，讲述了论文先被 ICLR 拒绝、到写邮件申诉、最终被接收的曲折经历。这启示我们，有价值的研究成果可能不会立即获得认可，但随着时间的推移，真正优秀的工作最终能够经受时间的考验，被学界发现并广泛应用。

三、大会演讲

3.1 Zico Kolter: Building Safe and Robust AI Systems

演讲者 Jeremy Zico Kolter 是卡内基梅隆大学的教授，并担任该校机器学习系主任。他的研究主要聚焦于模型优化与鲁棒性。在此次演讲中，演讲者系统回顾了过去十年他和他的团队在神经网络优化、可验证对抗鲁棒性、深度学习实践以及 AI 安全等方向的阶段性研究进展。演讲者针对各个研究阶段都分享了自己选择该研究方向的“动机”以及“方法”。其中，在关于 AI 安全的部分，演讲者提出了几个颇具启发性的观点：1. 当前被广泛关注的安全问题，可能会随着更强大模型的出现而被自然解决；2. 我们可能只能识别安全问题，却无法真正修复它们；3. AI 安全研究可能是当前学术界对人工智能发展影响最大的切入点。在演讲的最后，演讲者

针对不同阶段的研究人员提出了具体建议：1. 对于学生和初级研究者：不必担心你的研究路径看起来不够“战略性”。研究的真正进展往往来源于偶然的交流、对问题的好奇心以及自然的探索过程；2. 对更资深的教职员工/研究人员来说，请慷慨支持那些尝试走出常规路线、探索非传统方向的教职员工。

3.2 Song-Chun Zhu: Framework, Prototype, Definition and Benchmark

演讲者朱松纯教授是北京通用人工智能研究院院长、北京大学讲席教授、清华大学基础科学讲席教授。此演讲系统介绍了北京通用人工智能研究院与北京大学在通用人工智能 (AGI) 方面的研究成果。有趣的是，朱教授指出汉字“通”不仅有通用人工智能的含义，也正好包含了 A、G、I 三个英文字母，因此也将通用人工智能称作“TongAI”。演讲者首先展示了一个数字智能体原型——一个名为“通通”的生活在高度真实物理模拟环境中的虚拟小女孩，她能够在多物理属性和社会互动交织的情境中持续学习和成长。通通具备自我驱动的价值系统，拥有欲望与目标，能够据此生成计划并采取行动。她所体现的智能体行为不是外部指令控制的产物，而是基于内在动机和价值的自主行为。围绕这一智能体原型，演讲者进一步阐述了其背后的理论框架，即由三个核心组成部分构成的系统：认知架构 (C)、潜能函数 (U) 和价值函数 (V)。各种 AGI 系统可以被视为三维空间 (C,U,V) 中的一个点。这个框架代表了一种范式转变：从以大数据驱动小任务的统计范式转向以价值驱动的小数据、大任务的范式。演讲还介绍了团队提出的 TongTest——一个用于评估通用智能的全新测试标准与平台。TongTest 不仅远超图灵测试的复杂性，还融合了心理学与人类学的理论成果，评估智能体在复杂多模态、具身任务中的表现。目前，TongTest 认为通通的智能水平可类比人类 3 至 4 岁的儿童。此外，演讲还展示了一些近期在类人机器人及其实际应用方面的研究进展，并探讨了东方哲学视角下关于人与智能的本质，以及道德与社会规范如何在 CUV 框架下自然涌现，从而为 AGI 的安全性提供理论基础。总体而言，此次演讲不仅提出了一个具备理论完备性与实践原型的 AGI 研究框架，也结合认知科学、哲学与技术系统展示了一

种跨学科的、面向长期通用智能目标的研究路径。

3.3 Yi Ma: Pursuing the Nature of Intelligence

演讲者马毅教授是香港大学穆斯克特基金会数据科学研究所 (HKU IDS) 及计算机科学系的讲席教授, 其研究方向涵盖计算机视觉、高维数据分析与智能系统。在此次演讲中, 马教授从历史、科学、数学和计算等多个维度系统阐述了智能的多层次机制与本质。演讲伊始, 马教授回顾了智能在自然界中的演化历程——从系统发育、个体发育再到社会层面的演进, 以及人工智能的发展轨迹。在此基础上, 他提出了智能学习的三个核心问题, 并逐一给出解答: 1. “学什么?” : 智能应从外部世界中学习那些具有可预测性的数据。所有可预测的数据都可被编码于分布 $p(x)$ 中, 而该分布实际存在于一个低维的“子空间”里。尽管我们所观察到的数据是高维的, 但其中有用和有规律的信息往往隐藏在一个更低维度的空间中。因此, 学习应聚焦于最终的低维表征。2. “如何学?” 演讲者强调, 学习低维结构的一个基本且统一的机制是: 通过压缩来降低观察到的 (噪声) 数据分布的熵。3. “如何判断正确性?” 演讲者指出, 低维结构学习的正确性可以通过双向的“编解码”过程来验证——即既能压缩信息, 又能解码还原。这种双向验证机制为学习结果的可靠性提供了坚实基础。在探讨当前智能发展的基础上, 演讲者展望了“下一代智能”的发展方向, 即自动智能 (Autonomous Intelligence)。他强调, 我们应思考如何构建能够持续且自动学习连续表征的人工智能系统。对此, 自然智能提供了丰富的启发。例如, 视觉皮质已具备某种形式的离散编码机制, 这为人工系统的构建提供了生物学依据。在演讲结尾, 演讲者提出了关于智能的核心观点: “智能的核心在于如何编码和改进信息, 以更好地预测世界。” 总体而言, 该演讲提出了一个基于“压缩编码—解码”原理的数学框架, 作为理解与解释智能本质的理论基础。

3.4 Dawn Song: Towards Building Safe and Secure AI: Lessons and Open Challenges

演讲者 Dawn Song 教授是加州大学伯克利分校“负责任去中心化智能研究中心 (RDI)”的教员联合主任, 同时也是伯克利人工智能研究实验室 (BAIR Lab)

的一员。她的研究工作主要聚焦于深度学习、安全性以及区块链技术。在此次演讲中, Dawn Song 教授深入探讨了构建与部署人工智能和大语言模型智能体所面临的多重风险, 并分享了相应的缓解策略。她重点介绍了团队在隐私保护、鲁棒性、防止幻觉生成和公平性等方面的研究进展, 强调必须采取全生命周期的安全防护机制。具体的, 在前沿科技探索阶段, 需研究能够从根本上提升模型安全性的技术范式; 在模型对齐阶段, 应加强 AI 系统的防护能力, 以有效抵御多种攻击方式; 在模型评估阶段, 需深化对模型能力和行为的理解来确保它们以可信的方式来运作。在模型部署阶段, 必须保障系统按预期安全、稳定地运作。演讲者还指出, 前沿技术的迅猛发展, 正在促使网络攻击数量显著上升。在防御实践方面, 演讲者分享了她们在使用深度学习检测物联网脆弱性、以及利用大型语言模型进行零日漏洞识别方面的最新研究成果。最后, 演讲者指出要确保一个安全的 AI 未来, 我们必须采取一种社会技术协同 (sociotechnical) 的方法。演讲者还分享了其最近提出的一项基于科学与证据的 AI 政策建议, 重点阐述三个关键优先事项: 加深对 AI 风险的理解、制定有效的缓解措施, 以及推动制定更加健全的 AI 政策。

3.5 Danqi Chen: Training Language Models in Academia: Challenge or Calling?

演讲者陈丹琦副教授是普林斯顿大学计算机学院副教授, 领导普林斯顿自然语言处理组, 近期研究方向为训练、适配并理解语言模型, 尤其注重如何让学术界更容易地训练和使用这些模型。此演讲的主题为如何在学术界训练语言模型。随着语言模型的发展, 训练超大规模模型已成为机器学习的重要方向。然而, 由于这一领域依赖大规模计算资源、专有数据和工程基础设施, 因此几乎完全由工业界领导。相比之下, 学术界面临着显著的劣势, 包括计算资源有限、缺乏基础设施支持, 以及对高质量训练数据的访问受到限制。这些劣势使得学术界难以直接训练大语言模型。尽管如此, 演讲者强调, 学术界在语言模型训练中仍具备不可替代的价值。通过从训练过程本身出发, 学术研究可以帮助我们更深入地理解模型行为, 并在资源受限的条件下开发出更高效、可解释的算法与系统。演讲者分享了在过去两年中,

其实验室在有限预算下进行的预训练与后训练工作。这些研究关注于训练动态的分析、模型性能与效率之间的权衡，以及资源受限条件下的技术创新，同时将所得模型与工具以开放形式贡献给社区。此外，演讲者还指出了三个值得学术界持续投入的研究方向：1. 开发小而强的模型，在模型尺寸受限的条件下实现良好性能；2. 理解与改进训练数据，探索数据质量对模型训练效果的影响；3. 推进基于开源模型的后训练方法，包括对齐、指令微调和偏好优化等技术。总体而言，该演讲不仅呈现了在学术资源限制下进行语言模型训练的现实路径，也强调了学术界在方法创新、理论探索与社区共享方面的独特贡献。演讲呼吁学术界继续参与这一领域的研究，并探讨与工业界协作的可能形式，以促进语言模型技术的持续发展与应用。

3.6 Tim Rocktaeschel: Open-Endedness, World Models, and the Automation of Innovation

演讲者 Tim Rocktaeschel 教授是 Google DeepMind 的首席科学家，也是伦敦大学学院计算机系人工智能中心的教授，工作主要重点在于通用人工智能、开放性和自我改进智能。此演讲探讨了通往人工超级智能 (ASI) 的一条关键路径——从传统的目标导向优化转向“开放性进化” (Open-Endedness) 这一研究范式。“开放性进化”由 Stanley、Lehman 与 Clune 等学者提出，其核心理念是构建能够持续产出新颖且可学习成果的系统，而非仅仅追求单一性能指标的最优化。在这一框架下，系统不断自我生成和探索新的任务与解决方案，模拟出更接近自然智能的创新过程。演讲者重点介绍了其团队在大规模基础世界模型方面的研究。这类模型能够生成丰富多样的环境，为训练更具泛化能力

和鲁棒性的智能体提供支持。这些世界模型不仅扩展了可供学习的任务空间，也为推动更广义人工智能奠定基础。进一步地，演讲者提出一个重要论点：开放性进化与基础模型的结合，可能是实现自动化创新的关键机制。这一融合已在实践中展现出初步成果，例如自动提示工程、自动红队测试、大语言模型中的 AI 辩论等。这些自我迭代、自我挑战的能力，标志着人工智能系统正在逐步具备推动自身进步与创新的潜力。总体而言，这场演讲不仅展示了基础模型和开放性研究的新进展，也提出了一个深远的愿景：未来的人工智能或将具备自我生成目标、自我改进能力，成为推动创新本身的主体力量。这一方向为构建更具创造力与自主性的智能系统提供了理论基础与实践路径，值得广泛关注。

四、总结展望

此次 ICLR 会议上，“AI 安全”是一个备受关注的热点话题。事实上，杰出论文与大会演讲中 AI 安全话题占据 1/3，这足以说明此次 ICLR 对 AI 安全的重视。的确，随着大模型技术的迅速发展，如何有效管控 AI 系统、防范潜在风险，已成为技术进步所带来的关键社会议题之一。从数据收集的可靠性、到训练过程中的偏见控制、再到部署阶段的安全防护，AI 安全问题贯穿了整个技术链条。在当前监管框架尚未成熟的背景下，学术界的前瞻性探索显得尤为重要。ICLR 作为顶级学术会议，其重视安全议题不仅体现了研究界的高度敏感性，也预示着 AI 发展趋势正在从“能力竞赛”向“可信落地”转变。未来，技术与治理的深度融合，将成为 AI 安全发展的必经之路。

责任编辑 张杰

参考文献

- [1] Qi X, Panda A, Lyu K, et al. Safety alignment should be made more than just a few tokens deep [C]/ICLR. 2025.
- [2] Ren Y, Sutherland D. Learning Dynamics of LLM Finetuning [C]/ICLR. 2025.
- [3] Fang J, Jiang H, Wang K, et al. AlphaEdit: Null-Space Constrained Knowledge Editing for Language Models [C]/ICLR. 2025.
- [4] Wang J, Mittal P, Song D, et al. Data Shapley in One Training Run [C]/ICLR. 2025.
- [5] Ravi N, Gabeur V, Hu Y, et al. SAM 2: Segment Anything in Images and Videos [C]/ICLR. 2025.
- [6] Narasimhan H, Jitkrittum W, Rawat A, et al. Faster Cascades via Speculative Decoding [C]/ICLR. 2025.

[7] Kingma D, Ba J. Adam: A Method for Stochastic Optimization [C]//ICLR. 2015.

[8] Bahdanau D, Cho K, Bengio Y. Neural Machine Translation by Jointly Learning to Align and Translate [C]//ICLR. 2015.



徐逸峰

中国科学院计算技术研究所硕士研究生，导师为山世光研究员，主要研究方向为视觉内容生成。
Email: yifeng.xu@vipl.ict.ac.cn



王中琦

中国科学院计算技术研究所硕士研究生，导师为张杰副研究员，主要研究方向为人工智能安全。
Email: wangzhongqi23s@ict.ac.cn



何振梁

中国科学院计算技术研究所特别研究助理。主要研究方向为计算机视觉、生成模型、视觉的生成与理解协同。谷歌学术引用 1400 余次，AttGAN 工作曾入选 ESI 高被引论文。曾获得 2024 年中国图象图形学学会优秀博士论文提名、2020 年中科院计算所所长特别奖。

Email: hezhenliang@ict.ac.cn

新疆大学库尔班·吾布力教授访谈

2025年5月26日,《CCF-CV专委简报》在线采访了新疆大学博士生导师库尔班·吾布力教授。下面是采访实录。

库尔班老师,您好!首先,请您分享一下您的个人学习和研究经历。

1992年以优异成绩考入新疆大学少数民族科技班,在新疆大学完成预科阶段学习后,1994-1997年在电子科技大学大学就读通信工程专业,完成学业后回新疆大学信息科学与工程学院任教。2006年考入本院信息与通信工程专业攻读硕士学位,2009年顺利毕业获得硕士学位。2011-2012年获得国家留学基金委的资助到加拿大卡尔顿大学访问学习。2013年9月考入本院计算机应用专业攻读博士学位。博士阶段,以签名识别与鉴别方向作为课题,主要研究包括模式识别、计算机视觉和生物特征识别,于2018年获得博士学位,长期从事维吾尔文手写签名识别与鉴别、维吾尔文文档检索和多文种文档识别方面的研究和开发工作。

您在计算机视觉与自然语言理解,尤其是文本分类等领域内已有所建树,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?

这些年来,我们边探索边实践,针对特定领域进行了一些尝试性研究,在集体协作下收获了阶段性成果。

我们团队在计算机视觉和自然语言理解领域,主要围绕文档分析开展相关研究,由于文档类型的复杂性和多样性,涉及了多项研究内容,例如多文种文本检测与识别、签名识别与鉴别、文档图像检索、自然语言理解等。2022年我们的研究成果《多文种离线手写签名识别与鉴别关键技术及应用》通过了中国图象图形学学会组织的权威鉴定,被认定为整体技术达到国际先进水平。2025年我们的研究成果《多模态手写签名识别与鉴别关键技术及应用》通过中国计算机学会组织的成果鉴定、评价委员会一致认为该项成果整体技术达到国际先进水平,其中基于视频的空中签名鉴别技术、视听多模态签名识别与鉴别技术达到国际领先水平。目前主持了新疆维吾尔自治区“天山英才”科技创新领军人才计划项目《基于深度语义的多文种复杂文档识别与理解技术及应用》,结合新疆特色资源与先进技术解决复杂文档中存在的诸多科学问题。上述工作已在国内外期刊和会议上发表了100余篇论文,并获得了10余项授权专利。建议研究者们始终以解决实际问题为导向,保持对最新论文的持续追踪,在工程实现中注重算法效率与可解释性,既要追求学术创新也要重视技术落地,并用扎实的工作与批判性思维推动领域发展,为社会创造价值。

您现在是贵校的二级教授和博士生导师,也是新疆维吾尔自治区“天山英才”科技创新领军人才,有丰富的科研经历且取得了很多科研荣誉,能否跟大家分享一下您的成长历程,以及您的成长感悟?

我的成长过程是漫长的。1997 年留校任教时，本人本科学历、助教职称，也就是起步很低，并且学校职称评审制度中也没有破格评审的机制。因此从助教走到目前的教授（二级）付出了 28 年努力。特别是，助教、讲师阶段教学任务很重、每周至少完成 16 学时的教学任务。作为一名高校教师，还得从事科学研究，我的成长历程经历了从潜心基础研究到服务国家需求的转变过程。在工作期间，始终坚持以国家战略需求和新疆地区信息技术的发展为导向，带领团队围绕文档分析和理解、目标跟踪与识别、生物特征识别和三维重建等应用方向开展深入研究。在自治区党委和政府的大力支持下，我们攻克了签名识别与鉴别等多项关键技术，研发的系统已成功应用于新疆公安和司法领域，相关成果获得中国图象图形学学会科技进步奖等荣誉。作为“天山英才”科技创新领军人才，我深感责任重大，始终把培养青年科技人才作为重要使命。这些成绩的取得，要归功于党和国家对西部科技工作者的关心支持，也得益于团队成员的共同努力。我深刻体会到，只有将个人研究方向与国家发展需求紧密结合，科研工作才能真正创造价值。未来，我们将继续面向新疆经济社会发展主战场，为推动边疆地区科技进步和产业升级贡献智慧和力量。

您发表了 300 多篇高水平学术论文，能跟大家分享一下您是如何做到持续产出高水平论文的么？您在学术影响力方面做了哪些努力呢？

我认为持续产出高质量论文的关键在于团队一起努力。首先聚焦人工智能、机器学习与多文种文档分析与理解等方向，选择兼具理论价值与应用潜力的课题，积极争取国家和自治区各项项目保障研究经费，确保研究可持续性。注重学科交叉（如数学+计算机+其他科学），开辟创新空间，避免同质化竞争。然后开展问题驱动型研究，从实际科学难题（如新疆多文种文档分析与理解、多语言 NLP）出发，提炼理论问题，增强论文价值。最后，团队踏踏实实做科研、论文撰写方面采用迭代式写作方式，初稿-反馈-修订等环节循环，采取团队成员互评以及外部专家评审结合的方式，严控论文质

量。在学术影响力方面，作为 CCF 计算机视觉和人工智能专委执行委员，积极参加国内外领域的学术会议、加入 CCF、CAA、CSIG、CAAI 等学会人工智能领域的专委会、向优秀的专家学者学习。我积极推动学术资源向边疆靠拢，在新疆组织“CCF-CV/AI 走进高校”等系列活动，稳步推进新疆 NOI 事宜，并成功在乌鲁木齐承办多场全国性学术会议，如 PRCV2024、CCFAI2023 等。这些平台有效促进了东西部学者的思想碰撞，既为本地人才培养注入前沿学术活力，又通过产学研对话加速技术创新落地。

您获得了中国图象图形学学会科技进步二等奖、自治区自然科学优秀论文二等奖等奖项 20 余项。请问能否详细谈谈这些获奖背后的故事，以及您是如何通过不懈努力与创新思维，在这些科研领域取得这些显著成就的？

本人在多文种（特别是汉、英、维、哈、柯等）文档分析与理解领域取得了一些成绩，获得中国图象图形学学会科技进步二等奖、自治区自然科学优秀论文二等奖等奖项 20 余项。科研成就源于我们对国家、自治区以及边疆需求的深刻洞察、跨学科融合的创新，以及在艰苦（极端环境）下的长期坚守。

例如签名识别与鉴别研究，其中每一个环节都凝聚着团队的心血与坚持。我们从 2011 就开始做这项研究，并获得了两项国家自然科学基金以及自治区级项目的资金支持。为此，我们自主研发了改进版在线签名采集系统，以 120Hz 高采样率手写板精准捕捉书写动态，构建了国内首个少数民族多文种在线签名数据库和多文种（汉、维、哈、柯）离线签名数据库。为增强研究普适性，团队还融合了国际公开的 CEDAR、GPDS 等拉丁文数据集，奠定了跨文种研究的坚实基础。在数据支撑下，我们从离线和在线签名的特征提取、时序建模等基础问题突破，发表多篇论文，申请多项软著及专利；继而将算法优化成果转化为应用，指导学生团队开发出适用于新疆地区的多文种签名识别和鉴别系统，在大学

生创新竞赛中脱颖而出，并在新疆公检法部门和多家企业中推广和应用。最终，这项贯穿数据采集、理论创新到技术落地的系统性研究，分别通过了中国图象图形学会组织和计算机学会组织的成果鉴定，被认定为整体技术达到国际先进，部分技术达到领先水平，不仅见证了从基础研究到实际应用的跨越，也为多民族地区信息安全提供了原创性解决方案。未来，团队将继续扩大数据规模，探索结合多模态和小样本学习的多文种签名识别与鉴别技术，相关技术和成果拓展至中、西亚、“一带一路”沿线国家和地区。

您获批了多项科研项目，这实在令人钦佩。请您跟大家分享一下您在项目申请方面的宝贵经验和心得体会？相信您的真知灼见对我们大家都会是极大的启发和帮助？

科研项目申报的成功关键在于精准把握“需求导向、问题驱动、技术突破”三大核心要素，通过系统化的论证框架实现创新价值的有效传递。在选题设计环节，我们采用“痛点分析-需求验证-方案设计”的研究路径，以新疆少数民族文字手写签名这一亟待突破的技术空白为例，创新性地构建了传统学习与深度学习相融合的技术体系。申报材料撰写注重“三位一体”的可视化呈现策略：研究框架图需清晰展现科学问题的递进关系，技术模型图应完整呈现创新点，文字表述则坚持“问题-方法-价值”的三段式结构。在关键问题凝练和文档审核上，我们建立了“多轮评审-交叉修改-细节优化”的质量控制机制，通过团队成员的协同打磨，确保项目书达到学术严谨性与表达精准性的统一标准。

您是新疆维吾尔自治区青少年科普专家，并兼任乌鲁木齐市水磨沟区中小学科技副校长，这两个职务是对您的莫大信任和肯定，可否仔细谈谈您遇到的趣事？另外，这一荣誉是您甘为人梯的结果，可否分享您在教书和育人方面的心得？

作为自治区科普专家和乌鲁木齐市水磨沟区中小学科技副校长，多次到中小学做人工智能方面的讲座，

发现新疆地区的师生对新技术特别是人工智能领域的新技术的关注度很高，并且渴望学习相关技术。2025年2月份，我作为科普宣讲人被邀请到乌鲁木齐市第三十九小学举办“星耀苍穹，梦想启航”科技节系列活动。我以《人机共生时代，你准备好了吗？》为题目，介绍了人工智能的发展历程、前沿成果以及未来趋势。在问答环节时，孩子们向我提出了很多有趣、有意义的问题，如“人工智能能帮我写作业吗？”、“人类是否能完全控制人工智能？”，我对同学们的表现给予充分肯定与鼓励，直到讲座结束，还有不少同学围在我身边提出了很多疑问和困惑。我认为教育不仅是传授知识，还需注重启发式教学，培养学生的批判思维 and 创新能力。针对不同类型的学生，需要因材施教，耐心引导他们发现自己的潜力，促进学生的全面发展。此外，我始终以立德树人为核心，通过言传身教传递价值观，与学生们建立信任，让他们在求知中明理，在成长中成人。

您担任了多个学术职务、期刊编委和大会主席等，这必然占据您大量时间，能否跟大家分享一下您是如何协调本职工作和兼职工作的？能分享一下您的经验吗？

在协调学术本职与各类兼职工作的过程中，本人形成了系统化的工作机制。首先建立优先级评估，将任务按紧急程度与价值度进行划分，确保重点工作高效推进。充分利用时间，将科研、审稿、会议等不同性质的工作合理分配至特定时段，既保证研究深度又兼顾学术服务。尤其是注重资源整合与学生培养，通过指导学生参与部分编务工作、将学术服务转化为科研合作机会等方式提升整体效率。我认为青年学者应结合自身学术发展定位，有选择地参与符合学科发展前沿的战略性学术服务工作，通过建立科学的时间管理机制和资源整合策略，促进科研创新与服务效能的双向提升。

作为“新疆维吾尔自治区教育工委优秀共产党员”这个光荣称号的获得者，这对您个人意味着什么？这份荣誉对您今后的工作提出了哪些新的要求？

2021年7月在中国共产党建党100周年之际，本人有幸获得“新疆维吾尔自治区教育工委优秀共产党员”这一光荣称号，对我个人而言，是党组织对我工作的充分肯定和激励。这不仅是一份荣誉，更是一份责任和使命。这份荣誉将会激励我在教育工作中更加坚定地贯彻党的方针，特别是在推动新疆人工智能教育发展方面，我将致力于培养更多本土AI人才，促进智能技术与教育教学深度融合，为新疆数字化转型提供人才支撑。同时，我会继续发挥党员先锋模范作用，在促进民族团结、维护边疆稳定中作出更大贡献，用实际行动践行共产党员的初心使命。

可否请您谈一下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？

在第三代人工智能时代，计算机视觉的发展将集中于三个主要方向：从感知到认知的跨越、多模态深度融合和生成式视觉革命。首先，通过神经符号系统和物理引擎的结合，计算机视觉将不仅限于模式识别，而是能够理解场景逻辑关系。例如，NVIDIA的Voyager模型通过物理模拟预训练，帮助机器人掌握物体的隐含知识。其次，多模态深度融合将实现视觉、语言和声音等信息的统一表征，基于ViT与LLM的端到端架构不断涌现，使视觉系统在具身智能中的应用更为深入，能够有效指导机器人行动决策。最后，生成式视觉革命将推动可控内容的生成，支持自然语言指令，特别是在工业质检和医疗数据合成等领域，生成技术将有效提升稀缺数据的利用率。

尽管前景广阔，这一时代仍面临可解释性、小样本学习、动态环境适应和伦理安全等挑战。可解释性的问题在于CV模型的黑箱特性，导致信任危机；小样本学

习在数据稀缺情况下难以训练出有效模型；动态环境适应性在复杂天气条件下的表现不足；而伦理与安全风险方面，深度伪造和数据偏见等问题亟需解决。

我认为，未来的研究应聚焦于脑启发视觉计算、视觉-语言-行动的统一、量子视觉感知和隐私保护视觉计算等方向。脑启发视觉计算可模拟人脑处理机制，降低功耗；视觉-语言-行动的统一能够提升AI的指令理解和行动规划能力；量子视觉感知将通过量子技术提高图像匹配效率；而隐私保护视觉计算则可在保证数据安全的前提下完成模型训练。在未来3-5年，计算机视觉技术将在智能制造、智慧医疗、元宇宙和农业等领域广泛应用，创造显著的经济价值，其核心竞争力将从数据规模转向知识密度，神经符号计算与具身视觉的交互研究将展现出强大的发展潜力。

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

团队采用分组管理模式，按研究方向或项目需求划分为若干小组，每组有相关的教师、博士和硕士生，确保学科交叉和年级梯度搭配，并动态调整。各小组实行目标管理，制定年度、季度考核任务，博士生担任组长，同时建立组内和组间学术汇报促进思维碰撞。研究生培养实施分层要求：新生重点训练文献研读与实验复现能力，中期研究生需独立承担子课题，高年级研究生则负责指导低年级成员并参与跨组协作。问题处理采用分级机制：组内问题优先由组长协调，重大问题提交团队管理层。该管理模式通过优化资源配置、强化过程管理，在提升科研效率的同时系统培养研究生的团队协作能力和独立科研素养。

如果吐露研究工作者的心声，您最想说的是什么呢？

作为一名研究工作者，无论未来是面对机遇挑战还是困难险阻，始终不应忘记最初做科研的初心，保持热

爱，诚信务实。另外，需要阅读大量的文献并积极参与国内外学术会议，保持与时俱进的思想，进一步将模式识别，计算机视觉等创新型应用转变为实用型产品与服务，从而满足个人、社会乃至国家的迫切需求。

责任编辑 赵振兵 余焯



库尔班·吾布力

博士，新疆大学二级教授、博士生导师。中国计算机学会(CCF)常务理事、中国图象图形学学会(CSIG)理事、CCF 乌鲁木齐副主席、NOI 新疆竞赛委员会主席、GESP 和 PTA 新疆大学考点和服务中心的负责人、自治区教育工委优秀共产党员、新疆维吾尔自治区“天山英才”科技创新领军人才、自治区青少年科普专家，乌市水磨沟区中小学科技副校长。中国自动化学会模式识别与机器智能专业委员会(CAA-PRMI)副秘书长，CCF、CAA、CSIG、CAAI 等学会在人工智能领域的多个专业委员会委员，CCF/CSIG 杰出会员，IEEE、IAPR、IAENG 会员，新疆大学计算机学科群“双一流”科研平台——大规模场景实时三维重建与理解平台负责人，新疆大学“场景认知计算和智能系统研究中心”主任。主要研究方向为模式识别、图像处理、计算机视觉、生物特征识别和自然语言处理等。IEEE TPAMI、Nerocomputing、IEEE THMS 等期刊的审稿人。担任 NLP AI2020-2025 的大会主席、AIHCIR 2022-2024 联合主席、CCBR2018/PRCV2024 程序委员会主席、PRCV2019/IJCB2021 领域主席、NCIG2020/CPCC2022/CCFAI2023 的本地主席；多次担任 CVPR、ICPR、ICDAR、CCFAI 等会议的程序委员会委员或审稿人。主持国家重点研发课题 1 项、国家自然科学基金 4 项，省部级和其他项目 10 余项，在国内外期刊和会议上发表论文 300 余篇，出版学术专著 3 部，专利 10 余项，计算机软件著作权 60 余项，获得中国图象图形学学会科技进步二等奖、自治区自然科学优秀论文二等奖等省部级以上奖项 20 余项。

委员好消息

🌟 2025年4月24日，2024年度中国航天基金会奖颁发，CCF-CV专委会2位执行委员、西北工业大学**张艳宁**和北京空间飞行器总体设计部**王大轶**荣获“钱学森杰出贡献奖”。“钱学森杰出贡献奖”主要授予近10年来在航天领域基础科学研究、关键技术攻关、重大工程实施等方面做出杰出贡献的科技工作者，本年度全国仅18人获此殊荣。

🌟 2025年4月29日，CCF-CV专委会常务委员、北京邮电大学**马占宇**荣获北京市先进工作者荣誉称号。

🌟 2025年5月21日，北京市人民政府关于2024年度北京市科学技术奖励的决定发布，CCF-CV专委会7位执行委员获奖，中科院自动化所**张兆翔**获青年科学家北京市杰出青年中关村奖，中科院自动化所**王亮、黄岩、谭铁牛**等完成的“深度认知神经网络理论与方法”、北京科技大学**马惠敏**等完成的“认知启发的视觉计算理论与方法”获自然科学一等奖，北京交通大学**朱振峰**等完成的“数据和知识驱动的智慧诊疗关键技术及产业化应

用”、清华大学**黄高等**完成的“深海可控式交互作业机器人关键技术与应用”获科技进步一等奖。

🌟 2025年5月23日，ICRA 2025最佳论文揭晓，CCF-CV专委会执行委员、上海交通大学**卢策吾**团队的论文 Human-Agent Joint Learning for Efficient Robot Manipulation Skill Acquisition 获人机交互最佳论文奖。

🌟 2024年6月19日，ACM顶会SIGGRAPH 2024最佳论文揭晓，CCF-CV专委会常务委员、上海科技大学**虞晶怡**等完成的论文 CLAY: A Controllable Large-scale Generative Model for Creating High-quality 3D Assets 和 DressCode: Autoregressively Sewing and Generating Garments From Text Guidance 获荣誉提名。

责任编辑 刘海波

多模态医疗图像辅助诊断开源代码

东北大学 贾同 贾娜娜

多模态数据在医疗诊断中具有重要且不可替代的作用。单一模态往往只能捕捉疾病的某一方面特征，而不同模态（如 MRI、PET、CT、临床指标、基因数据等）反映的是疾病从结构、功能、生化、分子等多个层面的信息。通过多模态数据的融合，不仅可以提高疾病识别的准确性和鲁棒性，还能对疾病机制更全面的理解，提升个体化诊疗水平，尤其阿尔茨海默症、肿瘤等复杂疾病的早期筛查与分型预测中表现出显著优势。因此，有效整合多模态数据已成为现在智能医疗诊断系统发展的关键方面。本文主要从三种疾病类型（阿尔茨海默病分类、脑肿瘤分割、乳腺癌诊断）介绍该领域的研究成果。

1、Multi-fusion and Disease-induced Learning (MDL-Net)

MDL-Net 利用 Structural Magnetic Resonance Imaging (sMRI) 模态和 Positron Emission Tomography (PET) 模态融合用于阿尔茨海默病的诊断，融合模块主要包括全局感知学习模块、局部感知学习模块和潜在空间学习模块。其中，全局感知模块通过自适应 Transformer (SAT) 学习模块之间的全局关系；全局感知模块构建了局部感知卷积结构，用于捕捉模态之间的局部关联；潜在空间学习模块通过外积操作引入潜在信息，进一步增强特征表示能力；MDL-Net 融合了基于梯度权重的疾病驱动区域感知学习 (DRL) 模块，从未提升模型的可解释性，该模块通过迭代学习权重矩阵来识别阿尔茨海默病相关的大脑区域。完整的网络结

构如图 1 所示。

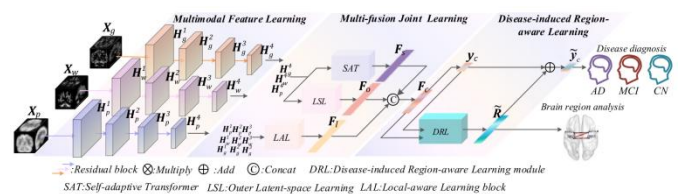


图 1 MDL-Net 架构图

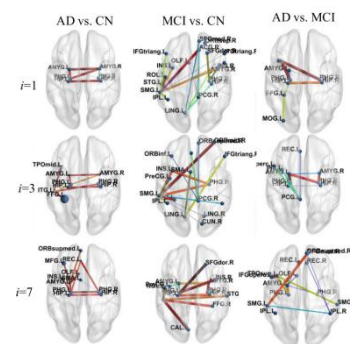


图 2 三个任务在不同迭代次数下的大脑连接网络

图 2 为结合大脑区域的多项功能分区，对提取出的前 10 各显著大脑区域进行分析的结果。实验结果表明 MDL-Net 在阿尔兹海默病诊断中取得了较高的性能，证明了所提出融合方法融合 MRI 模态和 PET 模态的有效性（例如，MDL-Net 取得了 96.37% 的准确率，优于最新的对比方法）。

更多有关 MDL-Net 的详细内容可参考发布该方法的论文“3D Multimodal Fusion Network With Disease-Induced Joint Learning for Early Alzheimer’s Disease Diagnosis”。

多模态医疗图像辅助诊断开源代码

在脑肿瘤挑战赛分割数据集 BraTS-2018 上进行了大量实验，结果表明，在模态缺失的情况下，所提出的框架相比现有的先进方法具有明显优势。如图 4 所示为脑肿瘤分割的可视化结果。

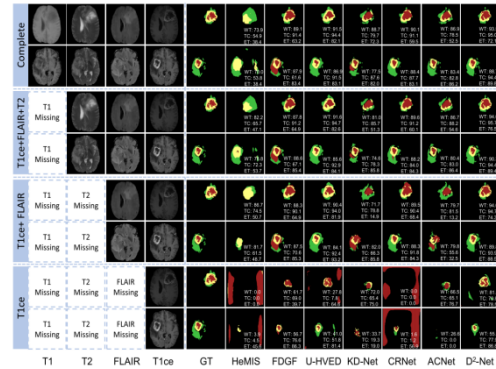


图 4 脑肿瘤分割结果图

更多有关 D²Net 的详细内容可参考发布该方法的论文 “D2-Net: Dual Disentanglement Network for Brain Tumor Segmentation With Missing Modalities”。

论文地址:
<https://ieeexplore.ieee.org/document/9775681>
代码地址:
<https://github.com/CityU-AIM-Group/D2Net>

3、Trusted Dynamic Feature Fusion (TDF-Net)

超声成像是乳腺癌诊断中的关键成像技术。然而，多模态乳腺超声的诊断过程通常耗时且劳动强度大，严重依赖医生的丰富经验。因此，开发乳腺癌的计算机辅助诊断系统具有重要意义。现有的诊断系统未能充分考虑不同超声模态对诊断结果的不同影响，并且在实际临床中难以应对模态缺失的问题。为此，提出了一种用于不完整多模态超声数据的乳腺癌诊断方法——可信动态特征融合网络 (TDF-Net)。该方法首先引入了双分支特征提取模块，用于捕捉模态特异性信息。同时设计了一种对比聚类损失，用以施加一致性约束，确保每个样本的不同模态特征在语义空间中的一致性。此外，本文

论文地址:
<https://ieeexplore.ieee.org/document/10498133>
代码地址:
<https://github.com/qzf0320/MDL-Net>

2、Dual Disentanglement Network (D²Net)

多模态磁共振成像 (MRI) 能够为脑肿瘤的自动分割提供互补信息，这对于疾病的诊断与预后具有重要意义。然而，在实际临床中，模态缺失的情况十分常见，这会导致许多依赖完整模态数据的现有方法失效。当前最先进的方法通过融合多模态图像和特征来学习肿瘤区域的共享表示，从而应对模态缺失的问题，但这些方法通常忽视了对模态间及模态与肿瘤区域之间关联关系的显式建模。受不同模态信息在分割不同肿瘤区域中具有不同作用这一事实的启发，可以显式地挖掘多种模态特异性信息与肿瘤特异性知识之间的相关性，从而提升分割性能。因此，提出 D²Net 用于处理模态缺失情况下脑肿瘤分割的双重解耦网络，该网络包括模态解耦阶段 (MD-Stage) 和肿瘤区域解耦阶段 (TD-Stage)。在 MD-Stage 中，设计了一种空间-频率联合的模态对比学习机制，以直接从 MRI 数据中解耦模态特定信息。为了分离肿瘤特异性并提取判别整体特征，在 TD-Stage 中提出了一种基于亲和引导的稠密肿瘤区域知识蒸馏机制，通过将解耦后的二值教师网络与一个整体学生网络的特征对齐来实现。通过显式的挖掘模态与肿瘤区域之间的关系，即使在某些模态缺失的情况下，所提出的模型也能够学习到足够的分割信息，其结构如图 3 所示。

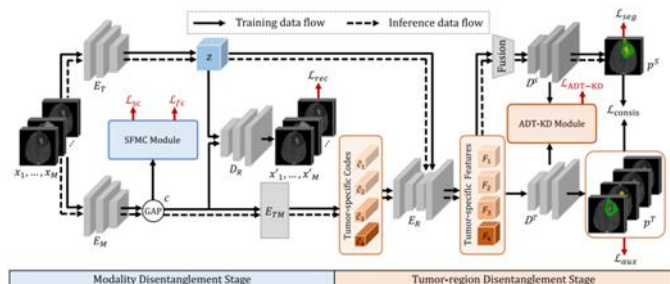


图 3 D²Net 结构图

< 0.05。如图 6 所示为可视化结果图。

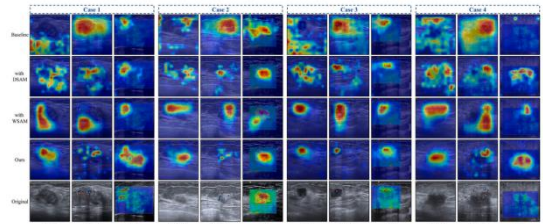


图 6 可视化结果图

更多有关 TDF-Net 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://doi.org/10.1016/j.inffus.2024.102592>

代码地址:

https://github.com/TimesXY/TDF_Net

提出了一种基于可逆神经网络的模态恢复方法，用于建立不同模态之间的映射关系，从而实现缺失模态的重建。最后，提出了一种基于 Dirichlet 分布的可信动态特征融合模块，用于结合不确定性信息量化每种模态对诊断结果的贡献，从而在不同样本之间实现模态特征的动态融合，总体结构如图 5 所示。

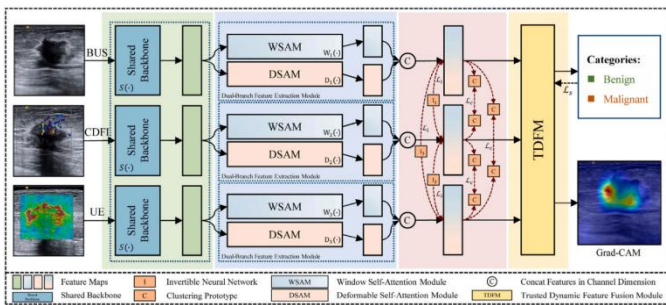


图 5 TDF-Net 结构图

TDF-Net 在一个已有的多模态乳腺超声数据集上进行了验证，诊断性能优于现有方法，平均 AUC 达 98.31%，95%置信区间为[96.98%, 99.64%]，且 p 值



贾同

东北大学教授、博士生导师，国家高层次领军人才，未来技术学院/机器人科学与工程学院党委书记。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



贾娜娜

博士研究生，东北大学信息科学与工程学院，研究方向为医学影像处理
电子邮箱: 2010284@stu.neu.edu.cn

责任编辑 李策 樊鑫

多模态三维目标检测数据集

兰州理工大学 李策 王宗顺

三维目标检测 (Multi-modal 3D Object Detection) 是计算机视觉领域中的一项关键任务，主要从三维点云数据中精确地识别和定位三维空间中的物体。随着自动驾驶、机器人感知和智能监控等应用的需求日益增长，三维目标检测技术正朝着多模态融合、长序列时序建模和多人交互等方向发展。然而，现有的三维目标检测数据集多集中于单一模态或特定场景，缺乏多模态、多场景、多交互的综合性数据支持。因此，构建涵盖多模态传感器数据（如 LiDAR、摄像头、毫米波雷达等）的大规模、多样化的三维目标检测数据集，成为推动该领域发展的关键。

本文将重点介绍具有代表性的多模态三维目标检测公开数据集，涵盖自动驾驶恶劣天气感知等应用场景，展示多模态融合技术在三维目标检测中的最新进展。

1、Cross Mechanism Dataset

CMD (Cross Mechanism Dataset, CMD)数据集由厦门大学团队于 2024 年发布。包含来自五类七种传感器的数据,分别为 128 线和 32 线机械 LiDAR、固态 LiDAR、4D Radar 和三台相机。所有传感器通过高精度的 PTP 时间同步。

数据涵盖 50 段场景，总计 10,000 帧/传感器。每个目标被标注为 9 自由度的三维包围框，并附带遮挡、运动状态与追踪 ID 等信息。CMD 数据集包含来自五种类型、共七种传感器的数据：包括一台 128 线

的机械旋转 LiDAR、一台 32 线的 MS LiDAR、一台固态 LiDAR、一台 4D 毫米波雷达，以及三台摄像头，如图 1 所示。

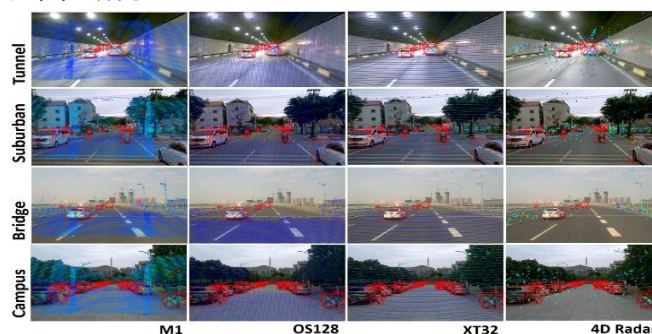


图 1 CMD 数据集中不同传感器采样数据示例

数据集下载地址:

<https://github.com/im-djh/CMD>

相关论文链接:

CMD: A Cross Mechanism Domain Adaptation Dataset for 3D Object Detection.

https://fq.pkwyx.com/default/https/www.ecva.net/papers/eccv_2024/papers_ECCV/papers/07443.pdf

2、TUMTraf-V2X

TUMTraf-V2X 是由慕尼黑工业大学与弗劳恩霍夫交通系统与智能交通研究所联合发布在 CVPR 2024 的高质量真实世界车-路协同 (Vehicle-to-Everything, V2X) 多模态三维点云数据集，专为自动驾驶中的合作三维目标检测与跟踪任务设计。该数据

集包含来自 5 个路侧传感器和 4 个车载传感器的 2,000 个标注点云和 5,000 张标注图像，共计 30,000 个带轨迹 ID 的三维边界框。标注涵盖 9 个类别，支持 OpenLABEL 标准格式，适用于多模态感知算法的训练与评估。



图 2 TUMTraf-V2X 数据集中 3D 框标签示例

数据集下载地址:

<https://tum-traffic-dataset.github.io/tumtraf-v2xD>

相关论文链接:

TUMTraf V2XCooperative Perception Dataset.

https://openaccess.thecvf.com/content/CVPR2024/papers/Zimmer_TUMTraf_V2X_Cooperative_Perception_Dataset_CVPR_2024_paper.pdf

3、HoloVIC

HoloVIC (Holographic Vehicle-Infrastructure Cooperation) 是由清华大学 SenseAuto 研究团队于 2024 年在 CVPR 上发布的大规模多传感器全息交叉口与车路协同数据集，旨在提升自动驾驶系统在复杂交通环境中的感知能力。

该数据集采用三种不同类型的传感器：摄像头 (Camera)、激光雷达 (LiDAR) 和鱼眼相机 (Fisheye)，并在不同的十字路口布局中使用四种传感器组合。每个交叉口配备了 6 至 18 个传感器，以捕获同步数据。数据集包含超过 100,000 帧同步数据，涵盖多种交通场景。为每帧数据提供了 3D 边界框标注，支持车辆、行人和其他交通参与者的检测与跟踪。基于 HoloVIC 数据集，定义了五个任务：单目 3D 检测 (Mono3D)、激光雷达 3D 检测、多人跟踪 (MOT)、

多传感器多人跟踪 (MSMOT) 和车路协同感知 (VIC Perception)。

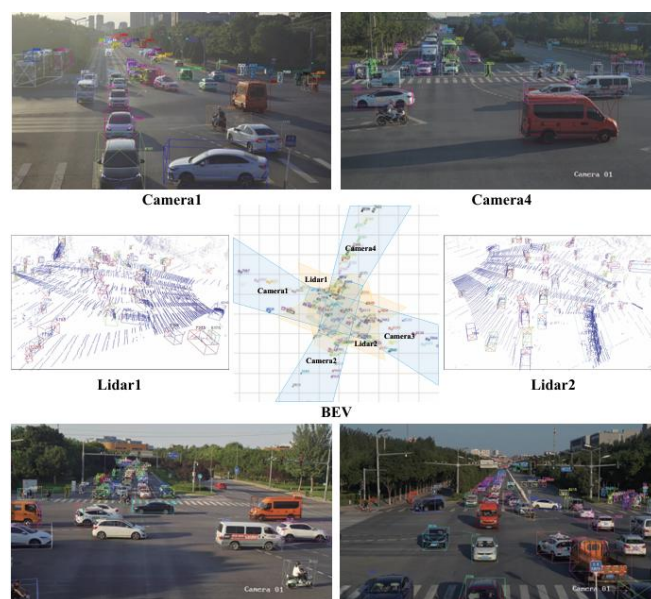


图 3 HoloVIC 数据集带注释的 3D 框标签示例

数据集下载地址:

<https://holovic.net>

相关论文链接:

HoloVIC: Large-scale Dataset and Benchmark for Multi-Sensor Holographic Intersection and Vehicle-Infrastructure Cooperative

https://openaccess.thecvf.com/content/CVPR2024/papers/Ma_HoloVIC_Large-scale_Dataset_and_Benchmark_for_Multi-Sensor_Holographic_Intersection_and_CVPR_2024_paper.pdf

4、MSU-4S

MSU-4S (Michigan State University Four Seasons) 是由密歇根州立大学于 2024 年在 CVPR 会议上发布的大规模多模态自动驾驶数据集，旨在填补现有公开数据集中在恶劣天气和季节变化下的感知盲点。

该数据集涵盖四个季节（春、夏、秋、冬）和多种天气条件（晴朗、雨天、雪天），在密歇根州立大学校园及周边地区采集，包含超过 100,000 帧的同步数据，涵盖相机、激光雷达（LiDAR）、雷达、全球导航卫星系统（GNSS）、车轮速度和转向角度等多种传感器数据。MSU-4S 的独特之处如图 4 所示，几乎每个场景及其对应的传感器数据都在四个不同季节中采集，提供了丰富的跨季节和跨天气条件的感知数据，适用于研究物体检测、领域适应和多模态融合等任务。数据集的标注包括二维和三维物体检测框，适用于训练和评估感知算法。



图 4 MSU-4S 数据集中不同天气、季节数据示例

李 策



教授，博士生导师，兰州理工大学电气工程与信息工程学院从事教育与科研工作，任信息化建设与管理处（信息化推进办公室）处长。研究方向为计算机视觉、医学影像分析，智能机器人等。

王宗顺



兰州理工大学电气工程与信息工程学院 博士研究生，研究方向为计算机视觉，三维场景感知等。

责任编辑 王田 贾同

好文推荐

清华大学和南开大学共同完成的最新成果“Referring Camouflaged Object Detection”发表在 IEEE TPAMI 2025。

论文: Xuying Zhang, Bowen Yin, Zheng Lin, Qibin Hou, Deng-Ping Fan, and Ming-Ming Cheng. Referring Camouflaged Object Detection, IEEE TPAMI, 47 (5): 3597-3610 (2025)

伪装目标检测 (Camouflaged Object Detection, COD) 旨在分割视觉上隐藏在周围环境中的目标。近年来, 该研究方向受到了越来越多的学术关注。伪装目标检测在医学图像分割、表面缺陷检测、农业病虫害检测等领域具有重要的实际应用价值。在真实应用场景中, 同一画面往往存在多个伪装目标, 但研究人员仅希望从画面中找出面向应用的特定的伪装目标。在这种设定下, 如果研究人员可以事先采集到这些伪装目标的参考图

片 (References), 那么该伪装目标的定位过程将变得具有方向性且简单可执行。本文研究人员将这类具有参考图像的伪装目标检测简称为 Ref-COD。

为了研究参考信息在 Ref-COD 中的作用, 文章作者首先建立了一个名为 R2C7K 的大规模数据集。该数据集由 7000 张图像组成, 覆盖了现实世界场景中的 64 个对象类别。接下来, 文章设计了一种双分支网络结构, 命名为 R2CNet, 用于伪装目标的检测。如图 1 所示, R2CNet 包括参考分支 (Reference Branch) 和分割分支 (Segmentation Branch)。其中, 参考分支用于从参考图像中嵌入目标物体的共同表示。分割分支则在共同表示的指导下识别和分割伪装物体。进一步而言, 文章设计了一个参考掩模生成模块来生成像素级先验掩模, 还提出一个参考特征富集模块来增强对指定伪装目标的识别能力。大量实验表明, Ref-COD 方法在分割特定伪装目标和识别目标主体方面优于同类方法。

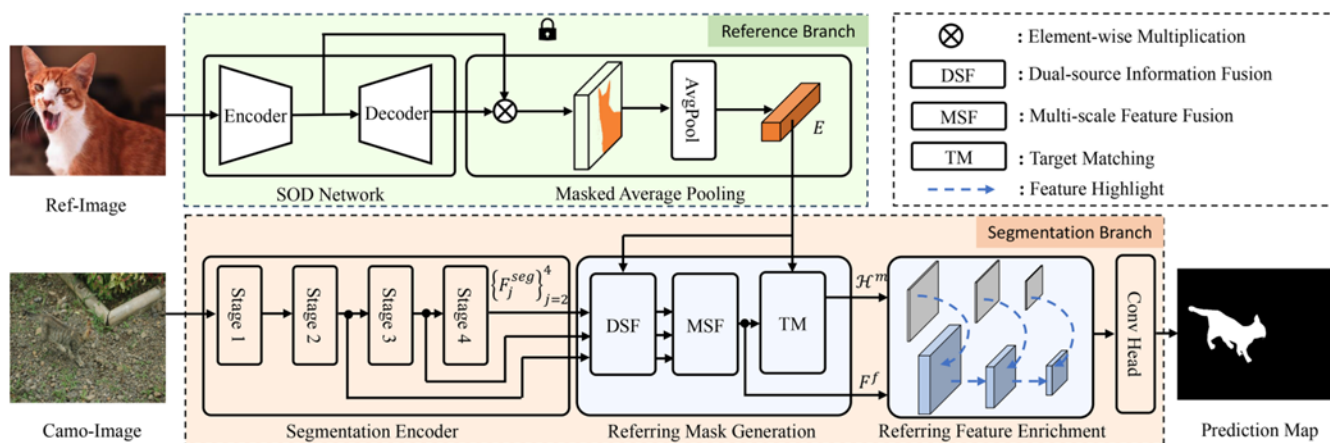


图 1 R2CNet 模型结构流程图

责任编辑 贾同 樊鑫

好文推荐

中国人工智能研究所 (TeleAI) 和大连海事大学的最新成果 “Spatial Residual for Underwater Object Detection” 发表在 IEEE TPAMI 2025。

论文: Jingchun Zhou, Zongxin He, Dehuan Zhang, Siyuan Liu, Xianping Fu, and Xuelong Li. Spatial Residual for Underwater Object Detection, IEEE TPAMI, 47 (6): 4996-5013 (2025)

水下目标检测 (Underwater Object Detection, UOD) 旨在检测水下环境中的目标。与陆地成像不同, 水下成像受到水下环境固有的多种复杂耦合退化因素的严重影响。这些因素会扭曲统计特性, 改变结构信息, 模糊真实数据特征, 影响视觉感知, 极大挑战了现有基于深度学习的目标检测方法。

目标特征与退化因子的动态耦合导致特征漂移, 降低水下探测器的性能。本文将特征漂移重新定义为在求

解偏微分方程 (PDEs) 时目标特征在边界约束下的不稳定性。基于该理论基础, 本文提出了空间残差 (Spatial Residual, SR) 块, 使用 SkipCut 在网络宽度上建立有效的约束来求解偏微分方程, 并优化解空间。基于所提空间残差块, 本文进一步建立一个面向复杂特征场景的具有 5 阶段的通用骨干网络, 该网络简称为 BSR5。简单来说, BSR5 通过 SkipCut 提取离散通道切片, 其中每个切片的特征在适当的数据容量内进行解析。在梯度反向传播中, SkipCut 作为一条捷径, 优化信息流和梯度分配, 以提高性能和加速训练。在 RUOD 数据集上的实验表明, 使用 BSR5 的 DETRs 和 YOLOs 在传统检测器和端到端检测器上均取得了最先进的结果。与使用 ResNet-101 的 RT-DETR 相比, BSR5-DETR 提高了 1.3% 和 2.7% AP, 同时参数分别减少了 41.6% 和 6.6%。进一步验证突出了 BSR5 的强收敛性和鲁棒性。

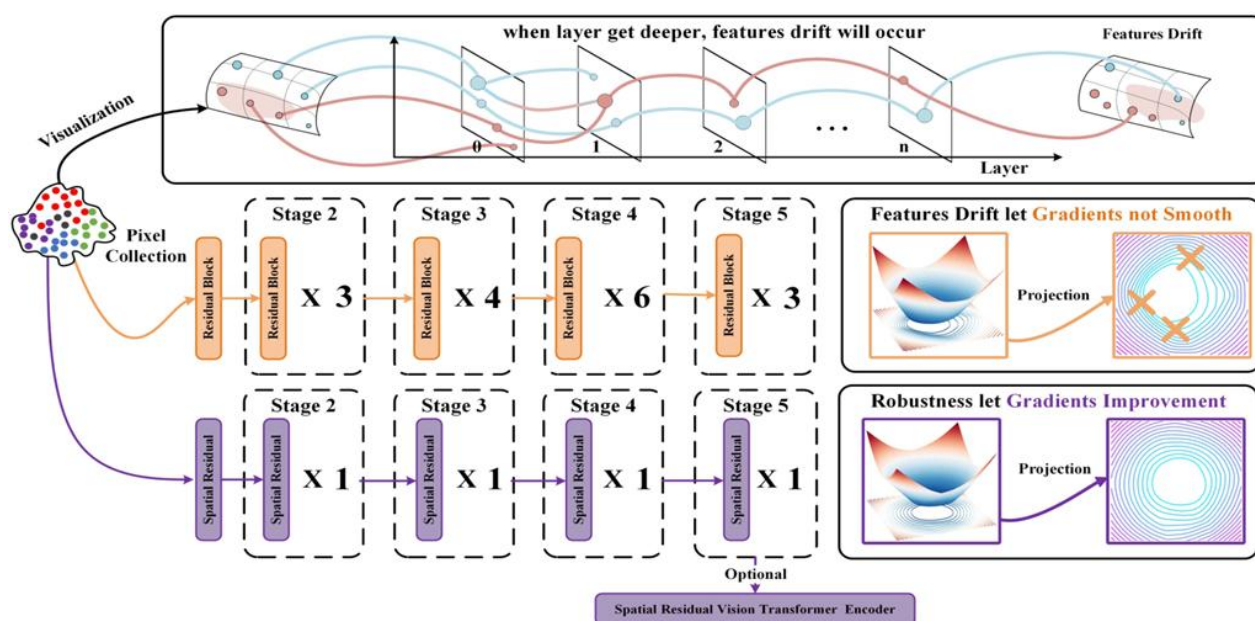


图 1 BSR5 模型结构流程图

责任编辑 李策 贾同

好文推荐

杭州电子科技大学的“Prophet: Prompting Large Language Models with Complementary Answer Heuristics for Knowledge-based Visual Question Answering”最新成果发表在 IEEE TPAMI 2025.

论文 Zhou Yu, Xuecheng Ouyang, Zhenwei Shao, Meng Wang, Jun Yu. Prophet: Prompting Large Language Models with Complementary Answer Heuristics for Knowledge-based Visual Question Answering, IEEE TPAMI, 2025, 10.1109/TPAMI.2025.3562422

基于知识的图像问答 (Knowledge-based VQA) 不仅要求模型分析图像和问题, 还需结合外部知识进行推理。早期方法依赖显式知识库检索相关信息, 易引入噪声, 影响性能。近期如 PICa 等工作, 利用大语言模型 (如 GPT-3) 作为隐式知识库, 通过图像转文本描述并用少样本提示引导回答, 取得了良好效果。但本文认为此方法仍存在瓶颈:

1) 输入 GPT-3 的图片信息往往不够充分, 忽略了图像中与问题相关的诸多细节, 2) 对示例的选择算法是次优的, 未充分利用问题和答案的语义信息。

为解决上述问题, 本文提出 Prophet 和 Prophet++, 通过答案启发辅助大语言模型完成知识型 VQA 任务。如图 1 所示, Prophet 训练轻量 VQA 模型, 提取两类启发信息: 1) 答案候选 (候选答案及置信度); 2) 答案感知示例 (基于特征相似度选取的标注样本)。如图 2 所示, Prophet++ 进一步引入多模态大模型生成第三类启发——答案支持原因, 并生成最终回答, 显著提升性能, 在 OK-VQA 和 A-OKVQA 上分别达 65.7% 和 68.0%。

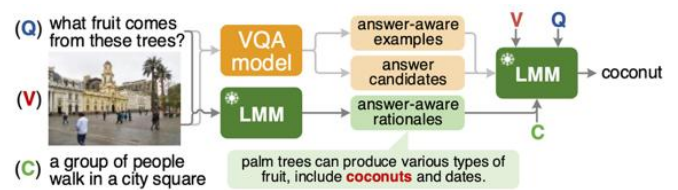


图 1 Prophet++ 的框架图

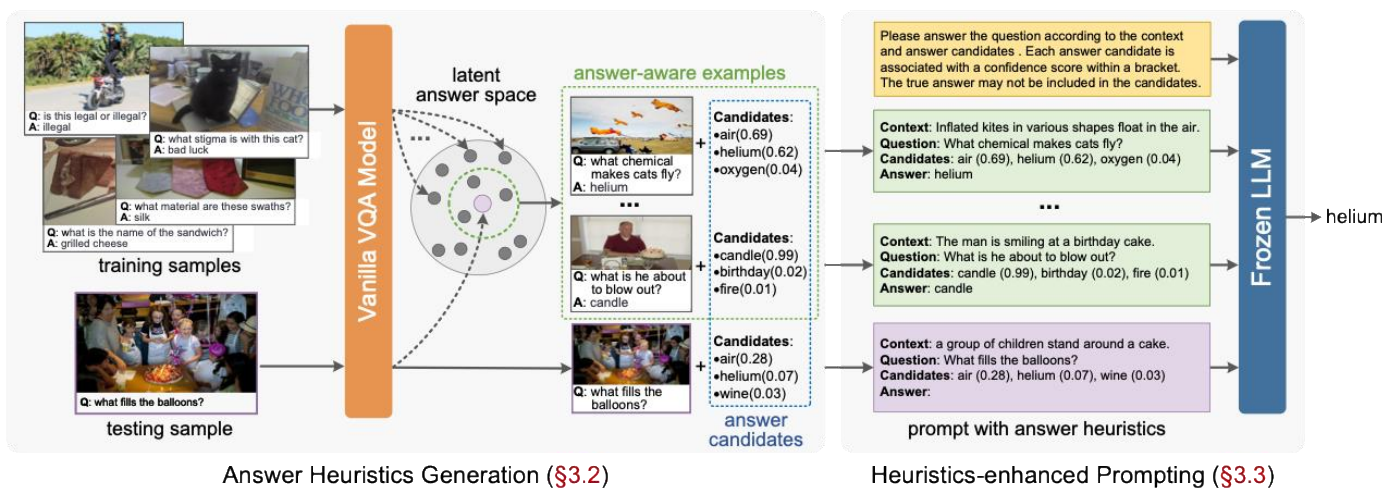


图 2 Prophet 的框架图

责任编辑 樊鑫 王田

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 Selected Topics in Signal Processing, Computerized Medical Imaging and Graphics, Computer Vision and Image Understanding 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。

第八届 PRCV 将于 2025 年 10 月 16 日至 10 月 19 日在上海举办，由上海交通大学承办。本届会议将秉持团结模式识别与计算机视觉领域科技工作者的宗旨，进一步推动开放合作，广泛吸引学术界和工业界的人才，提升会议的国际化水平，力求打造一个高品质的学术交流平台。大会的举办将为学术界与工业界提供更多产学研合作机会，推动模式识别与计算机视觉领域的协同创新和可持续发展。

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 |
|--------------------|---------------|----------------|------------|---|
| AIVRV 2025 | 2025.09.05-07 | Chengdu, China | 2025.07.12 | https://www.aivrv.org/ |
| AAAI 2026 | 2025.01.20-27 | Singapore | 2025.08.11 | https://aaai.org/conference/aaai/aaai-26/ |
| AIPCVT 2025 | 2025.12.12-14 | Nanjing, China | 2025.09.12 | https://ipcv.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 |
|---------------|--|---|------------|
| JSTSP | AI in Signal & Data Science -- Toward Large Language Model (LLM) Theory and Applications | iee.atyponrex.com | 2025.07.01 |
| JSTARS | Remote sensing information processing and dynamic monitoring in agriculture | https://www.grss-ieee.org/wp-content/uploads/2024/09/JSTARS_CfP_Remote-sensing-information-processing-and-dynamic-monitoring-in-agriculture.pdf | 2025.07.31 |
| CMIG | Generative Artificial Intelligence for Medical Imaging | https://www.sciencedirect.com/special-issue/319131/generative-artificial-intelligence-for-medical-imaging | 2025.09.30 |
| CVIU | Computer Vision for Sports and Winter Sports | https://www.sciencedirect.com/special-issue/321225/computer-vision-for-sports-and-winter-sports | 2025.09.30 |

COMPUTER VISION NEWSLETTER

02 2025
总第 44 期



计算机视觉专委会简报



CCF 计算机视觉
专委会