主办 CCF计算机视觉专业委员会

COMPUTER VISION NEWSLETTER

CCF 计算机视觉专委会简报

04 2025

总第 46 期

中国计算机学会 CCF

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 04 期

总第 46 期

主　办　CCF 计算机视觉专业委员会

编委会

CONTENTS
简报目录

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

CCF-CV 走进高校系列报告会

第 147 期 山东大学

2025 年 10 月 25 日下午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，山东大学浪潮人工智能学院承办的第 147 期 CCF-CV 走进高校系列报告会——“多模态智能感知与理解”学术论坛在山东大学中心校区浪潮人工智能学院 212 报告厅成功举行。本次报告会邀请了来自中国科学院自动化研究所**王亮**研究员、南京航空航天大学**秦杰**教授、哈尔滨工业大学**张盛平**教授、中国科学院计算技术研究所**宋新航**副研究员及西北工业大学**赵斌**副教授等多位顶尖专家学者。活动由山东大学浪潮人工智能学院助理教授**何科技**担任执行主席。报告由山东大学**何科技**助理教授、**丛润民**教授、**元辉**教授依次主持。

活动由**王亮**研究员、**张盛平**教授、**秦杰**教授、**宋新航**副研究员和**赵斌**副教授五位专家针对多模态感知、具身智能及视觉理解的前沿进展与未来趋势分别做主题报告。在圆桌讨论环节，其中三位专家围绕“具身智能与人类认知差异”“模型意识的判断标准”“学术内心驱动力”等议题展开深入交流。本次报告会集中展示了国内多模态智能感知与具身智能领域的最新研究进展，为高校师生提供了宝贵的学习与交流机会。

第 148 期 中国地质大学（武汉）

2025 年 11 月 15 日上午，由中国计算机学会计算机视觉专委会（CCF-CV）主办，中国地质大学（武汉）计算机学院承办的第 148 期 CCF-CV 走进高校系列报告会——“智能感知与计算机视觉前沿” 论坛在中国地质大学（武汉）未来城校区计算机学院 126 报告厅成功举行。本次报告会邀请了来自中国科学院计算技术研究所**山世光**教授、浙江大学**潘纲**教授、东南大学**张敏灵**教授、武汉大学**夏桂松**教授、西安电子科技大学**张向荣**教授、南京理工大学**张珊珊**教授、清华大学**赵思成**副教授及上海科技大学**马月昕**研究员等多位顶尖专家学者。本次活动由中国地质大学（武汉）教授**刘袁缘**担任执行主席。报告由中国地质大学（武汉）**张洪艳**教授、**胡成玉**教授、**龚文引**教授、**陈云亮**教授依次主持。

活动首先由中国地质大学（武汉）党委副书记、纪委书记**唐忠阳**致欢迎辞。接着，由 CCF-CV 专委会常委、中国科学院计算技术研究所**山世光**教授代表专委会致辞。特邀报告环节，由**张敏灵**教授、**夏桂松**教授、**张向荣**教授、**张珊珊**教授、**赵思成**副教授和**马月昕**研究员针对多模态感知、遥感智能解译、具身智能及视觉理解的前沿进展与未来趋势分别做主题报告。在 Panel 讨论环

节，在中国科学院计算技术研究所**山世光**教授和浙江大学**潘纲**教授的话题引导下，专家围绕“大模型时代的计算机视觉未来路在何方”“视觉域世界模型构建方法”等议题展开深入交流。最后，中国地质大学（武汉）计算机学院院长张洪艳教授进行总结发言。他强调，本次活动作为计算机学院建院 20 周年暨计算机学科建设 40 周年系列学术活动的首场报告会，营造了浓厚学术氛围，为院庆学术季奠定了良好开端，学院将以此次论坛为起点，持续推出高水平学术活动，强化学科内涵建设，助力智能感知与计算机视觉领域实现高质量发展。

责任编委 毋立芳、朱安娜

第 25 期 空间智能

CCF-CV 视界无限系列研讨会

2025 年 12 月 6 日，由中国计算机学会计算机视觉专委会主办、东南大学自动化学院和东南大学计算机科学与工程学院承办的第 25 期 CCF-CV “视界无限”系列活动——“空间智能”专题研讨会在东南大学举行。

研讨会由东南大学**朱鹏飞**教授主持，CCF-CV 专委会副主任、南京邮电大学副校长**刘青山**教授和 CCF-CV 专委会常务委员、东南大学研究生院常务副院长**耿新**教授致辞。随后北京空间飞行器总体设计部**王大轶**研究员、中国科学院自动化研究所**张兆翔**研究员、国防科技大学**徐凯**教授、浙江大学**章国锋**教授、天津大学**田栢苓**教授、山东大学**宋然**教授做主题报告，东南大学**魏秀参**教授主持专题研讨并做会议总结。

CCF-CV 专委会副主任、南京邮电大学副校长**刘青山**教授和 CCF-CV 专委会常务委员、东南大学研究生院常务副院长**耿新**教授线上开场致辞，对各位专家到来表示感谢。

王大轶研究员以空间飞行器这一关键空间无人系统为研究核心，聚焦非自主诊断重构模式下故障后转安全模式等待地面处理的痛点，从控制学视角深入剖析无人控制系统故障状态的精准识别逻辑，系统探讨如何通过能力定量表征构建科学有效的故障应对方案设计体系。报告不仅展现了自主诊断与重构技术的核心突破方向，

更为保障空间无人系统的安全稳定飞行、推动其迈向自主行为提供了重要的理论支撑与技术参考，其研究成果对我国空间信息体系与智能装备升级具有关键意义。

张兆翔研究员系统梳理了空间智能的发展根基与演进逻辑。报告回溯了认知地图理论、心理旋转实验等经典研究成果，阐释了空间表征的模拟性质与认知地图的神经基础等核心发现，同时衔接具身智能、智能体学习等现代技术方向，探讨了空间智能从经典认知理论到多模态感知、自主交互等前沿应用的传承与创新。报告聚焦空间智能的内涵与外延，揭示其作为数字世界与物理世界融合桥梁的核心价值，为该领域的理论深化与技术突破提供了系统性视角。

徐凯教授深入阐述世界模型对具身智能发展的核心赋能价值。报告聚焦因果关系在具身系统动态建模中的关键作用，直指传统具身智能感知 - 执行脱节的核心问题。结合其 LaDi-WM 等研究成果，介绍通过融合几何与语义特征的潜在扩散建模，构建可精准预测未来状态的具身世界模型，借助因果推理实现高层规划与低层执行的协同优化，为长程操纵任务提供可靠决策支撑，推动具身智能从数据处理向认知推理跨越，彰显空间智能与世界模型融合的技术突破。

章国锋教授从原生 3D 场景生成技术入手，分享了三维场景的重建与生成技术的发展脉络与未来趋势。章国锋教授指出，重建和生成结合得越来越紧密，并逐步向端到端发展，生成场景的范围也越来越大，时空一致性越来越强。但目前高质量 3D 场景数据极为匮乏，极大制约了原生 3D 场景生成模型的发展。如何利用互联网影像数据和合成数据来解决数据匮乏问题是一个重要趋势。针对这个问题，章国锋教授提出了一个空间智能生成模型的有效实现路径，通过将重建与生成结合分三个阶段来解决 3D 场景数据匮乏的问题，最终实现原生 3D 场景生成。章国锋教授对未来发展趋势进行了总结，并指出了重建与生成的深度融合与生成的实时性与时空一致性的发展方向，为学术界和工业界在构建更真实、更高效的虚拟世界方面提供了宝贵的思路。

田栢苓教授聚焦于复杂非结构化环境下的机器人高速自主导航与灵巧跟踪问题，指出现有分层级联导航框架在实时性与误差累积方面的瓶颈。报告重点介绍了

YOPO 系列工作，提出了一种基于学习的一阶段端到端规划范式。该方法借鉴目标检测中的 Anchor 思想，将感知、检测与规划融为一体，通过指导学习策略利用环境梯度直接优化网络。报告展示了该框架在无人机丛林极速穿越、动态目标灵巧跟踪以及地面越野导航等场景中的卓越性能，强调了其在实现零样本虚实迁移及毫秒级极速响应方面的技术优势，为轻量化机载平台的敏捷自主飞行提供了新的解决思路。

宋然教授聚焦于“类人视觉感知驱动的机器人学习”这一前沿议题，指出机器智能若要实现符合人类伦理与行为准则的自主行动，必须首先建立与人类高度一致的主观感知能力。报告从“感知作为行为的前提”出发，深入探讨了当前机器视觉在数据受限与开放动态场景下面临的泛化瓶颈。结合在图表示学习与多模态视频理解领域的最新成果，重点介绍了如何利用信息论原则解决极小样本下的结构化数据学习问题，以及如何引入大语言模型的语义知识引导开放词汇的时序动作定位。特别强调了通过融合结构化先验与多模态语义，赋予机器人像人类一样“举一反三”的认知与泛化能力，为实现具身智能在复杂非结构化环境中的高效感知与决策提供了新的范式。

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 环节由东南大学计算机学院**魏秀参**教授主持，参与嘉宾包括北京空间飞行器总体设计部**王大轶**研究员、浙江大学**章国锋**教授、东南大学**朱鹏飞**教授、中国科学技术大学**毛震东**教授、山东大学**宋然**教授、同济大学**王曰英**研究员。

围绕空间智能的内涵、三维表征形式以及具备物理定律一致性的时空因果推理框架这三个核心问题。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。

最后由**魏秀参**（主持人）对 Panel 环节做了总结，并感谢各位专家的精彩解答。论坛在大家热烈的思想碰撞中落下帷幕。

责任编委 杨巨峰

2025 年度 CCF-CV 专委工作会议顺利召开

2025 年 10 月 17 日，中国计算机学会计算机视觉专委会（CCF-CV）2025 年度工作会议在上海国家会展中心成功召开，专委会秘书长、中国科学院计算技术研究所**王瑞平**研究员主持会议。

首先，专委会主任、中国科学院计算技术研究所所长**陈熙霖**研究员以线上形式致辞，感谢全体委员过去一年为专委会发展付出的努力，提及专委会组织的走进高校、走进企业、“视界无限”、讲习班、RACV 等多场精彩纷呈的学术活动，向活动承办单位与组织者表达谢意，并祝愿委员未来收获更多成果。

随后，CCF 专委工委委员、华南理工大学副校长**许勇**教授代表学会发言，对专委会换届后的工作成效给予肯定，特别指出 RACV 系列特色品牌活动在学会内部、兄弟专委会及计算机视觉领域从业者中形成了良好示范并赢得口碑，鼓励委员在专委会带领下继续进步，同时预祝本次工作会议顺利举行。

接下来，专委会秘书长**王瑞平**研究员向与会嘉宾和执行委员做了 2025 年度专委会工作报告。报告简要介

绍了目前专委会的组织结构，概述了常委会议及秘书处工作会议的内容，持续加强组织建设。

他还提到，本年度多位委员获国家级人才称号、国际学会 Fellow 等荣誉，多位委员牵头成果获国家及省部级科技奖，参与完成的论文获得了重要国际会议的最佳论文奖等；随后，他回顾学术交流（走进高校、走进企业、视界无限、RACV、前沿讲习班等）、企业交流与教育工作；还提及专委会简报、网站和公众号等宣传渠道效果显著；最后，王秘书长提出了未来工作计划，将秉承“一切为委员服务”宗旨，扩大活动范围、聚焦主题、开放 RACV 承办方征集，提升委员参与度。

随后，根据会议日程，工作会议进行了执行委员的新增选举工作，由专委会秘书长**王瑞平**研究员主持。本年度共有 66 名计算机视觉领域的学者和企业研究人员申请加入专委会。在委员增选环节，申请人逐一上台介绍了个人情况和亮点工作。经常委会无记名投票，共有 45 人被遴选为新委员，大家纷纷表示将积极参与专委会组织的各项活动，为专委会建设和发展贡献力量。

苏州科技大学**胡伏原**教授作为宣传委员会主席对 2025 CCF 科技创业大赛进行宣讲。提及大赛设多个专项赛与行业赛，同时招募“TEC 推荐人”共建科创生态。

委员建言献策环节由专委会常委、副秘书长、北京工业大学**毋立芳**教授主持。中国石油大学**刘伟锋**委员建议 CCF-CV 专委会能够发挥自身影响力，将国际顶刊、国内顶刊与 PRCV 会议相结合，通过推荐优秀论文、组织专题专栏等方式，助力国内期刊提升学术水平与国际影响力，推动国内计算机视觉学术期刊的发展。专委会常委、百度计算机视觉首席科学家**王井东**研究员建议进一步扩充执行委员名额，吸纳更多来自不同地区、不同单位、不同研究方向的优秀人才加入专委会，以增强专委会的工作力量、提升工作效率与覆盖面。

最后，专委会顾问委员、专委会往届主任、北京大学**查红彬**教授对会议进行了总结。查教授肯定了专委会过去一年取得的成绩，勉励委员们继续努力做出更有影响力的工作，推动专委会在稳步发展道路上迈上新台阶。

之后，查教授向出席会议的特邀嘉宾及全体执行委员表达了感谢。至此，CCF-CV 2025 年度工作会议圆满落幕。

责任编委 马伟

专题综述

二次高斯泼溅：基于二阶几何基元的高质量表面重建

张子钰[1] 黄彬彬[2] 姜翰青[3] 周立阳[3] 项骁骏[3] 申抒含[1]
[1]中国科学院自动化研究所 [2]香港大学 [3]商汤科技

本文是中国科学院自动化研究所、香港大学和商汤科技合作研究的成果，发表于ICCV 2025的工作二次高斯泼溅（Quadratic Gaussian Splatting，QGS）[1]。论文研究的问题是如何设计新型高斯泼溅基元来实现高质量的几何重建以及高保真的渲染。该任务要求使用高斯泼溅模型[7]重建室内外场景，恢复逼真的外观，并提取场景的三角网格。我们针对该问题提出了QGS，QGS用可形变的二次曲面基元（如椭圆、抛物曲面）替代传统静态的平面高斯基元，使单个基元即可刻画复杂曲率，从而在保持渲染效率的同时减少基元数量与显存占用。与以往利用欧氏距离建模密度不同，QGS采用基于测地距离的密度分布，使密度权重随曲率内在自适应，并在形变过程中（例如由平面圆盘过渡到弯曲抛物面）保持一致性。论文给出了二次曲面上测地距离的闭式解，并结合快速的光线—二次曲面相交，实现面感知的泼溅管线：在相同细节下，过去需要数十个平面基元的区域可由极少量二次基元表示，如图1所示。实验在DTU[2]、Tanks and Temples[3]、MipNeRF360[4]等数据集上表明，QGS在几何重建上达到最先进水平：在DTU上相对2DGS[5]的倒角距离降低33%，相对GOF[6]降低27%；同时保持与主流方法相当的外观质量，在几何精度与视觉保真之间取得平衡。消融实验验证了测地距离建模、闭式解以及二次曲面参数化对性能的贡献，并证明该表示与现有的优化框架、正则项和可微渲染器兼容，可直接替换传统平面基元表示；同时支持与法线约束、深度先验和遮挡一致性等多种几何信号联合优化，具有良好扩展性与工程可落地性。潜在局限包括在极度复杂的拓扑或大尺度场景中需要更精细的初始化与层次化调度，未来可结合自适应分辨率与学习型参数化进一步提升效率与泛化能力。

一、研究背景

近年来，随着人工智能与传感技术的快速迭代，三维重建已成为计算机视觉与图形学交叉领域的前沿课

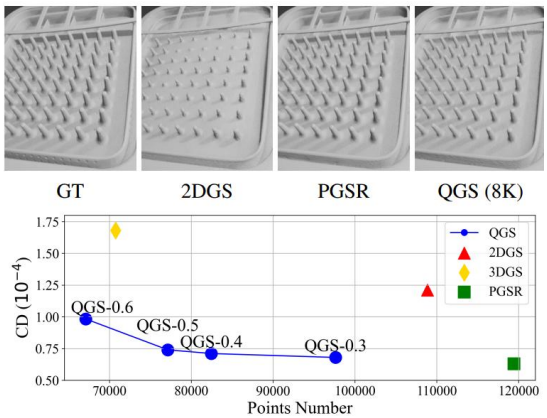


图 1　本文所提出的二次曲面高斯泼溅与平面基元方法的对比。结果表明，QGS 在使用更少基元的前提下获得更高的重建精度，尤其在高曲率区域能够恢复更连续、更贴合的表面；同时该表示具备良好的可扩展性与应用潜力。

题，其核心目标是在真实场景中重建兼具高几何精度与视觉真实感的三维模型。高斯泼溅方法[7]近年在室内外建模中展现出广泛适用性与优异效率，但多数工作将高斯基元建模为椭圆平面圆盘。作为一阶几何近似，平面圆盘的表达能力受限：面对高曲率结构，需要显著增密基元采样，既加重显存与存储开销，也带来渲染时延与优化不稳定等问题。

本文将椭圆圆盘的表示扩展为更一般的二次型框架。具体而言，我们用隐函数表示基元为曲面形式：

$$f(x,y,z) = (x,y,z,1)^T Q(x,y,z,1) = 0$$

其中Q描述了曲面的形状，当Q的惯性指数为$(1,1,0,0)$时，隐函数表达为平面结构；当惯性指数为$(1,1,-1,0)$时，隐函数表达为双曲面结构。因此该表示推广了三维高斯椭球基元，并将二维高斯圆盘作为特例纳入，同步实现对更高阶曲面的自适应。

二、QGS方法介绍

2.1 前言：高斯泼溅

Kerbl[7]等人提出使用三维高斯椭球来表达场景，并使用体渲染泼溅算法渲染图像：

$$C(\mathbf{p}) = \sum_{i=0}^{N-1} G_i(\mathbf{p})\alpha_i c_i \prod_{j=0}^{i-1}(1 - G_i(\mathbf{p})\alpha_j)$$

其中，α_i表示基元的不透明度，c_i表示高斯基元通过球谐系数（Spherical Harmonic, SH）建模的颜色，$\mathbf{p}$是像素坐标。最终，通过光度一致性损失优化高斯椭球。

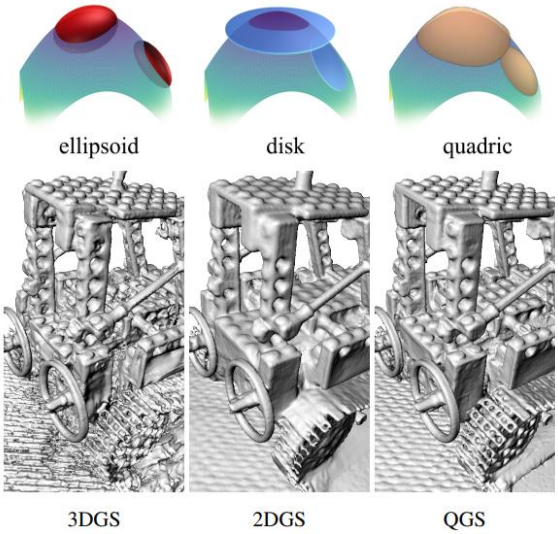


图 2　不同基元拟合表面示意图以及重建结果

随后，2DGS[5]提出使用二维高斯圆盘表征场景，并在引入法线一致性正则化损失后，结合深度图融合算法能够提取场景的三角网格。然而，平面圆盘面元仅仅是对场景表面的一阶线性近似，在高曲率区域容易产生过度平滑的重建，如图 2 所示。为了提升基元的几何拟合能力，我们提出使用将平面面元推广为二次曲面基元。我们接下来分三节介绍：其一，提出统一的二次曲面表征形式，采用二次型隐式参数化并兼顾可微渲染；其二，在二次曲面上构建基于测地距离的高斯分布，使密度与局部曲率一致，避免欧氏度量导致的失真；其三，设计面向重建与渲染的联合优化策略，实现几何与外观质量的协同提升。

2.2 二次曲面表征

给定齐次空间下的齐次坐标点$\mathbf{x} = [x,y,z,1]^T \in \mathbb{R}^3$一个空间的二次曲面可表示为：

$$\begin{aligned} f(x,y,z) &= [x,y,z,1]\begin{bmatrix} A & B & C & D \\ B & E & F & G \\ C & F & H & I \\ D & G & I & J \end{bmatrix}\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \\ &= \hat{\mathbf{x}}^T \begin{bmatrix} \frac{d_{11}}{s_1^2} & 0 & 0 & 0 \\ 0 & \frac{d_{22}}{s_2^2} & 0 & 0 \\ 0 & 0 & 0 & -\frac{d_{33}}{2s_3} \\ 0 & 0 & -\frac{d_{33}}{2s_3} & 0 \end{bmatrix} \hat{\mathbf{x}} \\ &= \frac{d_{11}}{s_1^2}\hat{x}^2 + \frac{d_{22}}{s_2^2}\hat{y}^2 - \frac{d_{33}}{s_3}\hat{z} = 0 \end{aligned}$$

其中 $\hat{}$ 表示高斯局部坐标系（参数坐标系）下的坐标。式中$d_{ii} \in \{0, \pm 1\}$用于控制抛物面的类型，决定其为椭圆抛物面、双曲抛物面或平面。然而，由于d_{ii}取值为离散变量，该表示形式无法在椭圆抛物面与双曲抛物面之间实现连续过渡。为此，我们在该参数上引入带符号的连续尺度因子：

$$f(x,y,z) = \frac{\text{sign}(s_1)}{s_1^2}\hat{x}^2 + \frac{\text{sign}(s_2)}{s_2^2}\hat{y}^2 - \frac{1}{s_3}\hat{z} = 0$$

通过构建符号因子：$s(x,t) = \tanh(t) \cdot \exp(x)$，使得基元尺度既能在正负方向上连续变化，又能保持可微性，从而实现椭圆抛物面与双曲抛物面之间的平滑过渡。

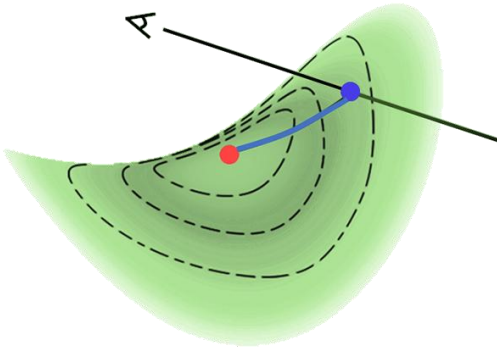


图 3　基于测地度量的高斯分布示意图

2.3 基于测地度量的高斯分布

将高斯分布的中心定义在二次曲面的中心，则曲面上任一点到中心的测地距离可定义为一条二次曲线的弧线长，如图 3 中蓝色粗线所示：

$$l(a,\rho_0)=\int_0^{\rho_0}\sqrt{1+(2at)^2}\,dt$$
$$=\frac{\ln\left(\sqrt{(2a\rho_0)^2+1}+2a\rho_0\right)+2a\rho_0\sqrt{(2a\rho_0)^2+1}}{4a}$$

其中a为该点所在二次曲线的二次项系数，ρ_0为该点所在极坐标系下的极线长度。然后，我们定义该点所在方向上的高斯分布标准差为：

$$\sigma_0(\theta_0)=\frac{s_1s_2}{\sqrt{(s_2\cos\theta_0)^2+(s_1\sin\theta_0)^2}}$$

于是，对于曲面上任一点，我们可以定义其高斯分布函数值为：

$$G\left(\hat{\mathbf{p}}_0(\theta_0,\rho_0)\right)=\exp\left(-\frac{\left(l(a(\theta_0),\rho_0)\right)^2}{2\left(\sigma(\theta_0)\right)^2}\right)$$

需要注意的是，当$|s_3|\to 0$时，二次抛物面的表达形式将退化为一个平面。进一步地，根据极限关系$\ln(1+x)\sim x$（当$x\to 0$时成立）。可知在$|s_3|\to 0$的情况下，有$a\to 0$，从而测地距离$l\to\rho_0$，这意味着，当抛物面逐渐平坦化时，曲面上的测地距离将渐近于欧氏距离。换言之，传统的 2DGS 可被视为本方法在平面极限下的一种退化形式。

2.4 对于二次曲面基元的优化策略

在传统 3DGS[7]与 2DGS[5]中，高斯椭球与高斯圆盘均属于封闭式几何表达，因此可以方便地在图像空间

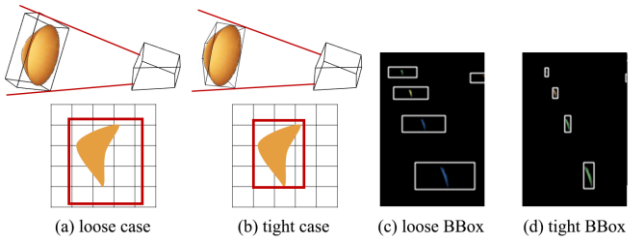


图 4　矩形包围盒与截台包围盒对比图

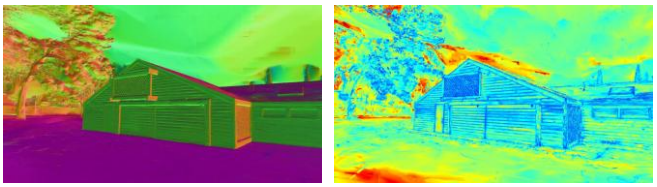


图 5　QGS 渲染的法线图（左）和曲率图（右）曲率图中颜色越偏蓝表示曲率越大，表面越弯曲。

中计算其主轴对应的包围盒。然而，本研究所提出的二次曲面基元包含凹凸曲面及鞍面等开放式形态，使得其在预处理阶段的可见区域计算变得更加复杂。为此，我们提出了一种截台型的包围盒，以减少无用的渲染区域，如图 4 所示。相比简单的矩形包围盒，截台包围盒能提升两倍的渲染速度。

另一方面，由于二次曲面是一种对场景的二阶拟合，因此除了常规的颜色、深度以及法线外，QGS 也能渲染出场景的曲率信息。为了简化说明，二次曲面的表达方程可以写为 $\hat{z}=\lambda_1\hat{x}^2+\lambda_2\hat{y}^2$，则二次曲面上任意一点的高斯曲率可解析算出：

$$\hat{K}_0(\hat{\mathbf{p}}_0)=\frac{4\lambda_1\lambda_2}{(1+4\lambda_1^2\hat{x}_0^2+4\lambda_2^2y_0^2)^2}$$

通过阿尔法混合技术，我们可以对给定视角渲染法线图与曲率图，如图 5 所示。

$$\{\mathbf{N},\mathbf{K}\}=\sum_{i=0}^{N-1}G_i\alpha_i\prod_{j=0}^{i-1}(1-G_j\alpha_j)\{\mathbf{n}_i,\hat{K}_i\}$$

图 6　传统质心排序（左）、逐瓦片排序（中）与逐像素重排序（右）机制下的法线图对比结果

CD (mm)↓	24	37	40	55	63	65	69	83	97	105	106	110	114	118	122	Mean	Time
NeuS	1.00	1.37	0.93	0.43	1.10	0.65	0.57	1.48	1.09	0.83	0.52	1.20	0.35	0.49	0.54	0.84	>12h
VolSDF	1.14	1.26	0.81	0.49	1.25	0.70	0.72	1.29	1.18	0.70	0.66	1.08	0.42	0.61	0.55	0.86	>12h
Neuralangelo	0.37	0.72	0.35	0.35	0.87	0.54	0.53	1.29	0.97	0.73	0.47	0.74	0.32	0.41	0.43	0.61	>128h
3DGS	2.14	1.53	2.08	1.68	3.49	2.21	1.43	2.07	2.22	1.75	1.79	2.55	1.53	1.52	1.50	1.96	11.2min
Gaussian surfels	0.66	0.93	0.54	0.41	1.06	1.14	0.85	1.29	1.53	0.79	0.82	1.58	0.45	0.66	0.53	0.88	6.7min
SuGaR	1.47	1.33	1.13	0.61	2.25	1.71	1.15	1.63	1.62	1.07	0.79	2.45	0.98	0.88	0.79	1.33	1h
2DGS	0.48	0.91	0.39	0.39	1.01	0.83	0.81	1.36	1.27	0.76	0.70	1.40	0.40	0.76	0.52	0.80	19.2min
GOF	0.50	0.82	0.37	0.37	1.12	0.74	0.73	1.18	1.29	0.68	0.77	0.90	0.42	0.66	0.49	0.74	1h
GSDF	0.59	0.94	0.46	0.38	1.30	0.77	0.73	1.59	1.29	0.76	0.59	1.22	0.38	0.52	0.51	0.80	32min
GS-pull	0.51	0.56	0.46	0.39	0.82	0.67	0.85	1.37	1.25	0.73	0.54	1.39	0.35	0.88	0.42	0.75	22min
Ours w/o MV	0.46	0.76	0.40	0.38	0.92	0.80	0.76	1.25	0.95	0.67	0.62	1.20	0.38	0.60	0.47	0.71	25min
PGSR	0.40	0.60	0.39	0.37	0.78	0.59	0.53	1.18	0.67	0.63	0.48	0.62	0.34	0.42	0.39	0.56	40min
Ours	0.38	0.62	0.37	0.38	0.75	0.55	0.51	1.12	0.68	0.61	0.46	0.58	0.35	0.41	0.40	0.54	48min

表 1　神经辐射场方法与高斯泼溅方法在 DTU 数据集上的几何定量对比，其中前三行为辐射场方法，后十行为高斯泼溅方法，最优结果用红色标出，次优结果用棕色标出，次次优结果用黄色标出。

在深度渲染方面，2DGS 提出使用中值深度进行深度图融合，即在进行阿尔法混合时，选取透射率达到 0.5 时对应的高斯基元交点深度作为像素的最终深度。该策略能够在一定程度上缓解体渲染中的深度偏差问题，从而获得更为准确的几何深度。然而，中值深度对渲染顺序极为敏感。当同一瓦片下的所有像素共享相同的渲染顺序，而各像素在混合过程中对应的中值深度不同步时，会导致多视角下的深度结果不一致，表现为条纹状伪影，如为解决这一问题，我们进一步拓展了 StopThePop[8] 中的逐瓦片排序与逐像素重排序机制到基于二次曲面基元的渲染管线中。具体而言，对于每个16×16像素的瓦片，我们选取距离二次曲面中心点投影点最近的像素所对应的视线，计算其与曲面的交点深度，并将该深度作为该瓦片的全局深度，用于后续瓦片级的全局排序。如图 6 所示，该策略能够有效消除条纹状不一致现象，但由于每个瓦片仅采用一条光线进行近似，仍会引入轻微的块状伪影。

为进一步提升一致性，我们借鉴 StopThePop 的思路，在瓦片排序的基础上增加了逐像素的局部重排序。具体做法是：在计算每个高斯基元的深度、法线及其他属性后，不是立即进行阿尔法混合，而是将这些属性暂存于长度为 8 的缓冲区中。当缓冲区填满后，从中选择距离相机最近的高斯进行混合更新，从而保证局部排序的精确性。同时为了在反向传播时保持与前向一致的传播顺序，我们将由远到近的梯度递归公式改写为由近到远的梯度递归公式：

$$\frac{\partial \hat{X}}{\partial \bar{\alpha}_i} = \left(X_i - \frac{\hat{X} - \sum_{j=0}^{i} X_j \bar{\alpha}_j T_j}{T_{i+1}} \right) T_i$$

最后，我们使用光度一致性损失、法线一致性损失、深度畸变损失和多视一致性正则化损失进行优化：

$$\mathcal{L} = \mathcal{L}_c + \lambda_d \mathcal{L}_d + \lambda_n \mathcal{L}_{Kn} + \lambda_{Mv} \mathcal{L}_{Mv}$$

三、实验结果

3.1 定量对比

本方法的完整渲染与优化管线均在 CUDA 核函数上实现。实验在三个主流公开数据集上进行，包括

F1-Score ↑	Geo-NeuS	N-angelo	2DGS	GOF	Ours w/o MV	PGSR	Ours
Barn	0.33	0.70	0.41	0.51	0.46	0.52	0.55
Caterpillar	0.26	0.36	0.24	0.41	0.32	0.38	0.40
Courthouse	0.12	0.28	0.16	0.28	0.26	0.26	0.28
Ignatius	0.72	0.89	0.52	0.68	0.79	0.77	0.81
Meetingroom	0.20	0.32	0.17	0.28	0.25	0.29	0.31
Truck	0.45	0.48	0.45	0.58	0.60	0.62	0.64
Mean	0.35	0.50	0.33	0.46	0.45	0.47	0.50
Time	>24h	>127h	34min	114min	43min	66min	75min

表 2　神经辐射场方法与高斯泼溅方法在 TNT 数据集上的几何定量对比

	Indoor scenes			Outdoor scenes		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
NeRF	26.84	0.790	0.370	21.46	0.458	0.515
Deep Blending	26.40	0.844	0.261	21.54	0.524	0.364
i-NGP	29.15	0.880	0.216	22.90	0.566	0.371
Mip-NeRF360	31.72	0.917	0.180	24.47	0.691	0.283
3DGS	30.52	0.921	0.199	24.45	0.728	0.240
SuGar	29.44	0.911	0.216	22.76	0.631	0.349
2DGS	30.39	0.924	0.182	24.33	0.709	0.284
GOF	30.80	0.928	0.167	24.76	0.742	0.225
Ours w/o MV	30.48	0.926	0.166	24.56	0.724	0.239
PGSR	30.35	0.924	0.176	24.29	0.718	0.236
Ours	30.45	0.919	0.184	24.32	0.706	0.242

表 3　神经辐射场方法与高斯泼溅方法在 Mip-NeRF 360 数据集上的外观定量对比

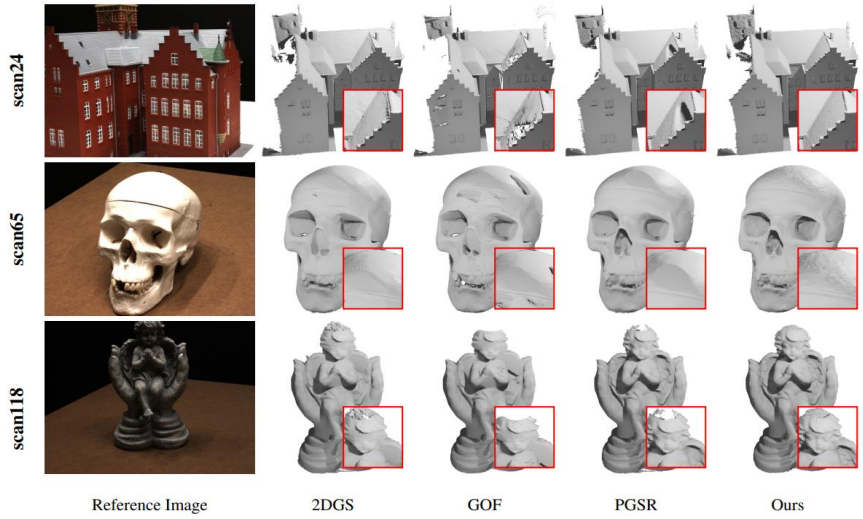


图 7　本方法与先进表面重建方法在 DTU 数据集上的定性对比结果

DTU[2]、Tanks and Temples[3]以及 Mip-NeRF 360[4]，涵盖了从室内小场景到复杂室外大场景的多种重建场景类型。在定量评估中，我们采用几何精度指标 Chamfer Distance (CD)与 F1 Score 来衡量重建网格的几何质量，同时使用外观指标 PSNR、SSIM 与 LPIPS 评估渲染图像的感知质量与一致性。为了全面验证本方法的有效性，我们与当前主流的神经辐射场类方法以及高斯泼溅类方法进行了系统对比。为了公平地比较曲面面元与平面面元的优劣，我们在所有的实验中都对比了使用或者不使用多视几何一致正则化的二次曲面基元方法，以全面地评估。我们将不使用多视几何一致正则化的方法记为 Ours w/o MV，使用了的记为 Ours。

在表 1 中，我们在 DTU 数据集上使用 CD 定量对比了本方法与神经辐射场类方法以及高斯泼溅类方法。本方法在 DTU 数据集上全面超过了以往的方法，并且训练时间也在前列。进一步地，即使在不使用多视几何一致正则化的方法中，本方法也超过了其他方法并有着有竞争力的训练速度。在表 2 中，我们在 TNT 数据集上通过 F1 Score 定量对比了本方法与其他方法。在表 3 中，我们在 Mip-NeRF 360 数据集上通过 PSNR，SSIM，LPIPS 评比了本方法与其他方法。在应用了多视几何一致正则后，本方法的渲染效果会有下降，但与其他方法相比，本方法依然能保持前列。

3.2 定性对比

图 7 展示了本方法和最先进高斯泼溅类方法的定性比较，结果显示本方法所提出的二次曲面基元能拟合更复杂的几何细节，并对弱纹理的高曲率区域有更佳的重

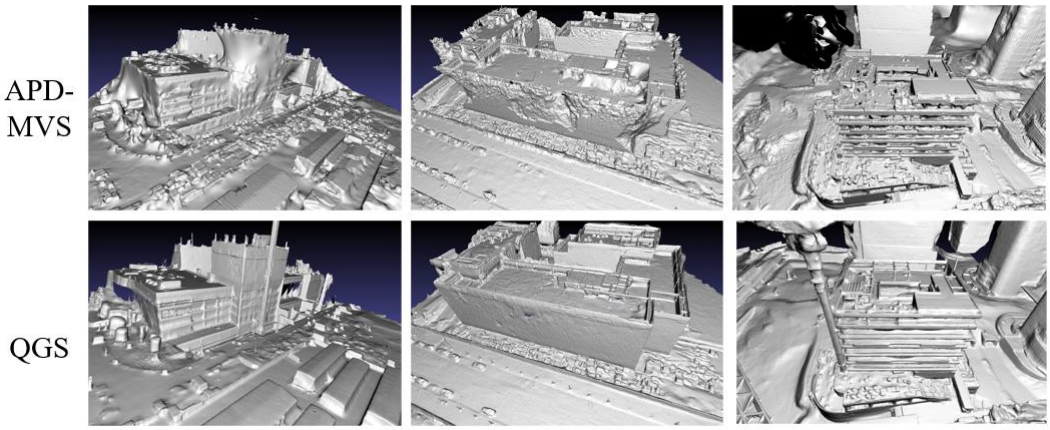


图 8　本方法与几何方法 APD-MVS[9]在室外建筑场景的定性对比结果

建效果。图 8 展示了在大规模复杂场景下，本方法与传统几何方法 APD-MVS[9]的定性对比，结果说明本方法对弱纹理、细长区域的重建更完整，且整体网格更平滑。

四、总结

本文提出二次曲面高斯泼溅（Quadratic Gaussian Splatting, QGS），作为高斯泼溅方法的扩展，用于精确重建场景几何并恢复细节结构。QGS 首次在高斯泼溅框架中引入二次曲面（quadric surfaces），并在非欧氏几何空间上定义高斯分布，以提升对二阶曲率的刻画与拟合能力，从而加强对复杂曲面的表达。基于多组室内与室外数据集的实验表明，所提出方法在几何重建方面达到当前最优水平，同时在渲染质量上保持具有竞争力的表现。

同时，本方法也以较大优势取得了 CAD/CG 2025 建筑场景高精度三维重建挑战赛的冠军。最后，本文所提出的 QGS 方法已开源于：https://github.com/will-zzy/QGS。

责任编委 崔海楠

参考文献

[1] Ziyu Zhang, Binbin Huang, Hanqing Jiang, Liyang Zhou, Xiaojun Xiang, Shuhan Shen. Quadratic Gaussian Splatting: High Quality Surface Reconstruction with Second-order Geometric Primitives. International Conference on Computer Vision, ICCV 2025.

[2] R. Jensen, A. Dahl, G. Vogiatzis, E. Tola, and H. Aanæs, 'Large scale multi-view stereopsis evaluation', in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 406–413.

[3] A. Knapitsch, J. Park, Q.-Y. Zhou, and V. Koltun, 'Tanks and temples: Benchmarking large-scale scene reconstruction', *ACM Transactions on Graphics (ToG)*, vol. 36, no. 4, pp. 1–13, 2017.

[4] J. T. Barron, B. Mildenhall, D. Verbin, P. P. Srinivasan, and P. Hedman, 'Mip-nerf 360: Unbounded anti-aliased neural radiance fields', in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 5470–5479.

[5] B. Huang, Z. Yu, A. Chen, A. Geiger, and S. Gao, '2d gaussian splatting for geometrically accurate radiance fields', in *ACM SIGGRAPH 2024 conference papers*, 2024, pp. 1–11.

[6] Z. Yu, T. Sattler, and A. Geiger, 'Gaussian opacity fields: Efficient and compact surface reconstruction in unbounded scenes', *arXiv preprint arXiv:2404. 10772*, 2024.

[7] B. Kerbl, G. Kopanas, T. Leimkühler, and G. Drettakis, '3D Gaussian splatting for real-time radiance field rendering', *ACM Trans. Graph.*, vol. 42, no. 4, pp. 139–131, 2023.

[8] L. Radl, M. Steiner, M. Parger, A. Weinrauch, B. Kerbl, and M. Steinberger, 'Stopthepop: Sorted gaussian splatting for view-consistent real-time rendering', *ACM Transactions on Graphics (TOG)*, vol. 43, no. 4, pp. 1–17, 2024.

[9] Y. Wang *et al.*, 'Adaptive patch deformation for textureless-resilient multi-view stereo', in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2023, pp. 1621–1630.

张子钰

中国科学院自动化研究所 23 级博士研究生，导师为申抒含研究员，主要研究方向为高斯泼溅。
Email:zhangziyu2021@ia.ac.cn

黄彬彬

现为香港大学计算机科学系博士研究生，师从高盛华教授。他的研究兴趣是计算机视觉和基础模型，近几年研究工作包括高斯泼溅（2DGS、Mip-Splatting），前馈神经网络（CUPID）。
Email: binbinhuang@connect.hku.hk

申抒含

2003 年和 2006 年分别获西南交通大学学士与硕士学位，2010 年获上海交通大学博士学位。现为中国科学院自动化研究所教授、中国科学院大学人工智能学院岗位教授。其主要研究方向为三维计算机视觉，重点涉及大规模场景三维重建、智能机器人三维感知以及三维语义重建等。在计算机视觉、摄影测量与机器人领域的重要期刊与会议上发表论文 100 余篇。IEEE 高级会员。曾获 2016 年 ACM 北京新星奖、2018 年中国图像图形学会科学技术二等奖、2023 年中国自动化学会自然科学一等奖、2023 年中国测绘学会科学技术一等奖、2024 年中国自动化学会科技进步一等奖等。
Email: shshen@nlpr.ia.ac.cn

热点追踪

模态感知学习与大小模型协同的多模态虚假新闻检测

北京工业大学 高祎菡 王博岳 鲍可馨 吴广超

本文主要介绍北京工业大学在多模态虚假新闻检测领域的研究成果，包括发表在IEEE Transactions on Neural Networks and Learning Systems (TNNLS) 2025的工作MoPeD[1]和最新工作进展TNNLS在审论文ReCoLV，代表了该领域在模态感知学习和大小模型协同方面的最新研究进展。

一、研究背景

在数字技术飞速发展的时代，在线社交媒体和互联网使得信息的传播变得空前便捷和迅速。然而这种便利性也为虚假新闻的广泛传播提供了可乘之机。虚假新闻，特别是融合了煽动性文本和伪造、无关图像的多模态虚假新闻，因其更强的迷惑性和情感冲击力，严重威胁着公共安全、误导社会舆论并破坏社会信任。因此开发高效、准确的多模态虚假新闻检测技术，已成为学术界和工业界亟需解决的关键挑战。

早期的虚假新闻检测研究主要集中在单模态分析。文本检测方法专注于分析语言学特征，例如夸张的文风、情感极性或特定的句法结构，并涌现了如 DRNN[3]、MDFEND[4]等模型。视觉检测方法则侧重于图像取证，通过分析像素域或频域特征来识别篡改痕迹，如MVNN[5]。然而，单模态方法存在根本局限，它们无法捕捉模态间的交互信息，尤其是虚假新闻中常见的图文不符或语义矛盾，而这恰恰是甄别高级谎言的关键线索。

为了克服这一瓶颈，研究重心转向了多模态检测。早期的多模态方法主要依赖特征融合。例如，SpotFake[6]和 EANN[7]将使用 VGG 或 BERT 提取的视觉和文本特征进行简单拼接或对齐，以进行联合分类。但这种浅层融合策略难以有效弥合不同模态间固有的“语义鸿沟”。随之而来的是基于跨模态关联的方法，它们通过更复杂的机制，如注意力，来显式地建模图文关系。例如，MCAN[8]使用共注意力网络来捕捉跨模态的相互依赖；CAFE[9]和 COOLANT[10]则进一步通过量化模态间的歧义性或使用对比学习来识别不一致性。

尽管这些方法取得了显著进展，但现有的多模态检测框架仍面临两大核心挑战。其一，是模态异质性（Heterogeneity）问题。现有模型大多平等地对待所有模态，但忽视了在特定样本中，决定性线索（Determinative Factor）可能仅存在于单一模态或者跨模态的交互特征中。模型缺乏一种自适应机制来动态感知并放大那个最关键的“破绽”来源。其二，是缺乏深度推理与知识验证能力。当前的数据驱动模型本质上仍停留在特征匹配和关联感知的层面，它们无法像人类一样，调用常识和外部知识来评估新闻内容的合理性。本文分别针对以上两种问题介绍该领域的两个工作。

图 1　多模态新闻示例阐释 MoPeD

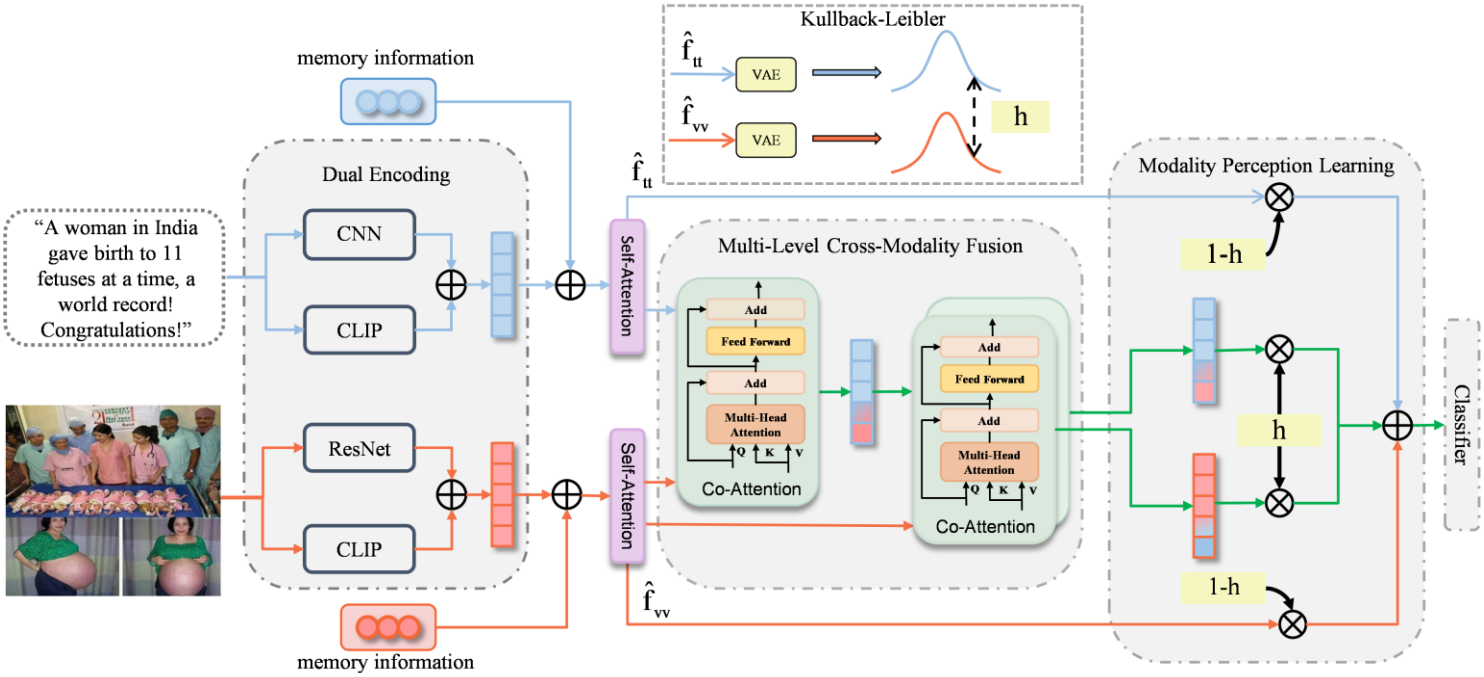


图 2　MoPeD 模型的整体架构

二、基于模态感知学习的决定性因素发现模型 MoPeD

2.1 模型简介

MoPeD (Modality Perception Learning-based Determinative Factor Discovery) 模型旨在解决现有方法在处理模态异质性、挖掘新闻中隐藏的决定性信息方面的不足。在真实的虚假新闻中，往往只有一小部分信息是具有欺骗性的，例如图 1 中“女子一次产下 11 胞胎”的假新闻，其图像与文本内容在语义上是高度一致的，但其文本内容本身因违背常识而暴露出虚假性；而在其他新闻中，则可能是图像存在明显篡改痕迹，或是图文内容毫不相关。MoPeD 的核心思想是设计一个能够自适应地感知并放大这些决定性因素的框架，无论它们存在于单个模态内部还是跨模态的交互特征中。该模型通过对特征提取、融合与聚合的全流程进行优化，实现了对决定性信息的精准捕捉，其主要能力包括：

(1) 全面的特征提取：通过创新的双编码模块，同时捕获模态特有（modal-specific）的细节信息和模态一致（modal-consistent）的语义信息。

(2) 深度的跨模态融合：受人类反复对比图文以理解复杂信息的启发，设计了多层次的跨模态融合模块，以充分挖掘图文间的深层隐式关联。

(3) 自适应的特征聚合：引入模态感知学习模块，通过动态计算图文内容的异质性得分，自适应地调整单模态特征与跨模态融合特征的权重，从而凸显在当前样本中最具判别力的信息。

2.2 模型架构

MoPeD 模型的整体架构如图 2 所示，其核心由三个创新的模块构成，旨在逐步揭示并利用多模态新闻中的决定性线索。

模块 1：双编码器（Dual Encoding）

为了全面地提取信息，MoPeD 摒弃了使用单一编码器的传统做法。它创新性地结合了**模态特有编码器**（如用于文本的 CNN 和用于图像的 ResNet）和**模态一致编码器**（预训练的 CLIP 模型）。前者专注于挖掘各个模态独特的表征，如文本的语言风格和图像的视觉伪影；后者则利用 CLIP 强大的跨模态对齐能力，将图文特征映射到统一的语义空间，有效缩减了模态间的语义鸿沟。此外，模型还引入了可学习的“记忆信息”向量 (memory information)，该向量在训练过程中不断更新，用于学习和调整单模态特征表示，从而捕捉数据集中关于真假新闻的潜在模式。

	Method	Accuracy	Precision	Recall	F1-score	Fake News			Real News		
						Precision	Recall	F1-score	Precision	Recall	F1-score
Weibo	EANN	0.770	0.770	0.770	0.770	0.774	0.757	0.766	0.766	0.782	0.774
	MVAE	0.715	0.723	0.715	0.712	0.765	0.616	0.682	0.682	0.812	0.742
	SpotFake	0.767	0.772	0.767	0.766	0.734	0.834	0.780	0.810	0.701	0.752
	SpotFake+	0.731	0.732	0.731	0.731	0.716	0.760	0.737	0.748	0.702	0.725
	SAFE	0.812	0.823	0.812	0.810	0.884	0.715	0.790	0.763	**0.907**	0.829
	CARMN	0.825	0.826	0.825	0.825	0.837	0.816	0.826	0.814	0.835	0.824
	MFAN	0.844	0.845	0.844	0.844	0.863	0.816	0.839	0.828	0.872	0.849
	CAFE	0.849	0.848	0.849	0.849	**0.898**	0.810	0.852	0.798	0.891	0.842
	COOLANT	0.843	0.843	0.843	0.843	0.826	0.853	0.839	0.859	0.834	0.846
	MoPeD	**0.883**	**0.883**	**0.883**	**0.883**	0.879	**0.887**	**0.883**	**0.887**	0.878	**0.883**
Weibo-19	EANN	0.827	0.831	0.827	0.828	0.760	0.819	0.788	0.876	0.832	0.854
	MVAE	0.722	0.718	0.722	0.719	0.663	0.595	0.627	0.754	0.804	0.778
	SpotFake	0.712	0.711	0.712	0.696	0.707	0.459	0.555	0.714	0.877	0.787
	SpotFake+	0.722	0.718	0.722	0.718	0.667	0.586	0.624	0.751	0.810	0.780
	SAFE	0.854	0.855	0.854	0.855	0.807	0.828	0.817	0.886	0.872	0.879
	CARMN	0.864	0.864	0.864	0.864	0.845	0.817	0.831	0.877	0.897	0.887
	MFAN	0.878	0.880	0.878	0.878	0.828	0.871	0.849	0.913	0.883	0.898
	CAFE	0.898	0.899	0.898	0.898	0.864	0.879	0.872	0.921	0.911	0.916
	COOLANT	0.881	0.881	0.881	0.881	0.853	0.846	0.850	0.899	0.904	0.902
	MoPeD	**0.929**	**0.929**	**0.929**	**0.929**	**0.899**	**0.922**	**0.911**	**0.949**	**0.933**	**0.941**
Pheme	EANN	0.782	0.774	0.782	0.777	0.649	0.558	0.600	0.826	0.875	0.850
	MVAE	0.792	0.784	0.792	0.784	0.685	0.540	0.604	0.824	0.897	0.859
	SpotFake	0.769	0.758	0.769	0.760	0.636	0.496	0.557	0.808	0.882	0.844
	SpotFake+	0.823	0.822	0.823	0.822	0.706	0.681	0.694	0.870	0.882	0.876
	SAFE	0.813	0.807	0.813	0.803	0.753	0.540	0.629	0.829	0.926	0.875
	CARMN	0.805	0.831	0.805	0.814	0.549	0.721	0.623	0.912	0.829	0.869
	MFAN	0.875	0.882	0.875	0.877	0.756	**0.850**	0.800	0.934	0.886	0.909
	CAFE	0.891	0.891	0.892	0.891	0.826	0.796	0.810	0.917	0.930	0.923
	COOLANT	0.894	0.895	0.894	0.894	0.796	0.833	0.814	**0.934**	0.917	0.925
	MoPeD	**0.904**	**0.903**	**0.904**	**0.903**	**0.851**	0.814	**0.833**	0.924	**0.941**	**0.932**

表 1　Weibo、Weibo-19 和 Pheme 数据集上 MoPeD 与最先进多模态假新闻检测方法的对比

模块 2: 多层次跨模态融合 (Multi-level Cross-Modality Fusion)

仅仅提取出高质量的单模态特征是不足够的，关键在于如何有效地融合它们以发现跨模态的线索。MoPeD 设计了一个包含两个层级的跨模态协同注意力融合模块。第一层融合模拟了人们“看图识文”的过程，即利用视觉特征来增强文本特征的理解。第二层融合则更为深入，它将第一层得到的视觉增强的文本特征与原始视觉特征再次进行双向交互。这种渐进式的深度融合机制，能够有效捕捉图文内容之间复杂的、非显性的关联，例如讽刺、暗示或矛盾等隐式信息，而这些信息往往是判断新闻真伪的关键。

模块 3: 模态感知学习 (Modality Perception Learning)

在得到单模态特征和跨模态融合特征后，如何确定哪部分信息在当前样本的判别中更重要，是一个核心问题。例如，当图文内容高度不一致时，跨模态特征可能比独立的单模态特征更有价值。为此，MoPeD 提出了模态感知学习模块。该模块首先利用变分自编码器(VAE)学习文本和视觉特征的分布，然后通过计算这两个分布之间的 KL 散度 (Kullback-Leibler divergence) 来量化它们的异质性得分。这个得分作为一个动态权重，用于自适应地调整最终用于分类的特征组合。当异质性得分较高，即图文差异大时，模型会更侧重于跨模态融合特征；当得分较低，即图文内容一致时，则更侧重于单模态特征，因为此时的决定性因素可能隐藏在单一模态的细节中，如文本的夸张表述。

2.3 实验结果

性能比较： 在三个广泛使用的公共数据集 Weibo[11]、Weibo-19[12]和 Pheme[13]上对 MoPeD 进行了全面评估。评估指标包括准确率、精确率、召回率和 F1 分数。如表 1 所示，MoPeD 在所有数据集上的准确率均显著优于当时的 SOTA 方法，包括 EANN[7]、MVAE[14]、Spotfake[6]、Spotfake+[17]、CARMN[15]、SAFE[16]、MFAN[2]、CAFE[9]和 COOLANT[10]。

	Method	Accuracy	F1-score	
			Fake News	Real News
Weibo	MoPeD	**0.883**	**0.883**	0.883
	w/o P	0.863	0.859	0.866
	w/o U	0.862	0.852	0.87
	w/o C	0.831	0.817	0.842
	w/o F	0.871	0.868	0.875
	w/o M	0.882	0.878	**0.886**
	w/o S	0.877	0.871	0.883
	w/o C+M	0.828	0.818	0.836
Weibo-19	MoPeD	**0.929**	**0.911**	**0.941**
	w/o P	0.909	0.873	0.928
	w/o U	0.898	0.870	0.916
	w/o C	0.868	0.843	0.886
	w/o F	0.919	0.896	0.933
	w/o M	0.922	0.901	0.936
	w/o S	0.902	0.874	0.919
	w/o C+M	0.847	0.816	0.870
Pheme	MoPeD	**0.904**	**0.833**	**0.932**
	w/o P	0.896	0.821	0.93
	w/o U	0.885	0.804	0.919
	w/o C	0.852	0.776	0.889
	w/o F	0.862	0.780	0.9
	w/o M	0.899	0.825	0.929
	w/o S	0.855	0.786	0.890
	w/o C+M	0.844	0.725	0.891

表 2　基于三个数据集的 MoPeD 架构设计消融实验

MoPeD 在 Weibo、Weibo-19 和 Pheme 数据集上分别达到 0.883、0.929 和 0.904 的最高准确率。这些结果相较于最先进方法，分别展现出 3.4%、3.1%和 1.0%的性能优势。此外 MoPeD 在所有测试的准确率、精确率、召回率和 F1 分数指标中均位列第一或第二，彰显了其在虚假新闻检测中的有效性。

消融实验：消融研究如表 2 所示进一步验证了模型设计的有效性。当移除模态感知学习模块 (w/o P) 时，

图 3　虚假新闻示例与其背景验证

模型性能在所有数据集上均出现显著下降，证明了自适应感知决定性因素的必要性。同时，移除 CLIP 模态一致编码器（w/o C）导致的性能下降幅度最大，这表明结合预训练模型的共享语义空间对于弥合模态鸿沟至关重要。

三、融合大小模型协同与反思性总结的类人多模态虚假新闻检测模型 ReCoLV

3.1 背景介绍

尽管像MoPeD这样的模型在特征挖掘和融合方面取得了显著进展，但它们仍然主要依赖于数据驱动的模式识别，缺乏对新闻内容进行背景知识验证、逻辑推理

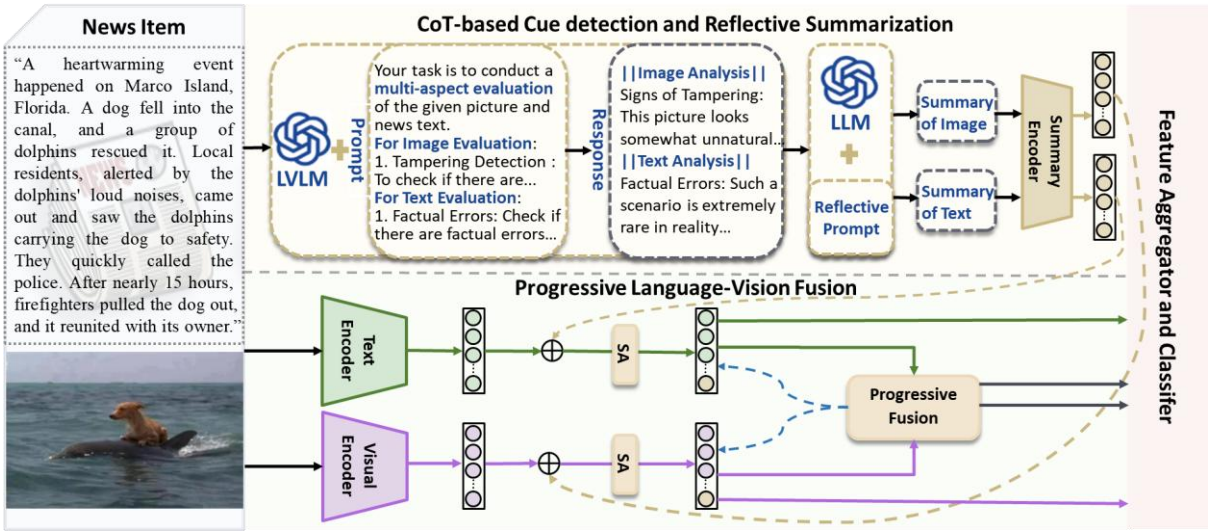


图 4　ReCoLV 模型的整体架构

和常识判断的能力，例如，模型无法主动验证如图3所示的“埃菲尔铁塔为巴基斯坦点亮绿灯”这一新闻事件的真实背景。然而人类在辨别虚假新闻时，往往会执行一个“分析-反思-总结”的认知过程：首先分析新闻中的各种线索，然后结合自身的知识和经验进行反思与验证，最后形成一个简洁的判断结论。这种基于推理的深度分析能力是当前大多数模型所欠缺的。

幸运的是，大型视觉语言模型（Large Vision-Language Models, LVLM），如Qwen-VL-Max，凭借其海量的预训练知识和强大的推理能力，为弥补这一差距提供了可能。它们能够理解复杂的上下文，检索外部知识，并进行逻辑推理，这与虚假新闻检测的核心挑战高度契合。基于此，提出了一个名为ReCoLV (Reflective Summarization and Large-Small Model Collaboration) 的创新框架，它通过协同一个大型视觉语言模型和一个轻量级检测模型，模拟人类的思维过程来检测虚假新闻。与直接将大型模型用作分类器不同，ReCoLV强调大小模型之间的协作，旨在实现语义深度、计算效率和模型可解释性之间的平衡。

3.2 模型架构

ReCoLV 模型的核心思想是将 LVLM 的推理能力作为一种“高级线索”提取器，辅助轻量级模型进行更精准的判断。其整体架构如图 4 所示，主要包含三大模块：

Prompt Template $\mathcal{P}_c$ for Cue Detection

你的任务是对给定的图片和新闻文本进行多方面的评估。 这是新闻文本：“**{news_text}**”
对于图片评估：
1. 篡改检测：要检查图片是否有**篡改迹象**，需仔细观察图片中的各个元素，比如颜色、形状、物体边缘等是否有不自然或者拼接的痕迹，注意任何看起来不自然或不一致的细节。
2. 来源可信度：核查图片来源和并评估其可信度。
在||图片分析||标签内给出关于图片是否被篡改、来源和可信度的分析结果。
对于新闻文本评估：
1. 事实性错误：是否有**事实性错误**，将文本中的知识与已知的知识进行核对。
2. 语言完整性：在评估**语言表达**是否合理、流畅时，检查句子结构是否完整、通顺，用词是否准确。
3. 逻辑一致性：检查是否有**矛盾、夸张、歪曲**或**误导性**的描述。
4. 情感倾向：分析**情感倾向**和是否存在偏见，它是否试图利用读者的情感反应来传播信息。
在||文本分析||标签内给出关于新闻文本语言表达、事实性错误、矛盾等情况以及情感倾向和偏见的分析结果。

图 5　基于思维链的线索检测

Prompt Template $\mathcal{P}_s$ for Reflective Summarization

你的任务是对一段新闻的图片和文本的分析内容进行总结。请仔细阅读以下内容：“**{news_analysis}**”
对于图片分析部分，你需要总结其核心要点，尽可能包含最多信息，然后将总结内容放在||**图片总结**||标签下。
对于文本分析部分，同样要提炼关键内容，确保信息详尽，将总结放在||**文本总结**||标签下。

图 6　基于思维链的反思性总结

模块 1：基于思维链的线索检测（CoT-based Cue Detection）

为了充分激发 LVLM 的潜力，设计了一套基于思维链（Chain-of-Thought, CoT）的提示（Prompt）策略 $\mathcal{P}_c$ 如图 5 所示，引导 LVLM 对新闻的图文内容进行结构化、多维度的“类人”分析。该分析过程被分解为两个并行的轨道：

图像评估（Image Evaluation）：不仅检查图像中是否存在不自然的颜色、拼接边缘等篡改痕迹，还利用其内部知识库验证图像来源的可靠性。

文本评估（Textual Evaluation）：从语言风格、情感极性、逻辑连贯性和事实准确性等多个维度对文本内容进行剖析，识别其中可能存在的夸大、误导或情感操控等欺骗性表述。

通过这种精心设计的提示，LVLM 能够生成一份详尽的、包含深度见解的分析报告，其粒度和深度远超传统特征提取器。

模块 2：基于思维链的反思性总结（CoT-based Reflective Summarization）

LVLM 生成的原始分析报告虽然内容丰富，但也可能存在冗长、重复或焦点不突出的问题，这不利于后续与轻量级模型的集成。为了解决这个问题，设计了反思性总结模块。该模块利用一个纯语言大模型（如 Qwen-Turbo），对前一阶段生成的冗长分析进行“反思”和“提炼”，如图 6 所示。通过专门设计的总结性提示 $\mathcal{P}_s$，引导大模型将视觉和文本分析的核心要点分别归纳成精炼、准确的摘要。这个过程类似于人类在深入思考后形成最终结论，在保留关键判别信息的同时，去除冗余噪声，生成高度浓缩且对下游任务更有价值的语义表征。

	Method	Accuracy	Precision	Recall	F1-score	Fake News			Real News		
						Precision	Recall	F1-score	Precision	Recall	F1-score
Weibo	EANN	0.770	0.770	0.770	0.770	0.774	0.757	0.766	0.766	0.782	0.774
	MVAE	0.715	0.723	0.715	0.712	0.765	0.616	0.682	0.682	0.812	0.742
	SpotFake	0.767	0.772	0.767	0.766	0.734	0.834	0.780	0.810	0.701	0.752
	SpotFake+	0.731	0.732	0.731	0.731	0.716	0.760	0.737	0.748	0.702	0.725
	SAFE	0.812	0.823	0.812	0.810	0.884	0.715	0.790	0.763	0.907	0.829
	CARMN	0.825	0.826	0.825	0.825	0.837	0.816	0.826	0.814	0.835	0.824
	MFAN	0.844	0.845	0.844	0.844	0.863	0.816	0.839	0.828	0.872	0.849
	CAFE	0.849	0.848	0.849	0.849	0.898	0.810	0.852	0.798	0.891	0.842
	COOLANT	0.843	0.843	0.843	0.843	0.826	0.853	0.839	0.859	0.834	0.846
	MoPeD	0.871	0.872	0.871	0.871	0.876	**0.863**	0.870	**0.867**	0.880	0.873
	Qwen-VL	0.815	0.828	0.784	0.805	0.828	0.784	0.805	0.803	0.844	0.823
	Ours	**0.879**	**0.880**	**0.879**	**0.879**	**0.904**	0.846	**0.874**	0.857	**0.911**	**0.883**
Weibo-19	EANN	0.827	0.831	0.827	0.828	0.760	0.819	0.788	0.876	0.832	0.854
	MVAE	0.722	0.718	0.722	0.719	0.663	0.595	0.627	0.754	0.804	0.778
	SpotFake	0.712	0.711	0.712	0.696	0.707	0.459	0.555	0.714	0.877	0.787
	SpotFake+	0.722	0.718	0.722	0.718	0.667	0.586	0.624	0.751	0.810	0.780
	SAFE	0.854	0.855	0.854	0.855	0.807	0.828	0.817	0.886	0.872	0.879
	CARMN	0.864	0.864	0.864	0.864	0.845	0.817	0.831	0.877	0.897	0.887
	MFAN	0.878	0.880	0.878	0.878	0.828	0.871	0.849	0.913	0.883	0.898
	CAFE	0.898	0.899	0.898	0.898	0.864	**0.879**	0.872	0.921	0.911	0.916
	COOLANT	0.881	0.881	0.881	0.881	0.853	0.846	0.850	0.899	0.904	0.902
	MoPeD	0.881	0.881	0.881	0.881	0.865	0.828	0.846	0.891	0.916	0.903
	Qwen-VL	0.816	0.723	0.847	0.780	0.723	0.847	0.780	0.892	0.797	0.842
	Ours	**0.932**	**0.933**	**0.932**	**0.932**	**0.953**	0.871	**0.910**	**0.921**	**0.972**	**0.946**
Pheme	EANN	0.782	0.774	0.782	0.777	0.649	0.558	0.600	0.826	0.875	0.850
	MVAE	0.792	0.784	0.792	0.784	0.685	0.540	0.604	0.824	0.897	0.859
	SpotFake	0.769	0.758	0.769	0.760	0.636	0.496	0.557	0.808	0.882	0.844
	SpotFake+	0.823	0.822	0.823	0.822	0.706	0.681	0.694	0.870	0.882	0.876
	SAFE	0.813	0.807	0.813	0.803	0.753	0.540	0.629	0.829	0.926	0.875
	CARMN	0.805	0.831	0.805	0.814	0.549	0.721	0.623	0.912	0.829	0.869
	MFAN	0.875	0.882	0.875	0.877	0.756	0.850	0.800	0.934	0.886	0.909
	CAFE	0.891	0.891	0.892	0.891	0.826	0.796	0.810	0.917	**0.930**	0.923
	COOLANT	0.894	0.895	0.894	0.894	0.796	0.833	0.814	0.934	0.917	0.925
	MoPeD	0.875	0.887	0.875	0.878	0.740	0.885	0.806	**0.948**	0.871	0.908
	Qwen-VL	0.644	0.196	0.087	0.121	0.708	**0.860**	0.777	0.196	0.087	0.121
	Ours	**0.906**	**0.907**	**0.906**	**0.907**	**0.835**	0.850	**0.842**	0.908	**0.930**	**0.934**

表 3　Weibo、Weibo-19 和 Pheme 数据集上 ReCoLV 与最先进多模态假新闻检测方法的对比

模块 3：渐进式语言-视觉融合 (Progressive Language-Vision Fusion)

这是实现大小模型高效协同的关键。该模块负责将 LVLM 提炼出的高级语义线索（来自反思性总结）与轻量级模型提取的浅层特征进行有效结合。首先采用与 MoPeD 类似的双编码器（Dual Encoding）策略，利用模态特有编码器（CNN 和 ResNet-50）和模态一致编码器（CLIP）提取浅层特征。随后，将反思性总结模块输出的摘要文本通过 CLIP 文本编码器转化为高级语义特征。

融合过程是渐进式的。它并非简单地在顶层将不同来源的特征进行拼接，而是通过一个基于 Transformer 的融合网络，让高级语义信息能够逐层渗透并指导浅层特征的表示学习。该网络结合了自注意力和跨注意力机制，确保了来自大模型的知识能够与来自小模型的感知充分互动，从而产生既有深度又有细节的最终新闻表征，极大地提升了模型的鲁棒性和检测精度。

3.3 实验结果

性能对比：ReCoLV同样在在相同的三个公共数据集Weibo、Weibo-19和Pheme上进行了评估，并与包括MoPeD在内的SOTA模型进行了比较。如表3结果显示，ReCoLV在所有数据集上都取得了目前最优的性能，尤其在Weibo-19数据集上，其准确率达到了惊人的0.932，显著超越了所有对比方法。

案例研究：为了更直观地展示模型的“类人”分析能力，进行了一个案例分析，如图7所示，在一个声称“母猪产下八个男婴”的典型中文虚假新闻案例中，ReCoLV的线索检测模块 $\mathcal{P}_c$ 首先启动分析。LVLM 成功地识别出多重不一致性：其文本分析模块指出“母猪产下人类婴儿是违背生物学常识的”，判定为严重的事实性错误；其图像分析模块也识别出图片中猪与婴儿的比例和位置极不自然（如图7中的 Clue-LVLM 输出）。随后，反思性总结模块 $\mathcal{P}_s$ 将这些发现提炼为“内容违背生物常识”和“图像合成痕迹明显”的核心摘要 。

图 7　中文虚假新闻案例研究

这些由 LVLM 提供的深度洞察（高级线索）与轻量级模型的底层特征进行渐进式融合，使模型能够轻松且自信地将该新闻判定为"虚假"。这一过程清晰地展示了 ReCoLV 框架在可解释性、准确性。

消融实验：消融实验如表4所示有力地证明了模型各组件的必要性。移除反思性总结模块、或整个线索检测与总结流程、或渐进式融合模块，均会导致模型性能的显著下降。这表明"分析-反思-总结"的类人流程以及大小模型的协同机制是ReCoLV取得成功的关键。

	Strategy	Accuracy	f1-score	
			Real News	Fake News
Weibo	ours	**0.879**	**0.883**	**0.874**
	w/o Summary	0.852	0.854	0.851
	w/o Summary&Cue	0.859	0.865	0.853
	w/o Progressive Fusion	0.867	0.876	0.857
Weibo-19	ours	**0.932**	**0.946**	**0.910**
	w/o Summary	0.868	0.887	0.841
	w/o Summary&Cue	0.861	0.877	0.840
	w/o Progressive Fusion	0.895	0.910	0.874
Pheme	ours	**0.907**	**0.934**	**0.842**
	w/o Summary	0.870	0.910	0.769
	w/o Summary&Cue	0.847	0.888	0.757
	w/o Progressive Fusion	0.881	0.914	0.803

表 4　基于三个数据集的 ReCoLV 架构设计消融实验

四、总结

本文介绍的两项工作，MoPeD 和 ReCoLV，从不同维度推动了多模态虚假新闻检测技术的发展。MoPeD 通过精巧的特征工程，设计了双编码、多层次融合和模态感知学习等模块，实现了对多模态新闻中隐藏的决定性判别因素的自适应挖掘，为数据驱动的检测方法树立了新的性能标杆。而 ReCoLV 则开创性地引入了大小模型协同的范式，通过模拟人类"分析-反思-总结"的认知过程，利用 LVLM 的知识推理能力，为模型赋予了近乎"常识"的判断力，显著提升了检测的深度和可解释性。

这两项工作的成功探索表明，未来多模态虚假新闻检测的研究将朝着两个方向深度发展：一是更精细化的多模态特征交互与对齐建模；二是更智能化、更接近人类思维的知识驱动与逻辑推理。将二者有机结合，有望开发出更强大、更可靠的下一代虚假新闻检测系统，为营造清朗的网络空间提供坚实的技术保障。

本文介绍的 MoPeD 和 ReCoLV 工作代码分别发布于：https://github.com/littlesunnywgc/MoPeD 和 https://github.com/ChloeGYH/ReCoLV。

责任编委　张 杰

参考文献

[1] B. Wang, G. Wu, X. Li, J. Gao, Y. Hu, and B. Yin, “Modality perception learning-based determinative factor discovery for multimodal fake news detection,” IEEE Transactions on Neural Networks and Learning Systems, vol. 36, no. 7, pp. 12 643–12 654, 2025.

[2] J. Zheng, X. Zhang, S. Guo, Q. Wang, W. Zang, and Y. Zhang, “MFAN: Multi-modal feature-enhanced attention networks for rumor detection,” in Proc. 31st Int. Joint Conf. Artif. Intell., Jul. 2022, pp. 2413–2419.

[3] T. Jiang, J. P. Li, A. U. Haq, and A. Saboor, “Fake news detection using deep recurrent neural networks,” in International Computer Conference on Wavelet Active Media Technology and Information Processing, 2021.

[4] Q. Nan, J. Cao, Y. Zhu, Y. Wang, and J. Li, “MDFEND: Multi-domain fake news detection,” in ACM International Conference on Information and Knowledge Management, 2021, pp. 3343–3347.

[5] P. Qi, J. Cao, T. Yang, J. Guo, and J. Li, “Exploiting multi-domain visual information for fake news detection,” in IEEE International Conference on Data Mining, 2019, pp. 518–527.

[6] S. Singhal, R. R. Shah, T. Chakraborty, P. Kumaraguru, and S. Satoh, “Spotfake: A multi-modal framework for fake news detection,” in IEEE International Conference on Multimedia Big Data, 2019, pp. 39–47.

[7] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, L. Su, and J. Gao, “EANN: Event adversarial neural networks for multi-modal fake news detection,” in ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2018, pp. 849–857.

[8] Y. Wu, P. Zhan, Y. Zhang, L. Wang, and Z. Xu, “Multimodal fusion with co-attention networks for fake news detection,” in Annual Meeting of the Association for Computational Linguistics, 2021, pp. 2560–2569.

[9] Y. Chen, D. Li, P. Zhang, J. Sui, Q. Lv, L. Tun, and L. Shang, “Cross-modal ambiguity learning for multimodal fake news detection,” in ACM Web Conference 2022, 2022, pp. 2897–2905.

[10] L. Wang, C. Zhang, H. Xu, S. Zhang, X. Xu, and S. Wang, “Cross-modal contrastive learning for multimodal fake news detection,” in ACM International Conference on Multimedia, 2023, pp. 5696—5704.

[11] Z. Jin, J. Cao, H. Guo, Y. Zhang, and J. Luo, “Multimodal fusion with recurrent neural networks for rumor detection on microblogs,” in Proc.

[12] C. Song, C. Yang, H. Chen, C. Tu, Z. Liu, and M. Sun, “CED: Credible early detection of social media rumors,” IEEE Trans. Knowl. Data Eng., vol. 33, no. 8, pp. 3035–3047, Aug. 2021.

[13] A. Zubiaga, M. Liakata, and R. Procter, “Exploiting context for rumor detection in social media,” in Proc. Int. Conf. Social Informat., 2017, pp. 109–123.

[14] D. Khattar, J. S. Goud, M. Gupta, and V. Varma, “MVAE: Multimodal variational autoencoder for fake news detection,” in Proc. World Wide Web Conf., May 2019, pp. 2915–2921.

[15] C. Song, N. Ning, Y. Zhang, and B. Wu, “A multimodal fake news detection model based on crossmodal attention residual and multichannel convolutional neural networks,” Inf. Process. Manage., vol. 58, no. 1, Jan. 2021, Art. no. 102437.

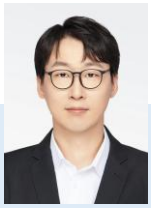
[16] X. Zhou, J. Wu, and R. Zafarani, “Similarity-aware multi-modal fake news detection,” in Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining, 2020, pp. 354–367.

[17] S. Singhal, A. Kabra, M. Sharma, R. R. Shah, T. Chakraborty, and P. Kumaraguru, “SpotFake+: A multimodal framework for fake news detection via transfer learning,” in Proc. AAAI Conf. Artif. Intell., 2020, pp. 13915–13916.

高祎菡

2024 年获得河北工业大学工学学士学位。目前正在北京工业大学攻读硕士学位，隶属于北京市多媒体与智能软件技术重点实验室。她的研究方向包括多模态虚假新闻检测和多语言大型语言模型（MLLMs）。

Email: aheadgyh@bjut.edu.cn



王博岳

2012 年毕业于河北工业大学计算机科学专业，获理学学士学位；2018 年毕业于中国北京工业大学，获博士学位。他现任北京工业大学北京市多媒体与智能软件技术重点实验室教授，研究方向包括多模态分析、知识图谱、流形学习及聚类分析。

Email: wby@bjut.edu.cn

吴广超

2025 年毕业于北京工业大学北京市多媒体与智能软件技术重点实验室，获得工学硕士学位。她的研究兴趣包括多模态假新闻检测、计算机视觉和模式识别。

Email: guangchaowu@emails.bjut.edu.cn

顶会观察

ICCV 2025

复旦大学 陈静静

国际计算机视觉大会（IEEE International Conference on Computer Vision，ICCV）是计算机视觉领域的顶级会议之一，与 CVPR 和 ECCV 并称为计算机视觉领域三大顶会。ICCV 为两年一届（奇数年举办），其学术影响力持续攀升，是中国计算机学会（CCF）推荐的人工智能领域 A 类国际学术会议，也是 Google Scholar 期刊与会议影响力榜单中的常客。本届 ICCV 大会于 2025 年 10 月 19 日至 23 日在美国夏威夷州檀香山的夏威夷会议中心举办。相比往届，ICCV 2025 的组委会阵容依旧强大且具有广泛的国际代表性，其中不乏华人学者的身影：上海科技大学的虞晶怡教授担任大会主席，同其他几位主席包括 Hilde Kuehne（图宾根大学）、Gerard Medioni（亚马逊）、Dimitris Samaras（石溪大学）和 Ramin Zabih（康奈尔大学）共同组织了本次盛会。6 名程序委员会主席中有 2 名华人学者：Google DeepMind 资深主任研究科学家孙德庆和百度视觉技术首席架构师王井东。

一、 会议亮点

回归热带与线下交流： 本届 ICCV 在风景秀丽的夏威夷檀香山举办，这也是继 CVPR 2017 之后，顶级视觉会议再次回归夏威夷。会议采用线下为主、线上结合的混合模式，为全球学者提供了绝佳的交流环境。据主办方统计，数千名研究人员亲临现场，享受了高密度的学术讨论与热带风情相结合的独特体验。

限投令与全员评审机制： 面对近年来 AI 会议投稿量爆炸式增长的挑战，ICCV 2025 首次引入了“每位作者最多投稿 25 篇”的硬性上限（Paper Cap），旨在抑制为了“刷量”而产生的低质投稿，鼓励研究者专注于更有深度的工作。同时，大会实施了更严格的义务评审制度，要求所有符合资格的作者（除非担任 AC 等职务）必须参与审稿，且对于未能按时提交审稿意见的作者实施了严厉的“Desk Reject”连坐惩罚，这一举措显著提升了审稿效率。

前沿技术爆发： 随着生成式人工智能（GenAI）的持续演进，本届会议上关于基础模型、3D 生成、视频生成以及具身智能的论文数量占据了主导地位。特别是结合大语言模型与视觉感知的多模态研究，成为了会场讨论最热烈的焦点。

二、 录用情况

投稿与录用： 大会共收到了 **11,239 篇**有效投稿，相较于 ICCV 2023 的 8,060 篇增长了约 **39.4%**。经过严格的评审，最终接收了 **2,701 篇**论文，录用率约为 **24.0%**，比上届的 26.8%有所下降。在录用的论文中，仅有 **64 篇**被选为 Oral Presentations，**Oral 率仅为 0.57%**，不仅远低于往年，也显著低于 CVPR（通常约 3%-4%）。这使得本届 ICCV 的 Oral 资格成为极具含金量的荣誉。此外，另有 **263 篇**论文被选为 Highlights，约占投稿总数的 2.3%。

评审团队： 为了处理海量投稿，会议组织了庞大的评审团，包括 6 位程序主席、**510 位**领域主席以及 **11,859 位**审稿人。每篇论文至少经过了 3 位独立审稿人的评估及一轮 Rebuttal。

中国力量： 来自中国（含港澳台）的投稿量继续领跑全球，占比约为 43%。在录用论文中，中国高校及企业表现亮眼。据不完全统计，清华大学以超过 90 篇的录用数量位居高校前列，上海科技大学在虞晶怡主席的带领下也有超 30 篇入选。企业方面，百度、华为（诺

亚方舟）、腾讯和阿里巴巴均有数十篇论文入选，显示了中国工业界在基础视觉研究上的深厚积累。

三、 热门研究方向

根据投稿和录用分布，ICCV 2025 呈现出以下技术趋势：

图像和视频合成和生成：这一方向的研究重心已不再满足于文本生成视频，而是集中在无需重训练的推理时引导与复杂的轨迹编辑上。例如，TITAN-Guide 和 TrajectoryCrafter 等工作实现了对摄像机运镜及物体运动轨迹的细粒度控制，解决了传统生成模型难以驾驭物理规律和长时连贯性的痛点。同时，动态帧率采样等高效生成技术被提出，旨在降低算力成本的同时保持视频的流畅度，标志着视频生成技术正逐步从“艺术创作”迈向可用于模拟真实物理世界的“世界模拟器”。

基于多视角与传感器的 3D 重建：3DGS 凭借其高效性彻底取代了 NeRF 成为主流，FreeSplatter 和 StreamGS 等前沿工作突破了对密集视角和预先相机标定的依赖，实现了从未标定稀疏视角到流式输入的极速重建。研究前沿正迅速向 4D 动态场景和语义特征融合拓展，这不仅让重建过程更加实时，还赋予了 3D 场景语义理解能力，实现了从“静态几何复制”到机器人可用的“动态实时感知”的跨越。

多模态大模型：社区开始系统性地验证原生多模态模型在参数扩展下的表现，证明了相比于简单的胶水式拼接，原生混合训练在算力充足时具有更优的扩展潜力。此外，视觉推理正从直觉式的模式识别向类似人类 System 2 的“慢思考”演进，通过引入长思维链来处理科学图表分析等复杂任务，推动模型从单纯的“看图说话”工具进化为具备深度逻辑能力的视觉智能体。

四、 热点论文

2025 年度最佳论文奖评审委员会由 6 名国际权威学者组成，包括 1 名华人学者：中国科学院计算技术研究所陈熙霖研究员。本年度大会的竞争异常激烈，最终评选出了 1 篇最佳论文，1 篇最佳学生论文，以及 2 篇最佳论文提名。

最佳论文 (Marr Prize)： Generating Physically Stable and Buildable Brick Structures from Text[1]，来自卡内基梅隆大学。在生成式 AI 从 2D 迈向 3D 物理世界的浪潮中，现有的文本生成 3D 模型往往忽略了物理约束，导致生成的物体无法在现实中制造或组装。本文提出了 BrickGPT，这是一个端到端的生成框架，能够根据文本提示生成既符合视觉语义又具备物理稳定性的乐高积木结构。作者构建了一个包含大规模物理稳定积木设计的数据集，并训练了一个自回归大语言模型来进行积木的序列预测。该研究创新性地引入了物理感知的回滚机制，在生成过程中实时修剪不稳定的结构，确保了生成的模型不仅美观，而且能够被机械臂或人类在现实中搭建出来。这一工作被认为是连接“生成式 AI”与“物理制造”的重要桥梁。

最佳学生论文： FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained Flow Models[2]，来自以色列理工学院。在基于文本的图像编辑领域，如何在修改图像语义的同时完美保留原图的结构细节（如物体轮廓、背景布局）一直是个棘手难题。现有的基于扩散模型的方法通常依赖于复杂 Inversion 过程来回溯噪声，这不仅计算昂贵，往往还伴随着重建伪影和由于数值误差导致的结构变形。本文基于流匹配 Flow Matching 理论，提出了一种创新的“免反演”编辑框架。该方法利用了预训练流模型天然的数学可逆性，通过求解 ODE 直接在源图像与目标文本条件之间建立精确的轨迹映射。实验表明，FlowEdit 彻底摒弃了传统方法中繁琐的优化与微调步骤，能够以极快的速度实现结构高度一致的图像编辑，为下一代交互式内容创作工具提供了基于流模型的高效新范式。

五、 大会获奖和竞赛奖

Helmholtz Prize： Helmholtz Prize 由 IEEE 模式分析与机器智能 (PAMI) 技术委员会设立，每两年在 ICCV 大会上颁发一次。该奖项以 19 世纪德国物理学家赫尔曼·冯·亥姆霍兹 (Hermann von Helmholtz) 命名，旨在表彰十年前发表于 ICCV、并对计算机视觉领域产生基础性且深远影响的研究成果。本年度共有两篇论文获奖：

Delving Deep into Rectifiers: Surpassing

Human-Level Performance on ImageNet Classification[3]，作者：何恺明、张祥雨、任少卿、孙剑。该研究基于 PReLU 网络，在 ImageNet 2012 分类数据集上取得了 4.94% 的 Top-5 测试错误率。相比 ILSVRC 2014 冠军模型 GoogLeNet 6.66% 的数值，其性能提升幅度达 26%。这也标志着在视觉识别挑战中，算法首次超越了 5.1% 的人类水平错误率。

Fast R-CNN[4] ，作者：Ross Girshick。该论文提出了多项革新，利用深度卷积网络对候选区域进行高效分类，在显著提升训练与推理速度的同时，也优化了检测精度。数据显示，Fast R-CNN 在训练 VGG16 网络时，速度是 R-CNN 的 9 倍，测试速度更是快了 213 倍；与 SPPnet 相比，其训练速度快 3 倍，测试速度快 10 倍。此外，该模型在 PASCAL VOC 2012 数据集上也取得了更高的平均精度均值。

Everingham Prize：Everingham Prize 旨在表彰为计算机视觉社区做出重大且持续性贡献的个人或团队。该奖项为纪念 Mark Everingham 而设立，意在激励后继者推动社区的整体发展。该奖项每年颁发一次，奇数年于 ICCV、偶数年于 ECCV 颁发。今年的获奖者分别是 SMPL Body Model 团队以及 VQA 团队。

Azriel Rosenfeld 终身成就奖：Azriel Rosenfeld 终身成就奖表彰的是在整个职业生涯中为计算机视觉领域做出重大贡献，并对该领域的发展产生非凡影响的研究学者。本年度获奖者 Rama Chellappa 是约翰霍普金斯大学电气与计算机工程及生物医学工程专业的彭博杰出教授，并兼任数据科学与人工智能专项计划的临时主任。他在计算机视觉、模式识别及机器学习领域的卓越建树，已对生物识别、智能汽车、法医学以及面部、物体和地形的二/三维建模等方向产生了深远影响。

六、 总结展望

ICCV 2025 不仅是一次学术成果的集中展示，更是对未来计算机视觉发展方向的风向标。从本届会议可以看出，计算机视觉正加速从“看懂世界”向“生成世界”和“操作世界”演进。

展望未来，随着计算能力的提升和算法的迭代，视觉模型将更加通用化和物理化。研究者们正致力于解决大模型在真实物理环境中的落地问题，以及如何让视觉系统具备类似人类的常识推理能力。对于中国学者而言，在保持论文数量优势的同时，在底层基础理论和原创性架构上的突破将是下一阶段的重要目标。

责任编委 **魏秀参**

参考文献

[1] Ava Pun, Kangle Deng, et al. Generating Physically Stable and Buildable Brick Structures from Text.. ICCV 2025 (Marr Prize).
[2] Vladimir Kulikov, et al. FlowEdit: Inversion-Free Text-Based Editing Using Pre-Trained Flow Models. ICCV 2025.
[3] Kaiming He, et al. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. ICCV 2015.
[4] Ross Girshick, et al. Fast R-CNN. ICCV 2015.

陈静静

复旦大学智能机器人与先进制造创新学院副教授，国家高层次青年人才计划入选者。主要研究方向为多媒体内容分析与理解、生成式人工智能安全、计算机视觉等。
Email:chenjingjing@fudan.edu.cn

天津大学王旗龙教授访谈

2025 年 8 月 25 日，《CCF-CV 专委简报》在线采访了天津大学博士生导师王旗龙教授。下面是采访实录。

王老师，您好！首先，请您分享一下您的个人学习和研究经历。

大家好！我 2007 年考入黑龙江大学计算机科学与技术专业，大三的时候由于一次竞赛的契机，非常有幸的遇到了我硕士和博士导师李培华教授。当时的比赛是利用 CUDA 在 GPU 上实现基于 EMD 的目标跟踪算法，由于没有任何可参考的工作，整个过程既辛苦又十分有成就感，现在回忆起来也是往事历历在目，这次经历让我对计算机视觉产生了浓厚的兴趣，后面也有幸保研到李老师的课题组，一直围绕图像视频识别任务开展研究。

一个非常重要的节点是硕士阶段后期，李老师为了开拓我的学术视野，安排我去香港理工大学张磊教授课题组进行交流学习，在那里我有幸结识很多学术生涯中非常重要的前辈，包括哈尔滨工业大学的左旺孟教授，天津大学胡清华教授等，在我科研研究的道路上提供了非常珍贵的支持和帮助，也坚定了我要继续走学术道路的决心。

2014 年我跟随李老师来到大连理工大学攻读博士学位。2018 年博士毕业后，在前辈们的鼓励和引荐下，我来到天津大学从事博士后研究工作，合作导师为胡清华教授。在此之后，一直在天津大学从事教学和科研工作。后面分别于 2021 年和 2024 年晋升为副教授和教授。

您的核心研究方向是“鲁棒神经网络架构”与“开放环境视觉感知”。能否谈谈您选择这一领域的初衷？您认为它在当前人工智能发展中的关键挑战是什么？

视觉表示学习是计算机视觉领域的核心问题之一，我研究生阶段在李老师的指导下，一直围绕图像视频的识别和理解开展研究，深刻体会到了鲁棒的视觉表示在视觉感知任务中的重要性。基于这样的一个动机，我们的研究目标一直聚焦在如何学习有效的视觉表示。早期的研究工作，我主要探索在经典的人工设计特征上构建基于概率分布的高效视觉表示。随着研究进展的不断变化，视觉表示学习的主流范式从传统的人工设计结合统计学习逐渐转向了深层神经网络，我的研究方向也主要聚焦在如何有效结合深层神经网络和概率分布建模，构建鲁棒的神经网络架构，从而提升开放环境视觉感知任务的表现。

目前大模型成为人工智能主导技术，但是发展适用于连续、密集数据和非结构化数据的鲁棒神经网络架构仍是一个重要挑战，特别是如何有效建模空间、结构、语义等多维度的上下文信息，提升大模型在复杂开放环境中时空感知和生成能力也是比较重要的研究课题。

您提出的一系列“深层概率分布表征网络”与传统神经网络相比有何根本区别？它们在提升模型鲁棒性和泛化能力方面有哪些优势？

深层概率分布表征网络本质上也是一种神经网络模型，需要建立在目前主流的神经网络模型基础上。最核心的区别是深层概率分布表征网络在传统神经网络的末端，为每个输入的数据样本构建深度特征的概率分布表征，从而代替原来的高维向量表征。为了实现这一目标，也面临一些新的挑战，包括高维深度特征如何鲁棒概率分布建模、深度学习框架复杂分布表征如何学习、概率分布表征中的密集计算如何优化。

传统神经网络将输入数据抽象成高维向量，难以建模数据的随机性。开放环境数据普遍存在随机性（自然噪声，跨域，外分布等），显著影响神经网络数据表示学习的性能。概率建模具备表示和修复不确定性的能力，深层概率分布表征的提出主要为了实现神经网络由深层向量表征学习转向深层概率分布表征学习，建模数据不确定性，从而提升神经网络的表征学习能力。

您的研究在自动驾驶环境感知和遥感影像解译中得到了实际应用，并获天津市科技进步奖。能否分享一下您的科研成果转化之路？以及一个技术落地过程中遇到的挑战及其解决方案？

在科研成果转化之路上我仍是一个初学者，不敢谈经验分享。这里我首先要感谢天津大学团队给搭建的平台和提供的机会。天津大学本身比较注重科研成果转化，在我刚入职天津大学的时候，学部的领导就不断带领我们到天津市相关企业进行交流讨论，并积极建立联合实验室，给年轻老师和企业牵线搭桥，达成合作的机会。

考虑到我的研究方向聚焦于开放环境感知，胡清华老师和朱鹏飞老师介绍我和中汽研和中水北方等天津市的企业合作，一同解决企业的技术需求，包括如何对自主采集的海量自动驾驶环境感知数据进行智能化管理和自动标注，为后续驾驶环境感知技术研发提供数据和平台支撑；如何利用拍摄的遥感影像解决水利水电领域关键目标和特殊事件的检测和分析。由于与我的研究方向比较切合，经过几年持续不断的合作，研发的技术在企业相关应用也取得了不错的效果，最终也十分幸运地获得了天津市科技进步奖。

一个技术落地过程中，我个人感触最多的是细节为王，需要持续不断地和企业对接和细化真实需求，甚至主动帮助企业明确具体需求，避免最终呈现结果和预期出现较大偏差的情况。此外，更多的时间需要花在技术部署阶段，包括接口对接、效率优化这些细节部分。

您下一步计划从哪些角度继续深化“鲁棒神经网络架构”和“开放环境感知”方向的研究？

我前期主要围绕概率分布表征和深层神经网络的结合开展研究。目前多模态大模型在开放环境视觉感知与推理任务上展现出强大的能力，但如何构建鲁棒的多模态表征仍是一个比较重要的问题。在未来工作中，我计划将概率分布表征研究从深层神经网络扩展到多模态大模型，研究多模态大模型通用概率化表征，解决多模态大模型中跨模态特征对齐难、鲁棒多模态表征构建难、跨任务表征协同难等问题，从而提升多模态大模型复杂场景下的视觉理解和生成能力。

您曾获中国人工智能学会优秀博士论文奖，并入选博士后创新人才支持计划。对于青年研究者如何规划早期学术生涯，您有哪些建议？

我自己也是教育和科研领域的晚辈，目前科研发展情况也和我学术生涯早期存在天差地别的变化，不敢说建议。借用前辈们建议我的话，学术生涯早期关注基础知识的储备和科研能力培养，不急于成果产出。最好瞄准一个方向持续深入研究，树立个人学术标签。这些建议对我而言非常重要，我也在持续感悟和学习。

您荣获了 2024 年天津市自然科学一等奖和吴文俊人工智能优秀青年奖。这些奖项对您和您的团队意味着什么？能否分享一下您的获奖历程及感悟？

对于我而言，我最大的感受是科研路上离不开团队平台的支持和坚持不懈的努力。我非常荣幸有机会能在目前的团队进行学习和工作。胡老师的团队围绕多模态

学习这一主题持续研究了多年，主要目标是可以从海量低质量多模态数据中获取关键信息，用于后续的推理和决策等任务。针对我的研究方向，胡老师建议我在低质量数据的表征构建方面进行研究，旨在实现对低质量数据进行鲁棒建模。因此，我的研究内容聚焦在利用深层概率分布表征对低质量数据进行建模，从而更好的处理低质量数据中噪声等数据随机性问题对数据分析带来的影响，为后续的多模态协同表征构建保驾护航。围绕多模态学习的持续研究很荣幸得到专家的认可，获得2024 年天津市自然科学一等奖。作为团队一份子，我荣幸和感恩有机会参与到这项工作中，也激励着我向坚持的目标继续努力。

您担任 CVPR 等领域主席和 IJCAI/AAAI 资深程序委员，这些服务对您自身的学术视野有何影响？您如何看待学术社区服务对学者成长的作用？

作为青年学者，我非常荣幸能参与学术社区服务。这些服务对我而言更多是学习和成长的机会。通过参与学术服务，我有幸接触到许多前沿工作与优秀学者，这让我更清晰地看到领域发展的脉络与潜在挑战，也促使自己不断反思研究的深度与价值。在我看来，学术社区服务本质上是一种双向滋养。它让我从“个人研究”转向“共同体建设”，在审稿、讨论和协作中，学习如何更全面、包容地评价工作，如何平衡创新与严谨。这些经历也让我更理解年轻学者的不易，从而更珍惜建设性反馈的意义。如果说有些许收获，是学会了在“专注自身研究”与“服务社区”之间寻找平衡，并以更谦逊的心态参与学术对话。学术生涯的真正意义，或许不仅在于发表，更在于成为知识生态中一个积极、互助的节点。

如果吐露研究工作者的心声，您最想说的是什么？

坚守对科研的热情，始终保持感恩的心，在追问未知的窄路上，做一束不灭的微光。

责任编委 余烨 赵振兵

王旗龙

王旗龙，天津大学智算学部教授，博士生导师。主要围绕鲁棒神经网络架构和开放环境视觉感知开展研究，探索了一系列深层概率分布表征网络，相关研究发表 SCI 一区/CCF-A 类论文 45 篇，其中一作和通讯论文 23 篇，谷歌学术引用 14700 余次，2020 年第一作者发表在 CVPR 的论文被引用 9600 余次；授权发明专利 5 项。先后获吴文俊人工智能优秀青年奖，中国人工智能学会优秀博士论文、CVPR 2020 最有影响力论文、图像处理旗舰会议 ICIP 2015 Top 10% Paper 等奖励。入选博士后创新人才支持计划，获国家自然科学基金面上/青年项目、科技委基础创新项目、CCF-百度松果基金、CAAI-华为 MindSpore 学术奖励基金（优秀结题项目）等支持。获 2024 年天津市科自然科学一等奖。部分技术应用于自动驾驶环境感知数据分析与遥感影像智能解译，分别获 2022 年和 2023 年天津市科学技术进步二等奖（第一完成人、第三完成人）。受邀担任 CCF-A 类会议 CVPR 2024/2025/2026 领域主席（AC）、IJCAI 2020 和 AAAI 2021 资深程序委员（Senior PC），担任多个国内会议的组织主席和出版主席等。

委员好消息

✪ 2025 年 9 月 21 日，2025 年度 CCF 科技成果奖评选结果揭晓，CCF-CV 专委会 7 位执行委员获奖。北京航空航天大学**李甲**等完成的“极端环境下近邻嵌入的微弱目标视觉计算关键技术及应用”获技术发明一等奖，中国科学院计算技术研究所**张杰**、**山世光**、**陈熙霖**等完成的“在线视觉检测技术及在高铁故障监测中的应用”获科技进步二等奖，北京师范大学**白璐**完成的“结构模式识别的理论、方法及其在金融人工智能的应用”、华南农业大学**黄栋**、中山大学**赖剑煌**等完成的“异构数据与模型协同融合表征理论与方法”获自然科学三等奖。

✪ 2025 年 9 月 24 日，CCF-CV 专委会执行委员、中山大学**林倞**获得由 CCF 与 ACM 联合评出的 2025 青年科技奖，表彰他在大规模多模态内容理解的创新算法和系统方面所做出的杰出贡献。

✪ 2025 年 10 月 18 日，2025 年度 CCF 推荐教材公布，CCF-CV 专委会执行委员、东南大学**魏秀参**等编著的《解析深度学习（第 2 版）》入选。

✪ 2025 年 11 月 7 日，2024 年度北京市科学技术奖公布，CCF-CV 专委会 5 位执行委员获奖，中科院自动化所**张兆翔**获杰出青年中关村奖，中科院自动化所**谭铁牛**、**黄岩**、**王亮**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。

✪ 2025 年 11 月 21 日，CCF-CV 专委会执行委员、西北工业大学**张艳宁**当选中国科学院院士。

✪ 2025 年 11 月 26 日，2025 年度中国图象图形学学会科学技术奖和激励计划评选结果揭晓，CCF-CV 专委会 25 位执行委员获奖。微软亚洲研究院（北京）**王井东**等完成的“分布外泛化多媒体表征学习”、浙江大学**赵洲**等完成的“跨模态内容高效生成理论与方法”、清华大学**代季峰**、中科院自动化所张**兆翔**等完成的“面向物体识别的视觉表征建模与学习方法研究”、南京信息工程大学（现南京邮电大学）**刘青山**、中国科学院自动化研究所**张一帆**等完成的“身体语言的视觉感知与理解”、大连理工大学**卢湖川**等完成的“显著目标检测深度建模与优化理论方法”获自然科学一等奖，北京理工大学**付莹**等完成的“低光通量影像复原解析关键技术及应用”获科技进步一等奖，哈尔滨工业大学**左旺孟**等完成的“动态环境低质量视觉信息增强与理解方法”、东南大学**桂杰**、中国科学院自动化研究所**李琦**等完成的“复杂场景下视觉数据的特征建模与识别”、武汉大学**罗勇**等完成的《跨域协同分析理论与方法》获自然科学二等奖，大连理工大学**杨鑫**等完成的《复杂陆战场目标威胁感知技术》获技术发明二等奖，中山大学**林倞**等完成的“立体化感知驱动的智能决策关键技术与应用”、中山大学**李冠彬**等完成的“面向智能显示设备的多模态感知与交互关键技术及应用”获科技进步二等奖。湖南大学**佃仁伟**、四川大学**胡鹏**、中国科学院计算技术研究所**张杰**获青年科学家奖。北京邮电大学**马占宇**指导的《数据受限场景下的细粒度视觉特征学习方法研究》、中科院自动化所**张兆翔**指导的《面向大范围场景的高效点云目标检测》、四川大学**彭玺**指导《深度学习驱动的图像复原：从基础模型构建到开放场景泛化》、中国科学院大学**黄庆明**指导的《面向不完备类别信息场景的分类指标优化研究》、华中科技大学**白翔**指导的《目标识别中的标注样本受限问题研究》入选博士学位论文激励计划。北京理工大学**付莹**指导的《面向复杂光照环境的图像增强研究》获博士学位论文激励计划提名。北京理

工大学**付莹**指导的《基于物理噪声建模的滤波阵列式光谱图像去噪方法研究》、大连理工大学**刘日升**指导的《场景驱动的低光照目标检测方法研究》、厦门大学**孙晓帅**指导的《面向端到端的叙事全景分割研究》、中国科学院计算技术研究所**闵巍庆**指导的《零样本食品图像检测研究》入选硕士学位论文激励计划。苏州科技大学**胡伏原**指导的《基于深度连续学习的多标签图像分类方法》获硕士学位论文激励计划提名。

✪ 2025 年 11 月 27 日，IEEE 公布了 2026 年度 IEEE Fellow 名单，CCF-CV 专委会 4 位执行委员入选，华中科技大学**白翔**因对文档图像处理和理解的贡献、江西财经大学**方玉明**因对视觉显著性检测和感知质量评估的贡献、北京大学**彭宇新**因对跨媒体分析系统和细粒度视觉识别的贡献、浙江大学**杨易**因对多媒体信号处理、感知和检索的贡献入选。

✪ 2025 年 12 月 6 日，2025 年中国通信学会科学技术奖奖励名单揭晓，CCF-CV 专委会 2 位执行委员获奖。北京邮电大学**欧中洪**等完成的“视听融合立体化故障检测智能技术及规模化应用”获科技一等奖，华南理工大学**陈俊颖**等完成的“5G 通信结合多模态环境感知的智能移动机器人自主控制技术与应用”获科技二等奖。

✪ 2025 年 12 月 10 日，2025 年度虚拟教研室试点建设典型名单公布，CCF-CV 专委会 4 位执行委员：哈尔滨工程大学**刘海波**负责的“智海 AI 课程虚拟教研室”、北京航空航天大学**王蕴红**负责的“重点领域模式识别课程群虚拟教研室”获典型虚拟教研室，北京航空航天大学**王蕴红**、**张永飞**完成的“‘思政铸魂-教学固基-实践创智’三维联动的人工智能高层次人才培养新范式”、西安电子科技大学**董伟生**等完成的“‘学-研-创-思’数智融合：‘计算智能导论’混合式教学创新实践案例”获典型教研方法。

✪ 2025 年 12 月 17 日，教育部公布第三批国家级一流本科课程认定结果，CCF-CV 专委会 6 位执行委员的课程入选。中国科学院计算技术研究所**王瑞平**、**陈熙霖**等主讲的《概率论与数理统计》、北京航空航天大学**刘祥龙**等主讲的《数据结构与程序设计（信息类）》、中南大学**赵于前**等主讲的《数字图像处理》、苏州科技大学**胡伏原**等主讲的《人工智能基础》、浙江大学**赵洲**主讲的《机器学习：模型与算法》入选。

✪ 2025 年 12 月 18 日，2025 年 CCF 博士学位论文激励计划评选结果揭晓，南开大学**程明明**《复杂场景的自适应视觉感知》获 CCF 博士学位论文激励计划提名。

✪ 2025 年 12 月 24 日，中国计算机学会公布了 2025 年度 CCF 会士名单，14 人入选。CCF-CV 专委会 2 位执行委员当选。北京大学**林宙辰**教授因在数据低秩建模、深度学习和表示学习理论方面取得了突出成果并为 CCF 计算机视觉专委的发展做出了重要贡献、北京大学**彭宇新**教授因在细粒度视觉分类、跨媒体分析研究上取得了突出成果并积极服务 CCF、为学会发展做出了重要贡献入选。

✪ 2025 年 12 月 25 日，黑龙江省教育厅公布了黑龙江省优质本科课程认定结果，CCF-CF 专委会执行委员、哈尔滨工程大学**刘海波**主讲的《走进人工智能》入选。

责任编委 刘海波

掌纹识别开源代码

兰州理工大学 孟存宁 李策

掌纹识别作为生物特征识别的重要分支，凭借其独特的纹理结构与高稳定性，在身份认证领域占据重要地位。然而，相较于传统的接触式采集，新兴的无接触式成像虽然在卫生与便捷性上具有显著优势，却面临着特征判别力瓶颈、非受控环境下的域适应鸿沟以及高维数据稀缺等严峻挑战，导致单一模态在复杂场景下的泛化能力受限。

近年来，深度学习技术的引入为突破上述瓶颈提供了新范式。通过构建多阶竞争机制的表征学习网络、设计端到端的空间自适应对齐策略以及引入可控生成模型进行数据扩增，不仅能从高维流形中重构掌纹的细微纹理与拓扑结构，还能有效解决非受控环境下的特征漂移与样本长尾分布问题。本文将重点介绍该领域极具代表性的开源成果，涵盖深度综合竞争机制（CCNet）、非受控环境下的自适应对齐（EE-PRnet）以及基于扩散模型的可控数据生成（PalmDiff），展示了从特征表征到数据合成的全链路技术突破。

1、Comprehensive Competition Mechanism in Palmprint Recognition (CCNet)

竞争编码（Competitive Code）是掌纹识别中的经典方法，但其本质上仅关注通道维度的胜者选择，严重忽略了特征的空间分布信息及高阶纹理特征的判别力。针对这一局限，CCNet(Comprehensive Competition Network)提出了一种全新的深度综合竞争机制。该网络创新性地设计了多阶纹理特征提取模块，通过并行分支同时捕捉一阶梯度与二阶纹理细节，显著丰富了特征的表达能力。

此外，CCNet 构建了“通道-空间”双重竞争机制。该机制不仅沿用了通道竞争以优选最佳纹理方向，更引入空间竞争策略，使模型能够自适应地锁定主线交汇处及纹理密集区等高判别力区域，显著增强特征表达的同时有效抑制背景噪声干扰。完整的网络架构如图 1 所示。

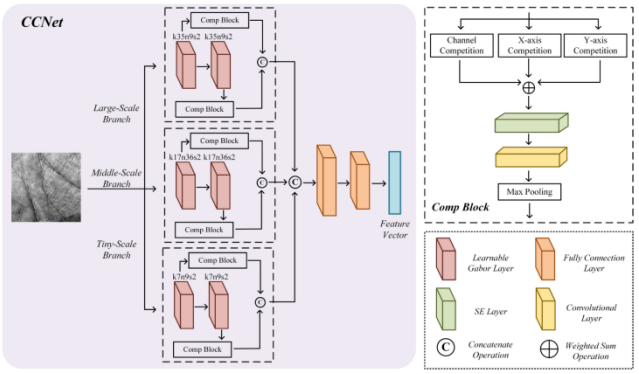


图 1 CCNet 结构图

相比于传统方法仅能激活零散的局部特征，CCM 机制显著扩展了高响应区域的覆盖范围。通过融合多维竞争信息，该机制能更有效地捕捉渐变主线与突变纹理等关键掌纹细节。(a)与(c)显示传统机制在不同阶特征上的关注区域均较局限，相比之下(b)与(d)展示了 CCM 更具判别性的热力图，可视化结果如图 2 所示。

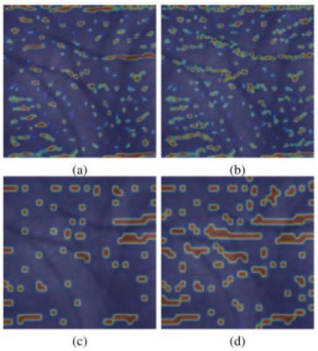


图 2 不同竞争机制的多阶纹理特征响应可视化对比

更多有关 CCNet 的详细内容可参考发布该方法的论文“Comprehensive Competition Mechanism in Palmprint Recognition”。

论文链接：
https://ieeexplore.ieee.org/document/10223233
代码链接：
https://github.com/Zi-YuanYang/CCNet

2、Palmprint Recognition in Unco-ntrolled and Uncooperative Envir-onment (EE-PRnet)

在非受控及非配合环境下，掌纹图像往往伴随复杂的背景噪声、剧烈的姿态变化及尺度差异，传统的基于人工设计特征的方法难以适用。EE-PRnet 提出了一种鲁棒的端到端掌纹识别框架，旨在解决从原始复杂图像到高精度身份认证的全链路自动化问题。

该框架包含两个级联阶段：第一阶段为 ROI 定位与自适应对齐网络（ROI-LAnet），其基于空间变换网络（STN）并结合薄板样条（TPS）非刚性变换，端到端回归手掌关键点以隐式学习空间对齐关系，在无需依赖传统显式关键点检测算法的情况下实现高精度的 ROI 自动裁剪与姿态校正；第二阶段为特征提取与识别网络（FERnet），从对齐后的掌纹图像中提取深度判别特征。通过联合训练策略，EE-PRnet 实现了特征对齐与识别性能的相互促进，其总体结构如图 3 所示。

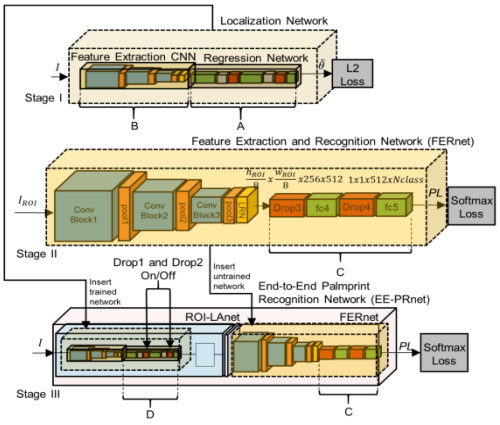


图 3 EE-PRnet 网络架构

即便面对光照昏暗、背景杂乱或手部姿态极度扭曲等极端干扰，EE-PRnet 仍能保持高度的特征语义一致性，将正确的目标身份精准锁定在检索结果的前列。在极具挑战性的 NTU-PI-v1 数据集上，该算法展现出了超越传统 SOTA 方法的显著性能优势，充分验证了其在开放场景下的实用潜力和技术领先性。真实场景下的 Top-10 检索可视化结果如图 4 所示。

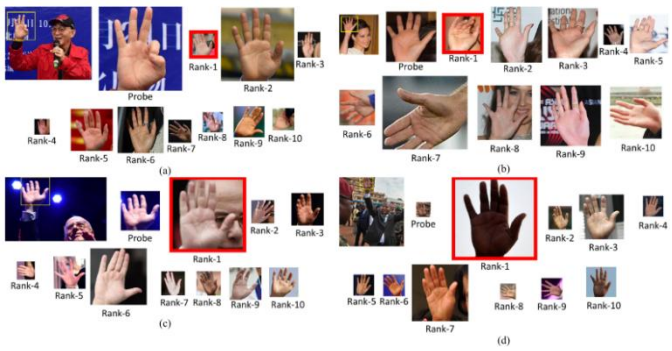


图 4 真实场景下的 Top-10 检索可视化结果

更多有关 EE-PRnet 的详细内容可参考发布该方法的论文“Palmprint Recognition in Unco-ntrolled and Uncooperative Environment”。

论文链接：
https://ieeexplore.ieee.org/document/8854829
代码链接：
https://github.com/matkowski-voy/Palmprint-Recognition-in-the-Wild

3、PalmDiff:When Palmprint Gene-ration Meets Controllable Diffusi-on Model

大规模公开数据集的匮乏已成为制约掌纹识别技术发展的关键瓶颈。传统的生成对抗网络（GANs）在生成样本时往往面临模式崩溃和泛化能力不足的问题，难以保留掌纹独特的主线结构与细微纹理。受扩散模型在生成领域优异表现的启发，本文提出了 PalmDiff，一种基于可控扩散模型的高质量掌纹生成方法。

PalmDiff 设计了专用的扩散过程，有效平衡了生成过程中的噪声抑制与纹理细节保留。同时，网络引入了多尺度聚焦线性注意力机制（MFLA），增强了骨干网络对细微纹理的捕捉能力。针对生成样本中可能出现的身份一致性退化问题，PalmDiff 引入 ID 一致性损失函数，并通过自适应层归一化模块将身份特征显式注入生

成过程，确保生成的掌纹在保持边缘结构可控的同时，维持严格的身份一致性。完整的网络架构如图 5 所示。

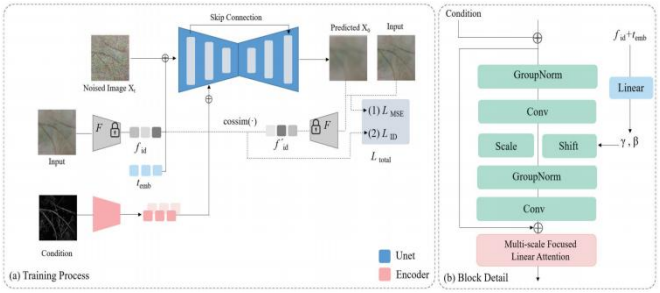


图 5 PalmDiff 网络架构

较于 CycleGAN 等传统对抗生成方法，PalmDiff 在主线结构连贯性与纹理细节保留方面表现更为稳定，并有效减少生成伪影。在定量评估中，其在 FID 等指标上取得较优结果，表明生成分布与真实数据具有更高一致性。将其生成数据用于识别模型预训练，在跨数据集测试中进一步降低了识别错误率，其 ROI 图像生成结果对比图如图 6 所示。

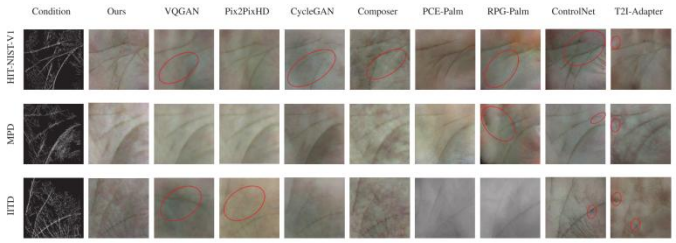


图 6 ROI 图像生成结果对比图

传统扩散过程在高频纹理建模上易产生不稳定性，导致生成掌纹出现色彩偏移与纹理平滑现象，其中 VP 与 VE 过程（图 7(b)、(c)）生成的样本主线模糊且高频细节缺失。相比之下，PalmDiff 采用的 EDM 确定性采样策略（图 7(a)）通过优化噪声方差调度，有效降低累积误差，在保持色彩一致性的同时更好地重建了掌纹结构与高频纹理，其无条件生成结果如图 7 所示。

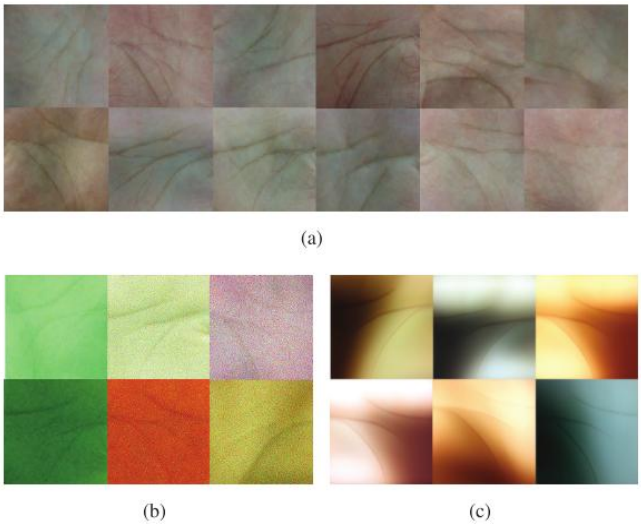


图 7 不同扩散采样过程下的无条件掌纹生成效果对比

更多有关 PalmDiff 的详细内容可参考发布该方法的论文“PalmDiff: When Palmprint Generation Meets Controllable Diffusion Model”。

论 文 链 接：
https://ieeexplore.ieee.org/document/11114788
代 码 链 接：
https://github.com/Ukuer/Diff-Palm

责任编委 贾同 王田

孟存宁

兰州理工大学自动化与电气工程学院，硕士研究生，研究方向为计算机视觉，掌纹识别等。

李 策

教授，博士生导师，兰州理工大学微电子现代产业学院院长，长期从事教育与科研工作。研究方向为计算机视觉、医学影像分析，智能机器人等。

X线违禁品检测数据集

东北大学 贾同 林舒扬

1、SIXray数据集概述

SIXray（Security Inspection X-ray）数据集是当前安检违禁品识别领域中应用最广泛的公共数据资源之一，可同时支持违禁品的分类与定位等多种视觉任务。该数据集由中国科学院大学模式识别与智能系统开发实验室于2019年发布，旨在推动面向真实安检环境的智能检测技术研究。

SIXray数据集的构建初衷在于模拟真实安检场景中显著的类别极度不均衡特性。该数据集采集自地铁安检系统的实际X线包裹图像，共包含1059231张样本，其中仅有8929张含有至少一种违禁品，正样本占比约为1%。这种极端不均衡分布真实反映了现实安检场景中“安全物品远多于违禁品”的固有特征。图1给出了部分包含违禁品的图像示例。所有图像均以JPG格式存储；对于正样本，数据集同时提供XML注释文件，用于记录违禁品的类别信息及其在图像中的位置。整个数据集覆盖六类常见违禁品，包括手枪、刀具、扳手、钳子、剪刀和锤子。其中，锤子类仅包含60个样本，由于数量过少，多数研究中不将其纳入实验分析。

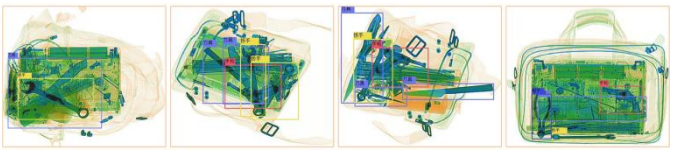


图1 SIXray数据集图像示例

SIXray数据集呈现出多方面的复杂性，在视觉识别任务中具有较高挑战性。首先，数据中的X线图像主要来自对手提包、背包和行李箱等实际物品的安检扫描。如图2所示，每张X线图像可视为由多幅透明子图相互

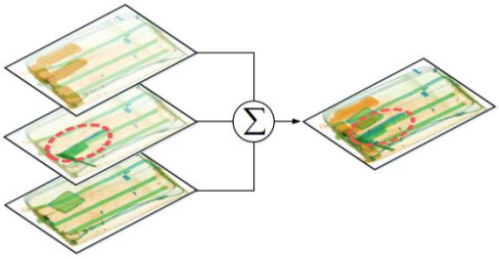


图2 X线图像可视为多幅透明子图像的叠加

叠加而成，X线的穿透成像属性导致被遮挡物体仍然可见，形成显著的物体重叠现象。其次，违禁品在X线图像中的尺度、形变、视角、材质和细粒度结构差异大，类内差异性高，显著增加检测难度。此外，由于大多数安全物品无法明确标注具体类别，使模型难以准确区分背景区域与潜在的目标区域。

如前所述，该数据集中正样本比例极低。在未进行特殊处理的训练易受负样本占比过高的影响而出现类别偏置，模型仅通过持续预测负类即可获得较高准确率，对训练的稳定性与学习有效性构成挑战。为研究数据不平衡对算法性能的影响，SIXray数据集构建了三个具有不同正负样本比例的子集，分别命名为SIXray10、SIXray100和SIXray1000，数字表示负样本与正样本数量比。SIXray10与SIXray100均保留全部正样本，分别匹配10倍、100倍数量的负样本。且SIXray100的数据分布最接近真实安检场景。为评估算法处理极端不平衡情况的能力，SIXray1000子集随机选取1000张正样本，并与全部1050302张负样本组合而成。各子集均以4:1的比例划分训练集与测试集。

数据集下载地址：
https://github.com/MeioJane/SIXray

2、OPIXray数据集概述

SIXray 数据集作为当前违禁品识别领域的重要公共资源，研究者能够深入探索违禁品分类与检测任务在高度复杂背景、严重遮挡和极端类别不平衡条件下的关键挑战，从而推动更高精度、更强鲁棒性的安检智能检测技术发展。

OPIXray（Occluded Prohibited Items X-ray）数据集是安检违禁品检测领域的重要资源，由北京航空航天大学复杂关键软件环境全国重点实验室于 2020 年发布，旨在应对安检场景 X 线图像存在的物体重叠问题。

OPIXray 数据集采集自北京首都国际机场，背景图像均来自真实安检设备扫描结果，而违禁品则通过专业软件合成到这些背景中。所有违禁品均由机场专业安检人员进行人工标注，采用边界框形式提供精确的位置信息。尽管图像采用合成方式生成，但仍保留了 X 线成像的关键特性，严重遮挡情况下仍能呈现物体的外形结构，以及依据材料属性赋予不同的伪彩色。与 SIXray 数据集类似，OPIXray 为每张图像提供独立的标注文件（.txt 格式），用于记录目标类别及其空间位置信息。

OPIXray 数据集共包含 8885 张 X 线包裹图像，覆盖五类刀具类违禁品：折叠刀，直刀，剪刀，美工刀和多功能刀。图像分辨率统一为 1225×954，均以 JPG 格式存储。数据集按约 4：1 的比例分为训练集（7109 张）图像，测试集（1776 张）图像。此外，数据集中有 35 张图像包含多个违禁品，其中训练集中 30 张，测试集中 5 张。图 3 所示为 OPIXray 数据集的部分样例图像：

图 3 OPIXray 数据集图像示例

为了研究不同遮挡程度对检测性能的影响，OPIXray 数据集将测试集进一步划分为三个子集，分别命名为遮挡等级 1（Occlusion Level 1, OL1）、遮挡等级 2（Occlusion Level 2, OL2）和遮挡等级 3（Occlusion Level 3, OL3），其中数字表示图像中违禁品的遮挡程度。如图 4 所示，OL1 中的违禁品几乎不存在遮挡或仅具有轻微遮挡，OL2 表现为部分遮挡，而 CLCXray（Cutters and Liquid Containers X-ray）数据集 OL3 子集则包含遮挡程度较高或完全被遮挡的样本，用于最大限度评估模型在高遮挡场景下的检测能力。

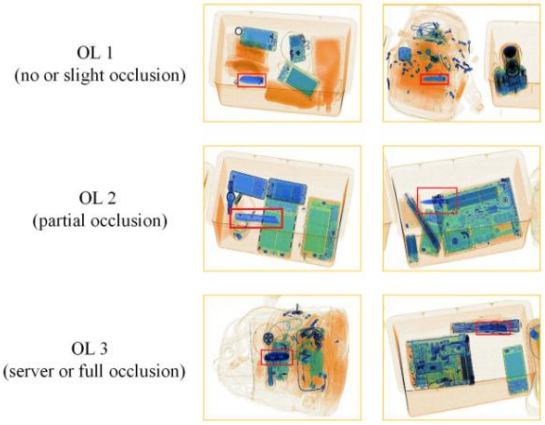


图 4 不同遮挡程度的图像示例

数据集下载地址：
https://github.com/OPIXray-author/OPIXray

OPIXray 数据集设计高度贴合真实安检场景的成像环境。其遮挡特性来源于现实中的行李安检情境，导致物体之间不可避免地发生重叠。此外，该数据集中各类别的样本数量明显不均衡，例如折叠刀和多功能刀类别的样本数量多于直刀类别，这主要是由于前两类刀具在实际旅客携带物品中出现频率更高。在不同遮挡等级中，OL3（高遮挡等级）样本数量远少于 OL1（低遮挡等级），原因在于刀具类违禁品体积较小、在行李中较易移动，因此在实际场景中完全遮挡的情况较为罕见。

OPIXray 数据集主要具备两个核心应用场景。首先，可用于评估模型在 X 线图像中检测违禁品，一个性能优越的模型应能在不同遮挡等级下均保持良好的检测效果。其次，可用于评估模型解决物体遮挡问题的能力，通过比较模型在不同遮挡等级下相较于其他方法的性能提升幅度，衡量其对遮挡问题的处理效果。当模型在高遮挡等级下的性能提升显著高于低遮挡等级时，表明其在应对物体遮挡问题上具有较强的有效性与鲁棒性。

OPIXray 数据集为研究物体重叠与遮挡场景下的检测算法提供了重要支撑。研究者可借助其系统评估模型在不同遮挡等级下的识别能力，分析遮挡对检测性能的影响，并探索提高模型在高遮挡、部分遮挡及复杂叠放条件下违禁品识别鲁棒性的有效方法。从而提升安检系统在复杂场景下的检测精度与安全保障。

3、CLCXray 数据集概述

据集由同济大学视觉与智能学习实验室于2022年发布，旨在应对X线包裹图像中物体极端重叠的检测挑战。

CLCXray数据集包含9565张X线图像，其中4543张为真实数据（来源于地铁实际安检场景），5022张为模拟数据（由人工设计的行李进行扫描生成）。图5展示了CLCXray数据集的部分图像示例：

图5 CLCXray数据集图像示例

CLCXray数据集包含12个类别，其中包括5种刀具和7种液体容器。刀具类别包括刃刀，匕首，菜刀，剪刀和瑞士军刀；液体容器包括易拉罐，纸箱饮料，玻璃瓶，塑料瓶，真空杯，喷雾罐以及罐头。该数据集共包含超过20000件潜在违禁品，每张X线图像平均包含超过两个潜在违禁品。图像分辨率在373×200至732×1280之间，所有标注已转换为COCO格式。CLCXray数据集被划分为训练集、验证集和测试集，比例为8:1:1。在构建测试集时，通过随机抽样将模拟数据与真实数据按照1:9的比例组合，其余样本则按照8:1的比例划分为训练集和验证集。值得注意的是，测试集中真实样本比例达到90%，显著高于训练集与验证集中的43%，从而更贴近真实安检场景的检测需求。

与SIXray和OPIXray相比，CLCXray数据集具有以下特性：首先，图像中物体重叠现象更加突出，主是每张图像中标注的目标数量更多，近60%的图像至少包含两个或以上前景目标。图6展示了CLCXray数据集中不同程度的物体重叠情况。其次，该数据集引入以往研究中尚未涉及液体容器类别。液体容器可能含有有毒、腐蚀性、易燃或易爆液体，这类违禁物品危险性高，但在实际安检中容易被忽视。第三，CLCXray提供了精确的边界框标注，使得模型更准确地学习目标定位。

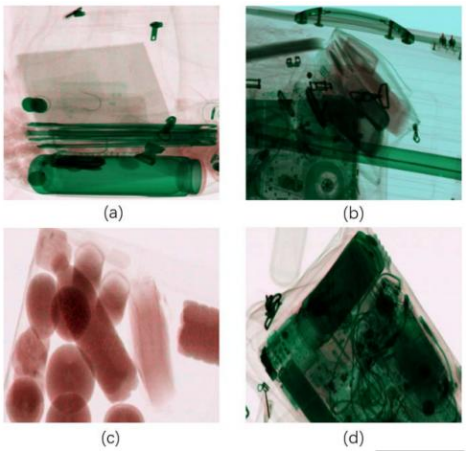


图6 CLCXray数据集中的重叠现象

数据集下载地址：
https://github.com/GreysonPhoenix/CLCXray

作为兼顾多物体重叠与多类别复杂性的典型公共数据资源，CLCXray数据集为研究X线图像中违禁品检测算法提供了重要的数据支撑。借助该数据集，研究者能够在更加复杂的场景下系统评估模型的检测性能，包括多目标重叠、液体容器识别及高密度前景情况下的表现，同时，CLCXray的精细标注可帮助研究者探索提升模型在多目标、高重叠及复杂物体组合条件下识别鲁棒性的有效方法，提高安检系统的安全保障。

贾 同

责任编委 李策 樊鑫

东北大学信息科学与工程学院教授、博士生导师，长江学者，未来技术学院、机器人学院党委书记（双肩挑）。研究方向：计算机视觉、模式识别与机器学习等。电子邮箱：jiatong@ise.neu.edu.cn

林舒扬

博士研究生，东北大学信息科学与工程学院，研究方向：计算机视觉、模式识别与机器学习等。电子邮箱：2210329@stu.neu.edu.cn

好 文 推 荐

东京大学、筑波大学以及北京理工大学合作的最新成果“EventHDR: From Event to High-Speed HDR Videos and Beyond”发表在 IEEE TPAMI 2025。

论文：Yunhao Zou; Ying Fu; Tsuyoshi Takatani; Yinqiang Zheng. EventHDR: From Event to High-Speed HDR Videos and Beyond, IEEE TPAMI, 47 (1): 32-50 (2025)

事件相机（Event Cameras）是一种新型的神经形态传感器，可异步捕捉场景动态。基于事件触发机制，此类相机记录的事件流相比传统相机具有更短的响应延迟和更高的光强灵敏度。基于这些特性，先前的研究尝试从事件数据中重建出高动态范围视频，但往往存在非真实感伪影或无法提供高帧率等问题。

在本文中，研究人员同步利用了事件相机的双重优势，即卓越的帧率与鲁棒的动态范围耐受性，旨在从事件流中重建出高速的高动态范围视频。如图 1 所示，针对重建此类视频存在的时间稀疏性问题，该文研究人员提出一种由关键帧引导的递归神经网络，该网络专为高速视频重建任务而设计。其所提出的深度学习模型能够沿高速帧序列提取时序关联信息。此外，为缓解稀疏事件流中的信息损失，本文引入了关键帧引导机制，向网络输入关键数据信息。他们还采用金字塔式可变形网络对连续事件帧间的特征进行对齐，施加时序一致性约束以增强序列整体的连续性。

在重建网络设计之外，本文提出首个真实配对的“事件-高速”的高动态范围数据集，称为 EventHDR。他们采用双高速相机配合高动态范围融合技术生成高质量真值数据，完整保留了动态场景中高速与高动态范围的核心特性。相较于先前重建方法，使用真实配对的训练数据可显著提升模型的重建质量。简单而言，先前方法由于训练数据真实性不足，导致重建结果存在明显伪影且与真实光强图像差异显著。而本文的模型能从事件流中生成视觉效果卓越的高速、高动态范围视频。大量实验表明在训练阶段引入真实配对数据能有效提升模型处理真实世界高动态范围场景的能力，为各类计算机视觉任务提供有力支撑。

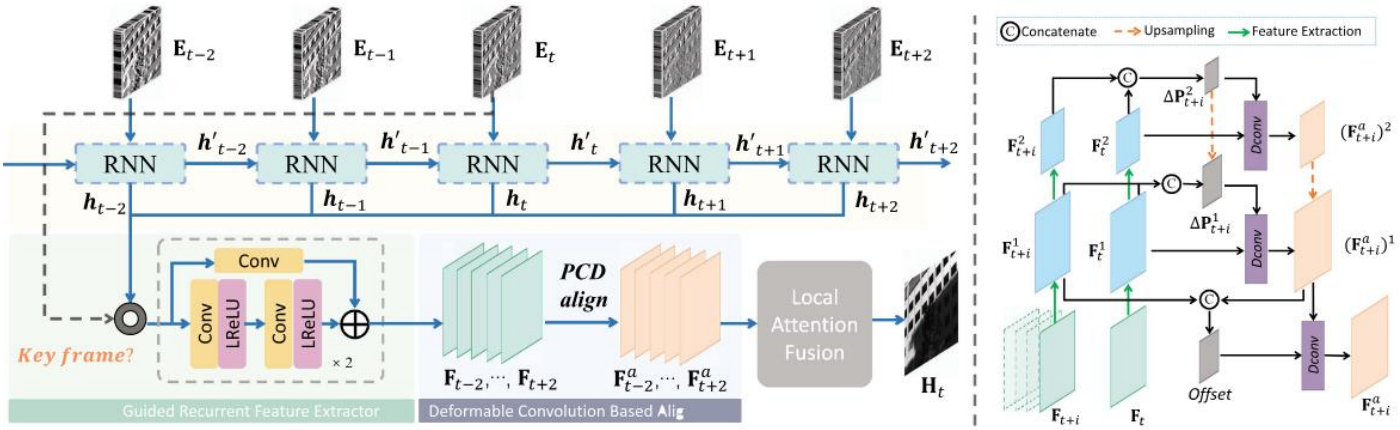


图 1 递归神经网络结构流程图

责任编委 贾同 樊鑫

好 文 推 荐

南京理工大学的”HORP: Human-Object Relation Priors Guided HOI Detection “

论文：Pei Geng, Jian Yang, Shanshan Zhang*. HORP: Human-Object Relation Priors Guided HOI Detection, CVPR 2025: 25325-25335.

人-物交互(HOI)检测旨在预测由“人-动作-物体”组成的三元组，其核心挑战在于准确识别每一对人-物交互关系。尽管近年来随着模型结构的不断改进，HOI检测性能有所提升，但整体仍不令人满意。本文首先进行了一些失败案例分析，发现“无交互”类别的准确率极低，严重制约了整体性能的提升。进一步分析误差类型后可以发现，无交互与有交互之间的混淆，本质上可通过引入人–物关系先验来缓解。为此，我们提出了两种人-物关系先验：（1）三维位置先验，表示人和物体之间的空间距离，用于区分无交互与直接交互。（2）注视区域先验，表示人是否能看到物体，用于区分无交互与间接交互。两类先验通过文本形式表达，再与原始视觉特征进行融合，为交互识别提供丰富的多模态线索。

具体来说，如图 1 所示，本文提出了一种基于人-物关系先验（HORP）的 HOI 检测方法，旨在提升模型对不同交互类型的区分能力。首先，我们设计了三维位置先验和注视区域先验，由于仅依赖图像平面的二维位置信息往往具有误导性（例如，看似靠近的物体可能处于完全不同的深度），我们将描述空间关系的方式从 2D 扩展至 3D，以更准确地反映人–物位置关系。三维位置先验由两部分组成：图像平面中的相对二维位置，以及相对深度信息。考虑到使用外部深度模型开销过大，我们通过“人物像素数量+语义类别”与“物体像素数量+类别”粗略的估计相对深度。对于注视区域先验，我们引入 gaze 检测器来预测人的注视区域，以估计目标物体是否处于其注视范围内。之后，我们将上述两类先验分别转化为文本描述，并输入冻结的文本编码器以获得文本特征。将原始视觉特征与两类先验的文本特征进行融合，构建 HOI 解码阶段的查询机制，并最终预测交互类别。实验结果表明，我们的方法显著提高了无交互类别的识别精度，并提升了整体性能。

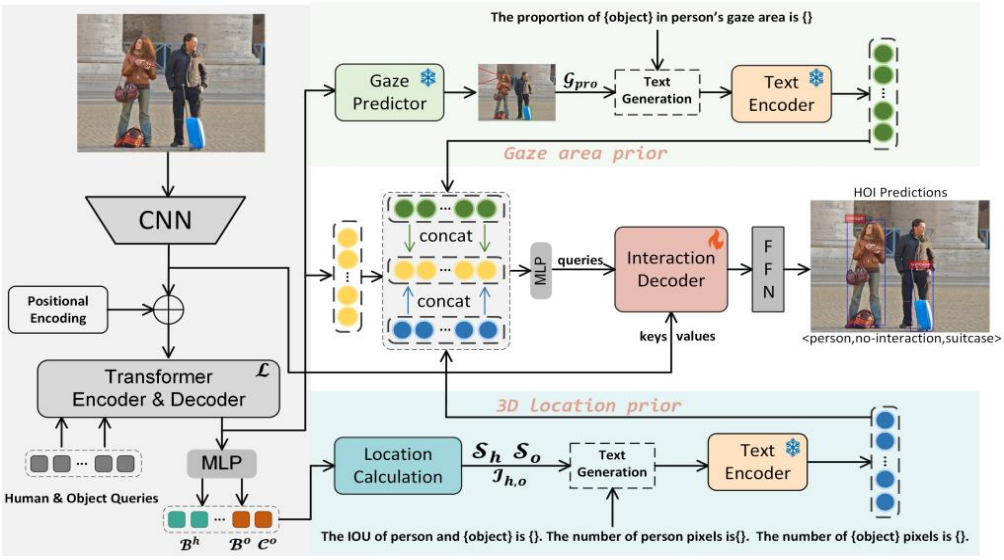


图 1 HORP 的整体架构

责任编委 王田 李策

好 文 推 荐

北京航空航天大学的" Understanding the Dimensional Need of Non-Contrastive Learning "

论文：Zhexiao Cao, Lei Huang, Tian Wang, Member, Yinquan Wang, Jingang Shi, Aichun Zhu, Tianyun Shi, Hichem Snoussi, , vol. 55, no. 9, pp. 4089-4102, Sept. 2025, doi: 10.1109/TCYB.2025.3577745

无负样本对比自监督学习解决了假负样本等对比学习的潜在危害，在视觉、语音、图结构和文本等多种模态上展现出强大的性能，但在实践中却普遍依赖高维表示空间，表现出明显的"维度低效"现象，其理论根源尚不清晰。本文从理论角度系统研究无负样本对比学习在表示维度上的需求，聚焦于"上游表示学习—下游分类性能"之间的传递机制。我们以两种典型的无负样本对比学习方法 Barlow Twins 和 SimSiam 为代表，分析其如何通过约束输出表示的相关结构来隐式放大类间距离，并在此基础上建立了一个将下游分类误差率与表示维度 d、潜在类别数 K 以及输出相关矩阵 Λ 紧密联系起来的误差上界理论，形式化地说明当 d 显著小于 K 时，下游任务性能会不可避免地显著劣化，从而为无负样本对比方法"需要高维表示"给出了严谨解释。

具体而言，如图 1 所示，本文提出以输出表示的相关矩阵 $\Lambda = E[f(x)f(x)^{\mathrm{T}}]$ 作为统一的结构指标，证明 Barlow Twins 的冗余约束损失以及 SimSiam 的学习动力学最终都在显式或隐式地最小化 $\|\Lambda - I_d\|^2$，并由此推导出一个可在预训练阶段计算的无标签度量 $L_{\text{bound}} = 2/d^2 \cdot \|\Lambda - I_d\|^2 + (K-d)/d^2$，用于上界下游最近邻分类误差率。我们进一步指出，无负样本对比学习优化的是"相关矩阵均匀性"而非高斯势能意义上的球面均匀性，使得未归一化表示倾向于沿一组近似正交基聚类，并由维度 d 与类别数 K 共同决定可分性。大量实验证明：在图像分类、目标检测与分割、图表示、音频表示及句子表示等多种模态和任务上，L_{bound} 与实际下游精度高度相关，当 d 大致不低于 $2K$ 时模型仍保持良好泛化，而过度压缩维度会使误差上界失去意义，从而为无负样本对比学习中"如何合理选取表示维度"提供了可解释的理论依据与实践指导。

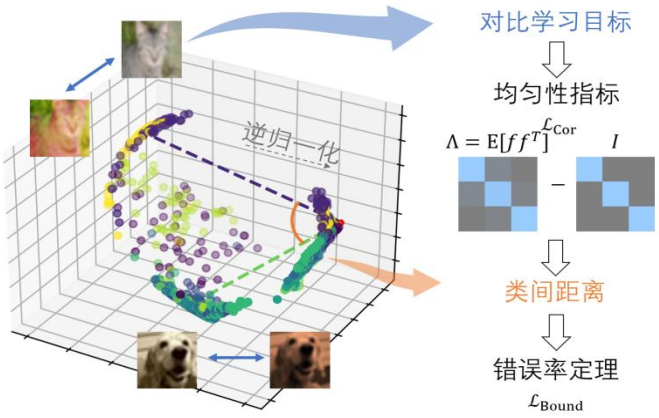


图 1 误差上界理论框架

责任编委 李策 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 Computers & Graphics， Pattern Recognition Letters, ISPRS J PHOTOGRAMM 和 Pattern Recognition。

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)，由中国图象图形学学会（CSIG）、中国人工智能学会（CAAI）、中国计算机学会（CCF）和中国自动化学会（CAA）联合主办。旨在汇聚全球学者展示前沿研究成果，推动技术创新与应用发展，已进入 CCF 分区（CCF-C）。会议通常包括主旨报告、专题论坛及产业交流环节，覆盖学术研究、智能驾驶、智能制造等热点方向。

第九届 PRCV 将于 2026 年由哈尔滨工程大学承办。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
IJCAI 2026	2026.08.15-21	Bremen, Germany	2026.01.19	https://2026.ijcai.org/
ICML 2026	2026.07.06-12	Seoul, South Korea	2026.01.28	https://icml.cc/Conferences/2026
ECCV 2026	2026.09.08-13	Malmö, Sweden	2026.03.06	https://eccv.ecva.net/

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
CG	Computer Graphics and Visual Computing Conference 2025 Special Issue	https://www.sciencedirect.com/journal/computers-and-graphics	2026.02.16
PRL	Special Section on the 5th International Conference on Intelligent Systems and Pattern Recognition	https://www.sciencedirect.com/special-issue/327209/special-section-on-the-5th-international-conference-on-intelligent-systems-and-pattern-recognition	2026.02.15
ISPRS P&RS	Advanced Uncrewed Vehicles solutions: at the crossroad of Remote Sensing, Computer Vision and Robotics	https://www.sciencedirect.com/special-issue/324110/advanced-uncrewed-vehicles-solutions-at-the-crossroad-of-remote-sensing-computer-vision-and-robotics	2026.03.31
PR	Foundation Models and Prompting for Visual Tasks in Harsh Conditions	https://www.sciencedirect.com/special-issue/322677/foundation-models-and-prompting-for-visual-tasks-in-harsh-conditions	2026.03.01

COMPUTER VISION NEWSLETTER

04 2025

总第 46 期

计算机视觉专委会简报

CCF 计算机视觉专委会