

陕西省自然科学奖公示信息

(2026年度)

一、项目基本情况

项目名称	多模态复杂行为意图预测理论及方法
主要完成人	魏平；杨进；李环；李嘉鹏
主要完成单位	西安交通大学

二、提名意见（适用于部门、机构提名）

提 名 者	陕西省知识产权局	提名等级	☐ 一等奖 ☒ 二等奖及以上
提名意见： “多模态复杂行为意图预测理论及方法”项目针对人工智能领域中复杂行为意图感知与预测的科学难题，构建了一套完整的原创性理论体系，实现了行为意图“可表征、可计算、可评价”的突破，具有重大的科学意义与应用价值。 该项目的核心贡献主要体现在三个维度的系统性创新：在表征层面，首创“人-意-物”时空交互模型，突破了意识意图在视觉数据中“隐式存在、难以捕捉”的瓶颈，实现了动态场景中跨时空、多粒度交互语义的统一建模；在计算层面，提出随机深度结构推理方法，创新性地解决了并发行为识别、注意力-意图联合推理等关键问题；在评价层面，建立了以“心智理论”为框架的意图预测评价体系，包含 IntentQA、TIA、COA 等数据集，填补了领域内“评价无基准”的空白。 项目成果的学术影响力与国际认可度显著。研究成果发表于 TPAMI、CVPR、ICCV 等顶刊顶会，被图灵奖得主、美国工程院院士等国际权威学者广泛引用并高度评价。在应用转化方面，技术成功部署于西安交通大学无人驾驶车辆等实际系统，充分验证了理论的工程落地价值。该项目的研究为智能系统理解真实场景的复杂时空结构提供了基础性理论支撑，其原创性、系统性与影响力均处于国际领先水平。郑重推荐其申报陕西省自然科学奖一等奖或二等奖。 说明：省科学技术奖一、二等奖项目，实行按等级标准提名、独立评审表决的机制。提名单者应严格依据省科学技术奖的标准条件，说明提名项目的贡献程度及等级建议。“仅提名一等奖”评审落选项目不再降格参评二等奖。提名项目正式提交后，提名等级建议本年度不得变更。			

二、提名意见（适用于专家提名）

[illegible]

三、项目简介

本项目聚焦多模态复杂行为意图预测理论及方法，面向自主无人系统、人机协作和公共安全等领域的行为理解与交互决策需求，研究如何综合利用 RGB 图像或视频、深度信息和自然语言等多模态数据，从人物行为、交互对象及任务情境中推断潜在行为意图。针对“异构信息关联难表征、动态行为意图难建模、行为意图评价无标准”的核心挑战，从理论表征、计算推理和评价体系三个层面开展系统研究，构建贯通多模态结构化表征、动态联合推理与多层次评价的理论方法体系，为多模态信息驱动的复杂行为意图的深层理解与可靠预测提供支撑。项目总体情况如图 1 所示，本项目成果三大科学发现点如下：

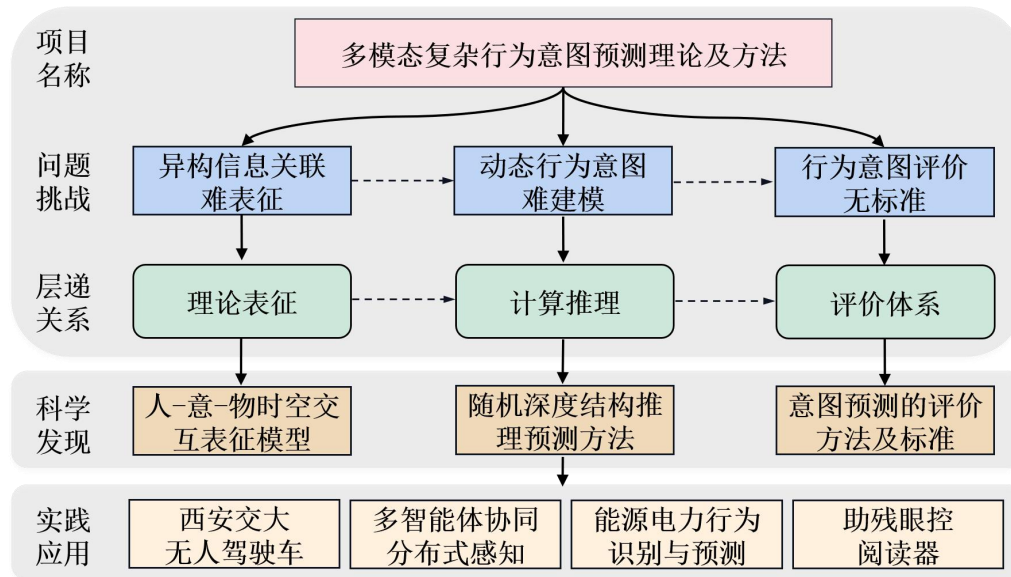


图 1.项目总体情况

1. 科学发现点一：提出面向动态场景的人-意-物三元时空组合隐结构表征理论，建立多模态观测与潜在行为意图之间的结构化联系，创造性地将不可见的意图转化为可见特征的时空交互结构关系，解决了感知数据的意图隐藏与行为并发难题，实现了从单一行为识别向多主体、多属性、多层级复杂时空结构建模的突破。

针对多模态观测中意图特征隐含、并发行为及深层语义关系难以刻画的问题，提出“人—意—物”三元时空组合隐结构表征理论，建立可观测行为线索与潜在意图之间的结构化联系。通过组合结构表征与复合时序逻辑刻画并发动作关系，突破单一动作标签的表征局限；以人体姿态、注意力、交互对象及其关系承载潜在意图，构建意图—行为—环境从视觉感知到认知推理的跨层关联；通过视频查询语言、对比学习与常识推理融合不同层次的上下文信息，实现视觉、语言与常识协同驱动的视频意图推理。相关成果为复杂行为意图的结构化表征与可解释理解奠定了理论基础。

2. 科学发现点二：提出面向复杂行为意图预测的层次化联合推理与跨任务协同方法，建立了四维时空交互解析、注意-意图推断以及与任务相关时序证据定位的方法体系，解决了复杂任务视频中的事件、行为、意图联合推理计算难题，实现了对复

杂动态场景视频的统一建模与深层语义解析。

针对复杂任务中事件、行为、注意与意图相互耦合以及交互关系动态演化的难题，提出层次化联合推理与跨任务协同方法。构建四维人-物交互随机分层图，采用动态规划束搜索实现事件分割、行为识别与物体定位的联合计算，形成行为与物体相互解释的推理机制；建立注意与意图相互约束的时序联合推断模型，利用人体姿态、对象关系和任务情境生成候选解释，并通过束搜索进行联合求解；提出任务解耦与跨任务反馈的多模态时序解析方法，通过提取任务特定与共享表征、交换任务间的互补信息以及设计任务相关联合损失，实现多任务协同优化，为复杂行为意图分析提供任务相关的时序证据与计算支撑。

3. 科学发现点三：构建覆盖并发行为、注意意图、四维交互与视频意图问答的多层次评价基准，形成从行为结构解析到潜在意图理解的分层验证路径，为多模态复杂行为意图预测提供数据与评价基础。

构建了以类“心智理论”为框架的复杂时空结构理解与预测标准化评价体系，解决了该领域长期缺乏统一测评标准的问题。设计了五类关键测试机制，覆盖认知问答、注意建模、并发行为与反事实推理核心维度。代表性成果包括：以动机-行为-结果链为核心的 IntentQA 数据集、聚焦注意-意图-任务三元关系的 TIA 基准、三维并发行为数据集 COA，以及 4D 人-物交互评估模块和反事实推理机制，全面提升了模型的可解释性与评估深度，为该领域建立系统化测评标准提供了基础。

本项目成果的核心理论与关键技术形成 5 篇代表性论文，其中 1 篇论文发表在中科院一区、人工智能领域国际顶级期刊 IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 上；4 篇论文分别发表在人工智能领域国际顶级、CCF-A 类会议 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) 和 IEEE/CVF International Conference on Computer Vision (ICCV) 上。研究成果被图灵奖得主、美国科学院/工程院等院士、IEEE/ACM Fellow 等顶尖学者引用并给予高度评价，在领域内形成了显著的学术影响力。图灵奖获得者、美国工程院及艺术与科学院院士、斯坦福大学 Hanrahan 教授指出本研究“通过学习人的行为来分类物体和行为”，突显其在人类行为理解中的应用价值；人工智能国际著名科学家、意大利巴勒莫大学 Cascia 教授引用本项目关于并发行为的原创工作，称其“处理了一个独特的、原创的问题”，并明确指出其为未来可能的研究方向之一。

相关理论成果在智能驾驶系统中成功部署，助力西安交通大学车队在智能车未来挑战赛中荣获城市道路组与高架道路组双料冠军；助力国家电网破解了新能源降碳监测与评估瓶颈，显著提升了能源运维精细度；同时，基于意图预测理论开发了助残电子书阅读器控制系统，显著提升了交互自然度；并且开发了多智能体分布式协同感知系统，大大提升了系统的任务完成效率。项目全面推动了多模态复杂行为意图预测领域的范式演进与理论构建，在国际学术界和工业界产生深远影响，确立了本团队在该领域的领先地位。

四、客观评价

该项目成果对应的相关论文已在人工智能领域国际顶级期刊 TPAM 与 CCF-A 类会议 CVPR 和 ICCV 上发表,被广泛引用与认可,产生了重要的学术影响。总体来看,相关评价不仅来自计算机视觉、多媒体理解、行为识别等学术领域的国际知名专家,也来自自动驾驶、长视频推理、三维视觉理解等应用方向的代表性研究团队。国际同行对该成果的评价也集中体现在三个方面:一是肯定其在复杂并发行为、组合语义表达和人-意-物交互建模方面提出了差异化的原创问题;二是肯定其通过潜在行为理解环境,并以分层图、四维人-物交互和序列建模实现了更全面的动态推理范式;三是肯定其构建的意图问答、因果推理和高级视频理解基准,对后续视频理解和具身智能任务评估产生了持续影响。

(一) 对发现点 1 的评价: 处理了一个独特的、原创问题

1. 图灵奖获得者、美国工程院院士、美国艺术与科学院院士、斯坦福大学 Hanrahan 教授指出该工作“通过学习人的活动来分类物体和行为”(代表性引文3);

2. IEEE & ACM & AAAI Fellow、Google DeepMind 研究科学家、CVPR Longuet-Higgins Prize 获得者、加州大学默塞德分校 Ming-Hsuan Yang 教授指出,本项目提出的 TaskWeave 方法“超越了传统的数据驱动联合建模方式,推进了视频时刻检索与高光检测的联合研究范式,并有效提升了时刻定位和显著性预测性能”(代表性引文7);

3. 人工智能国际著名科学家意大利巴勒莫大学 Cascia 教授在其行为综述经典论文中,将本项目关于并发行为的工作,作为原创代表性工作,在6处予以介绍和引用,评价其“处理了一个独特的、原创的问题”,并指出该研究为未来可能的研究方向之一(代表性引文4);

4. IEEE & IAPR Fellow、美国纽约州立大学视觉实验室主任 Lyu 教授在其工作中多处引用并指出该项目建立的并发行为数据库(COA)是“公开的唯一三维并发行为检测数据库”(代表性引文5)。

(二) 对发现点 2 的评价: 提出了一个更全面的程式、能够表征广义的人类交互

1. 美国工程院院士、ACM & IEEE Fellow、斯坦福大学教授 Guibas 在其论文中指出“提出了一个更全面的程式,从视频序列中联合分割物体和交互事件”(代表性引文1);

2. IEEE Fellow、ACM SIGGRAPH Academy 成员、加拿大西蒙菲莎大学杰出教授 Hao Zhang 指出,该成果能够“表征广义的人类交互”(代表性引文2);

3. 美国工程院院士、美国艺术与科学院院士、ACM & AAAI & IEEE Fellow、MIT CSAIL 主任 Daniela Rus 教授在其论文中将所建立的多视角3D事件数据集列入多模态人类活动数据集对比中,指出“这类数据集探索了多模态感知如何促进对人类行为的理解”(见附件);

（三）对发现点 3 的评价：体现出对高级视频能力的要求、专注于因果与时间推理

1. IEEE & ACM Fellow、哥伦比亚大学教授Shih-Fu Chang引用该工作并指出：“IntentQA 数据集展示了结构化事件图在视频推理中的优势”（代表性引文8）；

2. 美国工程院院士、美国艺术与科学院院士、ACM Fellow、Stanford HAI 创始联合主任Fei-Fei Li教授将该成果列为动作解析领域代表性工作，并指出此类研究能够“同时建模动作的多种特征”（代表性引文6）；

3. 丰田自动驾驶团队Woven by Toyota在其论文中引用该工作并指出：“结合上下文线索估算3D注视方向已被探索用于基于RGB(D)视频的优化方法中”，充分肯定了该方法在3D注视估计中的先导性与实践价值（见附件）；

4. IEEE Fellow、新加坡国立大学教授Wynne Hsu在其论文中引用该工作并指出：“IntentQA等基准体现出对高级视频处理能力的要求，如解释力、因果性、预测性与想象力”（见附件）；

5. Meta AI首席研究员Lorenzo在其论文中指出：“IntentQA被用于多任务微调训练”，为模型在长视频复杂推理任务中的能力提升提供了基础支持（见附件）。

五、代表性论文专著目录

(不超过 8 条, 其中代表性论文不超过 5 篇, 代表性专著不超过 3 部)

序号	论文专著名称	刊名	作者	年卷页码 (xx 年 xx 卷 xx 页)	发表时间	通讯作者	第一作者	国内作者	他引总次数	检索数据库	知识产权是否归国内所有
1	Modeling 4d human-object interactions for joint event segmentation, recognition, and object localization.	IEEE Transactions on Pattern Analysis and Machine Intelligence	Ping Wei, Yibiao Zhao, Nanning Zheng, Song-Chun Zhu	2017 年 39 卷 1165-1179 页	2017-06	-	魏平	魏平, 郑南宁	30	SCI, EI	是
2	Concurrent Action Detection with Structural Prediction.	Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).	Ping Wei, Nanning Zheng, Yibiao Zhao, Song-Chun Zhu.	2013 年 3136-3143 页	2013-12-01	-	魏平	魏平, 郑南宁	37	EI	是

3	Where and why are they looking? jointly inferring human attention and intentions in complex tasks.	Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).	Ping Wei, Yang Liu, Tianmin Shu, Nanning Zheng, Song-Chun	2018年 6801-6809 页	2018-06-18	-	魏平	魏平，郑南宁	44	EI	是
4	Task-driven exploration: Decoupling and inter-task feedback for joint moment retrieval and highlight detection.	Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition(CVPR).	Jin Yang, Ping Wei, Huan Li, Ziyang Ren.	2024年 18308-18318 8 页	2024-06	魏平	杨进	杨进，魏平，李环，任子洋	17	EI	是
5	Intentqa: Context-aware video intent reasoning.	Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV).	Jiapeng Li, Ping Wei, Wenjuan Han, Lifeng Fan.	2023年 11929-11940 0 页	2023-10-01	-	李嘉鹏	李嘉鹏，魏平，韩文娟，范丽	30	EI	是
合 计											
补充说明（视情填写）:											

六、主要完成人情况表

姓 名	魏平	排 名	1
行政职务	人工智能学院副院长		
技术职称	教授		
工作单位	西安交通大学		
完成单位	西安交通大学		
对本项目主要学术贡献：			
对科学发现点 1、 2、 3 均有贡献，主持和设计整个项目，提出了复杂行为意图表征、计算模型和评价标准，是代表性论文 1、 2、 3 的第一作者，代表性论文 4 的通信作者，代表性论文 5 的合作作者，也是发明专利（一种基于人眼注视点检测的电子书阅读器控制方法及系统）的第一完成人。			

姓 名	杨进	排 名	2
行政职务	无		
技术职称	助理教授		
工作单位	西安交通大学		
完成单位	西安交通大学		
对本项目主要学术贡献：			
对科学发现点 2 有贡献，提出了任务驱动的解耦表征、任务间互反馈与多任务协同优化方法，解决了多模态多任务表征难题和时序建模问题，是代表性论文 4 的第一作者，所提方法被 IEEE & ACM & AAAI Fellow 、Google DeepMind 研究科学家 Ming-Hsuan Yang 教授评价为“超越了传统的数据驱动联合建模方式，有效提升了时刻定位和显著性预测性能”。			

姓 名	李环	排 名	3
行政职务	无		
技术职称	助理教授		
工作单位	西安交通大学		
完成单位	西安交通大学		
对本项目主要学术贡献：			
对科学发现点 2 有贡献，提出了多任务联合优化方法，解决了多任务联合优化过程中导致的梯度冲突这一关键问题，是代表性论文 4 的合作作者。			

姓 名	李嘉鹏	排 名	4
行政职务	无		
技术职称	无		
工作单位	西安交通大学		
完成单位	西安交通大学		
对本项目主要学术贡献：			
对科学发现点 3 有贡献，提出了上下文感知的视频意图问答任务与推理模型，建立了意图预测的评估基准，是代表性论文 5 的第一作者。			

七、主要完成单位情况表

单位名称	西安交通大学
对本项目主要学术贡献： 西安交通大学作为成果的唯一完成单位，为本项目的研发与转化提供了全方位支撑。该成果依托人工智能与机器人研究所/人工智能学院的优势资源，学院拥有人机混合增强智能全国重点实验室以及国家人工智能产教融合创新平台两个国家级平台，长期专注于以计算机视觉与模式识别为基础的智能信息处理研究，在多模态学习领域积累了丰富的理论研究和实践经验，形成了成熟的研究体系。实验室配备了充足的高性能计算服务器和先进的处理设备，为项目提供强大的技术支持。依托国家级科研平台、先进的硬件设施以及高水平的人才队伍，为本成果提供全方位的平台支撑与专家建议，确保研究目标的实现与技术成果的落地。在项目实施期间，学校通过高性能计算平台与智能车实训场地，保障了多模态意图预测模型的大规模训练与实测，驱动了本成果在表征、计算、评价三个维度的系统性创新，实现了行为意图“可表征、可建模、可计算”的实质性跨越；更主导了成果的产学研转化，支持核心算法部署于本校无人驾驶车队及智能电网等重大工程。	

完成人合作关系说明

本项目第一完成人西安交通大学魏平教授，牵头负责本项目的研究工作，与其余三位完成人均存在合作关系。第一完成人在西安交通大学开展工作，主持国家自然科学基金等多项科研项目，指导研究生多名，与其他完成人合作完成多篇论文（合作证明材料见下表）。

本项目第二完成人西安交通大学杨进博士为第一完成人魏平教授的博士毕业生，二人合作开展多模态表征与视频时序理解的相关研究，合作发表学术论文 1 篇（合作证明材料见下表）。

本项目第三完成人西安交通大学李环博士为第一完成人魏平教授的博士毕业生，二人合作开展多模态学习的相关研究，合作发表学术论文 1 篇（合作证明材料见下表）。

本项目第四完成人西安交通大学李嘉鹏博士为第一完成人魏平教授的博士生，二人共同针对意图评价基准取得了一系列成果，合作发表学术论文 1 篇（合作证明材料见下表）。

完成人合作关系情况汇总表

序号	合作方式	合作者/项目排名	合作时间	合作成果	证明材料
1	论文合著	杨进, 2	2022年9月-至今	Jin Yang, Ping Wei, Huan Li, Ziyang Ren. Task-driven exploration: Decoupling and inter-task feedback for joint moment retrieval and highlight detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.	代表性论文
2	论文合著	李环, 3	2020年9月-至今	Jin Yang, Ping Wei, Huan Li, Ziyang Ren. Task-driven exploration: Decoupling and inter-task feedback for joint moment retrieval and highlight detection. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2024.	代表性论文
3	论文合著	李嘉鹏, 4	2021年9月-至今	Jiapeng Li, Ping Wei, Wenjuan Han, Lifeng Fan. IntentQA: Context-aware Video Intent Reasoning. International Conference on Computer Vision, 2023.	代表性论文
4	共同知识产权	李嘉鹏, 4	2019年8月-至今	魏平, 李嘉鹏, 唐浩雯, 王帅杰. 一种基于人眼注视点检测的电子书阅读器控制方法及系统 (授权号: CN 118826014 B)	发明专利