

丁子涵

邮箱: dingzihan737@gmail.com 电话: +86 18600022722



教育经历

2022/09 至今	北京航空航天大学, 人工智能研究院	博士
• 排名: 8/44		
2021/09 – 2022/06	北京航空航天大学, 计算机学院	硕士
• 导师: 刘偲		
2017/09 – 2021/06	北京航空航天大学, 高等理工学院	本科
• 排名: 4/213		

荣誉奖励

1. 腾讯犀牛鸟精英人才计划	2025/06
2. 国家奖学金	2023/09
3. 北航博士新生奖学金 (4/44)	2022/09
4. 北航十佳研究生提名奖 (全校博士研究生中 22 人)	2023/06
5. 北航优秀学术创新成果 1 项 (进入十佳优秀学术创新成果答辩)	2023/08
6. 校级三好学生	2023/09
7. 北航优秀研究生	2024/04
8. 北航优秀毕业生、优秀毕设	2021/06

研究经历

- 单流策略优化 (共同第一负责人)** **2025/08-2025/09**
 - 回归经典的 RL 范式并解决了其核心挑战。SPO 的创新点在于: 1) 用一个持久化的 KL 自适应价值追踪器替代了临时的组内基线, 提供了更稳定、低方差的学习信号; 2) 全局归一化优势函数, 提升了样本效率; 3) 天然支持优先采样, 实现了自适应课程学习。SPO 通过更根本的原理设计, 摒弃了复杂的工程变通, 为 LLM 的 RL 训练提供了更高效、可扩展的路径。在多个高难度数学推理基准上, SPO 的性能**平均超越 GRPO 3.4 个百分点** (在 BRUMO 25 等数据集上提升高达 7.3 百分点)。
 - 相关成果: 基于该算法的论文**被 NeurIPS 2025 Workshop: MATH-AI 接收, 投稿至 ICLR 2026 分数 (10/6/4/4)**。
- 面向歧义指令的人机协作 (第一负责人)** **2024/05-2024/07**
 - 我参与提出了直觉估计与知识规划 (IEKP) 框架。该框架创新性地将 LLM 的语境理解能力与结构化知识相结合: 1) 通过关联推理 (ARGoal) 初步估算并扩展人类意图; 2) 利用有限状态机 (FDPrune) 约束 LLM 的决策路径, 保证长期任务的鲁棒性; 3) 借助知识增强搜索 (K-Search) 高效定位目标。这一工作将 LLM 的“软”推理与“硬”知识库相结合, 显著提升了 AI 在模糊指令下的任务执行能力。
 - 相关成果: 基于该算法的论文**被 PRCV 2025 接收**
- 用于高效通信协作感知的紧凑型消息单元 (第一负责人)** **2023/06-2024/11**
 - 针对多智能体协同感知现有消息单元 (点云、框、BEV Map) 固有缺陷, 提出了新的消息单元——点云簇, 具备稀疏且同时包含语义和结构信息。基于此提出了协同感知的全面系统, 解决了现有方法在信息打包、信息聚合、位姿误差等方面的局限。该算法在协同联合感知相关 benchmark 上取得了 state-of-the-art 的结果。
 - 相关成果: 基于该算法的论文**发表 ICLR 2025**
- 全景叙事基础的冻结文本到图像扩散模型的动态提示 (第三负责人)** **2023/08 – 2024/04**
 - 全景叙事基础 (PNG) 需要对图像中的事物和材料对象进行全景分割, 并与叙事标题的名词短语相匹配。以前的判别方法依赖于全景分割的预训练, 这需要花费大量精力来注释标签, 并且缺乏图像-文

本对齐的先验知识，从而限制了它们的多模态学习能力。利用大规模文本到图像扩散模型固有的细粒度图像-文本对齐功能，将其生成性预训练知识转移到判别性 PNG 任务中。该算法在相关 benchmark 上取得了 state-of-the-art 的结果。

- 相关成果：基于该算法的论文发表 ACM MM 2024
- 5. 用于 RGB-T 语义分割的区域自适应和上下文互补交叉调制（第二负责人） 2022/06 – 2023/06
 - 前景-背景区分和互补信息挖掘是可见光-热红外（RGB-T）任务的两个关键挑战。我们提出了一个具有两个协作组件的交叉调制过程。第一个区域自适应通道调制（RACM）模块在细粒度区域级别进行通道注意，其中每个通道可以对前景和背景区域进行不同的调制。第二个上下文互补空间调制（CCSM）模块在编码阶段早期挖掘并传输两种模态之间的互补信息。该算法在相关 benchmark 上取得了 state-of-the-art 的结果。
 - 相关成果：基于该算法的论文发表于 Pattern Recognition 2024
- 6. 使用耦合的像素和对象上下文丰富短语的全景指代分割算法（第二负责人） 2022/11 – 2023/01
 - 提出使用短语-像素-对象转换器解码器来丰富具有耦合像素和对象上下文的短语，以实现更全面的视觉语言交互。此外，还提出了短语对象对比损失，以从更多与短语相关的对象标记中聚合更精确的对象上下文。该算法在全景指代分割 benchmark 上取得了 state-of-the-art 的结果。
 - 相关成果：基于该算法的论文已发表于 IJCAI 2023
- 7. 基于时空联合建模的视频目标指代分割算法（第三负责人） 2021/09 – 2022/02
 - 在工作 7 的基础上，我们对所提出的跨模态适应性调制模块进行改进，使得语言特征也被适应性的更新。在解码器阶段提出了一个语言感知的语义传播模块，通过语言采样将深层的语义信息传递到浅层，增强语言兼容的前景视觉特征。该算法在多个视频目标指代分割 benchmark 上取得了 state-of-the-art 的结果。
 - 相关成果：基于该算法的论文已发表于 TPAMI 2023
- 8. 针对开放词表检测的目标蒸馏金字塔算法（第四负责人） 2022/06 – 2023/03
 - 开放词表检测算法中非适应性候选框裁剪以及单级别特征模仿过程，导致知识抽取过程中目标信息被破坏和低效的知识迁移。为了解决这些问题，我们提出了一个目标蒸馏金字塔框架，适应性的变换目标候选框，采用目标掩码注意力来获得精确和完整的目标知识。使用全局和块蒸馏进行全面的知识迁移，补偿目标蒸馏中丢失的关系信息。该算法在 MS-COCO 上取得了 state-of-the-art 的结果。
 - 相关成果：基于该算法的论文已发表于 CVPR 2023
- 9. 针对视频目标指代分割的基于语言桥接的时空交互算法（第一负责人） 2021/09 – 2021/11
 - 针对视频目标指代分割任务，在编码器阶段利用语言作为中间桥实现显式和适应性的时空交互，精准定位描述性文本所指代的视频区域内容及实体。在解码器阶段利用通道注意力机制抑制语言无关的时间和空间特征，增强时空一致的通道。该算法在多个视频目标指代分割 benchmark 上取得了 state-of-the-art 的结果，且模型计算量和参数量显著降低。
 - 相关成果：基于该算法的论文已发表于 CVPR 2022，申请专利一项（实审阶段）
- 10. 基于像素-短语匹配的全景指代分割算法（第一负责人） 2022/01 – 2022/04
 - 针对全景指代分割算法，提出了一个简单有效的单阶段像素-短语匹配网络。并提出了一个语言兼容的像素聚合模块通过视觉上下文增强短语的可分辨性。该算法在全景指代分割 benchmark 上取得了 state-of-the-art 的结果。
 - 相关成果：基于该算法的论文已发表于 ACM MM 2022
- 11. 基于时空联合建模的视频目标指代分割算法（第四负责人） 2020/09 – 2020/11
 - 提出了时空联合建模的编码器-解码器框架，其中三维时间编码器从视频帧序列中识别和查询的动作，二维编码器抓取目标帧被查询目标的静态视觉上下文。在解码器阶段通过语言指导的特征选择模块灵活地集成时间和空间特征。在编码器阶段提出了一个跨模态适应性调制模块动态地重新组合每阶段空间/时间相关的语言特征。该算法在多个视频目标指代分割 benchmark 上取得了 state-of-the-

art 的结果。

- 相关成果：基于该算法的论文**已发表于 CVPR 2021**
- 12. **基于渐进式时空交互的视频目标指代分割算法（第一负责人）** **2021/04 – 2021/05**
 - 提出一个渐进多模态交互网络，在视觉骨架网络的每一个阶段进行跨模态的融合，在语言引导下实现视觉特征的渐进式学习。此外，我们利用视频目标分割网络进行后处理来对所有帧上的目标分割结果进行优化。
 - 相关成果：Youtube-VOS 视频目标指代分割赛道第二名（CVPR 2021 Workshop 技术报告）

项目参与

1. 科技创新 2030 重大项目-车路协同
2. 美团精细化图文理解项目（**2021-2022 年度卓越创新奖**）、美团开放图像理解项目
3. 航天创新研究院基于多模态大模型的空间情报态势分析项目
4. 航天三院多目标跟踪项目

实习经历

1. **美团视觉理解组算法实习生** **2021/09 – 2023/09**
 - 围绕精细化图文理解进行算法研究，包括视频指代分割、图像全景指代分割任务，相关成果发表 3 篇 CCF A 类论文，技术应用于美团优选商品首图推荐场景，在推荐精排层面优化 51.3% 的图片质量，提升了 0.43% 的线上千人曝光销售额，线上支付订单数相对提升 0.3%，并申请专利（进入实审）一项。
2. **腾讯游戏 AI 研发中心应用实习生** **2023/10 - 2024/03**
 - 研究利用世界模型提升强化学习及序列决策算法的通用性泛化性，在《元梦之星》游戏上进行测试验证。搭建基于 pytorch 的业务模型训练基座代码。

论文发表

1. Zhongwen Xu* and **Zihan Ding***. Single-stream Policy Optimization. **Accepted** by the 5th Workshop on Mathematical Reasoning and AI (MATH-AI), 2025.
2. **Zihan Ding***, Jinyu Chen*, Si Liu, Shifeng Zhang. **Intuition Estimation and Knowledge-Based Planning for Human-AI Collaboration. Accepted** by the 8-th Chinese Conference on Pattern Recognition and Computer Vision (PRCV), 2025.
3. **Zihan Ding**, Jiahui Fu, Si Liu, Hongyu Li, Siheng Chen, Hongsheng Li, Shifeng Zhang, Xu Zhou. Point Cluster: A Compact Message Unit for Communication-Efficient Collaborative Perception. **Published** at International Conference on Learning Representations (**ICLR**), 2025.
4. Hongyu Li, Tianrui Hui, **Zihan Ding**, Jing Zhang, Bin MA, Xiaoming Wei, Jizhong Han, Si Liu. Dynamic Prompting of Frozen Text-to-Image Diffusion Models for Panoptic Narrative Grounding. **Published** at the 32th ACM International Conference on Multimedia (**ACM MM**), 2024.
5. Fengguang Peng, **Zihan Ding**, Ziming Chen, Gang Wang, Tianrui Hui, Si Liu, Hang Shi. Region-adaptive and context-complementary cross modulation for RGB-T semantic segmentation. **Published** at Pattern Recognition (**PR**), 2024.
6. Tianrui Hui, **Zihan Ding**, Junshi Huang, Xiaoming Wei, Xiaolin Wei, Jiao Dai, Jizhong Han, Si Liu. Enriching Phrases with Coupled Pixel and Object Contexts for Panoptic Narrative Grounding. **Published** at the 32nd International Joint Conference on Artificial Intelligence (**IJCAI**), 2023.
7. Luting Wang, Yi Liu, Penghui Du, **Zihan Ding**, Yue Liao, Qiaosong Qi, Biaolong Chen, Si Liu. Object-Aware Distillation Pyramid for Open-Vocabulary Object Detection. **Published** by IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2023.
8. **Zihan Ding***, Zi-han Ding*, Tianrui Hui, Junshi Huang, Xiaoming Wei, Xiaolin Wei, Si Liu. **Published** at

the 30th ACM International Conference on Multimedia (**ACM MM**), 2022.

9. Tianrui Hui, Si Liu, **Zihan Ding**, Shaofei Huang, Guanbin Li, Guanbin Li, Wenguan Wang, Luoqi Liu, Jizhong Han. Language-Aware Spatial-Temporal Collaboration for Referring Video Segmentation. **Published** at IEEE Transactions on Pattern Analysis and Machine Intelligence (**TPAMI**), 2022.
10. **Zihan Ding**, Tianrui Hui, Junshi Huang, Xiaoming Wei, Jizhong Han, Si Liu. Language-Bridged Spatial-Temporal Interaction for Referring Video Object Segmentation. **Published** at IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2022.
11. Tianrui Hui*, Shaofei Huang*, Si Liu, **Zihan Ding**, Guanbin Li, Wenguan Wang, Jizhong Han, Fei Wang. Collaborative Spatial- Temporal Modeling for Language-Queried Video Actor Segmentation. **Published** at IEEE Conference on Computer Vision and Pattern Recognition (**CVPR**), 2021
12. **Zihan Ding**, Tianrui Hui, Shaofei Huang, Si Liu, Xuan Luo, Junshi Huang, Xiaoming Wei. Progressive Multimodal Interaction Network for Referring Video Object Segmentation. Published at the 3rd Large-scale Video Object Segmentation Challenge (CVPR 2021 Workshop Technical Report), 2021.