

China3DV 2025

面向人-机-环境协同共生的模型、系统及应用

汇报人：马月昕

上海科技大学

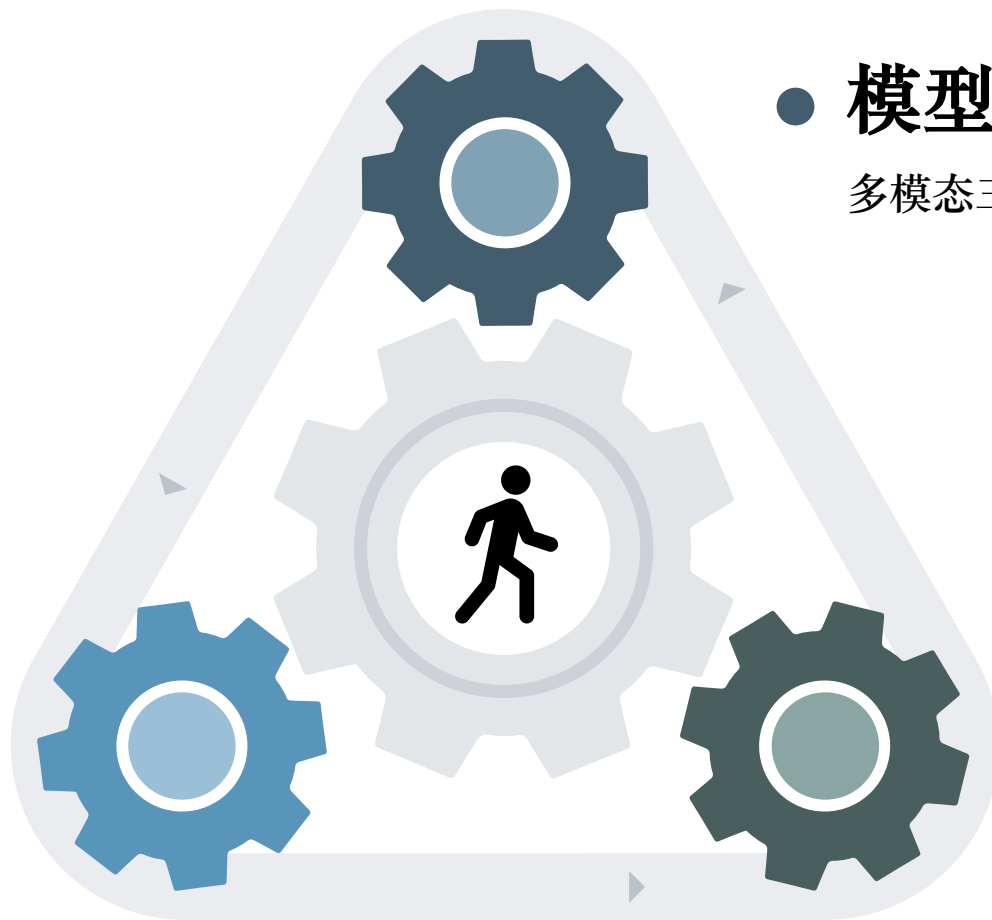
面向人-机-环境协同共生的模型、系统及应用 (2024年度工作)



面向人-机-环境协同共生的模型、系统及应用

● 系统

以人为中心的实时感知、
重建、预测系统



● 模型

多模态三维场景感知基础模型

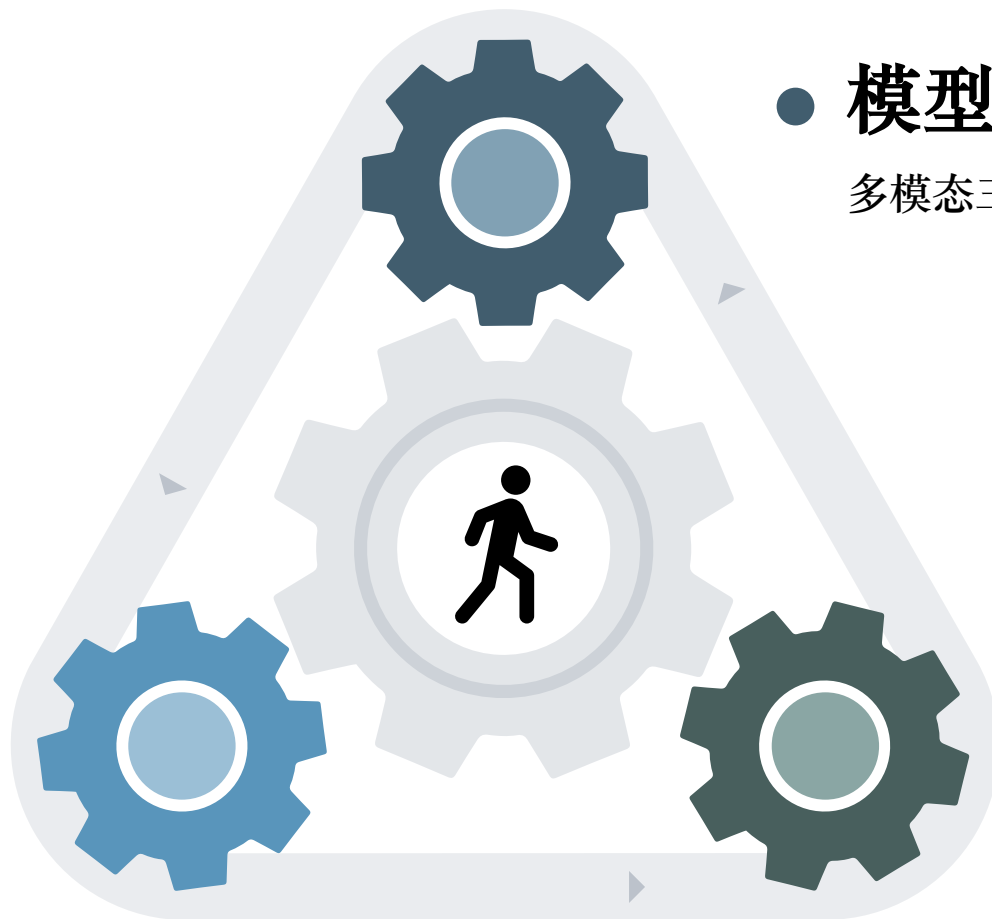
● 应用

类人具身智能应用

面向人-机-环境协同共生的模型、系统及应用

● 系统

以人为中心的实时感知、
重建、预测系统



● 模型

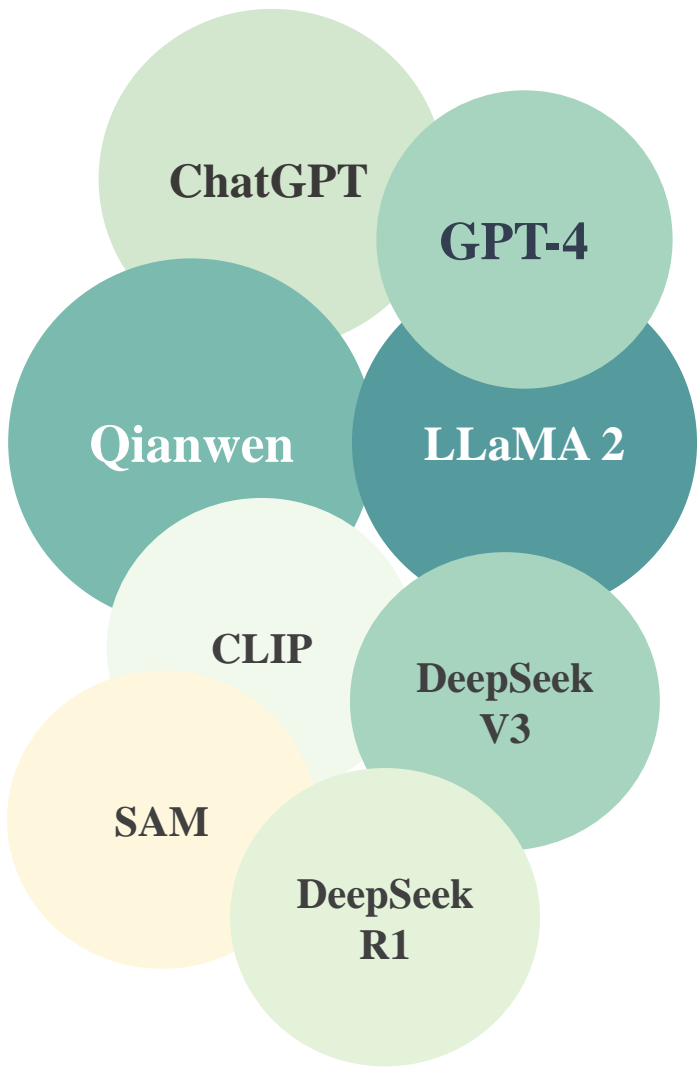
多模态三维场景感知基础模型

● 应用

类人具身智能应用

● 模型

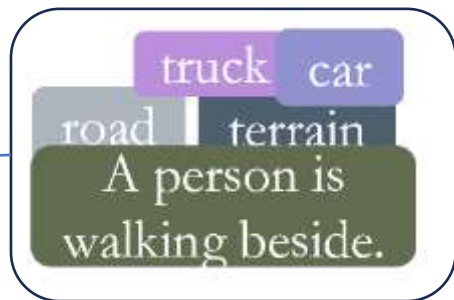
多模态三维场景感知基础模型



多模态特征空间对齐



提出有效的多模态特征空间对齐的技术范式，将大模型的通识能力迁移到三维点云，进行开放世界下的三维感知。



语言



图像

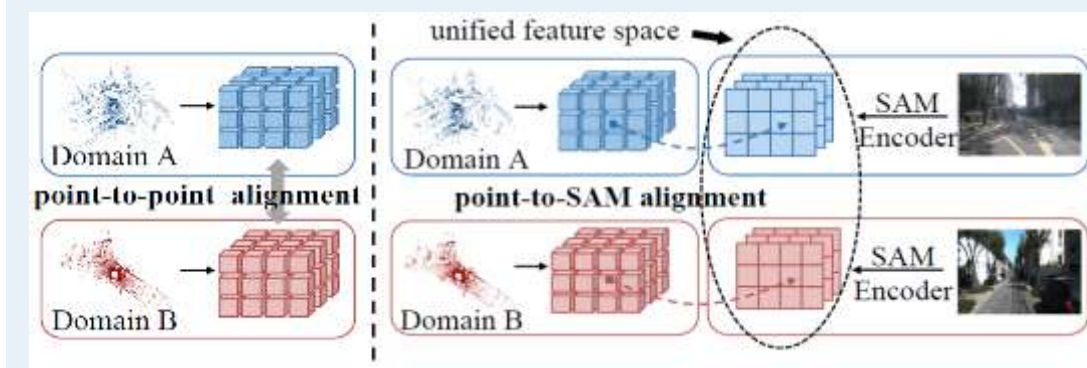


点云

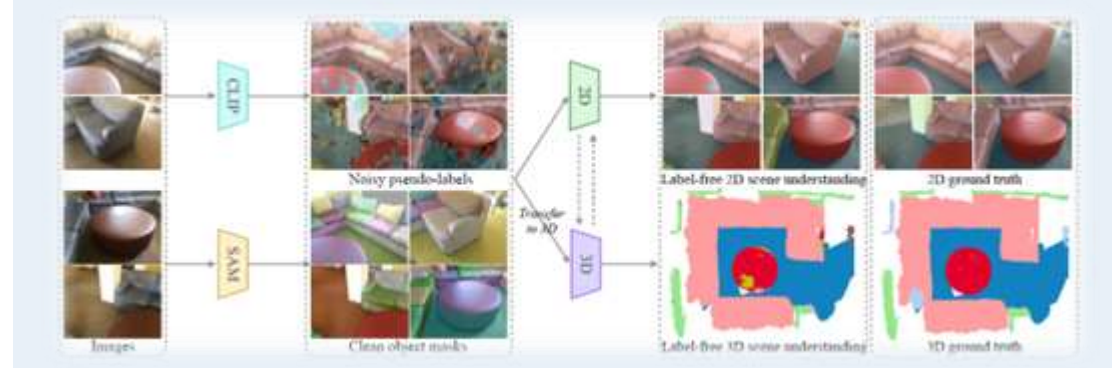
● 模型 - 突破三维数据瓶颈，迈向通用三维感知

多模态三维场景感知基础模型

多模态三维感知域迁移 (ECCV'24)



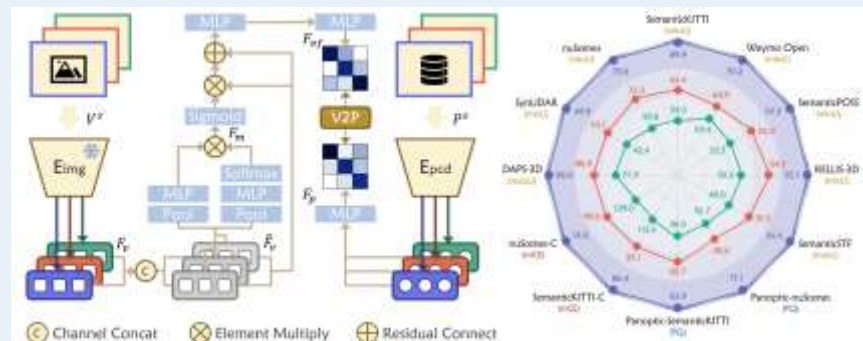
多模态无标签三维感知 (NeurIPS'24)



多模态三维视觉语言定位 (ECCV'24)



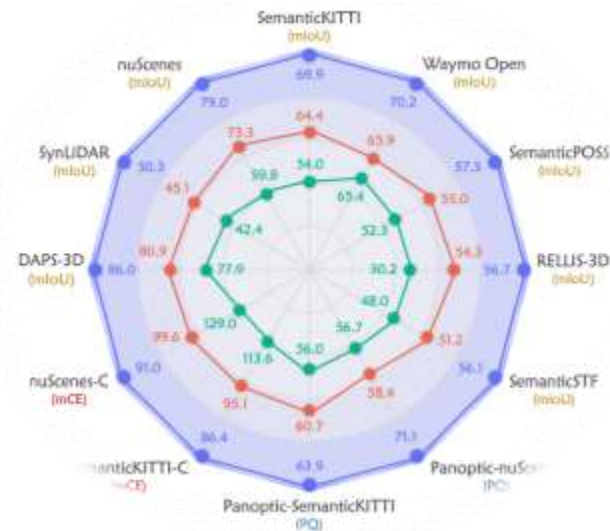
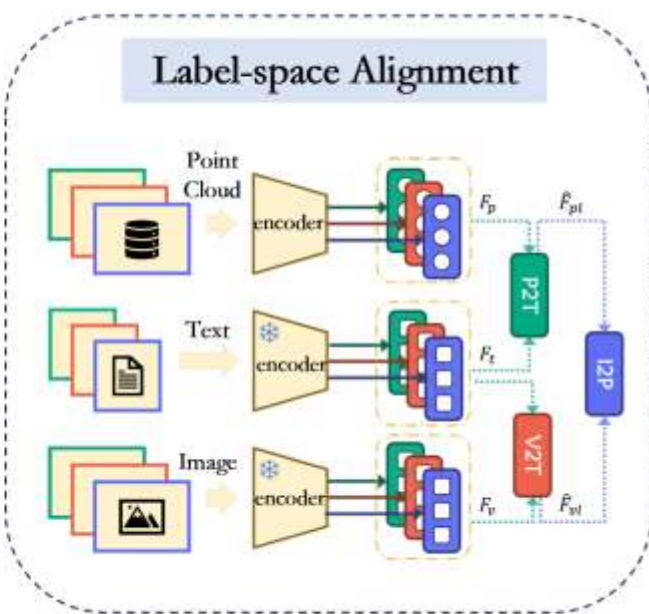
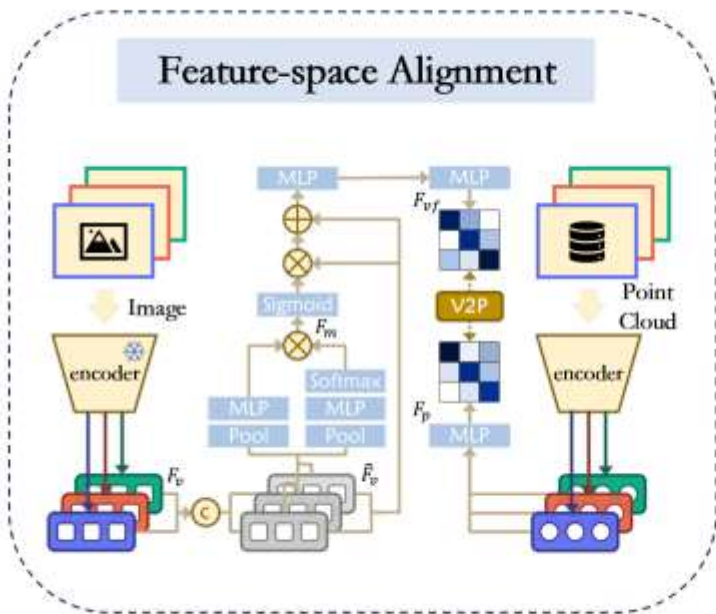
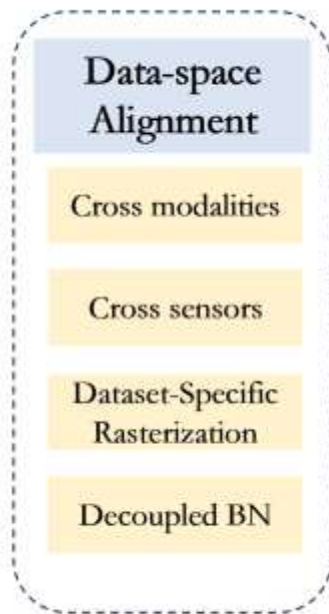
多模态三维感知基础模型 (CVPR'24)



● 模型 - 突破三维数据瓶颈，迈向通用三维感知

多模态三维场景感知基础模型

一个模型，一套参数解决多模态，多数据集，多任务下的三维场景感知

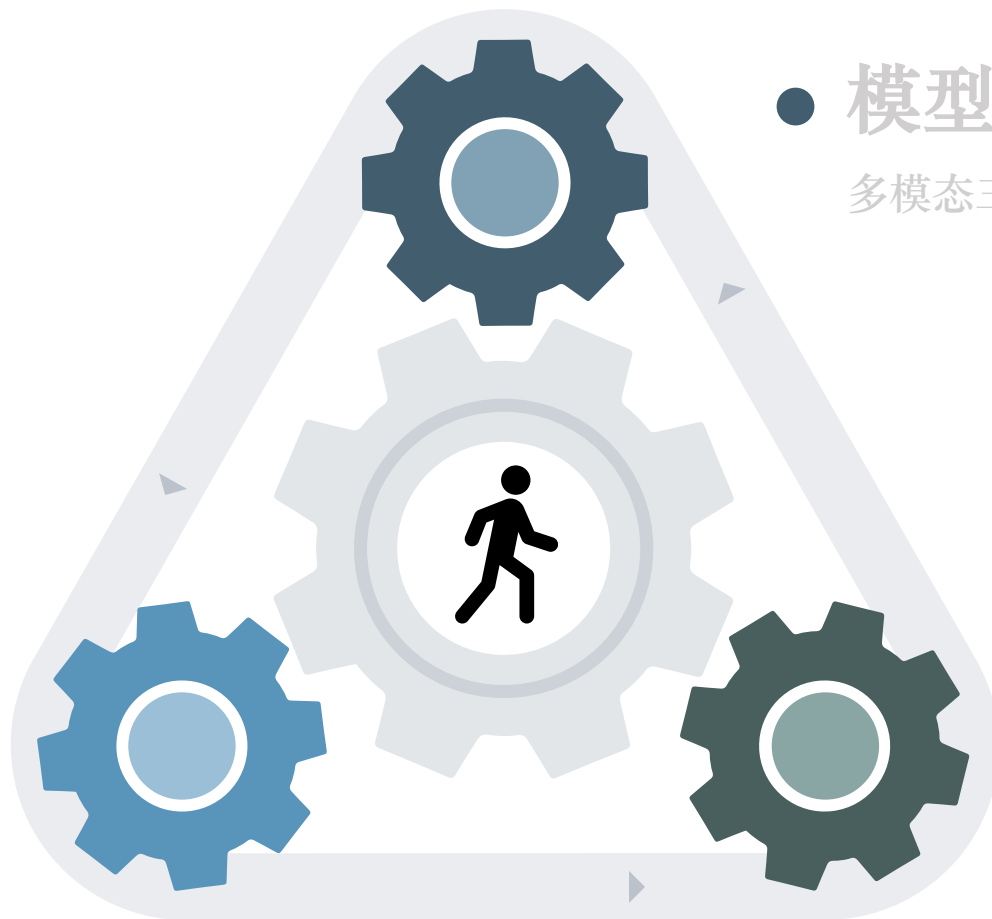


Multi-Space Alignments Towards Universal LiDAR Segmentation, Liu et al. CVPR, 2024

面向人-机-环境协同共生的模型、系统及应用

● 系统

以人为中心的实时感知、
重建、预测系统



● 模型

多模态三维场景感知基础模型

● 应用

类人具身智能应用

● 系统

以人为中心的实时感知、重建、预测系统

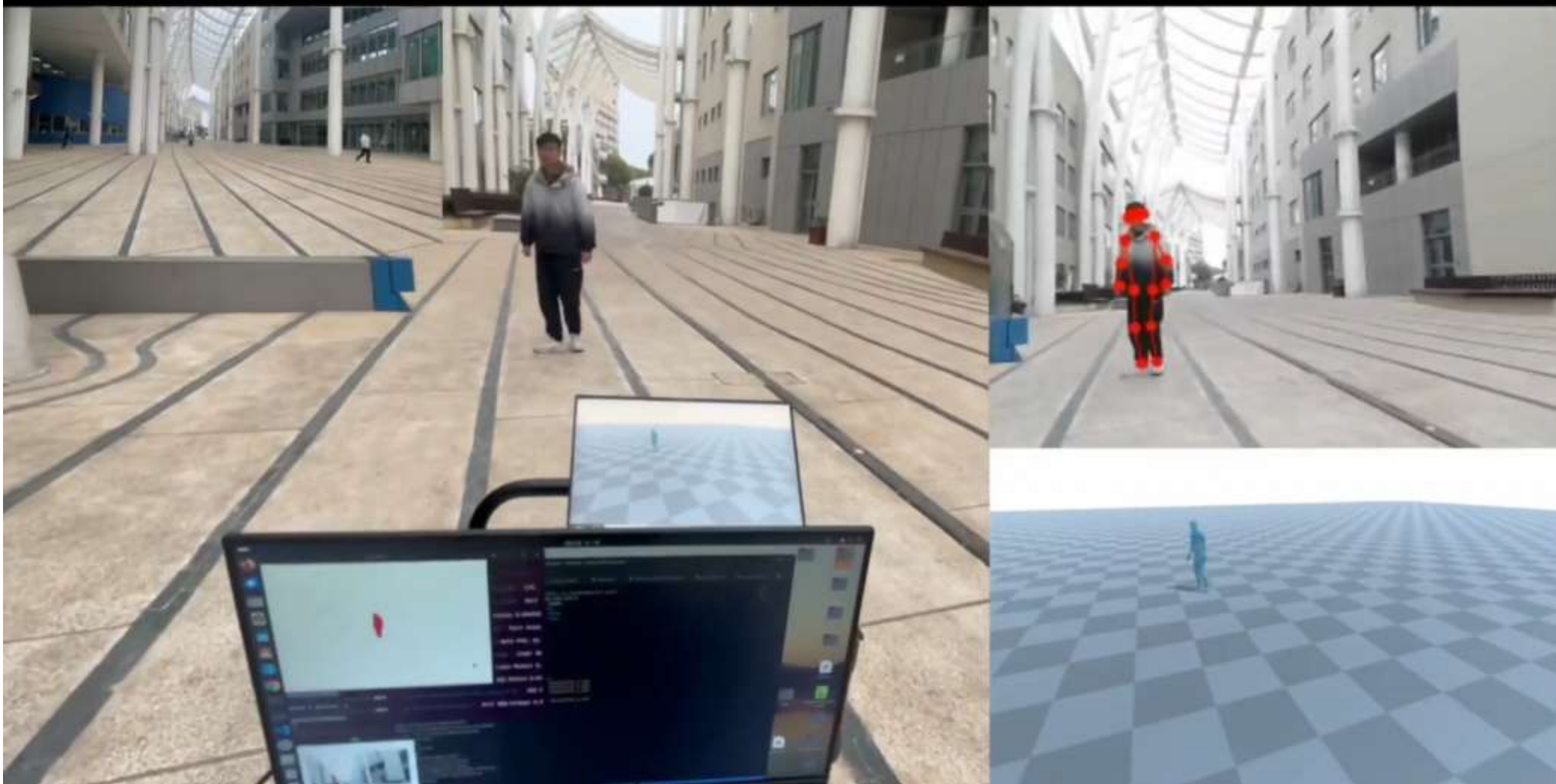
- 提出第一个面向以人为中心任务的通用人体表征 (CVPR'24)
- 提出第一个以人为中心的三维场景感知预训练模型 (CVPR'24)
- 提出第一个基于激光雷达点云的大场景人体步态识别算法 (ACM MM'24 Best Paper Candidate)
- 提出第一个基于激光雷达的大场景多人动作捕捉实时算法 (CVPR'24 Highlight, ECCV'24 Oral)
- 提出第一个三维场景下基于原始传感器信号的人体动作预测算法 (MM'24 Oral)



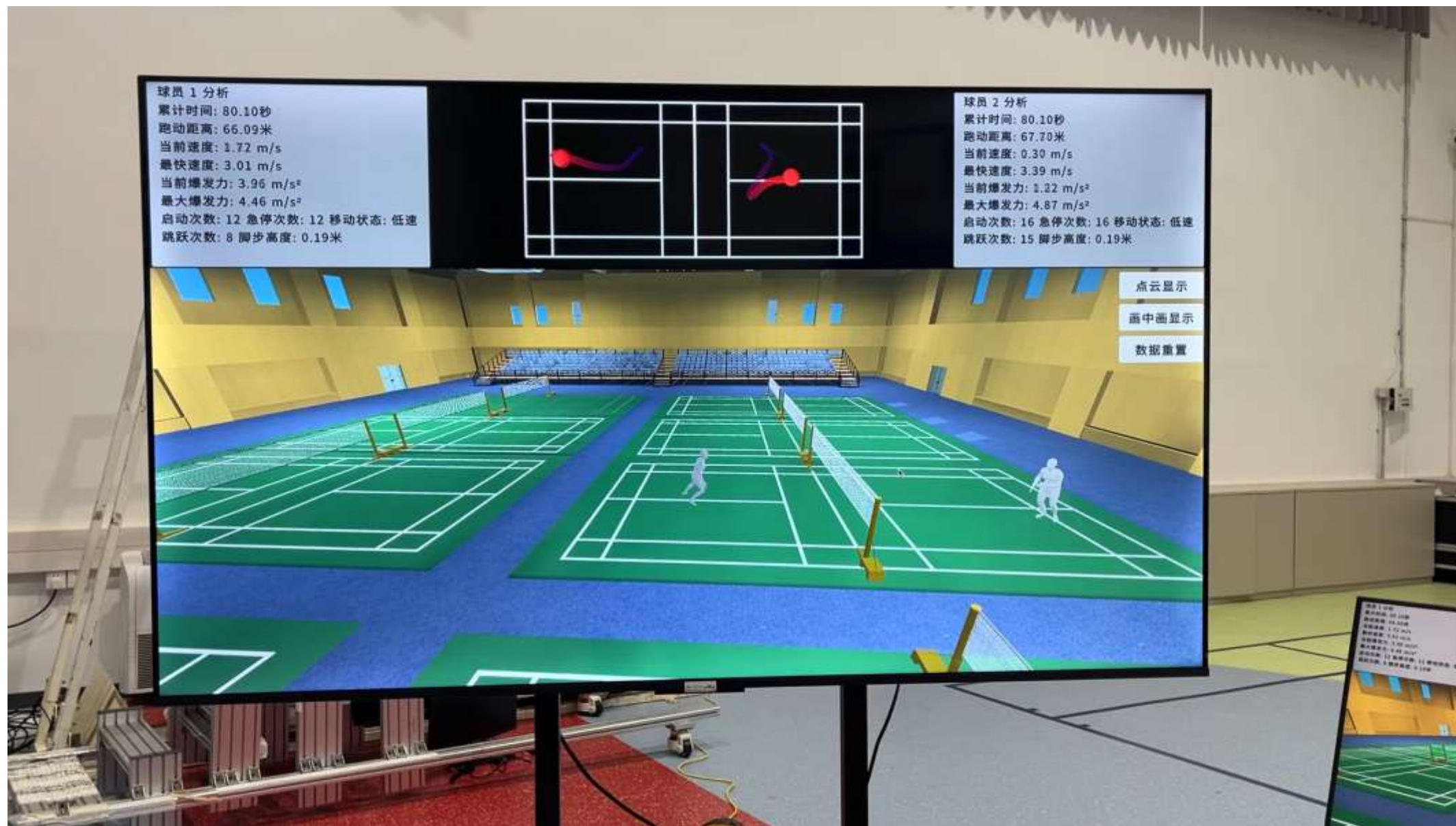
以人为中心的实时感知、重建、预测系统

● 系统展示 - 实时、鲁棒、精准、轻量化、易部署

Long trajectory tracking of human in robot view



落地应用展示 - 智慧体育 (已服务于国家队训练)



面向人-机-环境协同共生的模型、系统及应用

● 系统

以人为中心的实时感知、
重建、预测系统



● 模型

多模态三维场景感知基础模型

● 应用

类人具身智能应用

● 应用 - 人形机器人

具身智能应用

将捕获的人类行为迁移到类人机器人上，助力人机协同共生

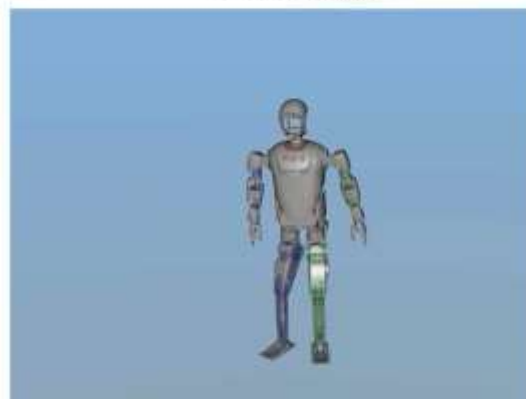


Real Person



Motion Captured Point Cloud

Speed 1.0x



Retargeted Robot Motion

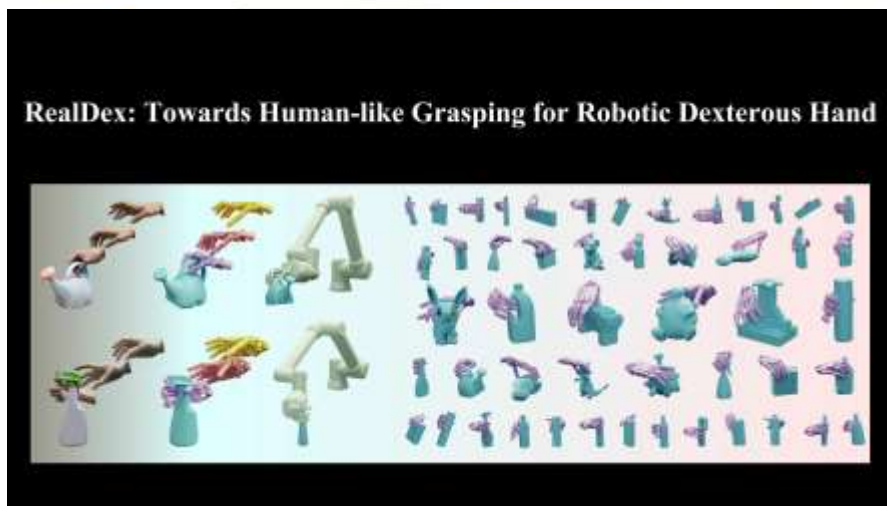


Physical Regulated Robot Motion

● 应用 - 五指灵巧手

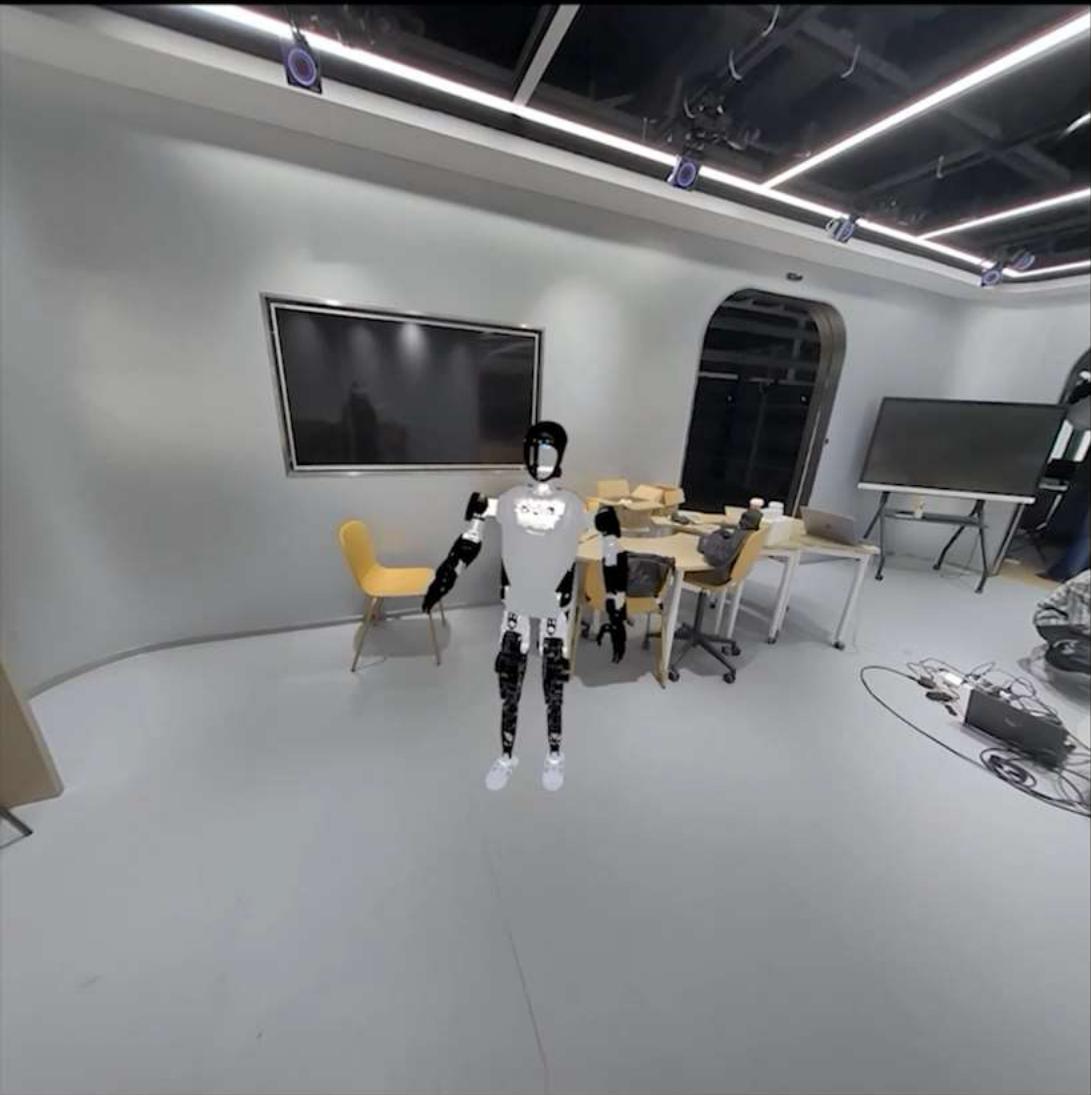
具身智能应用

将捕获的人类行为迁移到类人机器人上，助力人机协同与共生



(IJCAI'24 Oral)

Speed 1.0x

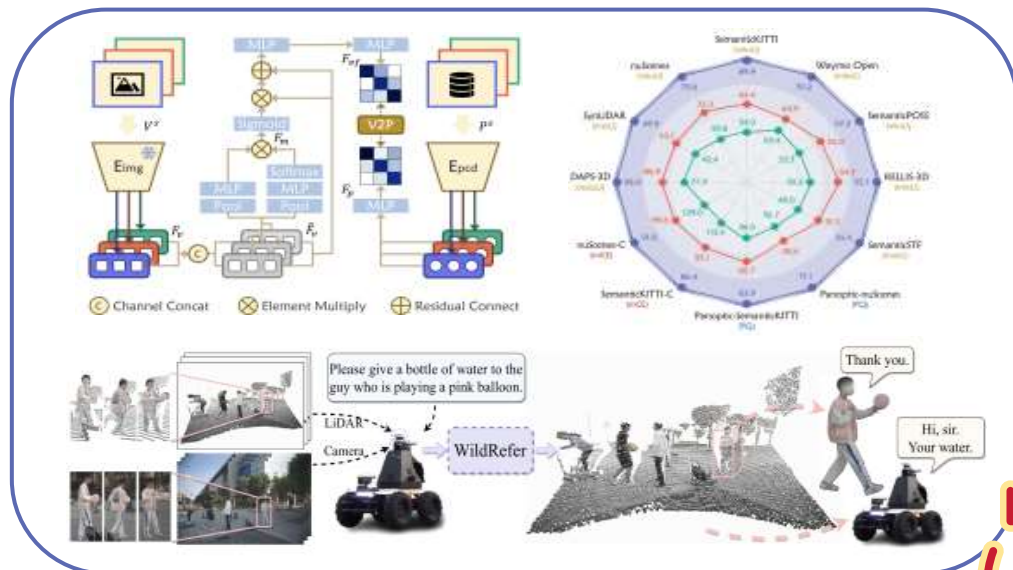


First-person Perspective

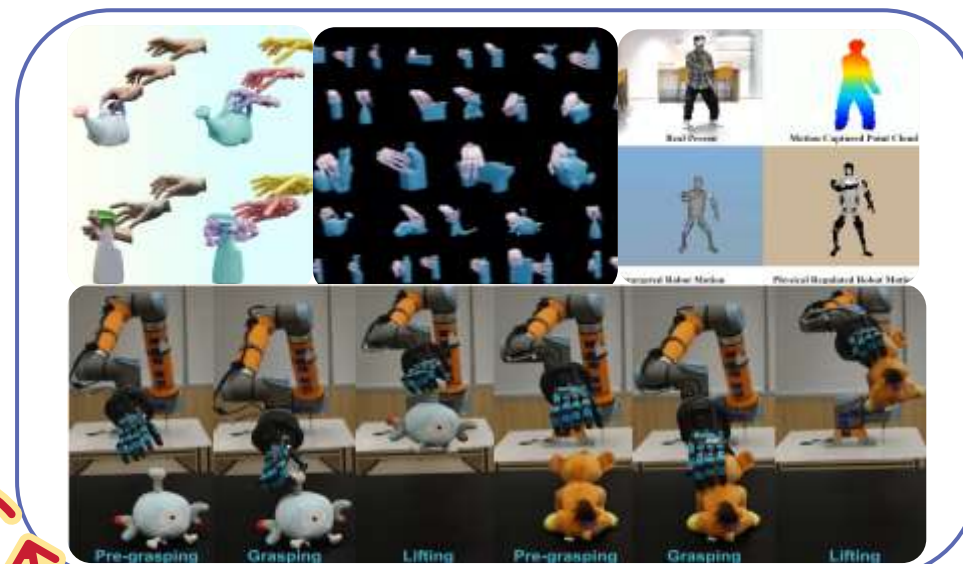


Third-person Perspective

面向人-机-环境协同共生的模型、系统及应用



1. 多模态三维场景感知基础模型



3. 具身智能应用



2. 以人为中心的实时感知、重建、预测系统

快乐科研
幸福生活

4DV Lab
— 4D VISION —

