



空间智能概述

彭思达
浙江大学

空间智能的定义

- 空间智能是**AI系统**在**三维空间**中感知、理解、推演、交互的能力

空间智能的关键研究内容

空间智能系统构建

感知

理解

推演

交互

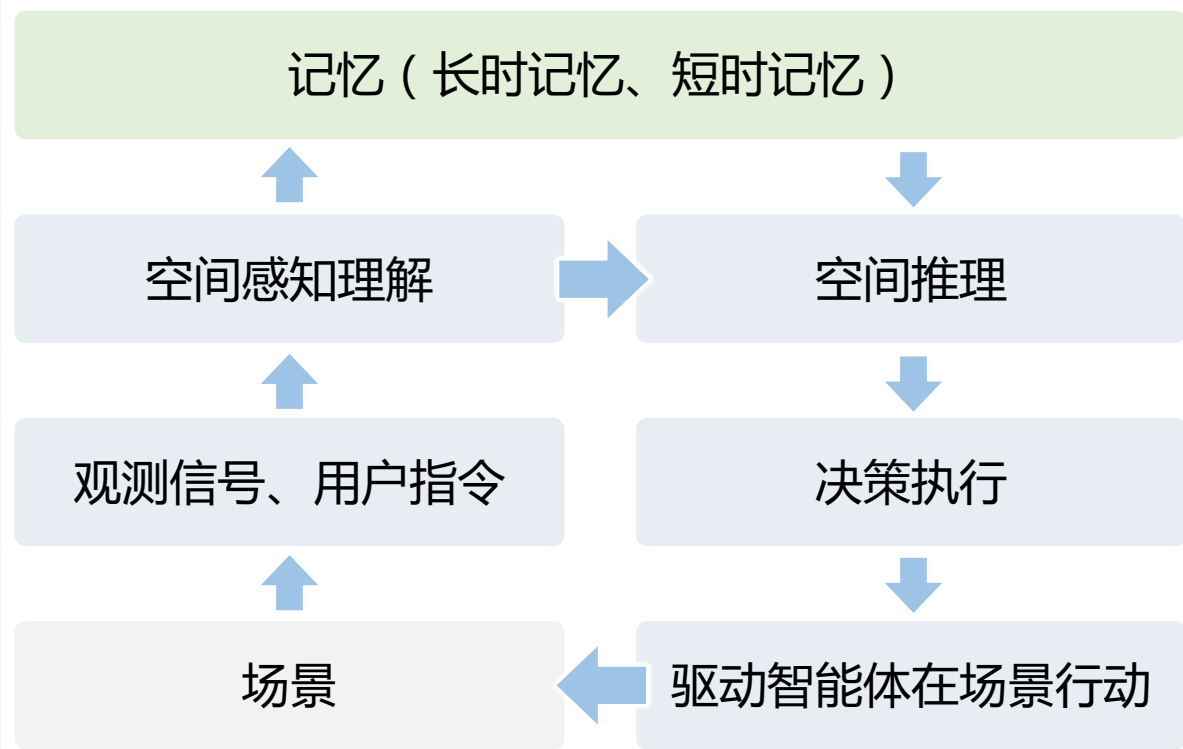
空间智能系统训练

仿真环境构建

模型仿真训练

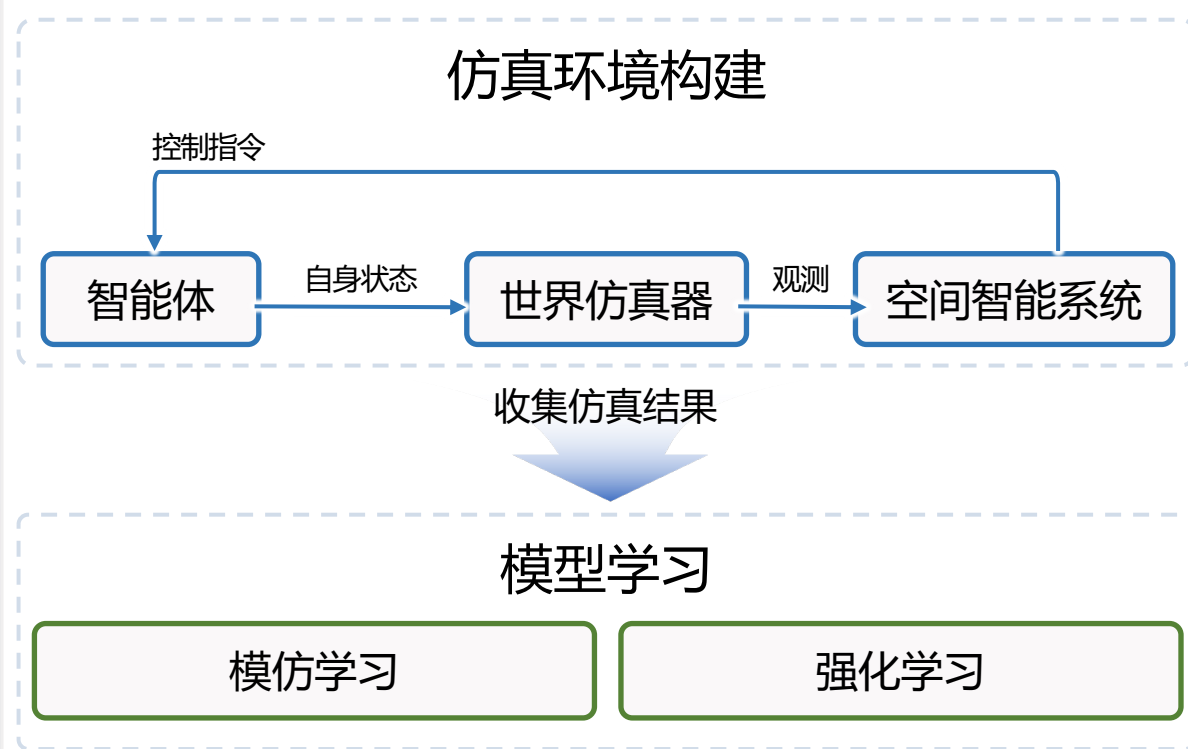
空间智能中各研究内容的关联

空间智能系统



该流程不一定分步实现，也可能内化在同一个模型中

空间智能系统训练



空间智能的价值

具备实体



人形机器人



无人车



无人机



AI眼镜

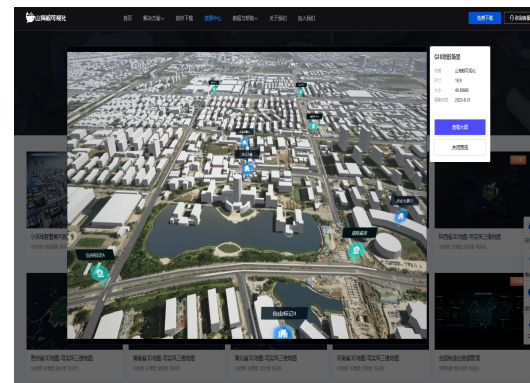
不具备实体



可控视频生成



游戏NPC



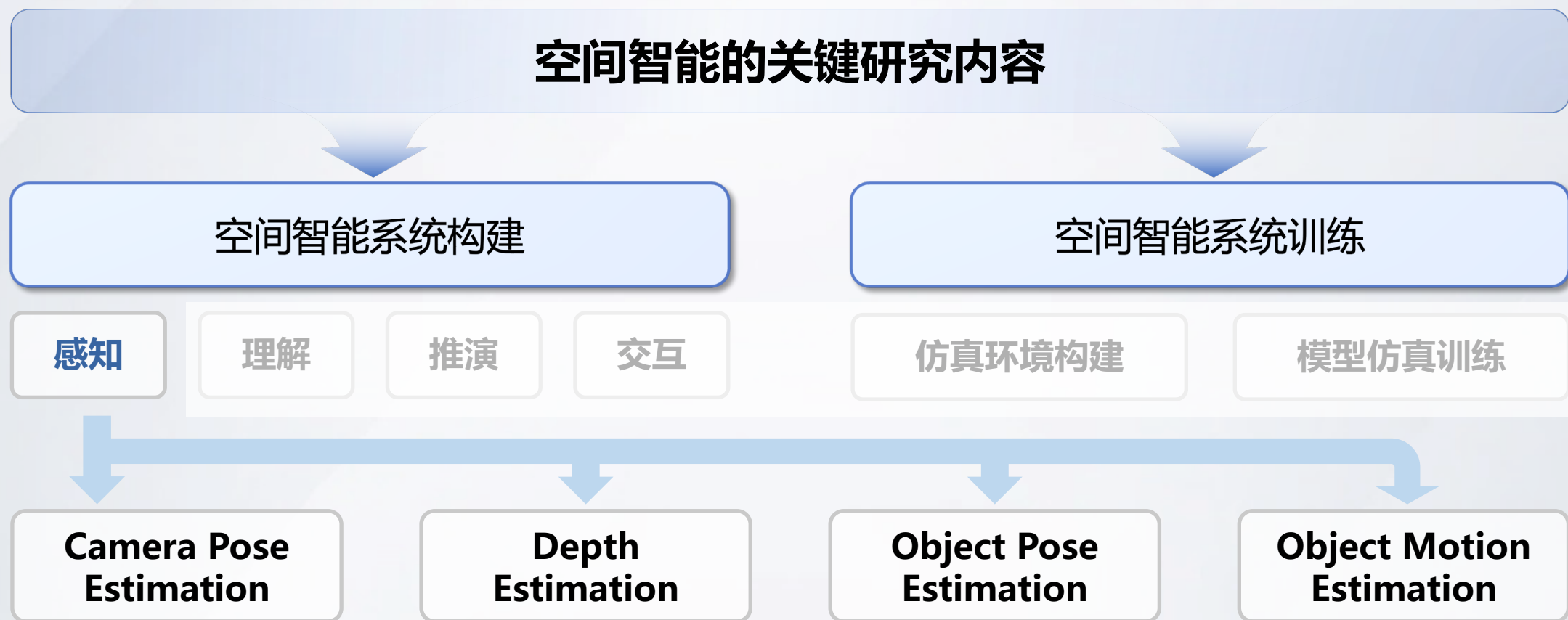
智能GIS系统



新一代地图

空间感知

- 致力于获取三维环境的底层信息，包括**位姿、距离、形状、运动**



空间感知热点回顾

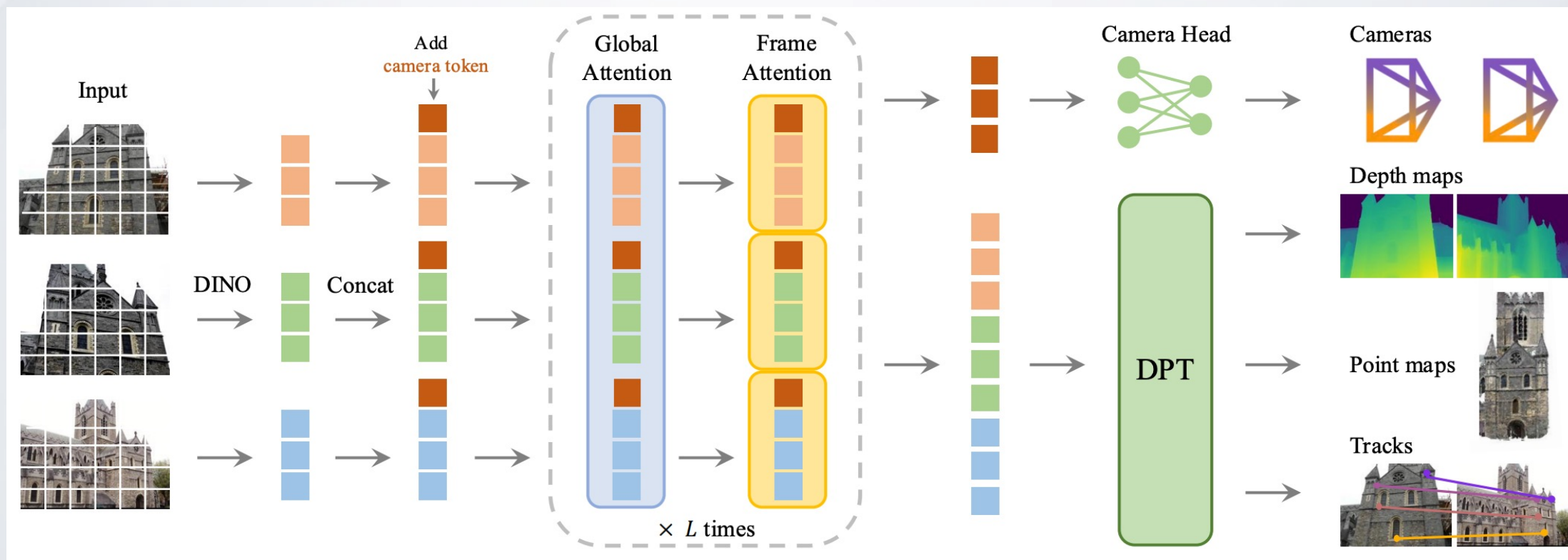
过去一年的技术趋势：基于ViT的视觉大模型

Camera Pose Estimation:
VGGT
(CVPR 2025)

Depth Estimation:
Depth Anything V2
(NeurIPS 2024)

Object Pose Estimation:
Cubify Anything
(CVPR 2025)

Object Motion Estimation:
Dynamic Point Maps
(arXiv 2025)



- 致力于从环境观测获取**包含空间关系的语义信息**

空间智能的关键研究内容

空间智能系统构建

感知

理解

推演

交互

空间智能系统训练

仿真环境构建

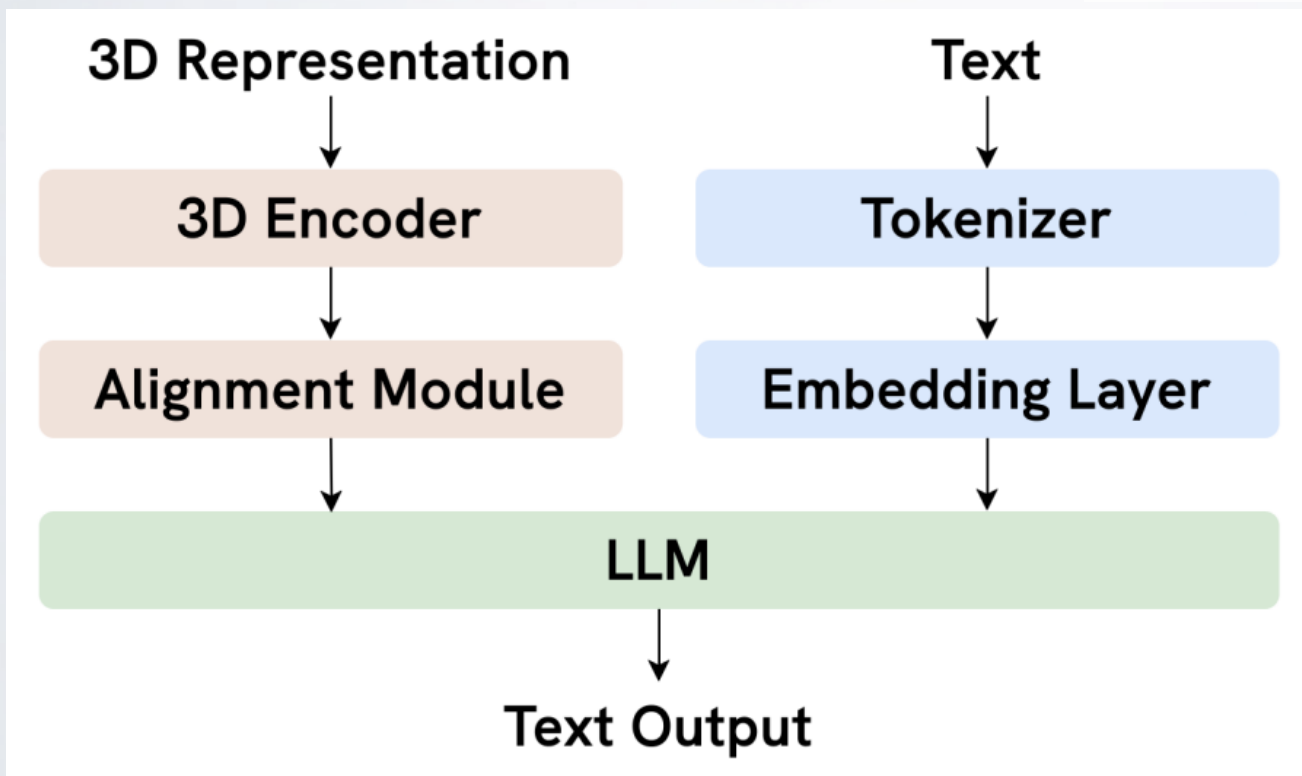
模型仿真训练

空间理解热点回顾

过去一年的技术趋势：多模态语言模型

范式1: 将场景表示为三维点云，使用VLM处理
When LLMs step into the 3D World
(arXiv 2024)

范式2: 将场景表示为视频，使用VLM处理
Apollo
(arXiv 2024)

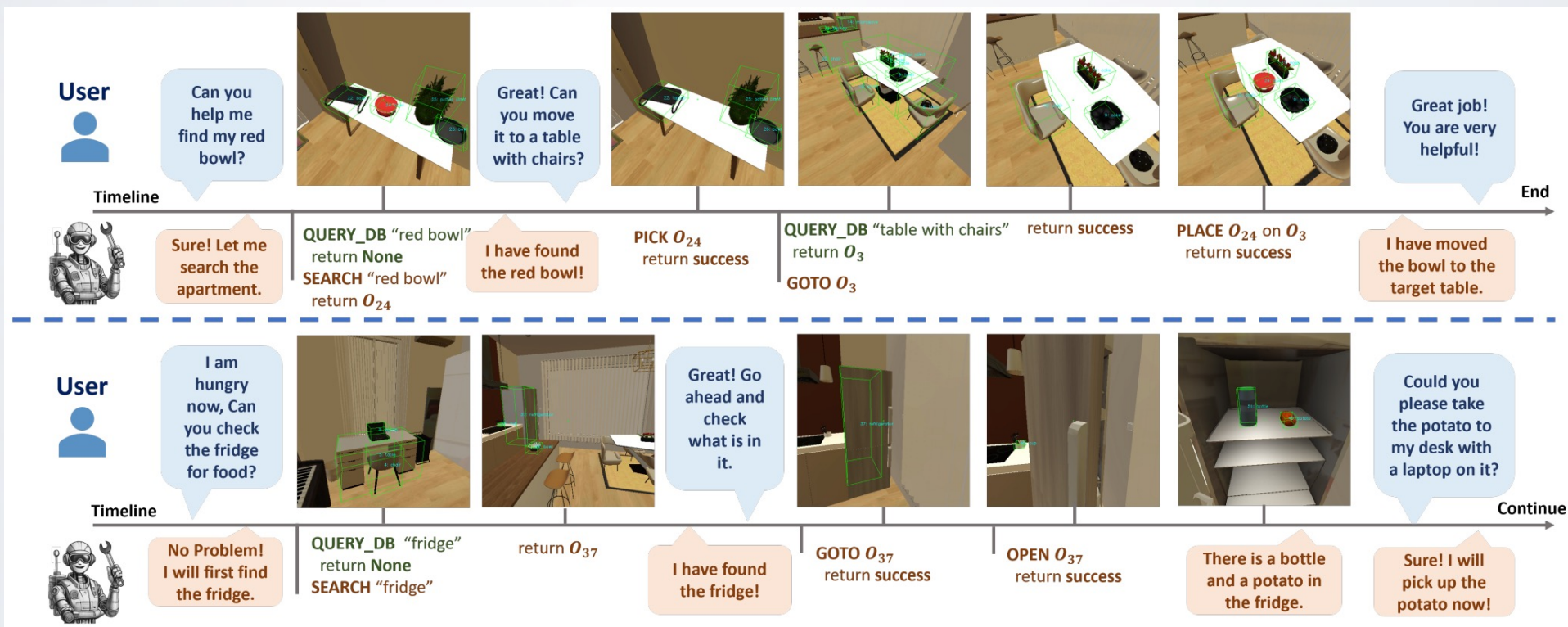


空间理解热点回顾

过去一年的技术趋势：多模态语言模型

范式1: 将场景表示为三维点云，使用VLM处理
When LLMs step into the 3D World
(arXiv 2024)

范式2: 将场景表示为视频，使用VLM处理
Embodied VideoAgent
(arXiv 2025)



- 致力于基于环境观测进行**空间信息相关的逻辑推理或预测**

空间智能的关键研究内容

空间智能系统构建

感知

理解

推演

交互

空间智能系统训练

仿真环境构建

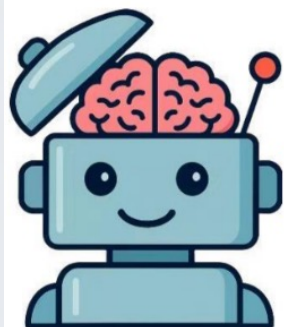
模型仿真训练

空间推演热点回顾

过去一年的技术趋势：思维链、视频生成模型

范式1: VLM with Thinking mode
Embodied-R
(arXiv 2025)

范式2: Video Generation
Navigational world model
(CVPR 2025)



Embodied-R



(Example) Question: Navigation Instruction given at initial position: [Observe around, then fly towards the road, then turn left and land on the roof of the building on the left]. What are you doing now?

Choose: A. I look around the square area. B. I fly towards the road.
C. I turn left and land on the roof of the building on the left.
D. I fly over the park. E. I land.

E. I land. **Fast-Thinking** ❌

Think: Let's analyze the event sequence and the instruction step by step:

1. Initial Position: I start by observing around the square area.
2. Next Step: After observing, I fly towards the highway.
3. Subsequent Step: Once at the highway, I should turn left.
4. Final Step: Finally, I should land on the roof of the building on the left.

From the video frames:

- The initial frames show buildings and a square area, indicating the phase.
- The subsequent frames show a transition from the square area to a broader view.
- The **later frames depict a clear view of the road** and surrounding buildings.

Answer: B. I fly towards the road. ✅

空间推演热点回顾

过去一年的技术趋势：思维链、视频生成模型

范式1: VLM with Thinking mode
Embodied-R
(arXiv 2025)

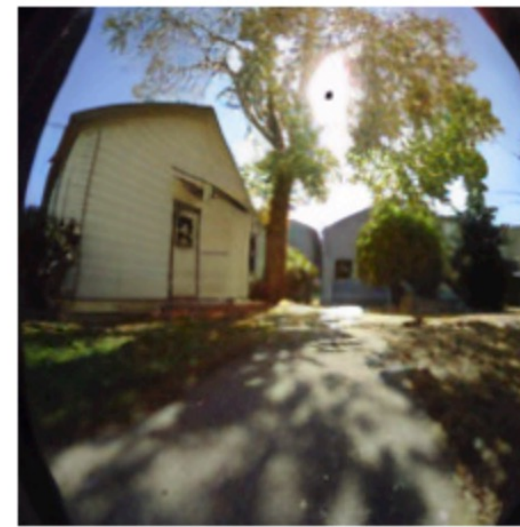
范式2: Video Generation
Navigational world model
(CVPR 2025)

navigation action and time
 $(\Delta x, \Delta y, \Delta \phi, k)$



Conditional
Diffusion
Transformer

model output



- 在物理或虚拟环境中执行某些动作，以实现**导航、操作、协作**等目标

空间智能的关键研究内容

空间智能系统构建

感知

理解

推演

交互

空间智能系统训练

仿真环境构建

模型仿真训练

空间交互热点回顾

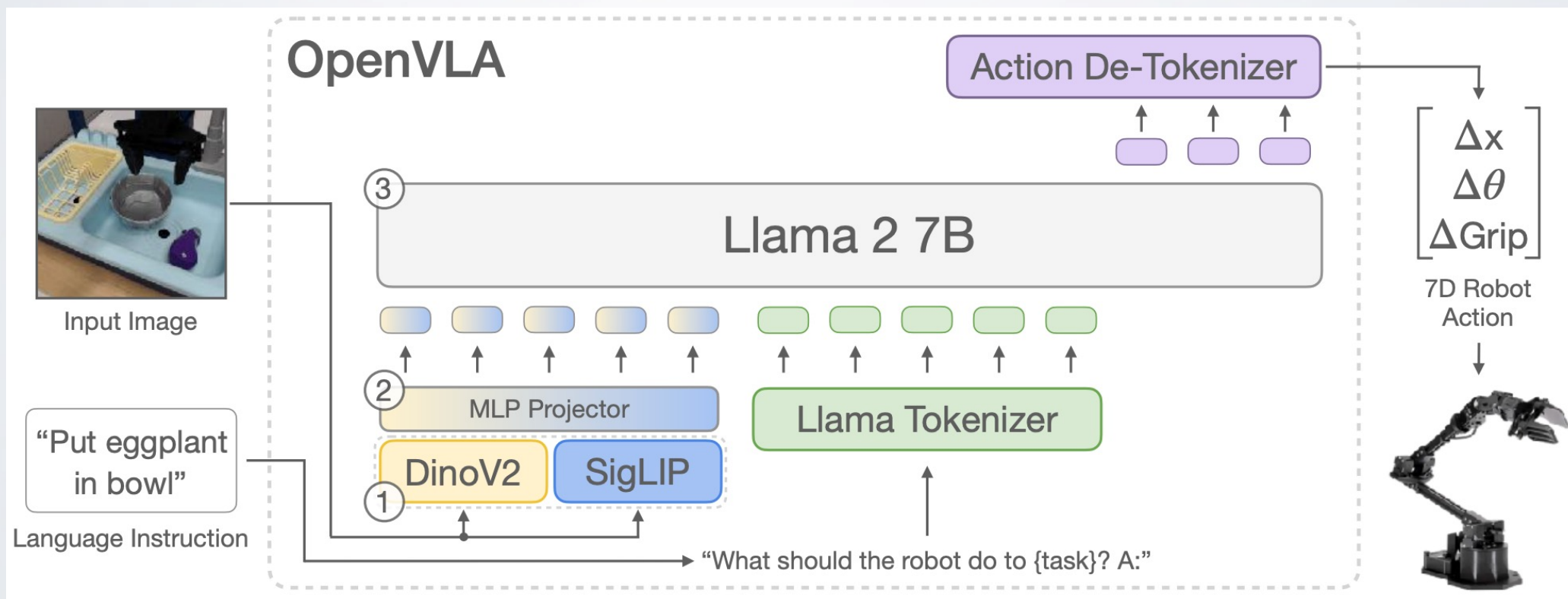
过去一年的技术趋势：基于VLM的端到端模型

机械臂:
OpenVLA
(CoRL 2024)

无人机:
OpenUAV
(ICLR 2025)

地面导航机器人:
LH-VLN
(CVPR 2025)

智能眼镜:
EgoLife
(CVPR 2025)



空间交互热点回顾

过去一年的技术趋势：基于VLM的端到端模型

机械臂:
OpenVLA
(CoRL 2024)

无人机:
OpenUAV
(ICLR 2025)

地面导航机器人:
LH-VLN
(CVPR 2025)

智能眼镜:
EgoLife
(CVPR 2025)

记忆（长时记忆、短时记忆）

空间感知理解

空间推理

观测信号、用户指令

决策执行

场景

驱动智能体在场景行动

空间智能的关键研究内容

空间智能系统构建

感知

理解

推演

交互

空间智能系统训练

仿真环境构建

模型仿真训练

空间智能系统训练

模仿学习

强化学习

收集仿真结果

空间智能系统仿真

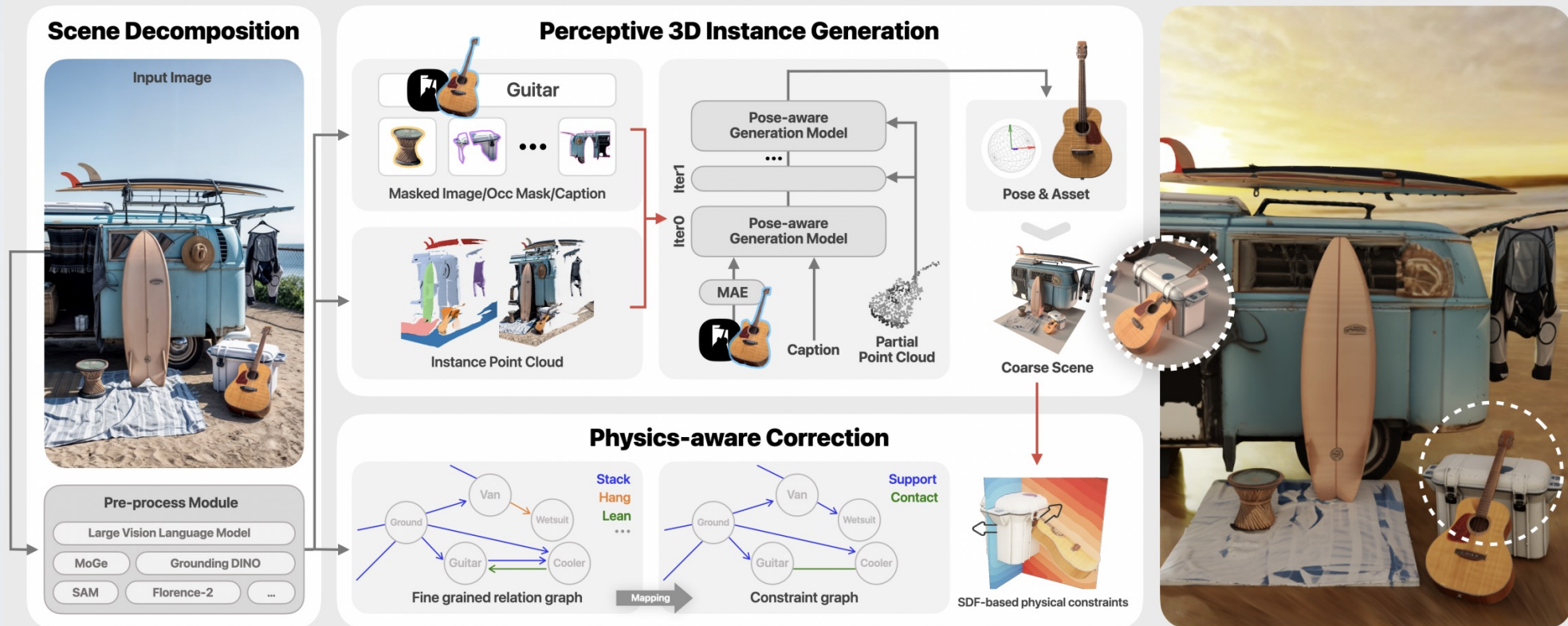


仿真环境构建热点回顾

过去一年的技术趋势：基于Generative AI构建仿真环境

范式1: 基于3D Generation的显式仿真器
CAST
(SIGGRAPH 2025)

范式2: 基于Video Generation的隐式仿真器
Genie-3
(Released in 2025)



仿真环境构建热点回顾

过去一年的技术趋势：**基于Generative AI构建仿真环境**

范式1: 基于3D Generation的显式仿真器
CAST
(SIGGRAPH 2025)

范式2: 基于Video Generation的隐式仿真器
Genie-3
(Released in 2025)



过去一年的技术趋势：**构建闭环仿真环境，对Agent进行强化学习**

- RL之父发表文章：Welcome to the Era of Experience

Welcome to the Era of Experience

David Silver, Richard S. Sutton*

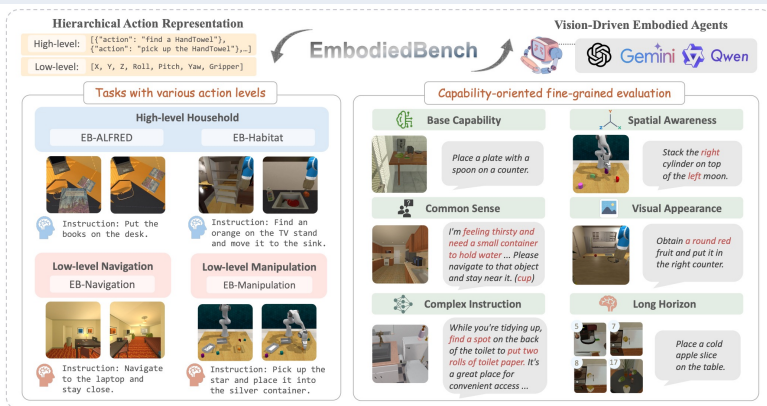
Abstract

We stand on the threshold of a new era in artificial intelligence that promises to achieve an unprecedented level of ability. A new generation of agents will acquire superhuman capabilities by learning predominantly from experience. This note explores the key characteristics that will define this upcoming era.

模型仿真训练热点回顾

过去一年的技术趋势：构建闭环仿真环境，对Agent进行强化学习

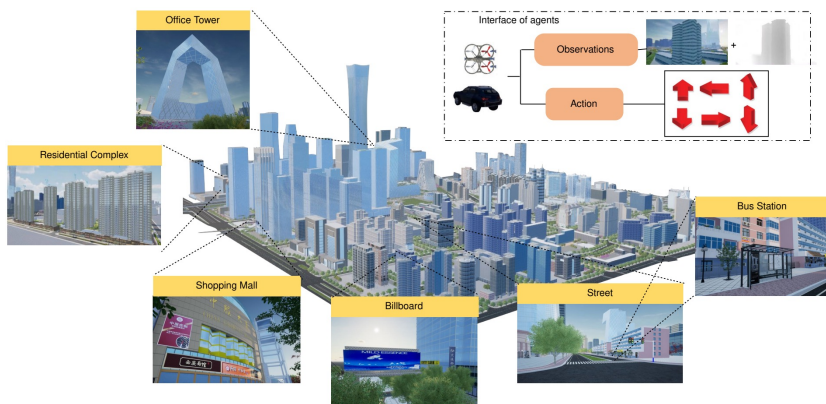
Embodied Bench (面向Spatial VLM)



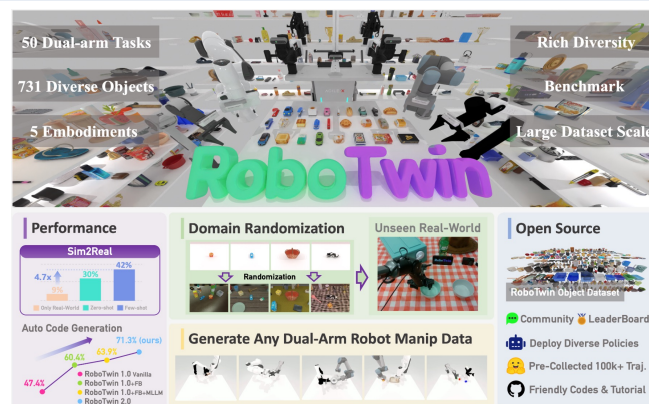
InternUtopia (面向各类机器人)



Embodied City (面向无人机)



RobotTwin (面向机械臂)





谢谢！

彭思达
浙江大学