智能机器人视觉感知：挑战和解决方案

——从数据角度出发

胡君杰
深圳市人工智能与机器人研究院

本报告涉及以下几个问题

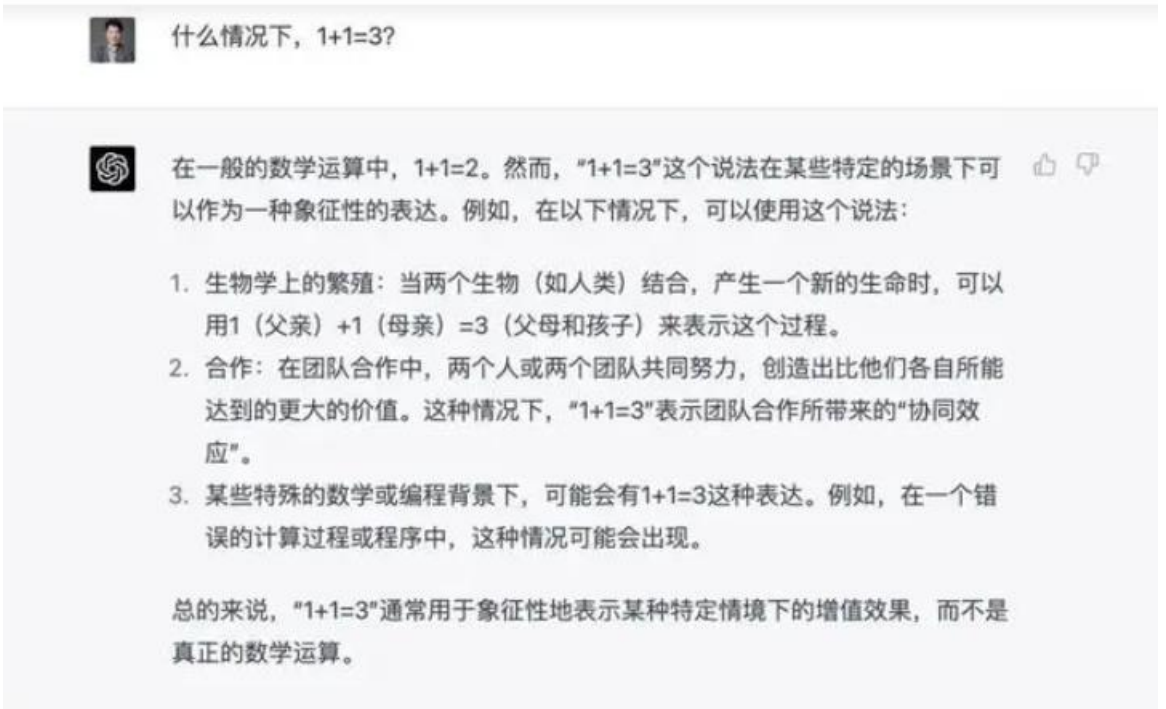
- 人工智能模型设计-解决方案
- 机器人视觉感知的挑战
- 机器人视觉感知的解决范式
- 机器人视觉感知未来趋势

人工智能研究前沿

大模型的数据闭环和解决方案

- Tesla纯视觉方案的无人驾驶Autopilot

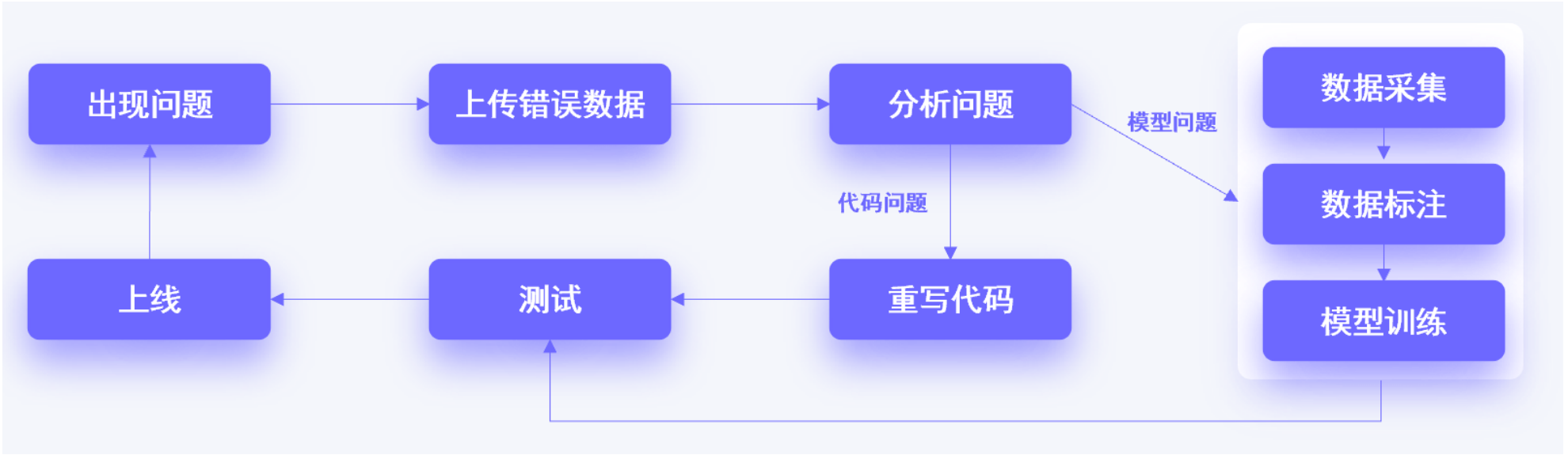
- OpenAI的网络机器人ChatGPT4



人工智能研究前沿

大模型的数据闭环和解决方案

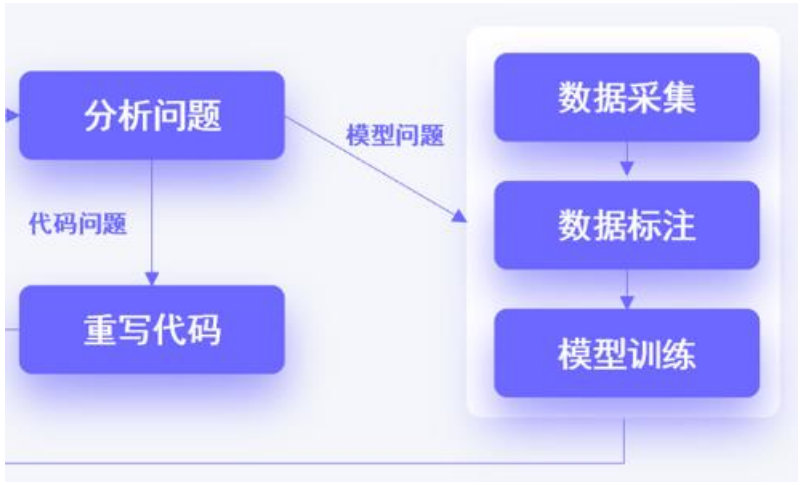
- Tesla纯视觉方案的无人驾驶Autopilot
- OpenAI的网络机器人ChatGPT4

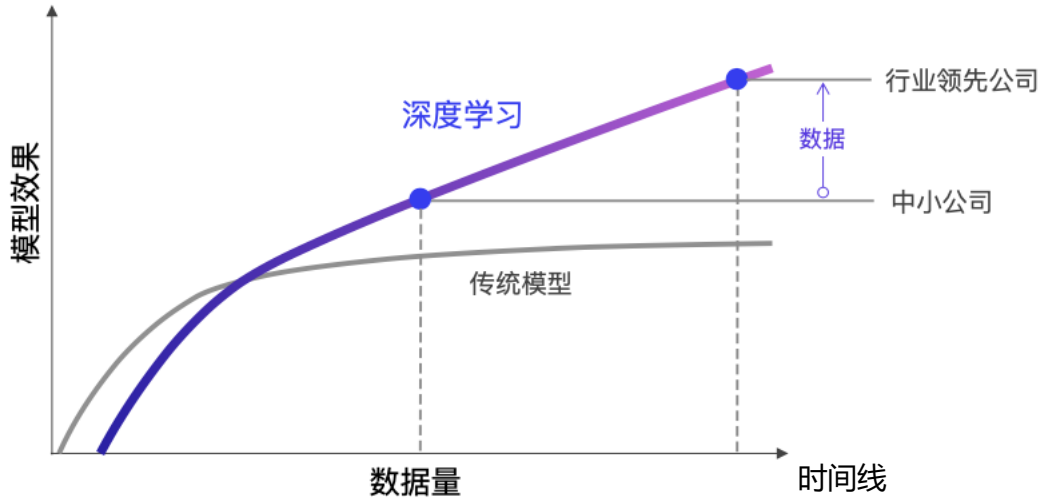


人工智能研究前沿

大模型的数据闭环和解决方案

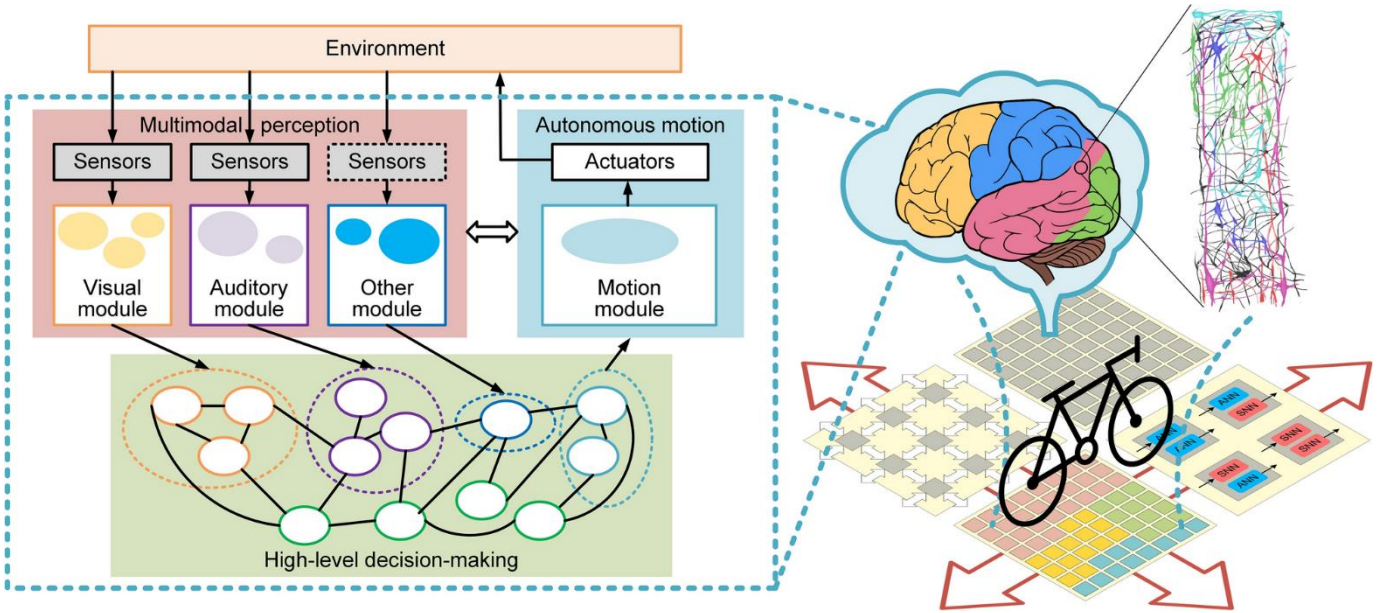
- 模型实现与数据成本的关系：随着数据量的上升，模型提升速度放缓
- 大公司依据数据闭环，通过不断提升数据量和数据质量，可以实现模型效果的持续提升，即量变产生质变





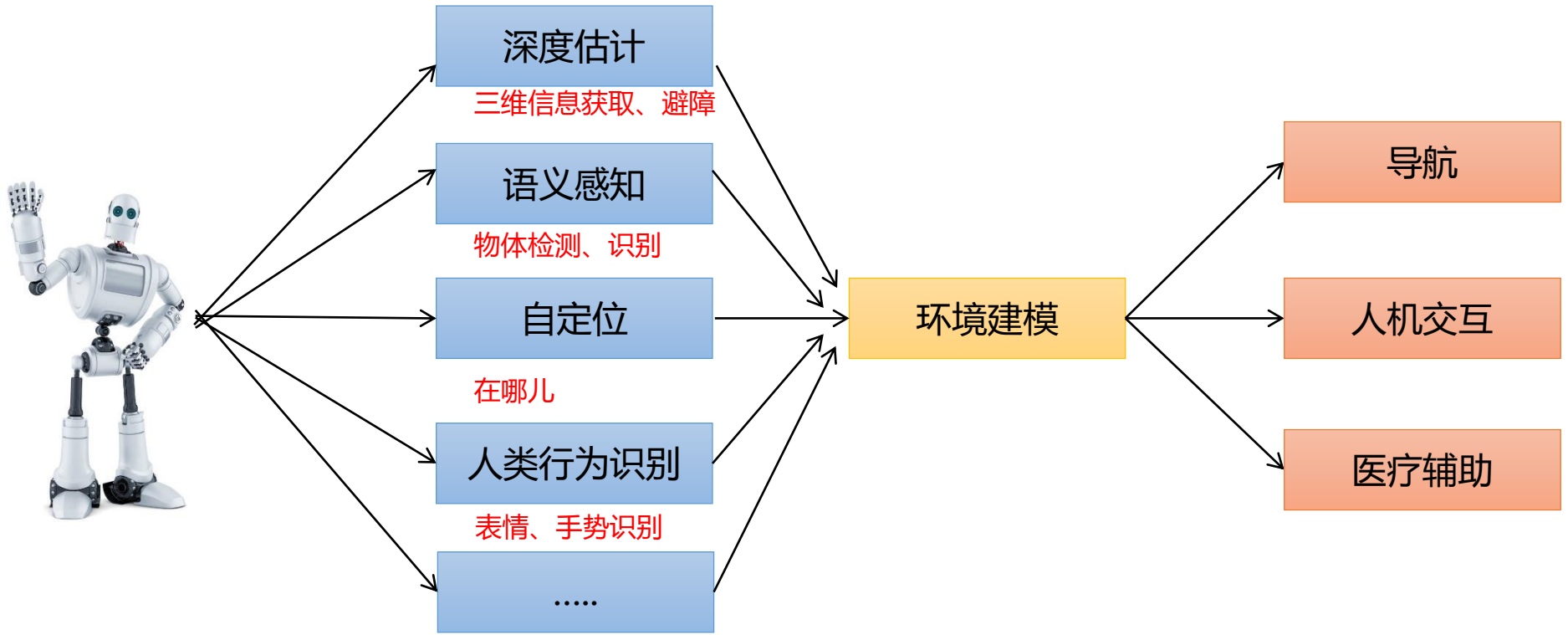
智能机器人-视觉感知

- 感知的目的在于机器人认识和建模真实三维世界
- 感知是机器人决策和控制的前提
- 视觉是感知最为重要和有效的途径



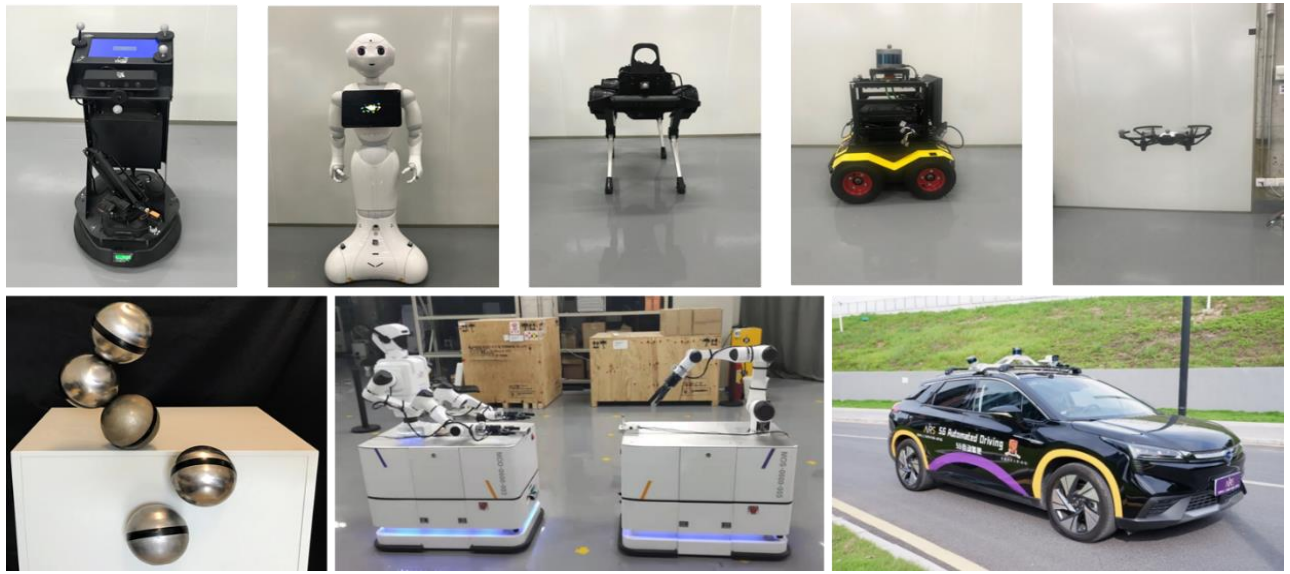
智能机器人-视觉感知

- 感知的目的在于使机器人认识和建模真实三维世界
- 感知是机器人决策和控制的前提
- 视觉是感知最为重要的组成部分



智能机器人感知：挑战

- 然而，以上数据闭环及模型实现方式，难以应用于实体机器人
  - 机器人类型差异



智能机器人感知：挑战

- 然而，以上数据闭环及模型实现方式，难以应用于实体机器人
  - 交互环境各式各样

城市

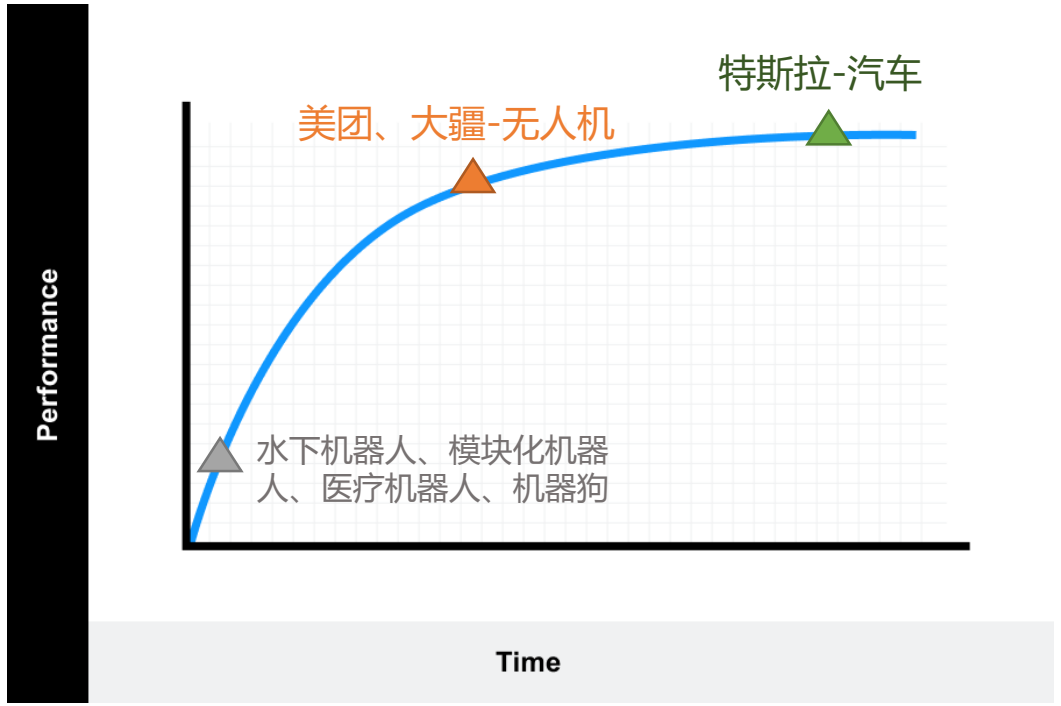
森林

海洋

智能机器人感知：挑战

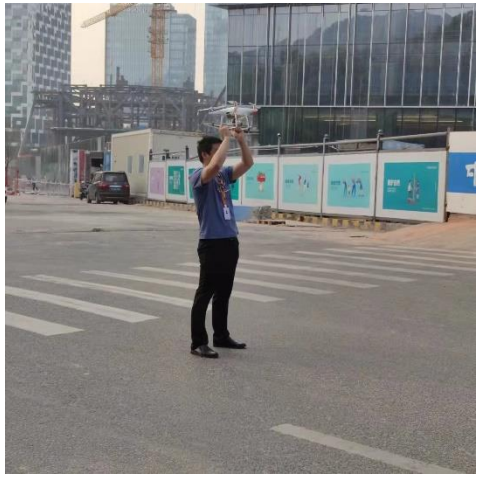
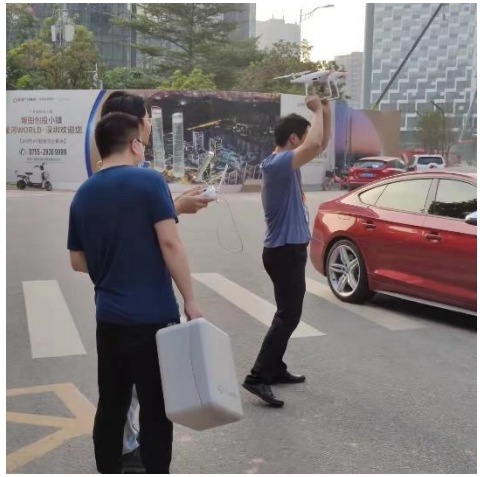
- 由于机器人类型和交互环境各样。数据闭环的方式由于成本原因，只能由大公司主导。当下的研究范式仍是
  - 基于特定机器人在特定交互环境下的算法设计。

数据	模型	问题
人工收集	有监督学习	需人工标注， 难以大量收集
在线交互	强化学习	机器人功能(如电池)限制，难以大规模收集， 只适用于部分任务



智能机器人感知：挑战

- 由于机器人类型和交互环境各样。数据闭环的方式由于成本原因，只能由大公司主导。当下的研究范式仍是
  - 基于特定机器人在特定交互环境下的算法设计。
  - 数据难以大量获取，以无人机在城市环境下环境语义感知为例：


智能机器人感知：从数据角度的解决方案

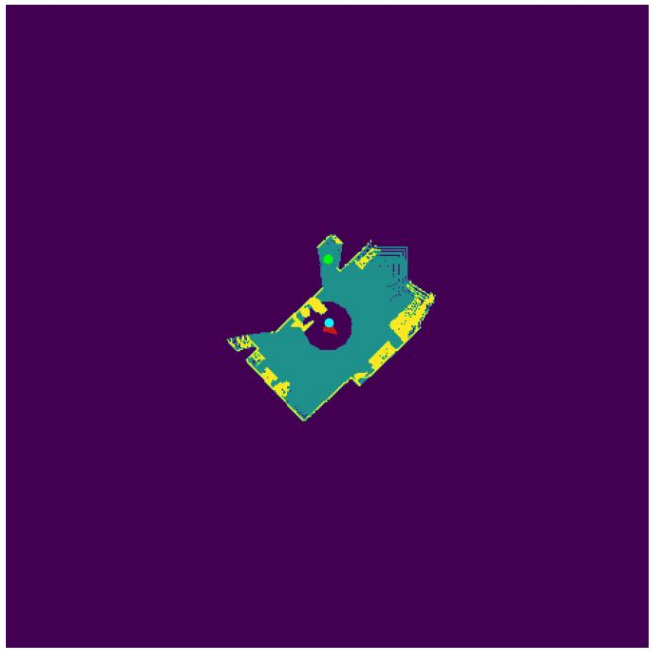
- 人类先验：手动设计规则/特征，只适用于部分任务。
- 无监督学习：无需人工数据标注，适用于low-level vision，如图像去噪、光照增强等。
- 半监督学习：仅需部分数据部分标注，适用于high-level perception，如语义分割、深度估计。
- 小样本学习：仅通过少量部分数据标注，进行深度模型学习。
- 知识蒸馏（模型迁移）：将大模型的感知能力赋予小模型，仅需图像输入。
- 终身学习：让模型持续学习，当有新数据时，仅需利用新数据进行模型训练并保留其原始感知能力。

智能机器人感知：从数据角度的解决方案

- **视觉SLAM**
  - 主动感知SLAM-人类先验
  - 黑暗条件下SLAM-无监督学习
  - 实时SLAM-知识蒸馏
- 机器人环境深度估计
  - 无监督学习
  - 知识蒸馏
  - 终身学习
- 无人机、多机器人环境语义识别
  - 半监督学习
  - 小样本学习
  - 多机器人语义SLAM

机器人主动SLAM-人类先验

- 手工设计各种规则，进行路径规划
- 门检测并构建拓扑地图以解决机器人路径震荡



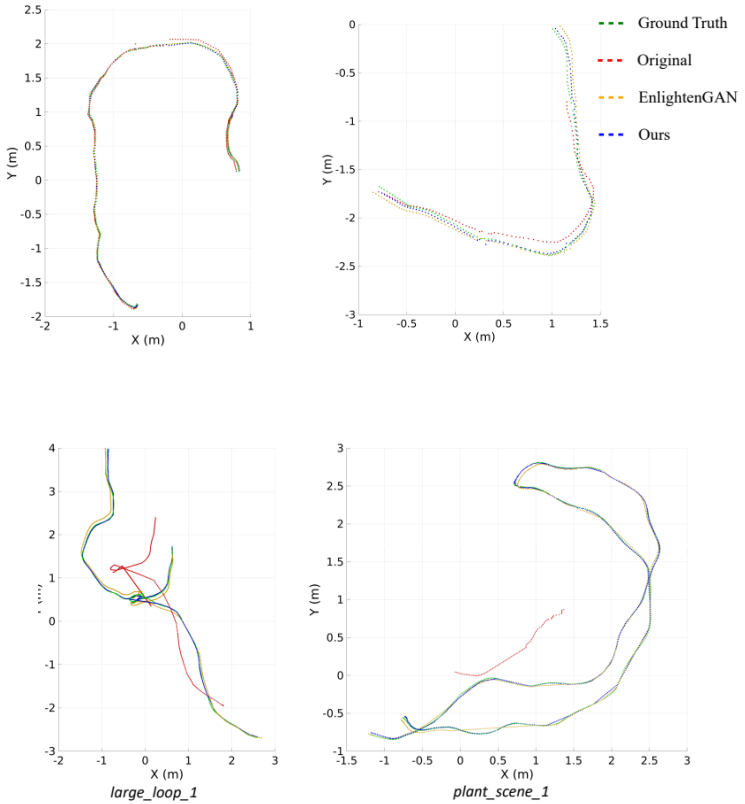
Bao, Hu et al. 2023

黑暗光照下SLAM-无监督光照增强

低光照环境下视觉SLAM

- 提出无监督对抗学习的低光照图像增强算法，

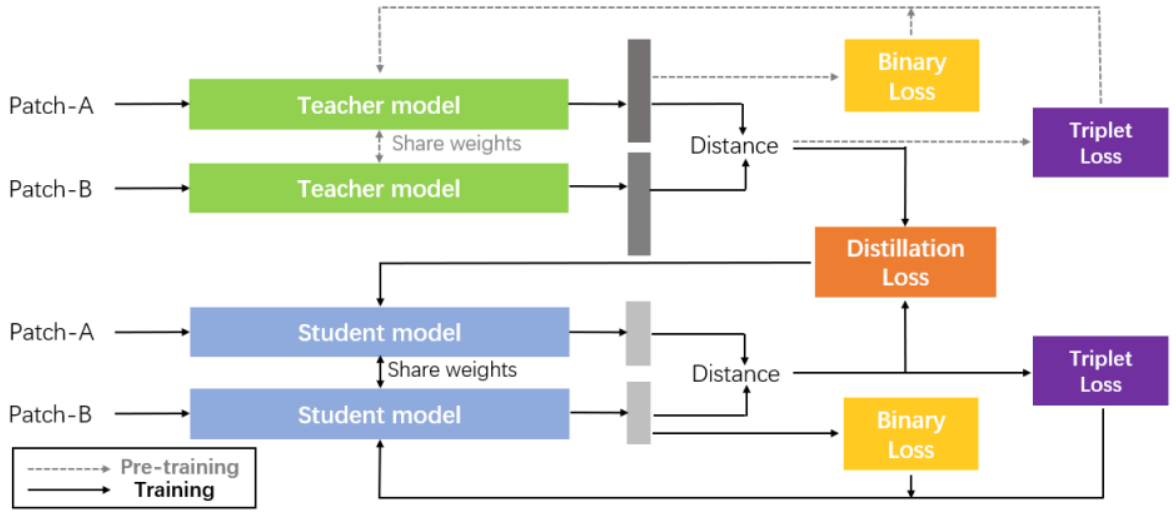
实现在黑暗/低光照环境下的SLAM。

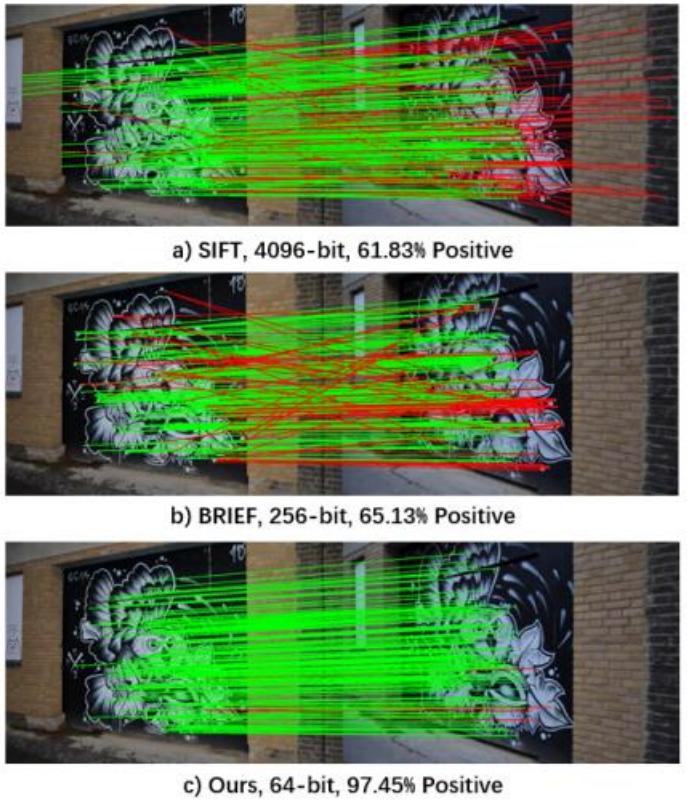


Hu et al. "A Two-stage Unsupervised Approach for Low light Image Enhancement." RAL 2021.

Efficient SLAM-基于知识蒸馏的描述子提取

- 从训练好的大模型（已有描述子抽取功能）进行迁移





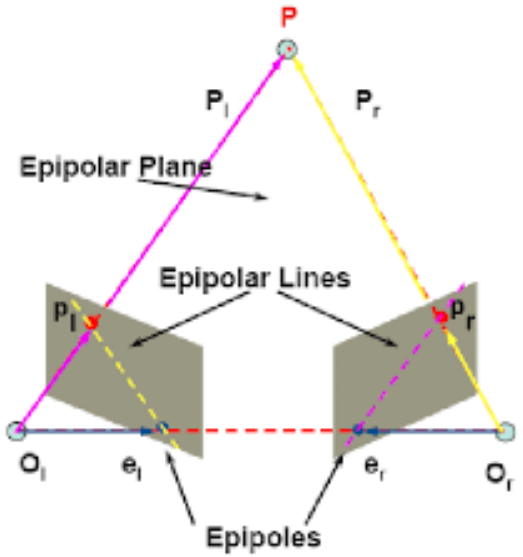
Guo, Hu et al. “Descriptor Distillation for Efficient Multi-Robot SLAM.” ICRA 2023.

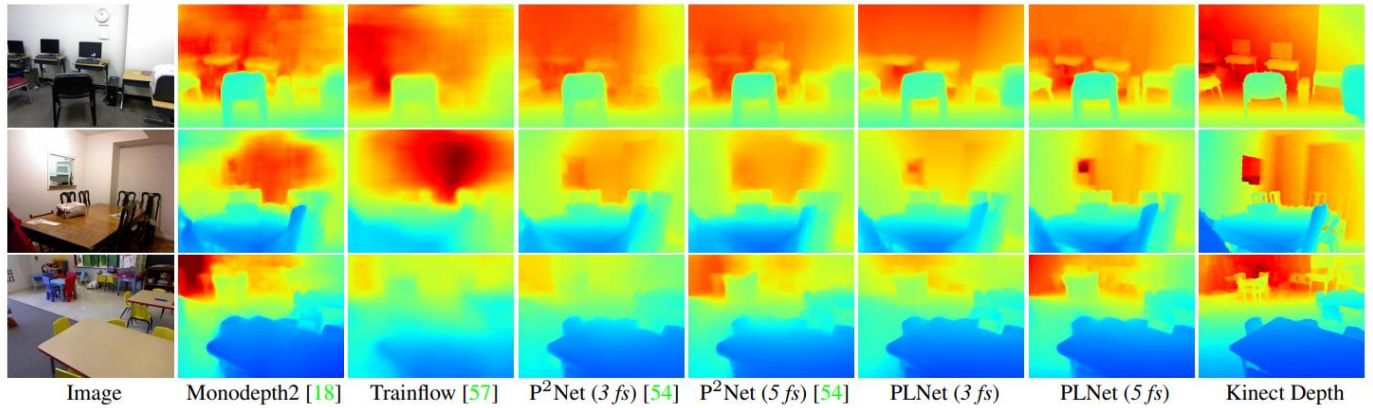
智能机器人感知：从数据角度的解决方案

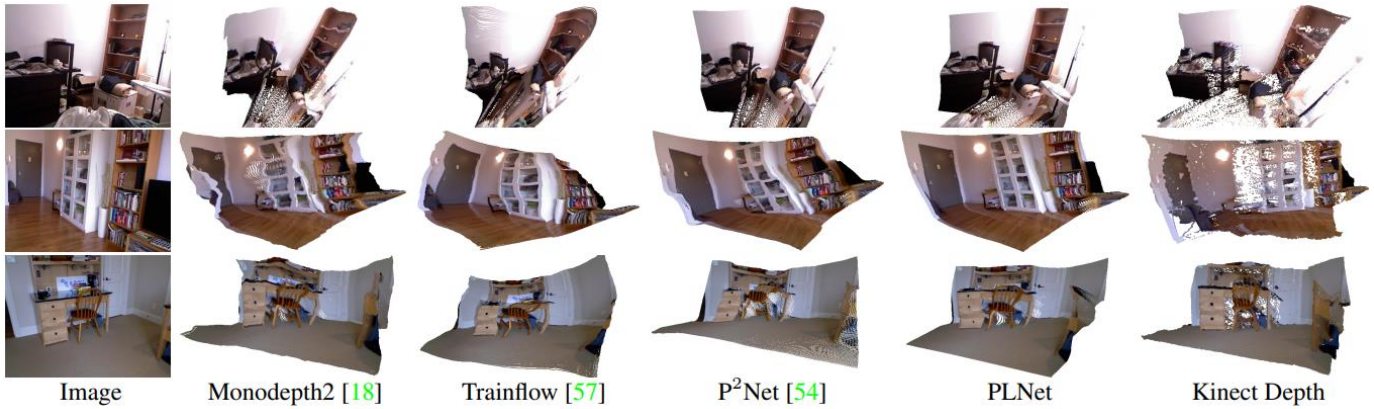
- 视觉SLAM
  - 主动感知SLAM-人类先验
  - 黑暗条件下SLAM-无监督学习
  - 实时SLAM-知识蒸馏
- **机器人环境深度估计**
  - 无监督学习
  - 知识蒸馏
  - 终身学习
- 无人机、多机器人环境语义识别
  - 半监督学习
  - 小样本学习
  - 多机器人语义SLAM

深度估计-无监督学习

- 根据多视角结合，计算相对深度，仅需图像视频仅需训练



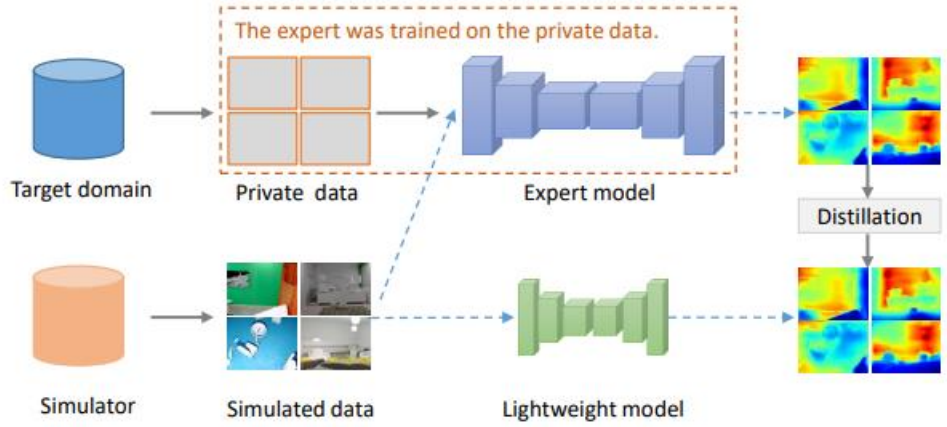




Jiang, Hu et al. “Descriptor Distillation for Efficient Multi-Robot SLAM.” 3DV, 2021.

深度估计-无数据知识蒸馏

- 解放对大规模训练数据的需求，无需原始真实世界训练数据，仅利用仿真数据进行模型迁移。

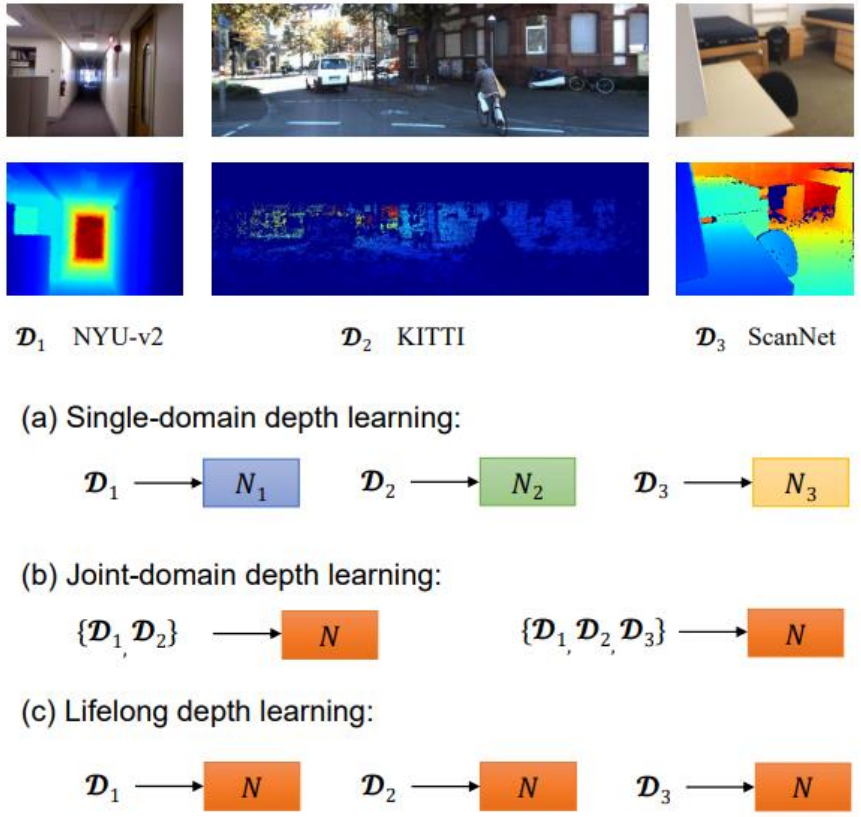


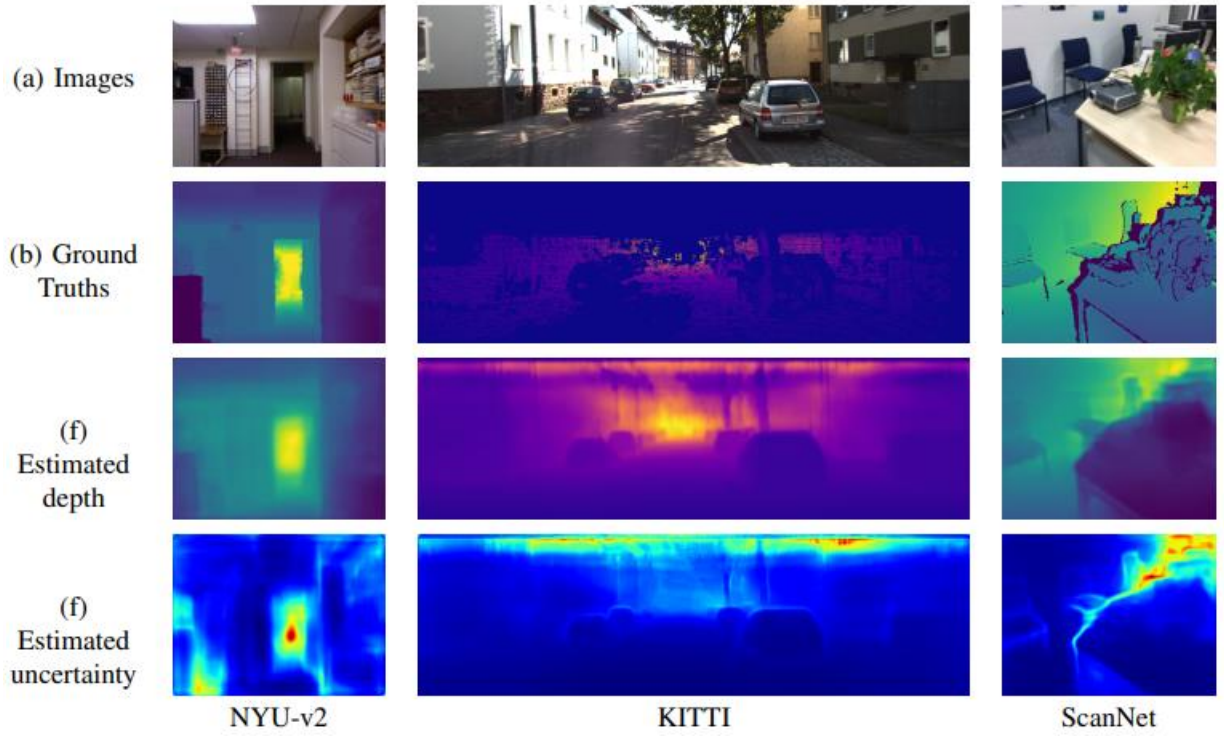
Teacher (Backbone) → Student (Backbone) Parameter Reduction		ResNet-34 [18] → ResNet-34 None		ResNet-34 [18] → MobileNet-v2 21.9 M → 1.7 M		ResNet-50 [25] → ResNet-18 63.6 M → 13.7 M	
Method	Data	REL ↓	δ_1 ↑	REL ↓	δ_1 ↑	REL ↓	δ_1 ↑
Teacher	NYU-v2	0.133	0.829	0.133	0.829	0.134	0.824
Student		0.133	0.829	0.145	0.802	0.145	0.805
Random noises	None	0.426	0.193	0.431	0.194	0.517	0.102
DFAD [12]		0.285	0.402	0.306	0.329	0.300	0.382
KD-OOD [16]	SceneNet $\mathcal{X}_1'$	0.164	0.753	0.175	0.712	0.188	0.660
Ours		**0.155**	**0.774**	**0.168**	**0.742**	**0.173**	**0.701**
KD-OOD [16]	SceneNet $\mathcal{X}_2'$	0.158	0.761	0.165	0.742	0.180	0.676
Ours		**0.151**	**0.789**	**0.157**	**0.778**	**0.165**	**0.726**

Hu et al. " Data-free Dense Depth Distillation." Arxiv 2022. submitted to TNNLS 2023.

深度估计-终身学习

- 实现了在单目深度估计任务上的终身学习算法，模型可持续的利用新数据进行更新，无需重新训练





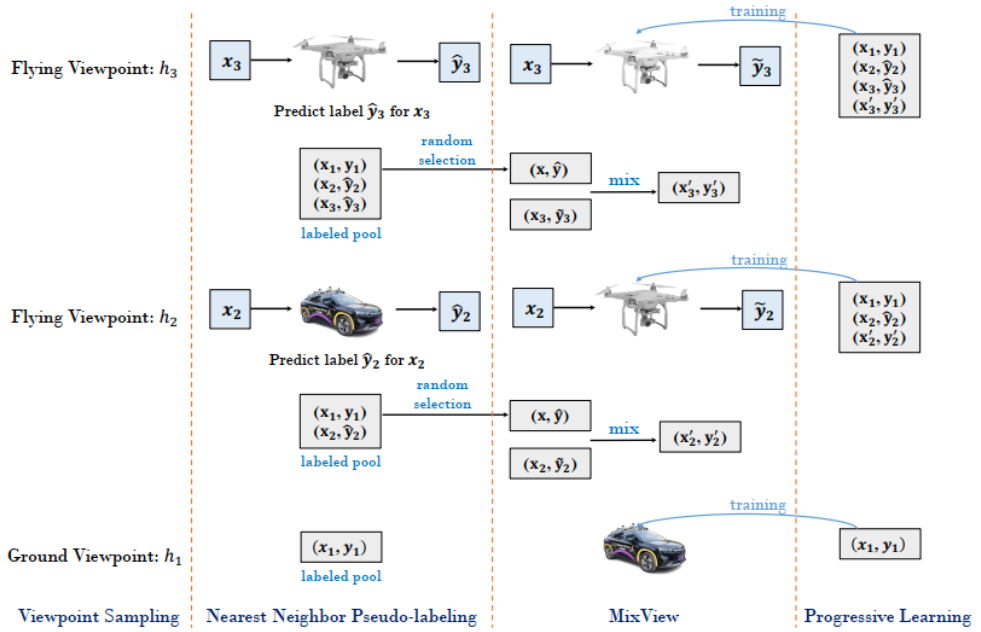
Hu et al. " Lifelong-MonoDepth: Lifelong Learning for Multi-Domain Monocular Metric Depth Estimation." Arxiv 2023. submitted to TNNLS 2023.

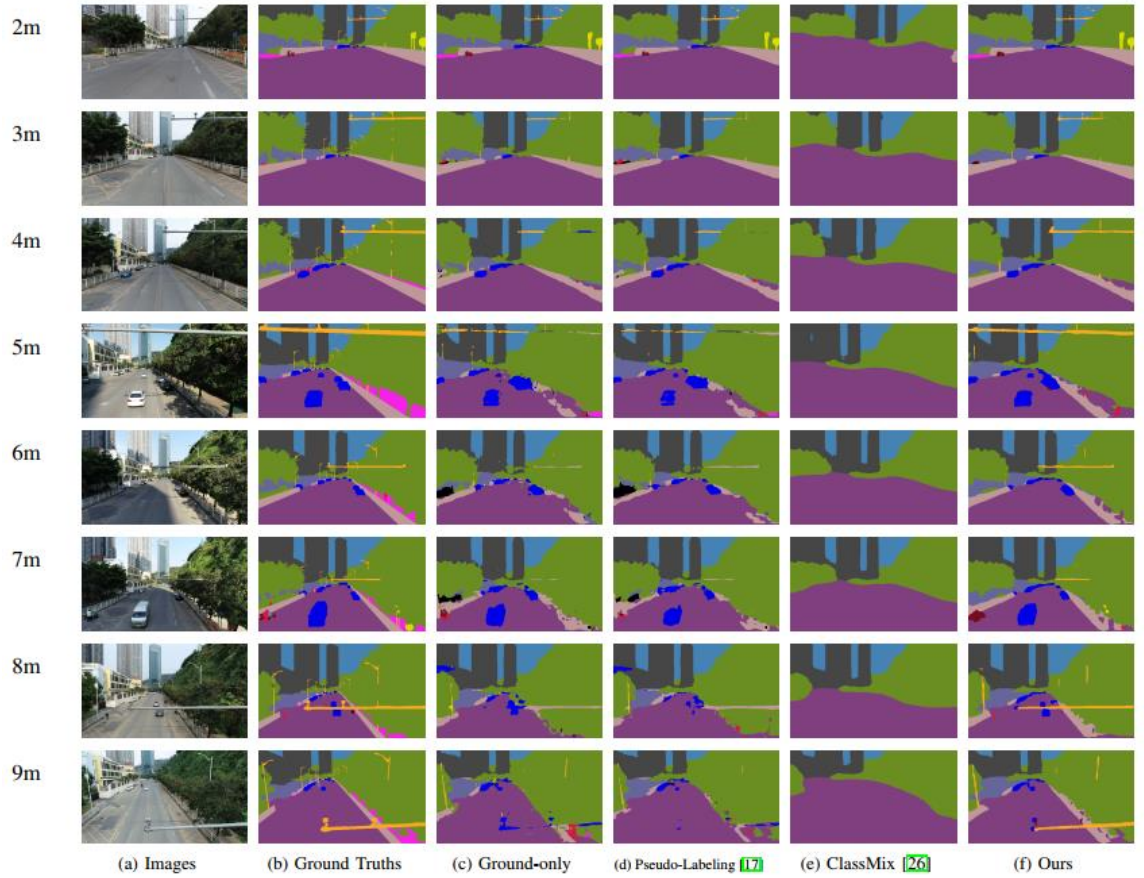
智能机器人感知：从数据角度的解决方案

- 视觉SLAM
  - 主动感知SLAM-人类先验
  - 黑暗条件下SLAM-无监督学习
  - 实时SLAM-知识蒸馏
- 机器人环境深度估计
  - 无监督学习
  - 知识蒸馏
  - 终身学习
- **无人机、多机器人环境语义识别**
  - 半监督学习
  - 小样本学习
  - 多机器人语义SLAM

无人机语义感知-半监督学习

- 提出仅利用地面标注数据和空中无标注数据的无人机感知实现算法



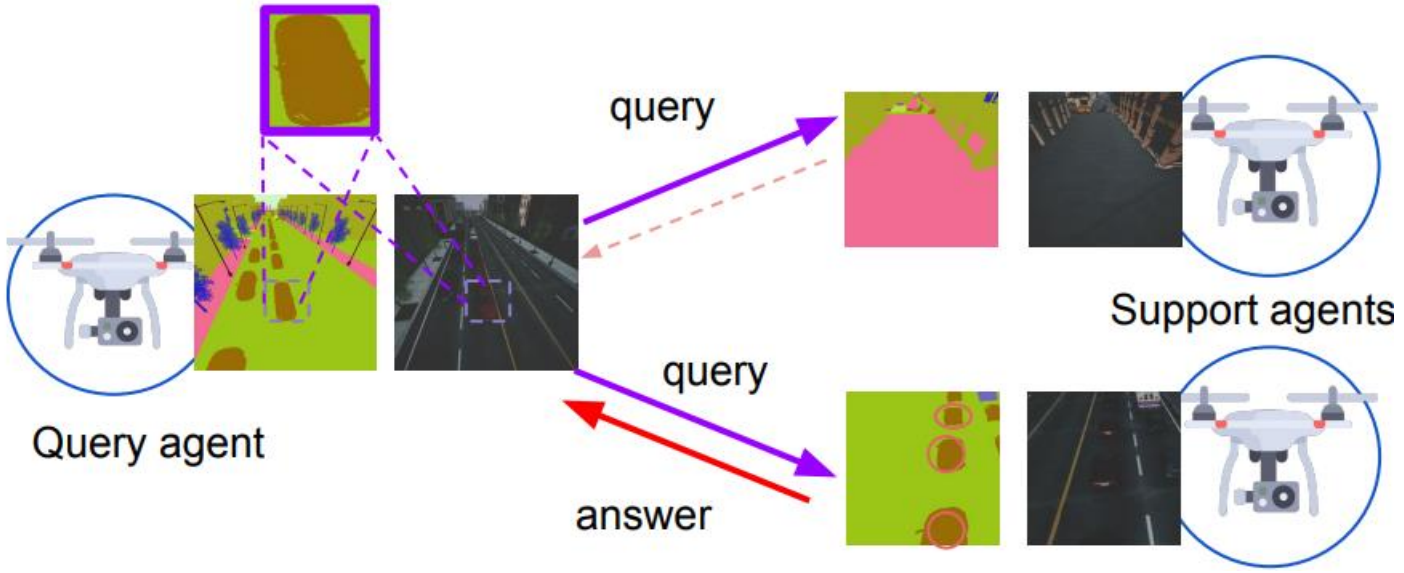


Hu et al. " Progressive Self-Distillation for Ground-to-Aerial Perception Knowledge Transfer." Arxiv 2022. submitted to TRO 2023.

多机器人协同感知-小样本学习

多机器人协同感知

- 查询机器人发生目标，支持机器人返回搜寻结果
- 小样本学习方式攻克真实世界数据缺乏挑战

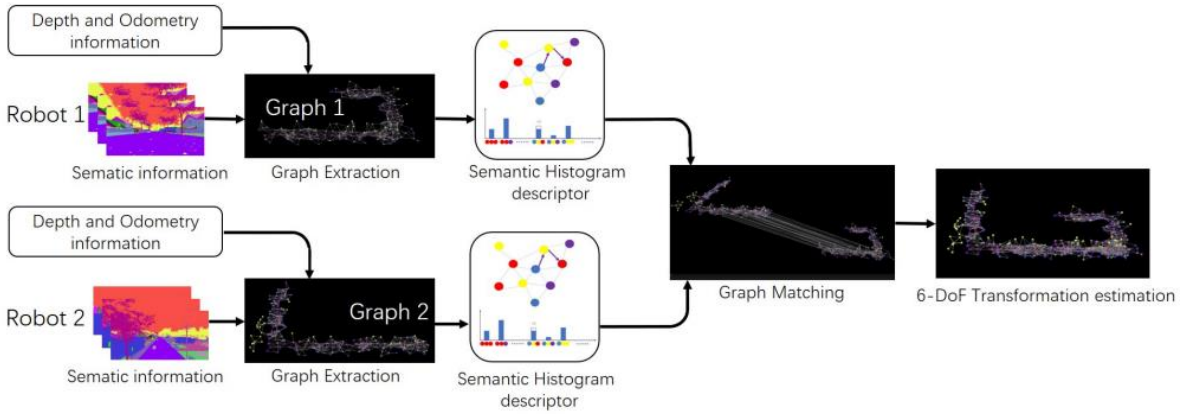


Fan, Hu et al. " Few-Shot Multi-Agent Perception.” ACM MM 2021.

多机器人语义SLAM-融合语义识别、深度估计、多机器人系统和SLAM

多机器人语义SLAM

- 提出基于语义的图匹配算法，实现多机器人地图配准与融合



(a) The trajectories of KITTI 08 dataset

(b) The successful multi-robots map fusion

TABLE III
THE TRANSLATION ERROR OF GLOBAL LOCALIZATION ON THE KITTI DATASET (IN METERS)

	Sequence 02	Sequence 08A	Sequence 08B	Sequence 19
Neighbor Vector	14.42±20.02	4.59±0.63	18.42±4.00	15.18±11.45
Random Walk	76.61±36.42	4.83±0.68	25.55±8.72	14.63±13.35
BoW	55.20 ± 42.01	74.12 ± 51.14	32.16 ± 20.79	108.83 ± 54.05
NetVLAD [17]	28.21 ± 19.35	35.02±21.04	24.52±14.41	55.11±20.96
Ours	**8.77±11.39**	**4.42±0.35**	**7.48±3.67**	**8.10±6.63**

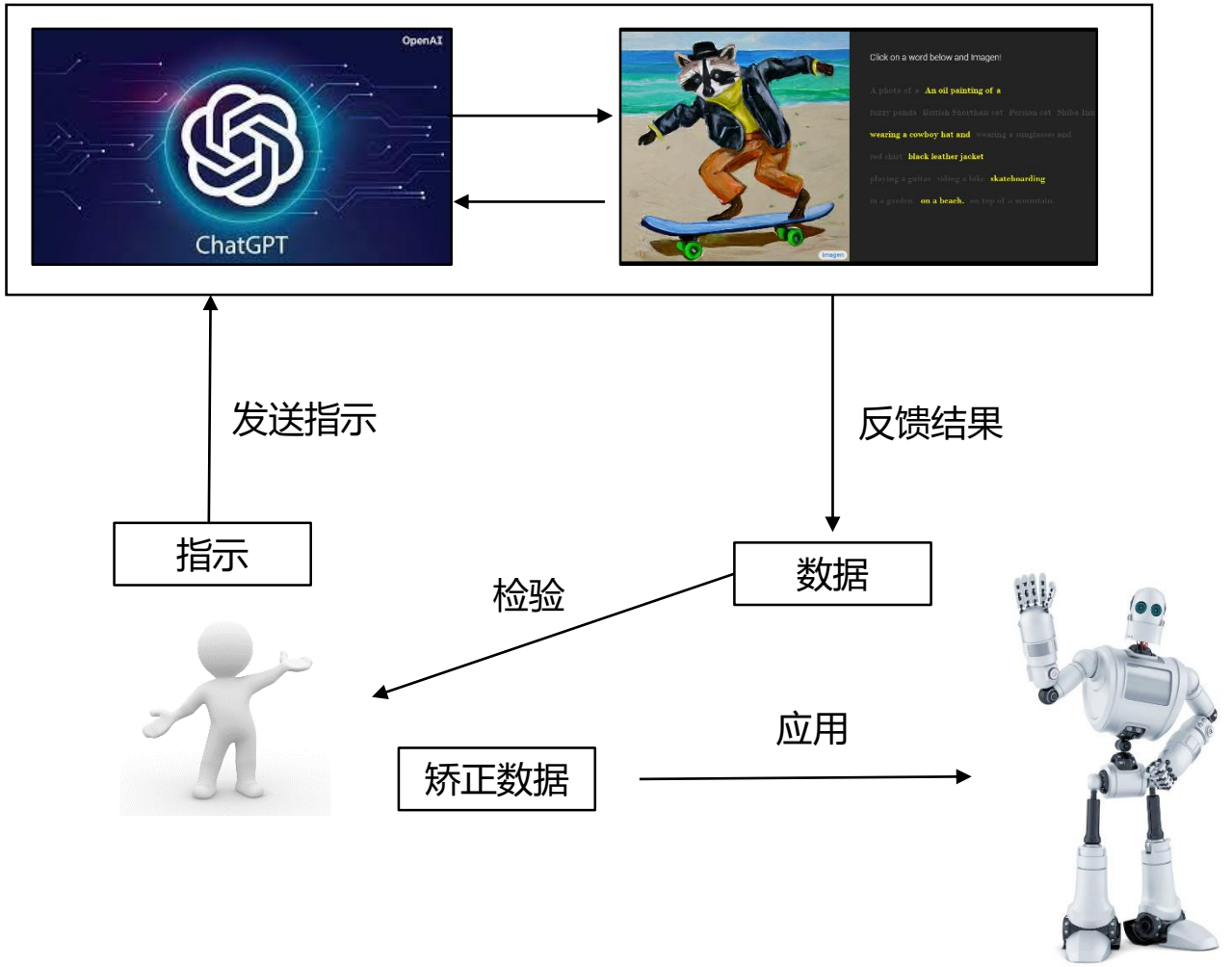
Guo, Hu et al. “Semantic Histogram Based Graph Matching for Real-Time Multi-Robot Global Localization in Large Scale Environment.” RAL 2021.

机器人感知未来趋势

- 数据闭环的解决方案往往由工业界主导，随着不断的进行数据采集、矫正、模型迭代，最终能够达到满意的效果。一般针对商业价值大的应用，如
  - 无人机：美团送餐
  - 扫地机器人
- 未来的较长时间内，基于无监督学习、半监督学习、小样本学习、强化学习等的技术方案，仍是机器人感知的重要实现方式。
- 学术界难以获取大量数据，且私有数据较为分散，难以共享。
  - 可能出现基于仿真的大数据、多任务数据集
  - 可通过知识蒸馏、迁移学习等技术方案，充分利用工业界的预训练人工智能大模型，实现在本地机器人感知任务上的小模型。

机器人感知未来趋势

- 利用大模型如ChatGPT-4收集训练数据并本地化训练。
  - 文本、视觉、3D大模型互通
  - 精心设计交互prompt，利用大模型生成目标域的图片、语言等并给出标注
  - 利用收集到的数据进行本地化训练，实现机器人部署



机器人感知未来趋势

- 基于视觉、文本提示的机器人导航将进一步发展