# 具身智能-11

刘华平

2025年5月7日

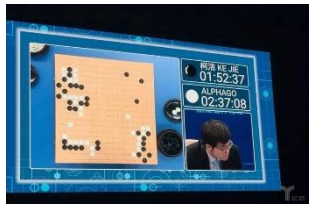
# 课程内容安排

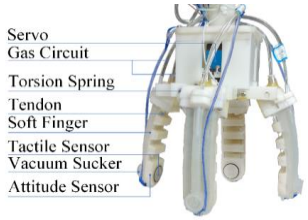
| 课次 | 周次 | 上课内容 | 软件 |
| --- | --- | --- | --- |
| 1 | 1 | 绪论 | |
| 2 | 2 | 深度学习 | |
| 3 | 3 | 强化学习1 | Gym, Mujuco |
| 4 | 4 | 强化学习2 | Gym, Mujuco |
| | 5 | 作业准备 | |
| 5 | 6 | 自监督与持续学习 | |
| | 7 | 开题 | Powerpoint |
| 6 | 8 | 形态智能 | Gym, Mujuco |
| 7 | 9 | 视觉导航：VLN | AI2THOR |
| 8 | 10 | 主动感知：VSN，EQA | AI2THOR |
| | 11 | 五一放假 | |
| 9 | 12 | 具身学习 | AI2THOR |
| 10 | 13 | 多体智能 | AI2THOR |
| 11 | 14 | 面向具身智能的AIGC | AI2THOR |
| | 15-16 | 成果准备与展示 | Powerpoint |

# 具身智能的体系

- 狭义与广义的具身智能

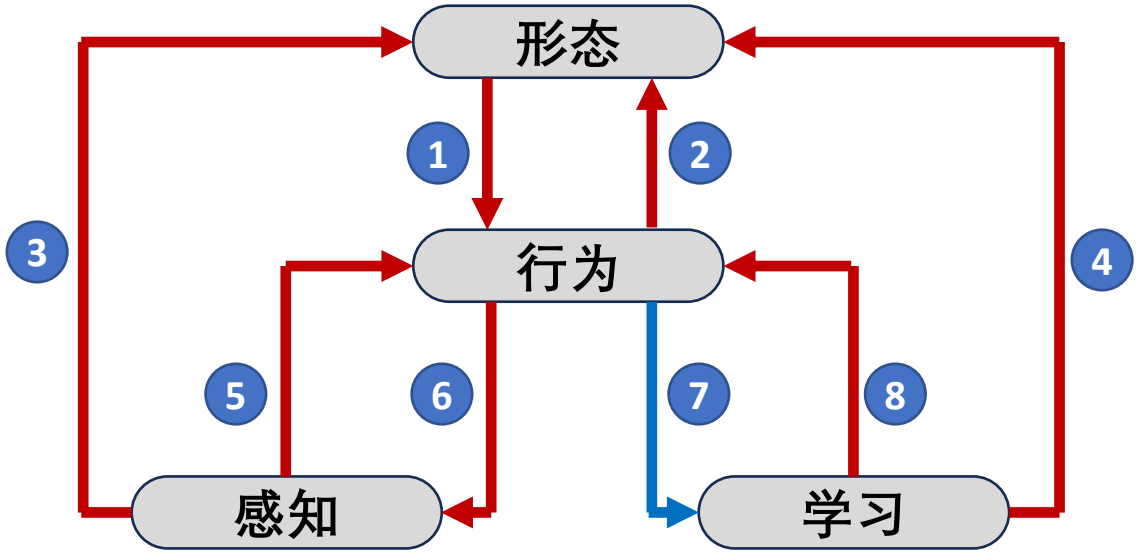

离 身 智 能 具 身 智 能



# 具身智能的体系

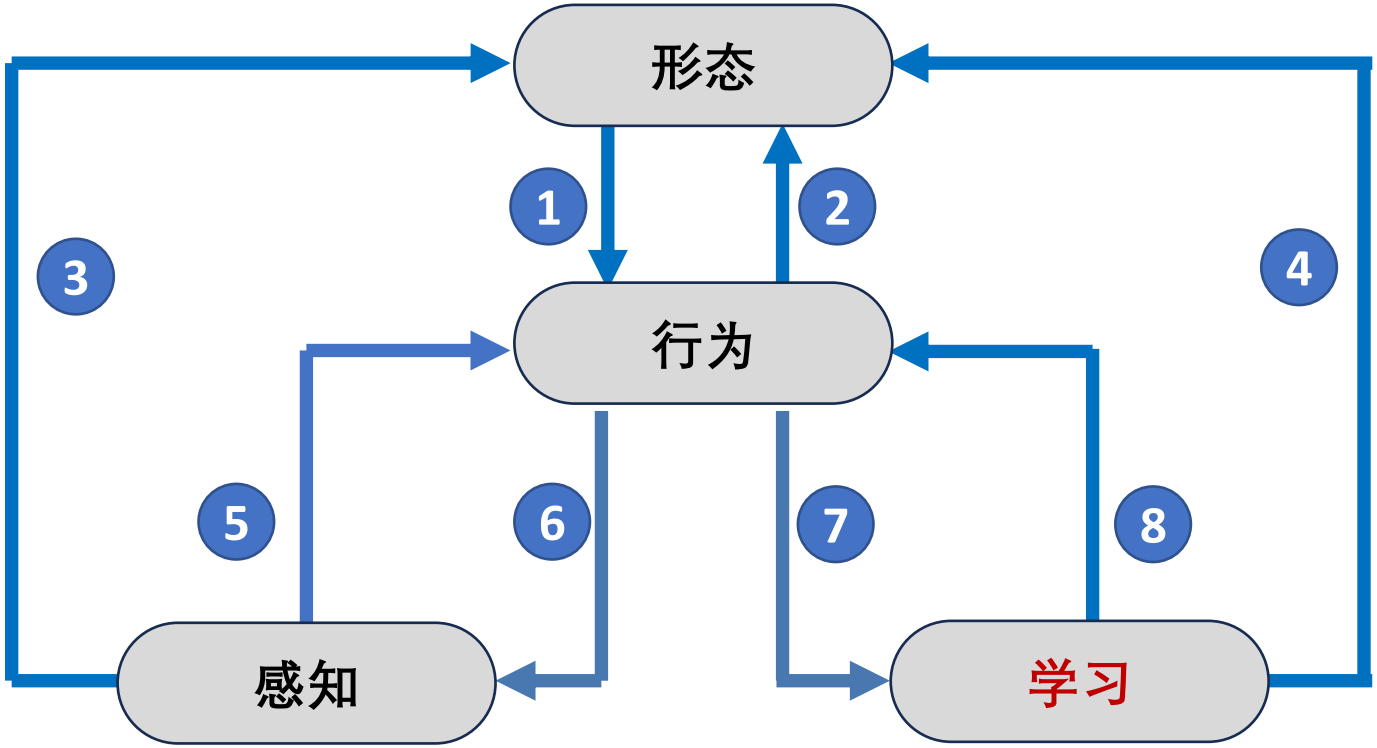
## ➢ 具身智能的体系结构





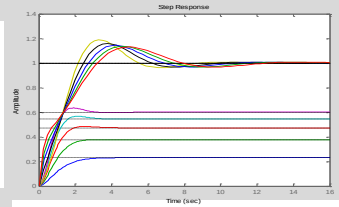
① 基于形态的行为生成　③ 基于感知的形态变换　⑤ 基于感知的行为生成　⑦ 基于行为的自主学习

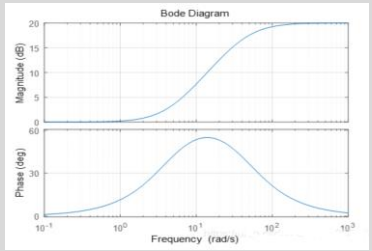
② 基于行为的形态控制　④ 基于学习的形态优化　⑥ 基于行为的主动感知　⑧ 基于学习的行为优化

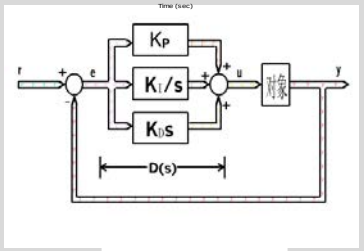

形态
1
2
3
4
行为
5
6
7
8
感知
学习

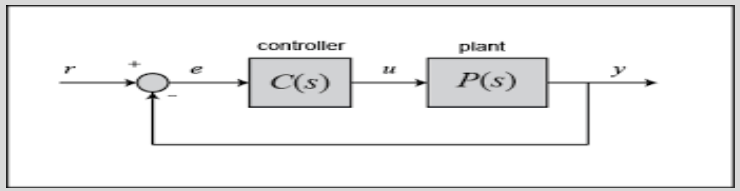
# 经典控制：经验

$$G_c(s) = \frac{s + \frac{1}{T}}{s + \frac{1}{\alpha T}} \quad (\alpha < 1)$$

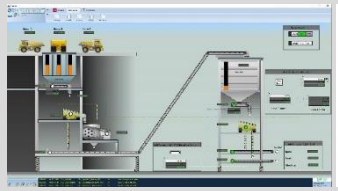


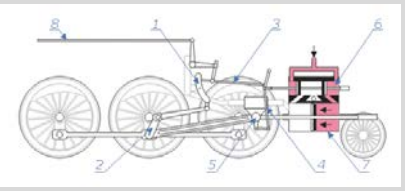




$$u_t = K_P e_t + K_I \int e_t dt + K_D \frac{de_t}{dt}$$



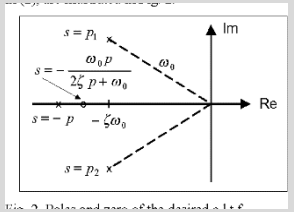


# 现代控制：模型

$$\text{state equations} \begin{cases} \dot{x}_1 = f_1(x_1, x_2, \ldots, x_n, u_1, \ldots, u_m) \\ \vdots \\ \dot{x}_n = f_n(x_1, x_2, \ldots, x_n, u_1, \ldots, u_m) \end{cases}$$

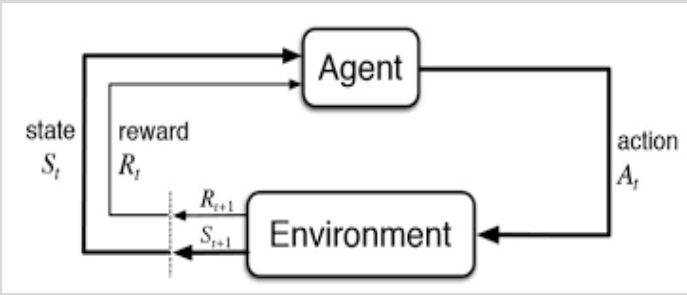
$$\text{output equations} \begin{cases} y_1 = h_1(x_1, x_2, \ldots, x_n, u_1, \ldots, u_m) \\ \vdots \\ y_p = h_p(x_1, x_2, \ldots, x_n, u_1, \ldots, u_m) \end{cases}$$

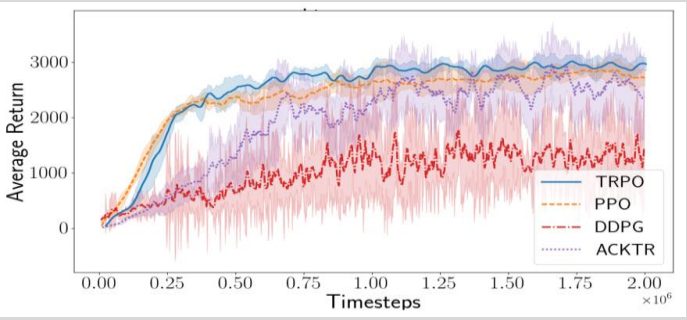
$$\begin{cases} \dot{x}_t = Ax_t + Bu_t \\ y_t = Cx_t + Du_t \end{cases}$$



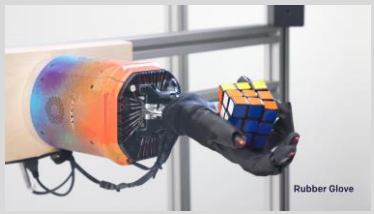

$$u_t = -Kx_t$$

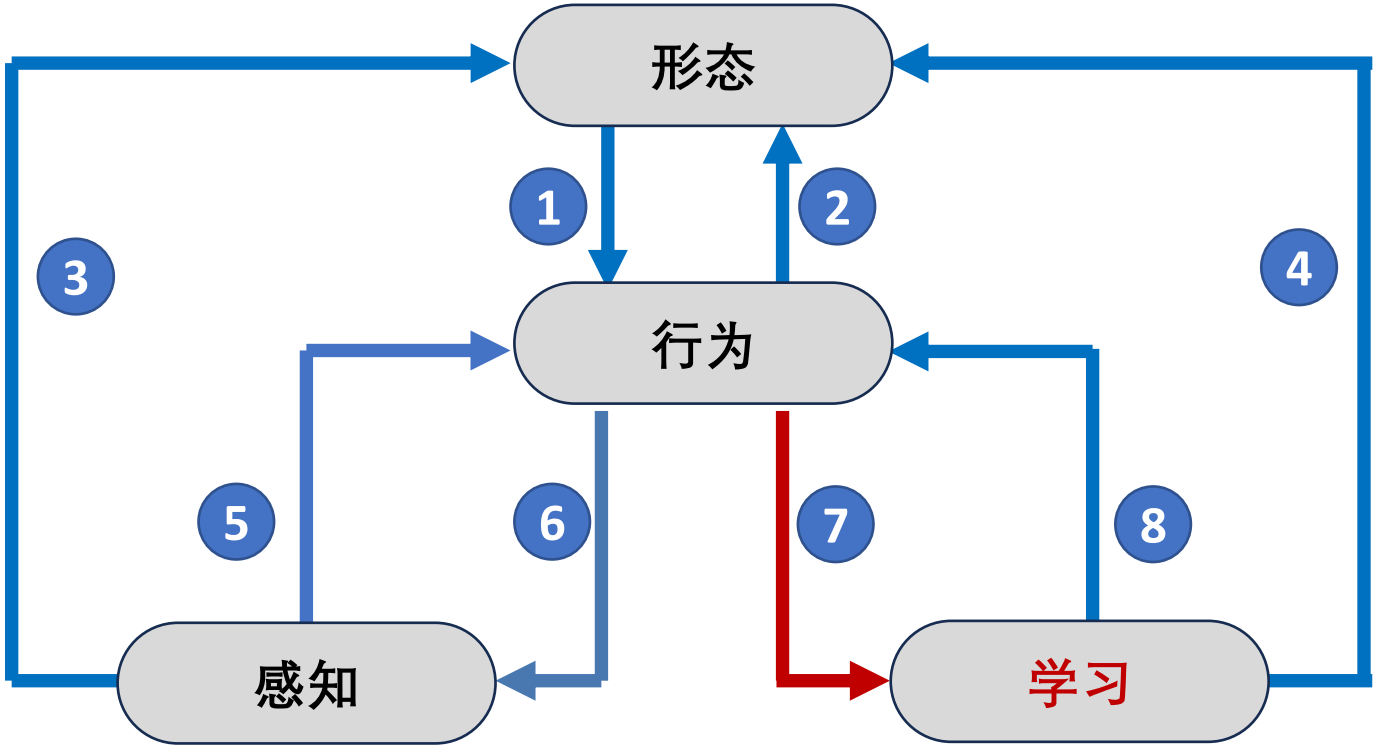
# 智能控制：学习




$$\pi_\theta(a_t \mid s_t)$$





形态
1
2
3
4
行为
5
6
7
8
感知
学习

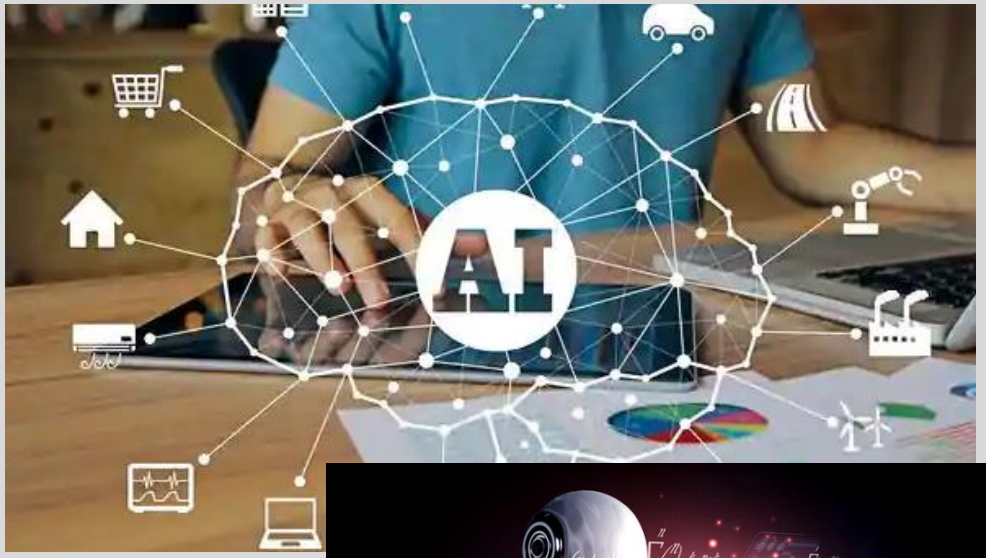
- 背景
- 具身学习
- 前沿

# 背景



## ➢ 主动学习：Why

### 采集数据




### 标注数据

# 背景



## ➢ 主动学习

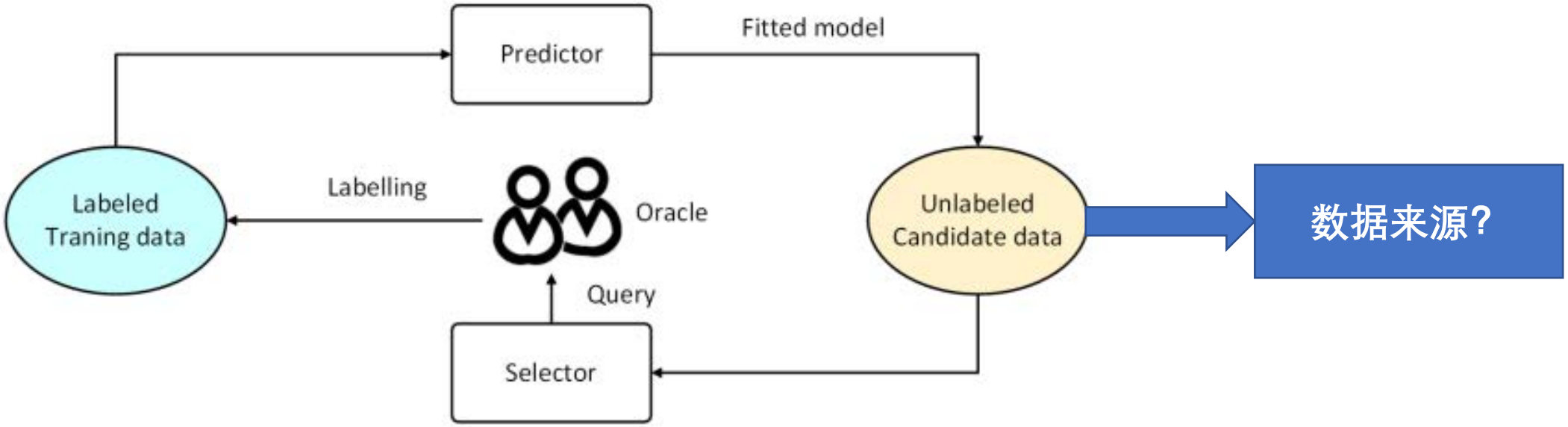



Fig. 1. Active learning trains the predictor with initial training samples, uses its selector to select a few of unlabeled samples, labels them, adds them into the training data set, and then re-trains the predictor.

Liu P, Wang L, Ranjan R, et al. A Survey on Active Deep Learning: From Model Driven to Data Driven[J]. ACM Computing Surveys (CSUR), 2022, 54(10s): 1-34.

# 背景

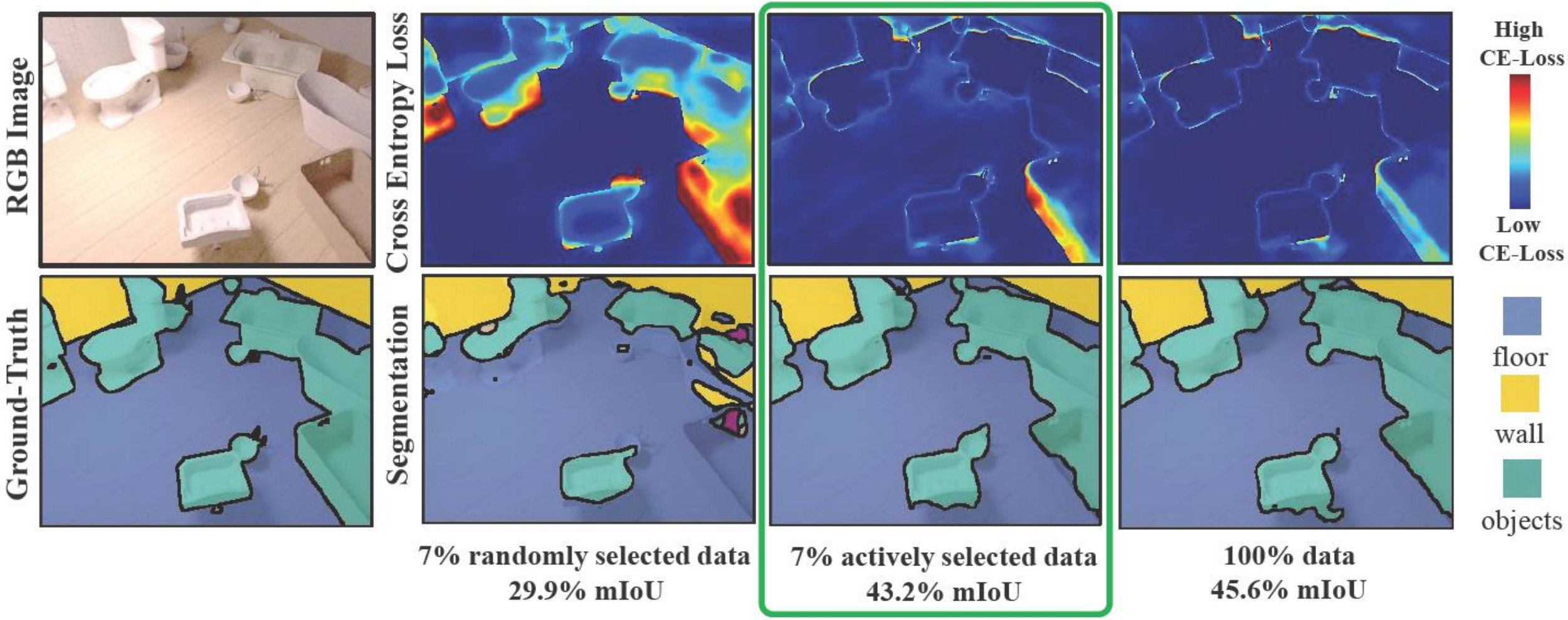


- 主动学习：标注

# 背景



- 主动学习：标注

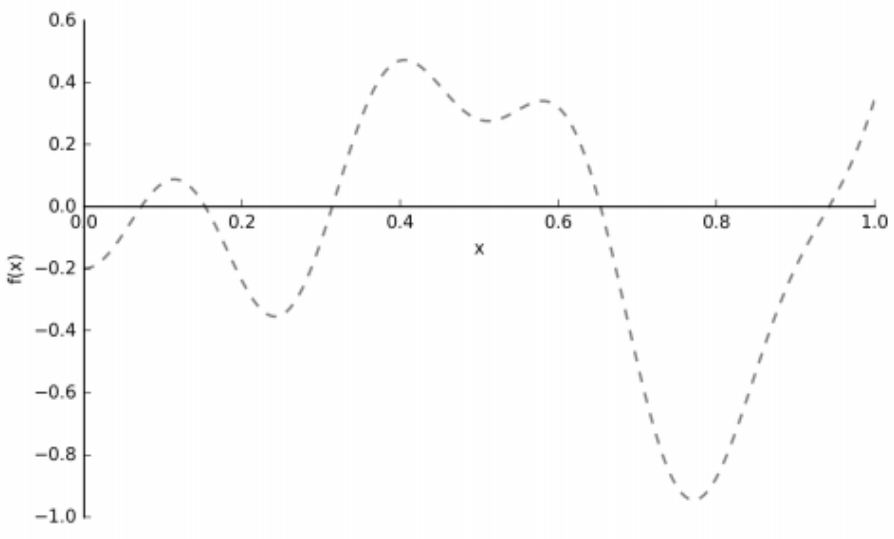

RGB Image
Cross Entropy Loss
Ground-Truth
Segmentation
High CE-Loss
Low CE-Loss
floor
wall
objects
7% randomly selected data
29.9% mIoU
7% actively selected data
43.2% mIoU
100% data
45.6% mIoU

# 背景



## ➢ 主动学习：标注

- Consider finding the optima $x_*$ (say minima) of a function $f(x)$




- Caveat: We don't know the form of the function; can't get its gradient, Hessian, etc

- Can only query the function's values at certain points (i.e., only "black-box" access)
  - The values may or may not be noisy (i.e., we may be given $f(x)$ or $f(x) + \epsilon$)

# 背景



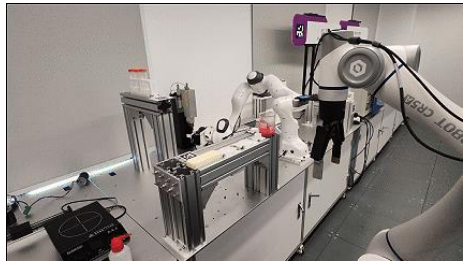
## ➢ 主动学习：标注

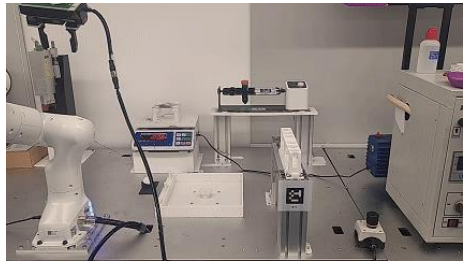
- Drug Design: Want to find the optimal chemical composition for a drug
  - Optimal composition will be the one that has the best efficacy
  - But we don't know the efficacy function
  - Can only know the efficacy via doing clinical trials
  - Each trial is expensive; can't do too many trials
- Hyperparameter Optimization: Want to find the optimal hyperparameters for a model
  - Optimal hyperparam values will be those that give the lowest test error
  - Don't know the true "test error" function
  - Need to train the model each time with different h.p. values and compute test error
  - Training every time will be expensive (e.g., for deep nets)
  - Note: Hyperparams here can even refer to the structure of a deep net (depth, width, etc)
- Many other applications: Website design via A/B testing, material design, optimizing physics based models (e.g., aircraft design), etc

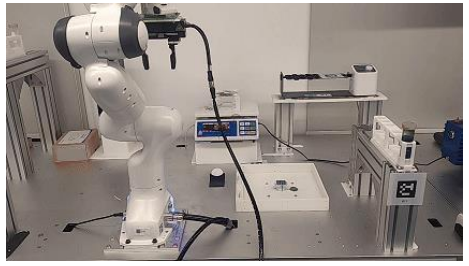
# 背景



## ➢ 主动学习：标注

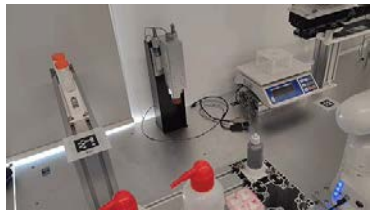






获取原料及模具

递送材料到操作区



获取试管

去除试管盖

放置试管到电子秤

添加硅胶

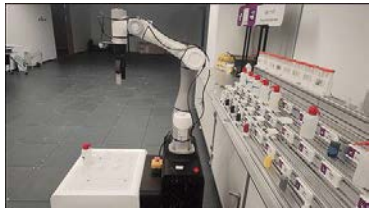
添加磁粉

添加试管盖

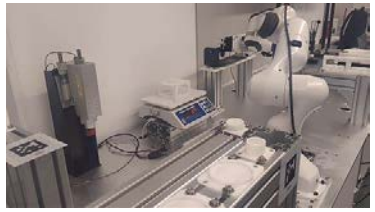
放置试管，开始搅拌

放置模具到电子秤

从工作区取回材料



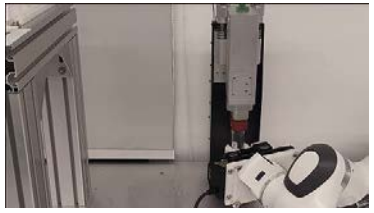
放回材料到材料区



从搅拌机取下试管

去除试管盖

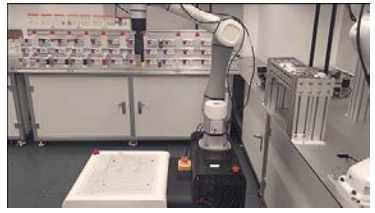
浇筑模型



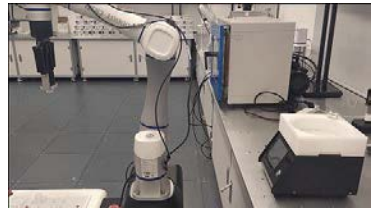
添加试管盖

放置废弃试管到废料区

取下浇筑好的模具



移动机器人获取模具



将模具放置到匀胶机

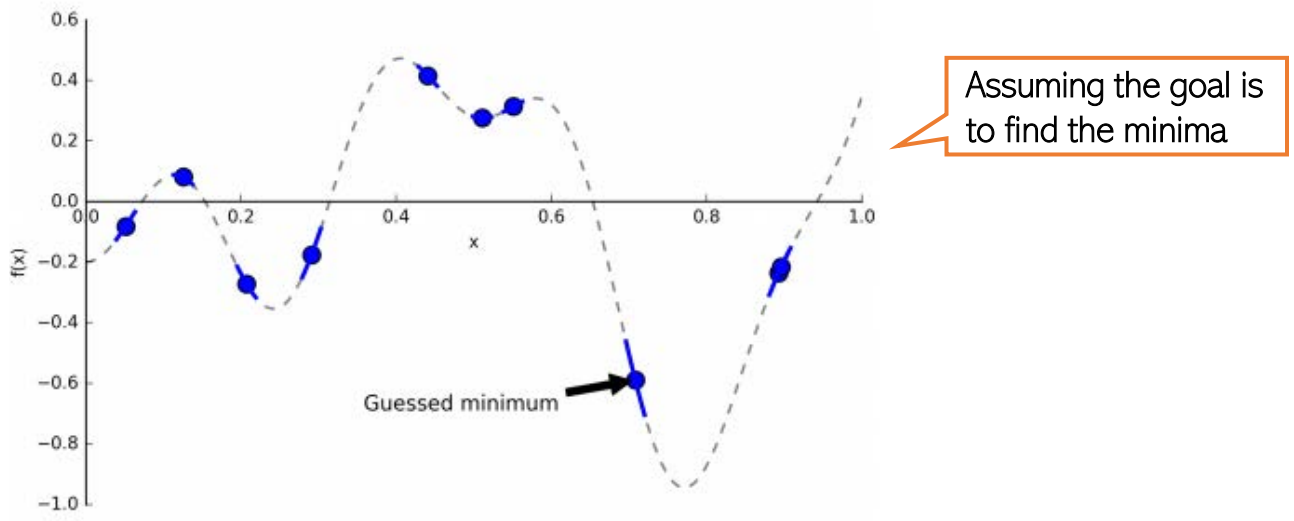
# 背景



## ➢ 主动学习：标注

- Can use BO to find maxima or minima

- Would like to locate the optima by querying the function's values (say, from an oracle)




- We would like to do so using as few queries as possible

- Reason: The function's evaluation may be time-consuming or costly

# 背景



## ➢ 主动学习：标注

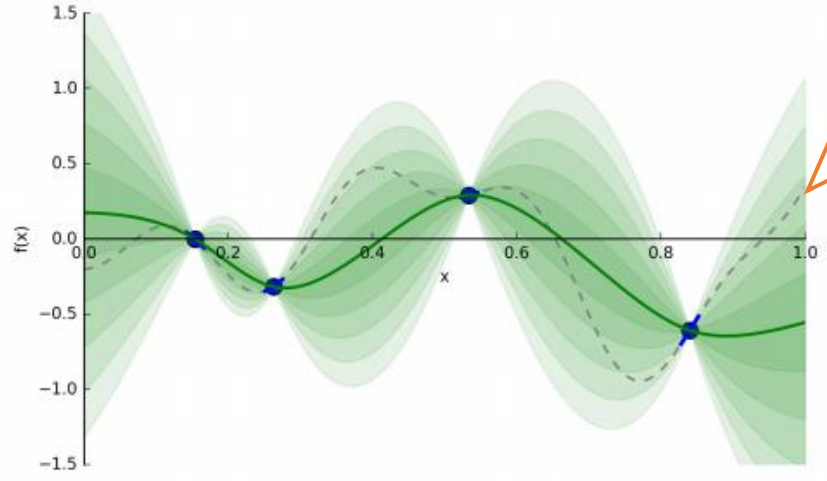
- Suppose we are allowed to make the queries sequentially
- This information will be available to us in form of query-function value pairs

Note: Function values can be noisy too, e.g., $f(x_n) + \epsilon_n$

$\{(x_n, f(x_n))\}_{n=1}^{N}$

- Queries so far can help us estimate the function

By solving a regression problem



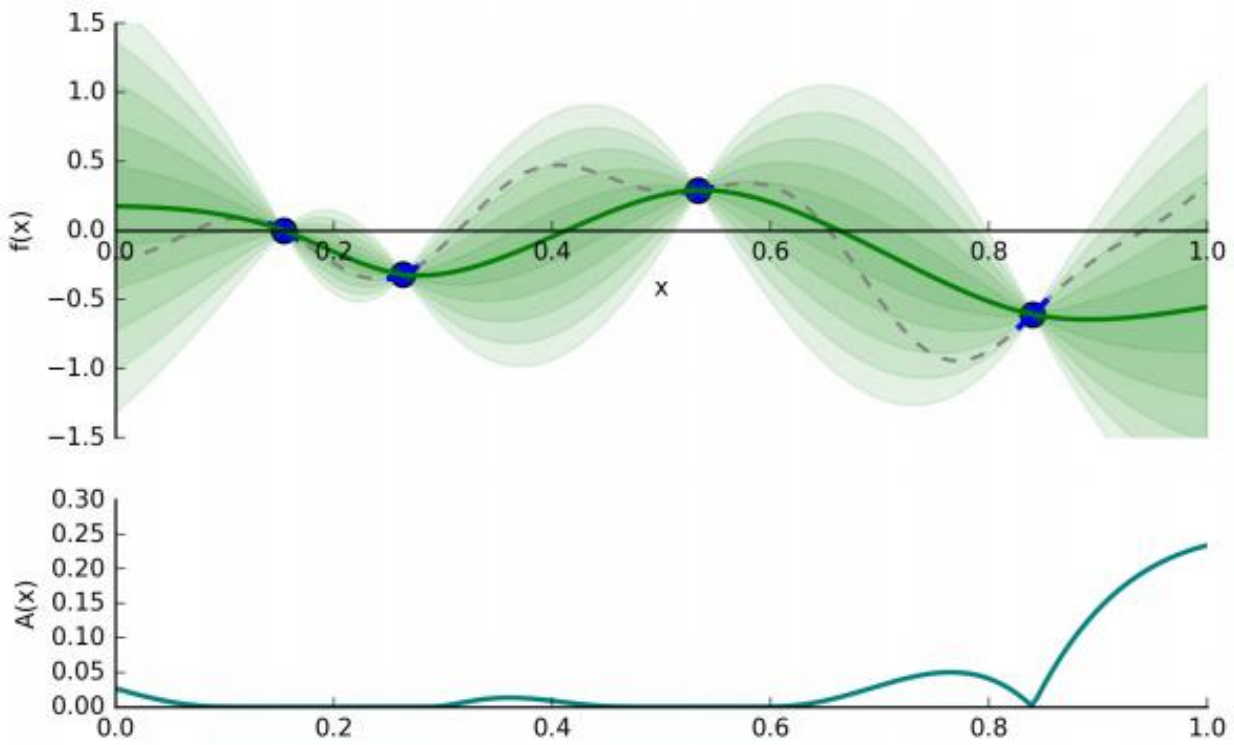

Dotted curve: True function
Green curve: Current estimate ("surrogate") of the function
Shaded region: Uncertainty in the function's estimate

- BO uses past queries + function's estimate+uncertainty to decide where to query next
- Similar to Active Learning but the goal is to learn $f$ as well as finds its optima

# 背景



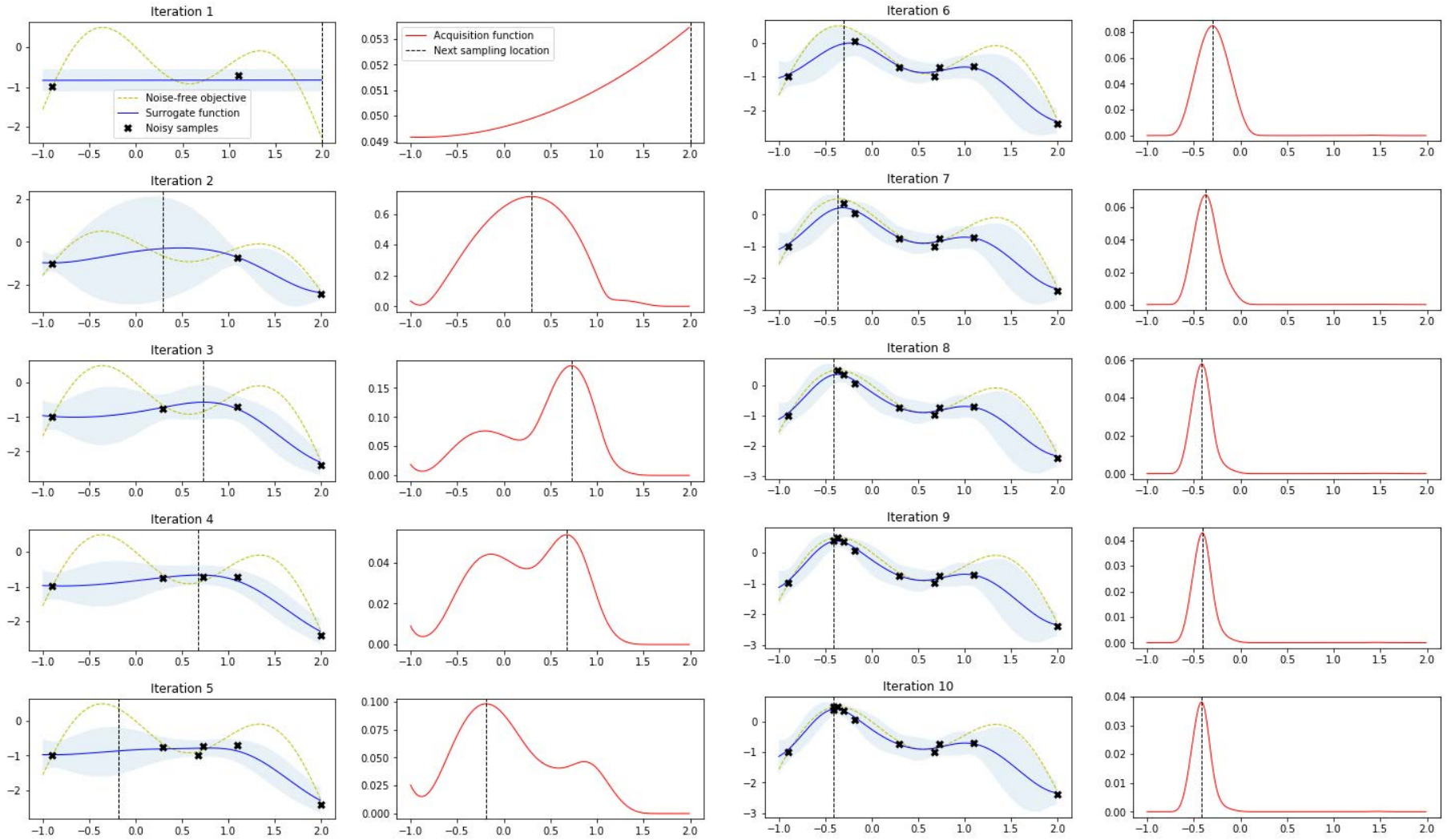
- 主动学习：标注


1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
f(x)
0.0
0.2
0.4
0.6
0.8
1.0
x
0.30
0.25
0.20
0.15
0.10
0.05
0.00
A(x)
0.0
0.2
0.4
0.6
0.8
1.0

# 背景



## ➢ 主动学习：标注


Iteration 1
Noise-free objective
Surrogate function
Noisy samples
Acquisition function
Next sampling location
Iteration 2
Iteration 3
Iteration 4
Iteration 5
Iteration 6
Iteration 7
Iteration 8
Iteration 9
Iteration 10

# 背景



## ➢ 具身学习

- 危险/复杂场景下的数据采集
- 协同提高数据采集效率


类别边界的模糊性


类别定义的模糊性

越野场景

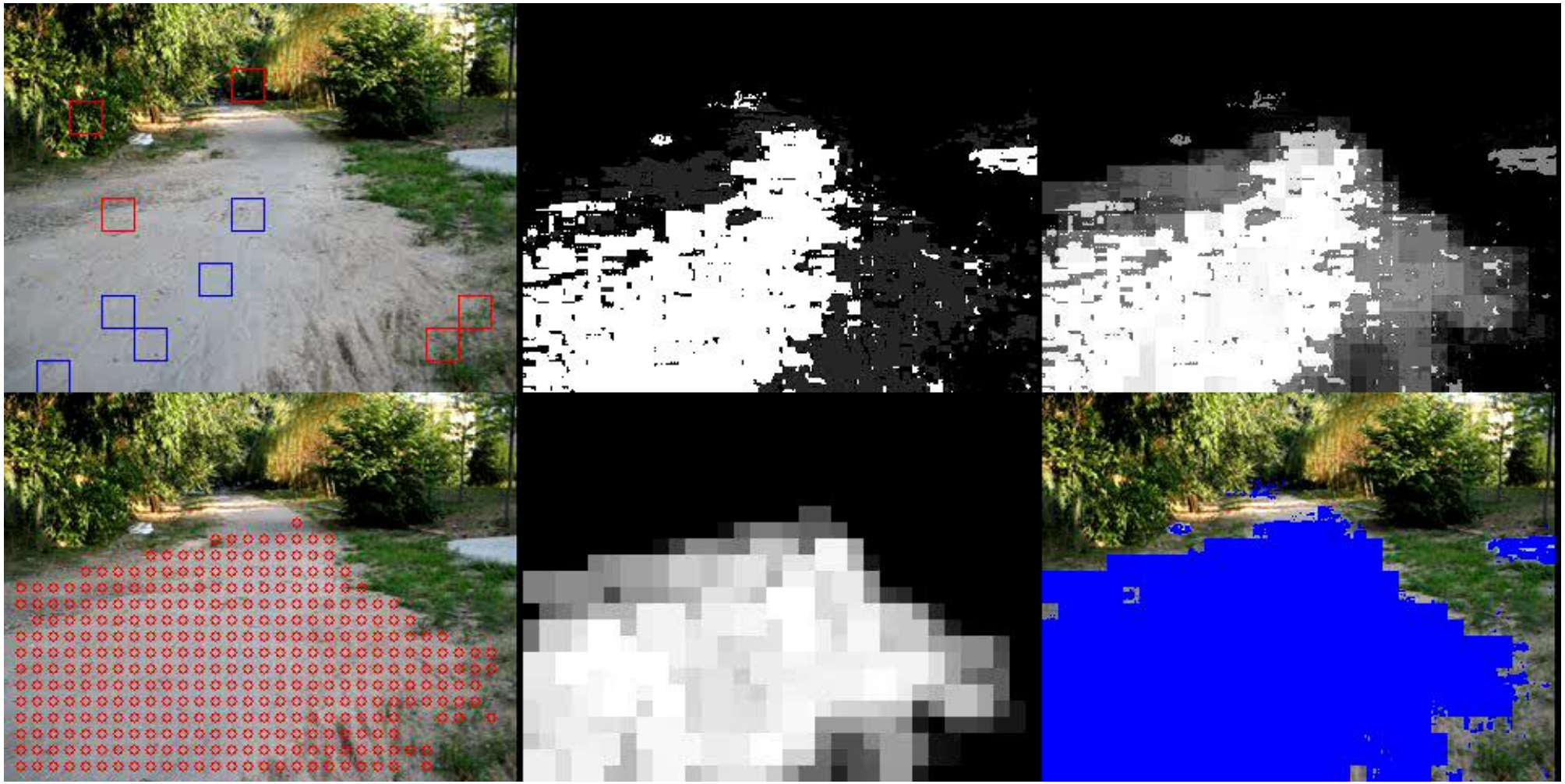
大型履带车

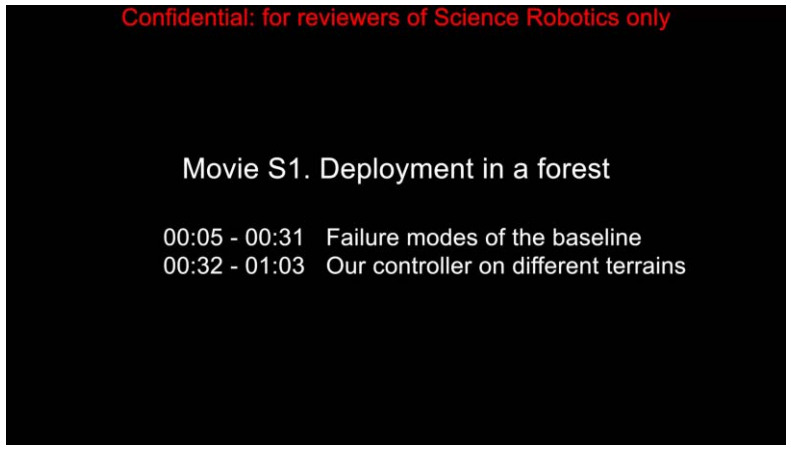
小型无人车

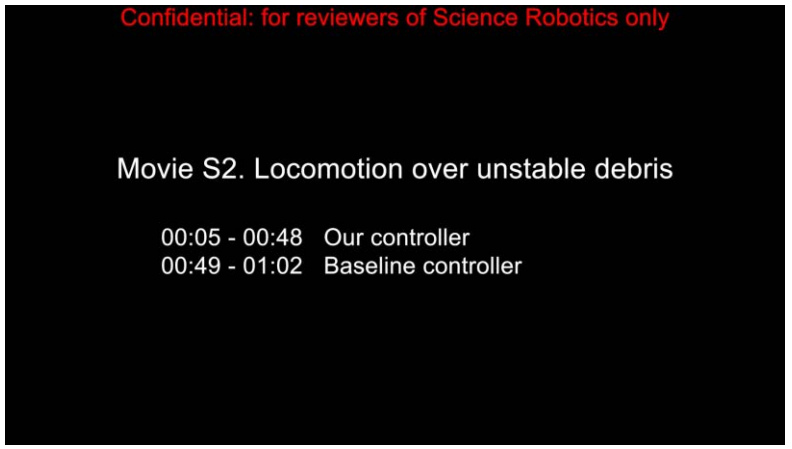
需求1：粗粒度语义分割

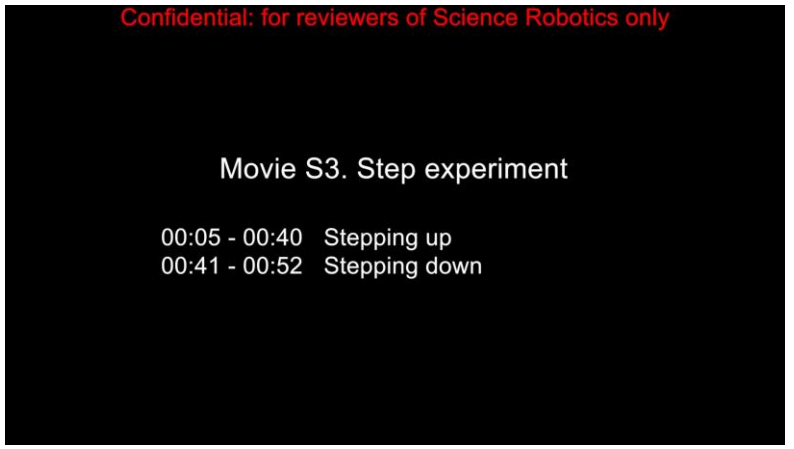

需求2：细粒度语义分割

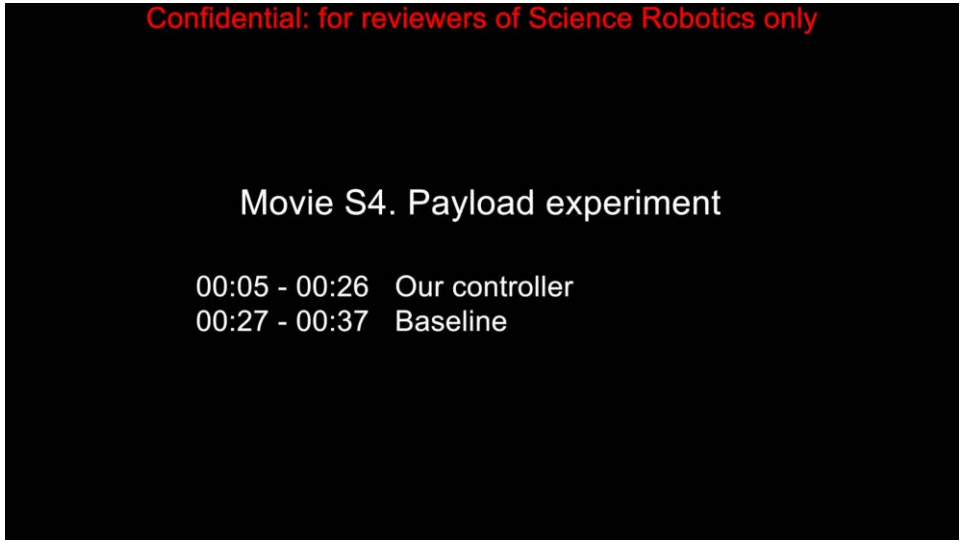

免费送
CAUTION








- Lee J, Hwangbo J, Wellhausen L, et al. Learning quadrupedal locomotion over challenging terrain[J]. Science robotics, 2020, 5(47): eabc5986.

Movie S1. Deployment in dry and wet beach sand.

Movie S2. Experiment on grass.

Movie S3. Experiment on athletic track.

Movie S4. Experiment on vinyl tile.

Movie S5. Experiment on air mattress.

Movie S6. Sudden terrain transition.

Choi S, Ji G, Park J, et al. Learning quadrupedal locomotion on deformable terrain[J]. Science Robotics, 2023, 8(74): eade2256.

# ➢ 地形适应的形态变换

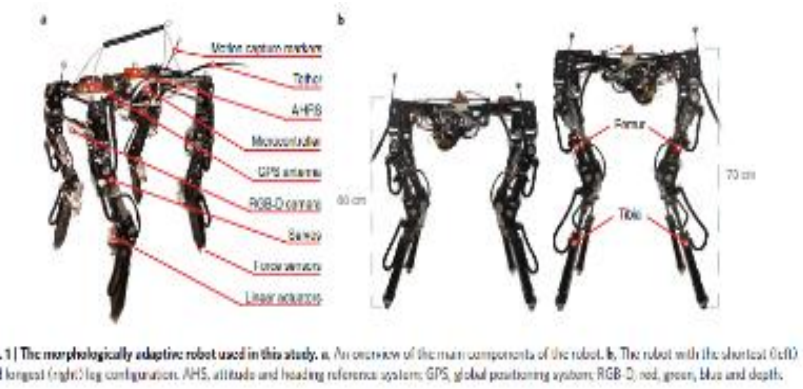
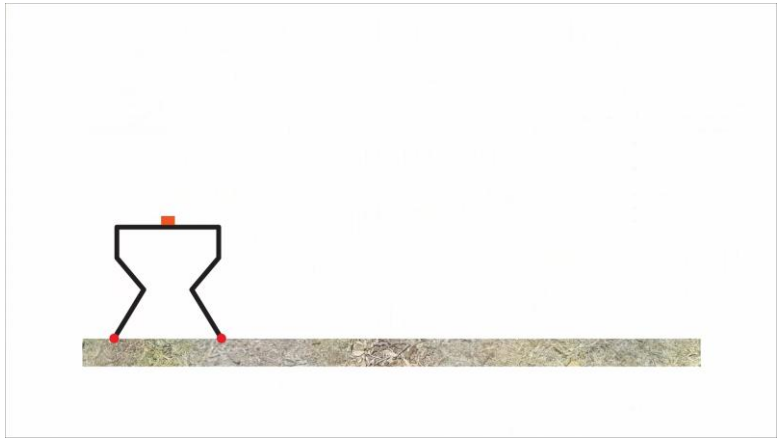
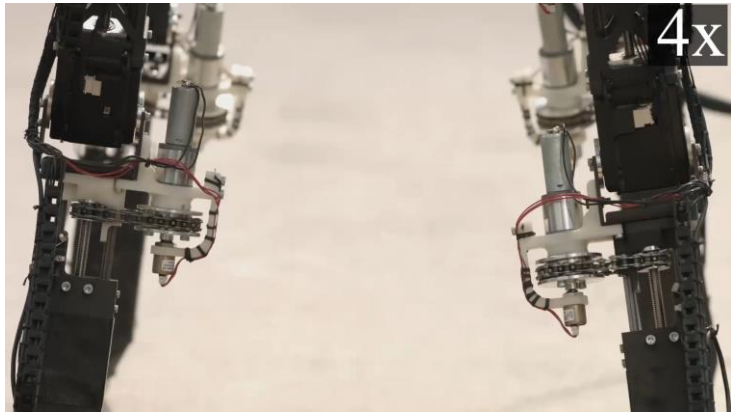
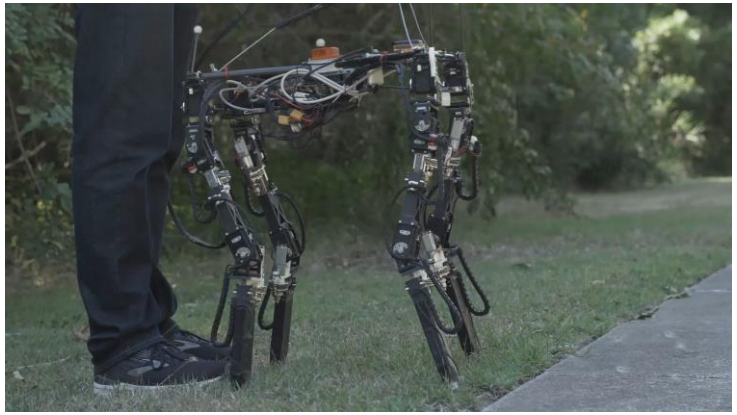


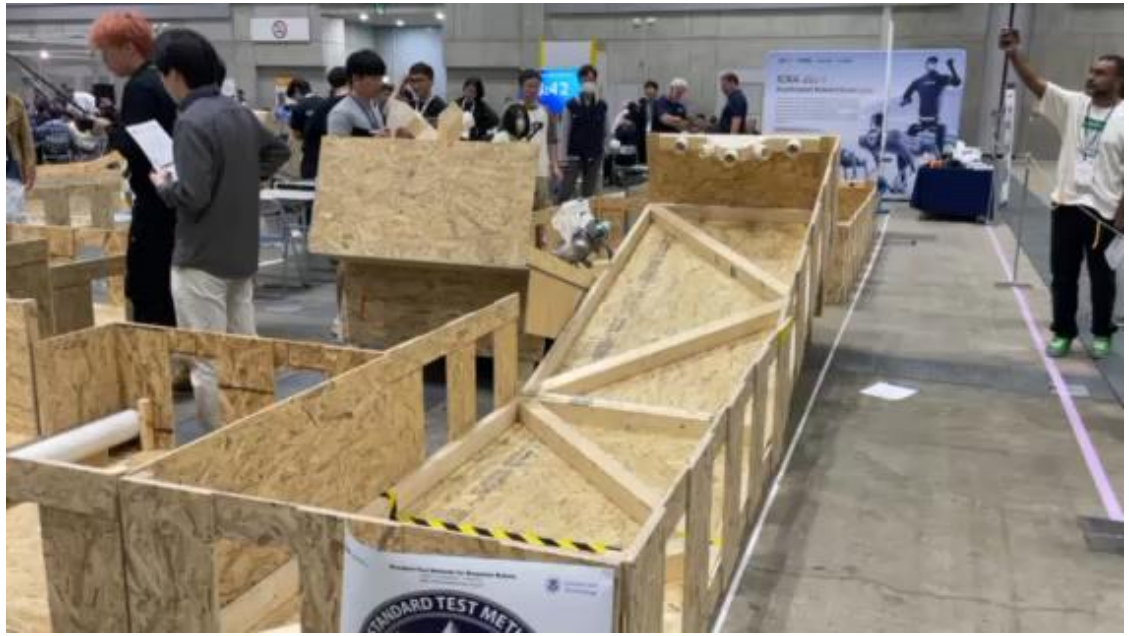

**Fig. 1 | The morphologically adaptive robot used in this study. a**, An overview of the main components of the robot. **b**, The robot with the shortest (left) and longest (right) leg configuration. AHRS, attitude and heading reference system; GPS, global positioning system; RGB-D, red, green, blue and depth.





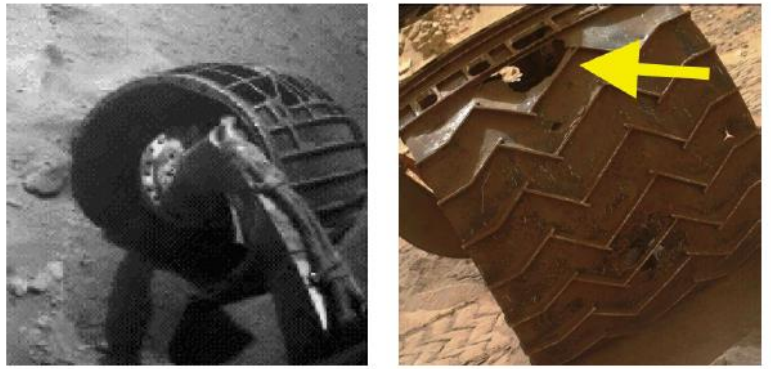

- Nygaard T F, Martin C P, Torresen J, et al. Real-world embodied AI through a morphologically adaptive quadruped robot[J]. Nature Machine Intelligence, 2021, 3(5): 410-419.



机遇号的轮子曾在2005年沉陷沙坑

2005年，机遇号的五个轮子陷入30厘米深的沙丘里，这些沙堆被称作“炼狱沙丘(Purgatory Dune)”。科研人员就曾模拟火星土壤帮助机遇号脱困。

Fig. 1: *Curiosity*'s experience in Hidden Valley highlights the need for an on-board terrain classification capability. (Image by NASA/JPL-Caltech)



(a) Wheel stuck in soft soil on MER *Spirit* rover. (b) Punctured wheels on MSL *Curiosity* rover.

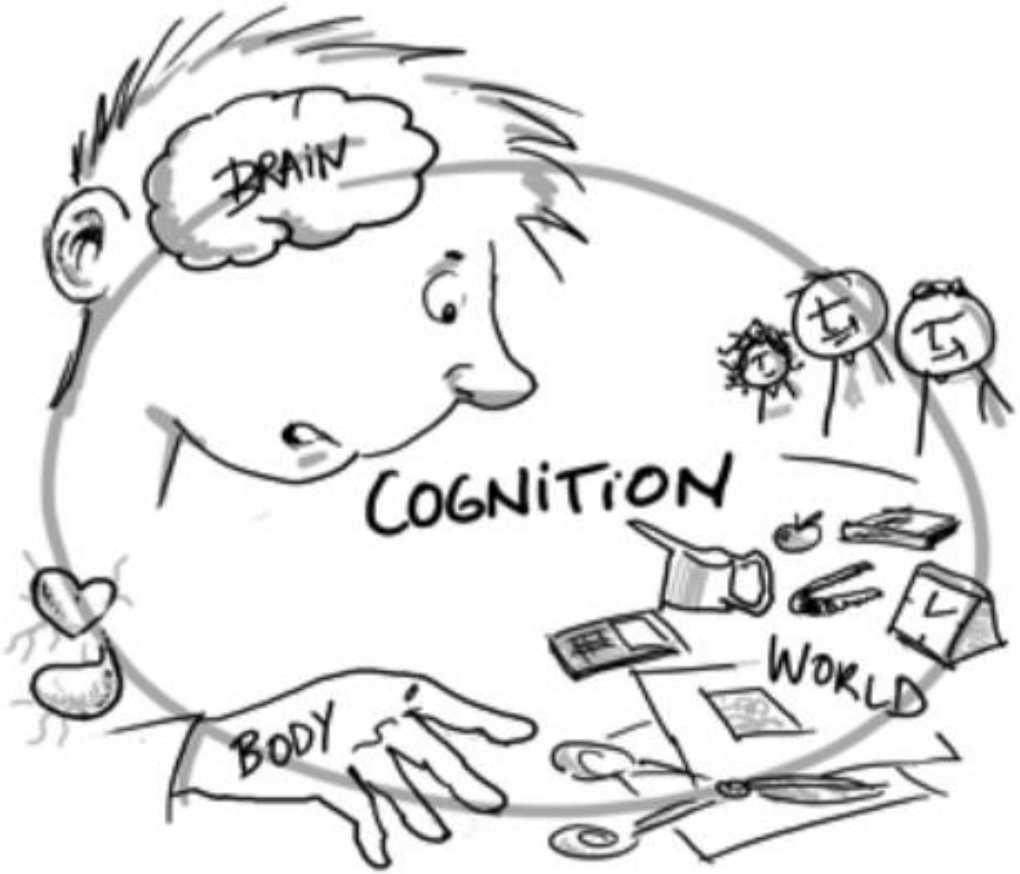
Fig. 2: Terrain risks on the Martian surface (Images by NASA/JPL-Caltech)

# 背景



## ➢ 具身学习

智能体（Agent）通过感知外部环境，产生思想并通过计算后，生成相应动作与环境交互，以此改变和影响环境，这个过程周而复始



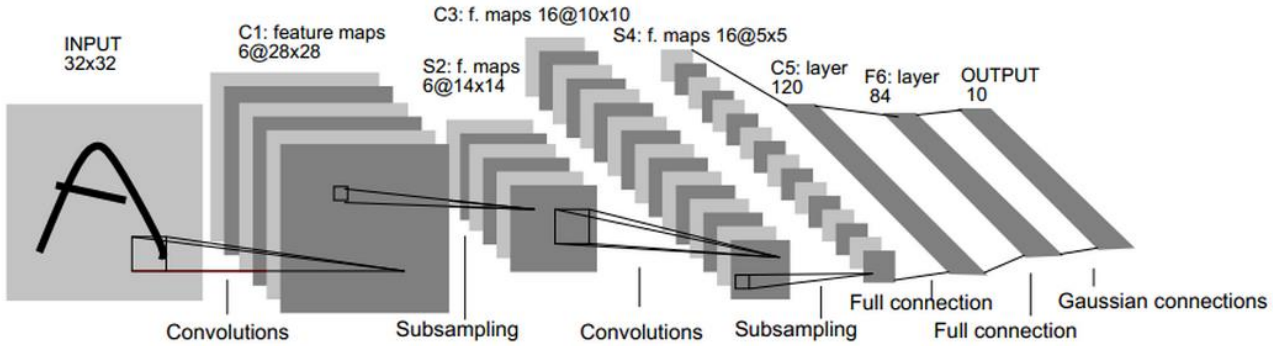

**具身学习强调身体在学习过程中的重要性**

# 背景



- 具身学习

## 监督式学习




$\{f(x_i), y_i\}$：特征与分类

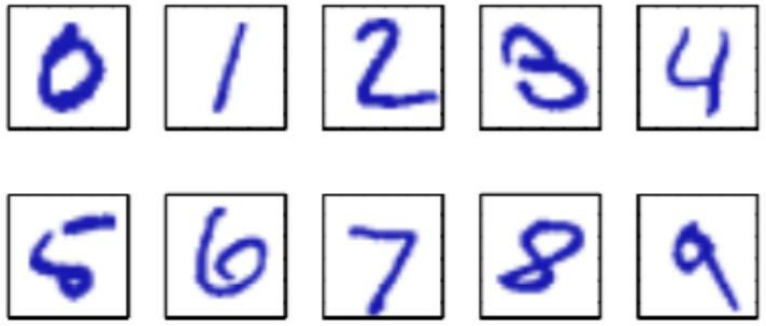
## 自监督学习


$\{f(x_i), y_i\}$：特征

- 背景
- 具身学习
- 前沿

# 行为→学习：具身学习

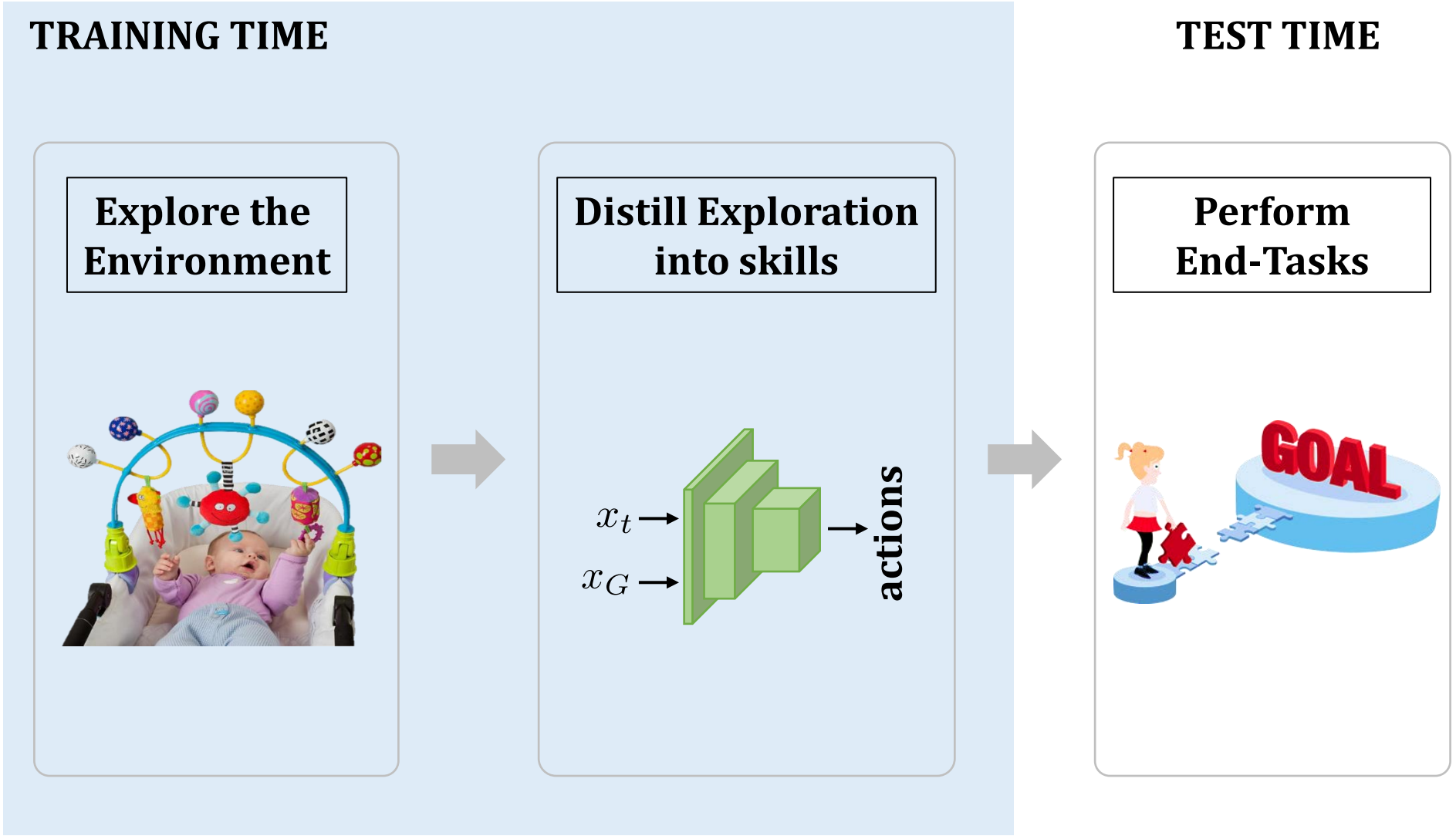
- 人是如何学习的？





# 行为→学习：具身学习

- 人是如何学习的？



# 行为→学习：具身学习

- 人是如何学习的？

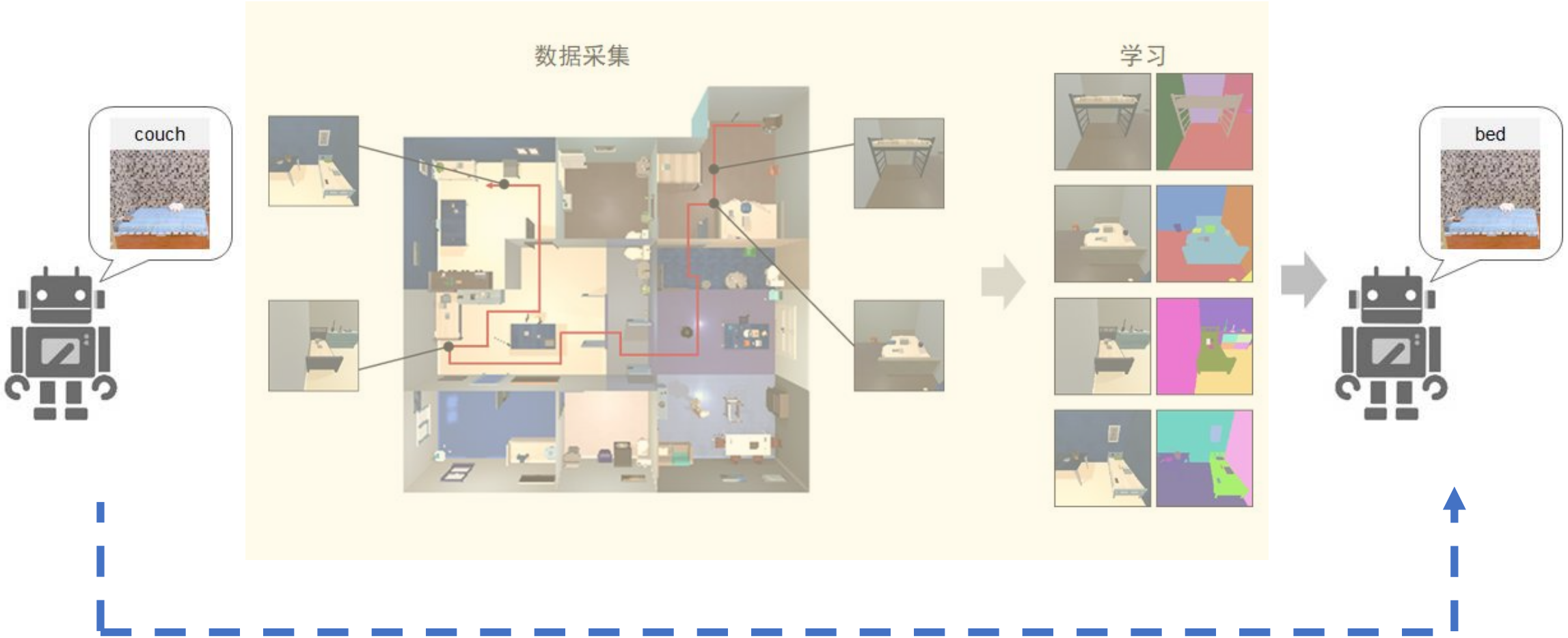
离线采集数据

vs.

在线采集数据

# 行为→学习：具身学习

- 问题描述

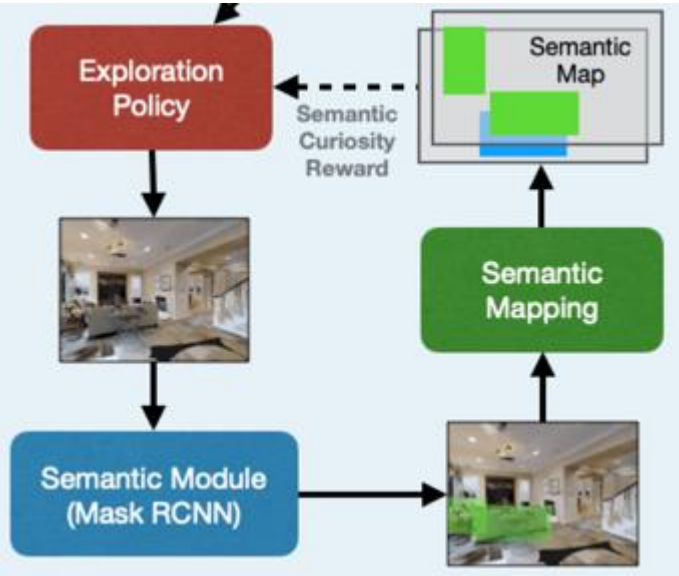


# 行为→学习：具身学习

## ➢ 问题描述

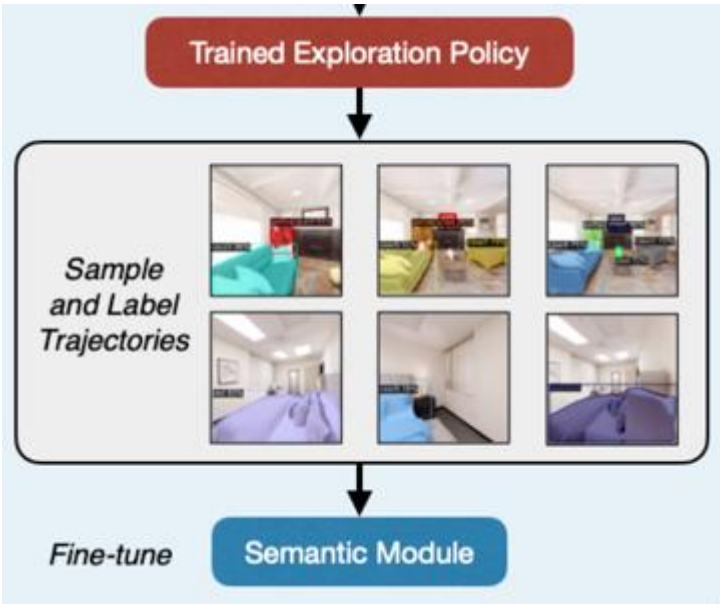
### 学习探索路径

学习如何探索环境才能获得更好的样本



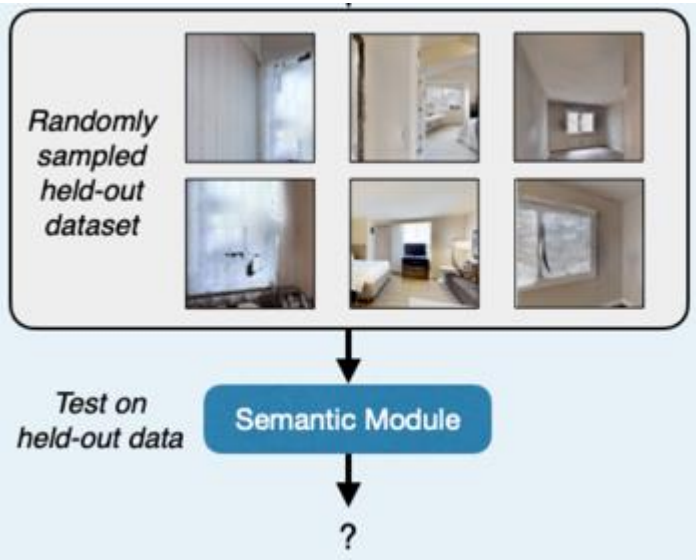

### 采集训练样本

利用探索策略获取样本与标签，提升感知器性能




### 部署应用

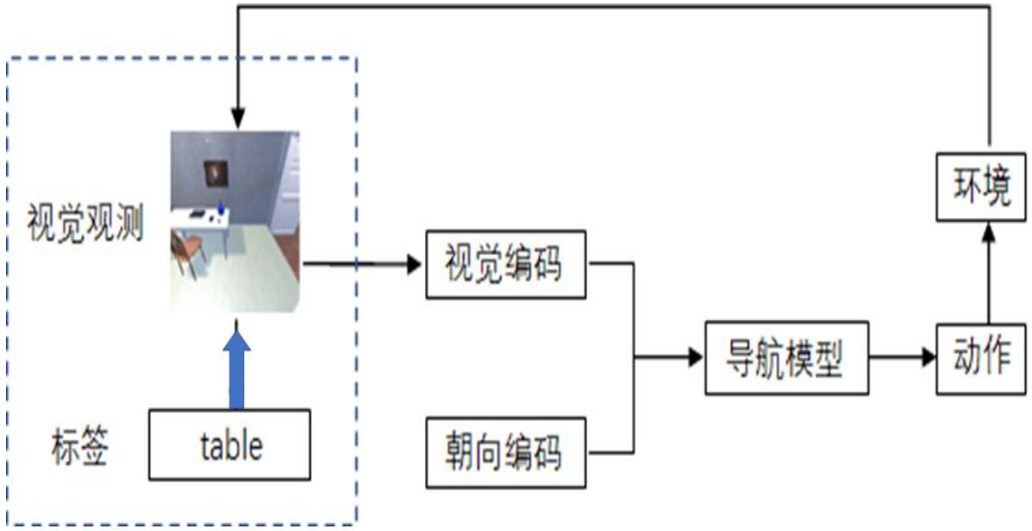
在新场景中对改进的感知器进行测试



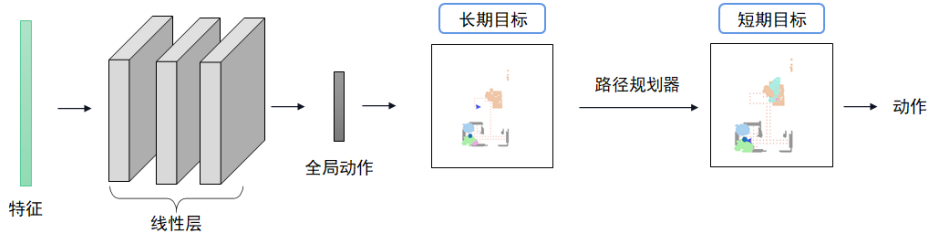
# 行为→学习：具身学习



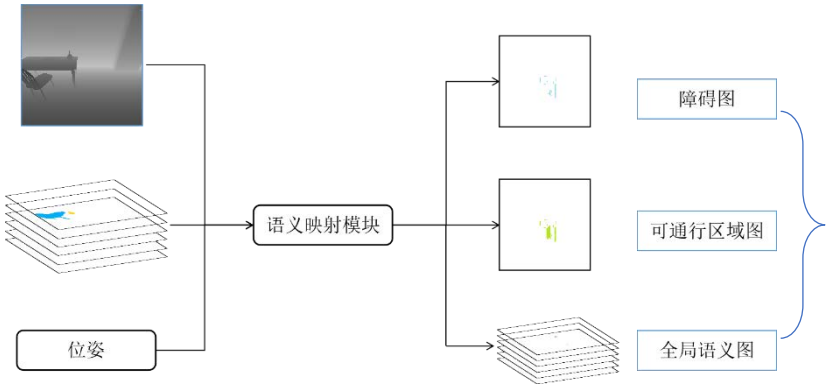
## ➢ 基本方法

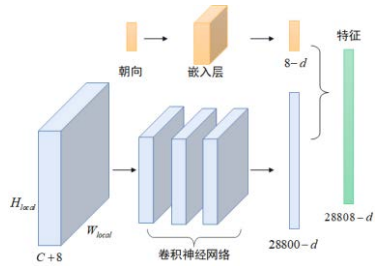


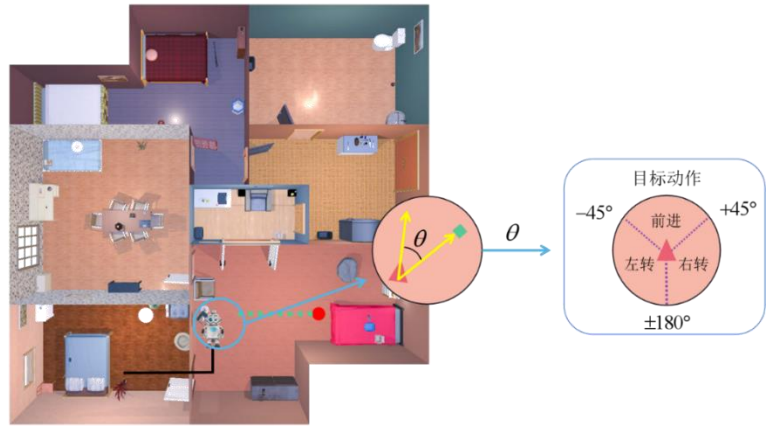

### 探索模型




### 视觉编码



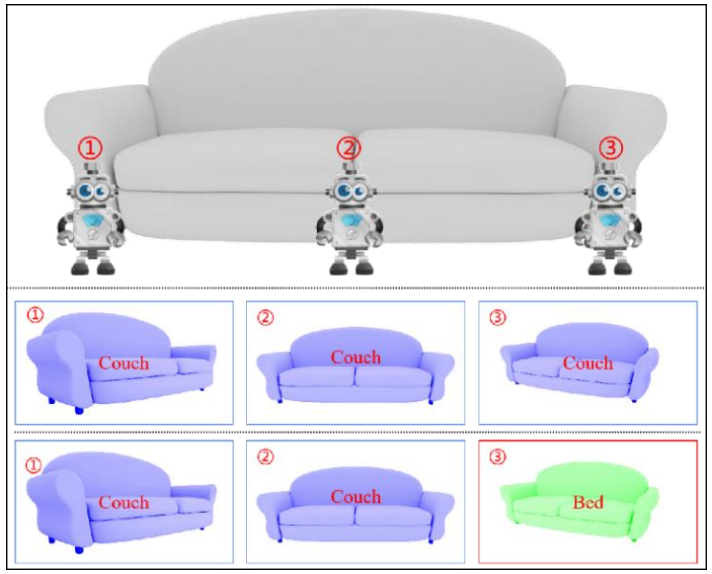



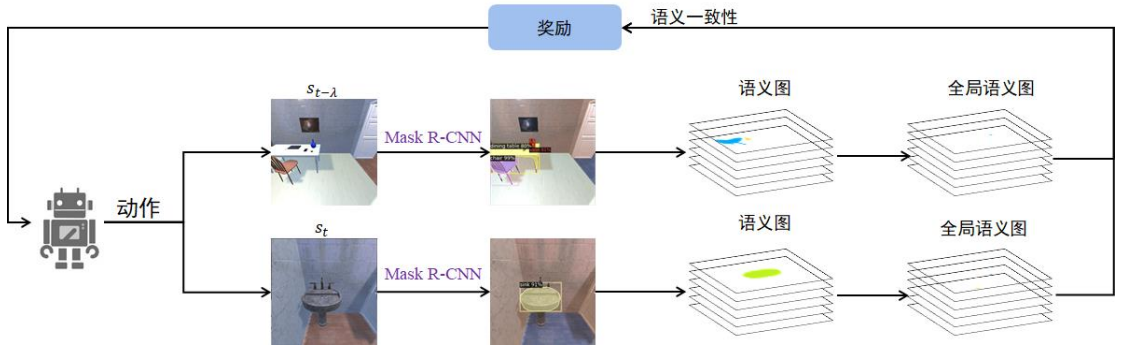



# 行为→学习：具身学习



## ➢ 模型训练




$$R = \sum_{n\in(1,2,\ldots);(i,j,c)\in(W_{full},H_{full},C)} abs(sem_{n\times\lambda}[i,j,c] - sem_{(n-1)\times\lambda}[i,j,c])$$

# 行为→学习：具身学习



## ➢ 实例

训练场景为房间：3、4、15、18、35、48、56、59、69、77、80、84、86、95、100、123、137、134、137、143；

验证场景为房间：146、148、151、152、162、169、170、172、173、177；

测试场景为房间：182、183、188、198、204、212、217、221、226、231；

在实验过程中，机器人的动作空间由平移运动和旋转运动构成，其中：

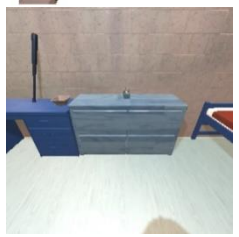
前进运动：向前的平移运动。每一次平移运动的移动距离固定为网格边长。

旋转运动：包括左转、右转两个方向的旋转运动。









$$ACC = \frac{TP + TN}{TP + TN + FP + FN}$$

$$IOU = \frac{\text{Target} \cap \text{Prediction}}{\text{Target} \cup \text{Prediction}}$$

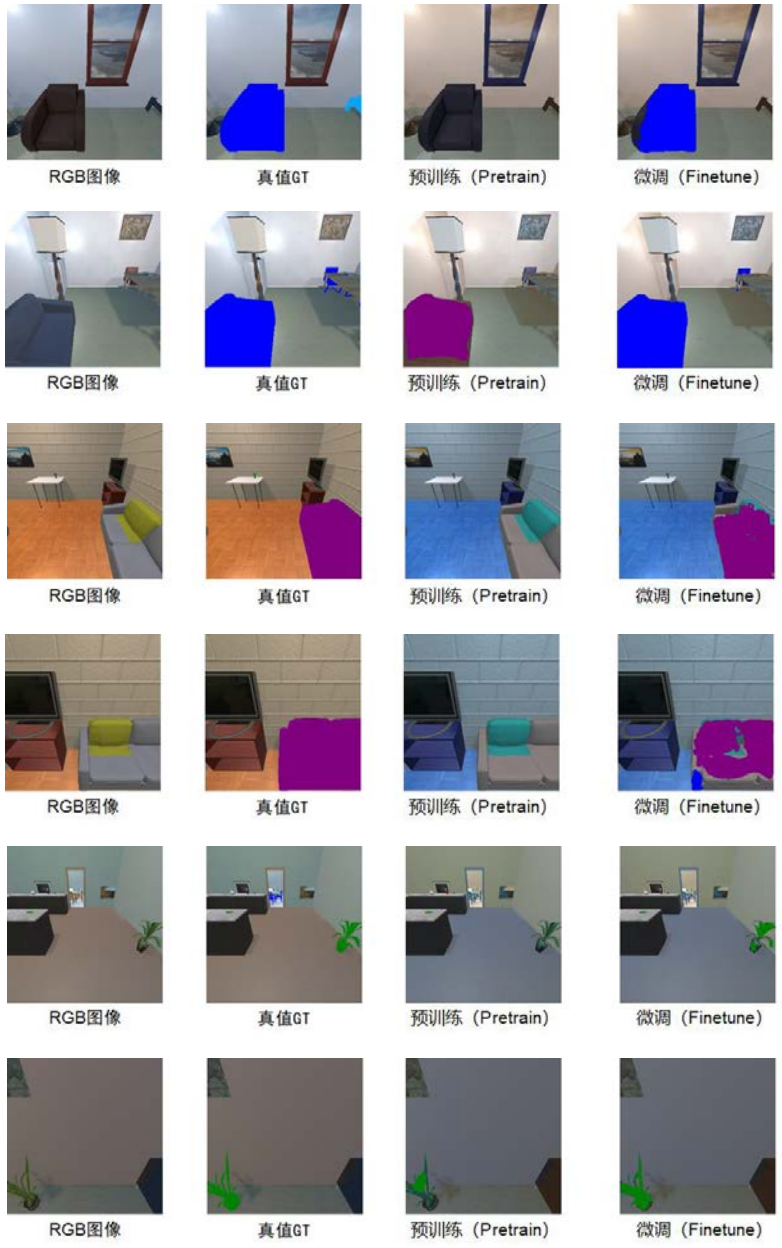
# 行为→学习：具身学习



## ➢ 实例

| 类别 | chair | couch | plant | bed | toilet | mACC |
|---|---|---|---|---|---|---|
| 预训练模型 | 35.24 | 4.73 | 58.03 | 2.35 | 37.07 | 27.48 |
| Random | 33.95 | 11.54 | 74.56 | 11.07 | 53.09 | 36.84 |
| Rule | 36.01 | 66.69 | 71.96 | 38.02 | 77.68 | 58.07 |
| Semantic | 60.89 | 68.31 | 77.35 | 35.75 | 83.52 | 60.89 |

| 类别 | chair | couch | plant | bed | toilet | mIOU |
|---|---|---|---|---|---|---|
| 预训练模型 | 30.40 | 4.31 | 57.77 | 2.32 | 31.37 | 21.03 |
| Random | 30.44 | 8.59 | 74.39 | 11.04 | 48.88 | 28.89 |
| Rule | 31.88 | 41.19 | 71.56 | 37.74 | 75.69 | 43.01 |
| Semantic | 45.22 | 42.45 | 77.18 | 35.49 | 81.50 | 45.22 |

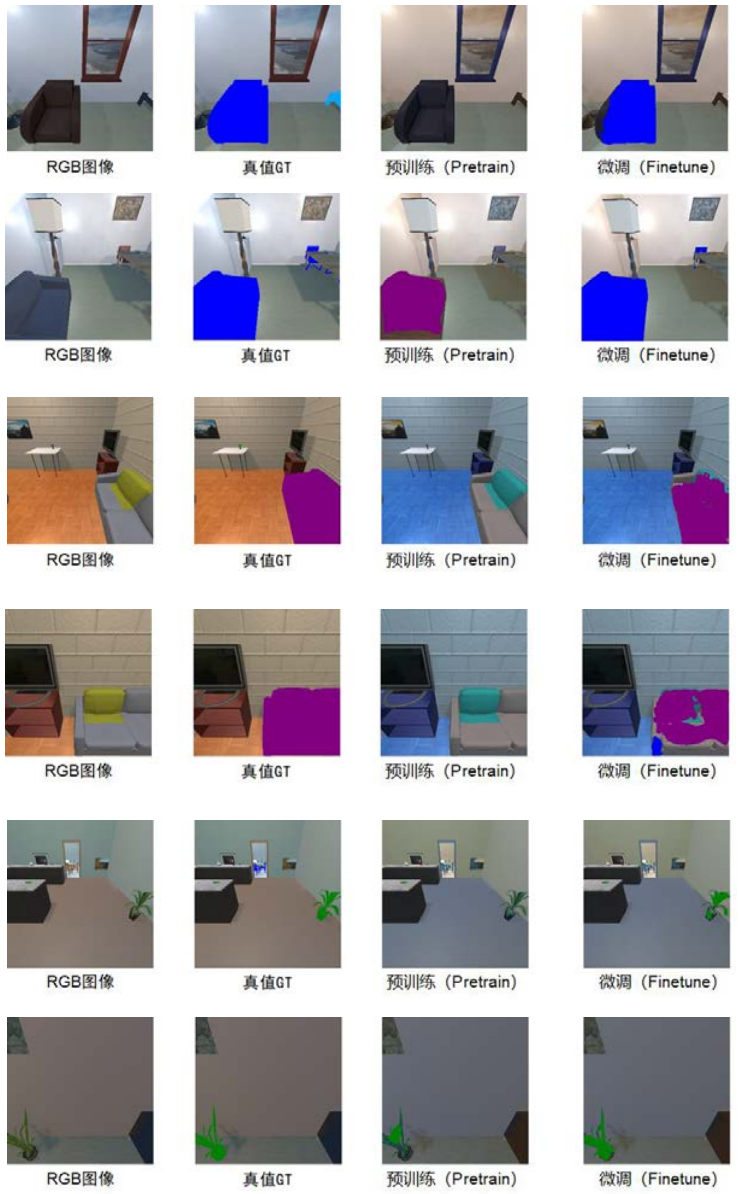
# 行为→学习：具身学习



## ➢ 实例

| 类别 | chair | couch | plant | bed | toilet | mACC |
|---|---|---|---|---|---|---|
| 预训练模型 | 35.24 | 4.73 | 58.03 | 2.35 | 37.07 | 27.48 |
| Random | 33.95 | 11.54 | 74.56 | 11.07 | 53.09 | 36.84 |
| Rule | 36.01 | 66.69 | 71.96 | 38.02 | 77.68 | 58.07 |
| Semantic | 60.89 | 68.31 | 77.35 | 35.75 | 83.52 | 60.89 |

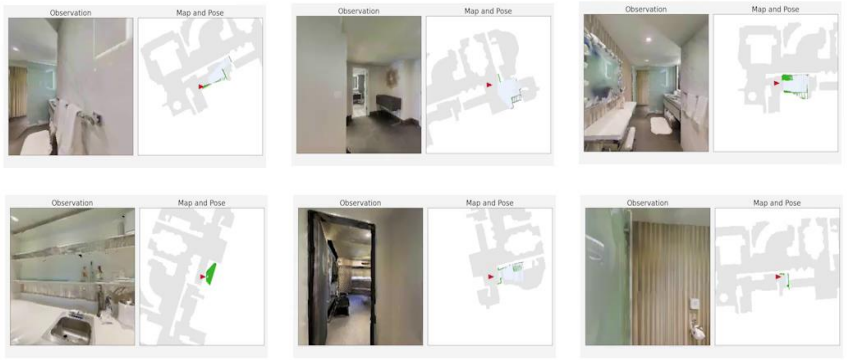
| 类别 | chair | couch | plant | bed | toilet | mIOU |
|---|---|---|---|---|---|---|
| 预训练模型 | 30.40 | 4.31 | 57.77 | 2.32 | 31.37 | 21.03 |
| Random | 30.44 | 8.59 | 74.39 | 11.04 | 48.88 | 28.89 |
| Rule | 31.88 | 41.19 | 71.56 | 37.74 | 75.69 | 43.01 |
| Semantic | 45.22 | 42.45 | 77.18 | 35.49 | 81.50 | 45.22 |

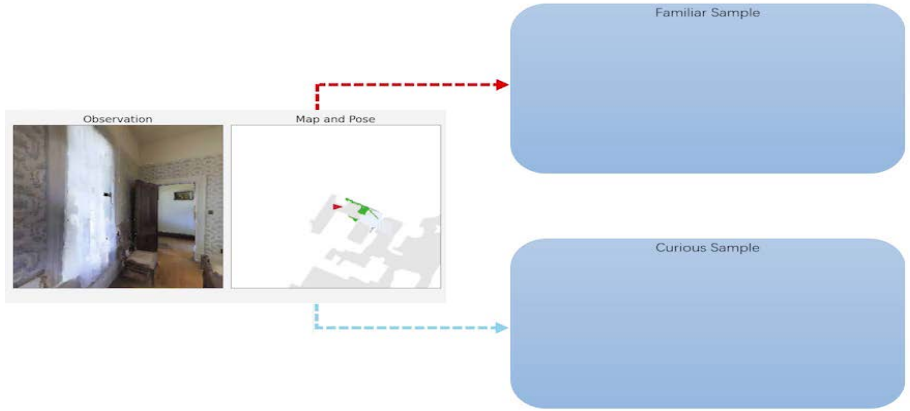



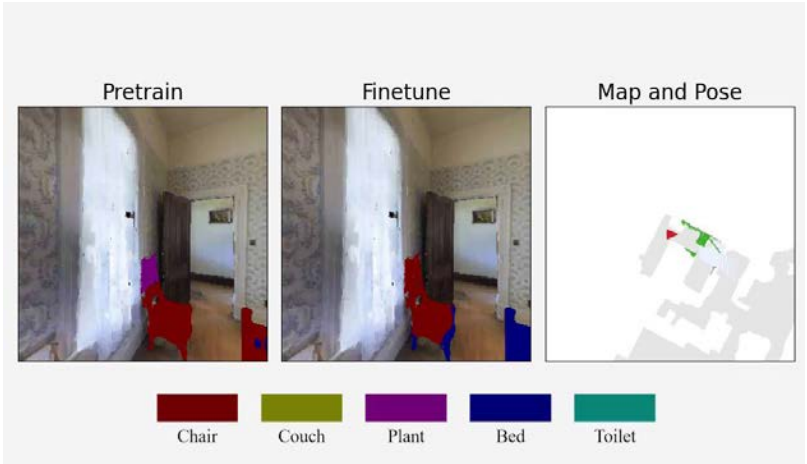
# 行为→学习：具身学习



- 实例




| Model | chair | couch | plant | bed | toilet | mACC |
|---|---|---|---|---|---|---|
| Pre-train | 53.58 | 38.74 | 68.15 | 19.99 | 0.00 | 36.09 |
| ANS | **75.38** | 11.71 | 68.09 | 25.42 | 0.00 | 36.12 |
| Random | 0.32 | 5.34 | 58.96 | 42.55 | 0.00 | 21.43 |
| Curious | 31.35 | 3.44 | 52.20 | **90.82** | 0.00 | 35.56 |
| Familiar | 54.44 | 12.60 | **76.20** | 60.94 | 0.00 | 40.84 |
| FC(Ours) | 53.82 | **39.91** | 72.74 | 52.84 | 0.00 | **43.86** |

- 背景
- 具身学习
- 前沿

1）端到端策略（直接生成动作）

2）层级探索策略（生成子目标）

3）自监督学习（生成伪标签）

4）增量学习（学习新类别、避免灾难性遗忘）

5）表示学习（学习特征表示，利用该特征表示提升感知能力）

## ➢ 端到端策略——语义好奇驱动的学习

- How should an exploration policy decide which trajectory should be labeled? One possibility is to use a trained object detector's failure cases as an external reward.
- However, this will require labeling millions of frames required for training RL policies, which is infeasible.
- Instead, we explore a self-supervised approach for training our exploration policy by introducing a notion of semantic curiosity.
- Our semantic curiosity policy is based on a simple observation ---- the detection outputs should be consistent.
- Therefore, our semantic curiosity rewards trajectories with inconsistent labeling behavior and encourages the exploration policy to explore such areas.
- The exploration policy trained via semantic curiosity generalizes to novel scenes and helps train an object detector that outperforms baselines trained with other possible alternatives such as random exploration, prediction-error curiosity, and coverage-maximizing exploration.

## ➢ 端到端策略——语义好奇驱动的学习

Several core research questions need to be answered:

(a) What is the policy of exploration that generates these observations?

(b) What should be labeled in these observations - one object, one frame, or the whole trajectory?

(c) How do we get these labels?

- Semantic curiosity for active visual learning, ECCV, 2020

## ➢ 端到端策略——语义好奇驱动的学习

- a good object detector has not only high mAP performance but is also consistent in predictions.
- the detector should predict the same label for different views of the same object.
- We use this meta-signal of consistency to train our action policy by rewarding trajectories that expose inconsistencies in an object detector.

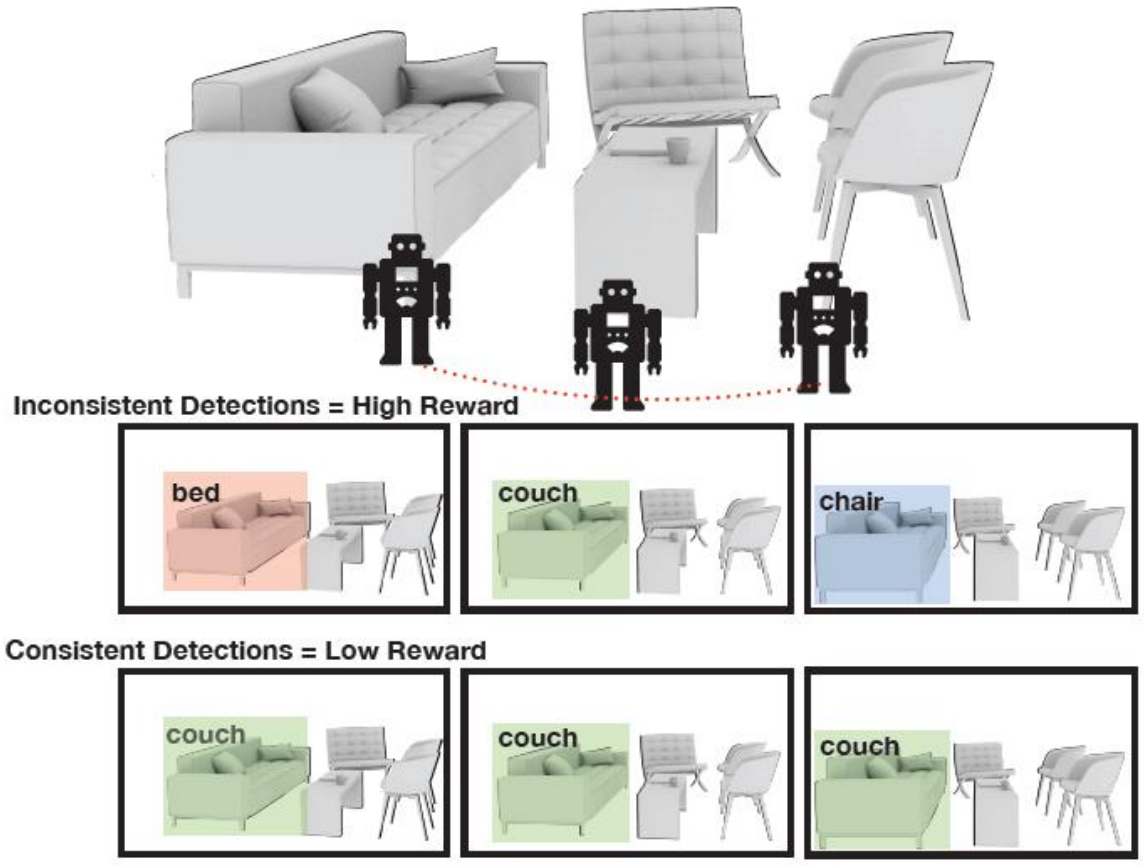
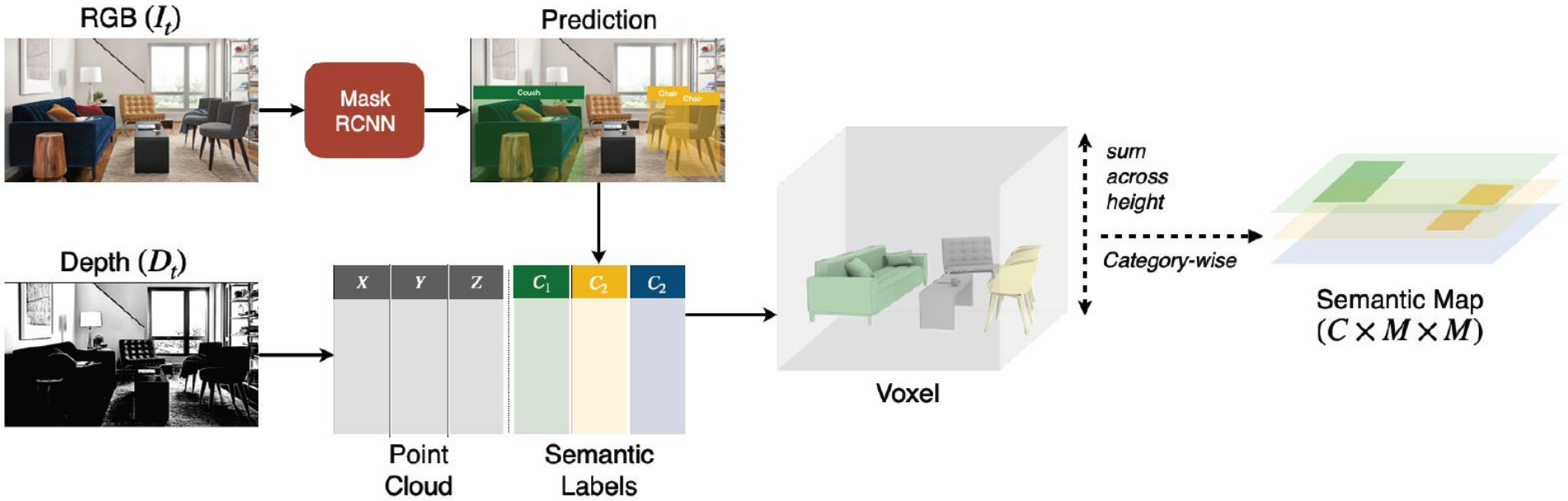



**Fig. 1: Semantic Curiosity**: We propose semantic curiosity to learn exploration for training object detectors. Our semantically curious policy attempts to take actions such that the object detector will produce inconsistent outputs.

## ➢ 端到端策略——语义好奇驱动的学习




The Semantic Mapping module takes in a sequence of RGB (It) and Depth (Dt) images and produces a top-down Semantic Map.

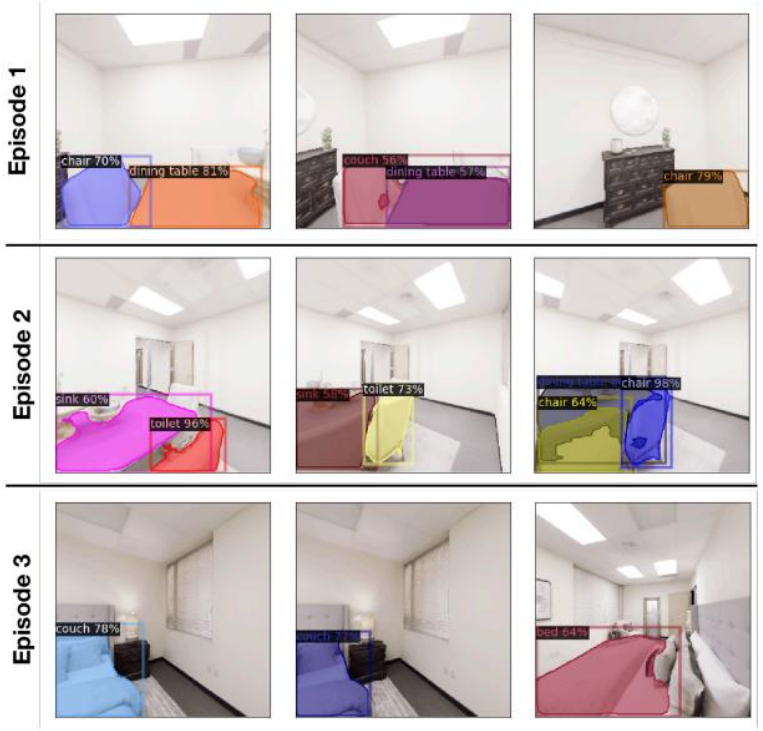
The semantic map allows us to associate the object predictions across different frames as the agent is moving. We dene the semantic curiosity reward based on the temporal inconsistency of the object predictions.

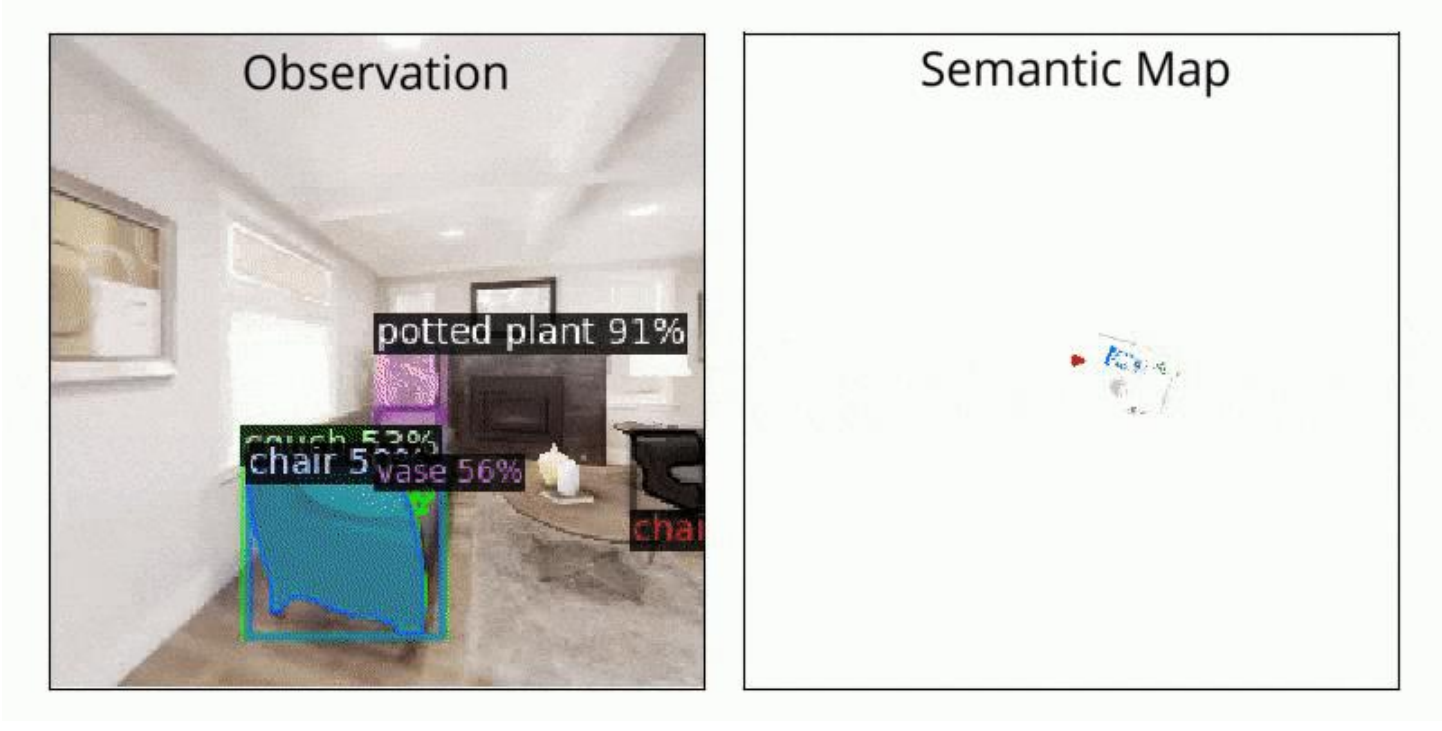
$$r_{SC} = \lambda_{SC} \Sigma_{(c,i,j)\in(C,M,M)} (M_t^{Sem}[c,i,j] - M_{t-1}^{Sem}[c,i,j])$$

# 具身导航学习前沿



## ➢ 端到端策略——语义好奇驱动的学习




Temporal Inconsistency Examples. Figure showing example trajectories sampled from the semantic curiosity exploration policy. We highlight the segmentation/detection inconsistencies of Mask RCNN. By obtaining labels for these images, the Mask RCNN pipeline improves the detection performance signicantly.

## ➢ 端到端策略——语义好奇驱动的学习

- Random：随机动作
- Prediction Error Curiosity：最大化前向预测
- Object Exploration：最大化Mak-RCNN检测数量
- Coverage Exploration：最大化探索区域
- Active Neural SLAM：最大化探索区域

**Table 1: Analysis.** Comparing the proposed Semantic Curiosity policy with the baselines along different exploration metrics.

| Method Name | Semantic Curiosity Reward | Explored Area | Number of Object Detections |
|---|---|---|---|
| Random | 1.631 | 4.794 | 82.83 |
| Curiosity [37] | 2.891 | 6.781 | 112.24 |
| Object exploration reward | 2.168 | 6.082 | 382.27 |
| Coverage Exploration [14] | 3.287 | 10.025 | 203.73 |
| Active Neural SLAM [9] | 3.589 | 11.527 | 231.86 |
| Semantic Curiosity | 4.378 | 9.726 | 291.78 |

从语义好奇奖励（不一致语义结果）、探索区域、检测到的物体数量三个维度分析对比

**Table 2: Quality of object detection on training trajectories.** We also analyze the training trajectories in terms of how well the pre-trained object detection model works on the trajectories. We want the exploration policy to sample hard data where the pre-trained object detector fails. Data on which the pre-trained model already works well would not be useful for fine-tuning. Thus, lower performance is better.

| Method Name | Chair | Bed | Toilet | Couch | Potted Plant | Average |
|---|---|---|---|---|---|---|
| Random | 46.7 | 28.2 | 46.9 | 60.3 | 39.1 | 44.24 |
| Curiosity [37] | 49.4 | 18.3 | 1.8 | 67.7 | 49.0 | 37.42 |
| Object Exploration | 54.3 | 24.8 | 5.7 | 76.6 | 49.6 | 42.2 |
| Coverage Exploration [14] | 48.5 | 23.1 | 69.2 | 66.3 | 48.0 | 51.02 |
| Active Neural SLAM [9] | 51.3 | 20.5 | 49.4 | 59.7 | 45.6 | 45.3 |
| Semantic Curiosity | 51.6 | 14.6 | 14.2 | 65.2 | 50.4 | 39.2 |

训练得到的轨迹中，各类物体检测结果
检测结果越差，代表采集到的数据越有利于提升检测器性能

**Table 3: Object Detection Results.** Object detection results in the Matterport domain using the proposed Semantic Curiosity policy and the baselines. We report AP50 scores on randomly sampled images in the test scenes. Training on data gathered from the semantic curiosity trajectories results in improved object detection scores.

| Method Name | Chair | Bed | Toilet | Couch | Potted Plant | Average |
|---|---|---|---|---|---|---|
| PreTrained | 41.8 | 17.3 | 34.9 | 41.6 | 23.0 | 31.72 |
| Random | 51.7 | 17.2 | 43.0 | 45.1 | 30.0 | 37.4 |
| Curiosity [37] | 48.4 | 18.5 | 42.3 | 44.3 | 32.8 | 37.26 |
| Object Exploration | 50.3 | 16.4 | 40.0 | 39.7 | 29.9 | 35.26 |
| Coverage Exploration [14] | 50.0 | 19.1 | 38.1 | 42.1 | 33.5 | 36.56 |
| Active Neural SLAM [9] | 53.1 | 19.5 | 42.0 | 44.5 | 33.4 | 38.5 |
| Semantic Curiosity | 52.3 | 22.6 | 42.9 | 45.7 | 36.3 | **39.96** |

- Semantic curiosity for active visual learning, ECCV, 2020

## ➢ 端到端策略——3D Embodied Learning

- 现有的一些具身AI平台是基于合成数据的，在这些平台上训练的模型在真实世界中会出现性能下降。因此，作者提出了一个用于机器人具身学习的真实3D具身数据集，该数据集由7个真实室内场景中的真实密集点云数据组成，通过模拟机器人在室内环境中的运动和交互，可以获得真实的视觉数据，且不会导致模型在现实环境中性能的下降。
- 数据规模：7个场景，采样方向为8个方向（每次转45°），采样点间距离为25cm。
- 动作空间：前、后、左、右、顺时针转45°、逆时针转45°

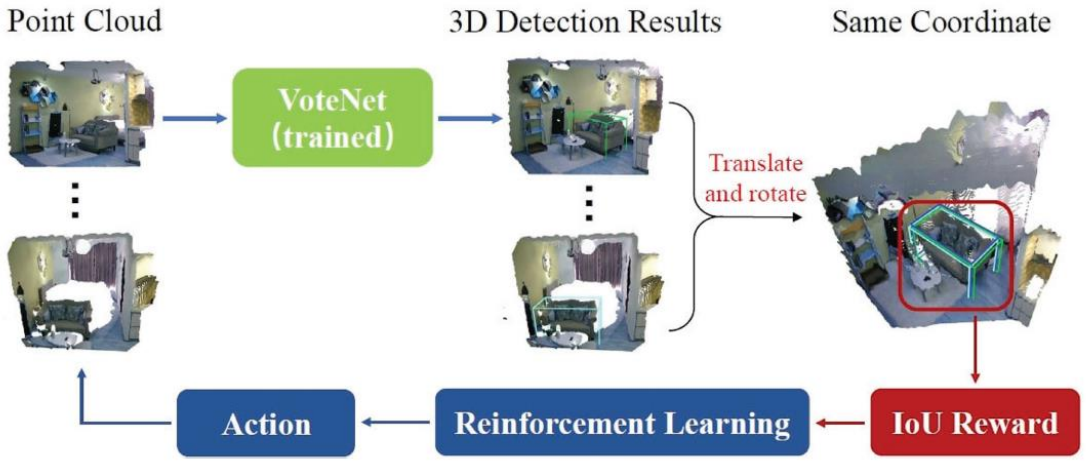
Fig. 2. Some samples from seven real world indoor scenes. There are noises and light reflections in these samples.

TABLE I
COMPARISON OF EMBODIED AI PLATFORMS AND OUR DATASET

| | Gibson | Replica | Ours |
|---|---|---|---|
| Scenes | 572 | 18 | 7 |
| Data | Synthetic data | Virtual room | Real data |
| Classes | - | 88 | 12 |
| Light reflections | no | no | yes |
| Real noises | no | no | yes |
| Annotations | 3D Semantic | 3D Semantic | 3D Detection |

Zhao Q, Zhang L, Wu L, et al. A Real 3D Embodied Dataset for Robotic Active Visual Learning[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 6646-6652.

## ➢ 端到端策略——3D Embodied Learning




Detection bounding boxes from different view positions are transferred into the same coordinate system through translation and rotation. Then, the 3D divergency reward formulation will calculate the difference between these boxes and use the reward to train the reinforcement learning model.


Fig. 7. The green line in the left figure represents the trajectory of the agent planned by our policy, and the six figures on the right correspond to the sampling data and annotations of the six representative red sampling points among the 50 sampling points on the left.

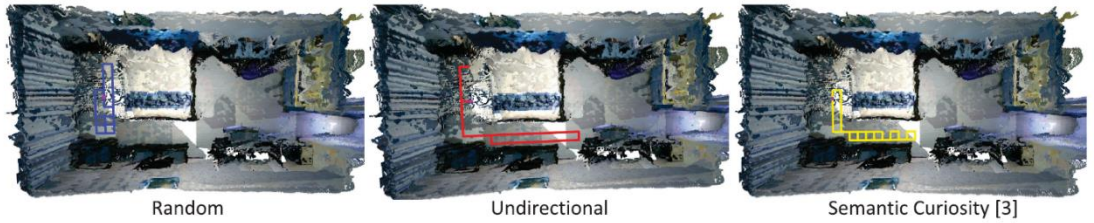



Fig. 8. The visualization of trajectories generated by other three baseline policies which share the same starting point (represented by purple dots).

$$r_t = \lambda \sum_c \left( n_t^c + n_{t-1}^c - IoU_{(t-1)(t)}^c \right),$$

## ➢ 端到端策略——3D Embodied Learning

- 探索策略（3D的语义好奇）：
  - 在不同的视角位置为同一物体产生一致的3D检测结果（大小、位置和标签）

TABLE II
PERFORMANCES OF THE RANDOM POLICY AND OUR POLICY IN UNSEEN (TEST) SCENES

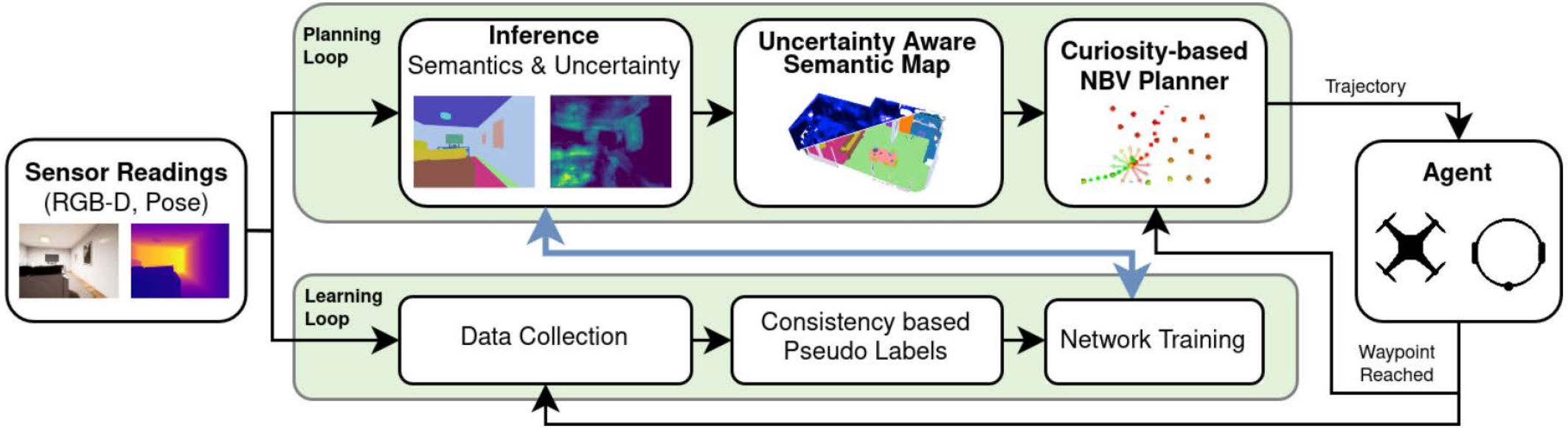
| Policy Name | mAP (in testing (unseen) scenes) | mAP (in training scenes) | 3D Divergency Reward | Var of mAP (in testing scenes) |
|---|---|---|---|---|
| Pretrained | 8.71 | 56.51 | - | - |
| Random Policy | 23.16 | **56.86** | 58.67 | ±14.24 |
| Unidirectional Policy | 26.20 | 55.98 | 62.16 | ±15.95 |
| Maximum Distance Policy | 25.68 | 56.07 | 62.59 | ±18.57 |
| Semantic Curiosity Policy [3] | 27.05 | 56.32 | 65.98 | ±17.13 |
| 3D Divergency Policy (ours) | **28.48** | 56.48 | **73.36** | ±16.86 |

- Random：随机采取行动
- Undirectional：朝着一个方向行动，碰到边界时会选择另外一个方向
- Maximum distance：选择最远的位置作为目标，到达目标后选择下一个最远的位置作为目标
- Semantic curiosity plilcy：利用2D语义好奇作为奖励

Zhao Q, Zhang L, Wu L, et al. A Real 3D Embodied Dataset for Robotic Active Visual Learning[J]. IEEE Robotics and Automation Letters, 2022, 7(3): 6646-6652.

## ➢ 端到端策略——信息路径规划

- Embodied active domain adaptation: 语义分割网络往往无法很好地推广到未知环境，因此会使用智能体收集新环境的图像，然后用于自监督的域自适应来更新语义分割网络。作者将这个问题形式化为一个信息路径规划问题，并提出了一种新的信息增益，利用从语义模型中提取的不确定性来收集相关数据。随着域适应的变化，不确定性会随时间变化，所提出的方法能快速反馈给智能体以采集不同的数据。
- Combining the advantages of classical and learning-based methods.




- Zurbrügg R, Blum H, Cadena C, et al. Embodied active domain adaptation for semantic segmentation via informative path planning[J]. IEEE Robotics and Automation Letters, 2022, 7(4): 8691-8698.

# 具身导航学习前沿

## ➢ 端到端策略——信息路径规划

TABLE I

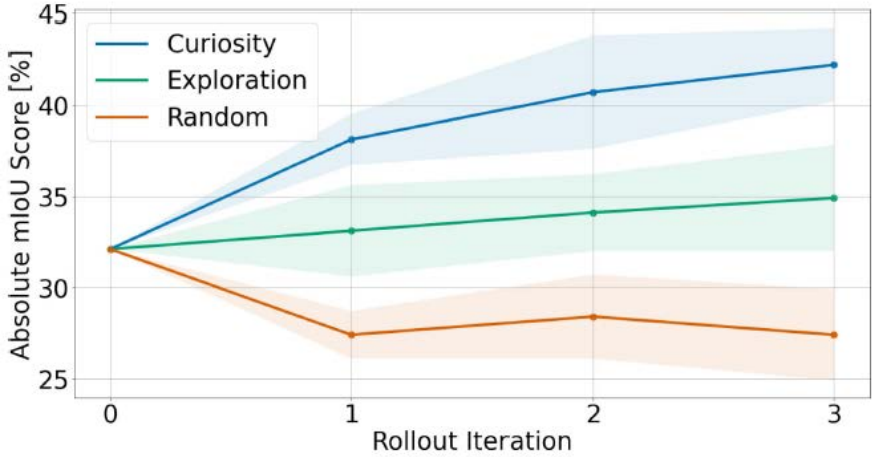
FINAL MIOU SCORE [%] OF THE NETWORK AFTER THREE AUTONOMOUS DOMAIN ADAPTATION CYCLES. WHILE EXPLORATION IS A STRONG BASELINE, OUR UNCERTAINTY-BASED INFORMATION GAIN CONSISTENTLY FURTHER IMPROVES THE MODEL

| Environment | Method | NYU 40 | Eigen 13 |
|---|---|---|---|
| Livingroom-Kitchen | Pretrained Model | 32.1 | 50.0 |
| | Random | 27.4 ± 2.5 | 55.5 ± 6.8 |
| | Exploration | 34.9 ± 2.9 | 68.5 ± 4.7 |
| | Curiosity (ours) | **42.2 ± 2.0** | **69.8 ± 2.8** |
| Bedroom-Office | Pretrained Model | 27.7 | 40.6 |
| | Random | 30.7 ± 3.8 | 46.1 ± 9.6 |
| | Exploration | 39.9 ± 2.2 | 59.6 ± 3.8 |
| | Curiosity (ours) | **40.6 ± 2.5** | **62.5 ± 4.6** |
| Bathroom-Other | Pretrained Model | 19.2 | 39.2 |
| | Random | 20.5 ± 2.6 | 43.8 ± 1.9 |
| | Exploration | 18.6 ± 1.3 | **51.8 ± 2.0** |
| | Curiosity (ours) | **21.2 ± 1.9** | 50.9 ± 1.2 |

TABLE II

MIOU (NYU40 / EIGEN13) [%] ON LIVINGROOM-KITCHEN: COMPARISON OF SELF-SUPERVISION METHODS USING THE DATA COLLECTED WITH DIFFERENT PLANNERS, AVERAGED OVER SEVEN RUNS. MOST METHODS BENEFIT FROM CURIOS DATA COLLECTION

| Method | Random | Exploration | Curiosity |
|---|---|---|---|
| Fourier Transfer [8] | 32.5 / 49.7 | 33.2 / 49.0 | 33.8 / 49.0 |
| Self-training[4] | **32.7** / 52.9 | 31.8 / 51.8 | 33.6 / 54.4 |
| Uncertainty Reduction [7] | **32.7** / 52.4 | 33.4 / 53.7 | 37.4 / 58.0 |
| Spatial Consistency (ours) | 27.4 / **55.5** | **34.9 / 68.5** | **42.2 / 69.8** |



## ➢ 端到端策略——信息路径规划

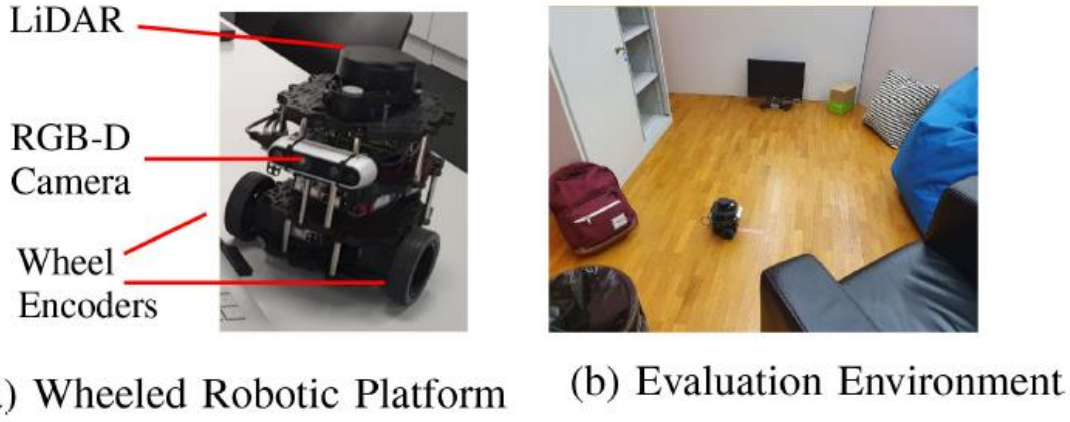



Fig. 5. Real world turtle bot setup.

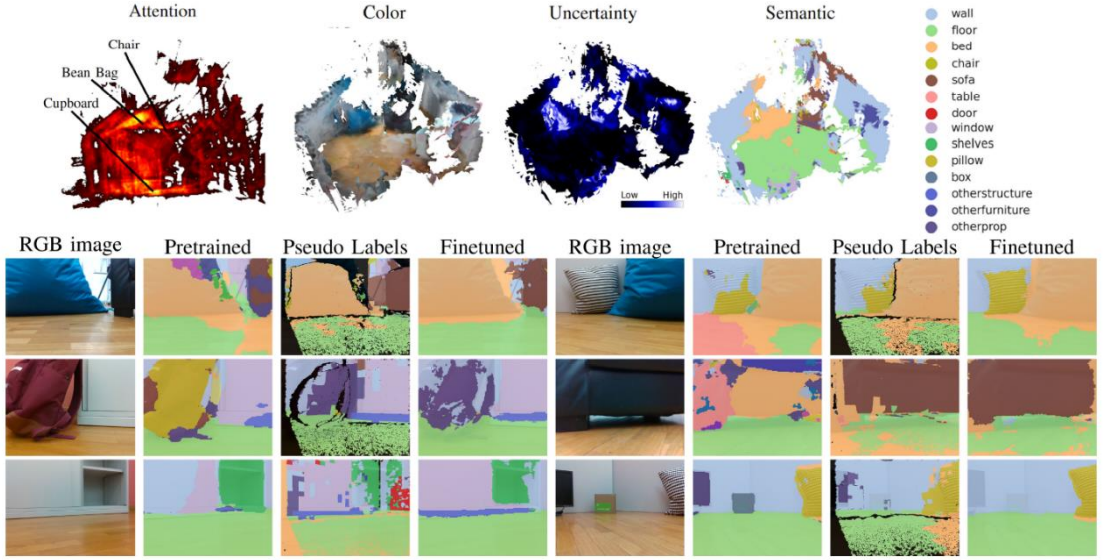



Fig. 6. Top row: Overview of the attention map and reconstructed meshes. In the attention map, points that occur in multiple views are colored brighter. Points with a height $z \leq 2$ cm are omitted for clarity. Uncertain objects such as the Pillow, Bean Bag, and Cupboard are scrutinized by our planner. The multiple surfaces, e.g. on the left, show the extent of noise in the state estimation. Bottom area: Selected images from the real-world experiment. Initially highly uncertain objects such as the Bean Bag (top row), indicated by noisy predictions, are thoroughly mapped by our active system and well represented after domain adaptation. Miss-classified objects (center row), e.g. the backpack labeled chair or chair labeled bed, are corrected over multiple observations and correctly recognized afterwards. Limitations of our method (bottom row) include a shelf, initially partly mislabeled as wall/window that is now only window, and small objects like the box not being captured in the map due to pose noise, which are then forgotten by the model. We observe a general trend of our method to improve the accuracy of object boundaries during autonomous domain adaptation.

## ➢ 层级策略——Embodied Learning

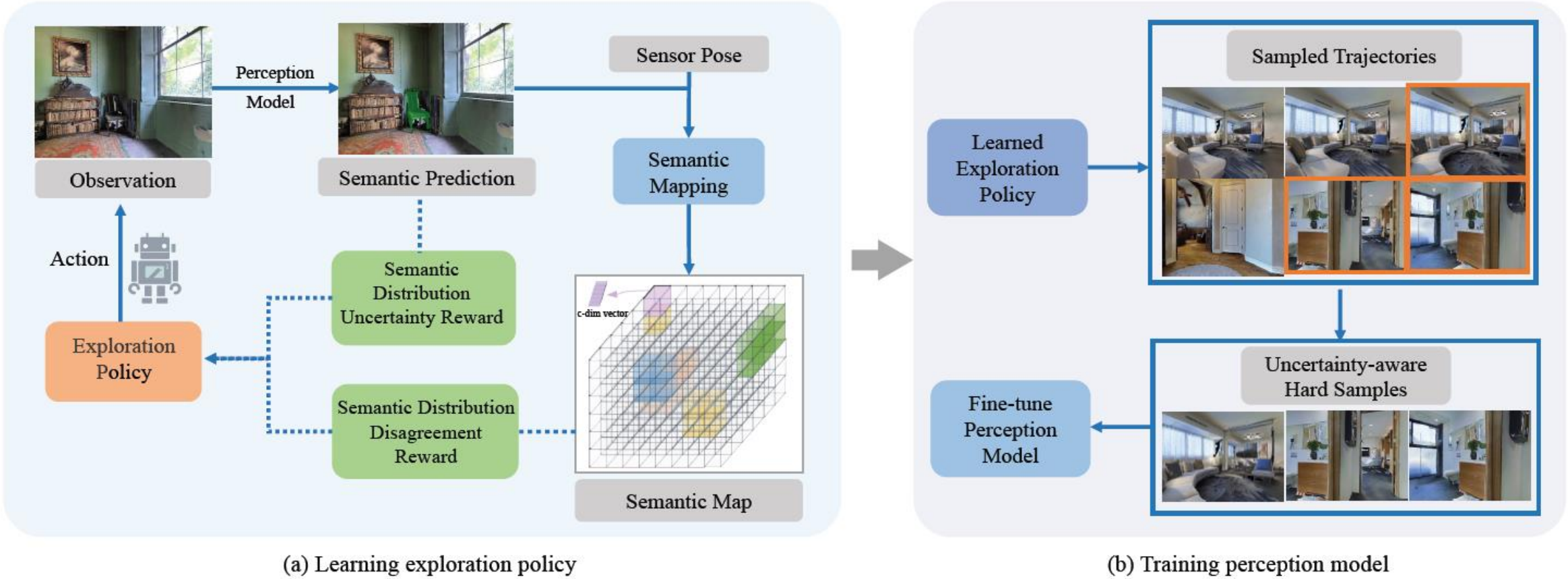



Fig. 2. The architecture of our proposed informative trajectory and sample exploration method. It contains two steps: the exploration policy aims to encourage the agent to explore the objects with semantic distribution disagreement or uncertainty, then the training stage aims at gathering hard samples on trajectories based on semantic distribution uncertainty to fine-tune the pre-trained model.

## ➢ 层级策略——Embodied Learning


Fig. 1. The illustration of semantic distribution disagreement, e.g., the bed is recognized to different objects/distributions across three viewpoints $(v1, v2, v3)$, and semantic distribution uncertainty, e.g., the probability of couch in observation $o1$ being predicted as bed and couch is relatively close. bg means background.

The semantic distribution disagreement reward is defined as the Kullback-Leibler divergence between the current prediction and the 3D semantic distribution map, which encourages the agent to explore the objects with different semantic distributions across viewpoints:

$$r_d = KL(m_t, M_{t-1}). \tag{2}$$

In addition, we propose a semantic distribution uncertainty reward $r_u$ to explore the objects whose predicted probabilities belonging to two categories are relatively close, as Eq. 4 explains.

$$r_u = \begin{cases} 1, & u > \delta \\ 0, & u < \delta \end{cases} \tag{3}$$

- 语义分布不一致：不同视图下，同一物体被识别为不同物体/分布（如左图bed被识别为couch）
- 语义分布不确定：同一物体被预测为概率接近的两个物体（如右图couch和bed概率相近）

# ➢ 层级策略——Embodied Learning

TABLE I

COMPARISON WITH THE STATE-OF-THE-ART METHODS FOR OBJECT DETECTION (BBOX) AND INSTANCE SEGMENTATION (SEGM) USING AP50 AS THE METRIC. N MEANS THE EXPLORATION POLICY IS PROGRESSIVELY TRAINED FOR N TIMES.

| Task | Method | Chair | Couch | Potted Plant | Bed | Toilet | Tv | Average |
|---|---|---|---|---|---|---|---|---|
| Bbox | Pre-trained | 21.05 | 25.23 | 22.58 | 24.24 | 22.62 | 29.67 | 24.23 |
| | Re-trained | 23.77 | 27.36 | 26.20 | 25.21 | 24.82 | 34.48 | 26.97 |
| | Random | 29.98 | 31.65 | 23.91 | 28.66 | 31.78 | 40.44 | 31.07 |
| | Active Neural SLAM [23] | 32.02 | 32.74 | 31.94 | 30.31 | 26.30 | 38.68 | 32.00 |
| | Semantic Curiosity [12] | 33.51 | 33.11 | 32.91 | 29.57 | 25.76 | 39.97 | 32.46 |
| | Ours (n=1) | 33.57 | 34.36 | 32.79 | 31.54 | 28.38 | 43.81 | 34.07 |
| | Ours (n=3) | 33.34 | 34.48 | 35.28 | 32.12 | 31.87 | 43.11 | **35.03** |
| Segm | Pre-trained | 12.72 | 22.98 | 16.71 | 23.82 | 23.85 | 29.75 | 21.64 |
| | Re-trained | 14.99 | 24.68 | 18.36 | 24.32 | 25.15 | 34.23 | 23.62 |
| | Random | 18.22 | 27.25 | 8.82 | 28.19 | 29.08 | 39.39 | 25.16 |
| | Active Neural SLAM [23] | 17.89 | 29.24 | 15.22 | 29.66 | 27.29 | 38.61 | 26.32 |
| | Semantic Curiosity [12] | 18.18 | 30.06 | 18.39 | 29.03 | 26.70 | 40.01 | 27.06 |
| | Ours (n=1) | 19.18 | 30.14 | 15.56 | 31.03 | 28.19 | 43.43 | 27.92 |
| | Ours (n=3) | 19.28 | 30.13 | 16.22 | 31.27 | 28.92 | 44.76 | **28.42** |

- Pre-trained：在原始COCO数据集上训练的模型
- Re-trained：使用COCO数据集中六类重新训练的模型
- Random：随机采取行动
- n=3：基于最新微调感知模型迭代式的训练探索策略三次

TABLE IV

EFFECTS OF PROGRESSIVELY TRAINING THE EXPLORATION POLICY FOR $n$ TIMES ON THE OBJECT DETECTION TASK.

| Method | Chair | Couch | Potted Plant | Bed | Toilet | Tv | Average |
|---|---|---|---|---|---|---|---|
| n = 1 | 33.57 | 34.36 | 32.79 | 31.54 | 28.38 | 43.81 | 34.07 |
| n = 2 | 32.18 | 33.32 | 36.06 | 31.38 | 30.81 | 44.53 | 34.71 |
| n = 3 | 33.34 | 34.48 | 35.28 | 32.12 | 31.87 | 43.11 | **35.03** |

- 迭代式训练n次时模型性能变化

- Learning to Explore Informative Trajectories and Samples for Embodied Perception. (Jing et al, ICRA 2023)

# ➢ 层级策略——Embodied Learning

TABLE II
EFFECTS OF SETTING DIFFERENT THRESHOLDS IN HARD SAMPLE SELECTION ON OBJECT DETECTION TASK.

| Method | Training Image | Chair | Couch | Potted Plant | Bed | Toilet | Tv | Average |
|---|---|---|---|---|---|---|---|---|
| $\delta$ = 0.1 | 20k | 33.57 | 34.36 | 32.79 | 31.54 | 28.38 | 43.81 | 34.07 |
| $\delta$ = 0.2 | 13k | 33.38 | 34.61 | 31.34 | 31.84 | 26.24 | 41.64 | 33.18 |
| $\delta$ = 0.3 | 9k | 32.79 | 35.03 | 31.10 | 31.81 | 22.83 | 41.11 | 32.44 |

- 不同阈值下困难样本对性能的影响

TABLE III
ABLATION STUDIES ON THE OBJECT DETECTION TASK. SDD AND SDU MEANS THE SEMANTIC DISTRIBUTION DISAGREEMENT REWARD AND THE SEMANTIC DISTRIBUTION UNCERTAINTY REWARD IN TRAJECTORY EXPLORATION, RESPECTIVELY. HSS MEANS THE HARD SAMPLE SELECTION. SC MEANS THE SEMANTIC CURIOSITY [12].

| Method | Chair | Couch | Potted Plant | Bed | Toilet | Tv | Average |
|---|---|---|---|---|---|---|---|
| Ours w/o SDD | 32.08 | 34.51 | 32.05 | 29.91 | 27.95 | 42.76 | 33.21 |
| Ours w/o SDU | 33.49 | 34.39 | 32.95 | 31.19 | 27.31 | 42.18 | 33.59 |
| Ours w/o HSS | 32.44 | 33.69 | 32.22 | 30.85 | 27.67 | 41.78 | 33.11 |
| SC + HSS | 33.87 | 33.52 | 32.69 | 31.08 | 27.93 | 41.08 | 33.36 |
| Ours | 33.57 | 34.36 | 32.79 | 31.54 | 28.38 | 43.81 | **34.07** |

- SDD：语义分布不一致
- SDU：语义分布不确定
- HSS：困难样本选择
- SC：语义好奇

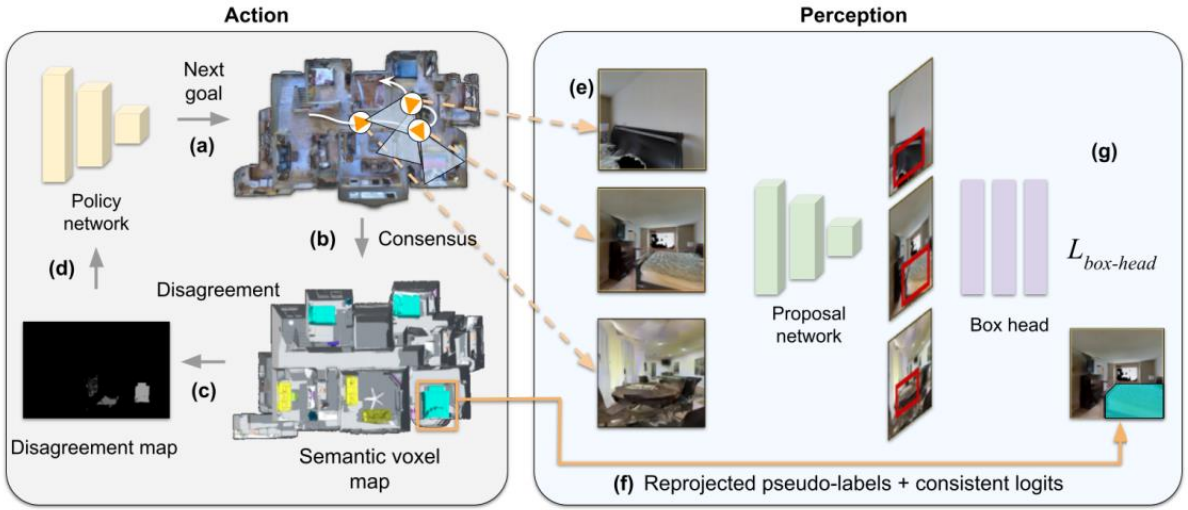

Fig. 4. Qualitative examples of instance segmentation by different models.

- 不同模型下的定性结果

Learning to Explore Informative Trajectories and Samples for Embodied Perception. (Jing et al, ICRA 2023)

## ➢ 层级策略——分歧探索

- 将目标检测结果投影到三维空间构建关于环境的语义一致体素图
- 将体素图进行垂直方向投影，计算每个cell的分歧分数（disagreement score），得到分岐图
- 利用分歧图学习探索策略
- 将构建的语义体素投影生成伪标签，实现self-training




- By assuming that pseudo-labels for the same object must be consistent across different views, we learn an exploration policy mining hard samples and we devise a novel mechanism for producing refined predictions from the consensus among observations.

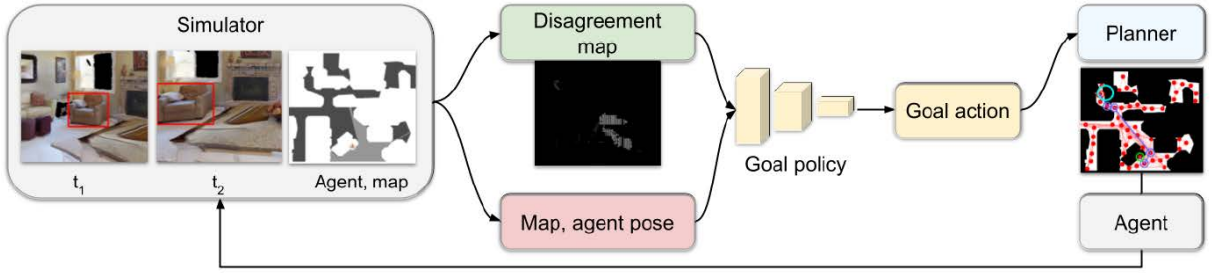



Fig. 3: Overview of policy training. Given all previous detections, a disagreement map is fed as input to the policy network together with the agent's pose in it. The policy predicts a new long-term goal in the map; a traditional planner extracts a series of sub-goals to the predicted goal and the agent moves to each sub-goal.

- Scarpellini G, Rosa S, Morerio P, et al. Look around and learn: self-improving object detection by exploration[J]. arXiv preprint arXiv:2302.03566, 2023.

## ➢ 弱监督学习——Embodied visual active learning

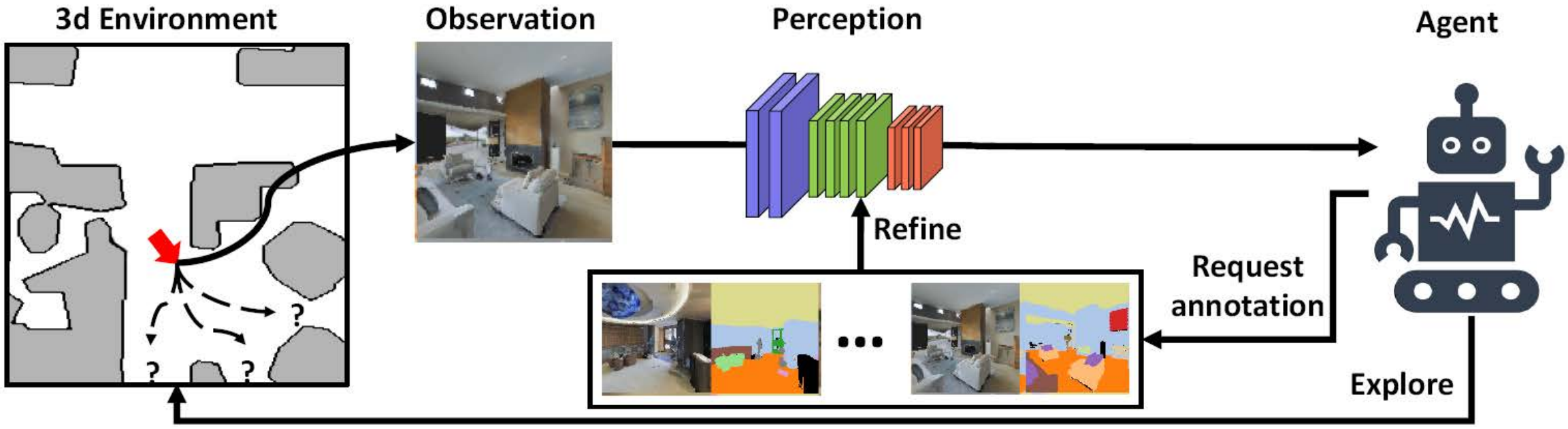



Figure 1: Embodied visual active learning. An agent in a 3d environment mus
in order to efficiently refine its visual perception. The navigation component ma
traditional active learning, where the data pool over which the agent queries anno
or pre-recorded video streams, is static and given.

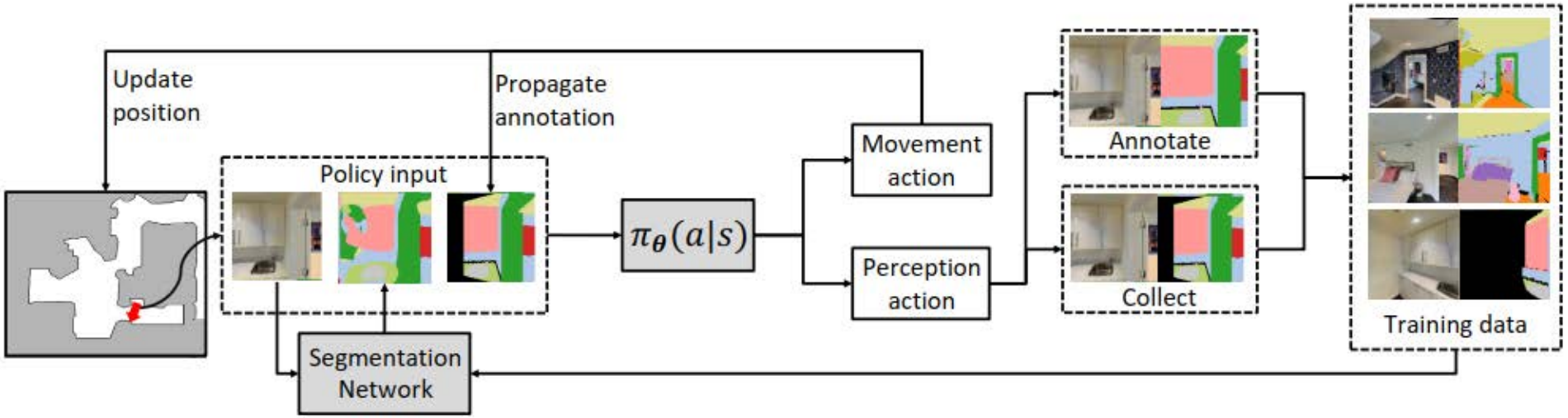
work by online retraining. The trainable method uses deep reinforcement learning with a reward function that balances two competing objectives: *task performance, represented as visual recognition accuracy*, which requires exploring the environment, and the necessary *amount of annotated data* requested during active exploration. We extensively evaluate

Nilsson D, Pirinen A, Gärtner E, et al. Embodied visual active learning for semantic segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(3): 2373-2383.

# 具身导航学习前沿



## ➢ 弱监督学习——Embodied visual active learning




**动作空间：**

- **movement：**前行、旋转
- **perception：**标注、采集

- 给定t时刻的ground-truth
- 智能体采取行动（movement和perception）得到t+1时刻观测，并使用光流法传播生成t+1时刻ground-truth
  1）采集：30% of the propagated labels are unknown ,采集t+1时刻传播生成的标注数据
  2）标注：85% are unknown,获取t+1时刻ground-truth数据
- rewards：性能提升+鼓励智能体探索新位置
- 感知：采集或标注行为后，执行refine

$$R_T = \text{mIoU}(\mathcal{S}_T, \mathcal{R}) - \text{mIoU}(\mathcal{S}_0, \mathcal{R}) \quad (1)$$

$$R_t^{ann} = \text{mIoU}(\mathcal{S}_t, \mathcal{R}) - \text{mIoU}(\mathcal{S}_{t-1}, \mathcal{R}) - \epsilon^{ann} \quad (2)$$

$$R_t^{col} = \text{mIoU}(\mathcal{S}_t, \mathcal{R}) - \text{mIoU}(\mathcal{S}_{t-1}, \mathcal{R}) \quad (3)$$

Nilsson D, Pirinen A, Gärtner E, et al. Embodied visual active learning for semantic segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(3): 2373-2383.

## ➢ 弱监督学习——Embodied visual active learning

- Random：采取随机行为
- Rotate：持续左转
- Bounce：一直直行直到遇见障碍物，随机旋转，继续直行
- Frontier exploration：构建地图指导智能体在参考视图半径内行动
- Space filler：基于最短空间填充曲线采取行动
- RL-agent：文中提出的方法

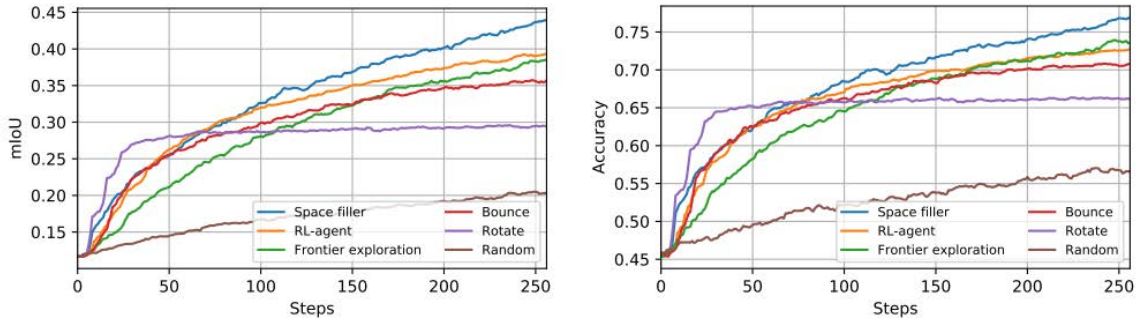



Figure 4: Mean segmentation accuracy and mIoU versus number of actions (steps), evaluated on the Matterport3D test scenes. The RL-agent was trained on 256-step episodes. This agent fairly quickly outperforms all other comparable pre-specified agents. *Rotate* is strong initially since it quickly gathers many annotations in a 360 degree arc, but is eventually outperformed by most other methods that move around in the houses. Frontier exploration yields similar accuracy as the RL-agent after about 170 steps, but uses significantly more annotations (cf. table 1) and assumes perfect pose and depth information. The space filler, which assumes full knowledge of the environment, yields the best results after about 100 steps.

相同step下模型性能对比

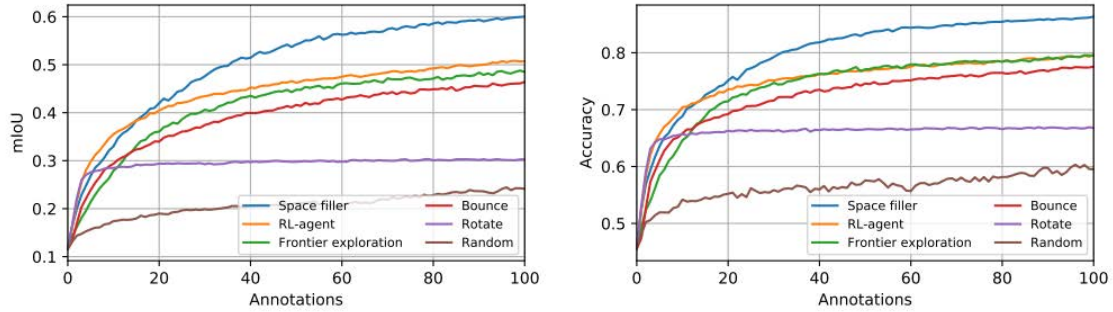



Figure 5: Mean segmentation accuracy and mIoU for a varying number of requested annotations evaluated on the Matterport3D test scenes. The RL-agent outperforms all comparable pre-specified methods (although frontier exploration matches it in accuracy after about 40 annotations), indicating that it has learnt an exploration policy which generalizes to novel scenes. The space filler, as expected, outperforms the RL-agent, except for less than 15 annotations. Thus the RL-agent is best before and around its training regime, where on average annotates 16.7 times per episode, cf. table 1.

标注相同数量数据时，模型性能对比

Nilsson D, Pirinen A, Gärtner E, et al. Embodied visual active learning for semantic segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(3): 2373-2383.

## ➢ 弱监督学习——Embodied visual active learning

Table 1: Comparison of different agents for a fixed episode length of 256 actions on the Matterport3D test scenes. The RL-agent gets higher mIoU using far fewer annotations than comparable pre-specified methods, implying that the RL-agent's policy selects more informative views to annotate.

| Method | mIoU | Acc | # Ann | # Coll |
|---|---|---|---|---|
| Space filler | 0.439 | 0.769 | 24.7 | 23.9 |
| RL-agent | 0.394 | 0.727 | 16.7 | 15.2 |
| Frontier exploration | 0.385 | 0.735 | 24.2 | 21.6 |
| Bounce | 0.357 | 0.708 | 29.6 | 26.0 |
| Rotate | 0.295 | 0.661 | 34.3 | 32.7 |
| Random | 0.204 | 0.566 | 29.1 | 19.5 |

RL-agent所需标注数据更少，一个episode平均标注16.7次

Table 2: Comparison of different agents for a fixed budget of 100 annotations on Matterport3D test scenes. The RL-agent gets a higher mIoU than comparable pre-specified agents, despite not being trained in this setting.

| Method | mIoU | Acc | # Steps | # Coll |
|---|---|---|---|---|
| Space filler | 0.600 | 0.863 | 1048 | 91 |
| RL-agent | 0.507 | 0.796 | 1541 | 94 |
| Frontier exploration | 0.485 | 0.796 | 998 | 84 |
| Bounce | 0.464 | 0.776 | 861 | 87 |
| Rotate | 0.303 | 0.668 | 752 | 96 |
| Random | 0.242 | 0.595 | 910 | 64 |

标注数据数量相同时，模型性能和所需step比较

Table 3: Ablation study of different RL-based model variants for 256-step episodes on the validation set. The full RL-agent outperforms all ablated models at a comparable or lower number of requested annotations.

| Variant | mIoU | Acc | # Ann | # Coll |
|---|---|---|---|---|
| Full model | 0.427 | 0.732 | 16.4 | 16.4 |
| No collect nor $\boldsymbol{P}_t$ | 0.415 | 0.727 | 17.9 | 0.0 |
| Only exploration | 0.411 | 0.727 | 16.1 | 14.4 |
| $R_t^{exp} = 0$ | 0.401 | 0.719 | 17.7 | 47.4 |
| No $\phi_{img}$ | 0.378 | 0.696 | 14.3 | 3.8 |
| No ResNet | 0.375 | 0.705 | 23.3 | 0.3 |

Full model性能最好，且要求标记的次数较低

Nilsson D, Pirinen A, Gärtner E, et al. Embodied visual active learning for semantic segmentation[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2021, 35(3): 2373-2383.

## ➢ 弱监督学习—— Embodied learning for lifelong visual perception

- 前一篇工作：

1）强化学习：基于模型性能提升多少设计奖励函数
2）action：基于光流法生成伪标签，根据伪标签的可信度判断是否采取annotate和collect

- 本文工作：

1）强化学习：模型提升（采取annotate时计算性能提升，其余时刻为0）
鼓励探索（选取frontier point作为长期目标，Dijkstra生成短期目标，计算到达短期目标距离）
2）action：annotate（语义分割准确率低于阈值时；固定频率（每20步）标注一次）

episodic：episode开始时重置分割网络
lifelong：保持前一个episode的网络参数

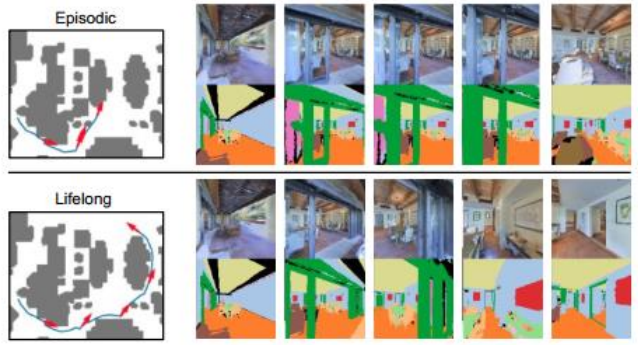



Figure 4: The first five requested annotations when evaluating the RL-agent in the episodic versus lifelong setup. The agent requests annotations at a sparser rate when evaluated lifelong in this scene. For more qualitative examples, please see the appendix.

| Model | Setup | $\Delta A$ / annot | $\Delta A$ / step | mIoU (1-50) | mIoU (51-100) |
|---|---|---|---|---|---|
| Accuracy oracle | Episodic | 1.162 | 0.056 | 0.306 | 0.429 |
| | Lifelong | 1.256 (+0.094) | 0.056 | 0.340 (+0.034) | 0.448 (+0.019) |
| RL-agent | Episodic | 0.876 | 0.057 | 0.286 | 0.397 |
| | Lifelong | 1.070 (+0.194) | 0.059 | 0.324 (+0.038) | 0.407 (+0.010) |
| Uniform | Episodic | 1.086 | 0.057 | 0.297 | 0.402 |
| | Lifelong | 1.086 ($\pm$0) | 0.057 | 0.311 (+0.014) | 0.409 (+0.007) |
| Random | Episodic | 1.114 | 0.057 | 0.296 | 0.395 |
| | Lifelong | 1.063 (-0.051) | 0.057 | 0.303 (+0.007) | 0.393 (-0.002) |

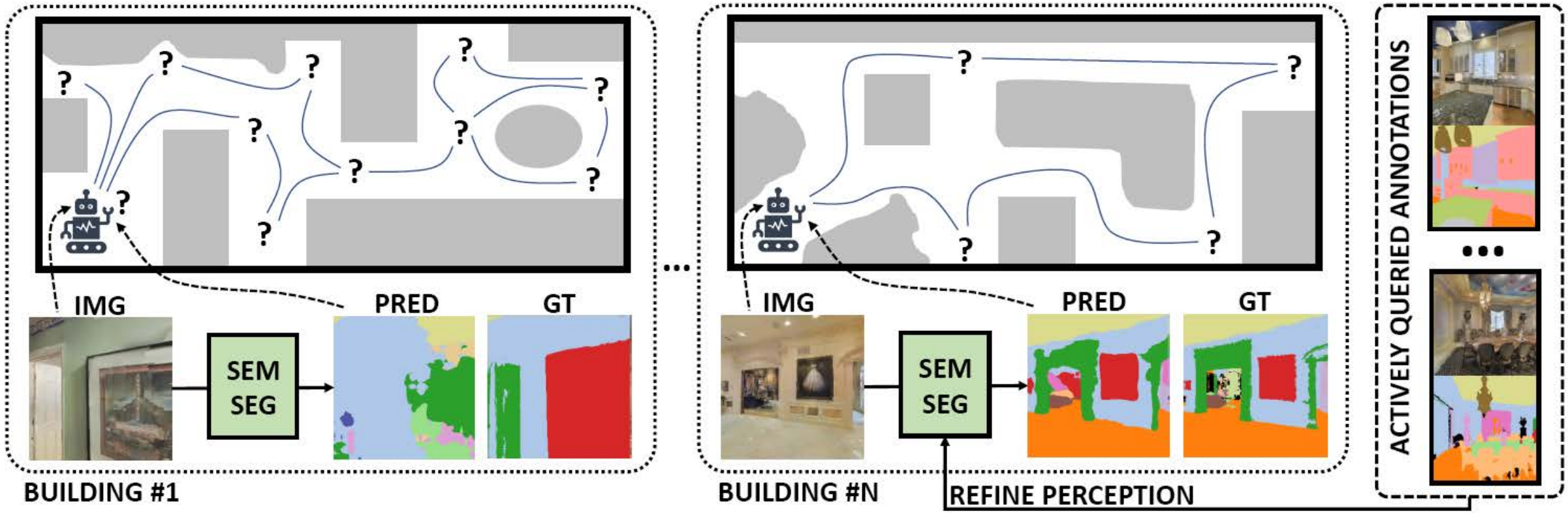
Table 1: Comparison between episodic and lifelong setups for the accuracy oracle, RL-agent and baselines (uniform, random). We see that the average area explored since the last annotation ($\Delta A$/ annot, $m^2$) when requesting a new annotation is higher on average in the lifelong compared to the episodic setup for both the accuracy oracle and the RL-agent, indicating adaptive behavior depending on the current segmentation accuracy. This is not the case for the two simpler heuristics. The average navigable area added to the mapper per step ($\Delta A$/ step, $m^2$) is identical for episodic and lifelong for the accuracy oracle and the two baselines, as is expected since they explore identically, but for the RL-agent we see that in the lifelong setup it on average explores faster. Finally, we see that the improvements of the average mIoU after annotations 1 - 50 and 51 - 100 in the lifelong setup compared to the episodic are larger for the accuracy oracle and RL-agent than for the baselines, and especially for the early part of episodes (annotations 1 - 50). This indicates that both the RL-agent and the accuracy oracle adapt their annotation strategies when evaluated in the lifelong setup.

Nilsson D, Pirinen A, Gärtner E, et al. Embodied Learning for Lifelong Visual Perception[J]. arXiv preprint arXiv:2112.14084, 2021.

# 具身导航学习前沿



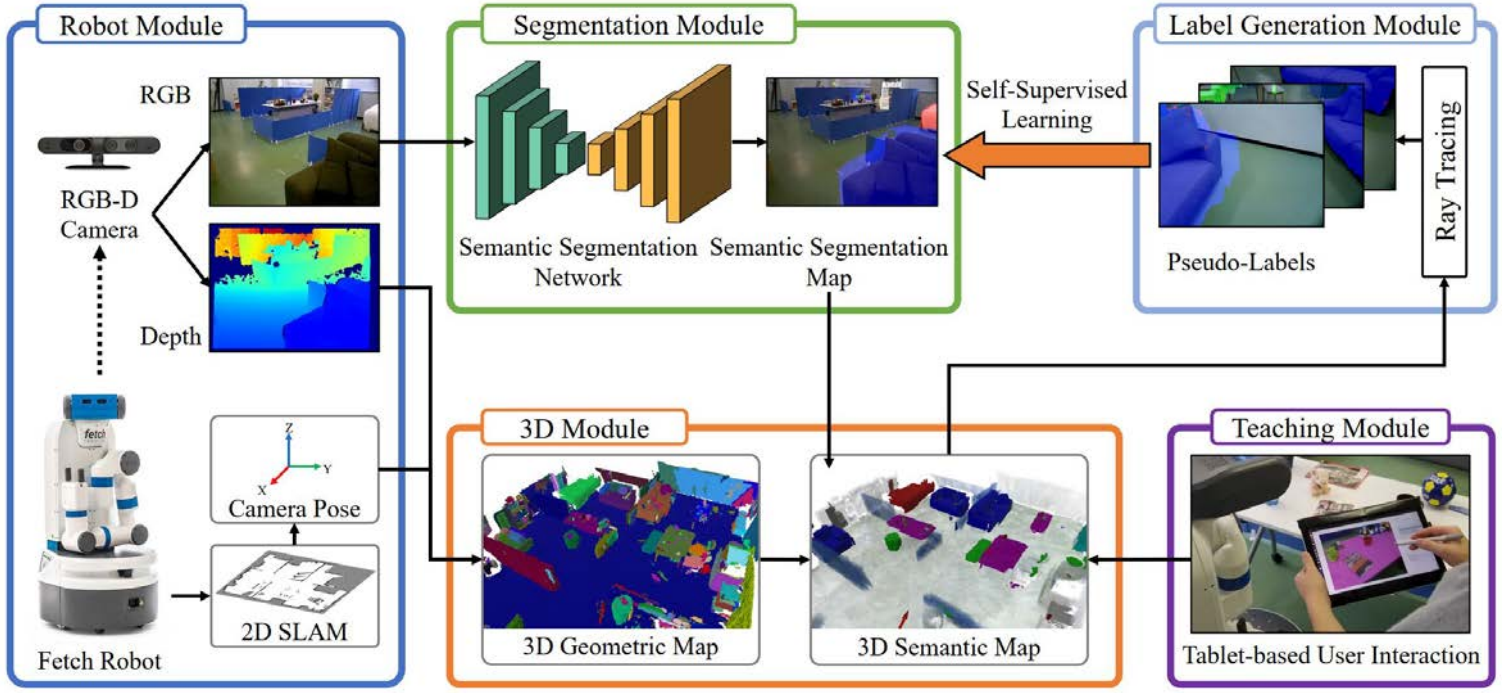
## ➢ 弱监督学习—— Embodied learning for lifelong visual perception




Nilsson D, Pirinen A, Gärtner E, et al. Embodied Learning for Lifelong Visual Perception[J]. arXiv preprint arXiv:2112.14084, 2021.

## ➢ 弱监督学习——交互

- 目前的语义分割方法需要大型数据集和昂贵的注释来实现准确的推理，没有大量额外的图像和注释，就无法处理所有可能的对象或环境变化。因此，作者提出了一种用于语义分割的学习系统，该系统将3D语义映射与自主移动机器人和用户之间的交互相结合。利用自主移动机器人自主构建3D语义图，通过交互获得少量额外的粗略注释，可以更准确地预测新的对象类别。




- Kanechika A, El Hafi L, Taniguchi A, et al. Interactive Learning System for 3D Semantic Segmentation with Autonomous Mobile Robots[C]//2024 IEEE/SICE International Symposium on System Integration (SII). IEEE, 2024: 1274-1281.

# 具身导航学习前沿

重要！

## ➢ 自监督学习——SEAL

- It utilizes perception models trained on internet images to learn an active exploration policy.
- The observations gathered by this exploration policy are labelled using 3D consistency and used to improve the perception model.

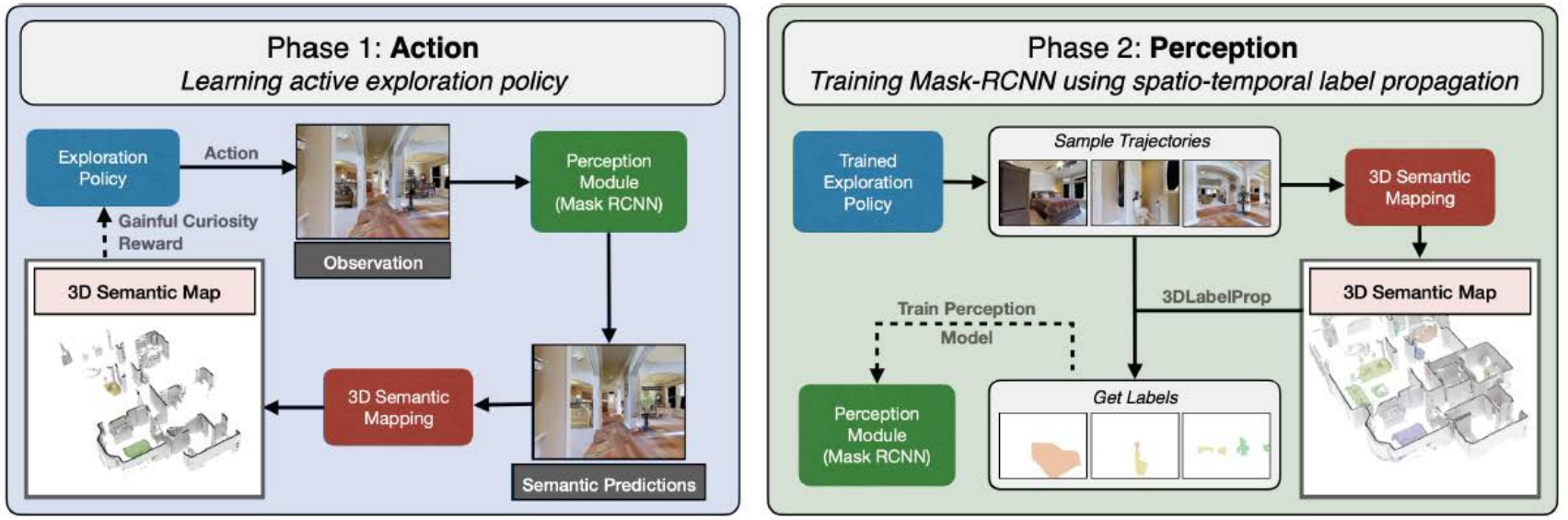



**Figure 1: Self-supervised Embodied Active Learning.** Our framework called Self-supervised Embodied Active Learning (SEAL) consists of two phases, *Action*, where we learn an active exploration policy, and *Perception*, where we train the Perception Model on data gathered using the exploration policy and labels obtained using spatio-temporal label propagation. Both action and perception are learnt in a completely self-supervised manner without requiring access to the ground-truth semantic annotations or map information.

- Seal: Self-supervised embodied active learning using exploration and 3d consistency, NeurIPS, 2021

## ➢ 自监督学习——SEAL

**Stage 1: Learning Action:** The Gainful Curiosity reward is then defined to be the number of voxels in the 3D Semantic Map having greater than ˆs score for at least one semantic category.
This reward encourages the agent to find new objects and keep looking at the object from different viewpoints until it gets a highly confident prediction for the object from at least one viewpoint.

奖励：3D语义图中，至少属于一个语义类别（高置信度）的体素的数量

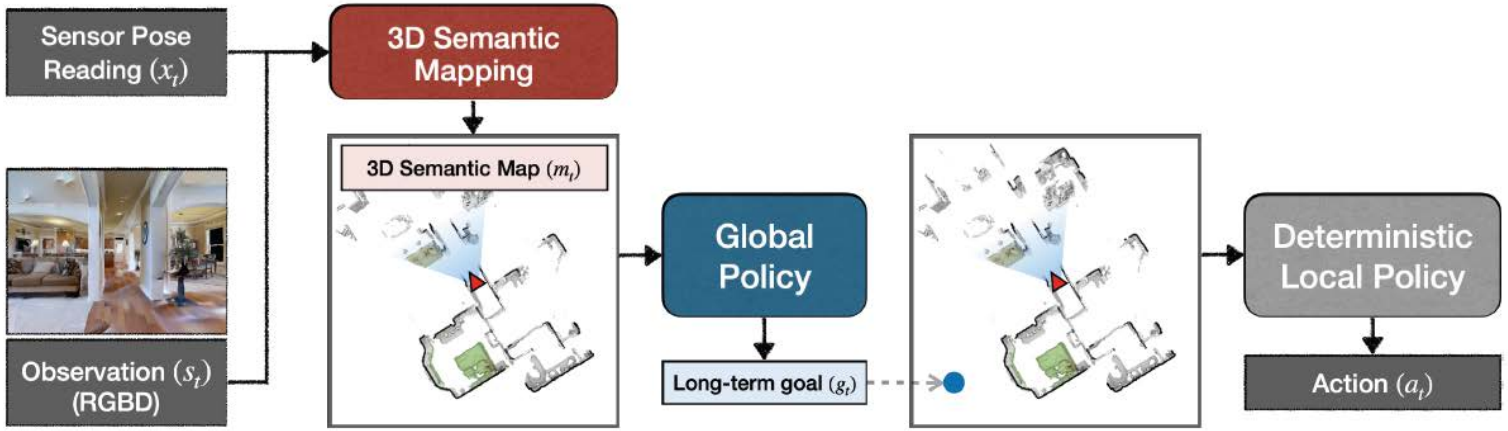



**Figure 3: Learning Action using Gainful Curiosity.** We use a modular architecture for the Gainful Curiosity Policy. The 3D Semantic Mapping module is used to construct and update the map, $m_t$, at each time step $t$. The Global Policy is used to sample long-term goal ($g_t$). A deterministic Local Policy is used to plan a path to the long-term goal and take low-level navigational actions. The Gainful Curiosity intrinsic motivation reward is computed using the 3D Semantic Map.

- Seal: Self-supervised embodied active learning using exploration and 3d consistency, NeurIPS, 2021

## ➢ 自监督学习——SEAL

**Stage 2: Learning Perception:** After labeling each voxel in the map, we find the set of connected voxels labeled with the same category to find object instances. We fill small holes (< 0.25m3) in object instances and remove small objects (< 0.025m3) to get the final labeled 3D semantic map. The instance label for each pixel in each observation in the trajectory is then obtained using ray-tracing in the labeled 3D map based on the agent's pose.

3D语义图反向投影至2D空间生成伪标签

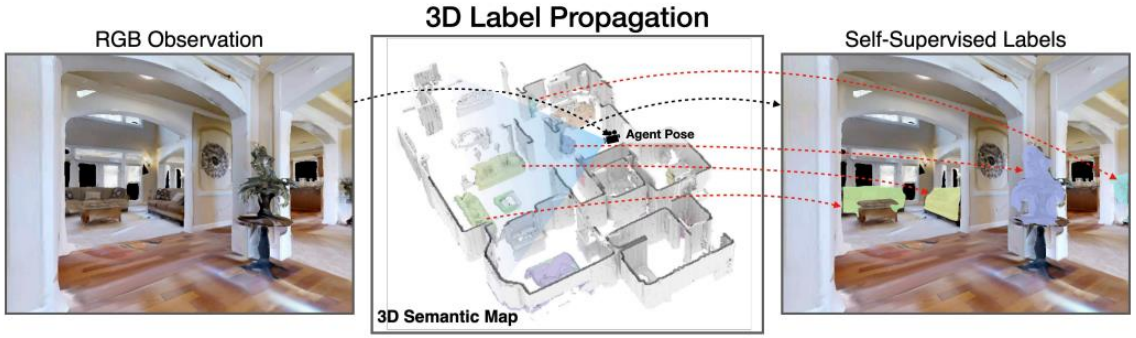
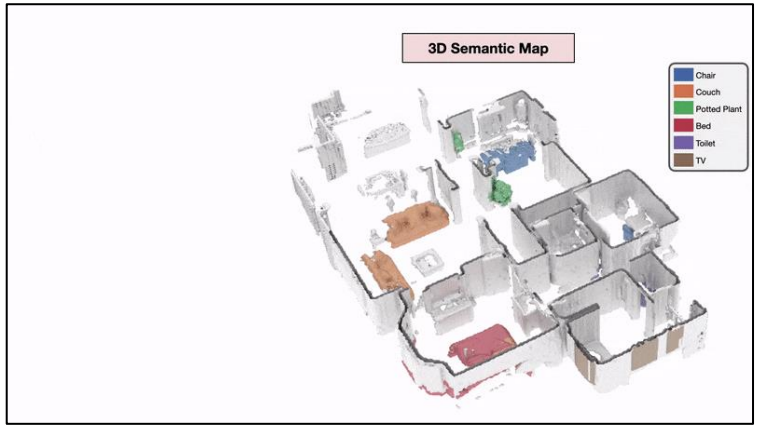



**Figure 4: Learning Perception using 3DLabelProp.** The agent trajectory is to used to create a semantic 3D map of the environment. The map is labelled in a self-supervised manner using 3D consistency. The label for each pixel in the agent trajectory is obtained using ray-tracing in the labeled map based on the agent's pose.




The semantic map is used to compute an intrinsic motivation reward for training the exploration policy and for labelling the agent observations using spatio-temporal 3D consistency and label propagation.

- Seal: Self-supervised embodied active learning using exploration and 3d consistency, NeurIPS, 2021

# ➢ 自监督学习——SEAL

- Random：随机采取行动
- Frontier Exploration：基于边界的探索
  启发式导航到最近的未探索点来探索看不见的环境
- Active Neural SLAM：最大化探索区域
- Self-Training：使用探索策略采集的数据训练Mask-RCNN
- Optical Flow：使用光流法传播标签。

**Table 1: Results.** Performance of all the baselines as compared to the proposed SEAL framework for both Generatlization and Specialization settings. We report bounding box and mask AP50 scores for Object Detection and Instance Segmentation.

| | Generalization | | Specialization | |
|---|---|---|---|---|
| Method | Object Detection | Instance Segmentation | Object Detection | Instance Segmentation |
| Pretrained Mask-RCNN | 34.82 | 32.54 | 34.82 | 32.54 |
| Random Policy + Self-training [52] | 33.41 | 31.89 | 34.11 | 31.23 |
| Random Policy + Optical Flow [22] | 33.97 | 32.34 | 34.33 | 32.22 |
| Frontier Exploration [53] + Self-training [52] | 33.78 | 32.45 | 33.29 | 32.50 |
| Frontier Exploration [53] + Optical Flow [22] | 35.22 | 31.90 | 34.19 | 32.12 |
| Active Neural SLAM [10] + Self-training [52] | 34.35 | 31.20 | 34.84 | 32.44 |
| Active Neural SLAM [10] + Optical Flow [22] | 35.85 | 32.22 | 35.90 | 33.12 |
| Semantic Curiosity [11] + Self-training [52] | 35.04 | 32.19 | 35.23 | 32.88 |
| Semantic Curiosity [11] + Optical Flow [22] | 35.61 | 32.57 | 35.71 | 33.29 |
| SEAL | **40.02** | **36.23** | **41.23** | **37.28** |

- Generalizaiotn：

训练：在一组训练环境中训练 1000 万帧；

测试：直接在一组没见过的测试环境中进行测试

- Specialization：

测试：允许智能体在每个测试环境中探索 300个步长

**Table 3: Weak Supervision Results.** Performance of a Mask-RCNN naively fine-tuned with a few frames of labelled data as compared using the proposed SEAL framework for label propagation. We report bounding box and mask AP50 scores for Object Detection and Instance Segmentation.

| | Fine-tuning Mask-RCNN | | SEAL | |
|---|---|---|---|---|
| Num labels | Object Detection | Instance Segmentation | Object Detection | Instance Segmentation |
| 0 | 34.82 | 32.54 | 41.23 | 37.28 |
| 5 | 34.22 | 31.67 | 41.44 | 37.65 |
| 10 | 35.14 | 32.52 | 42.63 | 38.48 |

- weak supervison：

人工标记某些帧，替换感知模块中的预测结果

- Num lables：

Mask-RCNN naively fine-tuned
with a few frames (k = 0, 5, 10) of labeled data

- Seal: Self-supervised embodied active learning using exploration and 3d consistency, NeurIPS, 2021

## ➢ 自监督学习——Multi-View

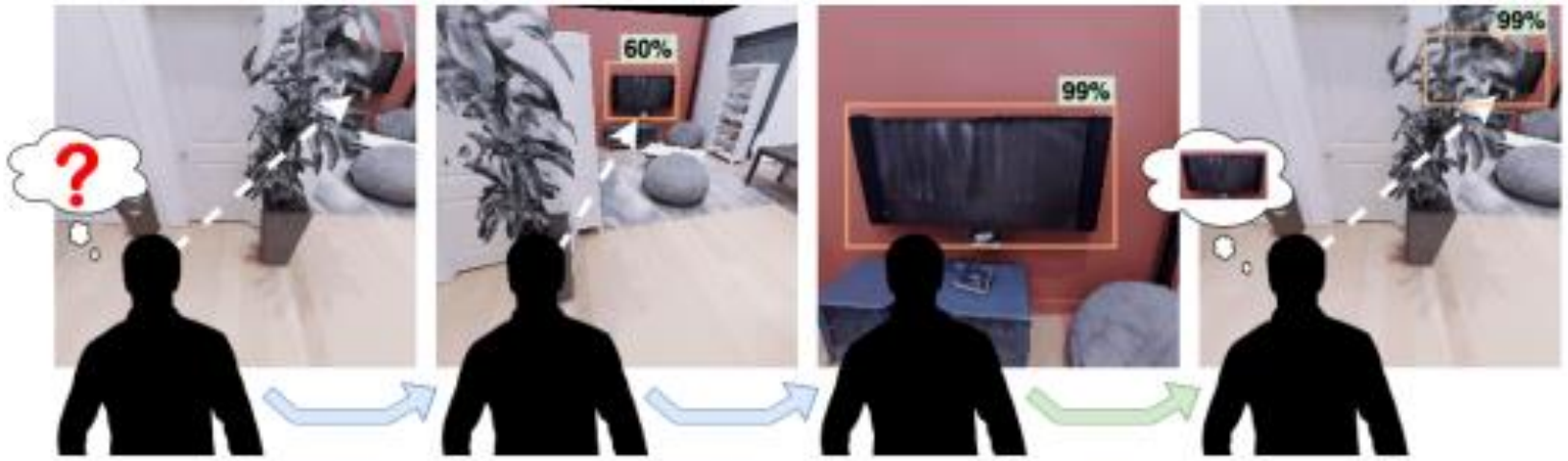



Figure 1. **Improving object recognition by moving.** An agent is viewing an object from an occluded, unfamiliar viewpoint. By moving to less occluded, more familiar viewpoints of the object (blue arrow), the agent can use the familiar viewpoints to self-supervise the previously unfamiliar viewpoints (green arrow).

- Move to see better: Self-improving embodied object detection[J]. arXiv preprint arXiv:2012.00057, 2020.

## ➢ 自监督学习——Multi-View

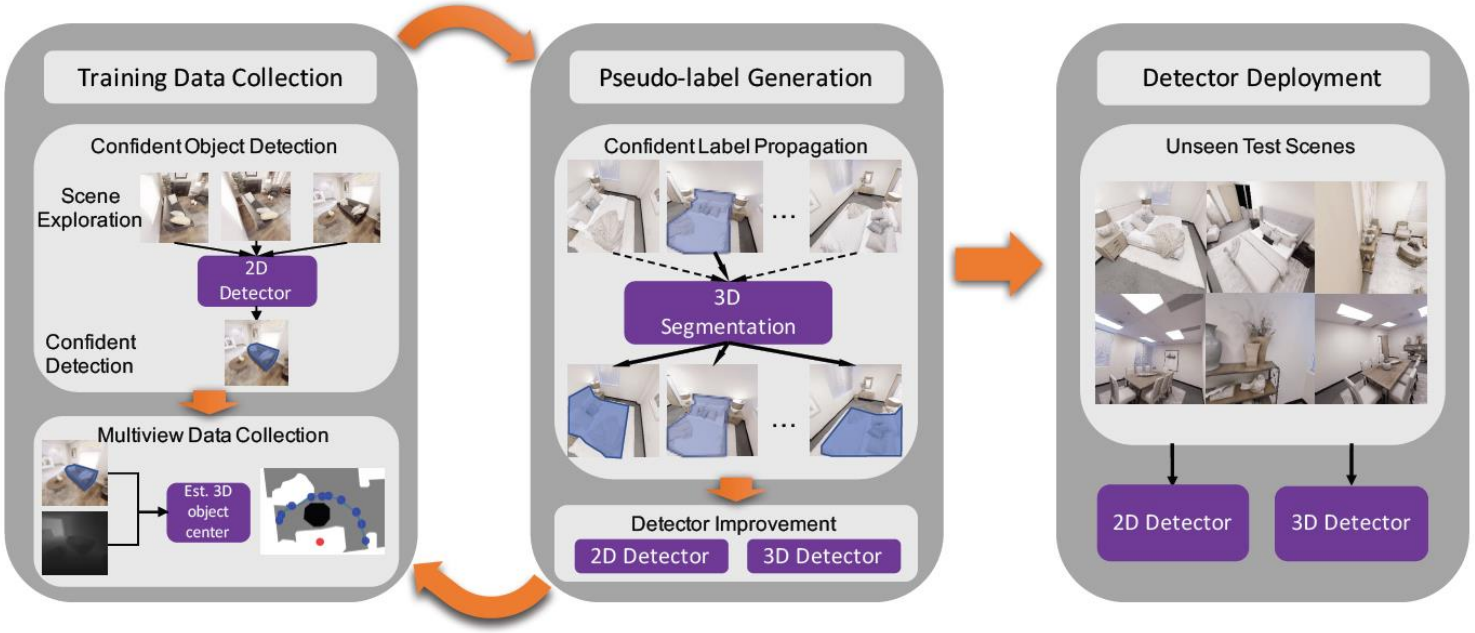



Figure 2. **Seeing by Moving (SbM).** We use confident detections of a pre-trained 2D object detector to guide self-supervised multi-view data collection and pseudo-label generation. Our 3D segmentation module segment the detected objects in 3D using aggregated RGB-D images. The 2D and 3D detectors fine-tuned on the pseudo-labels perform better on unseen test scenes.

(1) Data collection: randomly move in the environment to collect observations and occasional high-confidence detections, then plan paths to collect diverse posed RGB-D images of the detected objects;
(2) 3D segmentation: segment the detected objects in 3D using aggregated RGB-D images, and then re-project that segmentation to form 2D pseudolabels in all views;
(3) Detector improvement: fine-tune the pre-trained detector on the pseudo-labels.

- Move to see better: Self-improving embodied object detection[J]. arXiv preprint arXiv:2012.00057, 2020.

## ➢ 自监督学习——Multi-View

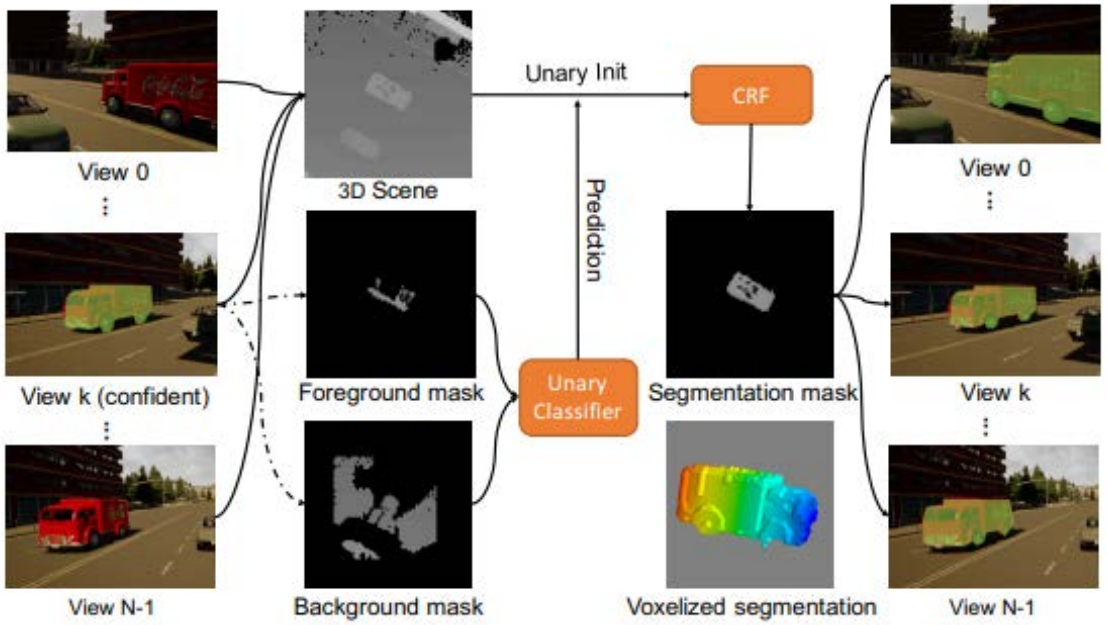



Figure 3. **3D Segmentation.** All images are unprojected into 3D space using depth and pose, as well as the segmentation mask from confident view $k$. We sample foreground and background to train a unary classifier, whose outputs are used to initialize the unary potentials of the CRF model. The final 3D segmentation is then reprojected to all views to obtain pseudo-labels.

**2D Detection Training** Since our estimated segmentations are in 3D, and since we have the source views and poses, we are able to create 2D pseudo-labels as well. We produce 2D pseudo-labels by re-projecting $P_{seg}$ to all views. For a point $(X, Y, Z)^T$ in $P_{seg}$, we can get its 2D pixel coordinate in the $i$-th frame with:

$$(u, v)^T = \mathbf{G}_i \left( f_x \frac{X}{Z} + c_x, f_y \frac{Y}{Z} + c_y \right) \tag{4}$$

The reprojected points in $P_{seg}$ are sparse in 2D, so we fit a concave hull to convert them into a connected binary mask. Our experiments show that these pseudo-labels are lower quality than ground truth labels, but still provide a valuable boost to a pre-trained detector.

- Move to see better: Self-improving embodied object detection[J]. arXiv preprint arXiv:2012.00057, 2020.

## ➢ 自监督学习——Multi-View

| Method | mAP@0.25 |
|---|---|
| LDLS [46] | 44.03 |
| SbM Self-Sup. F-PointNet (ours) | **83.87** |
| Supervised F-PointNet | 85.06 |

Table 6. **Fine-tuning with SbM labels outperforms self-supervised LDLS.** 3D object detection performance of LDLS [46], frustum PointNet trained on SbM segmentations, and GT-trained frustum PointNet on the CARLA test set.

- SbM性能超出LDLS
- SbM接近全监督性能

| Method | mAP@0.5 | mAP@0.3 |
|---|---|---|
| Pre-trained | 21.36 | 26.14 |
| SbM Labels w/ noise (ours) | **23.15** | **33.68** |
| SbM Labels w/o noise (ours) | 26.20 | 38.12 |

Table 4. **Pseudo-label accuracy with pose noise in the Replica training set.** We show that actuation noise weakens the data collection, yet our method is still able to produce pseudo-labels that are better than the pre-trained detectors' predictions.

- 在噪声下SbM仍能提升模型性能

| mAP@IoU | Method Name | Cushion | Nightstand | Shelf | Beanbag | Avg |
|---|---|---|---|---|---|---|
| 0.5 | SbM-ws Trained | **93.62** | **81.25** | **24.38** | 82.18 | **70.35** |
| | Limited GT Trained | 87.24 | 79.79 | 16.40 | **88.77** | 68.04 |
| 0.3 | SbM-ws Trained | **94.23** | **81.25** | **25.61** | 82.18 | **70.81** |
| | Limited GT Trained | 87.24 | 79.79 | 16.40 | **88.77** | 68.04 |

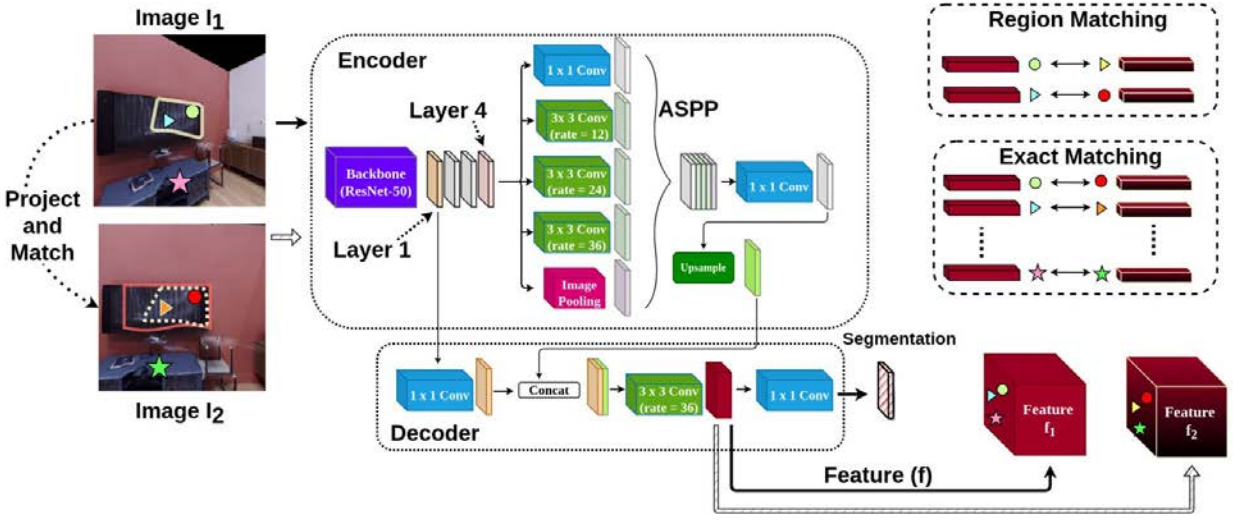
Table 5. **SbM-ws labels can be used to train detectors on novel categories.** We compare the performance of the detector trained on labels produced by SbM-ws with the detector trained on ground truth labels. The results show that the detector trained on SbM-ws labels outperforms the on trained on limited ground truth labels.

- 在伪标签上微调的检测器与在提供的相同真值标签上微调的检测器的性能对比


- Move to see better: Self-improving embodied object detection[J]. arXiv preprint arXiv:2012.00057, 2020.

## ➢ 自监督学习——时空一致性

- 利用移动机器人在新环境中移动和多视图观测的能力，即从不同的角度和位置捕获环境的视图。利用空间一致性（同一环境中的不同视图之间，相同物体或场景的像素应该具有相似的语义标签）和时间一致性（随着时间的推移，同一物体或场景的像素的语义标签应该保持一致）线索，使用高效区域匹配方法将不同视图进行像素关联。此外，还提出了一种对比学习的变体，使得模型可以在没有标签数据的情况下，通过自监督的方式进行预训练，并在微调阶段使用少量的标签数据，达到了很好的性能。




$$T_{1\to 2}(I_1) = \{K(T_2{}^{-1}(T_1(K^{-1}(\mathbf{X})) \quad \forall \mathbf{X} \in I_1\}$$

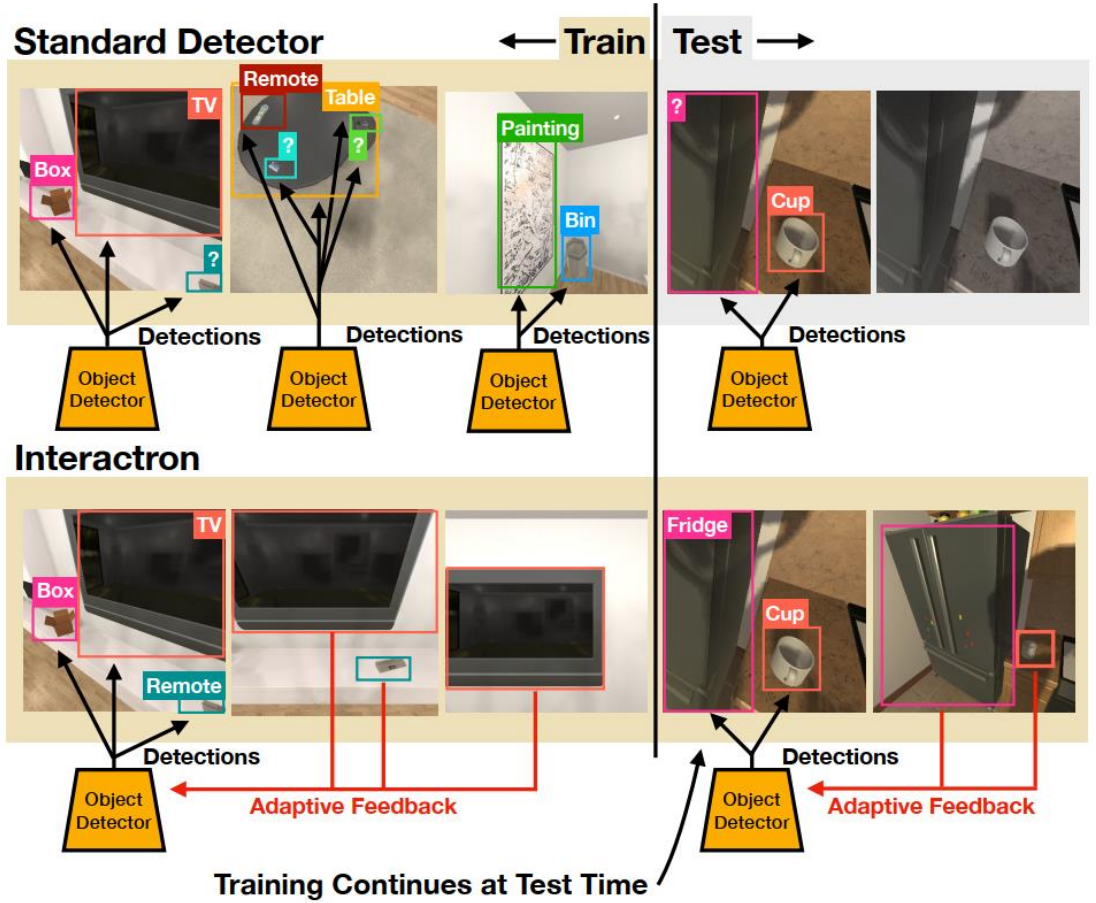
- Shrestha S, Li Y, Košecká J. Self-supervised pre-training for semantic segmentation in an indoor scene[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2024: 625-635.

# 具身导航学习前沿

重要！

## ➢ 自监督学习——**Interactron**: Embodied Adaptive Learning

- we propose Interactron, a method for adaptive object detection in an interactive setting, where the goal is to perform object detection in images observed by an embodied agent navigating in different environments.
- Our idea is to continue training during inference and adapt the model at test time without any explicit supervision via interacting with the environment.




- Kotar K, Mottaghi R. Interactron: Embodied adaptive object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14860-14869.

## ➢ 自监督学习——**Interactron**: Embodied Adaptive Learning

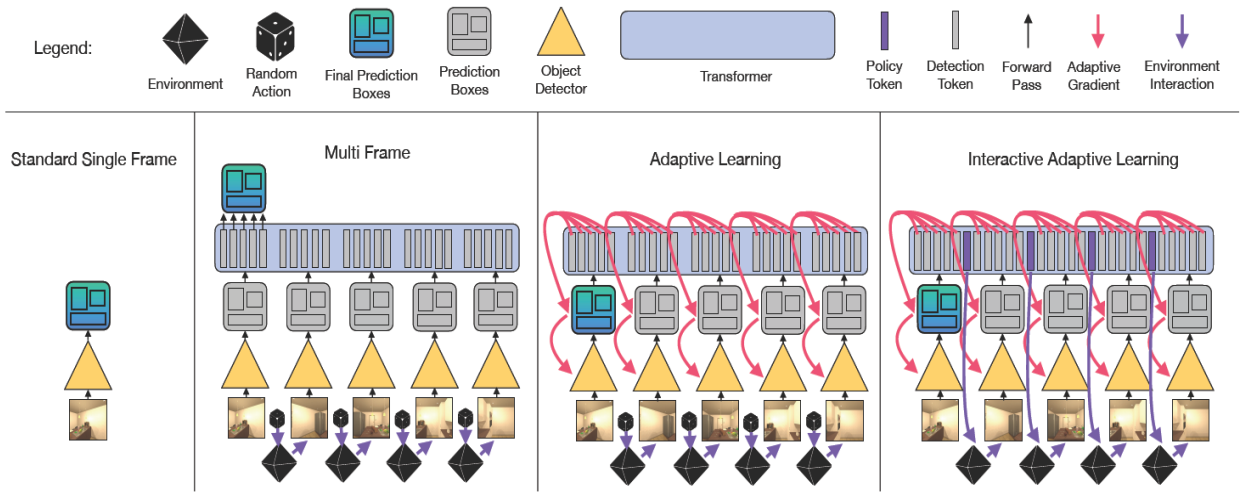



Figure 2. The architecture of the four main models presented in Section 4. A single frame baseline, which is just an off-the-shelf object detector, a multi-frame baseline which includes an inter-frame fusion Transformer and two INTERACTRON models (w/ and w/o a learned policy), which use a Transformer to learn a self-supervised loss function.

**Algorithm 1** Training $(d(\mathcal{T}, \mathbf{S}_{\text{train}}), \theta, \phi, \rho, \alpha, \beta_1, \beta_2, \beta_3, n)$

**while** not converge **do**
  **for** mini-batch of tasks $\tau_i \in d(\mathcal{T}, \mathbf{S}_{\text{train}})$ **do**
    $\theta_i \leftarrow \theta$
    $t \leftarrow 0$
    $\mathbf{F}_i \leftarrow [f_{S_i,p_i}]$
    **while** $t < n$ **do**
      Sample action $a$ from $\mathcal{P}^{\rho}_{int}(\mathbf{F}_i)$
      Take action $a$ and collected frame $f$
      $\mathbf{F}_i \leftarrow \mathbf{F}_i + [f]$
      $t \leftarrow t + 1$
    $\theta_i \leftarrow \theta_i - \alpha \nabla_{\theta_i} \mathcal{L}^{\phi}_{ada}(\theta_i, \mathbf{F}_i)$
  $\theta \leftarrow \theta - \beta_1 \sum_i \nabla_{\theta} \mathcal{L}_{det}(\theta_i, f_{S_i,p_i})$
  $\phi \leftarrow \phi - \beta_2 \sum_i \nabla_{\phi} \mathcal{L}_{det}(\theta_i, \mathbf{F}_i)$
  $\rho \leftarrow \rho - \beta_3 \sum_i \nabla_{\rho} \mathcal{L}_{pol}(\theta_i, \mathcal{P}^{\rho}_{int}(\mathbf{F}_i))$

- Kotar K, Mottaghi R. Interactron: Embodied adaptive object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14860-14869.

# 具身导航学习前沿



## ➢ 自监督学习——**Interactron**: Embodied Adaptive Learning

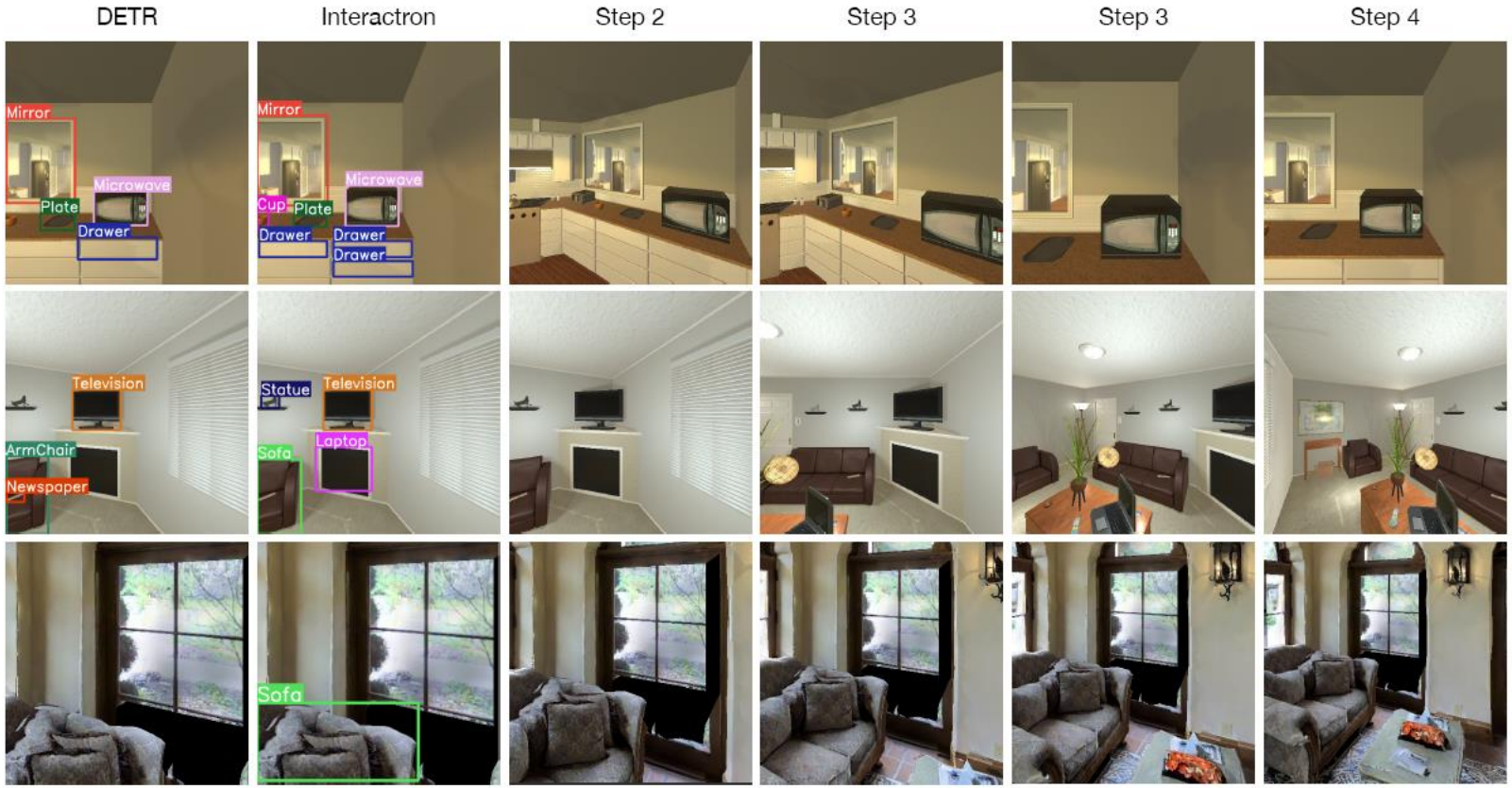



Figure 3. **Qualitative results.** in the AI2-iTHOR and Habitat environments. The first column displays the results produced by DETR [5], the 2nd column displays the results of INTERACTRON, and the subsequent columns depict the interactive steps that the model took. Note that for the result shown in the third row, INTERACTRON has never seen Habitat images during training.

- Kotar K, Mottaghi R. Interactron: Embodied adaptive object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2022: 14860-14869.

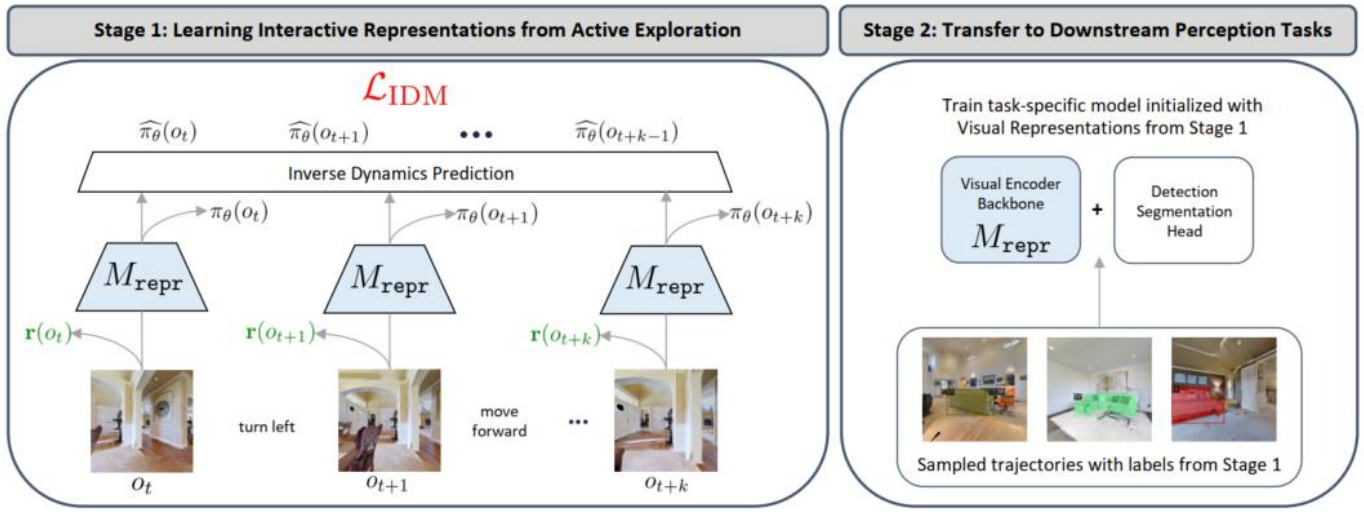
# 具身导航学习前沿

重要！

## ➢ 表示学习——动作

统一了两个任务

- 视觉表示学习：具身智能体的主动交互提供了图像外的重要信息——连续帧编码智能体的运动，可以训练视觉表示来捕捉帧之间的长视野动态信息。具体地，通过同时学习探索策略（目的是发现多样的新观测），以及预测给定一系列视觉观测时的多步动作，学习共享视觉表示
- 下游任务微调：使用第一阶段学到的视觉表示初始化模型，并根据下游任务，从第一阶段收集的数据中随机采样一部分数据微调预训练的网络




- Our framework consists of two stages. In Stage 1, the agent learns visual representations from interactions by actively exploring in environments from intrinsic motivation. Our visual representation learning approach directly considers embodied interactions by incorporating supervisions from action information. In Stage 2, we utilize both learned visual representations and label a small random subset of samples from explored trajectories to train downstream perception models.

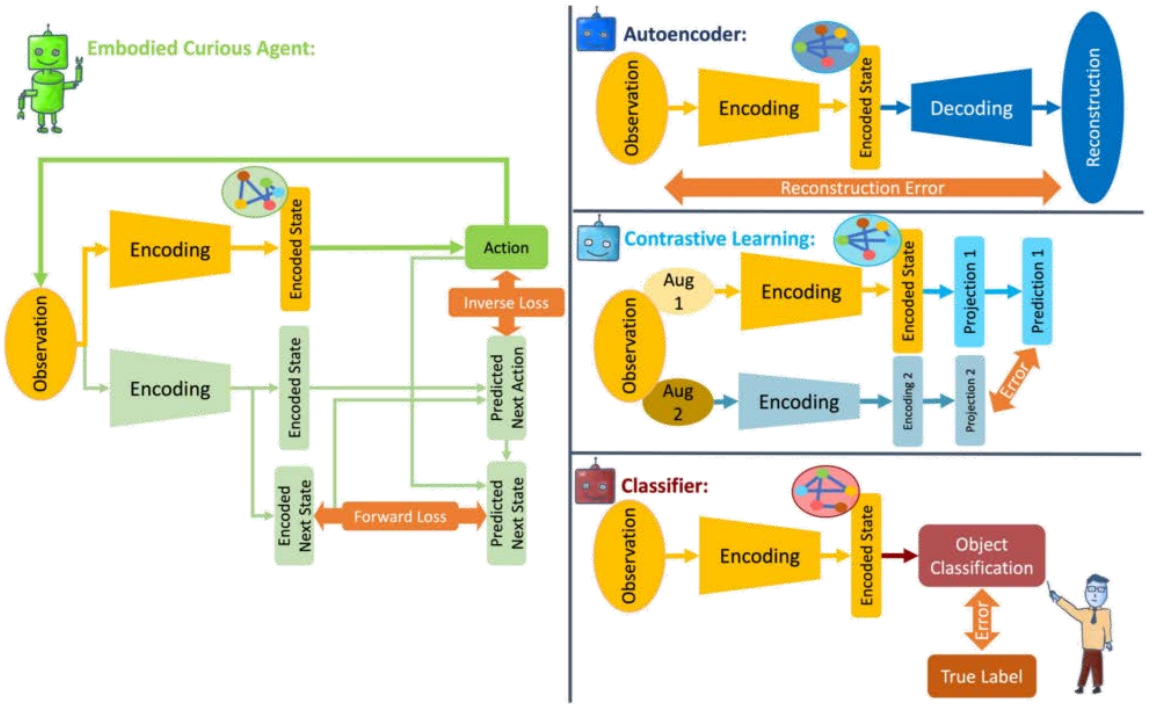

- Liang X, Han A, Yan W, et al. ALP: Action-Aware Embodied Learning for Perception[J]. arXiv preprint arXiv:2306.10190, 2023.

# 具身导航学习前沿

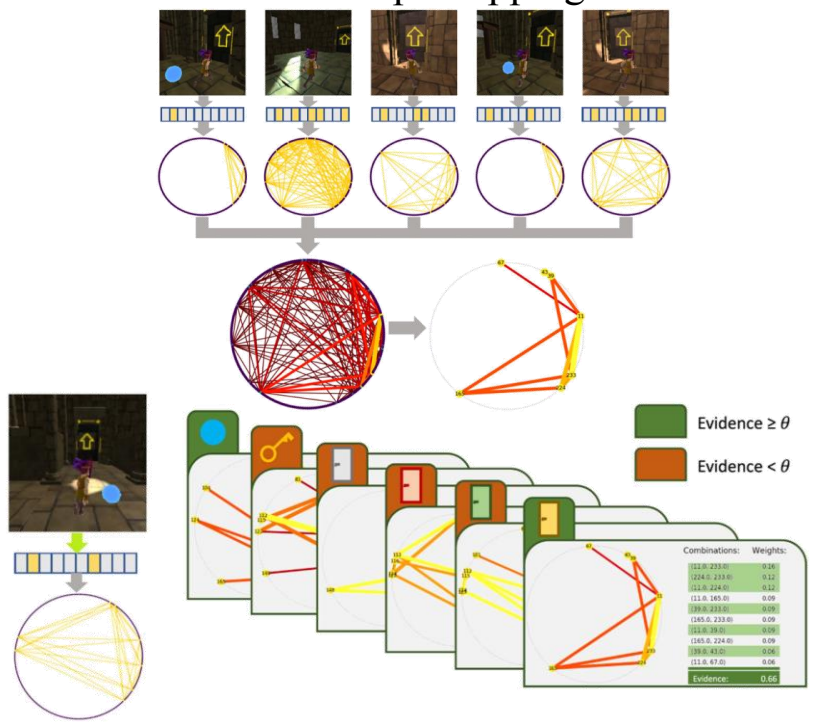
重要！

## ➢ 表示学习——少样本学习

- 大多数目标检测都需要全监督的数据进行学习。作者利用智能体与环境的交互来自主学习视觉表示。具体地，作者在仿真环境中通过自监督和好奇驱动的探索机制来学习表示，此外，还提出了快速概念映射（Fast concept mapping），利用在自监督交互学习过程中学到的表示，快速关联语义概念，因此只需要标记少量样本即可识别出多个物体。




Fast concept mapping

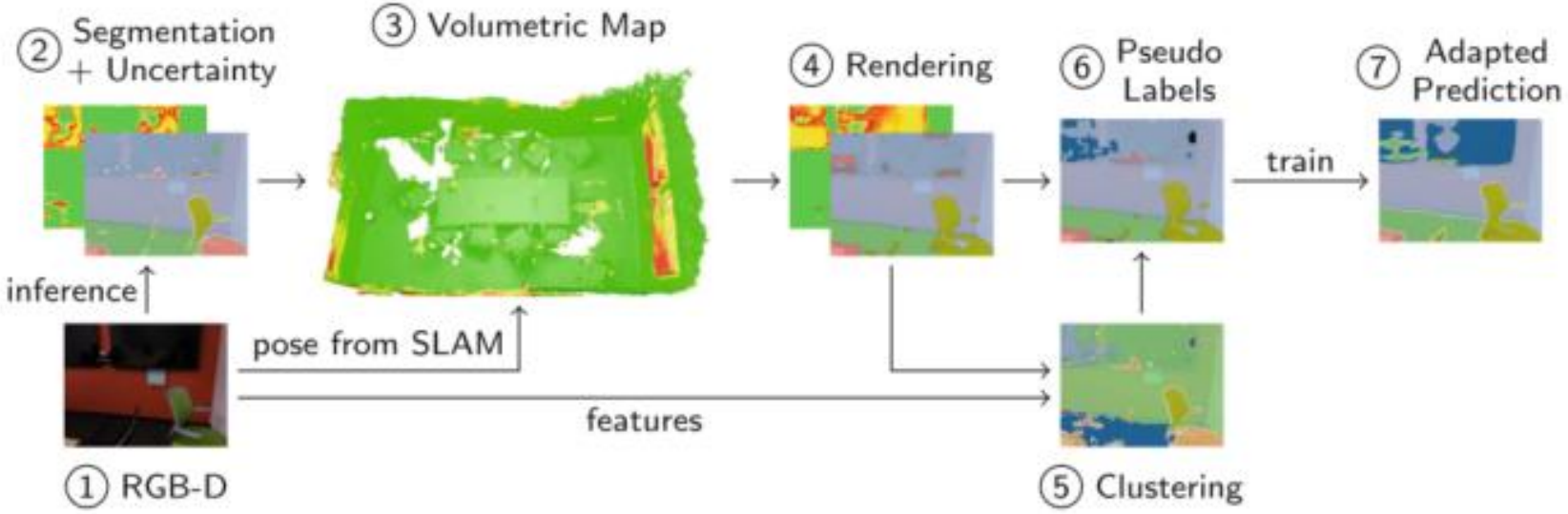



- Clay V, Pipa G, Kühnberger K U, et al. Development of Few-Shot Learning Capabilities in Artificial Neural Networks When Learning Through Self-Supervised Interaction[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2023.

## ➢ 增量学习——SCIM

作者提出了一个框架，该框架通过融合多种模态的观测信息，通过聚类、推理和映射的方法来实现自主发现新的语义类别并提高对已知类别的准确性

- 多模态观测信息：RGB-D、分割模型的预测及不确定性、深度特征、环境的体素图以及从该图中提取的几何特征。
- 聚类优化：观测用图结构表示，节点为像素，边为观测信息间的距离关系。通过图聚类算法进行聚类。
- 自监督学习信号：利用映射和聚类生成伪标签，以更新语义分割模型。




- Blum H, Müller M G, Gawel A, et al. SCIM: Simultaneous Clustering, Inference, and Mapping for Open-World Semantic Scene Understanding[C]//The International Symposium of Robotics Research. Cham: Springer Nature Switzerland, 2022: 119-135.

## ➢ 检测器提升

- We propose to use an automatic annotation procedure, which leverages on human-robot interaction and depth-based segmentation, for the acquisition and labeling of training examples.
- We fine-tune the Faster R-CNN network with these data acquired by the robot autonomously.

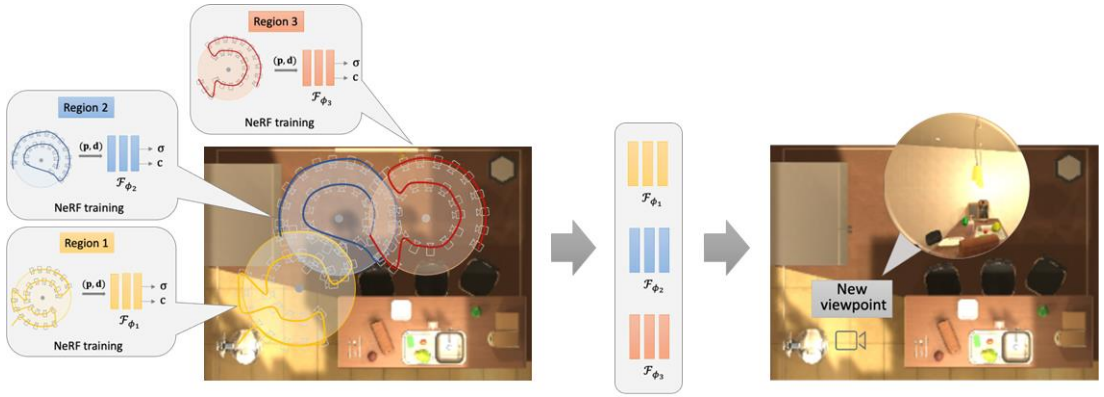
Data acquisition:
the teacher shows the object in front of the cameras of the iCub. A tracking routine, uses stereo vision, selecting the pixels from the depth map that are closer to the robot, thus segmenting the object from the background. A bounding box is estimated to surround it, and it is stored as annotation jointly with the label of the objects which is provided verbally by the teacher.

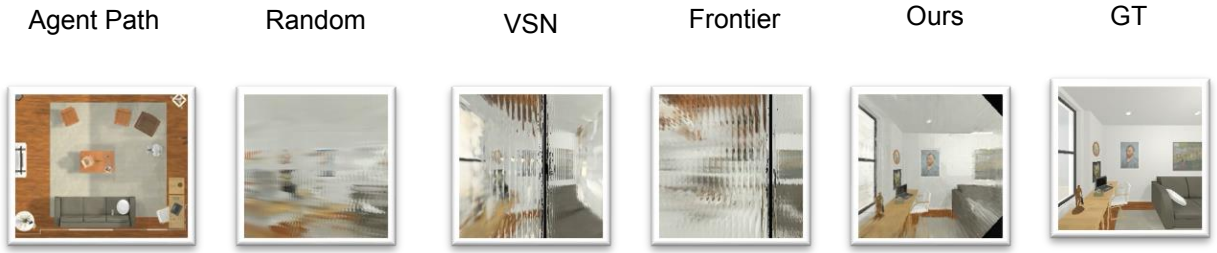
- Interactive Data Collection for Deep Learning Object Detectors on Humanoid Robots,

# 具身导航学习前沿

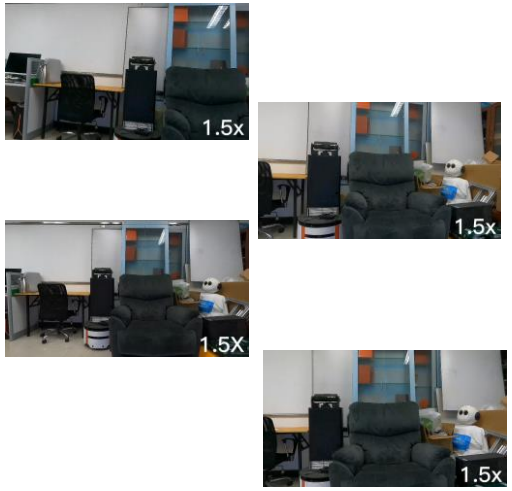


## ➢ 具身NeRF



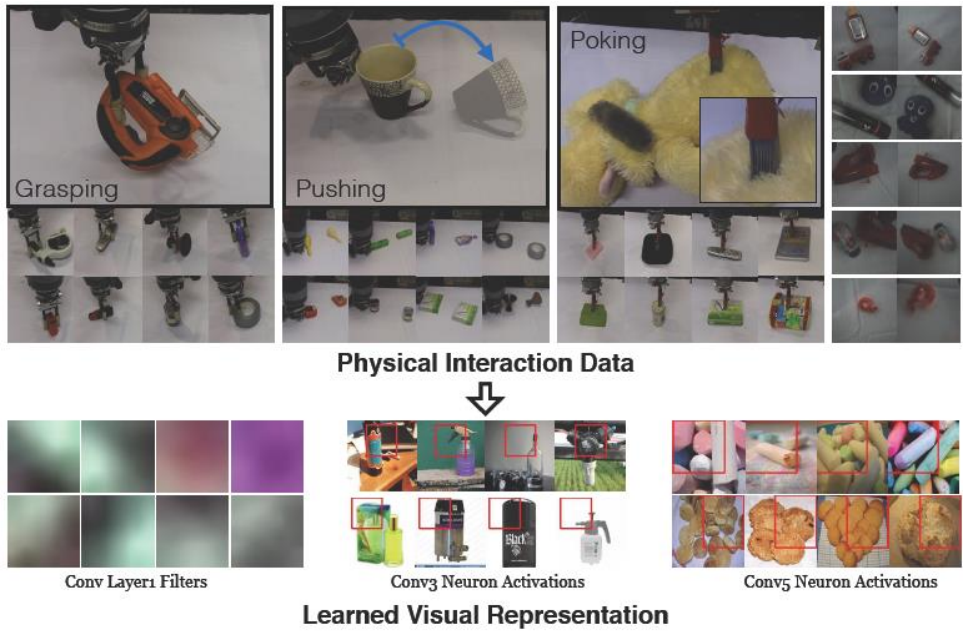


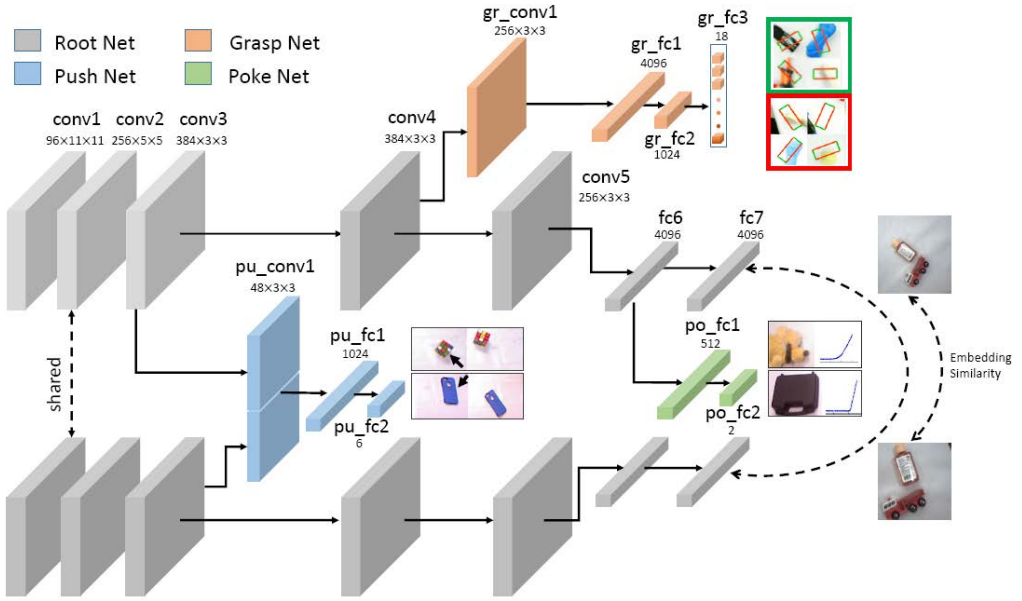



# 具身操作学习



## ➢ 好奇机器人（触觉学习）



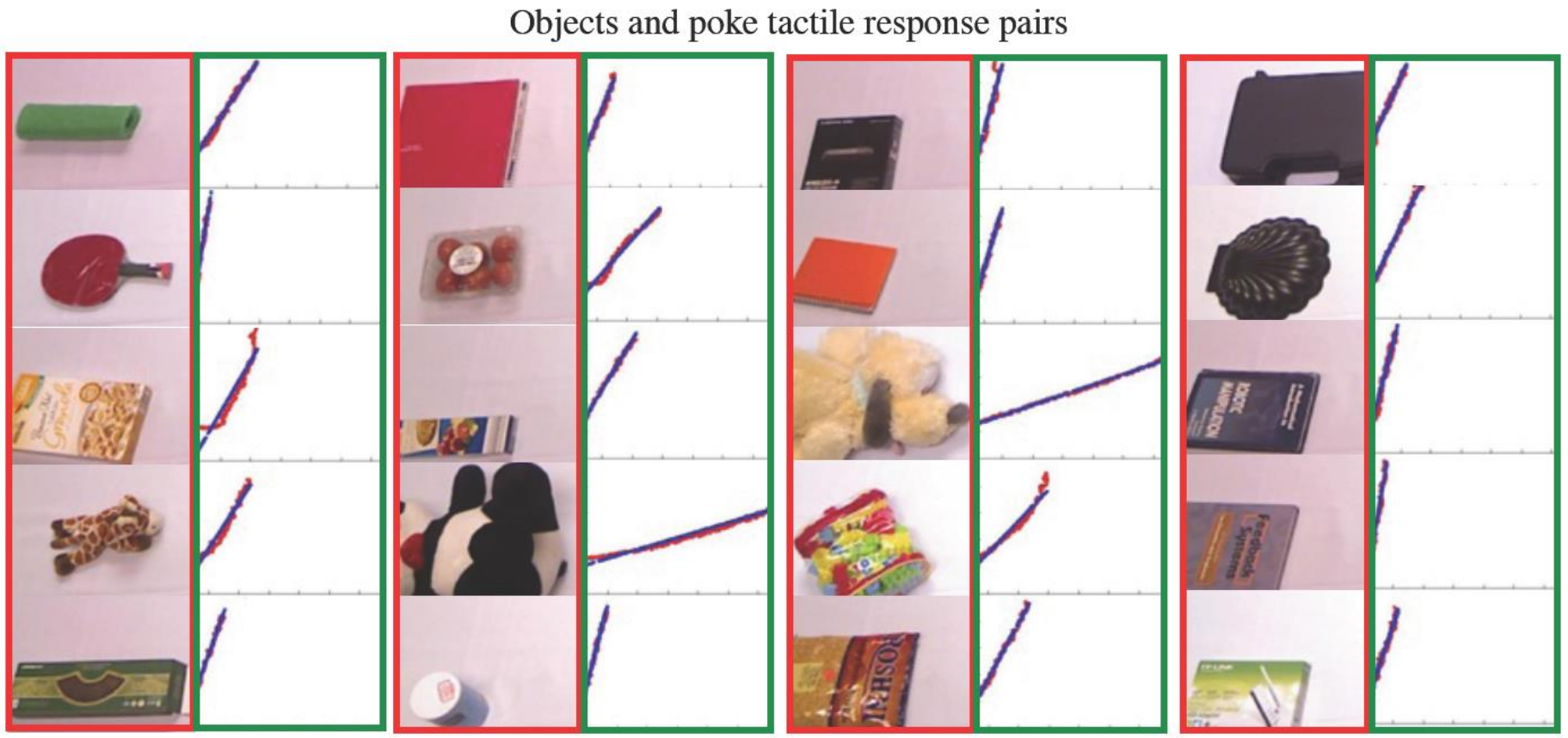

we build one of the first systems on a Baxter platform that pushes, pokes, grasps and observes objects in a tabletop environment. It uses four dierent types of physical interactions to collect more than 130K datapoints, with each datapoint providing supervision to a shared ConvNet architecture allowing us to learn visual representations.

- The Curious Robot: Learning Visual Representations via Physical Interactions, ECCV, 2016

# 具身操作学习



## ➢ 好奇机器人（触觉学习）




- The Curious Robot: Learning Visual Representations via Physical Interactions, ECCV, 2016

# 具身操作学习



## ➢ 好奇机器人（触觉学习）

**Table 1.** Classification accuracy on ImageNet Household, UW RGBD and Caltech-256

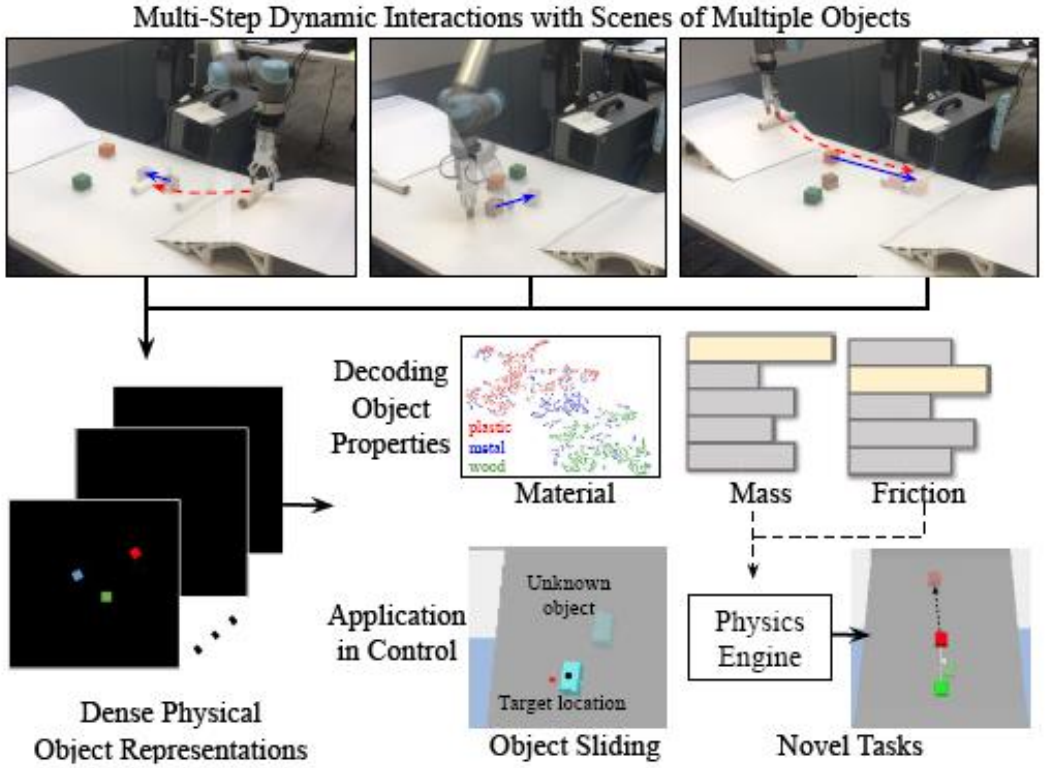
| | Household | UW RGBD | Caltech-256 |
|---|---|---|---|
| Root network with random init. | 0.250 | 0.468 | 0.242 |
| Root network trained on robot tasks (**ours**) | 0.354 | 0.693 | 0.317 |
| AlexNet trained on ImageNet | 0.625 | 0.820 | 0.656 |
| Root network trained on identity data | 0.315 | 0.660 | 0.252 |
| Auto-encoder trained on all robot data | 0.296 | 0.657 | 0.280 |

- The Curious Robot: Learning Visual Representations via Physical Interactions, ECCV, 2016

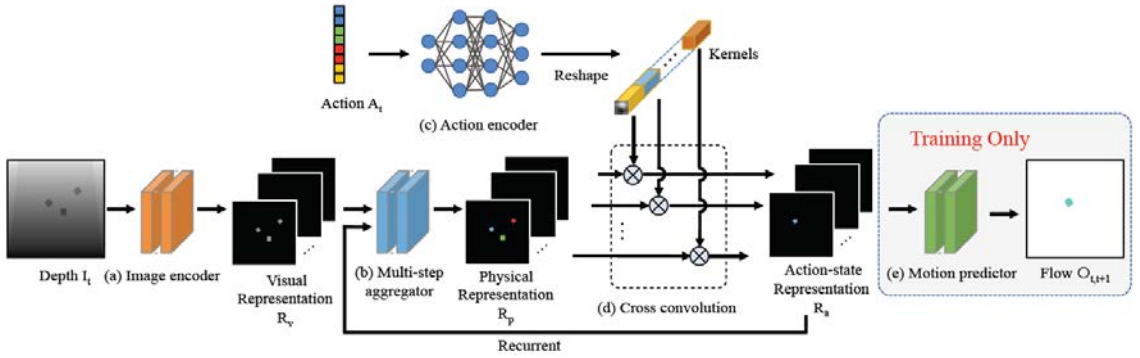
# 具身操作学习



## ➢ 学习物理性质




- We propose DensePhysNet, a system that actively executes a sequence of dynamic interactions (e.g., sliding and colliding), and uses a deep predictive model over its visual observations to learn dense, pixel-wise representations that reflect the physical properties of observed objects.

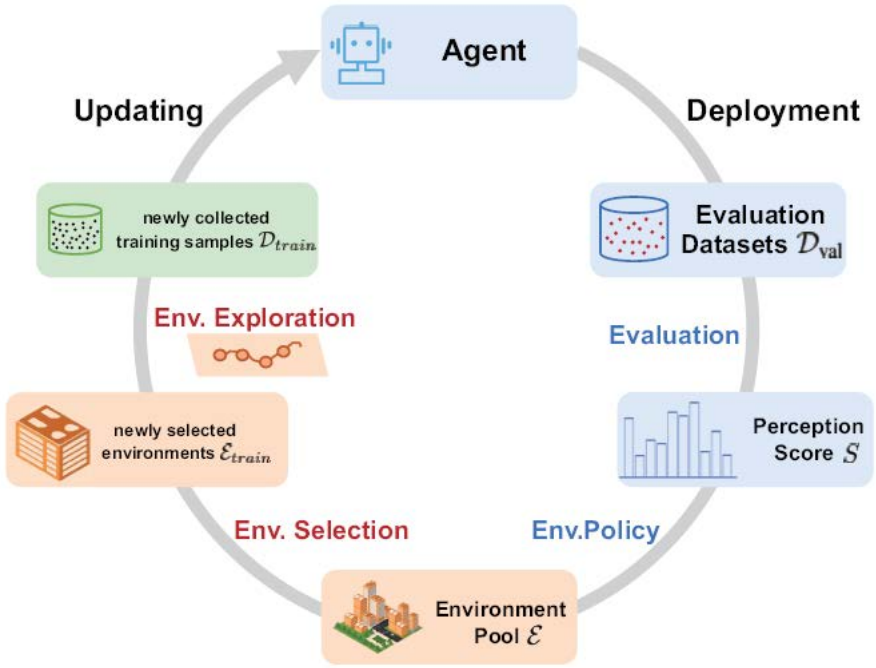


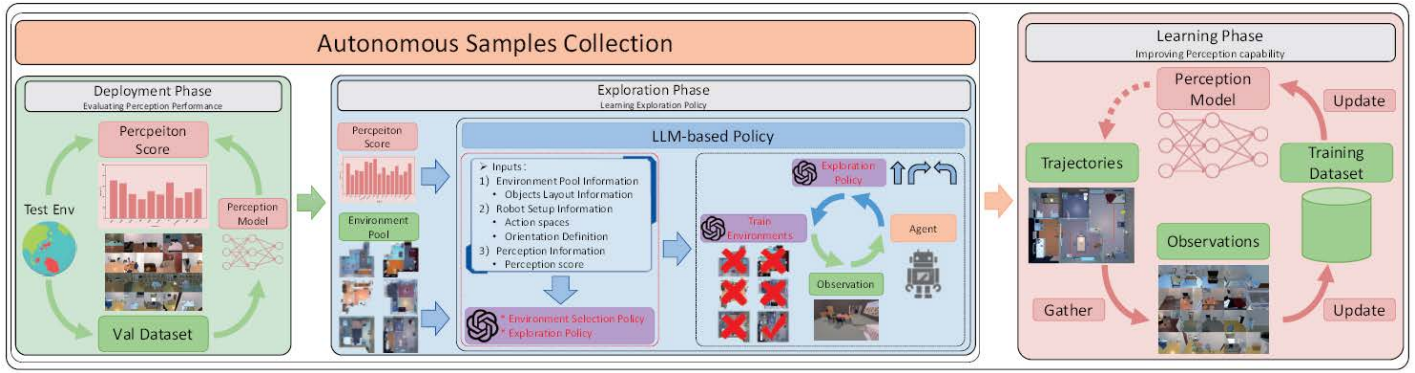

- Learning Dense Physical Object Representations via Multi-step Dynamic Interactions, RSS, 2020

# 具身操作学习

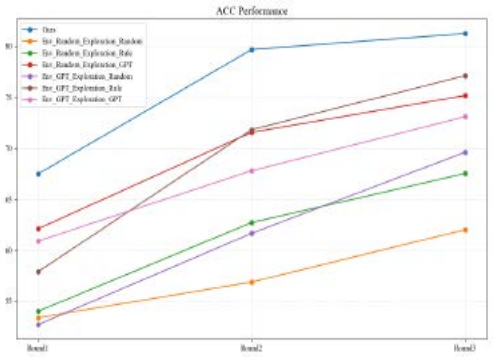


## ➢ 大模型



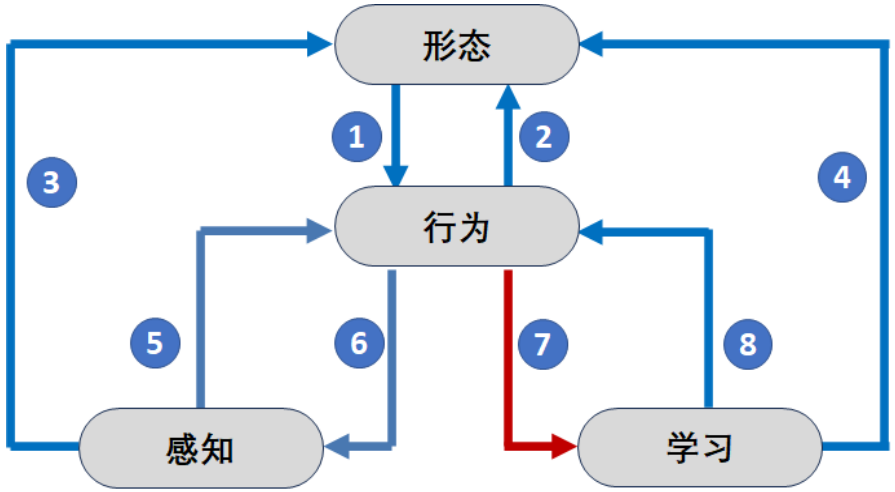

| Method | Round | Back. | Bed | Chair | Fridge | Micro. | Sink | Toilet | Table | Plant | Sofa | TV | mACC |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Pretrain | - | 97.92 | 1.21 | 0.00 | 0.03 | 0.01 | 0.03 | 50.08 | 0.00 | 0.00 | 0.03 | 0.00 | 13.27 |
| $Env_{Random} + Exp_{Random}$ | Round1 | 99.24 | 28.14 | 32.36 | 55.16 | 0.00 | 66.47 | 72.92 | 46.75 | 49.60 | 55.19 | 81.28 | 53.37 |
| | Round2 | 98.86 | 59.43 | 34.77 | 51.45 | 0.00 | 66.02 | 73.32 | 46.07 | 52.68 | 50.54 | 92.44 | 56.87 |
| | Round3 | 99.17 | 71.92 | 42.61 | 51.99 | 0.00 | 83.37 | 79.54 | 49.64 | 55.04 | 56.86 | 91.95 | 62.01 |
| $Env_{Random} + Exp_{Rule}$ | Round1 | 98.58 | 34.12 | 35.83 | 22.37 | 0.00 | 55.25 | 88.39 | 46.79 | 42.19 | 77.85 | 92.64 | 54.00 |
| | Round2 | 98.22 | 80.91 | 40.04 | 56.04 | 0.00 | 64.47 | 85.76 | 48.41 | 60.03 | 61.52 | 94.29 | 62.70 |
| | Round3 | 98.01 | 80.16 | 47.37 | 75.14 | 0.24 | 75.36 | 92.81 | 51.40 | 65.23 | 62.50 | 94.38 | 67.51 |
| $Env_{Random} + Exp_{GPT}$ | Round1 | 98.97 | 86.79 | 43.08 | 46.48 | 1.08 | 75.85 | 84.42 | 49.17 | 45.74 | 59.99 | 91.41 | 62.09 |
| | Round2 | 98.71 | 89.82 | 50.79 | 78.12 | 28.29 | 75.19 | 95.59 | 54.01 | 62.22 | 59.89 | 94.39 | 71.55 |
| | Round3 | 98.68 | 90.00 | 56.28 | 73.32 | 27.08 | 92.69 | 97.34 | 60.55 | 69.27 | 67.05 | 94.46 | 75.15 |
| $Env_{GPT} + Exp_{Random}$ | Round1 | 98.97 | 30.24 | 44.36 | 44.73 | 15.24 | 70.48 | 96.25 | 37.56 | 39.22 | 32.85 | 69.67 | 52.69 |
| | Round2 | 99.43 | 30.97 | 57.20 | 51.20 | 18.62 | 82.90 | 98.78 | 45.75 | 36.04 | 74.77 | 82.51 | 61.65 |
| | Round3 | 99.46 | 71.80 | 62.78 | 63.57 | 28.96 | 82.69 | 98.42 | 51.45 | 35.14 | 81.94 | 89.39 | 69.60 |
| $Env_{GPT} + Exp_{Rule}$ | Round1 | 97.82 | 56.06 | 44.67 | 35.98 | 6.49 | 88.74 | 97.51 | 39.59 | 39.34 | 42.01 | 88.45 | 57.88 |
| | Round2 | 98.63 | 72.53 | 61.87 | 50.24 | 64.56 | 88.30 | 97.99 | 44.92 | 55.70 | 62.58 | 92.60 | 71.81 |
| | Round3 | 99.32 | 73.06 | 66.05 | 80.19 | 80.85 | 87.64 | 96.87 | 56.39 | 52.36 | 61.20 | 94.11 | 77.10 |
| $Env_{GPT} + Exp_{GPT}$ | Round1 | 99.33 | 92.61 | 63.12 | 47.76 | 24.73 | 72.17 | 96.82 | 49.42 | 26.61 | 11.49 | 85.79 | 60.89 |
| | Round2 | 99.60 | 89.78 | 68.54 | 47.57 | 26.23 | 74.19 | 98.28 | 56.17 | 41.30 | 53.76 | 90.13 | 67.78 |
| | Round3 | 99.27 | 90.35 | 71.35 | 62.56 | 27.52 | 89.42 | 98.93 | 54.67 | 50.07 | 65.83 | 93.95 | 73.08 |
| SAIL | Round1 | 98.96 | 53.82 | 71.01 | 63.15 | 0.60 | 91.13 | 94.82 | 58.00 | 47.83 | 71.26 | 91.54 | 67.47 |
| | Round2 | 99.57 | 83.63 | 72.99 | 83.02 | 67.23 | 92.76 | 98.42 | 56.60 | 61.74 | 68.31 | 91.99 | 79.66 |
| | Round3 | 99.59 | 84.87 | 75.73 | 80.08 | 74.62 | 93.41 | 98.96 | 59.53 | 61.93 | 70.88 | 93.78 | 81.22 |



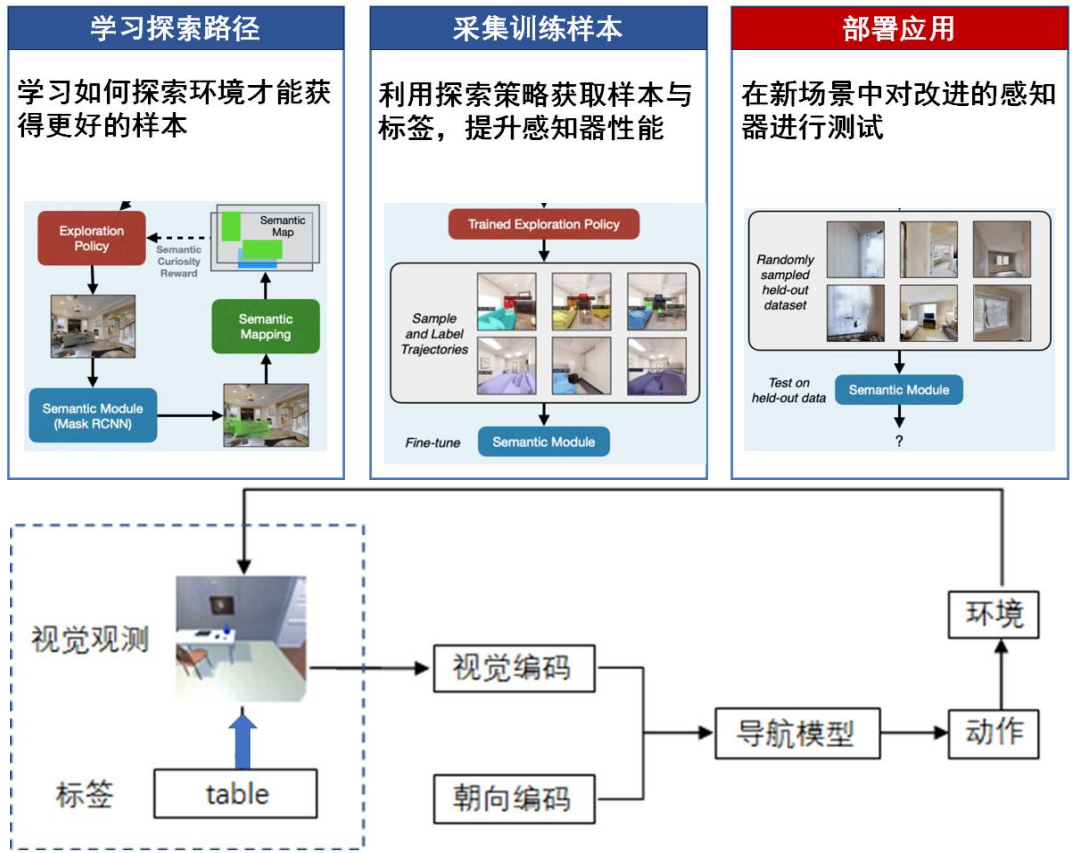
# 行为→学习：具身学习



## ➢ 小结





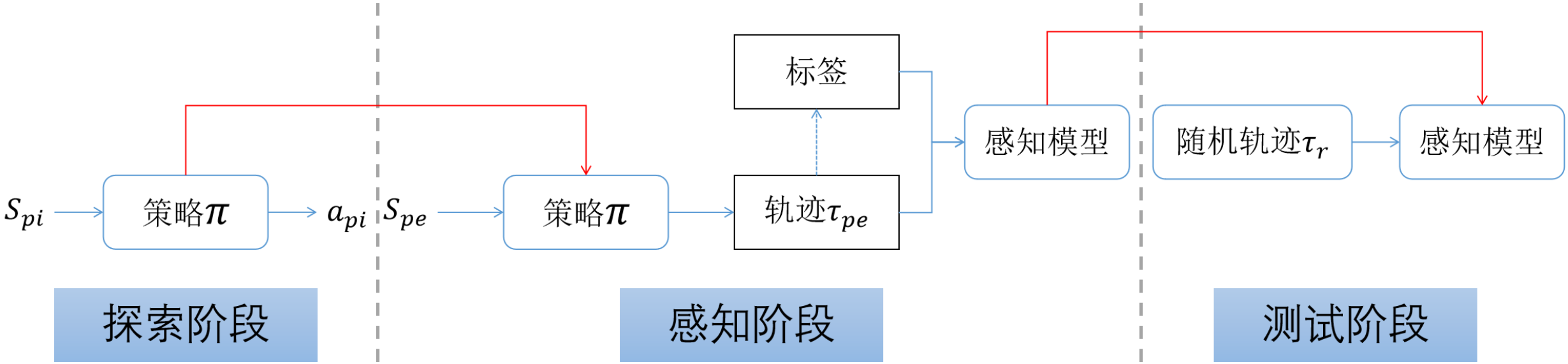


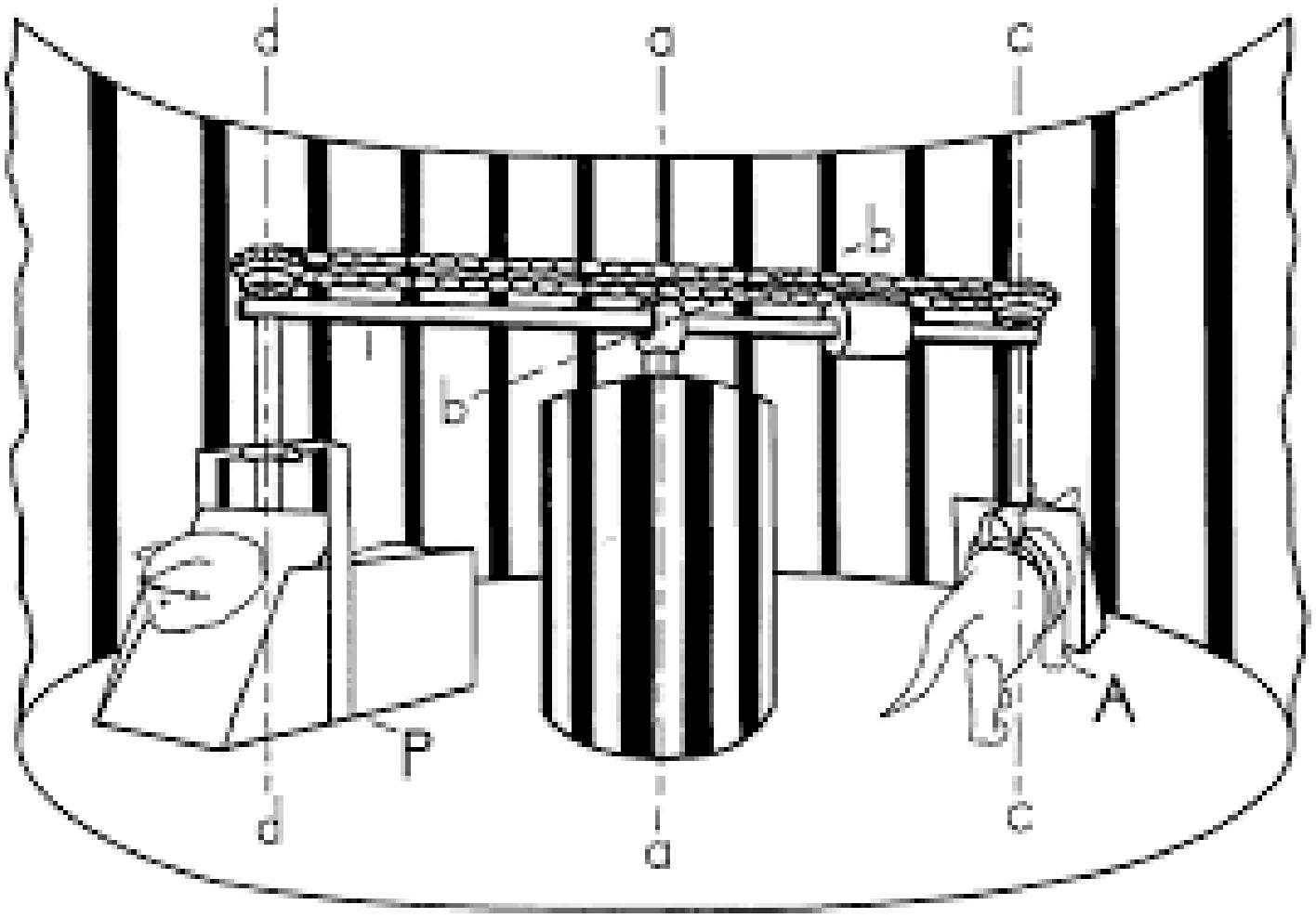

- 感知任务、高效探索，跨模态学习… …

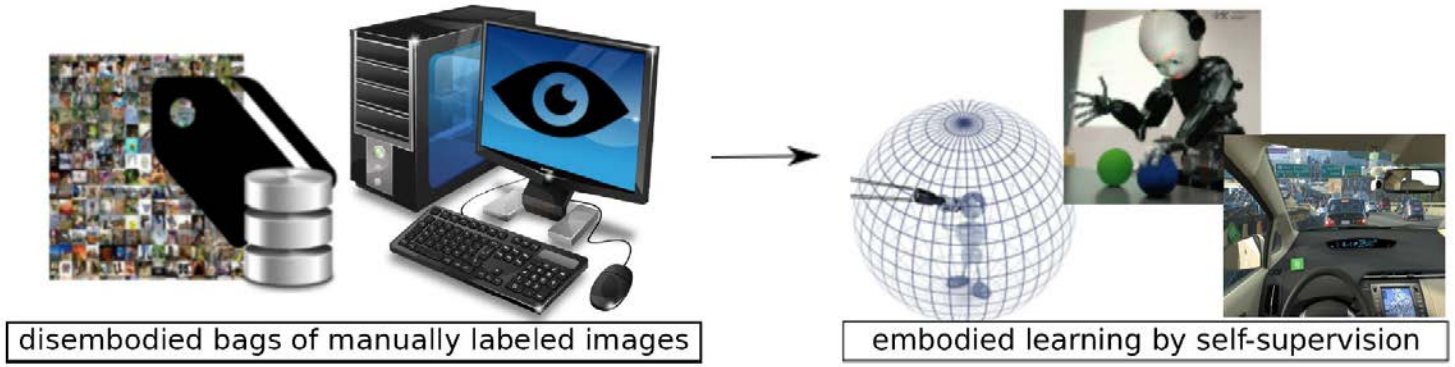
# 具身导航学习小结



## ➢ 具身学习




d
a
c
b
b
P
A
d
a
c


disembodied bags of manually labeled images
embodied learning by self-supervision

清華大學
Tsinghua University

# Towards Compositional Generalization for Robot Learning

5.12PM
15:30

## Abstract:

Intelligence is the ability to do the right thing in myriad unfamiliar situations. Robots today can do very well in complex, but narrow tasks, e.g., in-hand manipulation. However, home service robots (if any) are far from their human counterparts in abilities. What is the nature of the underlying gap? How would the latest advances in AI -- learning, planning, foundation models -- help to overcome this gap? In this talk, I will argue that integrating model-based planning and data-driven learning will lead to a data-driven, compositional learning architecture for scalable robot intelligence. The key issue here is the interplay, rather than the conflict, between structure and data. I will illustrate the general thinking with our work on robots navigating anywhere on our university campus, robots folding a variety of clothes, and robots aiming to operate in the open world.

## Biography:

David Hsu is a Provost's Chair Professor in the Department of Computer Science, National University of Singapore, the director of Smart Systems Institute, and also the founding director of NUS Artificial Intelligence Laboratory. He received BSc in computer science & mathematics from the University of British Columbia, Canada, and PhD in computer science from Stanford University, USA. He is an IEEE Fellow.

His research lies in the intersection of robotics and AI. In recent years, he has been working on robot planning and learning under uncertainty for human-centered robots. His work won multiple international awards, including, most recently, Test of Time Award at Robotics: Science & Systems (RSS) in 2021 and IJCAI-JAIR Best Paper Prize in 2022.

He has chaired or co-chaired several major international robotics conferences, including WAFR 2010, RSS 2015, ICRA 2016, and CoRL 2021. He served on the editorial boards of International Journal of Robotics Research and Journal of Artificial Intelligence Research. He is currently an Editor of IEEE Transactions on Robotics.