PreAfford: Universal Affordance-Based Pre-Grasping for Diverse Objects and Environments

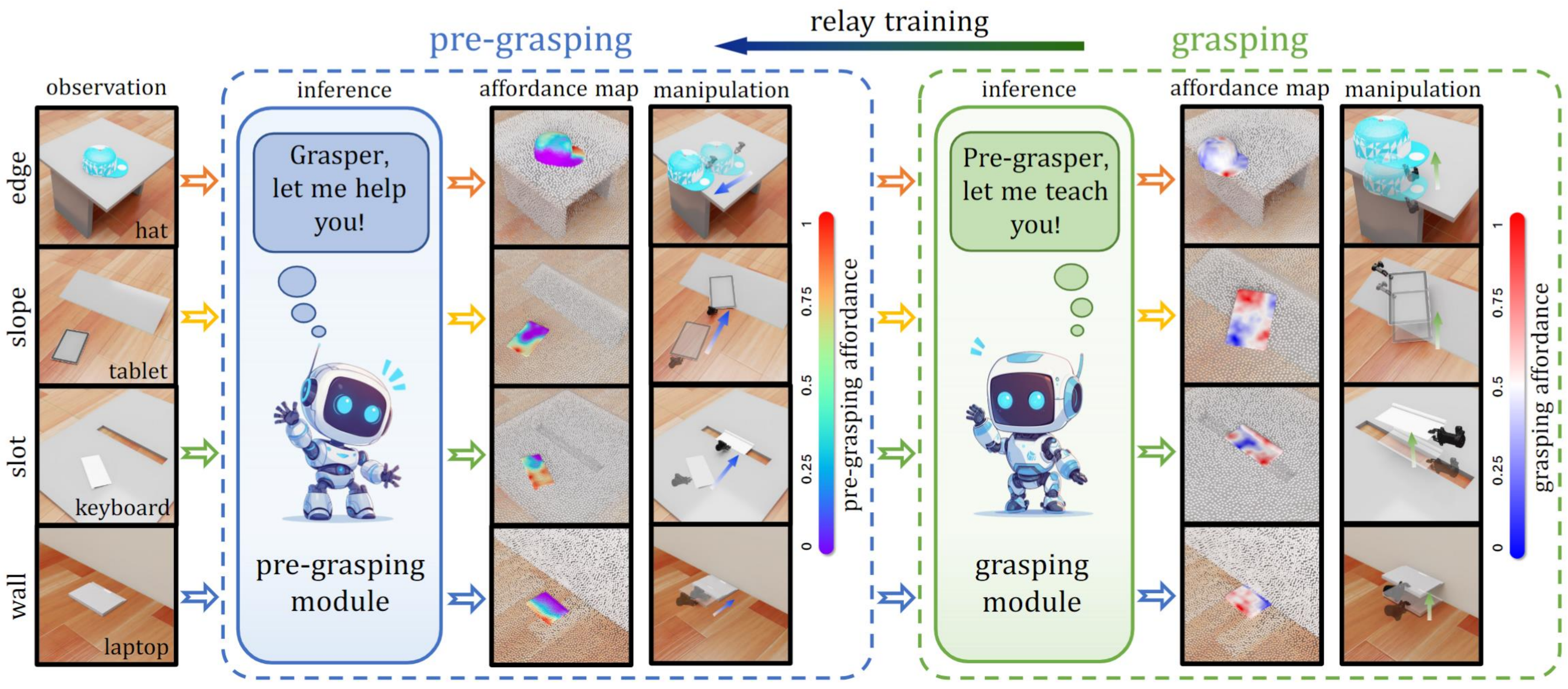
Kairui Ding[1,2], Boyuan Chen[1,2], Ruihai Wu[3], Yuyang Li[4], Zongzheng Zhang[1], Huan—ang Gao[1], Siqi Li[1], Guyue Zhou[1], Yixin Zhu[4] Hao Dong[3], and Hao Zhao[1]

[1]Institute for AI Industry Research (AIR), Tsinghua University, [2]Xingjian College, Tsinghua University, [3]CFCS, School of ComputerScience, Peking University, [4]Institute for Artificial Intelligence, PekingUniversity

IROS '24
ABU DHABI

Project website

I. Teaser: PreAfford architecture



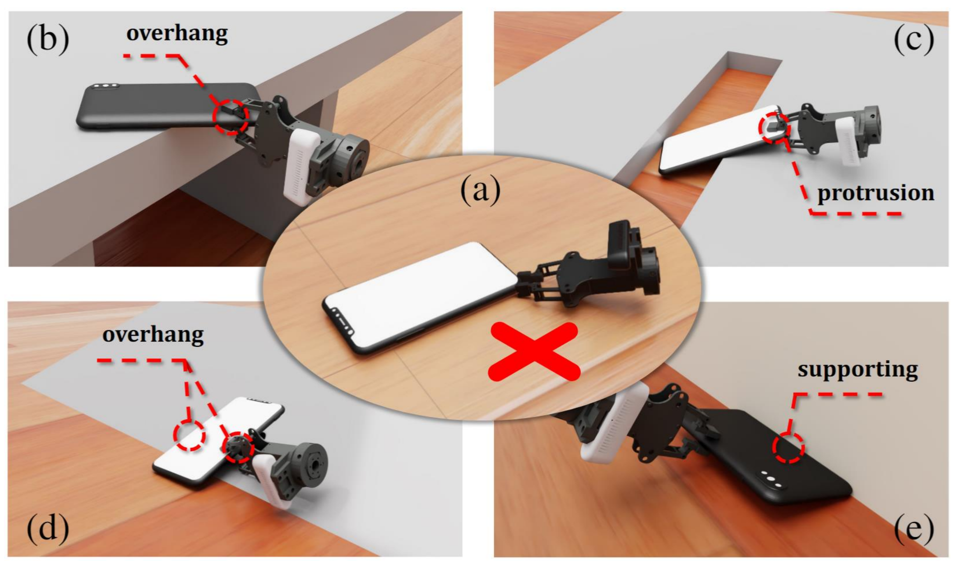
Employing a relay-training paradigm, two synergistic modules cooperate to facilitate the manipulation of objects typically considered ungraspable.

- Pre-grasping module: assessing environmental features such as edges, slopes, slots, and walls to propose strategic pre-grasping actions that enhance the likelihood of a successful grasp.
- Grasping module: evaluating these actions and provides feedback in the form of rewards, which are used to refine and optimize the pre-grasping strategies.

II. Preliminary: Pre-grasping

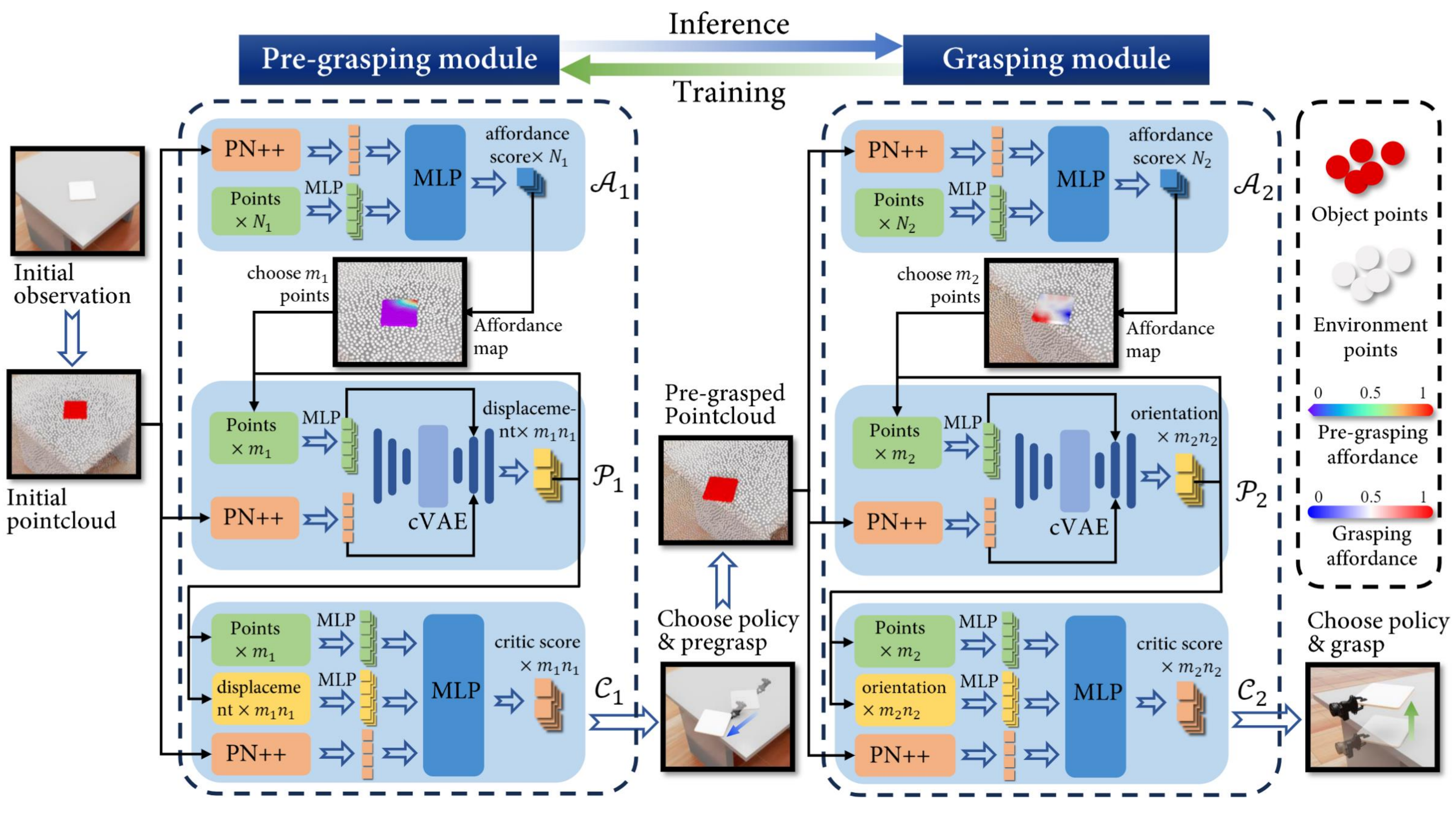
Some ungraspable objects exists in daily life, posing significant challenges for two-finger gripper to grasp directly.

For example:
Scissors Cap Phone Keyboard Laptop… ☹



But by repositioning the object before grasping, and leveraging environmental features like Edge, Slot, Slope, Wall, the object becomes graspable. ☺
This pre-manipulation on objects for easier grasping is called pre-grasping.

III. Method: PreAfford pipeline



Pre-grasping and grasping are managed by the pre-grasping module and the grasping module, respectively. Within each module, three specialized neural networks are employed: an affordance network $\mathcal{A}$, a proposal network $\mathcal{P}$, and a critic network $\mathcal{C}$.

Grasping network is trained first with following loss function, enabling critic network to judge success likelihood of grasping manipulation. $r = 1$ in a successful grasp, otherwise $r = 0$.

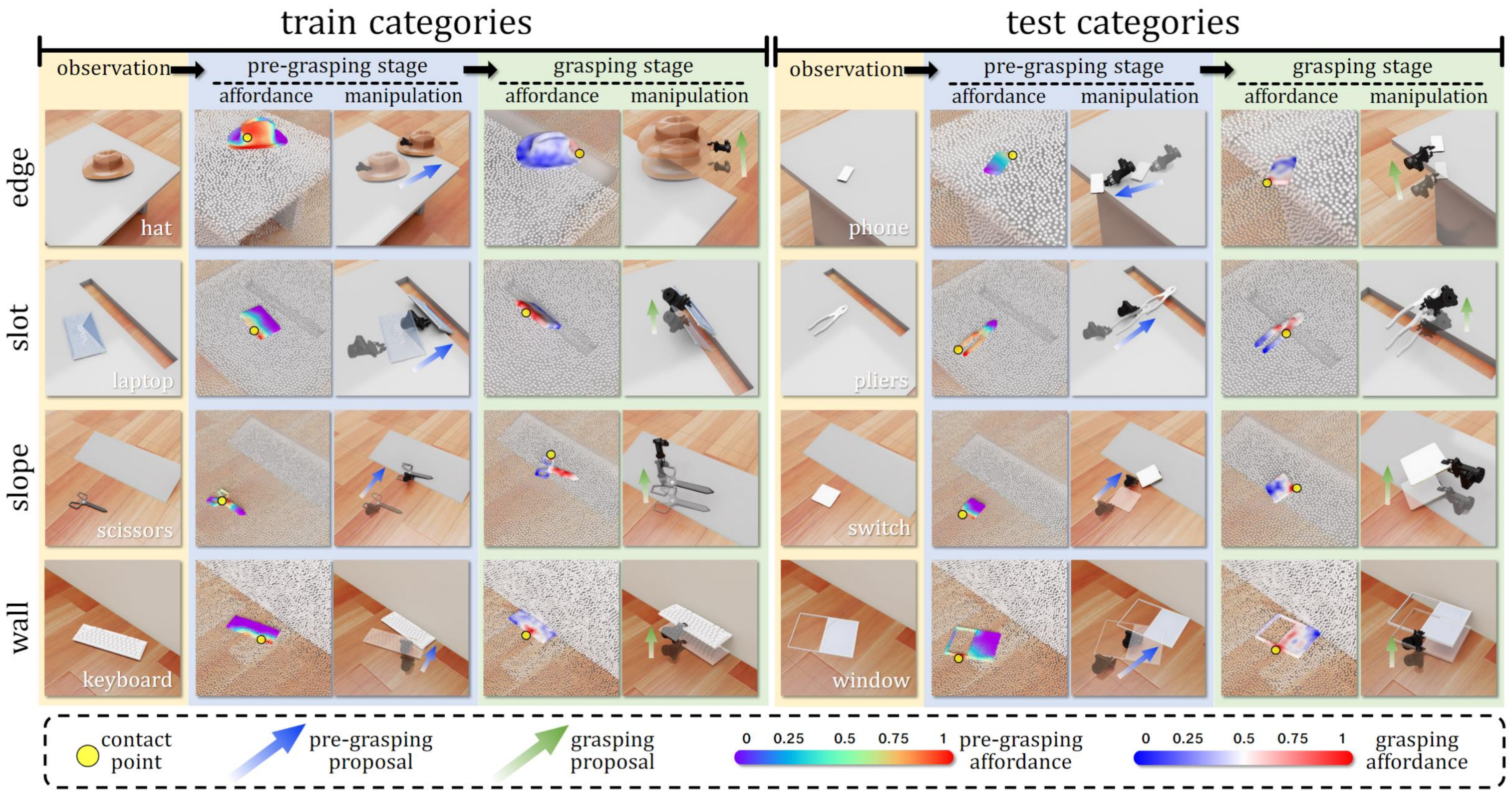
$$\mathcal{L}_{\mathcal{C}_2} = r\log(\mathcal{C}_2(p_2, \vec{\theta}_2)) + (1-r)\log(1-\mathcal{C}_2(p_2, \vec{\theta}_2)).$$

The training label for pre-grasping critic network is generated by the grasping module, which we refer to as relay-training. $\hat{c}_2^{\text{before}}$ and $\hat{c}_2^{\text{after}}$ are the success likelihood of grasping manipulation evaluated by grasping module, evaluating the increase of graspability with the pre-grasping manipulation. p is a penalty term introduced for safety reasons.

$$\mathcal{L}_{\mathcal{C}_1} = \left| \mathcal{C}_1(p_1, \Delta\vec{x}_1 | \mathcal{O}_1) - p \cdot (\hat{c}_2^{\text{after}} - \hat{c}_2^{\text{before}}) \right|.$$

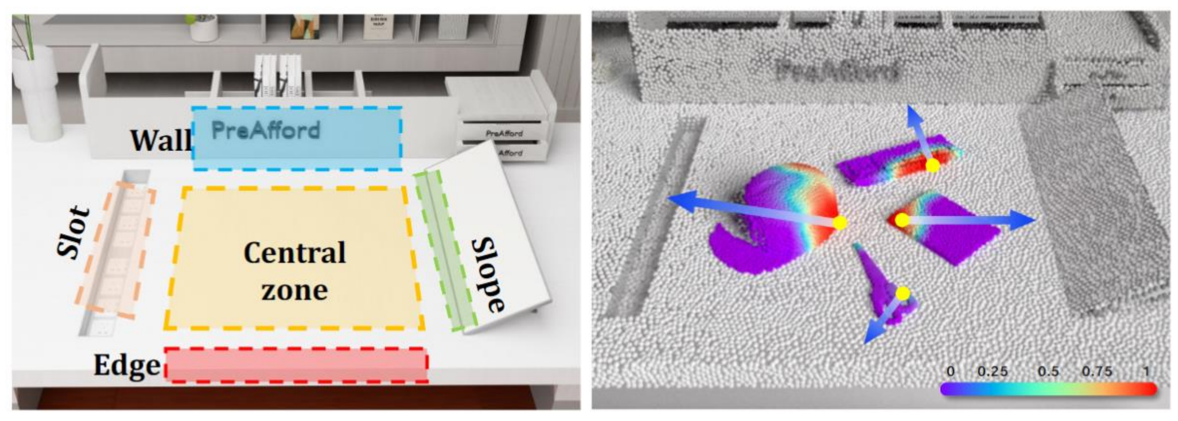
During the prediction phase, the two modules form a closed-loop control system. The pre-grasping module adjusts the object's pose until it is deemed suitable for the grasping module to apply a grasp.

IV. Experiment: Tests in simulation



We demonstrate pre-grasping manipulation on training and testing categories in four scenarios—edge, slot, slope, and wall. Affordance maps highlight effective interaction areas, showing PreAfford's capability to devise suitable pre-grasping and grasping strategies for various object categories and scenes, including both seen and unseen objects.

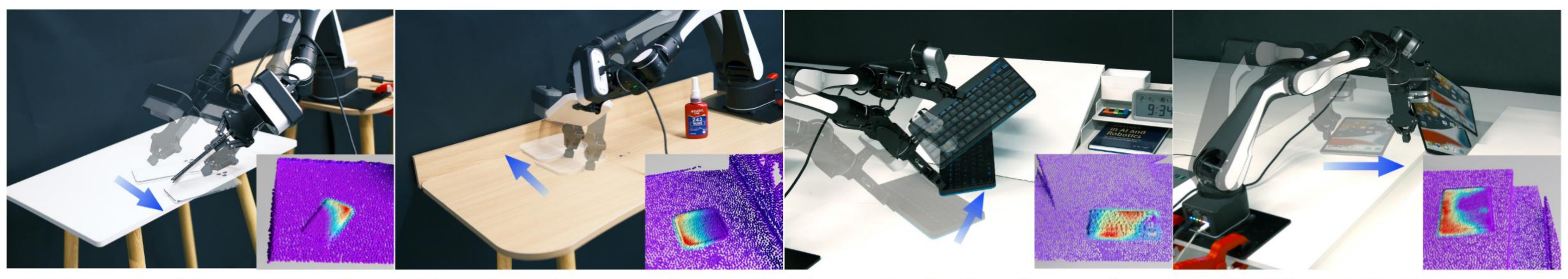
A complex scenario with multiple environmental features is also constructed. Quantitative results show that with closed-loop policy, PreAfford increases the success rate of grasping by 69 percent on test categories.



(a) A complex environment (b) Affordance heatmap

Setting	Train object categories						Test object categories					
	Edge	Wall	Slope	Slot	Multi	Avg.	Edge	Wall	Slope	Slot	Multi	Avg.
W/o pre-grasping	2.3	3.8	4.3	3.4	4.0	3.6	6.1	2.3	2.9	5.7	6.0	4.6
Random-direction Push	21.6	10.3	6.4	16.8	18.1	14.6	24.9	17.2	12.1	18.4	23.0	19.1
Center-point Push	32.5	23.7	40.5	39.2	39.0	35.0	25.1	17.4	28.0	30.2	21.5	24.4
Ours w/o closed-loop	67.2	41.5	58.3	76.9	63.6	61.5	56.4	37.3	62.6	75.8	55.4	57.5
Ours	**81.4**	**43.4**	**73.1**	**83.5**	**74.1**	**71.1**	**83.7**	**47.6**	**80.5**	**83.0**	**74.6**	**73.9**

V. Experiment: Tests in real-world



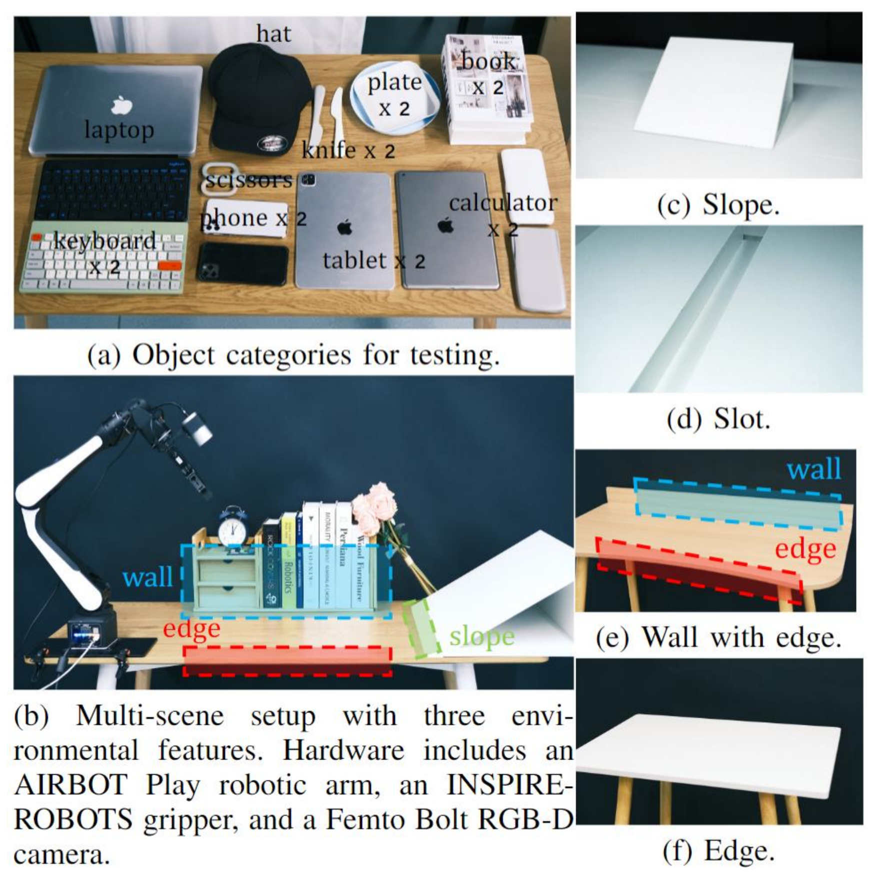
(a) move a tablet to table edge (b) push a plate towards a wall (c) push a keyboard up a slope (d) slide a tablet into a slot

Hardware setups:

- Robotic Arm: AIRBOT Play robotic arm, a compact six-degree-of-freedom manipulator designed by DISCOVER ROBOTICS
- RGB-D data: ORBBEC Femto Bolt laptop
- End-effector: an INSPIRE-ROBOTSEG2-4C gripper.

The experimental arrangement included ten object categories, split into five known and five unknown categories during training.

Experimental results show over 70 percent improvement on success rate of grasping manipulation in both seen and unseen categories.



(a) Object categories for testing.
(b) Multi-scene setup with three environmental features. Hardware includes an AIRBOT Play robotic arm, an INSPIRE-ROBOTS gripper, and a Femto Bolt RGB-D camera.
(c) Slope.
(d) Slot.
(e) Wall with edge.
(f) Edge.

Setting	Seen categories						Unseen categories					
	edge	wall	slope	slot	multi	avg.	edge	wall	slope	slot	multi	avg.
W/o pre-grasping	0	0	0	0	0	0	10	0	5	0	0	3
With pre-grasping	70	45	80	90	85	74	80	30	75	90	85	72

VI. Reference

Ding, K., Chen, B., Wu, R., Li, Y., Zhang, Z., Gao, H. A., ... & Zhao, H. (2024). PreAfford: Universal Affordance-Based Pre-Grasping for Diverse Objects and Environments. IEEE/RSJ International Conference on Intelligent Robots and Systems.