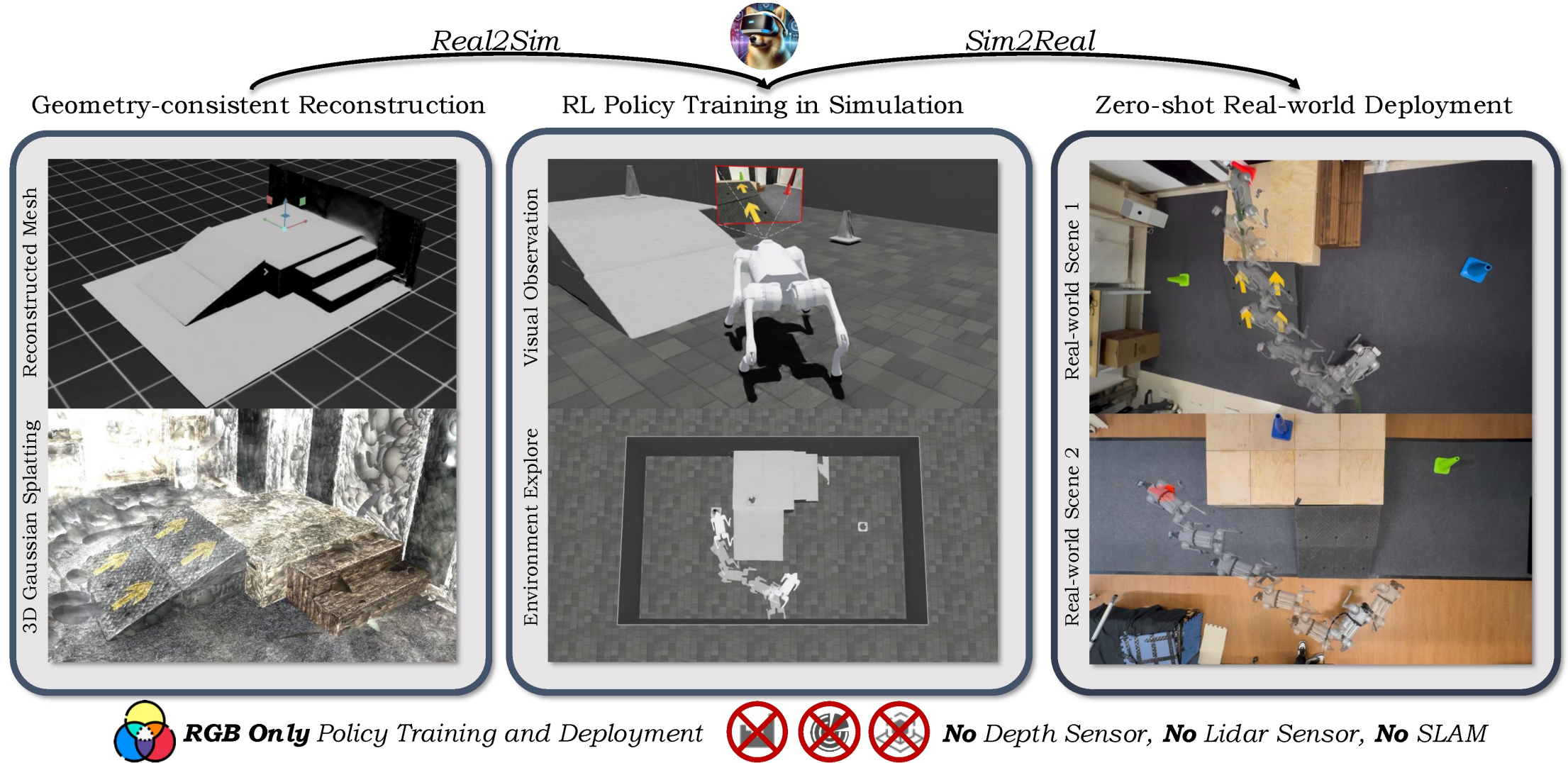


VR-Robo: A *Real-to-Sim-to-Real* Framework for Visual Robot Navigation and Locomotion

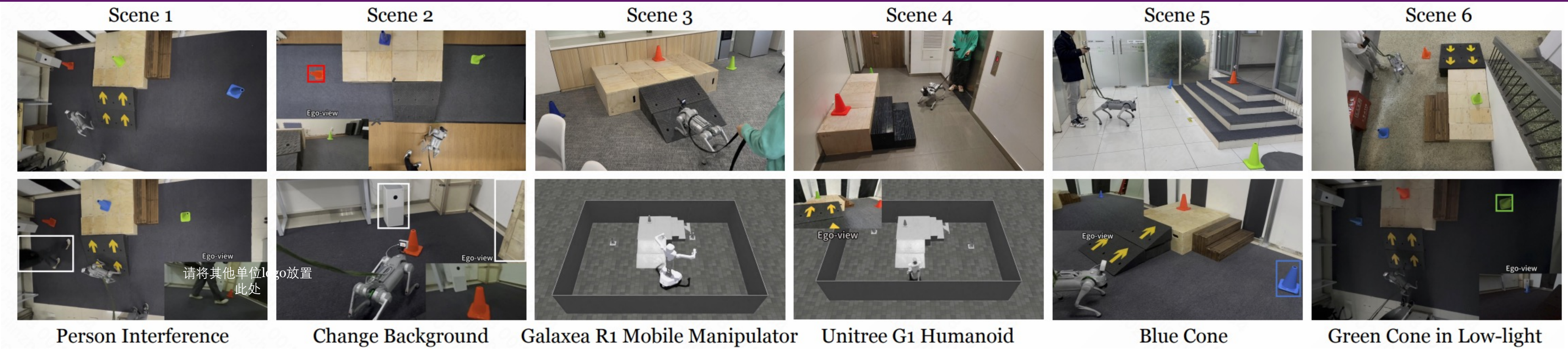
Shaoting Zhu\* Linzhan Mou\* Derun Li Baijun Ye Runhan Huang Hang Zhao<sup>†</sup>

Motivation and Introduction

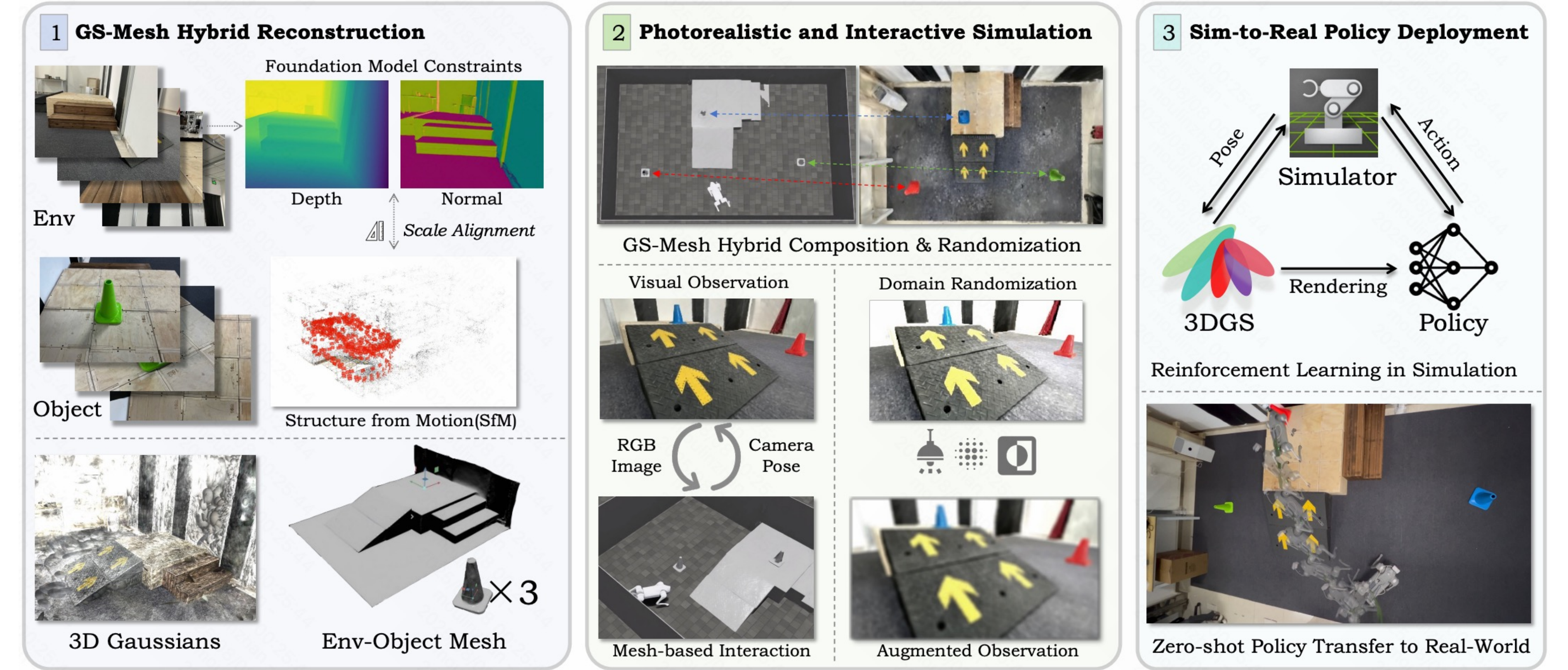
- RL policies trained in physical simulators often encounter challenges when deployed in real-world environments due to **sim-to-real gaps**.
- Simulators typically fail to replicate **visual realism and complex real-world geometry**. Moreover, the lack of realistic visual rendering limits the ability of these policies for **high-level tasks** requiring RGB-based perception like ego-centric navigation.
- We present a **Real-to-Sim-to-Real** framework that generates photorealistic and physically interactive "digital twin" simulation environments for **visual navigation and locomotion learning**.



Diverse Experiment Settings

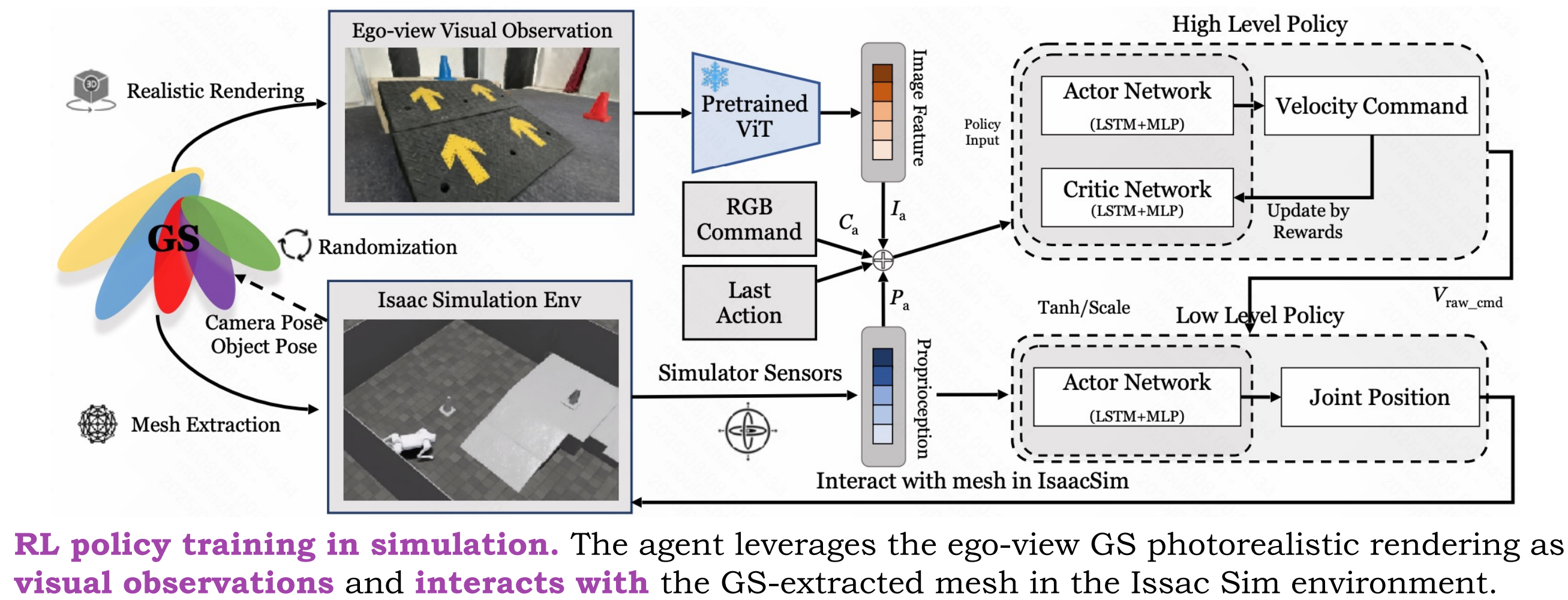


Overall Pipeline



**VR-Robo *real-to-sim-to-real* framework.** We build a **realistic and physically interactive** simulation environment with GS-mesh hybrid representation and occlusion-aware composition & randomization for policy training. Finally, we **zero-shot** transfer the RL policy trained in simulation into the real robot for **ego-centric visual navigation and locomotion**.

RL Policy Training



**RL policy training in simulation.** The agent leverages the ego-view GS photorealistic rendering as **visual observations** and **interacts with** the GS-extracted mesh in the Issac Sim environment.

Experiment Results

TABLE I: Comparison and ablation experimental results in the real-world setting.

Method	Exteroception	Success Rate $\uparrow$			Average Reaching Time (s) $\downarrow$		
		Easy	Medium	Hard	Easy	Medium	Hard
<b>Ours</b>	RGB	<b>100.00%</b>	<b>93.33%</b>	<b>100.00%</b>	<b>4.96</b>	<b>6.28</b>	<b>9.09</b>
Imitation Learning (IL)	RGB	0.00%	0.00%	0.00%	15.00	15.00	15.00
SARO [40]	RGB	66.67%	26.67%	0.00%	46.49	57.24	60.00
Textured Mesh	RGB	20.00%	6.67%	0.00%	12.90	14.90	15.00
CNN Encoder	RGB	73.33%	66.67%	6.67%	9.10	11.41	14.90
w/o Domain Randomization	RGB	53.33%	6.67%	0.00%	10.04	14.76	15.00

TABLE II: Comparison results in the simulation setting.

Method	SR $\uparrow$	ART (s) $\downarrow$
<b>Ours</b>	<b>100.00%</b>	<b>4.94</b>
Imitation Learning (IL)	8.67%	14.01
Random Background	43.33%	11.75

TABLE III: Ablation results in the simulation setting.

Method	SR $\uparrow$	ART (s) $\downarrow$
<b>Ours</b>	<b>100.00%</b>	<b>4.94</b>
Textured Mesh	22.00%	12.73
CNN Encoder	54.67%	10.04

