UniVTAC: A Unified Simulation Platform for Visuo-Tactile Manipulation Data Generation, Learning, and Benchmarking

Webpage

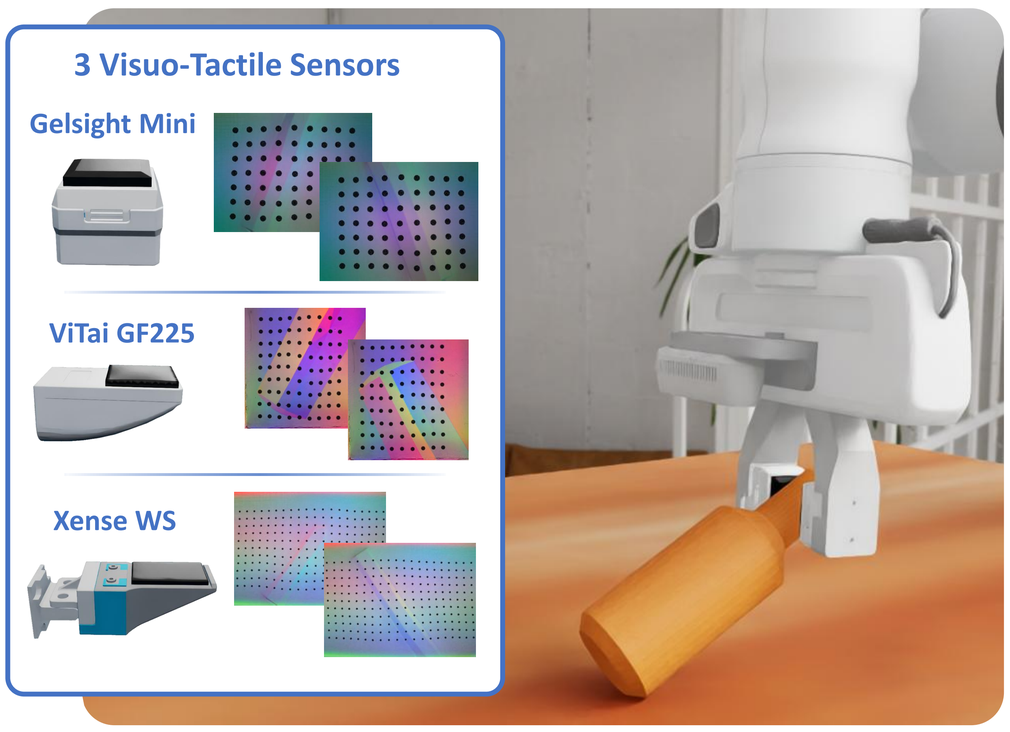
Baijun Chen[5,2*], Weijie Wan[6*], Tianxing Chen[4*], Xianda Guo[7,2*], Congsheng Xu[1], Yuanyang Qi[3], Haojie Zhang[3], Longyan Wu[8], Tianling Xu[1], Zixuan Li[6], Yizhe Wu[3], Rui Li[3], Xiaokang Yang[1], Ping Luo[4], Wei Sui[2†], and Yao Mu[1†]

[1] ScaleLab, Shanghai Jiao Tong University [2] D-Robotics [3] ViTai Robotics [4] The University of Hong Kong [5] Nanjing University [6] Shenzhen University [7] Wuhan University [8] Fudan University

1 Why Visuo-Tactile Manipulation

- Vision suffers from occlusion and poor contact observability.
- Tactile sensing provides local contact geometry, deformation, and shear cues.
- Large-scale tactile data and standardized benchmarks are scarce.

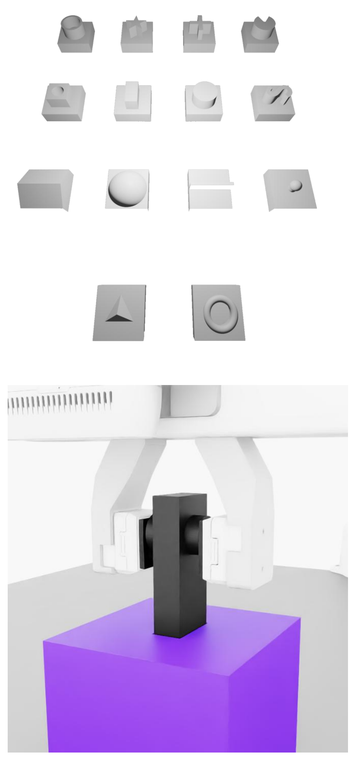
2 Scalable Visuo-Tactile Simulation

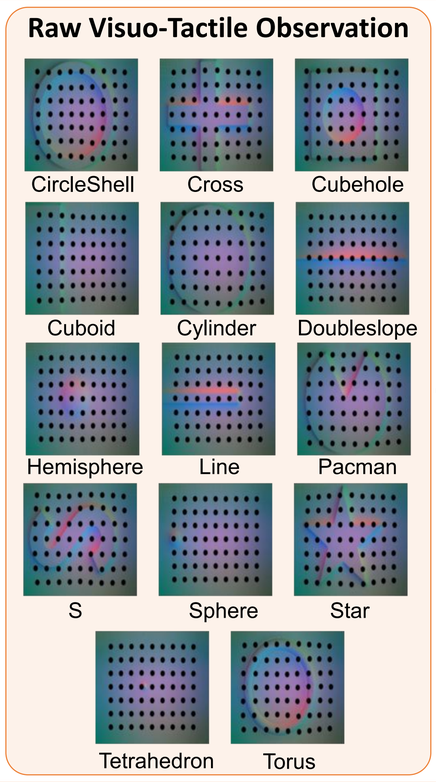


Tactile-Aware Manipulation Primitives

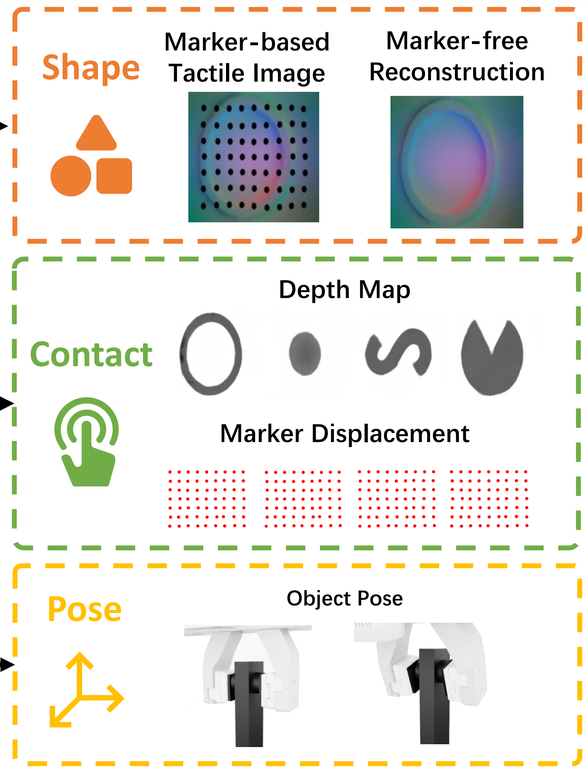
Unified simulation platform for visuo-tactile data generation across diverse sensors.

3 Privileged Supervision & UniVTAC Encoder





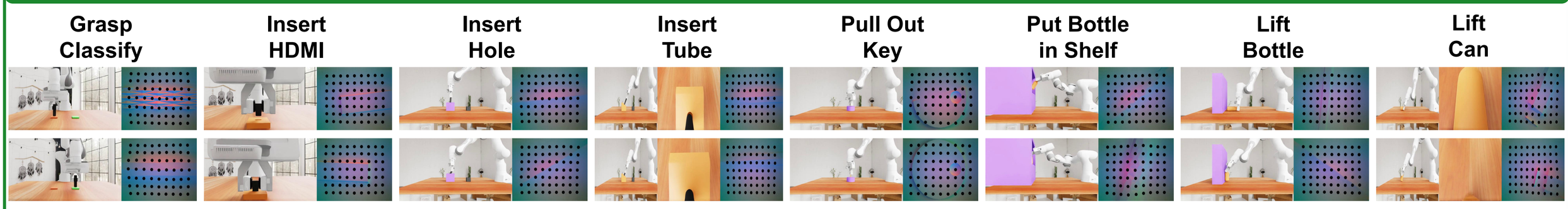
UniVTAC Encoder



Simulation provides privileged physical annotation that are difficult to obtain in the real world.

Pretrain in simulation, then integrate the encoder into downstream policy learning.

4 UniVTAC Benchmark



Shape preception | Pose reasoning | Contact-rich Interaction

5 Main Results

Simulation (8 Tasks) — ACT (Vision-Only) 30.9% → ViTAL(Strong Baseline) 40.5% → ACT+UniVTAC(Ours) 48.0% (+17.1%)

Method	Lift Bottle	Pull-out Key	Lift Can	Put Bottle in Shelf	Insert Hole	Insert HDMI	Insert Tube	Grasp Classify	Average
ACT	42	28	20	28	19	15	45	50	30.9
VITaL	**72**	**47**	8	**32**	**25**	6	34	**100**	40.5
Ours	71	46	**29**	31	24	**28**	**56**	99	**48**

Real-world (3 Tasks) — ACT (Vision-Only) 43.3% → ACT+UniVTAC (Ours) 68.3% (+25.0%)

Method	Insert Tube	Insert USB	Bottle Upright	Average
ACT	55	15	60	43.3
ACT+UniVTAC	**85**	**25**	**95**	**68.3**