InsertScale: A Benchmark towards Foundation Visuo-Tactile Policy Scalable Learning for Insertion Task

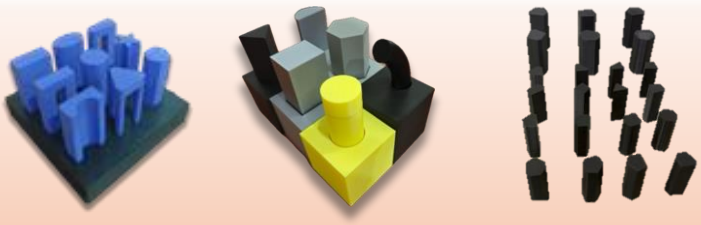
Congsheng Xu, Jingxiang Guo*, Baijun Chen*, Liuhaichen Yang, Zhen Zou, Yuzhang Li, Jieji Ren, Yiming Wang, Yichao Yan, Yao Mu†, Xiaokang Yang†*

Shanghai Jiao Tong University · National University of Singapore · Nanjing University

Abstract: Robotic visuo-tactile manipulation has become increasingly focused for its promising ability to push forward the performance boost for tasks that contains rich contact. Among all tasks, the insertion task is the most atomic one, so it is worth researching and benchmarking. While there is still a vacancy for a standard evaluation platform or a large foundational dataset for it. We first propose **Insert-100M**, a dataset that contains diverse insertion tasks with multiple shape parameters. We then propose **InsertScale**: a benchmark for generalized visuo-tactile insertion tasks with three novel visuo-tactile downstream tasks.

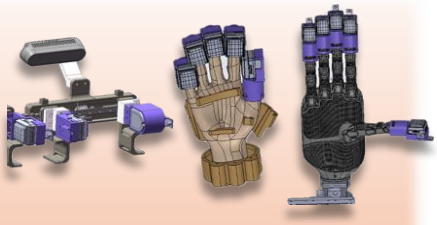
Insert-100M Dataset

Multi-Sensor Compatible, Multi-enbodiment compatible



Diversity in Shape with batched CodeGen:

1. Varies in Number of Sizes, Height and Ratio
2. Varies in Slot Margin and Slot Insertion Angle
3. Field Randomization Data Augmentation
4. Texture-Rich Augmentation



In-situ collection

Multi-embodiment

Multi-sensor

Cross-sensor

Real-world Collection Hardware Design

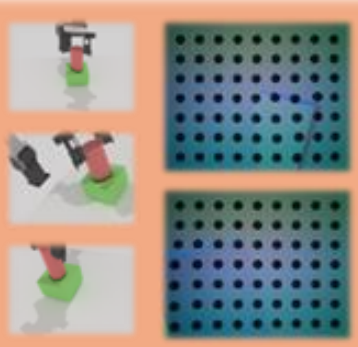
Multi-view Visual Info
Visuo-Tactile Info
Robotic State
Human In-situ Description

Optical MoCap

Multi-view and Multi-modal Collection

InsertScale Benchmark

Novel Down Stream Task, Towards General Model



Further Data Scaling

Portable DataGen

Parameterized Diversity

Scalable Simulation Collection

Zero-shot/Few-shot to Unseen Object

3D-based Visuo-tactile Learning

VTLA Residual Learning

Challenging Novel Downstream Tasks

VTLA | Tac-DP | Tac-ACT | Tac-SDP | ...

Practical Baselines