Vi-TacMan: Articulated Object Manipulation via Vision and Touch

Vision for global guidance • Touch for local precision

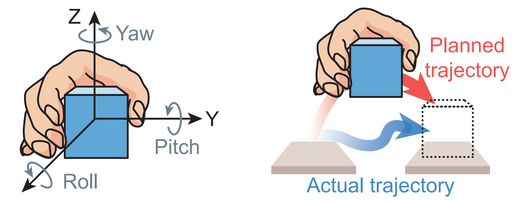
Leiyao Cui*, Zihang Zhao*†, Sirui Xie*, Wenhuan Zhang*, Zhi Han†, Yixin Zhu†

* Equal Contributor † Corresponding Author

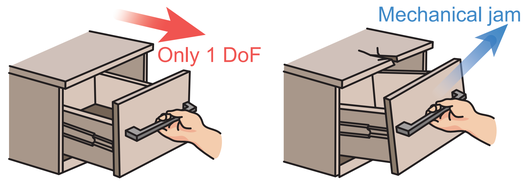
University of Chinese Academy of Sciences
Peking University

https://vi-tacman.github.io

Why is Articulated Object Manipulation Hard?



Floating (6 DoF) => **High tolerance**



Articulated (1 DoF) => **Zero tolerance**

Key Ideas

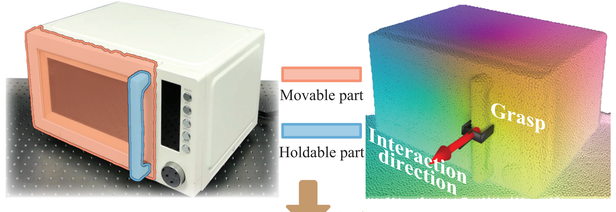
When Manipulating Unseen Articulated Objects

Vision-based methods can infer hidden kinematics but yield **imprecise** estimates.

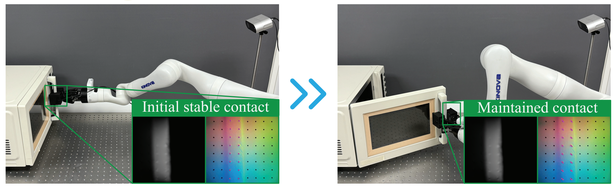
Touch-based methods achieve robust execution but require good **initializations**.

Core insight: coarse visual cues are sufficient when coupled with tactile feedback.

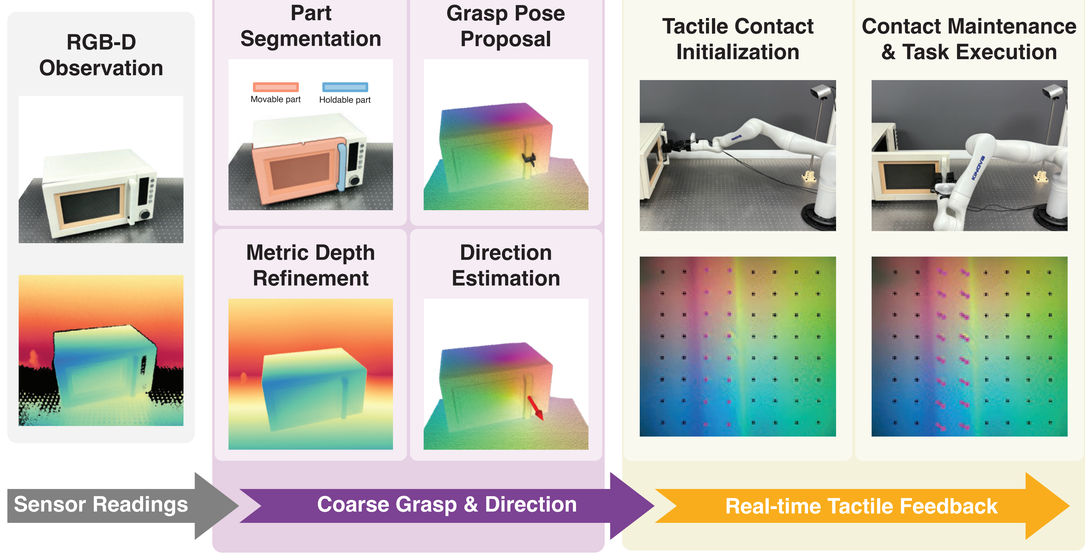
Vision-based high-level guidance



Tactile-informed low-level controller



The Vi-TacMan Framework



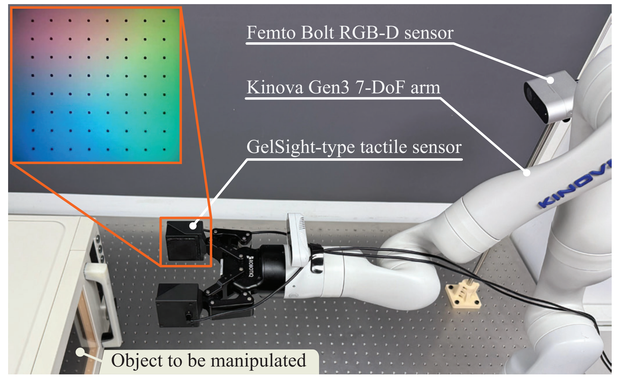
Synthetic Dataset & Real-World Setup

385 Articulated Objects from PartNet-Mobility

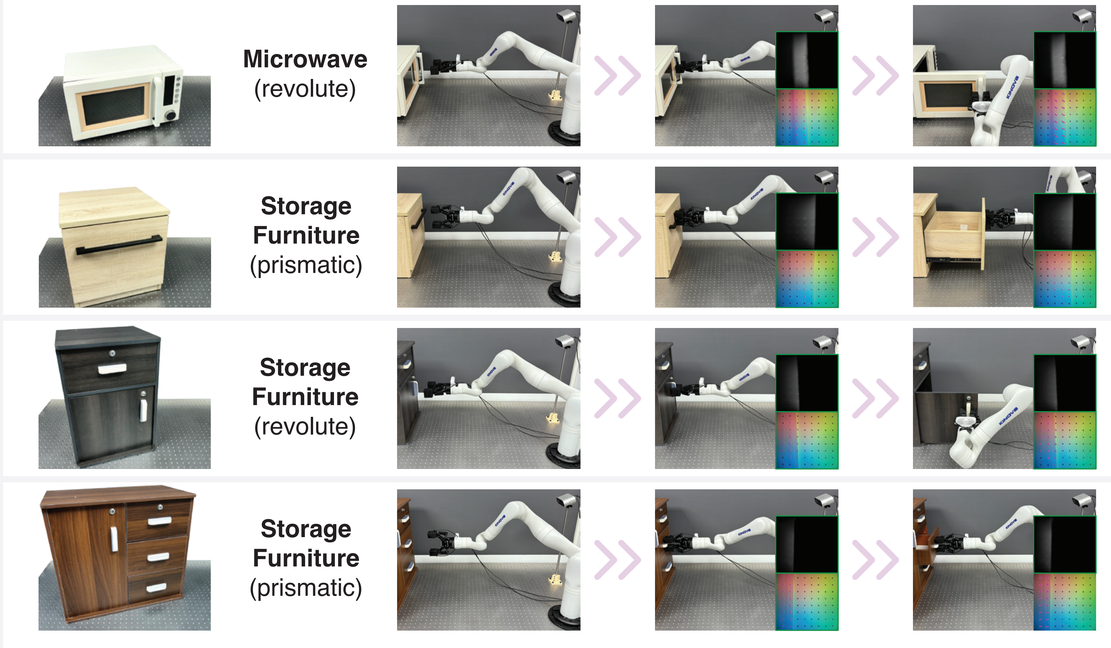
Microwave · Refrigerator
Storage Furniture · Trash Can

Dishwasher · Door
Oven · Table

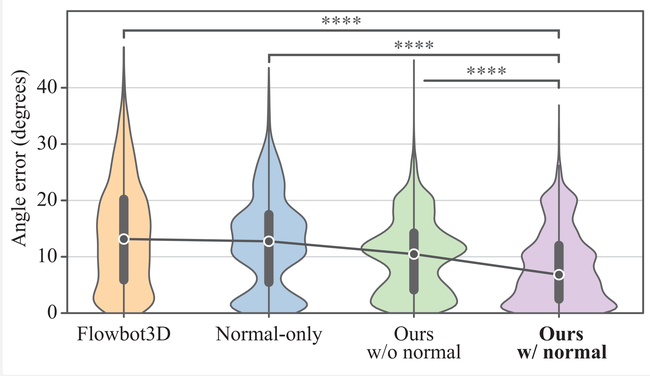
Training	Validation	Test
39524 samples	9881 samples	5836 samples



Real-World Validation



Direction Estimation on Unseen Objects



Website

Paper

Video