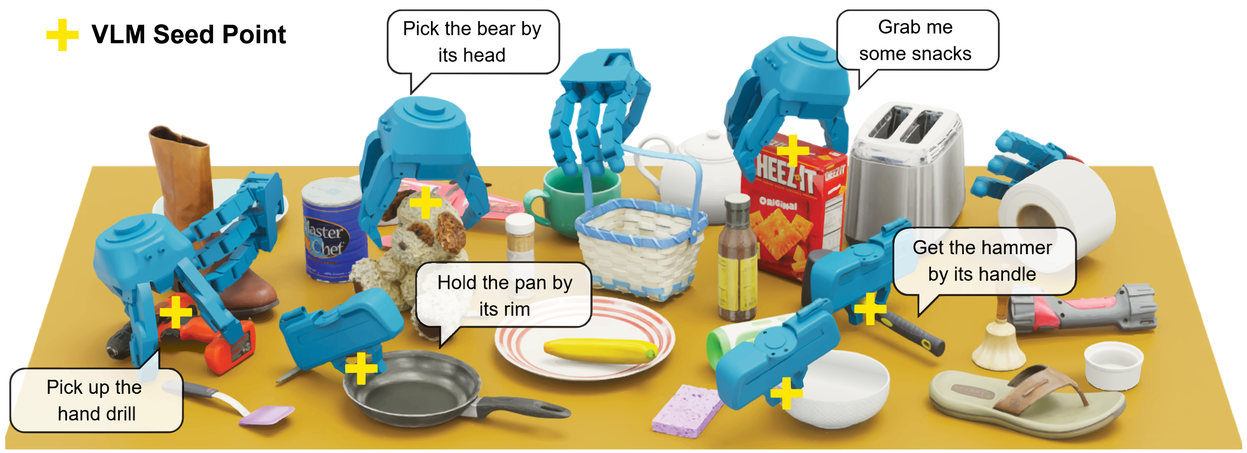
Control Robot Grasping With Natural Language

SeededGrasp: Language-Guided Grasping in Complex Scenes with Multiple Embodiments

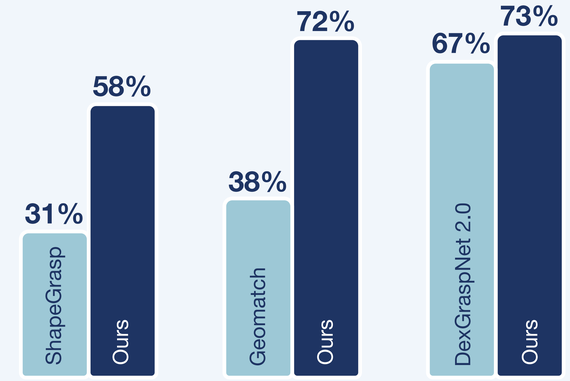
Overview

- Controlled grasp pose generation from point clouds.
- SeededGrasp contributions:

1. Natural language control.
2. Multiple robot embodiments.
3. Complex tabletop scenes.

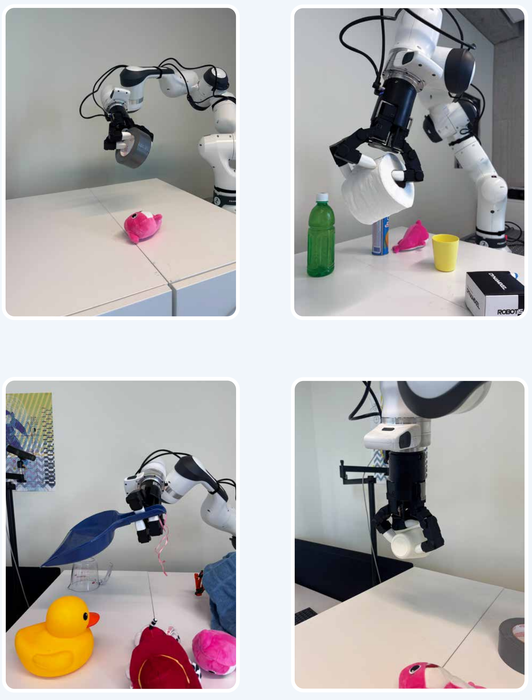


Results

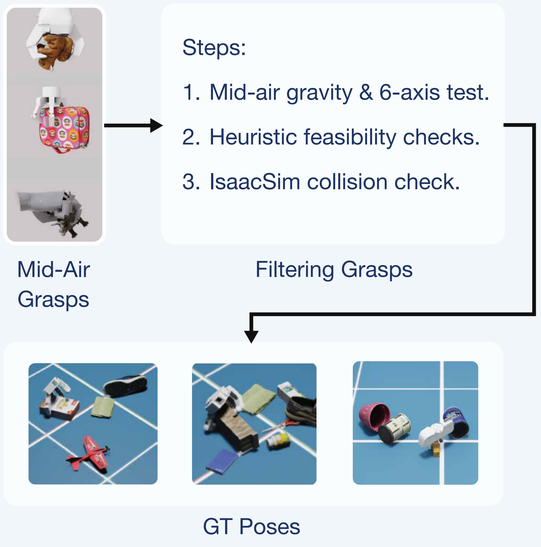


Success Rates In-Sim

78% Success Rate In-Real



Dataset

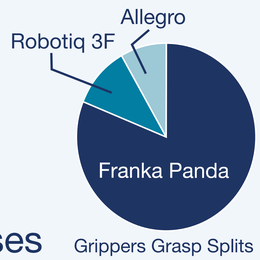


610 scenes

334 objects

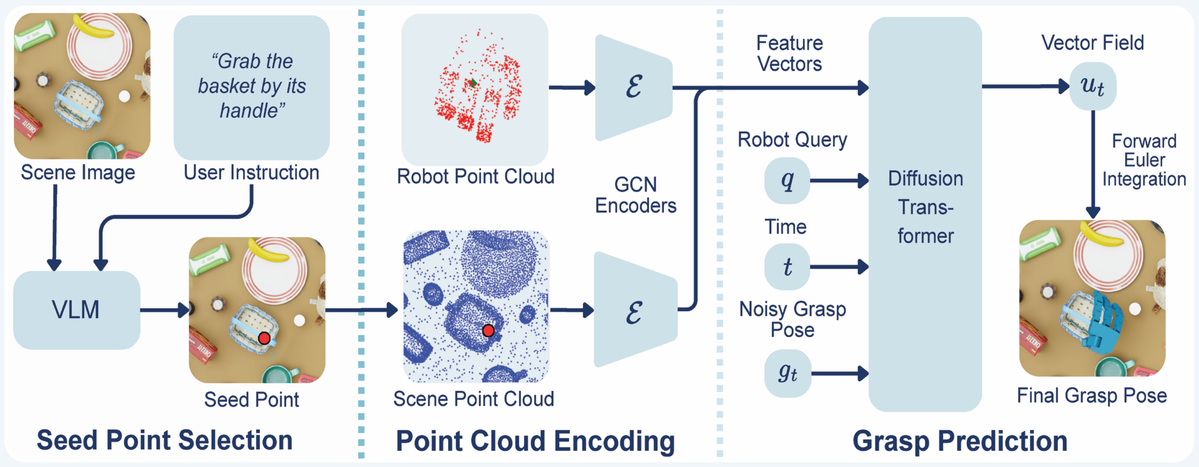
3 grippers

2.56M grasp poses



Method

- Train flow-matching model to generate grasps around 3D seed points.
- At inference, VLM predicts seed point from user commands.



UNIVERSITY OF TORONTO · VECTOR INSTITUTE · TORONTO INTELLIGENT SYSTEMS LAB

Yang Xu*, Gurpreet Mukker*, Raymond Wang*, Jasper Gerigk, Maria Attarian, Igor Gilitschenski