Embodied AI Tasks in ManiSkill and Visual Learning Challenges

Building and Working in Environments for Embodied AI (part VI)

CVPR 2022 Tutorial
ManiSkill Benchmark and Challenge

Generalizable Manipulation Skill Benchmark with Large-Scale Demonstrations

Tongzhou Mu*, Zhan Ling*, Fanbo Xiang*, Derek Yang*, Xuanlin Li*, Stone Tao, Zhiao Huang, Zhiwei Jia, Hao Su
Overview

We are going to talk about:

- What kind of manipulation tasks are difficult?
- Imitation learning, RL, and classical robotics, which one can solve embodied AI tasks better for now?
- Preview of ManiSkill Challenge 2022
Outline

- ManiSkill Benchmark
- Summary of ManiSkill Challenge 2021
- Preview of ManiSkill Challenge 2022
Outline

- ManiSkill Benchmark
- Summary of ManiSkill Challenge 2021
- Preview of ManiSkill Challenge 2022
Generalizable Manipulation Skills

- After learned to manipulate a category of objects, human can manipulate unseen objects of the same category, despite the large topological and geometric variations.
- We refer to such ability to interact with unseen objects within a certain category as generalizable manipulation skills.
Object-Level Generalization

Generalize Over
Object Geometry / Topology / Appearance
Object-Level Generalization

- Require to discover knowledge about object structure
  - Part? Keypoint? Skeleton? …

Generalize Over
Object Geometry / Topology / Appearance
What Is ManiSkill?
What Is ManiSkill?

- A benchmark for learning **generalizable manipulation skills**
What Is ManiSkill?

● A benchmark for learning generalizable manipulation skills
● And it features
  ○ Four manipulation tasks targeting at distinct challenges of short-horizon physical manipulation skills
What Is ManiSkill?

- A benchmark for learning **generalizable manipulation skills**
- And it features
  - Four manipulation tasks targeting at distinct challenges of *short-horizon physical* manipulation skills
  - Diverse objects with large topology, geometry, and appearance variations
What Is ManiSkill?

- A benchmark for learning **generalizable manipulation skills**
- And it features
  - Four manipulation tasks targeting at distinct challenges of *short-horizon physical* manipulation skills
  - Diverse objects with large topology, geometry, and appearance variations
  - Ego-centric 3D visual inputs from a panoramic camera mounted on a robot
What Is ManiSkill?

- A benchmark for learning **generalizable manipulation skills**
- And it features
  - **Four manipulation tasks** targeting at distinct challenges of *short-horizon physical* manipulation skills
  - **Diverse objects** with large topology, geometry, and appearance variations
  - **Ego-centric 3D visual inputs** from a panoramic camera mounted on a robot
  - **Large-scale demonstration datasets**
Manipulation Skills with Distinct Challenges

Tasks feature distinct object motions and manipulation skill properties
• Motion Constrained by Revolute Joint
Open Cabinet Drawer

• Motion Constrained by Prismatic Joint
• Motion Constrained by Plane
• Complex Underactuated System
• Dual-arm Collaboration
• Motion w/o Constraints
• Dual-arm Collaboration
• Hard Balancing
Generalization Over Diverse Objects

We manually processed 162 objects over 4 categories
Large topology, geometry, and appearance variations
(Including objects that cannot be procedurally generated)
3D Visual Inputs from Panoramic Camera: Point Cloud / RGBD Image
Demonstrations Provided

~36,000 demo trajectories in total
Over 1.5 million 3D video frames

Current observation \( s_t \)
Action \( a_t \)
Next observation \( s_{t+1} \)
Reward \( r_t \)

\[ \{(s_t, a_t, s_{t+1}, r_t)\}_{t=1}^n \]
Training-Evaluation Protocol

● For each task
  ○ Agent is trained on training environments (with training objects)
  ○ Evaluated on test environments (with test objects)
  ○ Metric is mean success rate

![Diagram showing the training and evaluation process]
Interesting Research Problems

● How to design better networks for physical manipulation?
  ○ Recent experiments: SparseConvNet achieves better training and test performance
  ○ Many 3D Shape Analysis techniques can be utilized
    ■ Keypoint Extraction? Part Discovery? …

● How to combine offline demo with online learning?
  ○ What if multiple modalities in demo?

● How to mitigate the occlusion issues during manipulation?
  ○ Active perception?

● …
Outline

● ManiSkill Benchmark

● Summary of ManiSkill Challenge 2021

● Preview of ManiSkill Challenge 2022
ManiSkill Challenge 2021

- Total Awards: 20,000 USD
- Challenge Award Ceremony held on ICLR 2022 Workshop on Generalizable Policy Learning in the Physical World
- The workshop has also received many works on generalizable policy learning.
  - [https://openreview.net/group?id=ICLR.cc/2022/Workshop/GPL#accept-poster-](https://openreview.net/group?id=ICLR.cc/2022/Workshop/GPL#accept-poster-)
Three Competition Tracks

- Demonstrations provided, and 3 tracks to choose from:
  - **No interaction track**: Learn from demonstration only, no interaction with the environment. (welcome, 3DV experts!)
  - **No external annotation track**: Interaction allowed on top of demonstrations, no additional data and environment annotations. (welcome, RL experts!)
  - **No restriction track**: Do whatever you want to solve the problem. (welcome, robotics experts!)
Competition Summary

● Timeline
  ○ 07/29/2021: Registration opens
  ○ 08/09/2021: Submission opens
  ○ 08/20/2021: Evaluation starts
  ○ 01/14/2022: Challenge ends

● 40+ registered teams

● 7 teams have received the final awards

● Awards worth $20,000 given to winners (sponsored by Qualcomm AI)
Which track is the hardest?

- Metric: 0.5 train set perf + 0.5 test set perf
- For each track, we average scores across tasks (by best performing team). Std represents task difficulty variation for each track.

![Graph showing success rates for different tracks](image)
Which track is the hardest?

- Metric: 0.5 train set perf + 0.5 test set perf
- For each task, we average scores across tracks (by best performing team). Std represents task difficulty variation for each track.
Train-Test Performance Difference By Track

The chart shows the success rate for different conditions:
- No Interaction
- No External Annotation
- No Restriction

The y-axis represents the success rate, ranging from 0.0 to 1.0. The chart indicates that the success rate is higher for the 'No Restriction' condition compared to 'No Interaction' and 'No External Annotation'.
Train-Test Performance Difference By Track
Outline

● ManiSkill Benchmark
● Summary of ManiSkill Challenge 2021
● Preview of ManiSkill Challenge 2022
ManiSkill Challenge 2022

● Follow the form of the 2021 challenge
  ○ Focus on object-level-generalizability
  ○ Three tracks

● We will add new tasks and provide a much faster system.
ManiSkill Challenge 2022

● ManiSkill 2021
  ○ 4 tasks for articulated object manipulation

● ManiSkill 2022
  ○ ~10 tasks for rigid body
  ○ 5 tasks for articulated body
  ○ 5 tasks for deformable body
Deformable Body Tasks

Pouring  Writer  Filling  Excavation  Rolling pin
ManiSkill Challenge 2022

- Launch in **July, 2022**
- Please watch the twitter of our group account for further updates:

https://twitter.com/HaoSuLabUCSD
QA & Links & Contact

- SAPIEN & ManiSkill: [https://sapien.ucsd.edu](https://sapien.ucsd.edu)
- Group page: [https://cseweb.ucsd.edu/~haosu/](https://cseweb.ucsd.edu/~haosu/)
- Contacts
  - Zhiwei Jia [zjia@eng.ucsd.edu](mailto:zjia@eng.ucsd.edu)
  - Fanbo Xiang [fxiang@eng.ucsd.edu](mailto:fxiang@eng.ucsd.edu)
  - Jiayuan Gu [jigu@eng.ucsd.edu](mailto:jigu@eng.ucsd.edu)
  - Rui Chen [ruichenthalu14@gmail.com](mailto:ruichenthalu14@gmail.com)

https://twitter.com/HaoSuLabUCSD