Natural Language Can Help Bridge the Sim2Real Gap

RSS 2024

Albert Yu, Adeline Foote, Ray Mooney, Roberto Martín-Martín
UT Austin

albertyu@utexas.edu
Large-Scale Real-World Data
[RT-1/2/X, RSS 2023]

Digital Twins
[Li, et. al., CoRL 2022]

Large # Target Domain Demos
[Kormushev, et. al. ICRA 2011]
Source Domain (sim)

Natural language

Target Domain (real)
Domain Invariant Representations

“gripper holding bread above coaster.”

“gripper holding carrot above yellow square.”

CNN

Learned Representation Space

π head

robot action
Domain Invariant Representations

“gripper holding milk next to coaster.”

“gripper wrapping blender wire.”
Image-Language Pretraining

Language-regularized CNN pretraining variants

A. Regression-based

\[ f_{\text{sim}} \rightarrow \hat{f}_{\text{sim}} \rightarrow \hat{f}_{\text{lang}} \approx f_{\text{sim}} \rightarrow \hat{f}_{\text{sim}} \rightarrow \hat{f}_{\text{lang}} \approx f_{\text{sim}} \]

B. Distance-based

\[ f_{\text{sim}} \rightarrow \hat{f}_{\text{sim}} \rightarrow \hat{f}_{\text{lang}} \approx f_{\text{sim}} \rightarrow \hat{f}_{\text{sim}} \rightarrow \hat{f}_{\text{lang}} \approx f_{\text{sim}} \]

- "gripper holding bread over square"
- "gripper holding carrot over yellow mat"
Multitask, Multidomain Behavioral Cloning

Task Instruction:
"Place carrot on yellow mat."

robot state

concat

+ 

robot action

FC Layers
Sim2Real Results by Task: 100 real trajs.

Language can bridge wide sim2real gaps with deformable objects.

Each bar: 2 seeds, 10 trials/seed

- No PT (real)
- No PT (sim+real)
- MMD
- CLIP
- R3M
- Ours (Lang Reg)
- Ours (Lang Dist)
Sim2Real Results by Task: 50 real trajs.

- Stack Object
- 2-step Pick and Place
- Wrap Wire

Each bar: 2 seeds, 10 trials/seed
Sim2Real Results by Task: 25 real trajs.

Each bar: 2 seeds, 10 trials/seed

Stack Object
- No PT (real): 20
- No PT (sim+real): 35
- MMD: 40
- CLIP: 30
- R3M: 25
- Ours (Lang Reg): 20
- Ours (Lang Dist): 5

2-step Pick and Place
- No PT (real): 45
- No PT (sim+real): 25
- MMD: 10
- CLIP: 15
- R3M: 20
- Ours (Lang Reg): 0
- Ours (Lang Dist): 5

Wrap Wire
- No PT (real): 45
- No PT (sim+real): 35
- MMD: 20
- CLIP: 25
- R3M: 5
- Ours (Lang Reg): 5
- Ours (Lang Dist): 5
2-step Pick and Place

Our $\pi$ rollouts ($5 \times$)

Successes

Failures
Wrap Wire
Our $\pi$ rollouts (5×)

Successes

Failures
Conclusion

• **Language can bridge the sim2real gap** with domain-invariant representations with 25-100 real-world demonstrations.

• **Our method enables leveraging low-fidelity sim data for sim2real transfer on deformable objects.**
Language

Domain-Invariant Representations

Sim

Real

PDF: https://arxiv.org/pdf/2405.10020
Website: https://robin-lab.cs.utexas.edu/lang4sim2real/