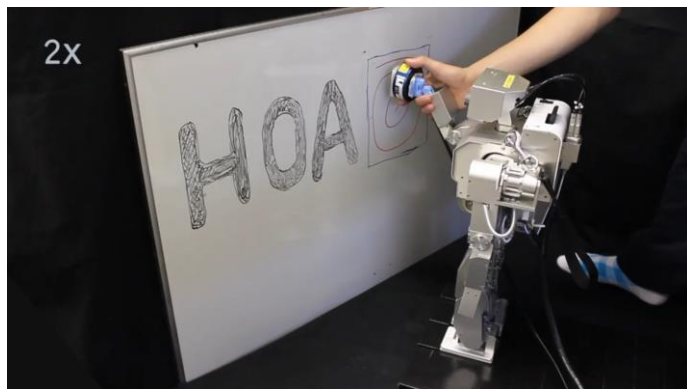
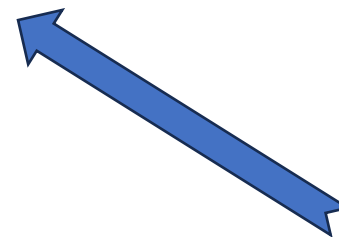
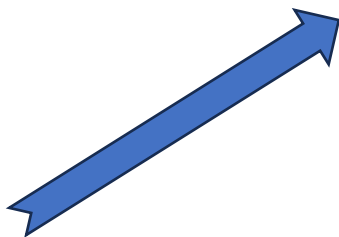


Natural Language Can Help Bridge the Sim2Real Gap

RSS 2024

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

albertyu@utexas.edu



Large # Target Domain Demos
[Kormushev, et. al. ICRA 2011]

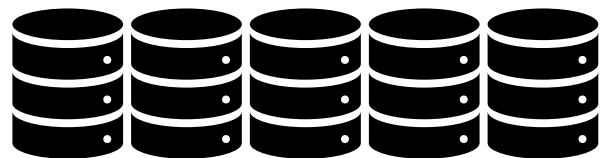


Large-Scale Real-World Data
[RT-1/2/X, RSS 2023]



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

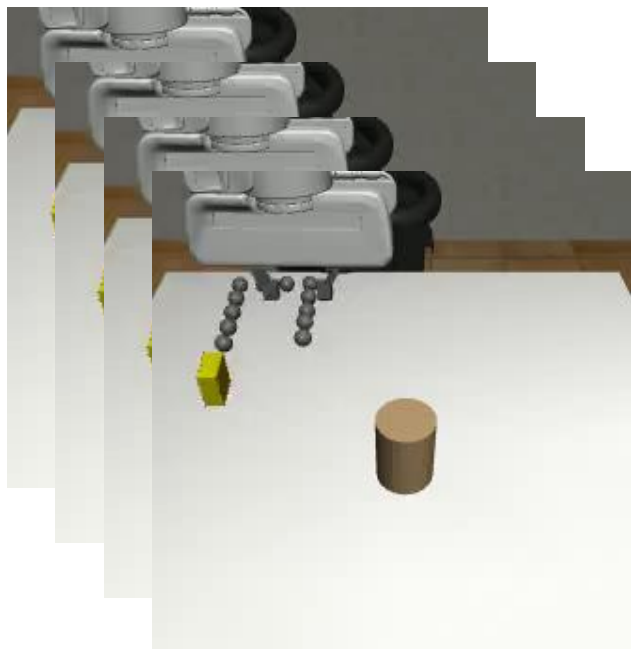
Source Domain (sim)



Target Domain (real)

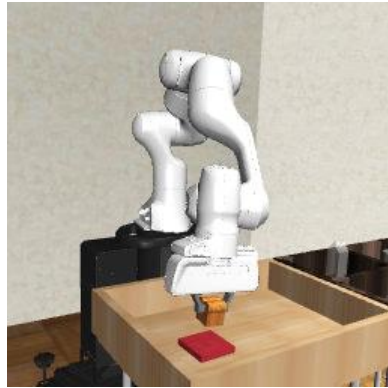


Natural language



Domain Invariant Representations

*“gripper holding bread
above coaster.”*

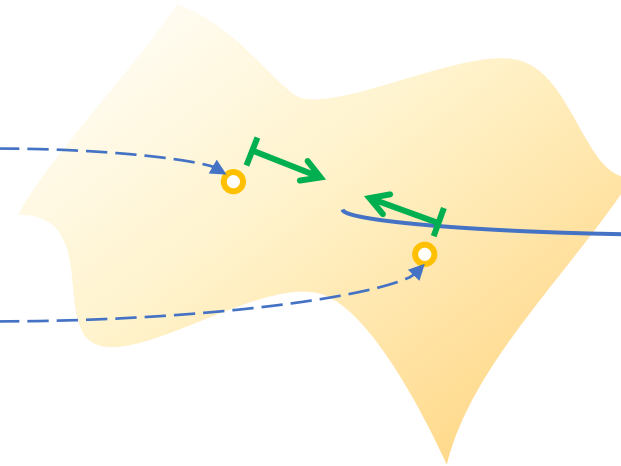


*“gripper holding carrot
above yellow square.”*

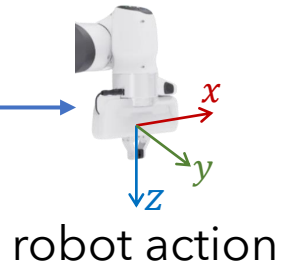


CNN

Learned Representation Space

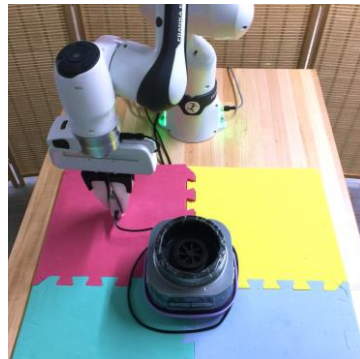
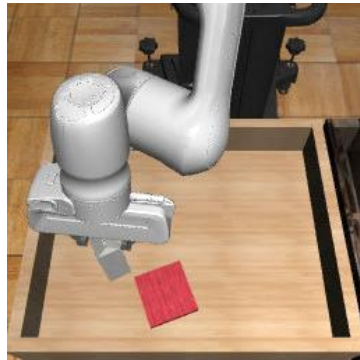


π
head



Domain Invariant Representations

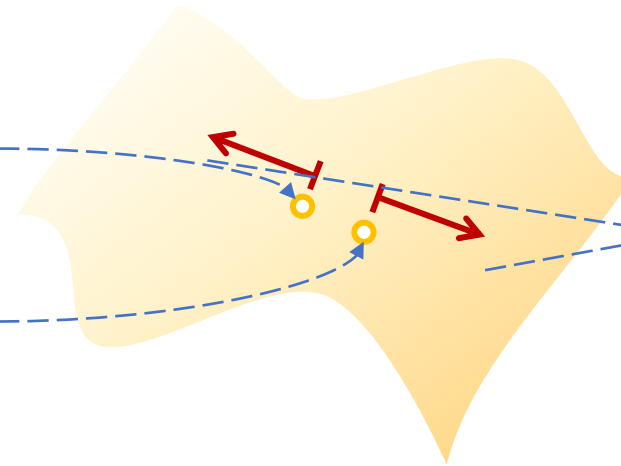
“gripper holding milk next to coaster.”



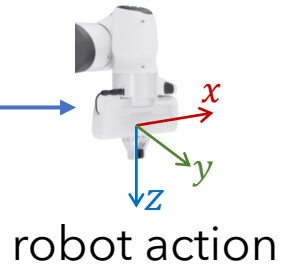
“gripper wrapping blender wire.”

CNN

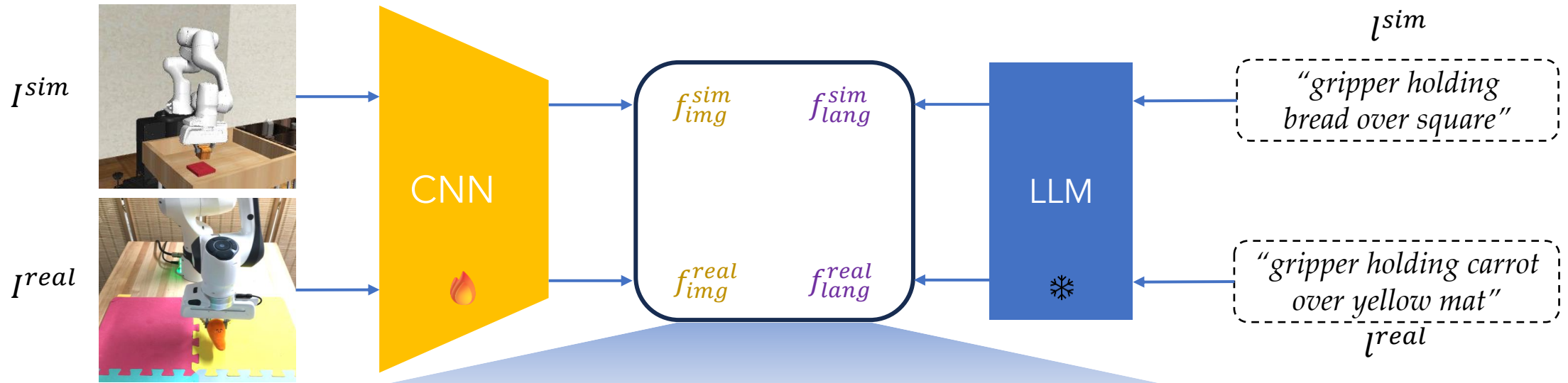
Learned Representation Space



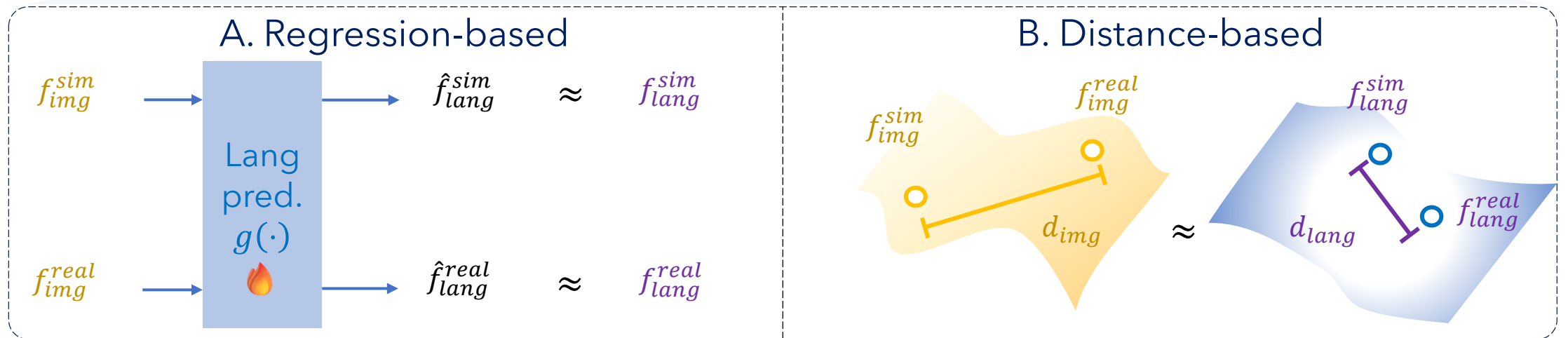
π
head



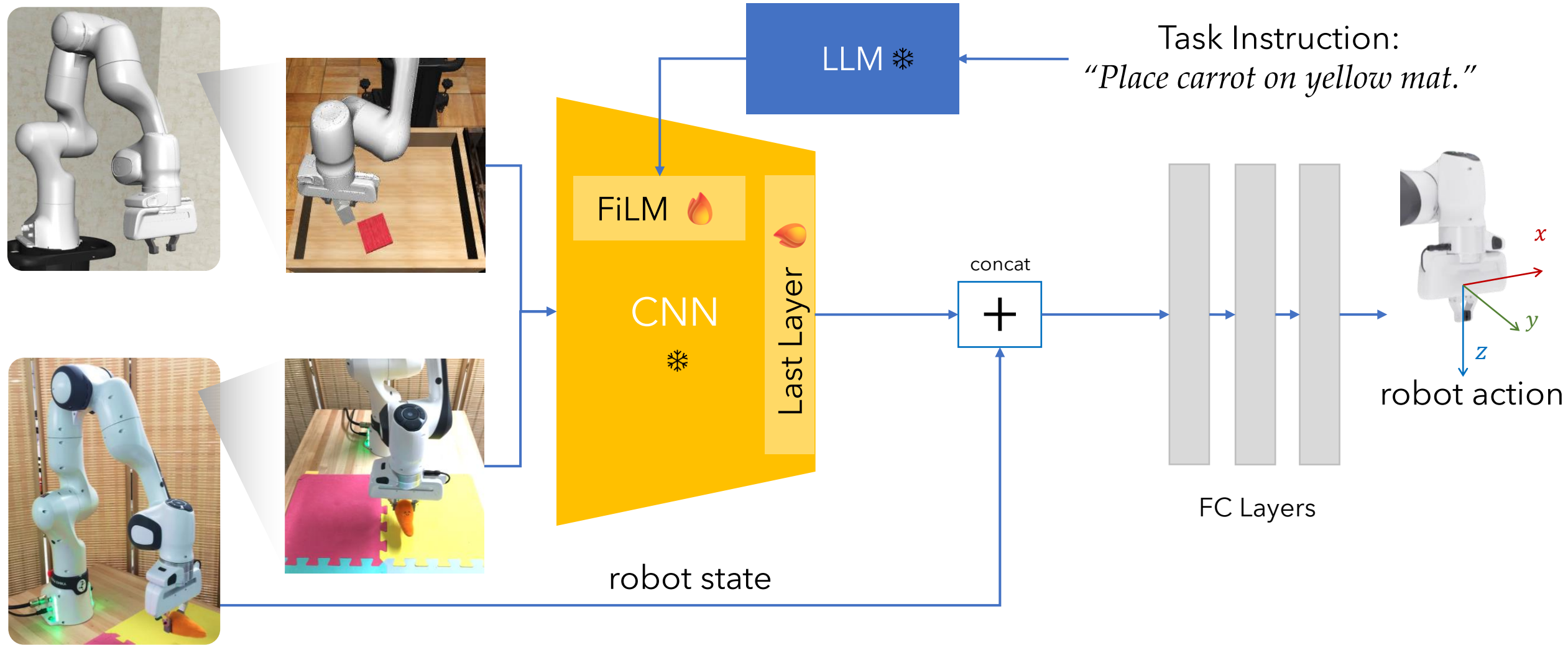
① Image-Language Pretraining



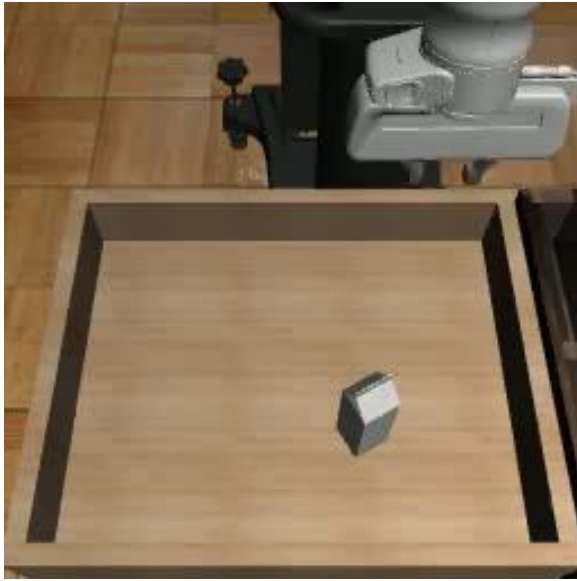
Language-regularized CNN pretraining variants



② Multitask, Multidomain Behavioral Cloning



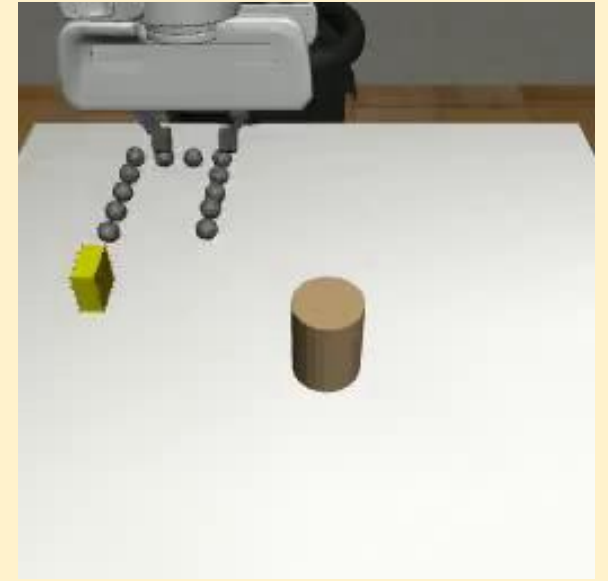
Stack Object



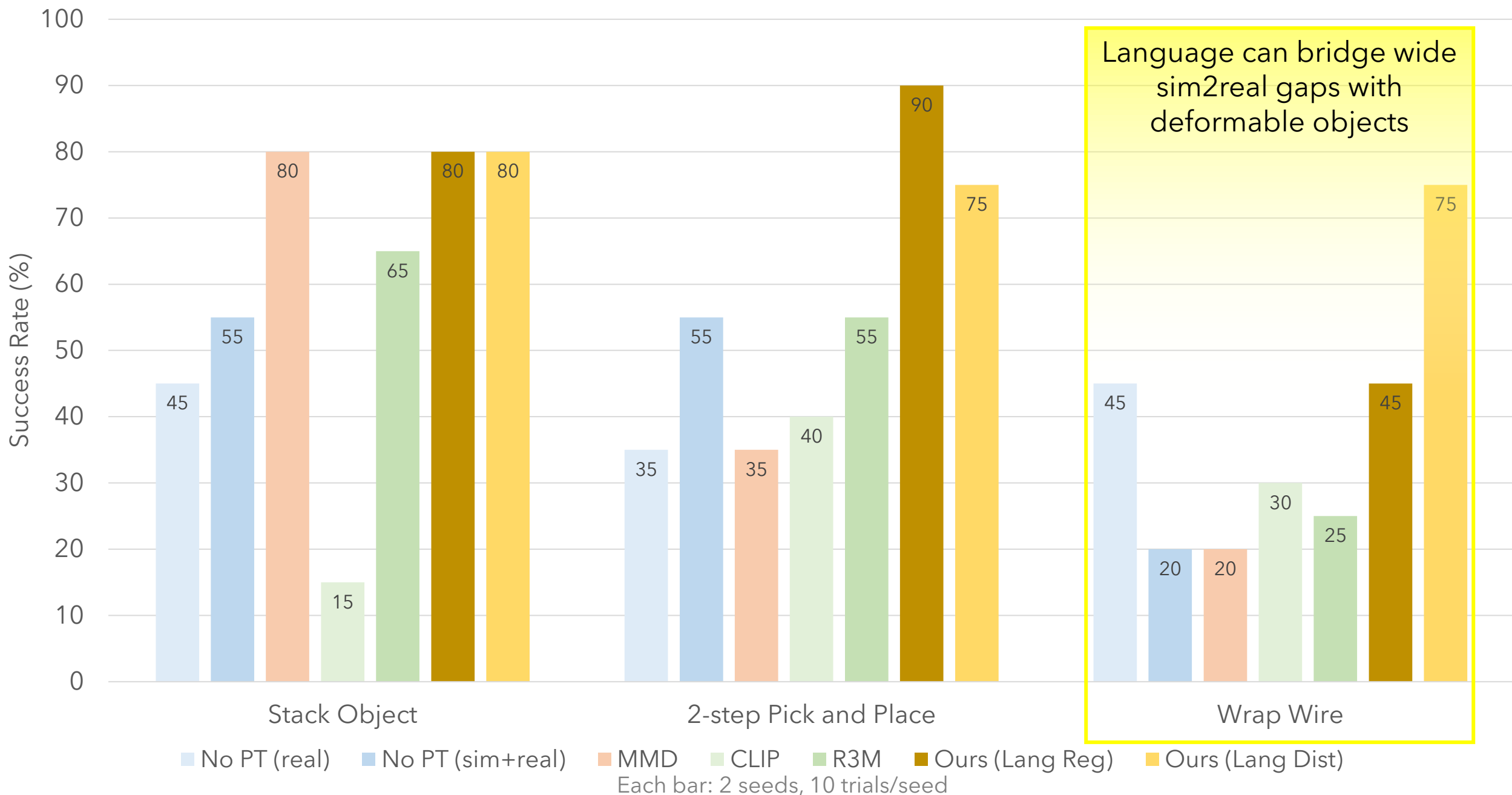
2-step Pick and Place



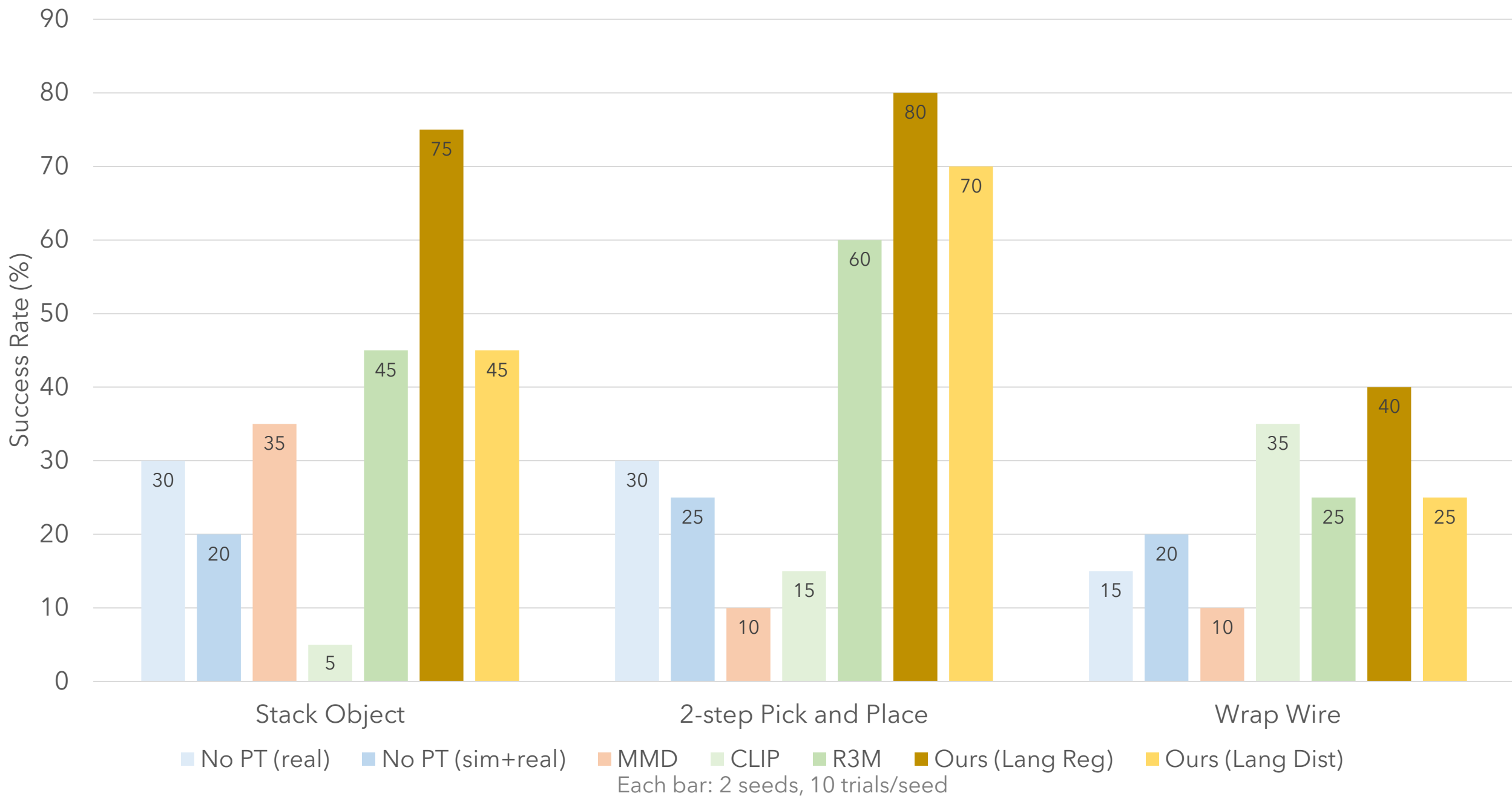
Wrap Wire



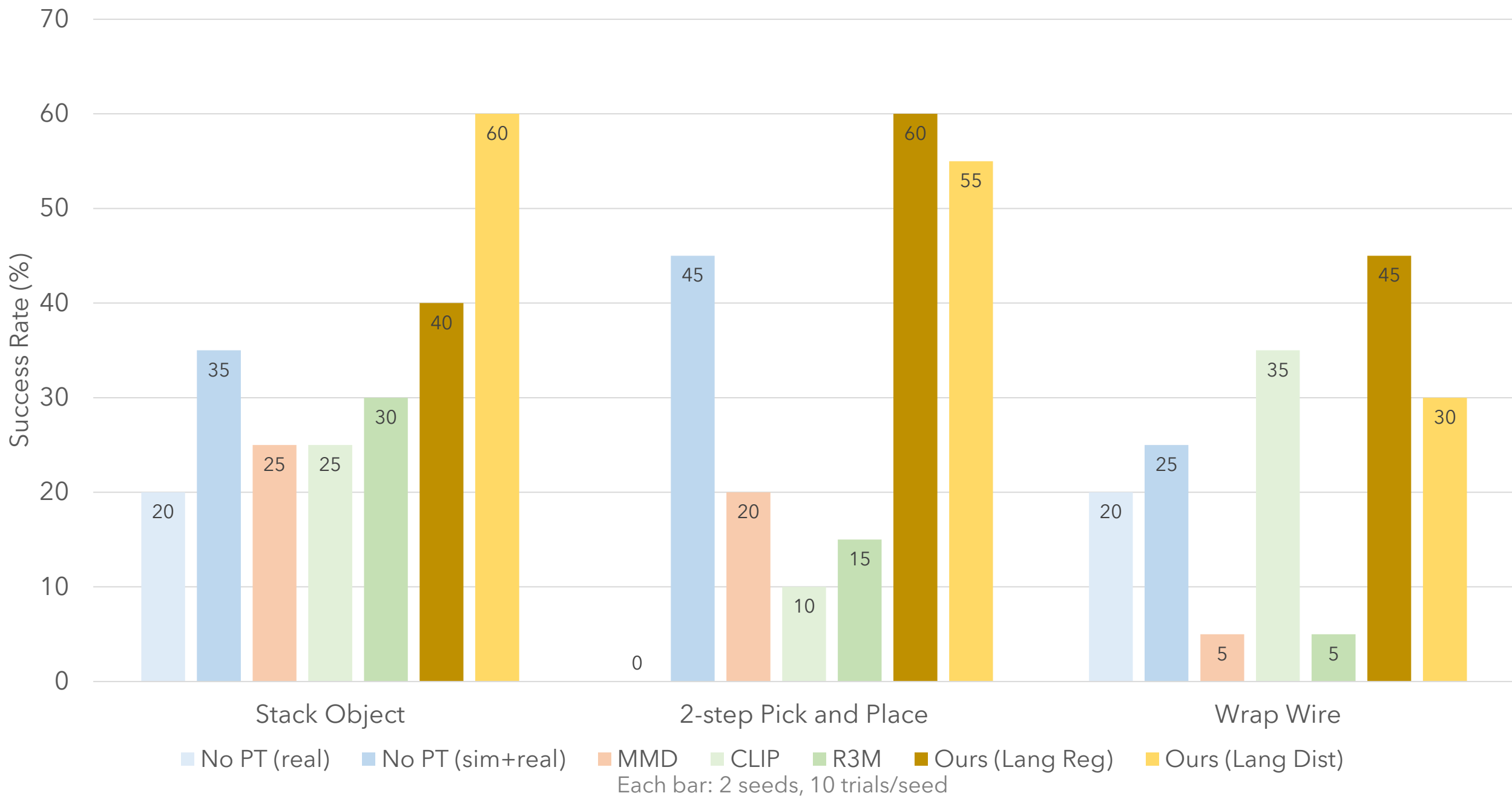
Sim2Real Results by Task: 100 real traj.



Sim2Real Results by Task: 50 real traj.



Sim2Real Results by Task: 25 real traj.

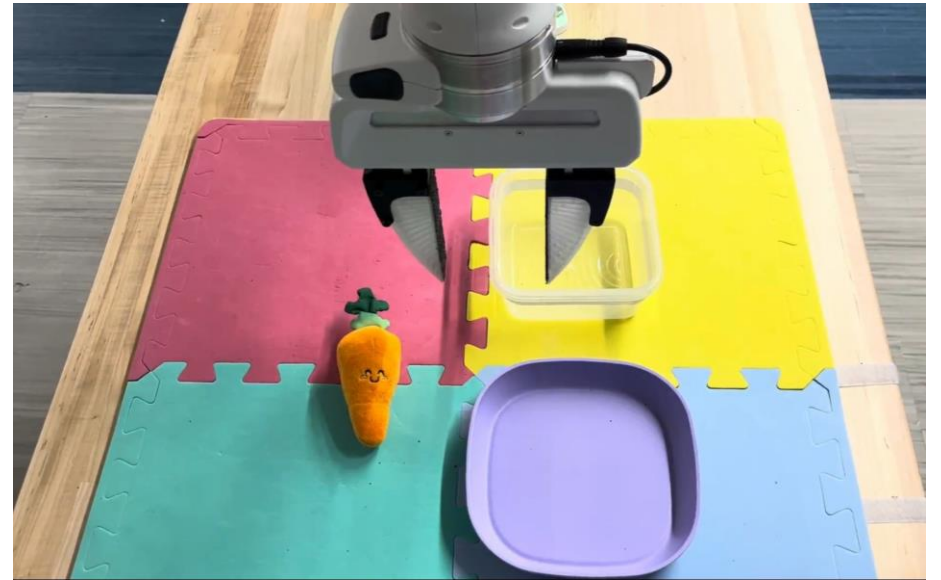
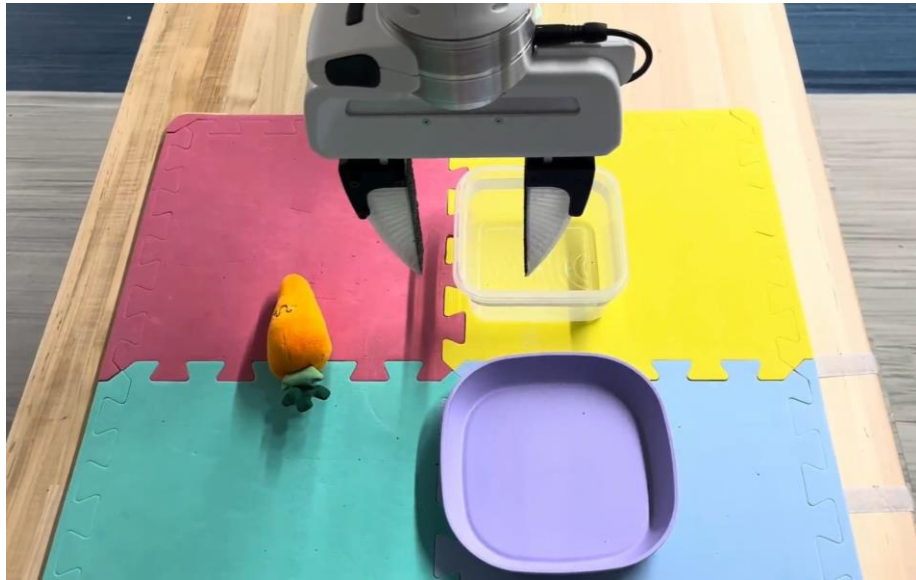


2-step Pick and Place

Our π rollouts (5 \times)

Successes

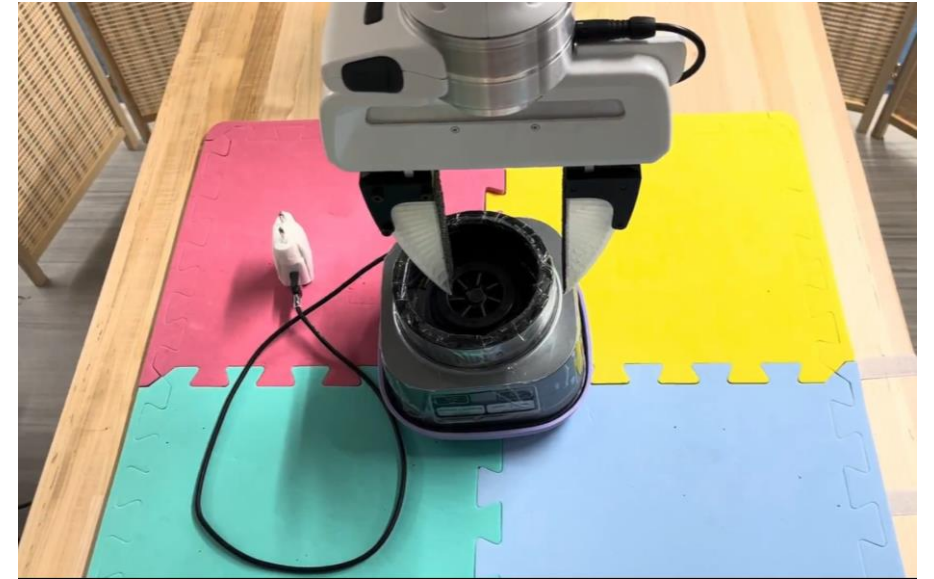
Failures



Wrap Wire

Our π rollouts (5x)

Successes



Failures



Conclusion

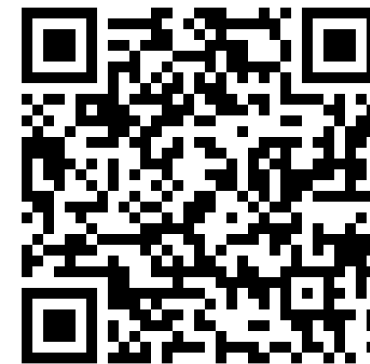
- **Language can bridge the sim2real gap** with domain-invariant representations with 25-100 real-world demonstrations.
- **Our method enables leveraging low-fidelity sim data** for sim2real transfer on deformable objects.



PDF

PDF: <https://arxiv.org/pdf/2405.10020>

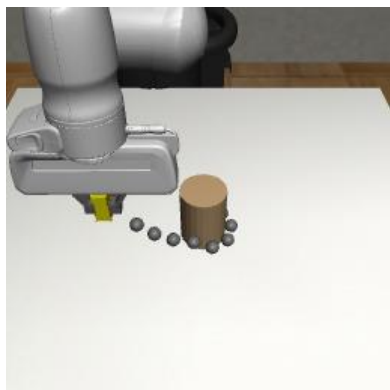
Website: <https://robin-lab.cs.utexas.edu/lang4sim2real/>



Website

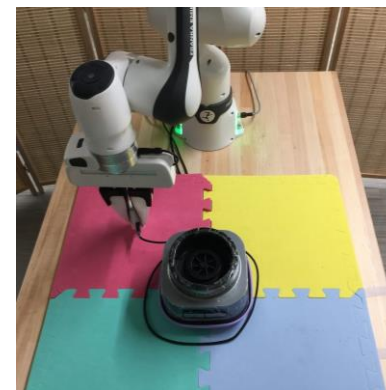
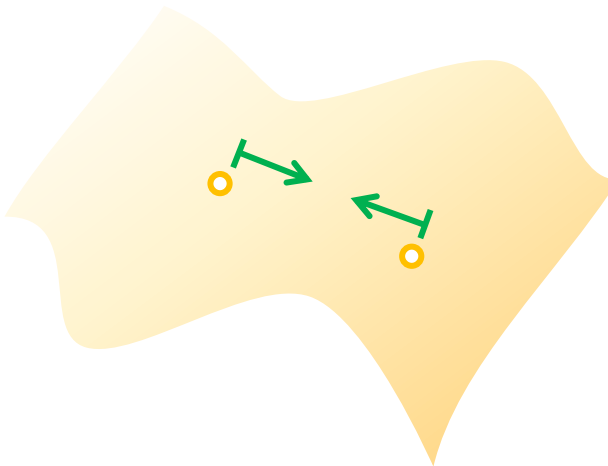


Language



Sim

Domain-Invariant Representations



Real