Object Assembly, a Spatial-Geometric Reasoning Pathway to Physical Intelligence

Guest Lecture – CS6301 – UT Dallas

Karthik Desingh
Assistant Professor, University of Minnesota
Minnesota Robotics Institute (MnRI)
Department of Computer Science and Engineering

Robotics:
Perception & Manipulation
(RPM) Lab

Why is Object Assembly such an important task to focus on in robotics?

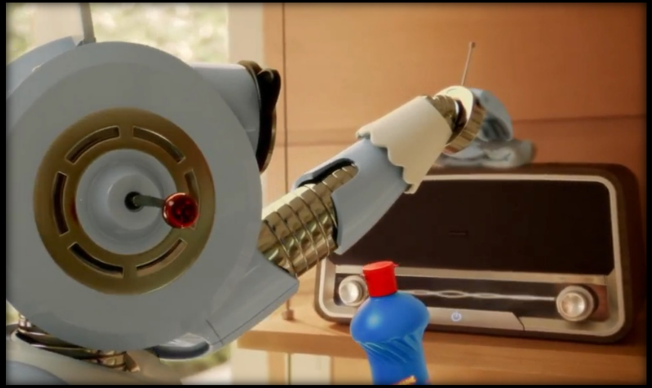
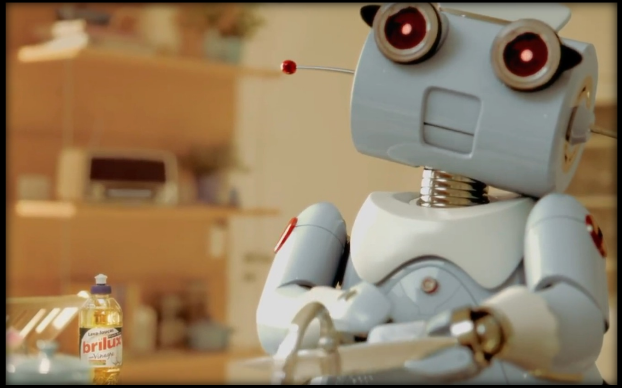
Why is Object-Object Interaction such an important task to focus on in robotics?

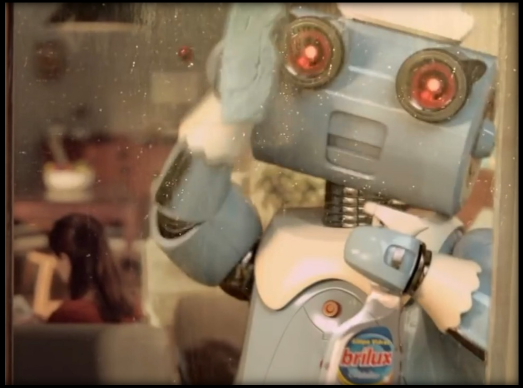
General Purpose Robot - *the dream*

[Brilux TV commercial]

Why is Object-Object Interaction such an important task to focus on in robotics?





Why is Object Assembly such an important task to focus on in robotics?

… could lead to robot's physical intelligence.

A form of physical intelligence is where the agent is able to interact with novel objects seamlessly.

We posit that **object assembly** task could lead to this **physical intelligence**.

How do we enable robot to perform object assembly tasks?

Francisco Suárez-Ruiz *et al.* Can robots assemble an IKEA chair?.
Sci. Robot. **3**,eaat6385(2018).DOI:10.1126/scirobotics.aat6385

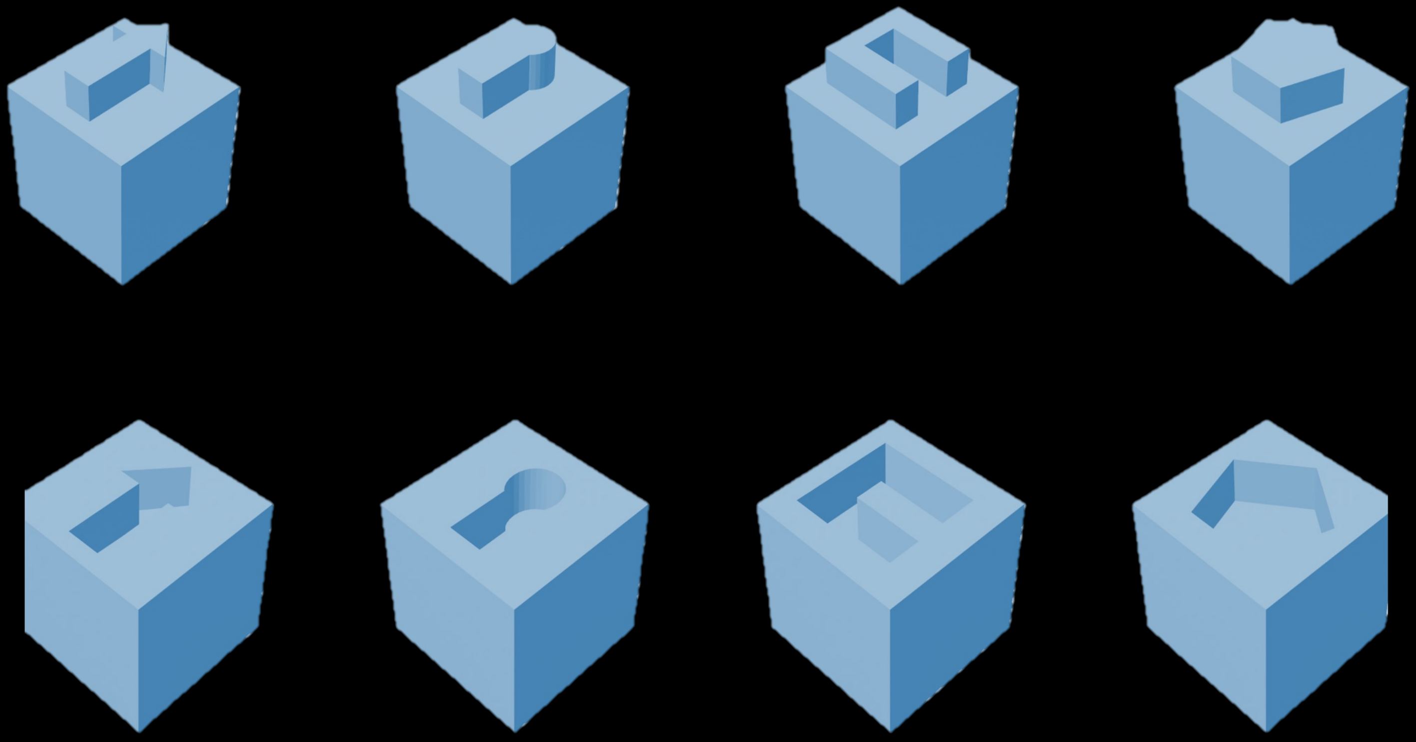
How do we enable robot to perform object assembly tasks on novel objects?

Evidence from neuroscience and cognitive science supports the notion that humans employ spatio-geometric features, mediated by specific neural pathways and cognitive processes, to perform object assembly tasks.

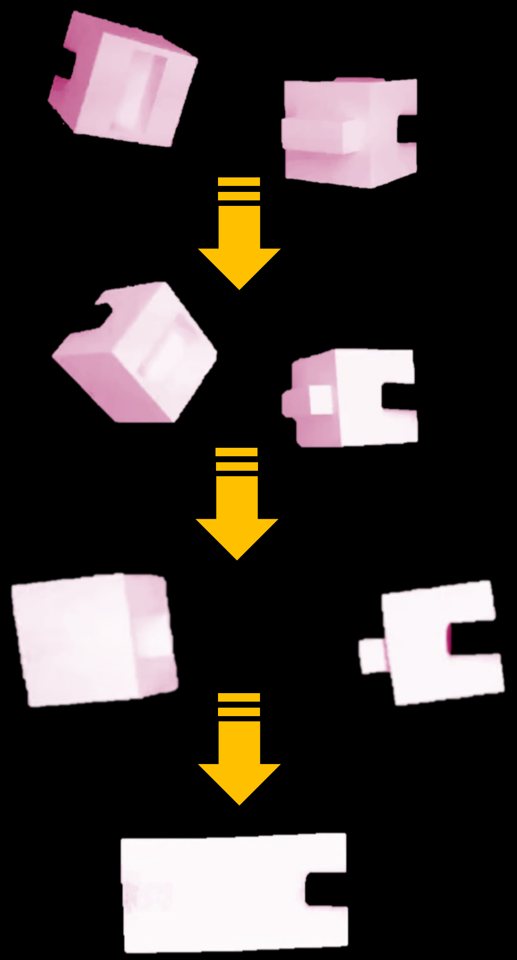
Evidence from neuroscience and cognitive science supports the notion that humans employ spatio-geometric features, mediated by specific neural pathways and cognitive processes, to perform object assembly tasks.

We posit that robots need representations that can capture spatio-geometric features to learn novel-object assembly skills from demonstrations.

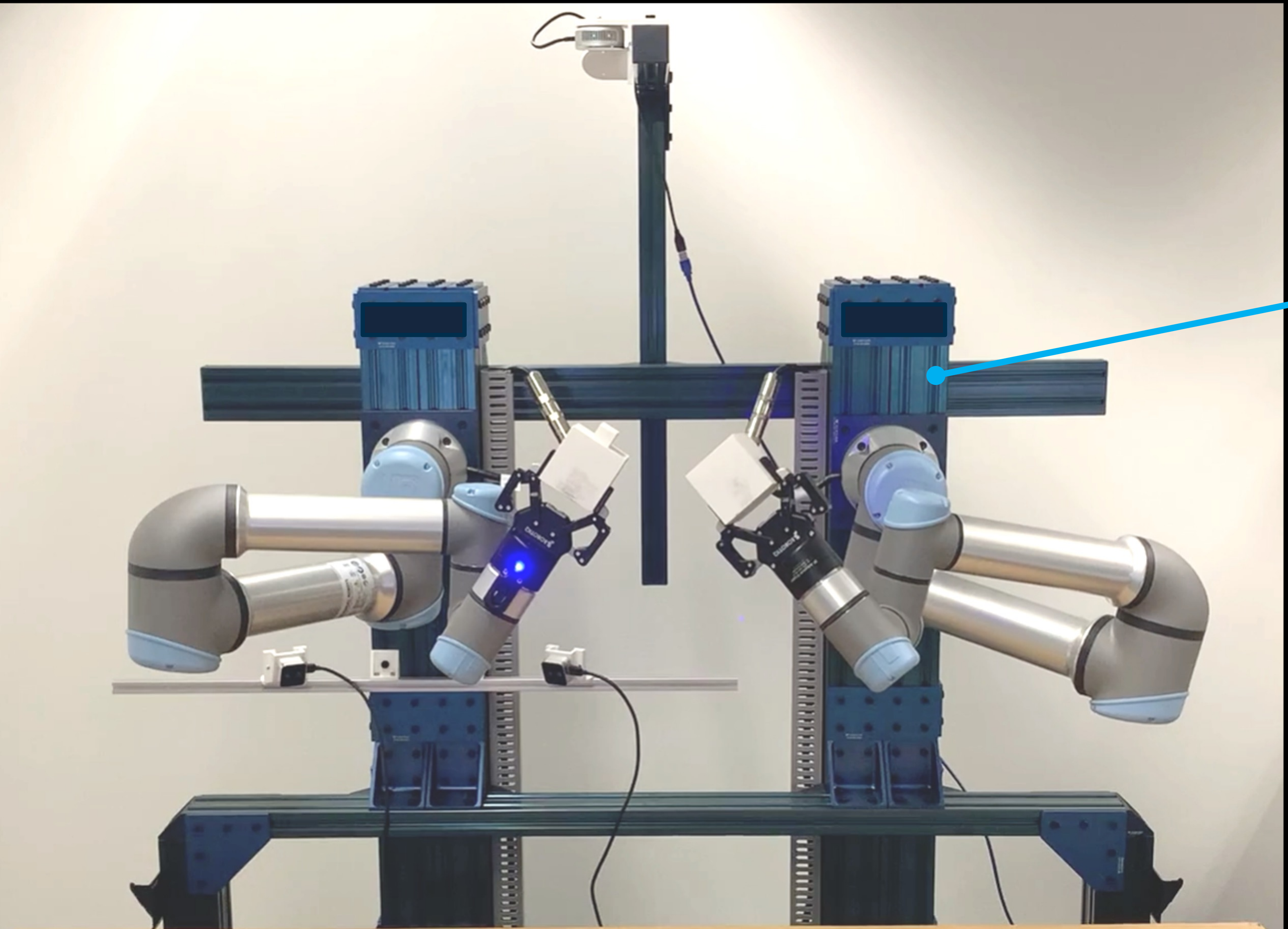
Inspired by LEGO puzzles we designed object pairs with various peg and hole geometries



Assembly Process



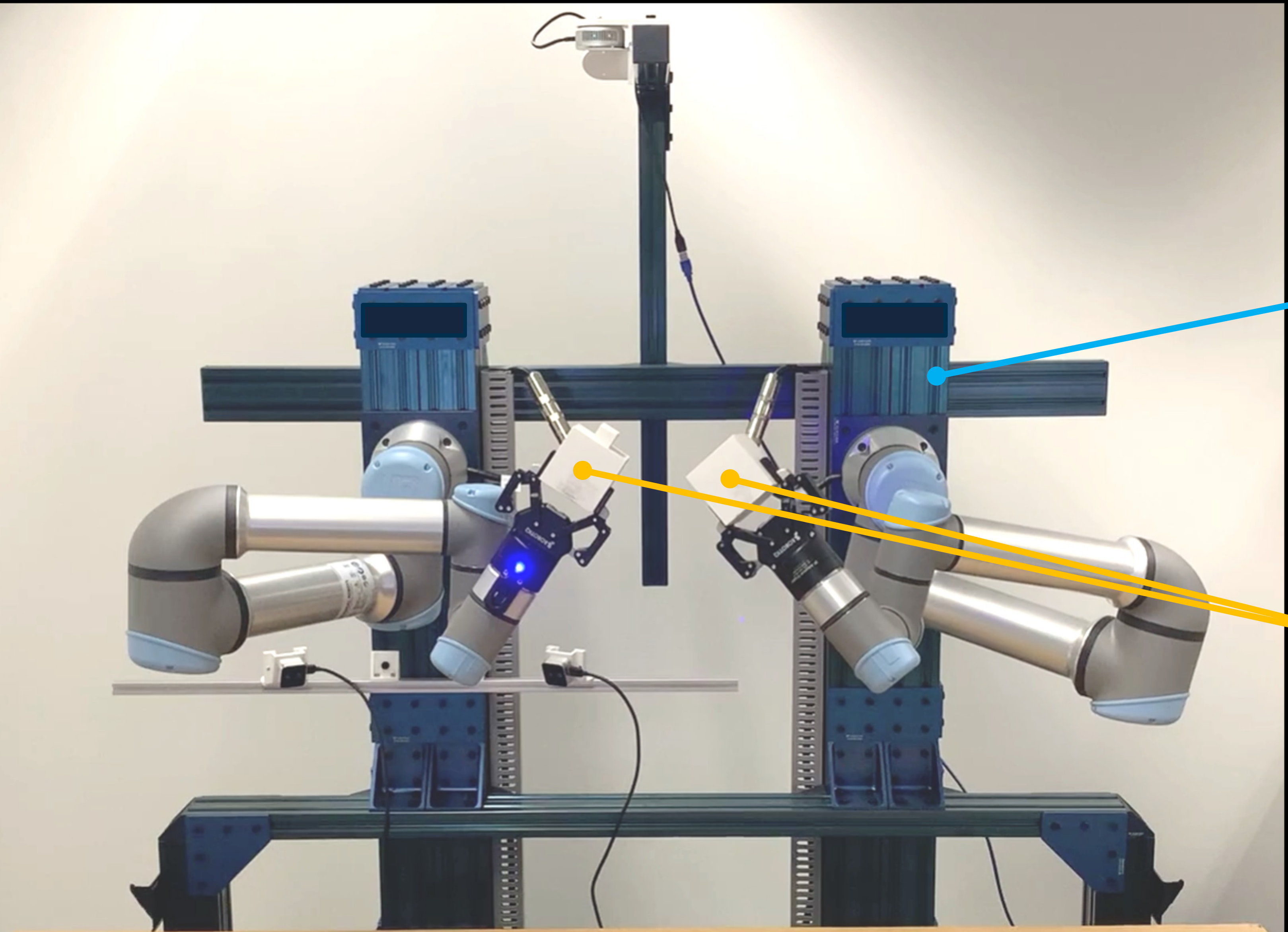
Task: Geometry Informed Object Assembly



Dual-arm Robot is tasked to assemble the object parts held by its grippers

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Task: Geometry Informed Object Assembly

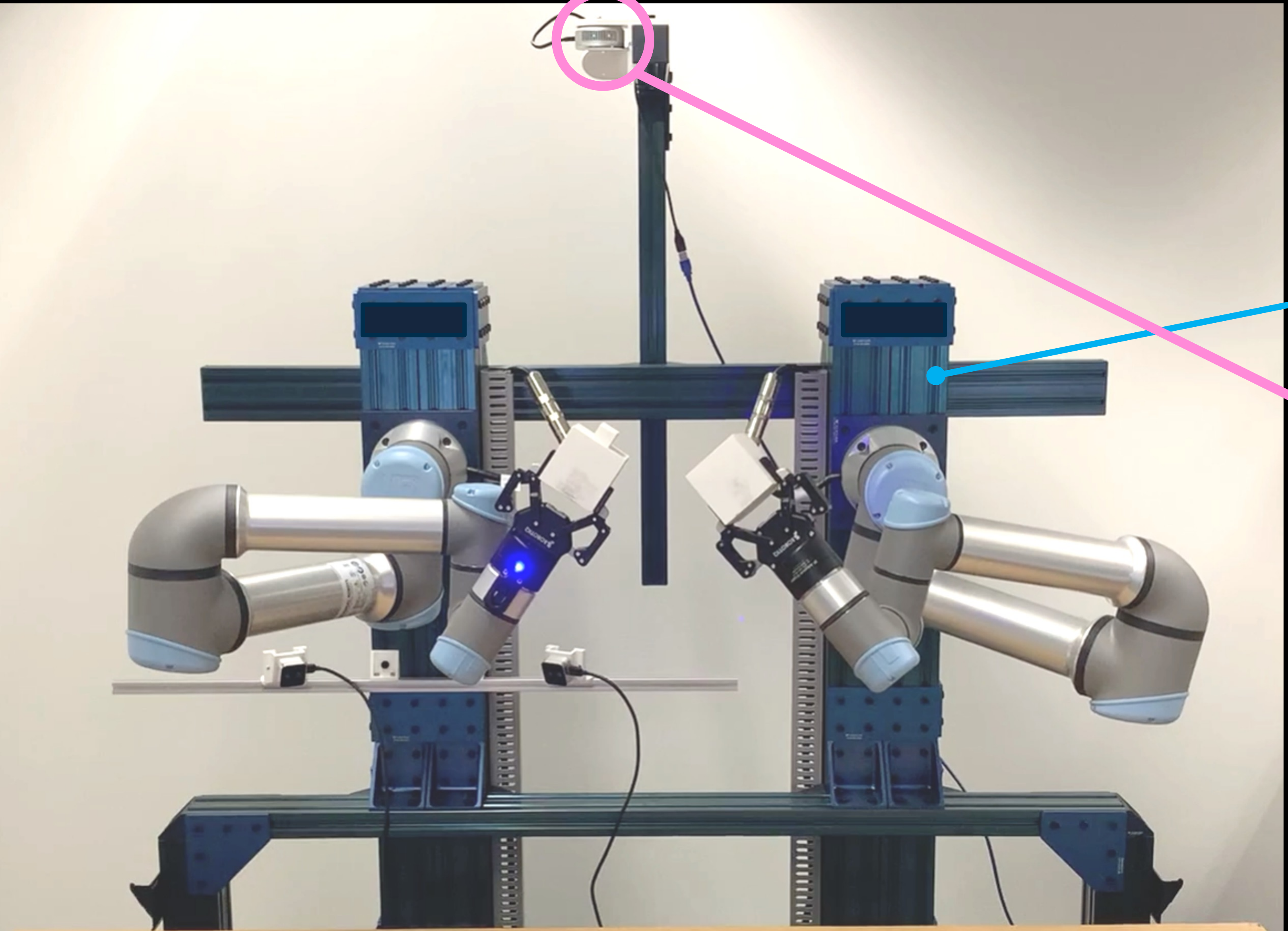


Dual-arm Robot is tasked to assemble the object parts held by its grippers

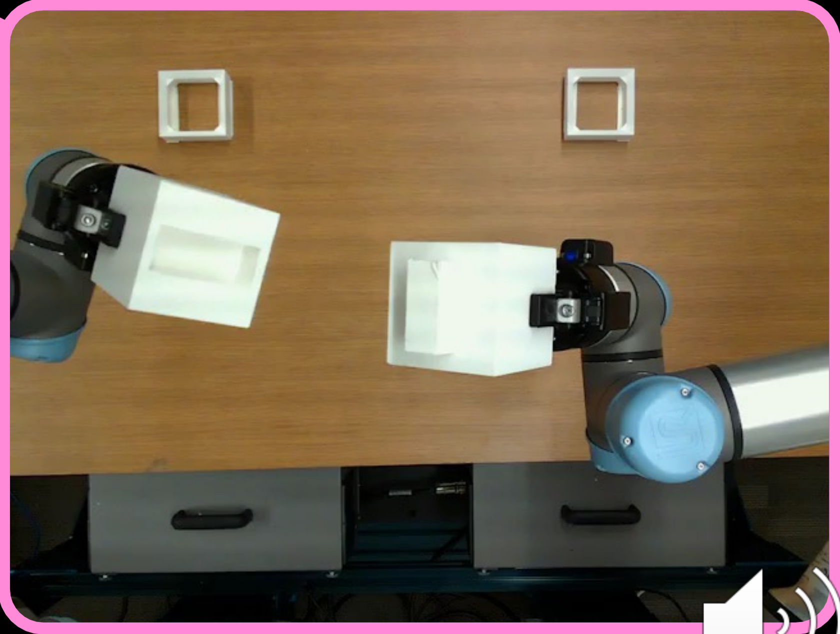
Pair of object parts with extruded and intruded geometries

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Task: Geometry Informed Object Assembly



Dual-arm Robot is tasked to assemble the object parts held by its grippers



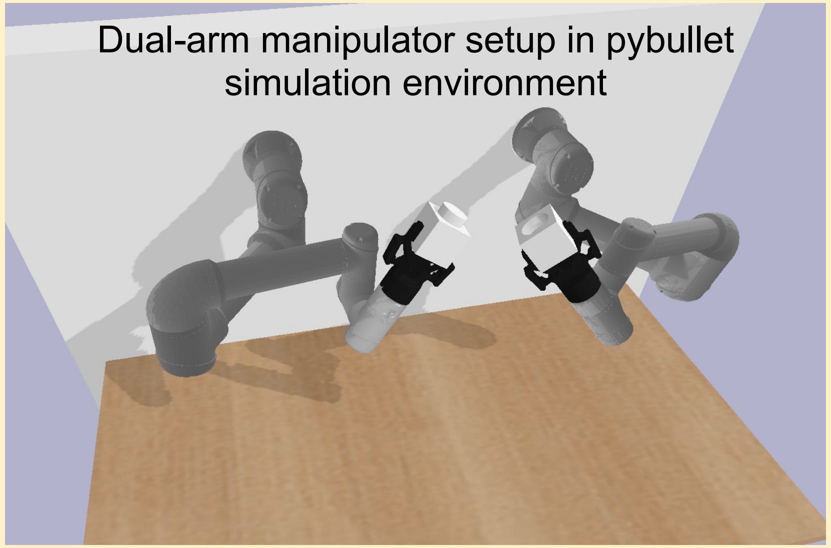
Sensor view

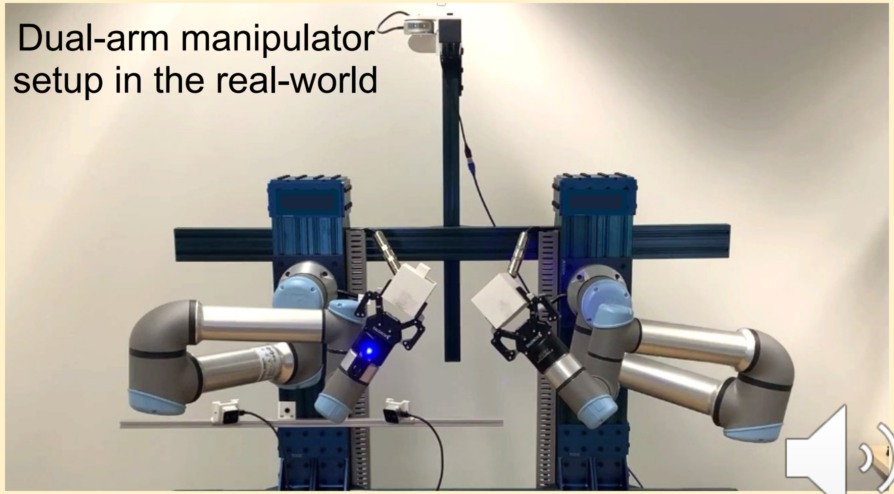
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Objectives of the project

Objectives of the project

To learn dual-arm manipulation policy for object assembly



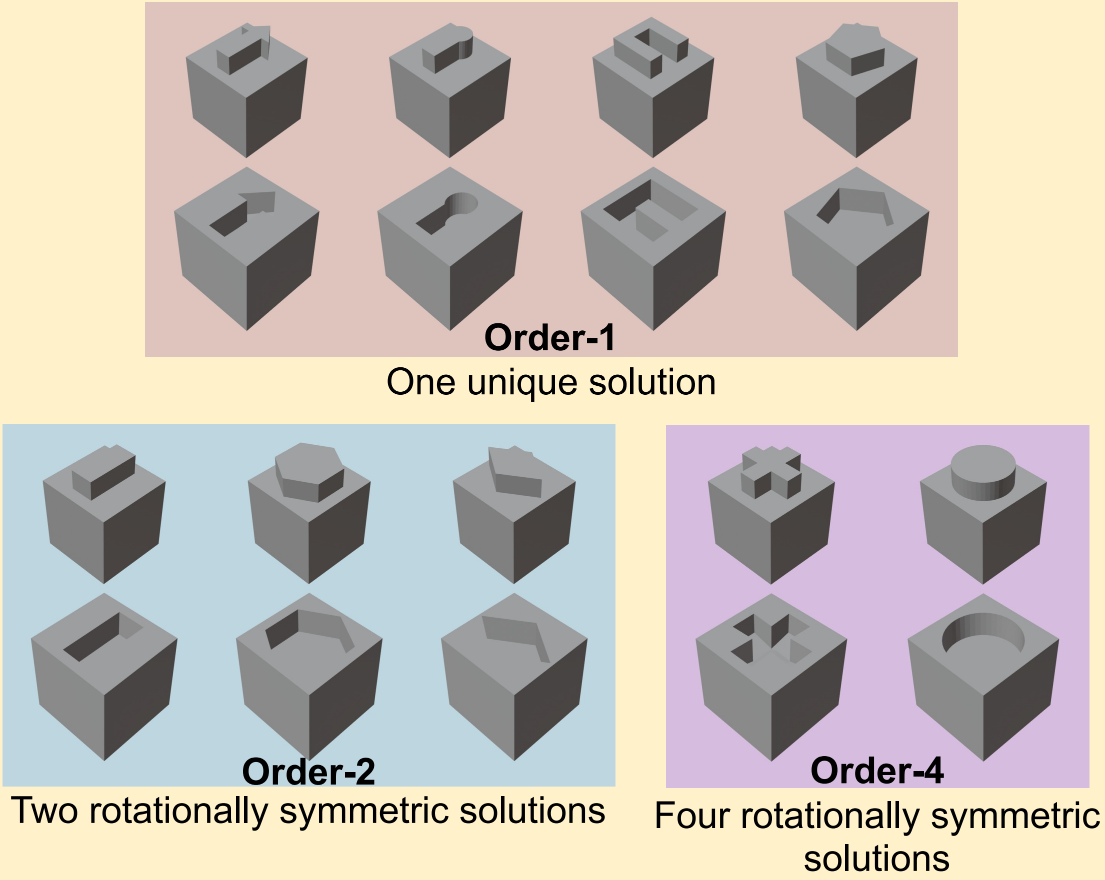


C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Objectives of the project

To learn dual-arm manipulation policy for object assembly

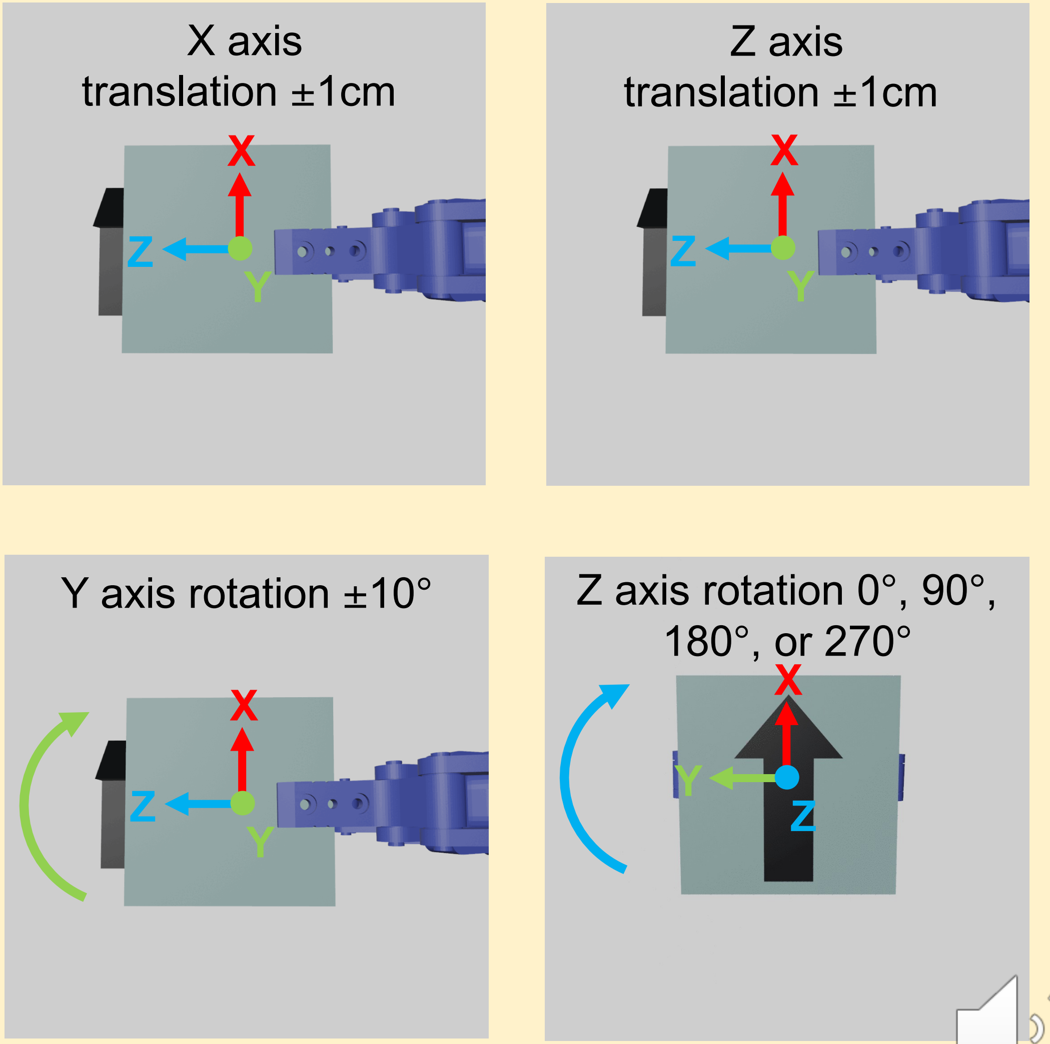
To implicitly perform spatio-geometric reasoning



C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Objectives of the project

- To learn dual-arm manipulation policy for object assembly
- To implicitly perform spatio-geometric reasoning
- To be robust to grasp variations



C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

We posit that robots need representations that can capture spatio-geometric features to learn novel-object assembly skills from demonstrations.

CLIP

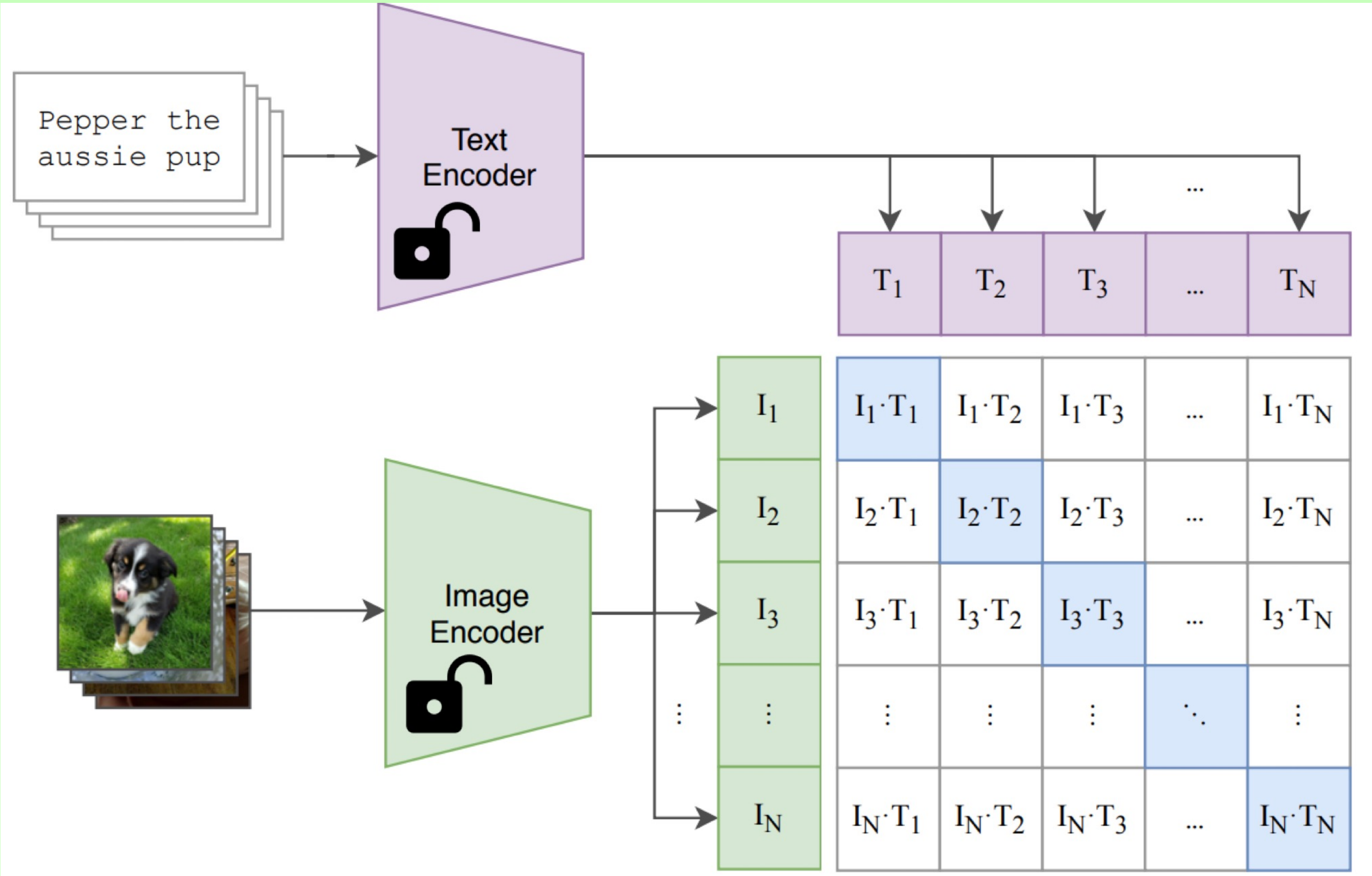
MAE

R3M

CLIP

MAE

R3M



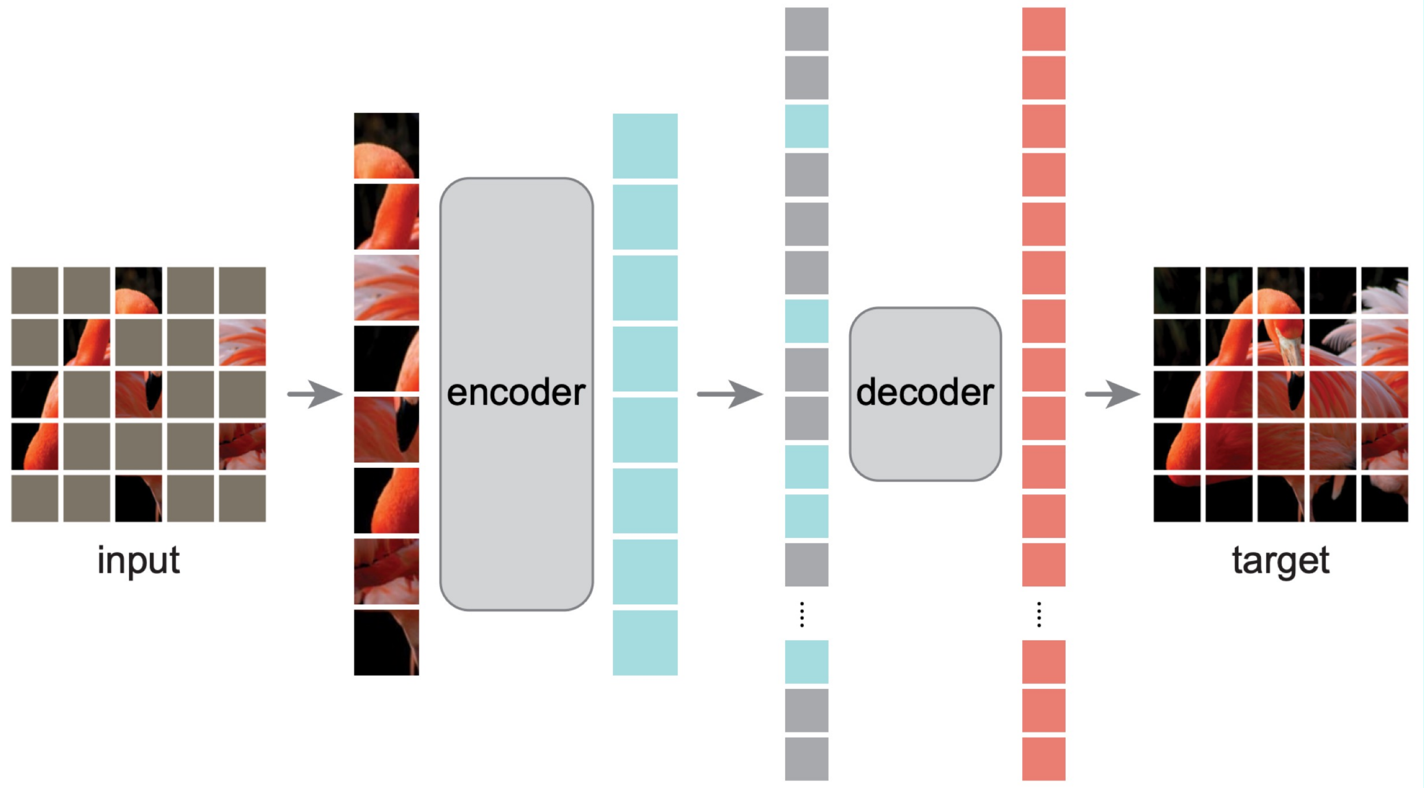
Contrastive **L**anguage **I**mage **P**re-training

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. Learning Transferable Visual Models From Natural Language Supervision. In International Conference on Machine Learning (ICML) 2021, Vol. 139. 8748–8763.

CLIP

MAE

R3M



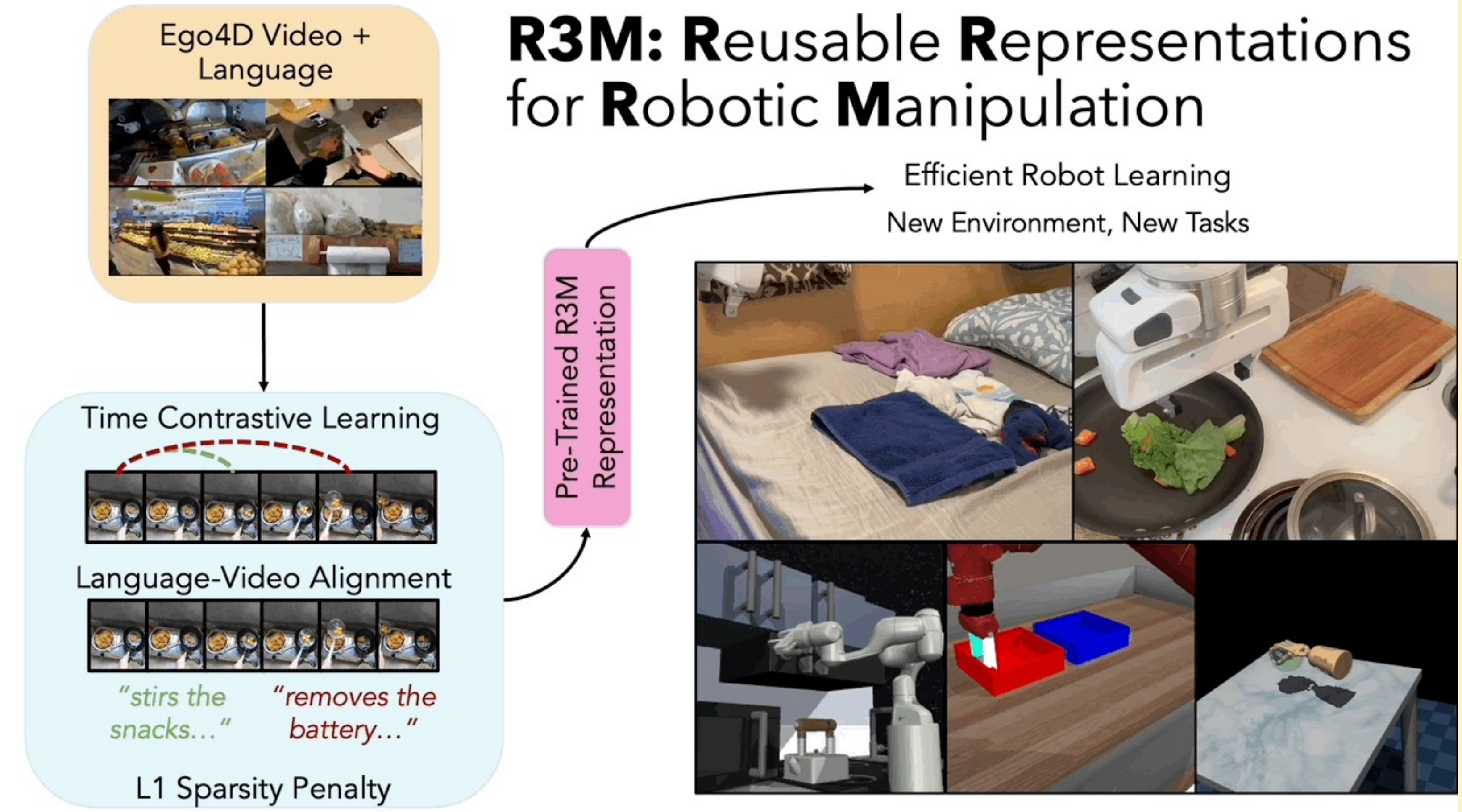
Masked **A**uto **E**ncoder

He, K., Chen, X., Xie, S., Li, Y., Dollár, P., & Girshick, R. (2022). Masked autoencoders are scalable vision learners. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)* (pp. 16000-16009).

CLIP

MAE

R3M



Nair, S., Rajeswaran, A., Kumar, V., Finn, C., & Gupta, A. R3M: A universal visual representation for robot manipulation. In Conference on Robot Learning (CoRL) 2022.

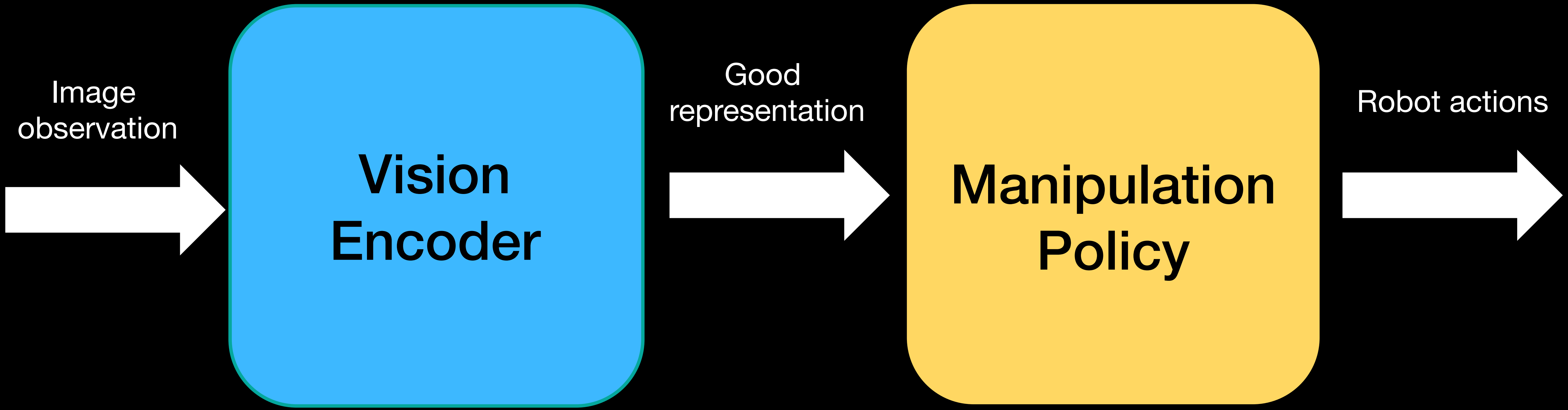
CLIP

MAE

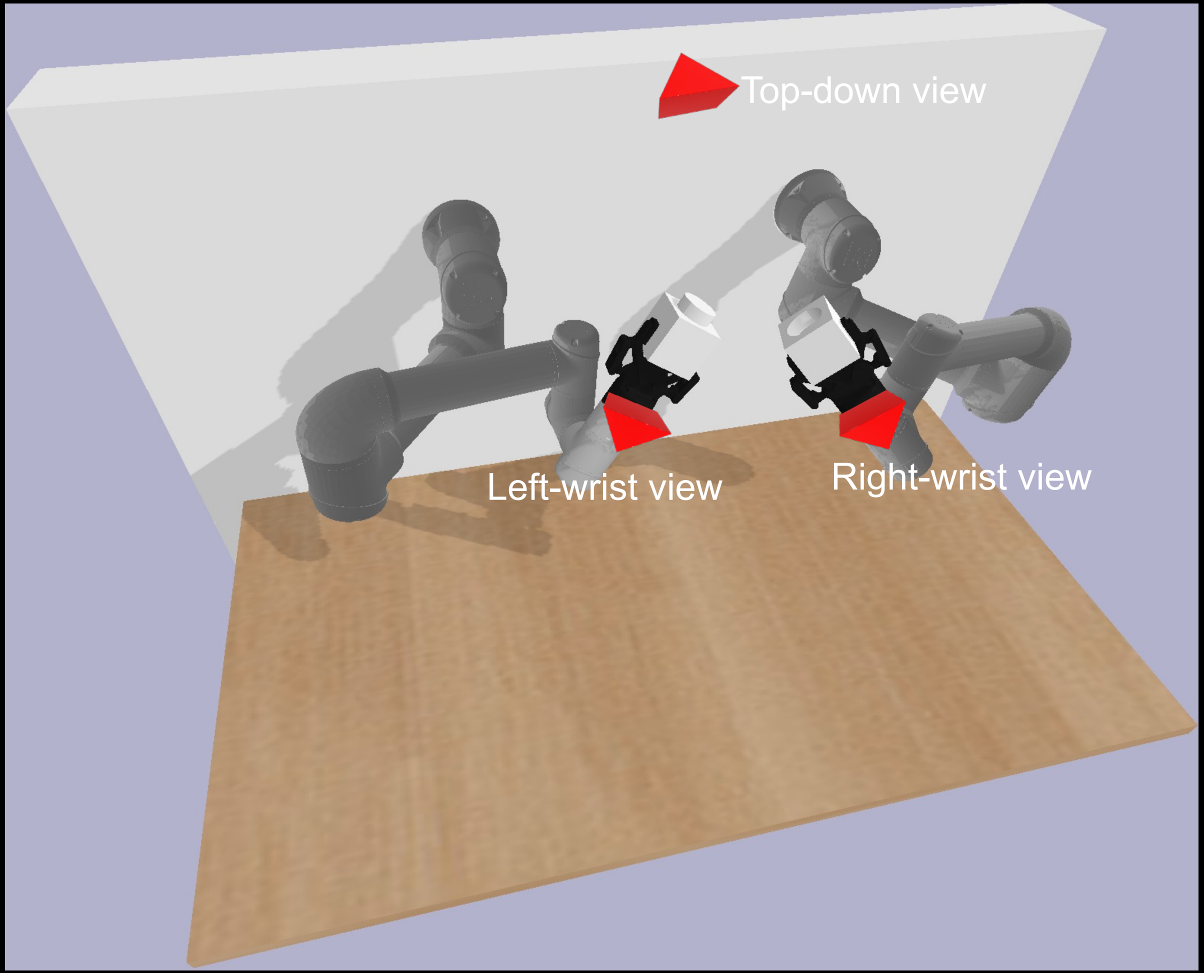
R3M

Vision
Encoder

Behavior cloning from demonstrations

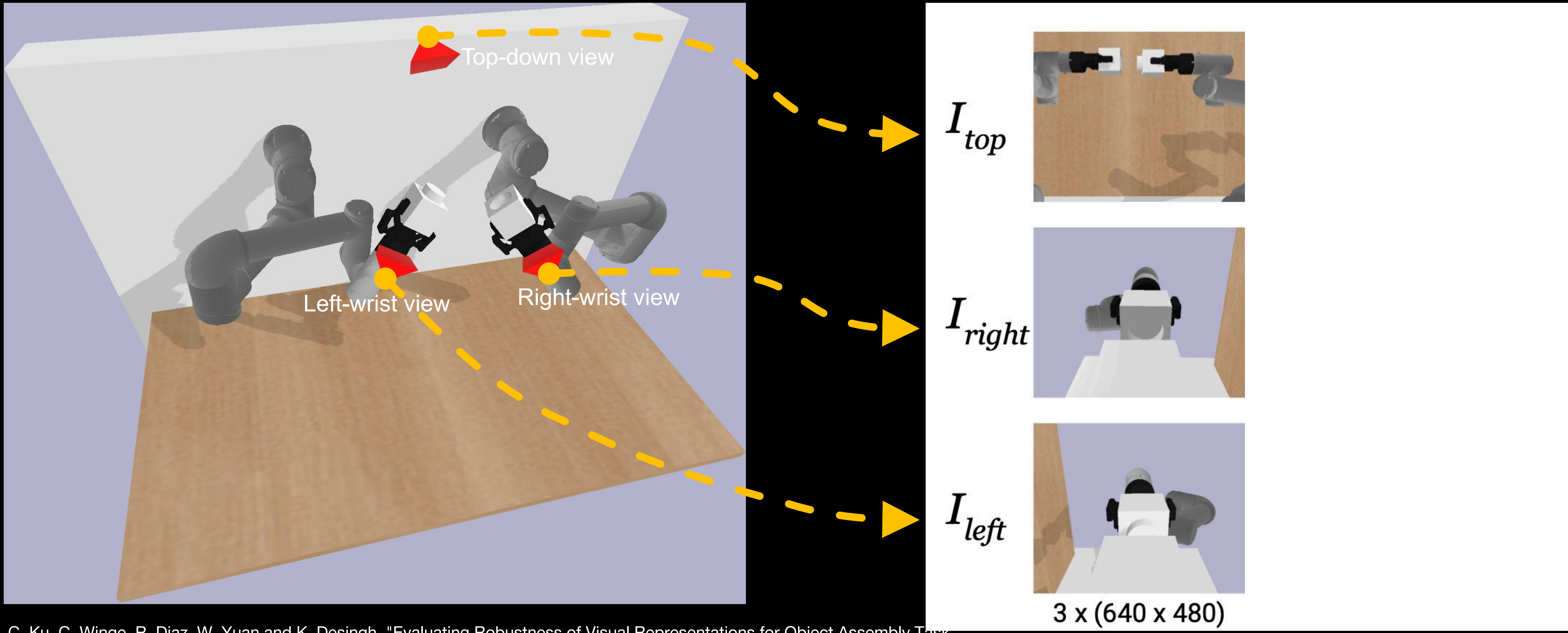


Dual-arm Manipulation Policy Learning Framework



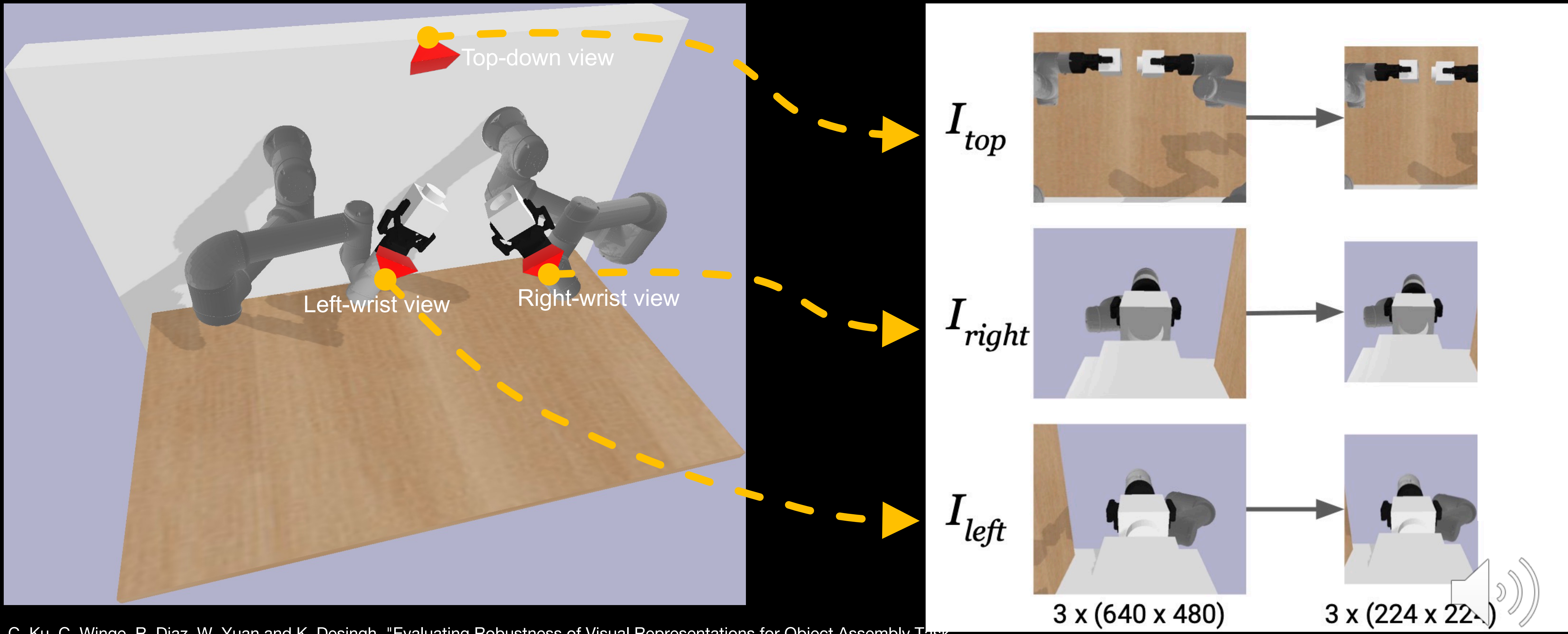
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Dual-arm Manipulation Policy Learning Framework



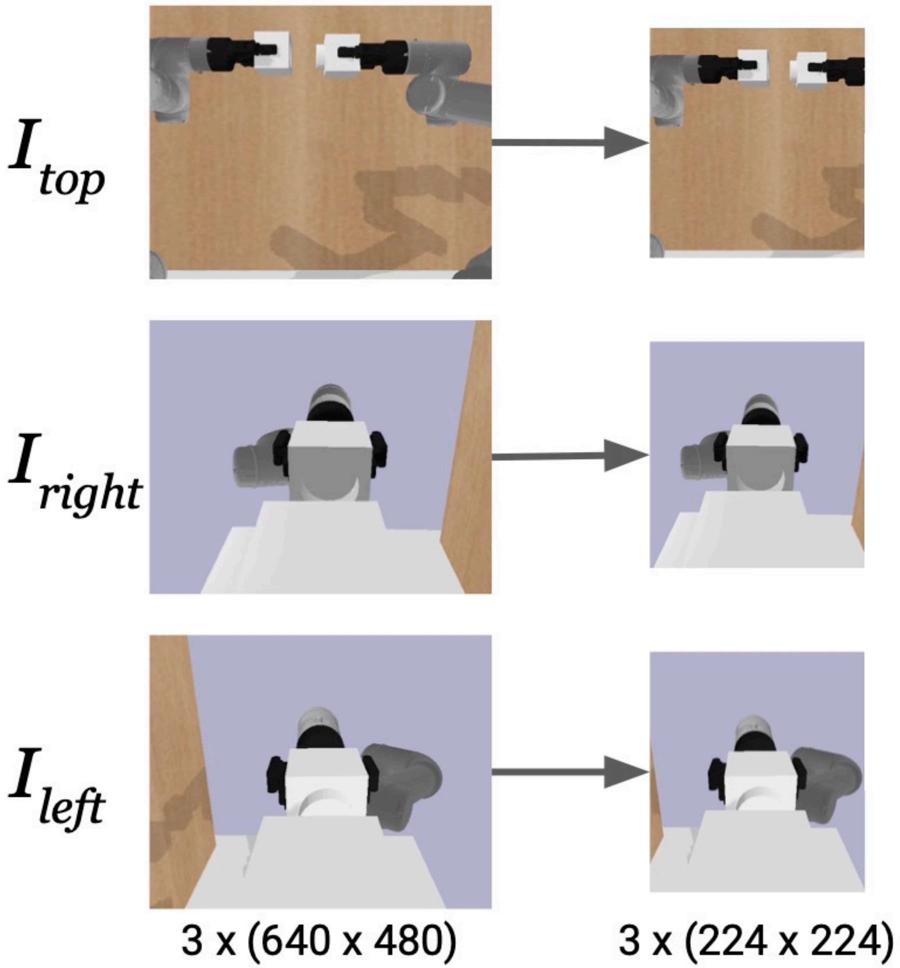
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Dual-arm Manipulation Policy Learning Framework

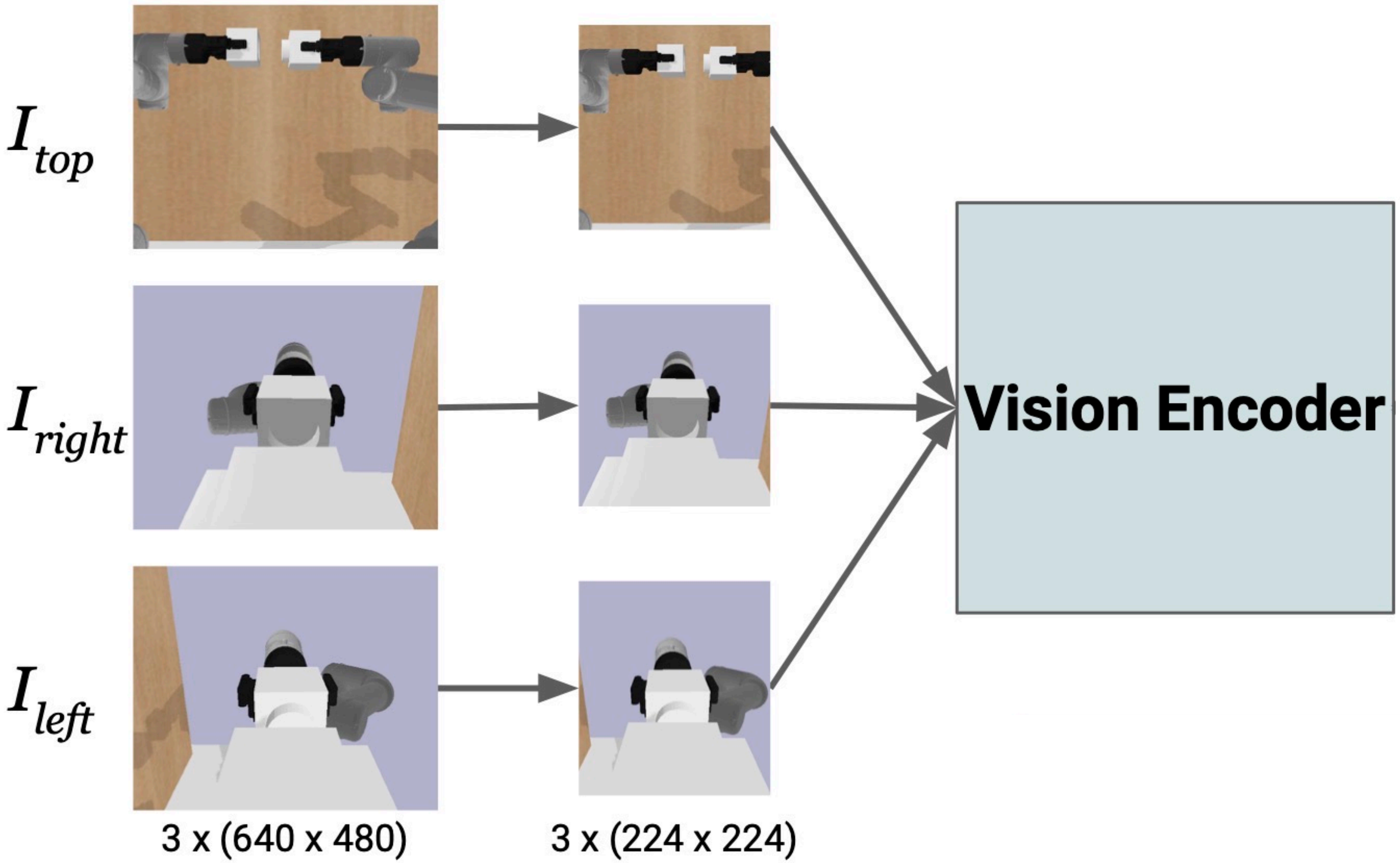


C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

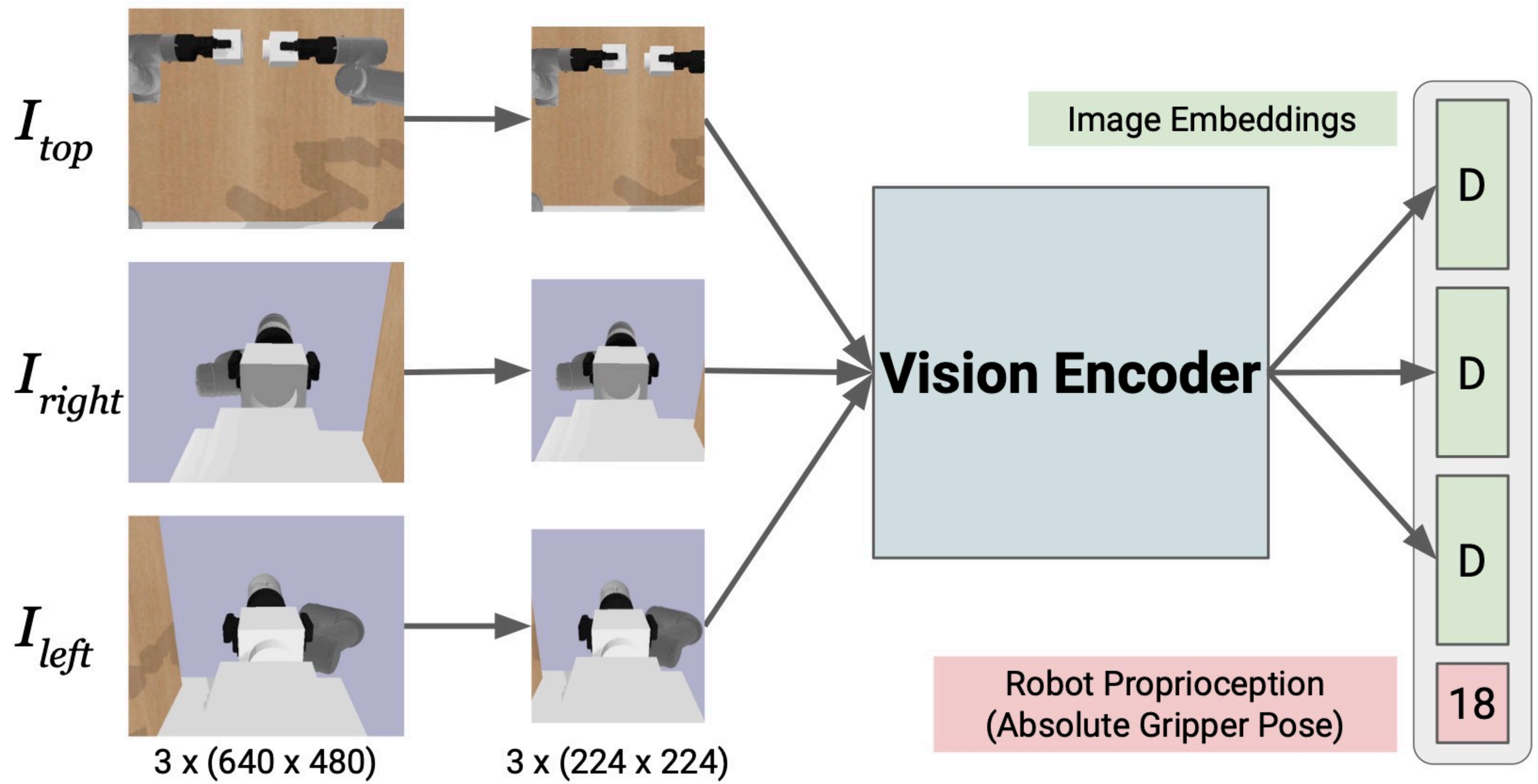
Dual-arm Manipulation Policy Learning Framework



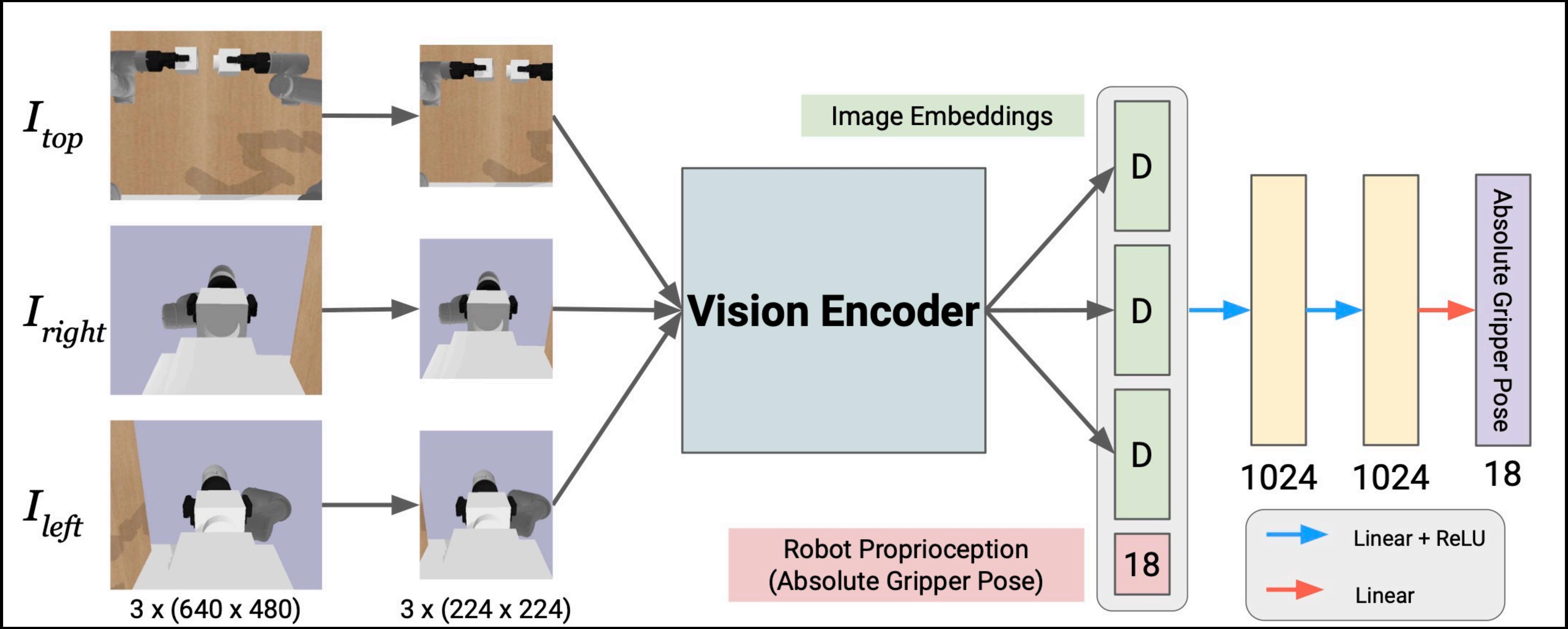
Dual-arm Manipulation Policy Learning Framework



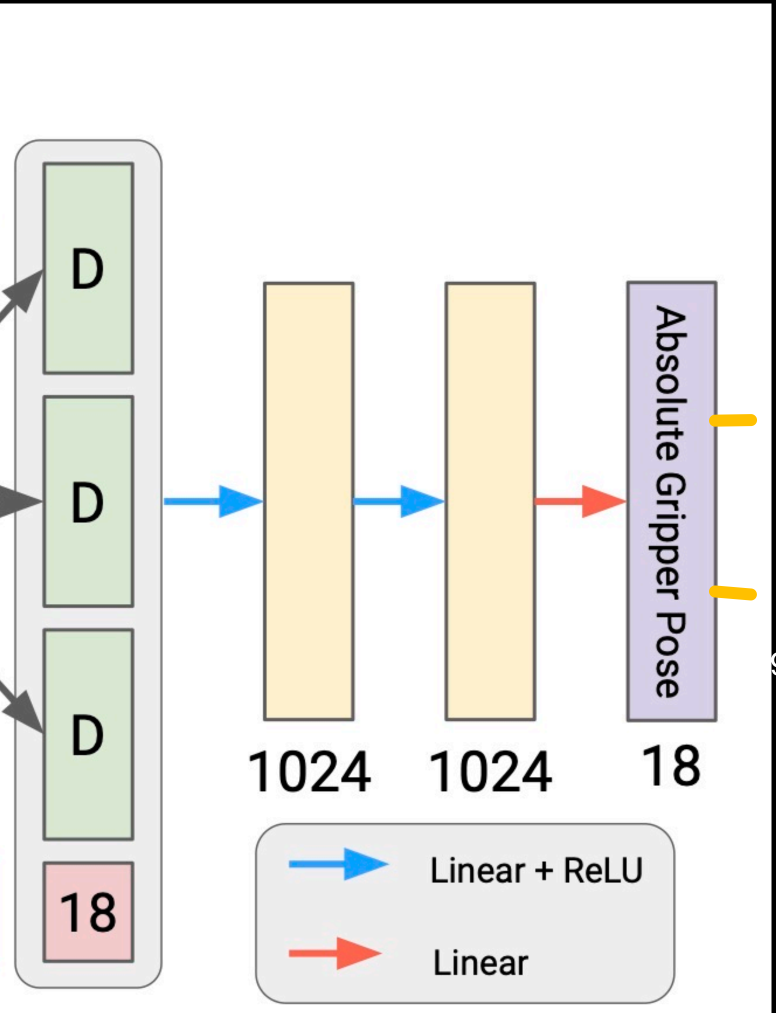
Dual-arm Manipulation Policy Learning Framework



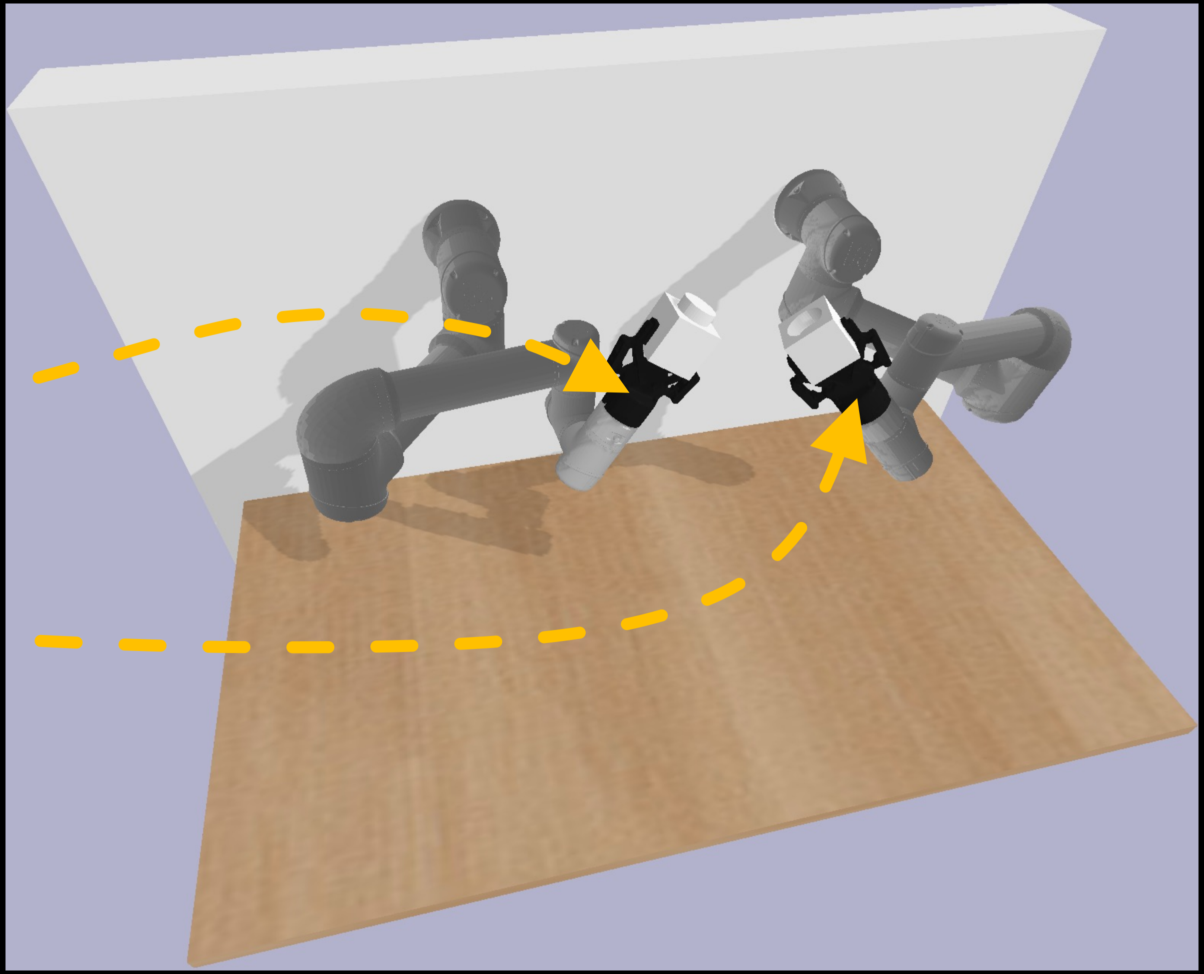
Dual-arm Manipulation Policy Learning Framework



Dual-arm Manipulation Policy Learning Framework



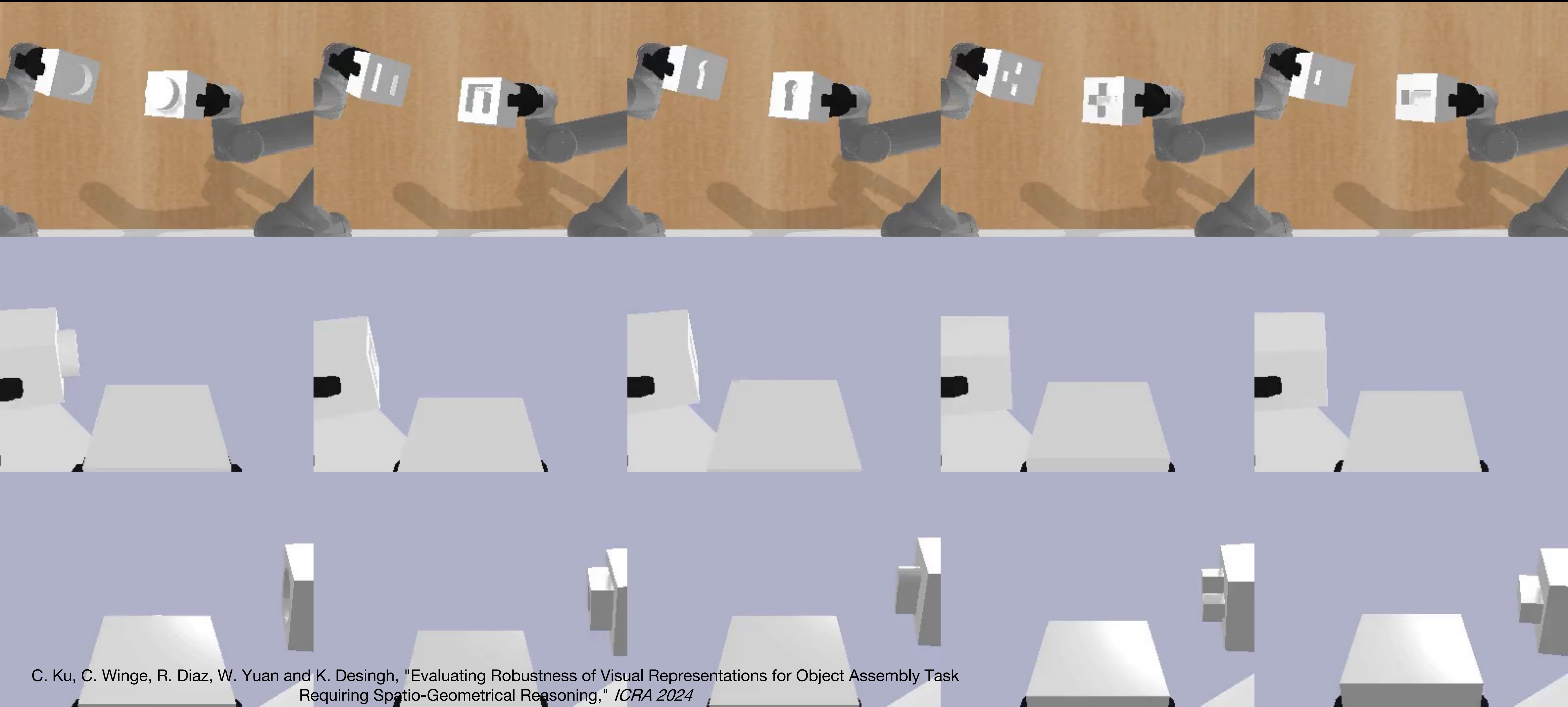
9 values representing the gripper pose with 3 values for x, y, z and 6 values from first two columns of rotation matrix [Zhou et al. 2019]



Y. Zhou, C. Barnes, J. Lu, J. Yang, and H. Li. On the continuity of rotation representations in neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5745–5753, 2019.

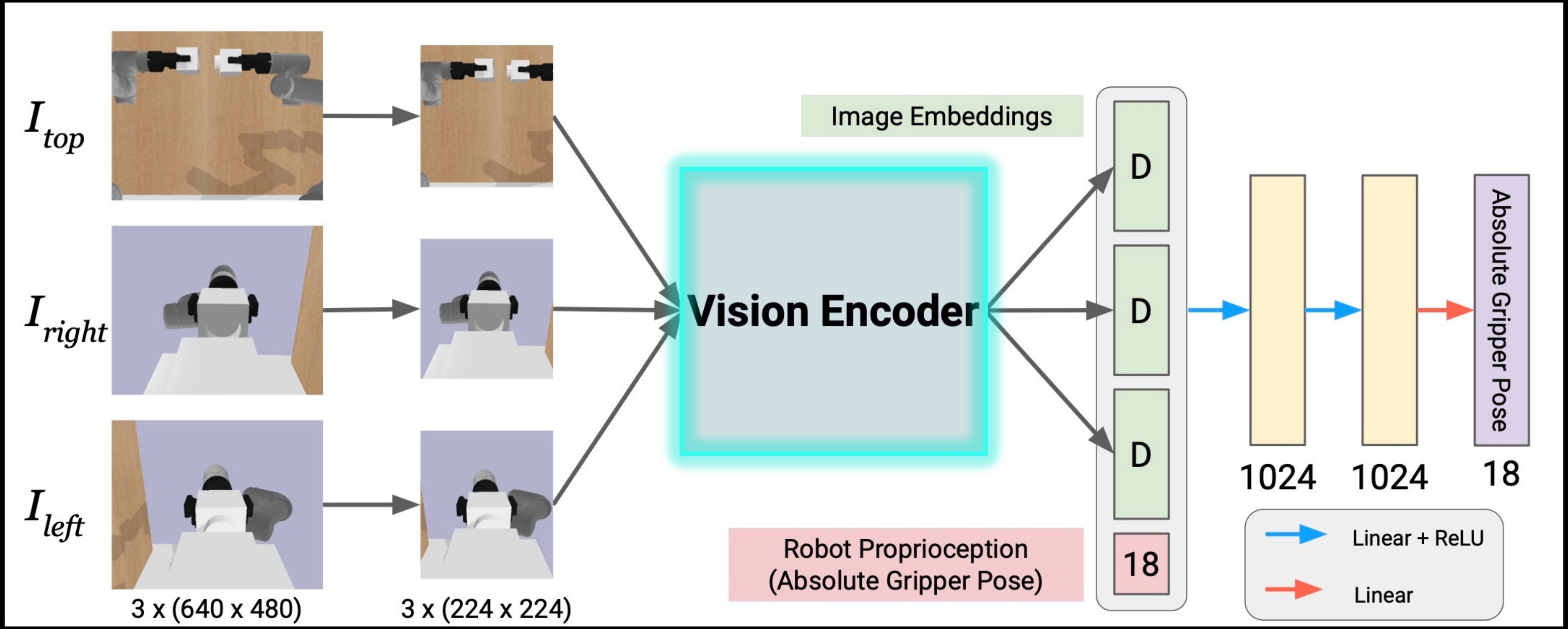
Demonstration Data in Simulation Experiments

Sampled videos. Note: 3 views are collected in simulation experiments



C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Evaluations with Existing Visual Encoders



To our surprise, the pre-trained representations did not do well than a ResNet trained from scratch.

MAE ViT-B/16

R3M ResNet-50

ViT-B/16

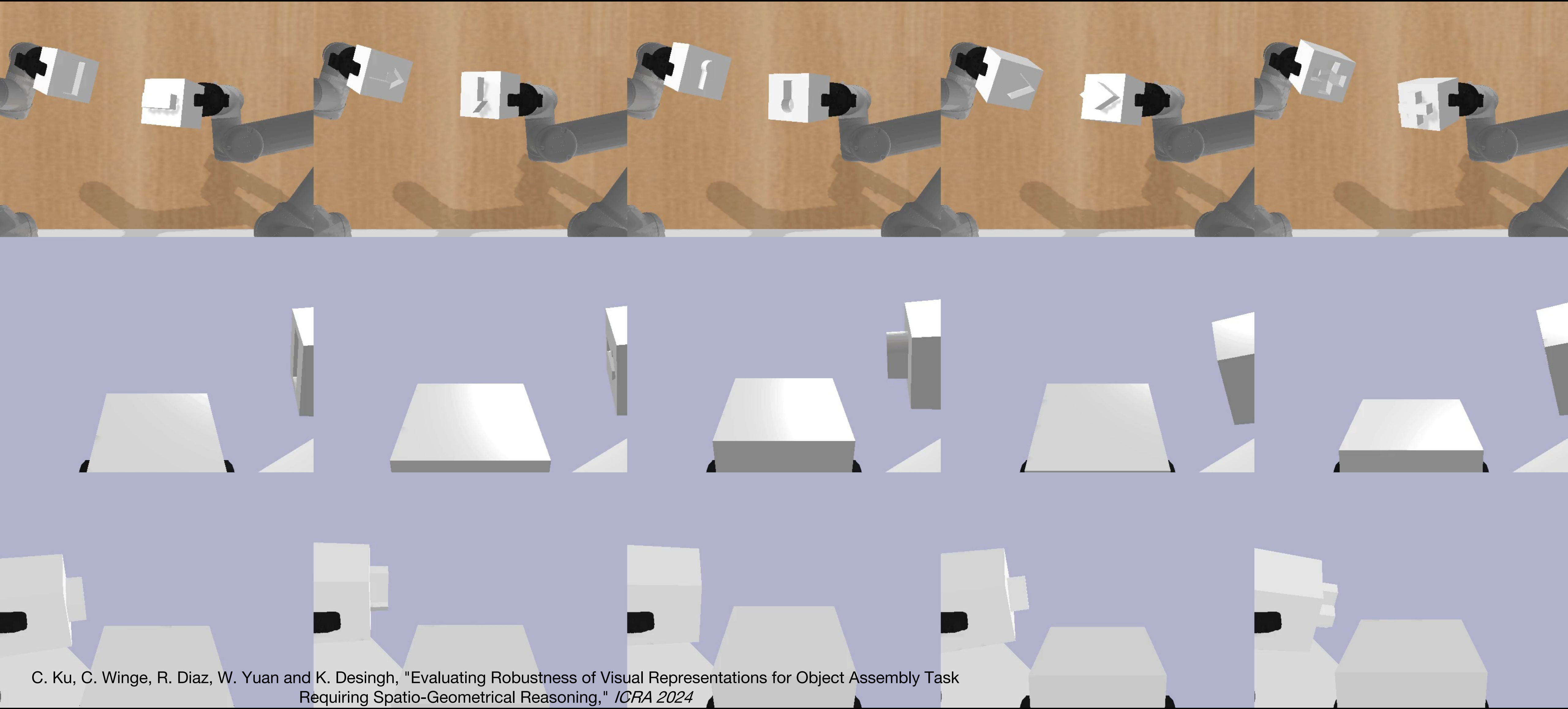
ResNet-50

Non-pretrained ResNet-50

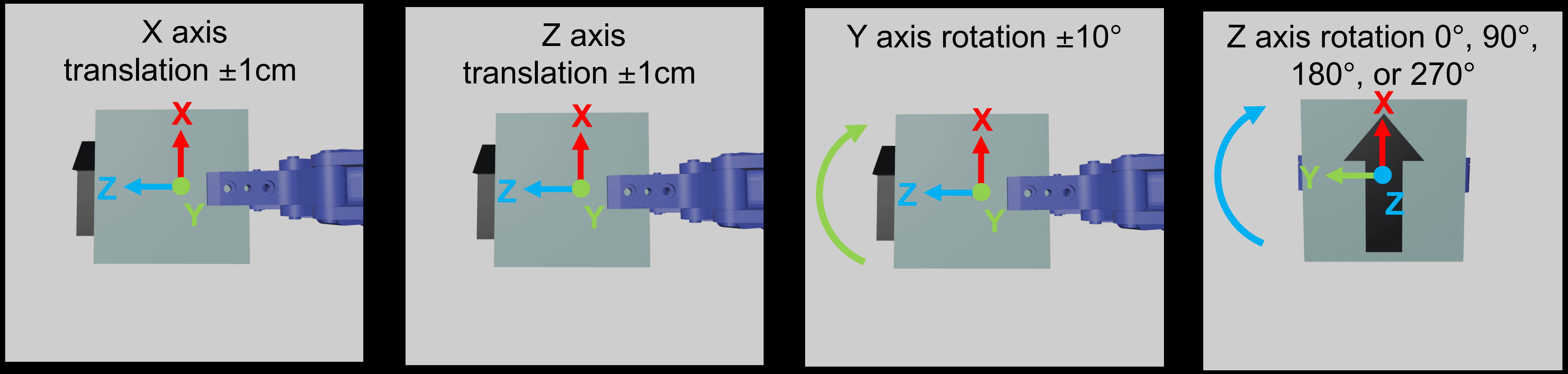
Non-pretrained ResNet-18

Evaluation in Simulation

Sampled results from successful results with Non-pretrained ResNet-18 model.



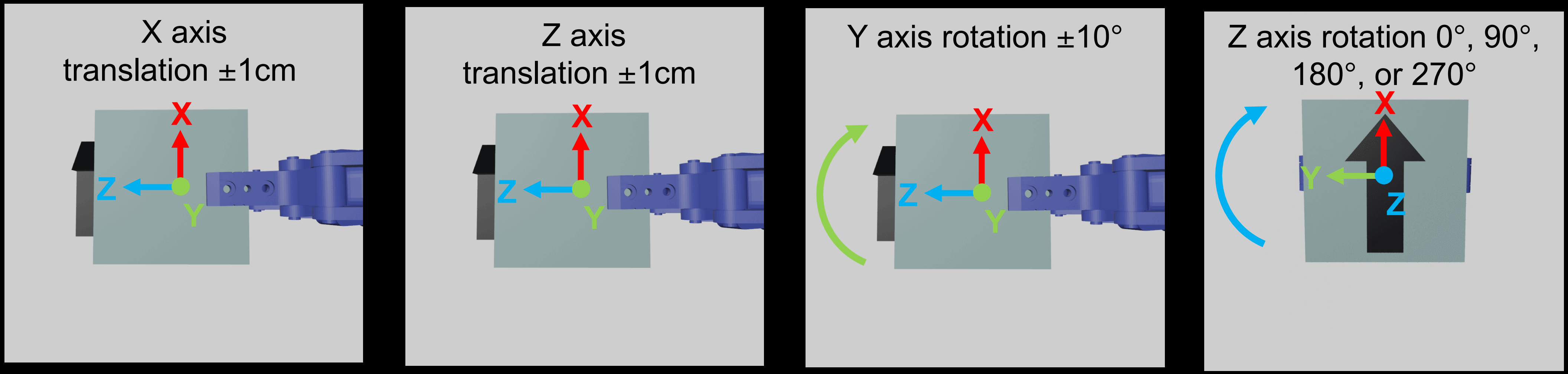
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*



	XT	ZT	YR	ZR
Non-pretrained ResNet-18	**1.000**	**1.000**	**1.000**	**0.775**
Non-pretrained ResNet-50	**1.000**	**1.000**	**1.000**	**0.825**
ImageNet ResNet-50	1.000	1.000	0.925	0.425
R3M ResNet-50	0.950	1.000	1.000	0.275
CLIP ResNet-50	1.000	1.000	0.975	0.625
ImageNet ViT-base	0.950	1.000	0.975	0.450
CLIP ViT-base	1.000	1.000	0.900	0.575
MAE ViT-base	1.000	1.000	0.925	0.350

TABLE I: Success rates of all visual representations trained with 100 demonstrations of indicated task variation. Non-pretrained ResNets clearly outperform pretrained models on ZR.

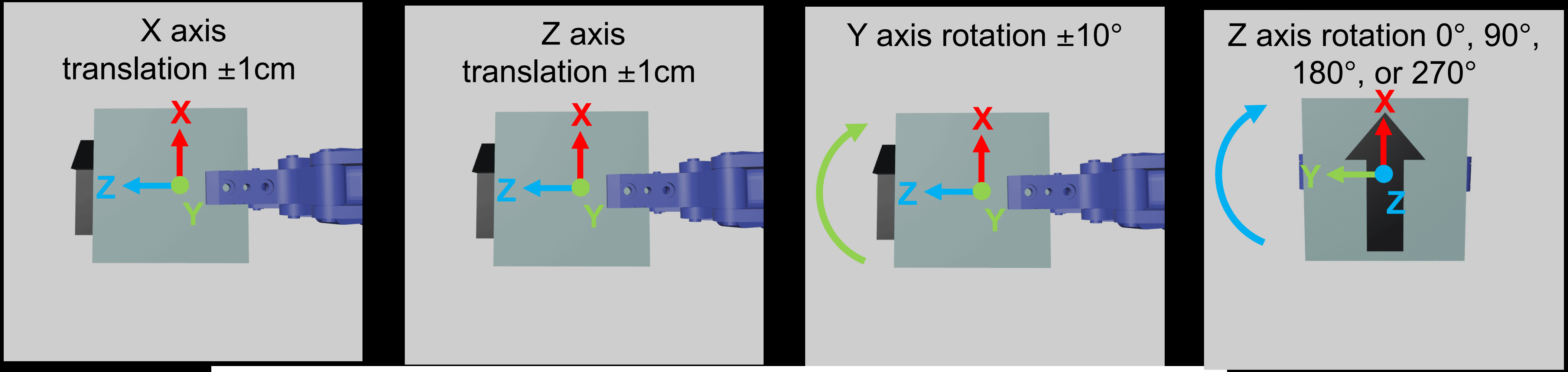
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*



	XTZR	ZTZR	YRZR	XZTYZR
Non-pretrained ResNet-18	**0.825**	**0.825**	**0.675**	**0.275**
Non-pretrained ResNet-50	0.425	0.775	0.300	0.075
ImageNet ResNet-50	0.225	0.225	0.175	0.050
R3M ResNet-50	0.150	0.275	0.05	0.050
CLIP ResNet-50	0.500	0.575	0.250	0.150
ImageNet ViT-base	0.150	0.300	0.225	0.025
CLIP ViT-base	0.300	0.250	0.200	0.050
MAE ViT-base	0.375	0.25	0.175	0.050

TABLE II: Success rates of all visual representations trained with 1000 demonstrations of indicated task variation using all objects.

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*



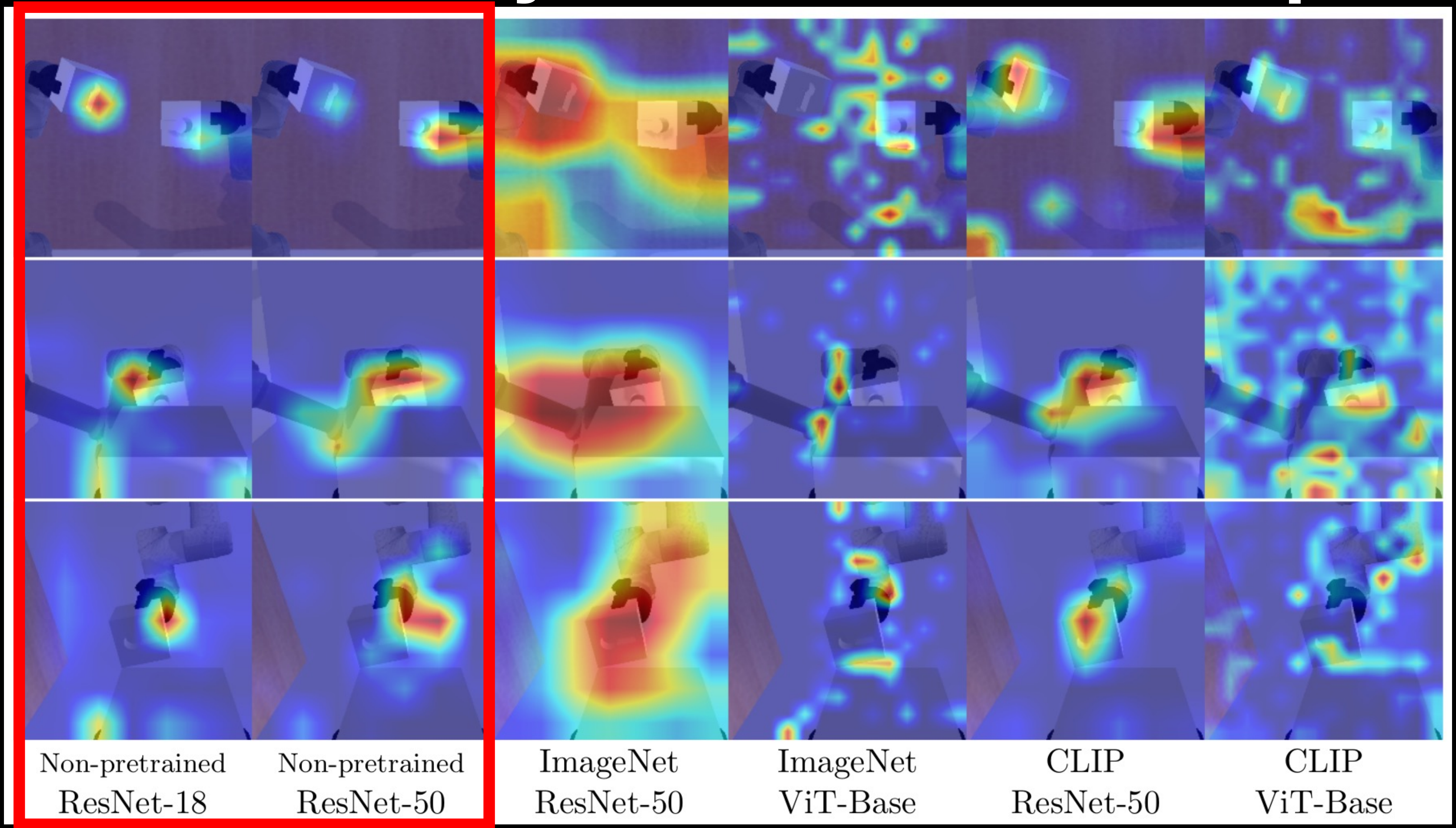
Decreasing
Order of Symmetry

Objects	XTZR	ZTZR	YZR	XZTYZR
circle	0.85±0.07	1.00±0.00	0.83±0.05	0.43±0.07
plus	0.93±0.04	1.00±0.00	0.77±0.01	0.38±0.04
minus	0.80±0.03	0.98±0.00	0.44±0.10	0.33±0.10
diamond	0.77±0.06	1.00±0.00	0.33±0.08	0.34±0.08
hexagon	0.71±0.08	1.00±0.00	0.38±0.08	0.30±0.08
u	0.37±0.08	0.54±0.06	0.12±0.06	0.17±0.03
pentagon	0.34±0.10	0.56±0.04	0.10±0.07	0.18±0.07
arrow	0.38±0.07	0.66±0.08	0.18±0.03	0.17±0.05
key	0.38±0.07	0.66±0.08	0.17±0.04	0.19±0.05
all	0.61±0.02	0.82±0.03	0.37±0.05	0.28±0.03

TABLE III: Success rates of Non-pretrained ResNet-18 trained on 1000 demonstrations including all objects. Mean and standard deviations over 3 different evaluations of 40 randomized rollouts.

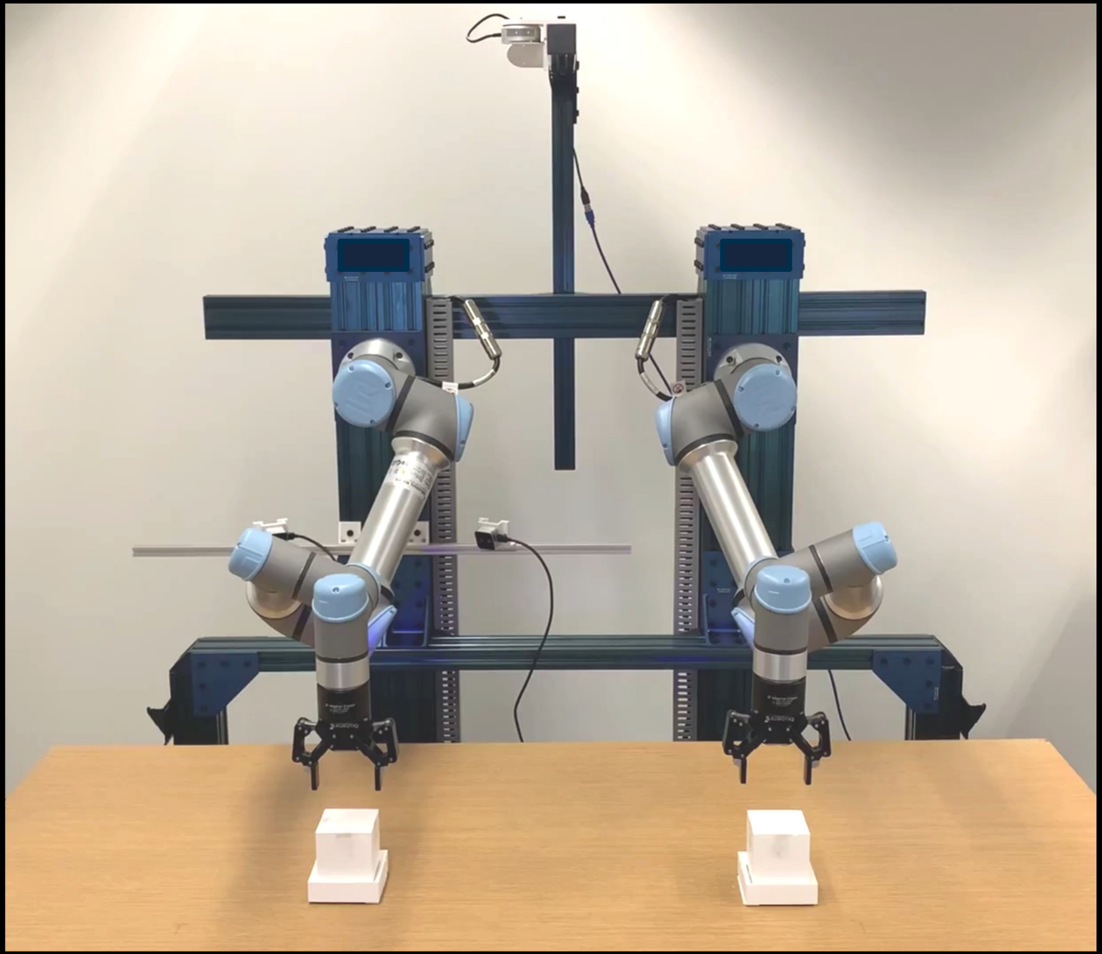
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Qualitative Analysis of Activation Maps



C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Real world setup for data collection & evaluation

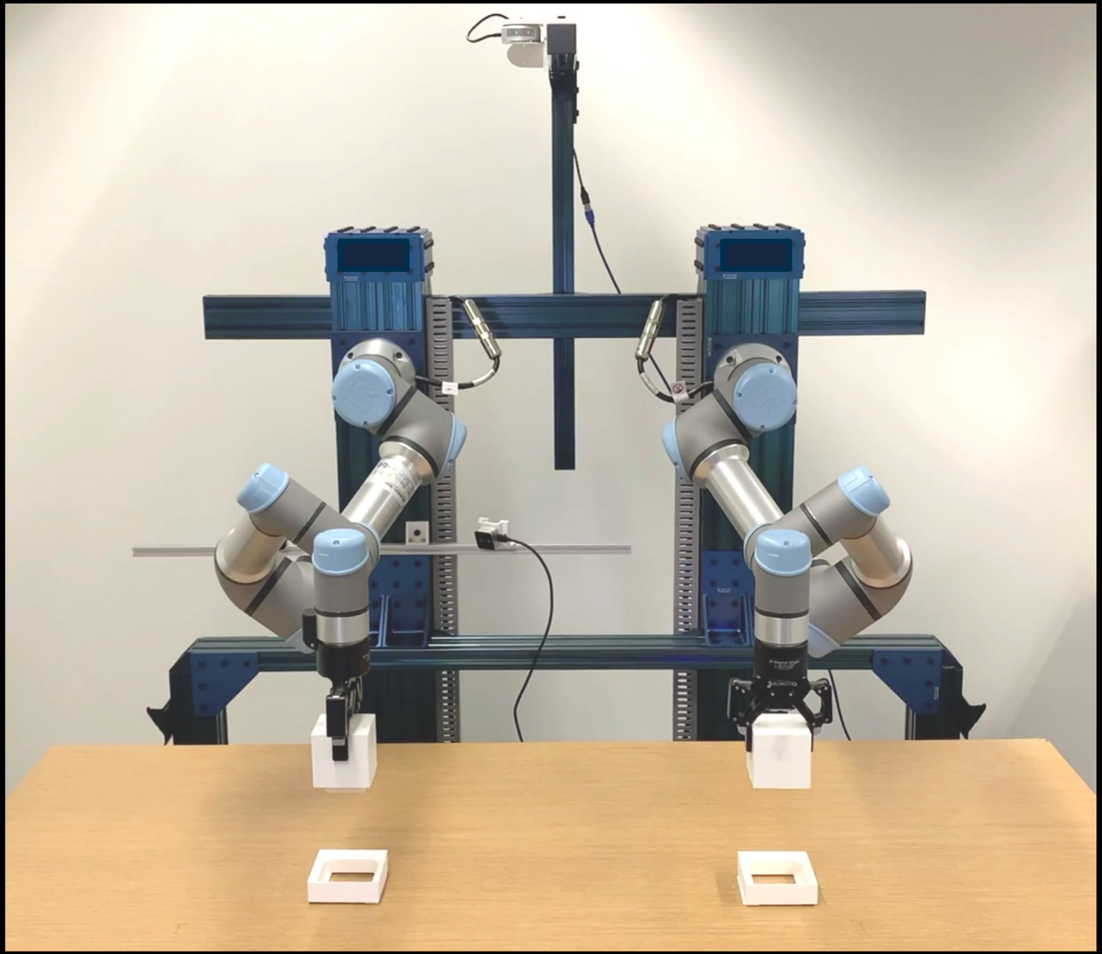


For each episode, objects are picked up with a random grasp variation

Note: the geometrical information is not available to the robot during grasping until seen by the top-view camera

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Real world setup for data collection & evaluation



For each episode, objects are picked up with a random grasp variation

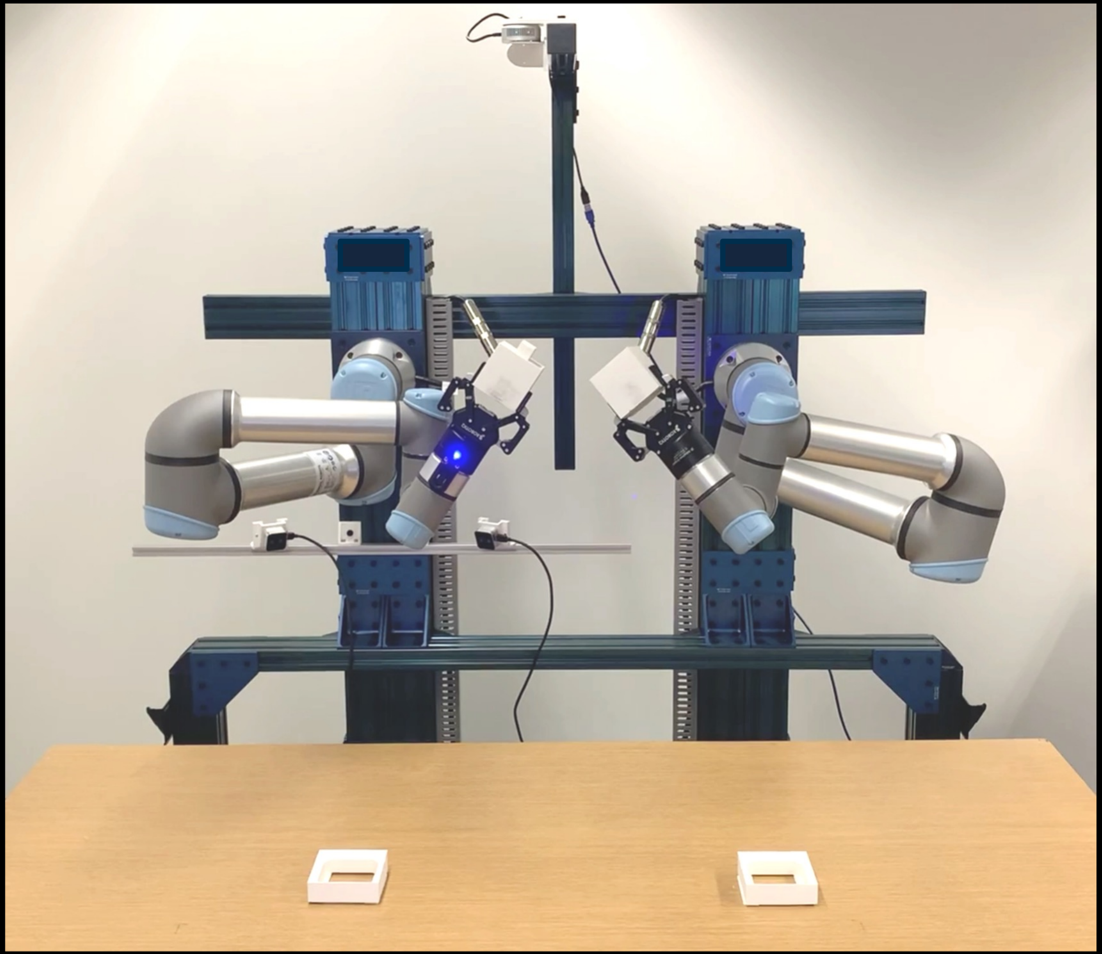
Note: the geometrical information is not available to the robot during grasping until seen by the top-view camera

The object parts are shown to the top-view camera.

Note: the extrusions and intrusions are randomized between left and right grippers

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Real world setup for data collection & evaluation



For each episode, objects are picked up with a random grasp variation

Note: the geometrical information is not available to the robot during grasping until seen by the top-view camera

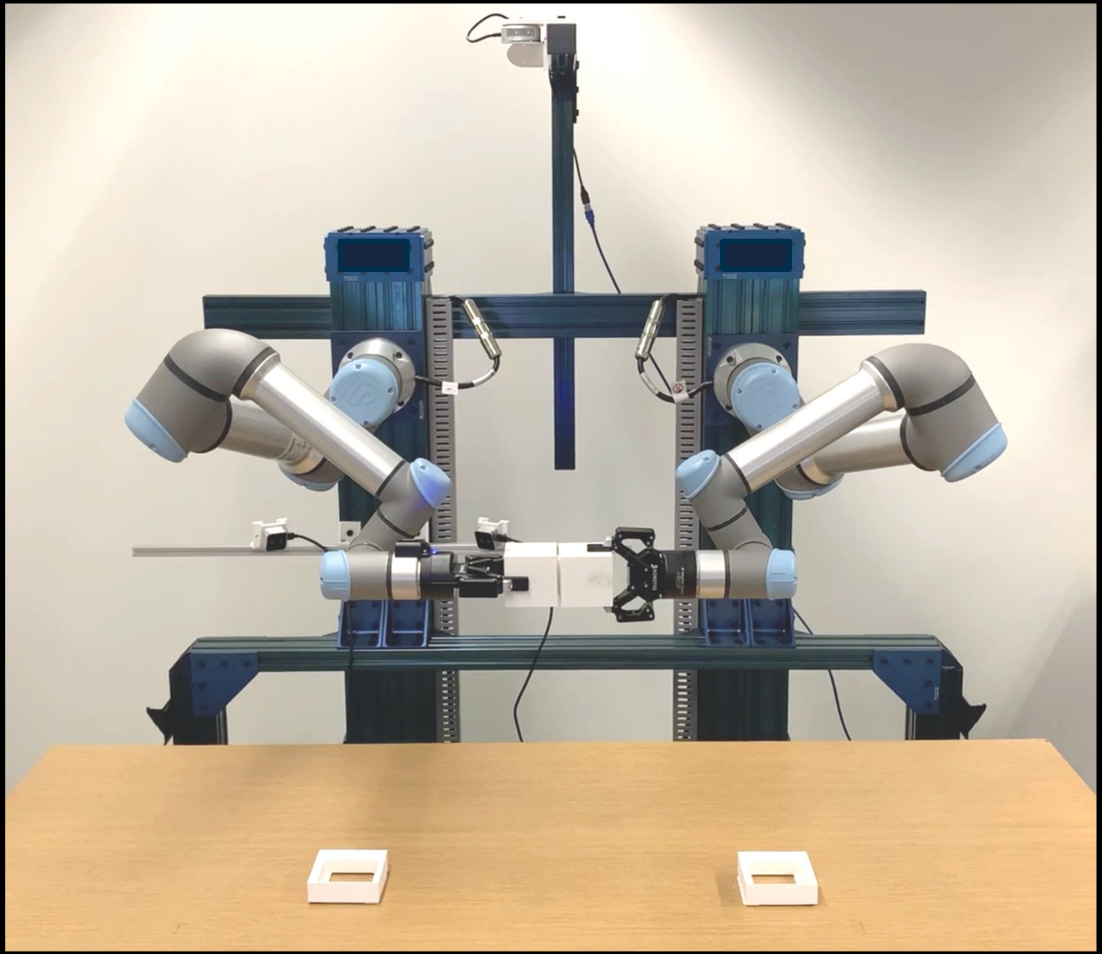
The object parts are shown to the top-view camera.

Note: the extrusions and intrusions are randomized between left and right grippers

Scripted expert trajectories are used to perform the assembly task to collect the demonstrations

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Real world setup for data collection & evaluation



For each episode, objects are picked up with a random grasp variation

Note: the geometrical information is not available to the robot during grasping until seen by the top-view camera

The object parts are shown to the top-view camera.

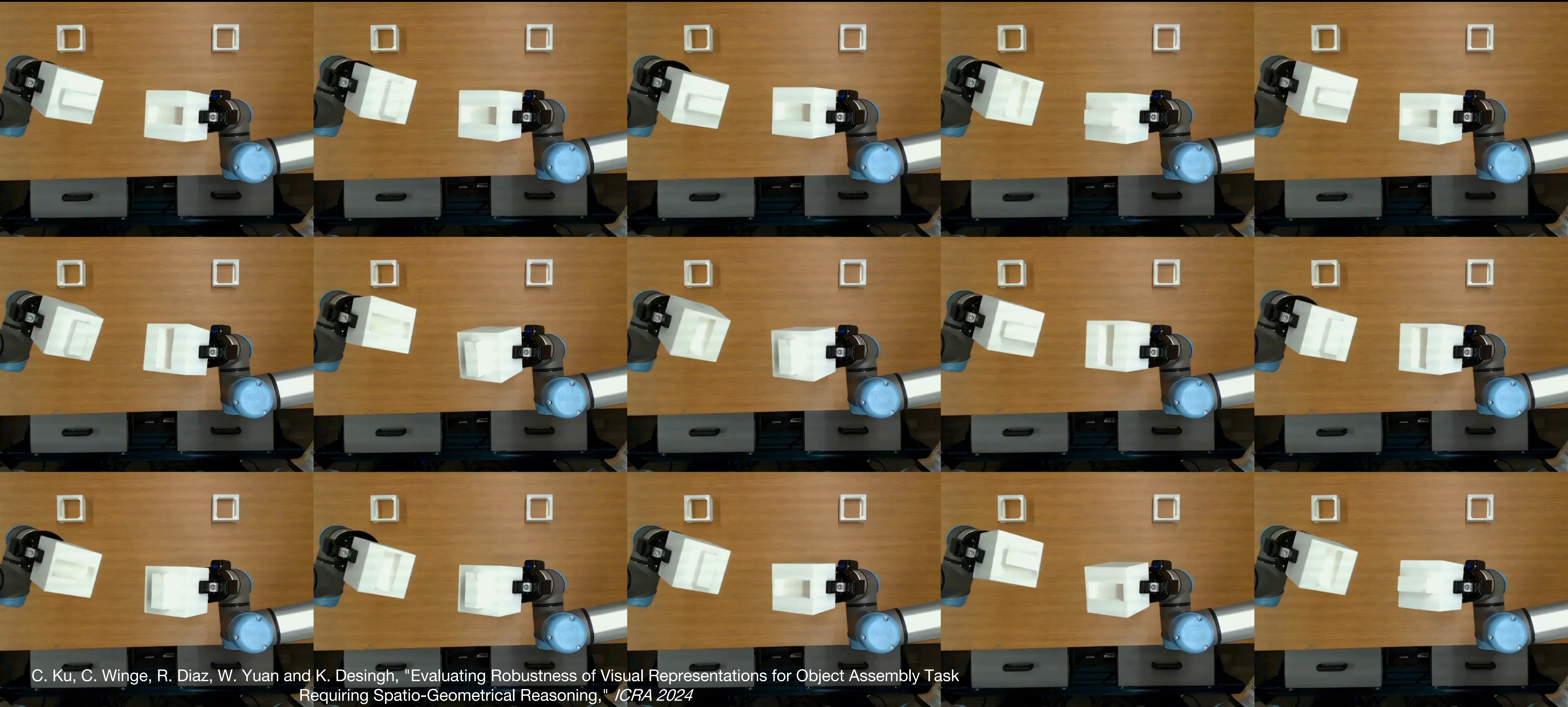
Note: the extrusions and intrusions are randomized between left and right grippers

Scripted expert trajectories are used to perform the assembly task to collect the demonstrations

Object parts are placed back on the supporting wedges before grasping for the next episode in data

C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

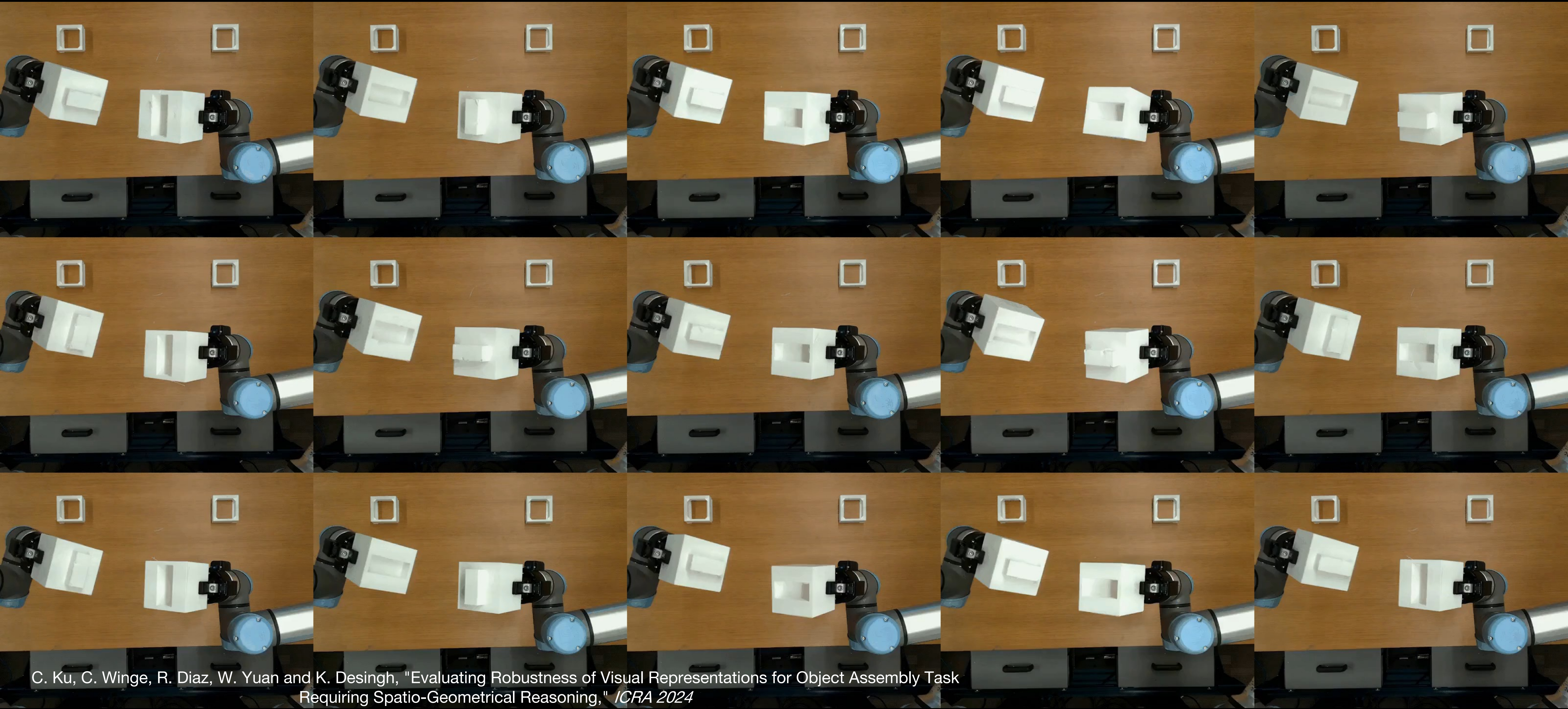
Real World Demonstration Data Examples



C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

Real World Evaluation

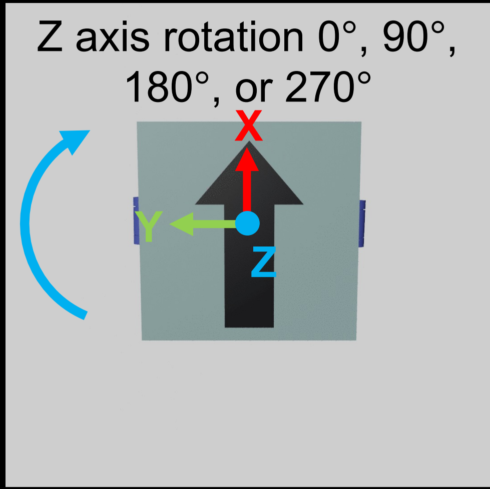
Sampled results from successful experiments. Note: real world experiments do not use wrist camera views

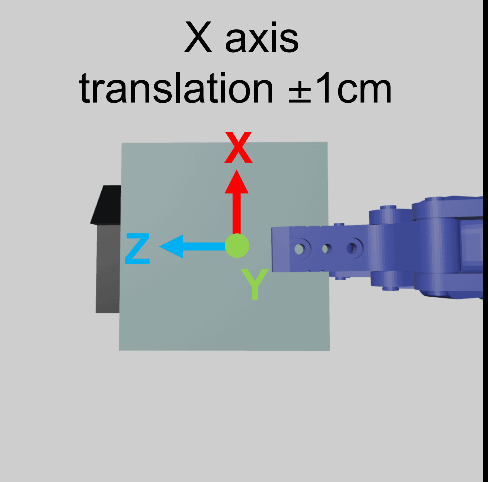


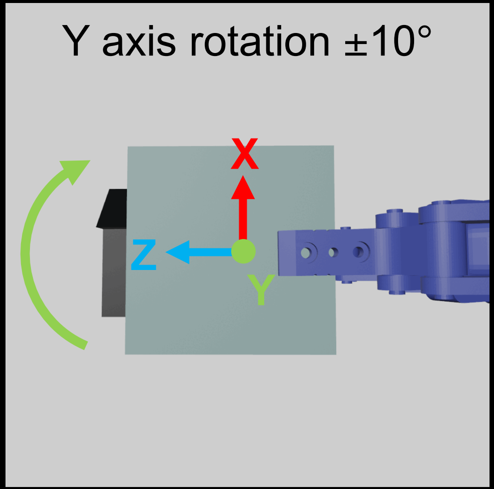
C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

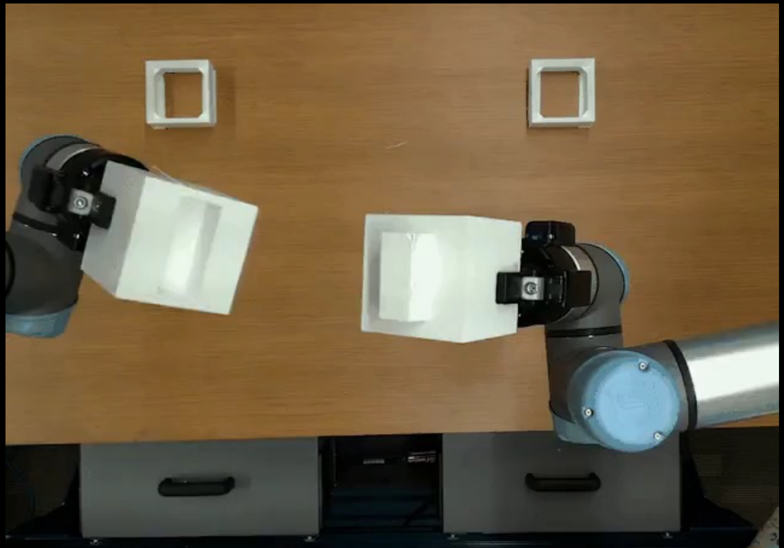
Real World Failure Examples

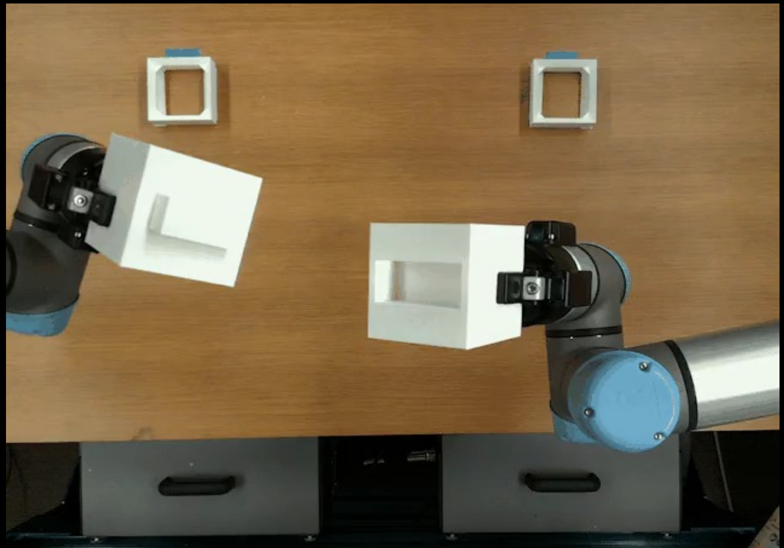
Sampled results from failed experiments. One from each grasp variation is shown.

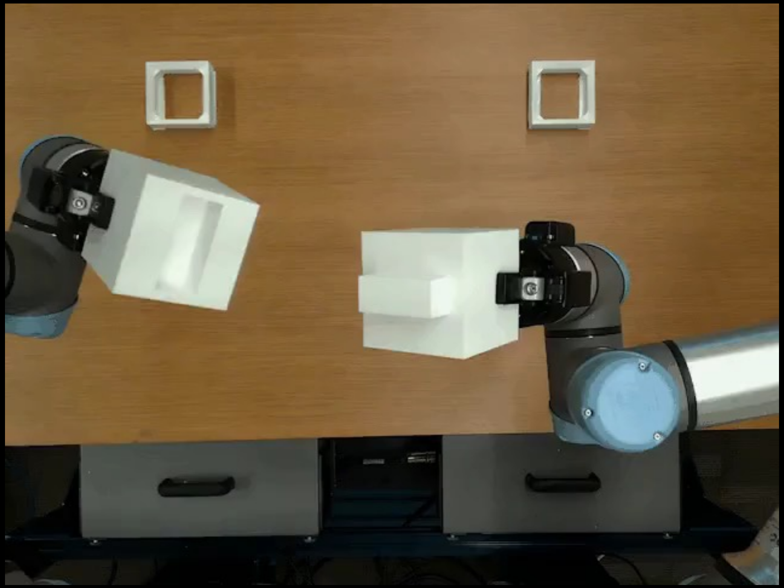












C. Ku, C. Winge, R. Diaz, W. Yuan and K. Desingh, "Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning," *ICRA 2024*

We experimented on a number of things!

- What if we can finetune the pre-trained models?
- What if we give more data?
- How much impact does proprioception have?
- Does object texture help in performance?
- How does the model perform when the geometries are perturbed from the training distribution?

2024 IEEE International Conference on Robotics and Automation (ICRA)
May 13-17, 2024. Yokohama, Japan

Evaluating Robustness of Visual Representations for Object Assembly Task Requiring Spatio-Geometrical Reasoning

Chahyon Ku[1], Carl Winge[1], Ryan Diaz[1], Wentao Yuan[2] and Karthik Desingh[1]

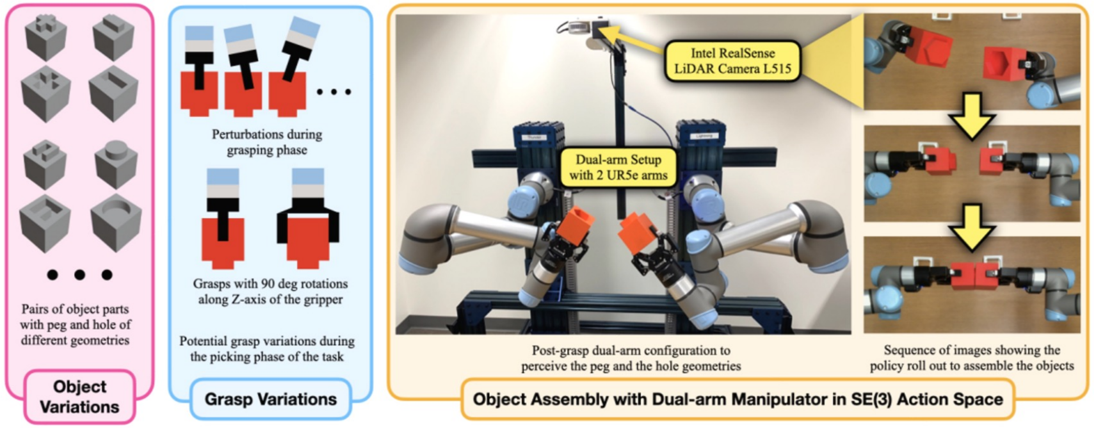


Fig. 1: An overview of our benchmarking setup. Benchmarking robustness under object variations (left) and grasp variations (center) of visual policy learning methods on object assembly task with a dual-arm manipulator in SE(3) action space (right)

We posit that robots need representations that can capture spatio-geometric features to learn novel-object assembly skills from demonstrations.

- No explicit object-geometric knowledge
- Maybe visual information is good enough ← **Requires F/T sensing**
- Maybe pretrained visual representations are good enough to give us these features.

Not necessarily true due to distributional shift

AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks

Ryan Diaz[1], Adam Imdieke[1], Vivek Veeriah[2], Karthik Desingh[1]

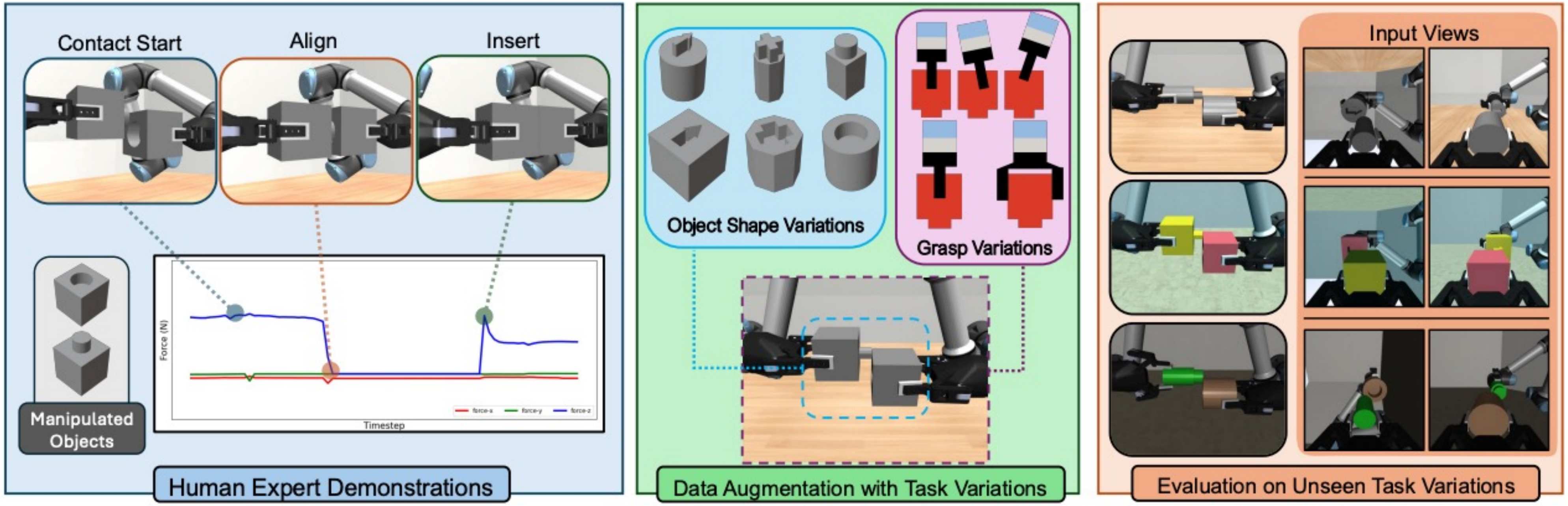
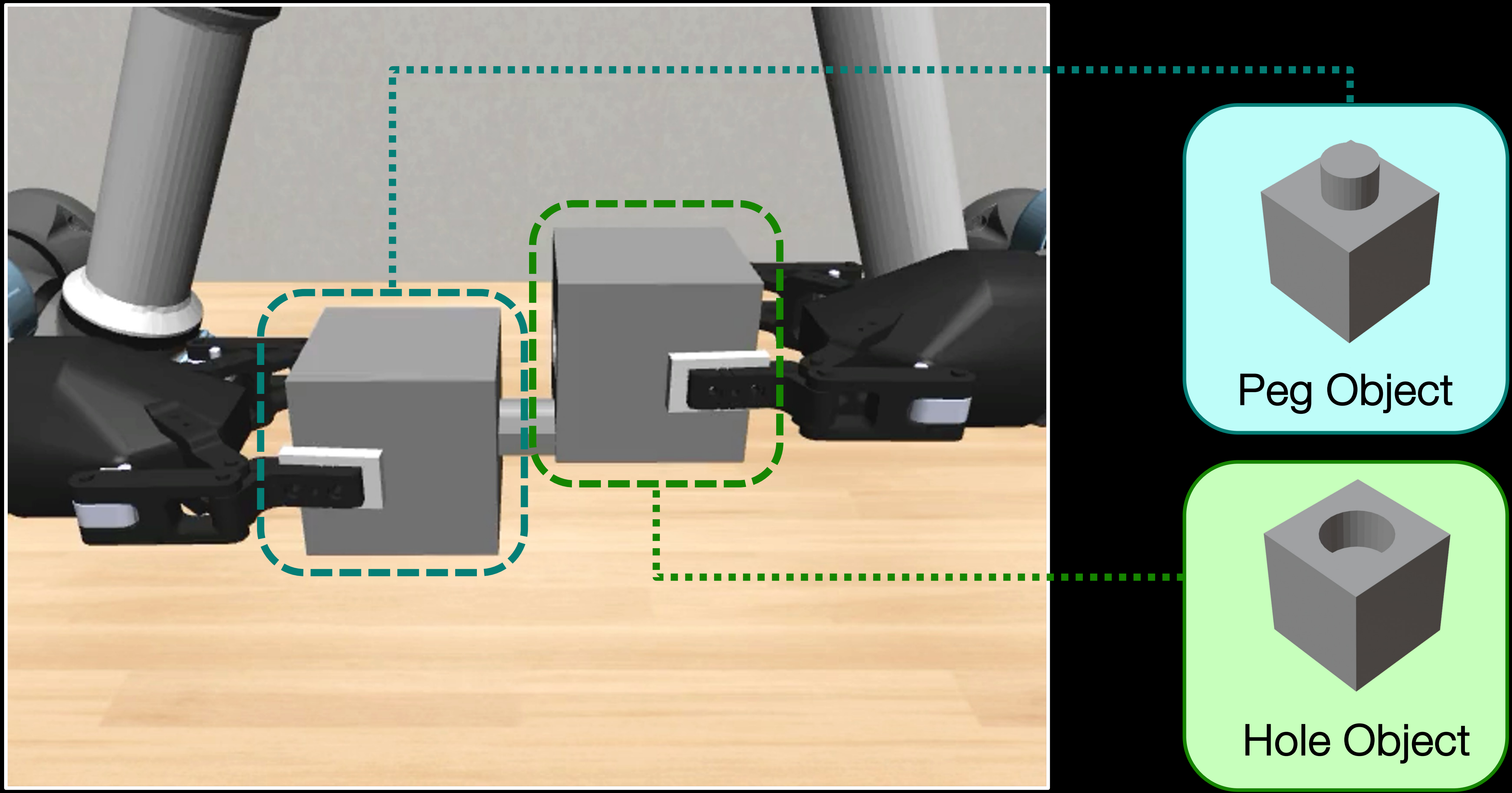
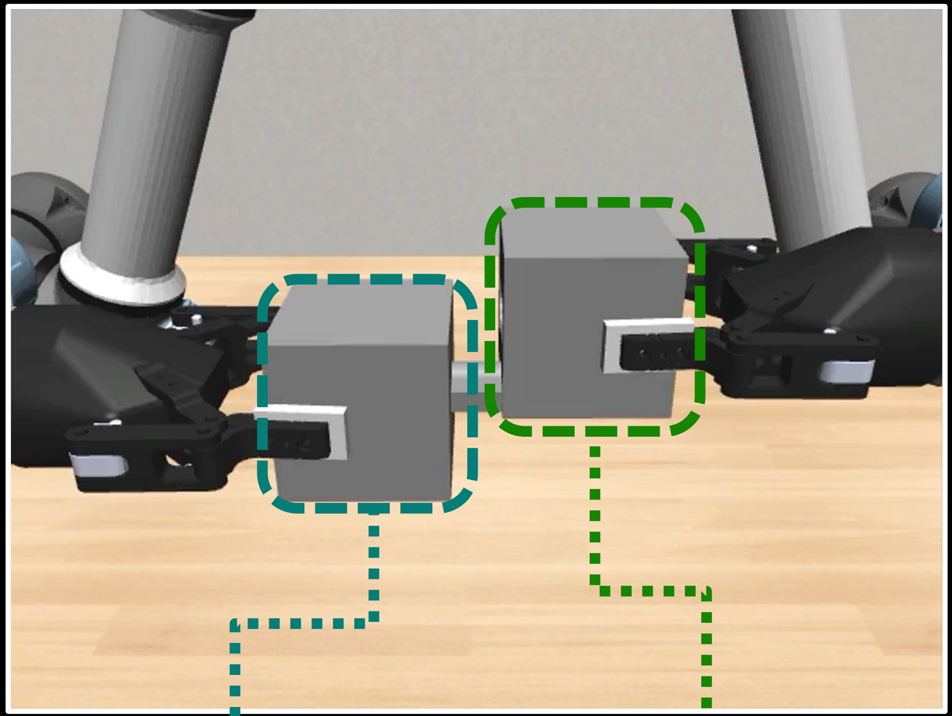
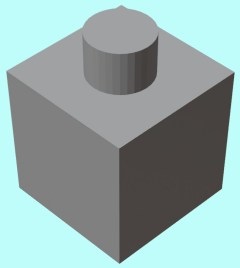


Fig. 1: AugInsert is a data collection and policy evaluation pipeline aimed towards analyzing the robustness of a multisensory (vision, force-torque, and proprioception) model with respect to different observation-level task variations in object shape, grasp pose, and visual environmental appearance. Our framework introduces task variations to a dataset of human-collected demonstrations through a system of online data augmentation.

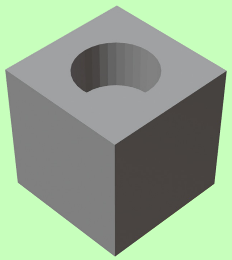


Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).





Peg Object

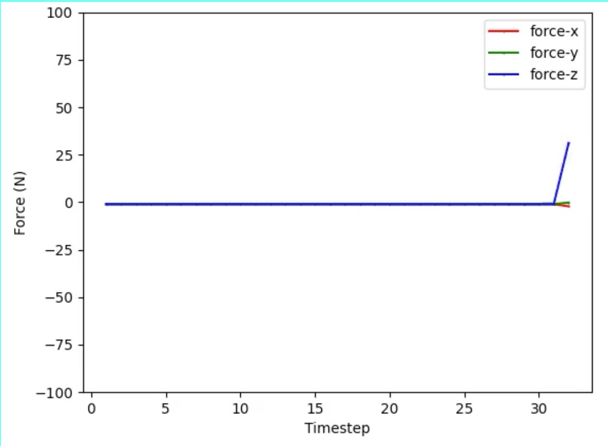


Hole Object

Sensory Inputs

Left Wristview

Right Wristview



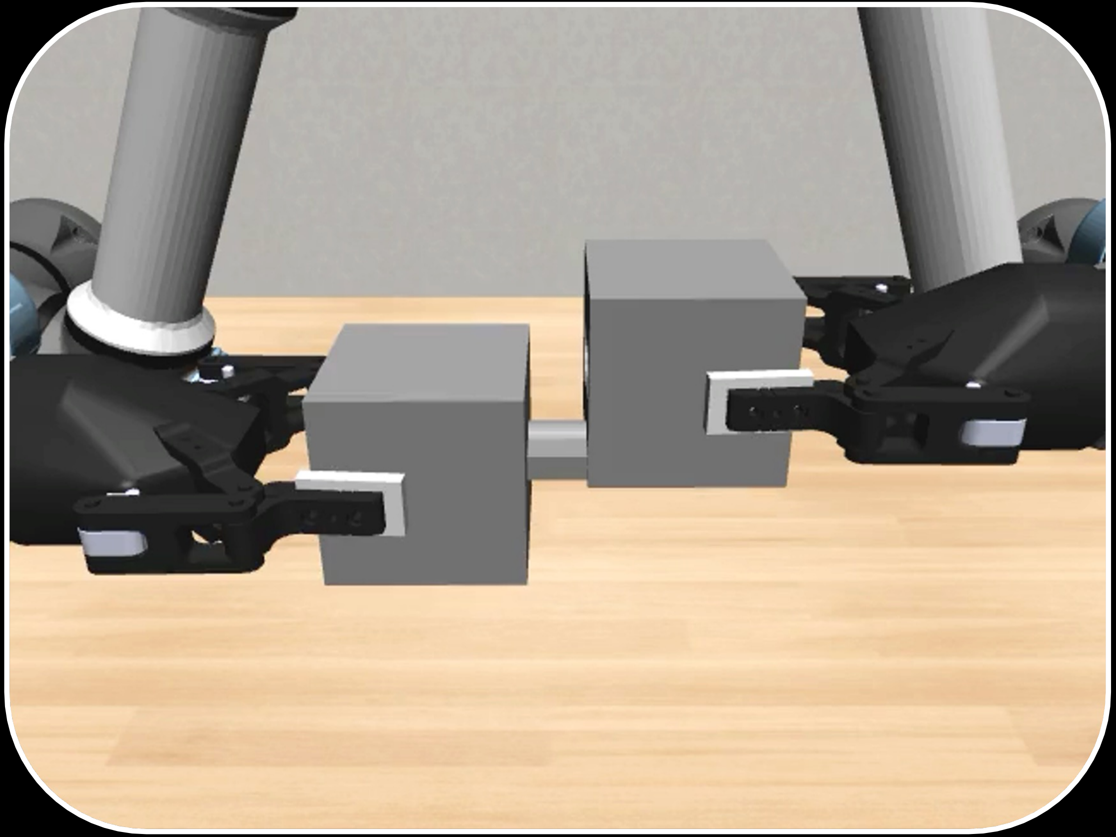
Force-Torque

Proprioception

End-Effector Poses

Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

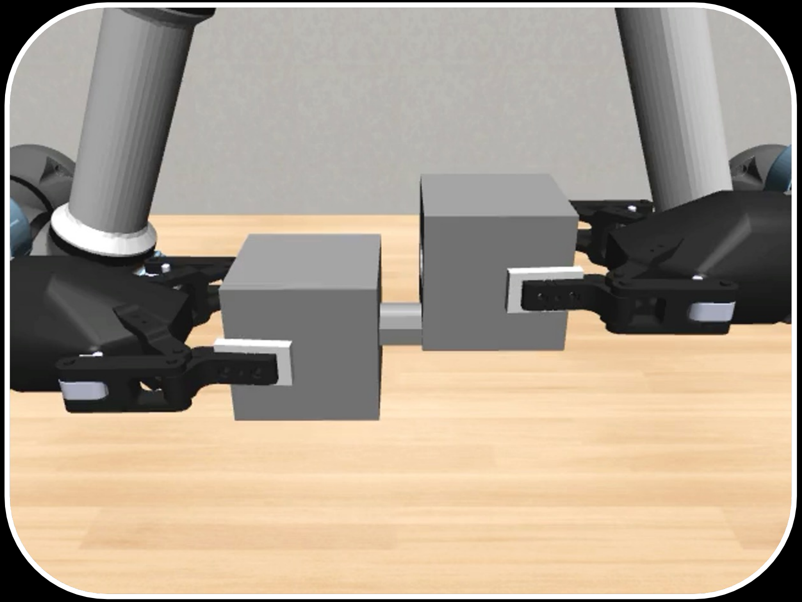
Canonical (no variations)



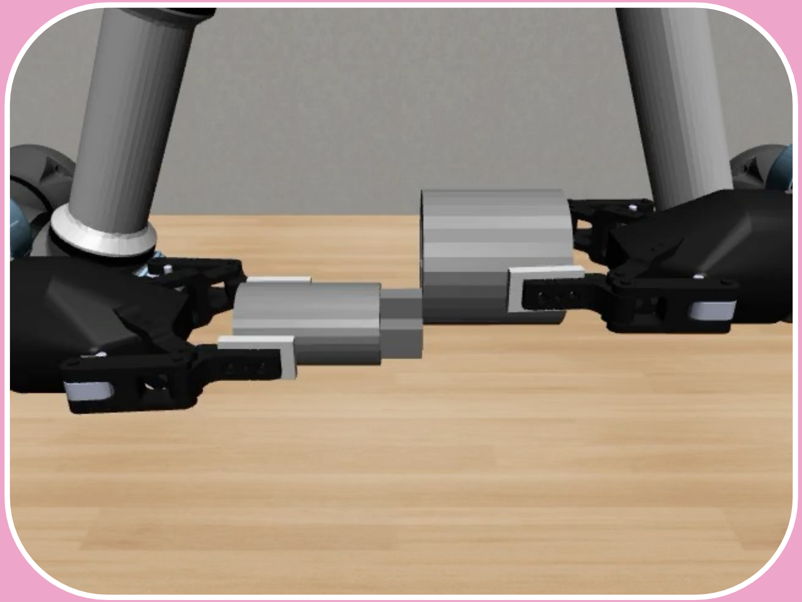
Introduce observation-level task variations

Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

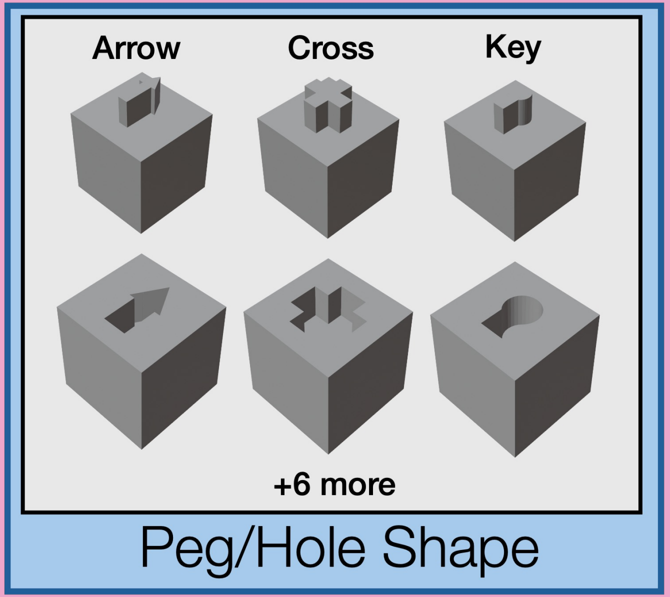
Canonical

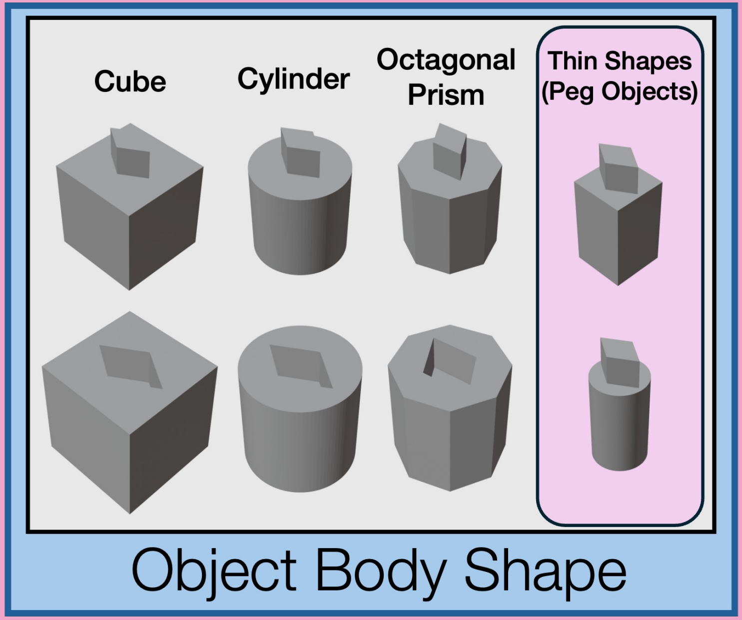


Task Variation



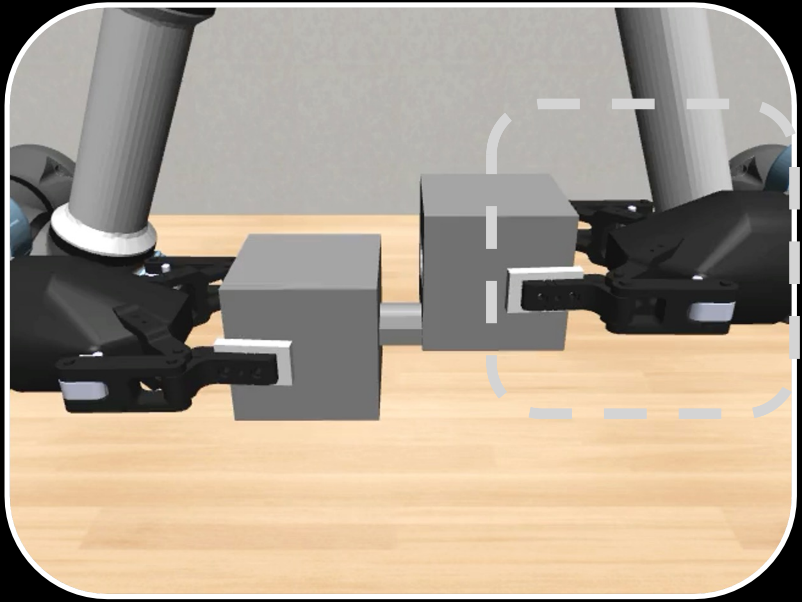
9 peg/hole shapes
6 object body shapes
(3 full size, 3 thin)
54 total object shapes



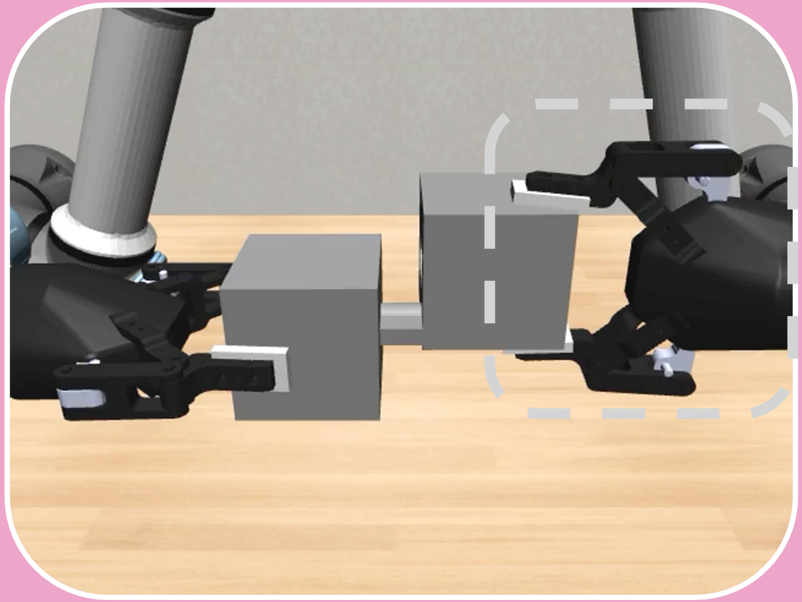


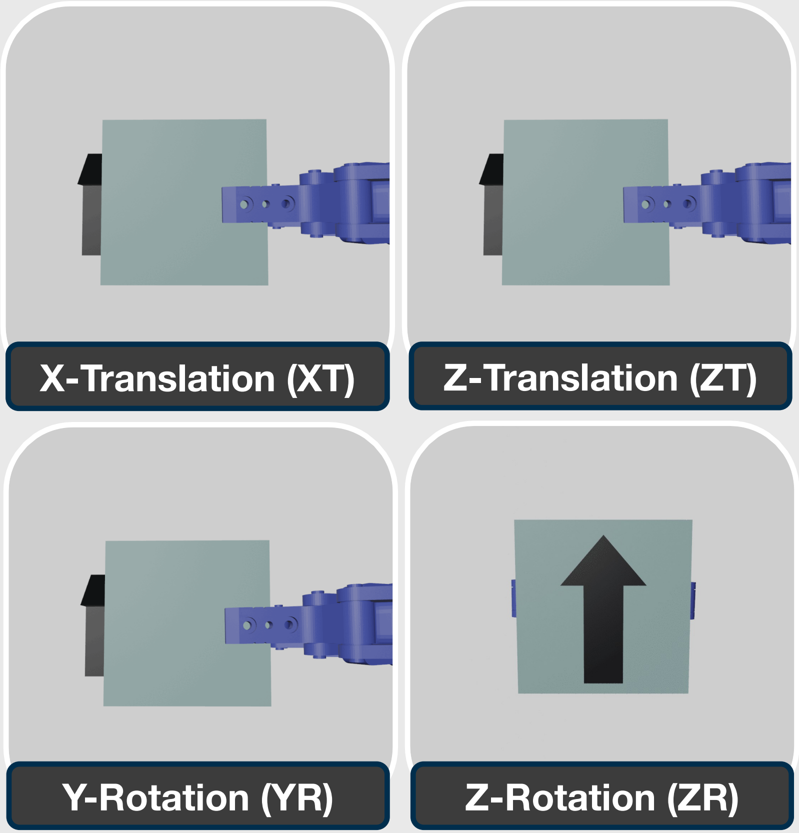
Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

Canonical



Task Variation

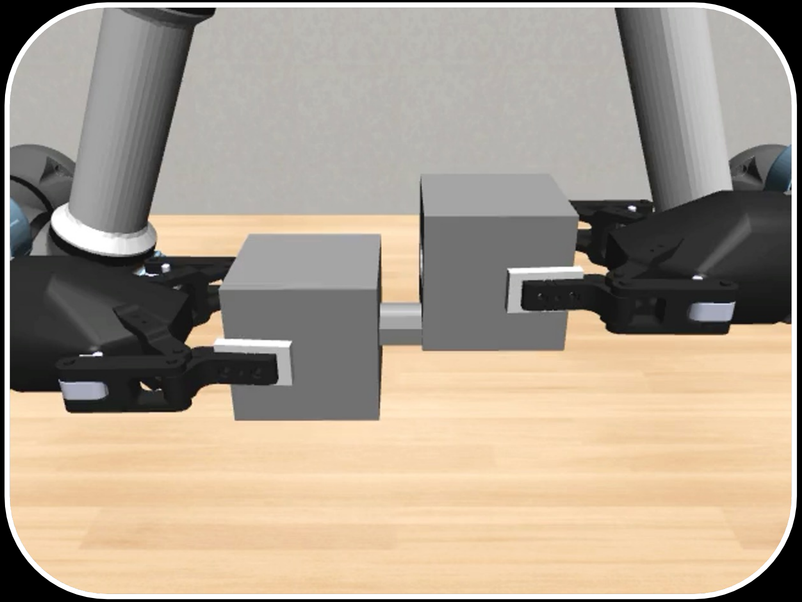




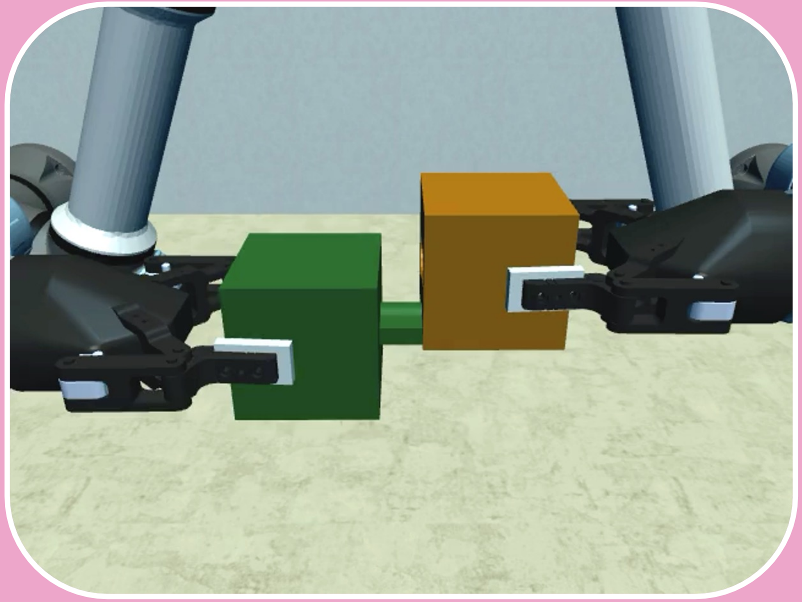
Grasp Pose

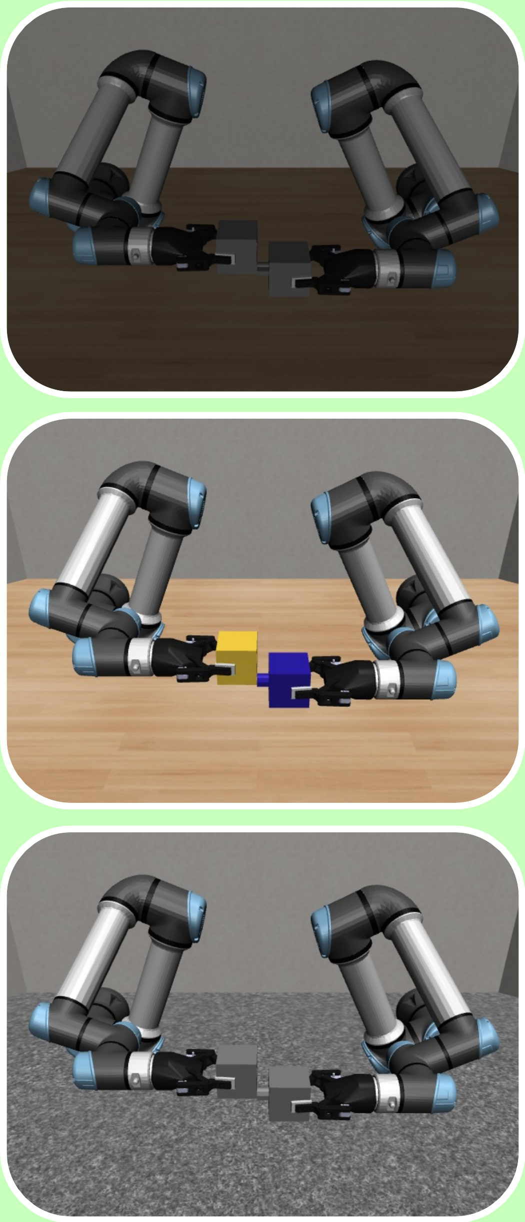
Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

Canonical



Task Variation





Scene Appearance

Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

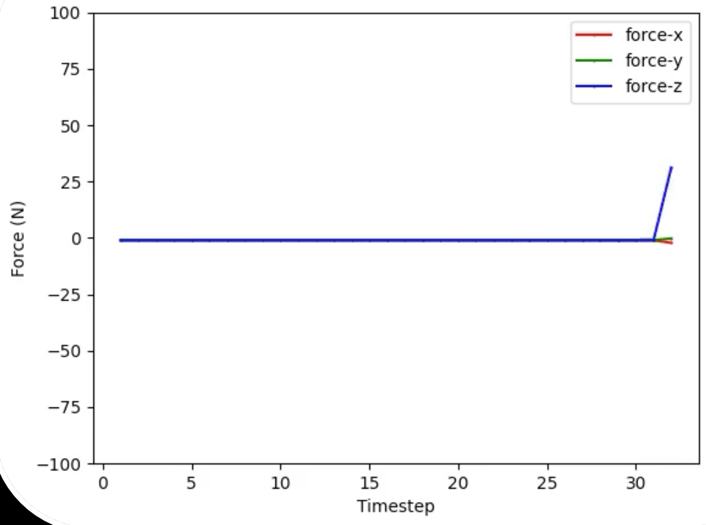
Canonical

Task Variation

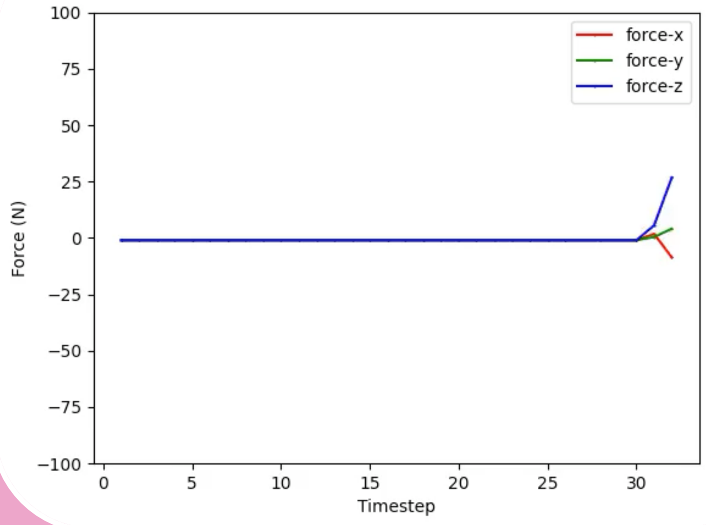
Camera Angle

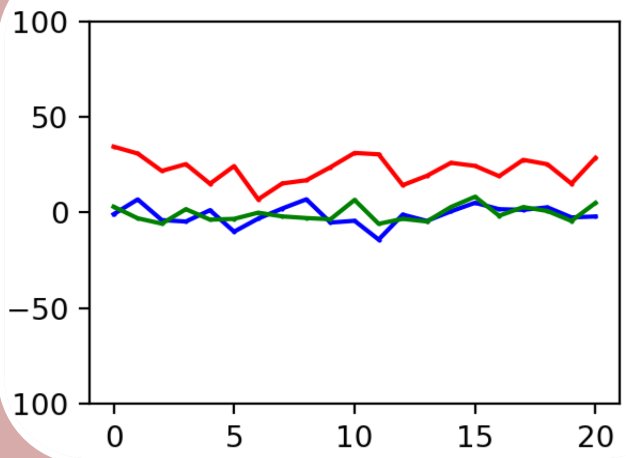
Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

Canonical



Task Variation





Sensor Noise

Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

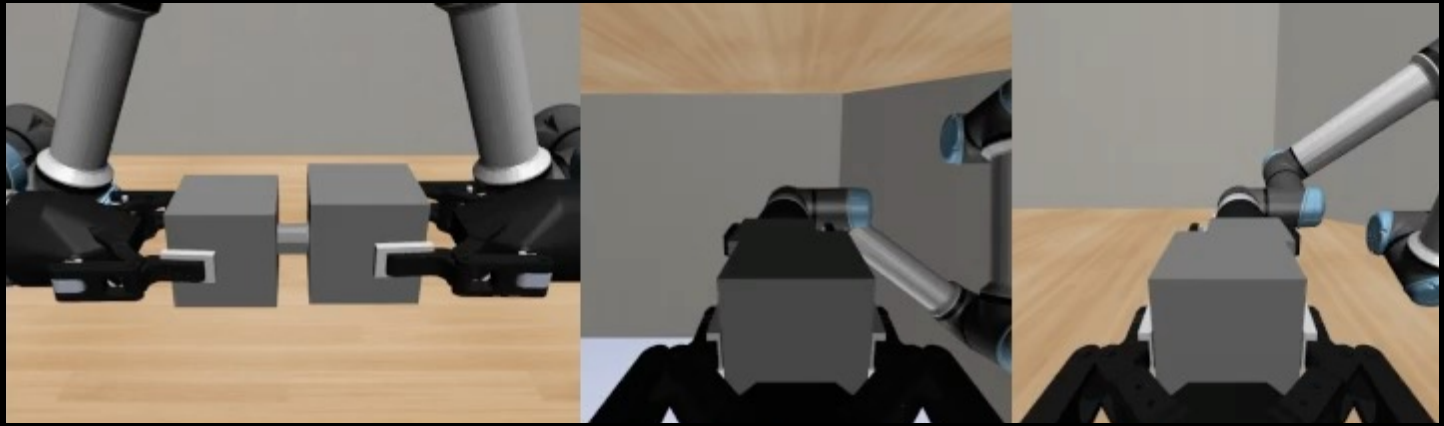
Data Collection

Teleoperated demonstrations collected in "canonical" (no variations) environment

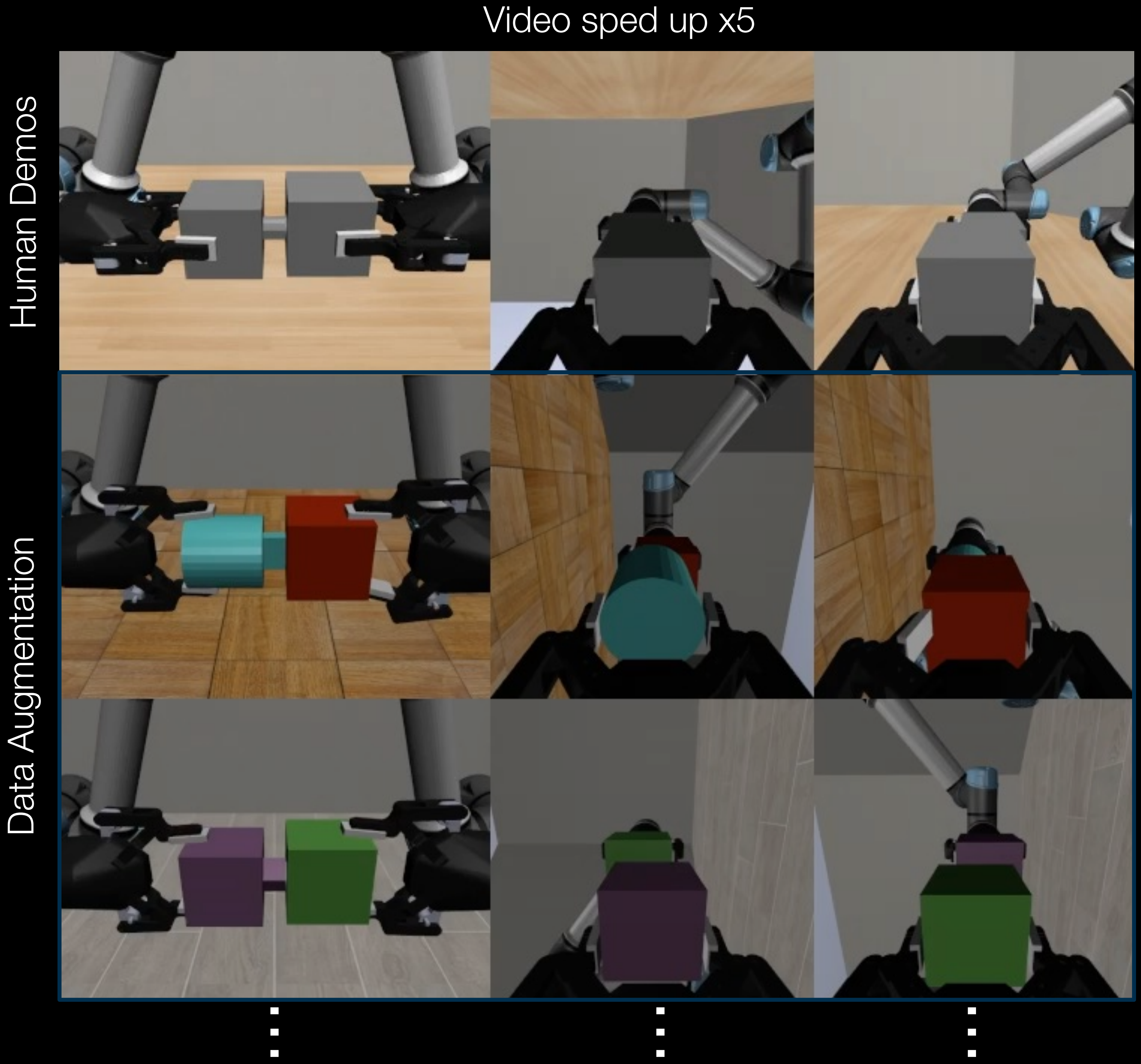
Video sped up x4

Video sped up x5

Human Demos



Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).



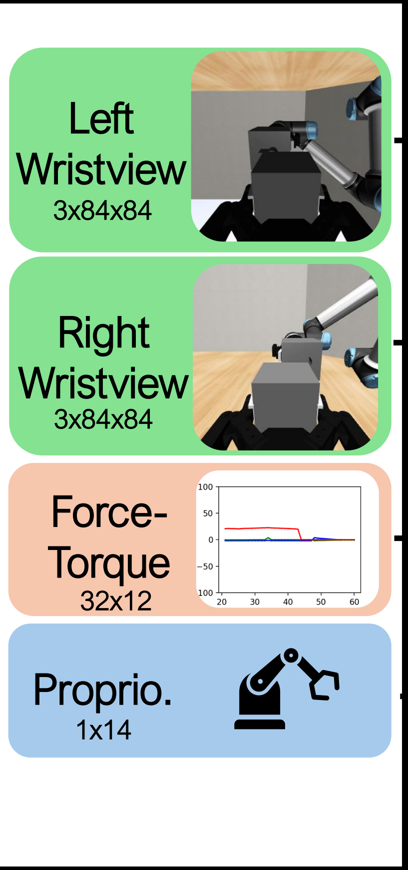
Data Collection

Online data augmentation via
trajectory replay on environments with
task variations applied

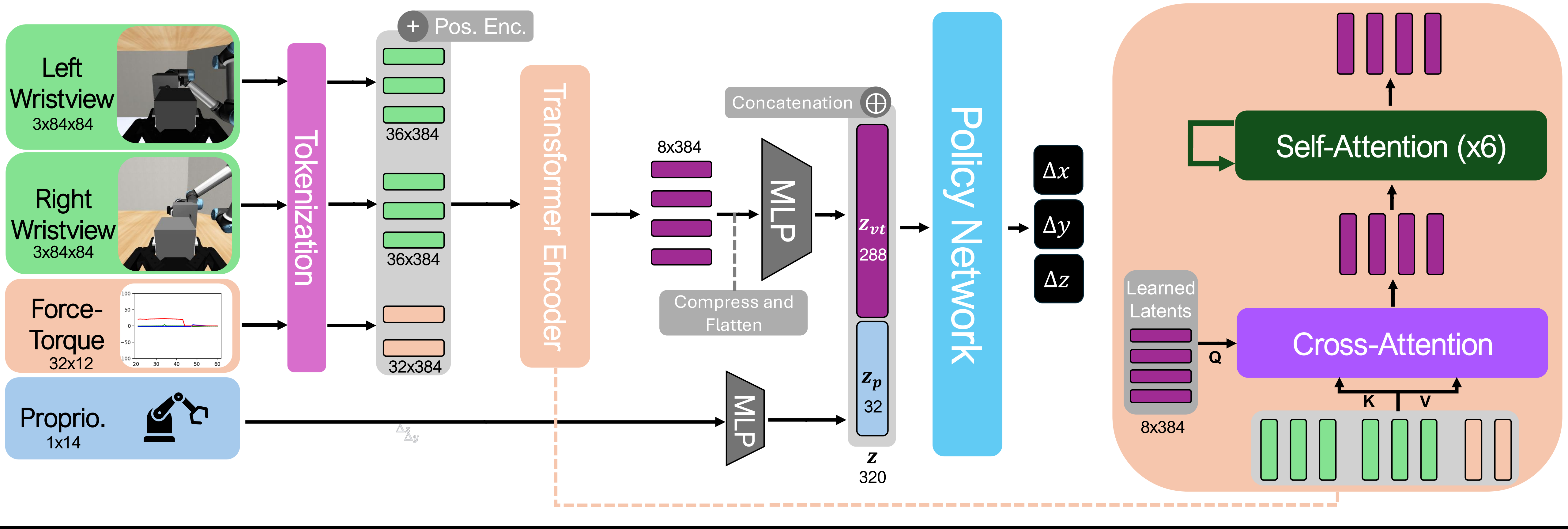
Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

Model Architecture

Model Architecture



Model Architecture

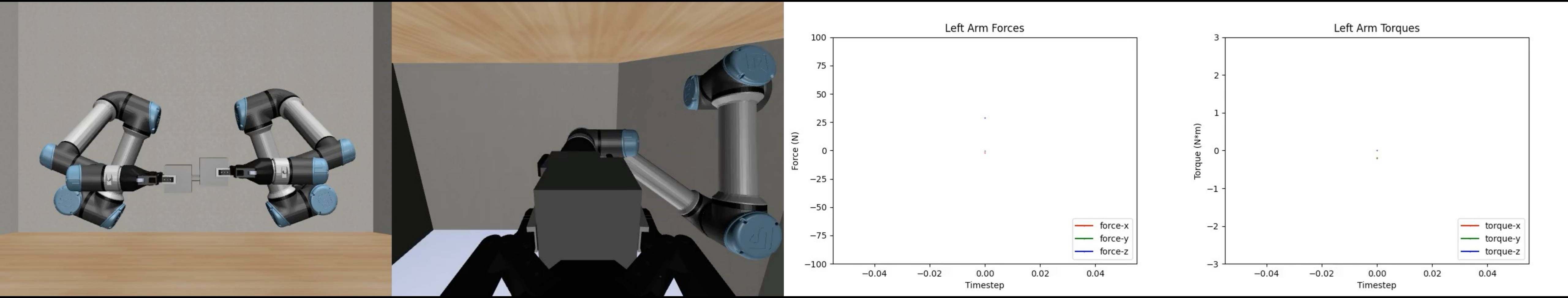


Feed output embedding into MLP policy network, predict actions

Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

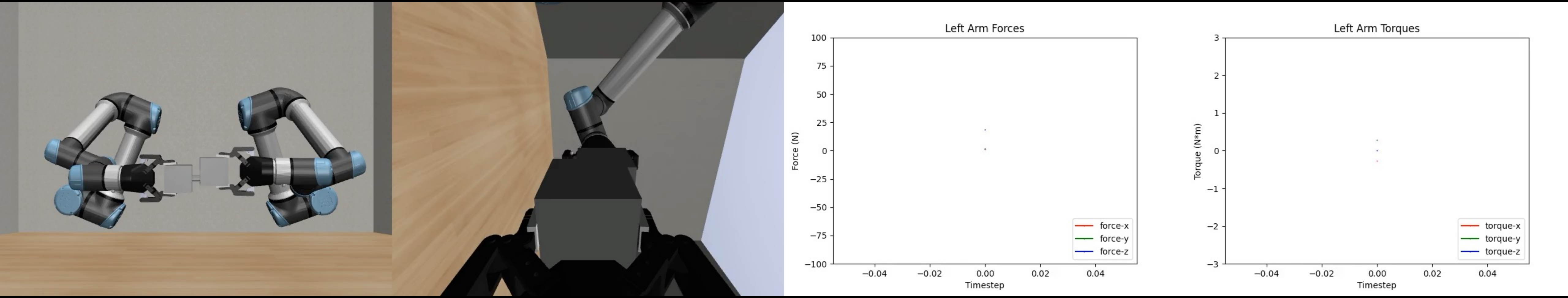
Policy Trained on No Variations

Frontview (Visualization) | Wristview (Part of Policy Input) | Force-Torque (Part of Policy Input)



Rollouts in *Canonical* environment: **0.980** mean success rate*

*Success rates over 50 rollouts averaged over 6 training seeds

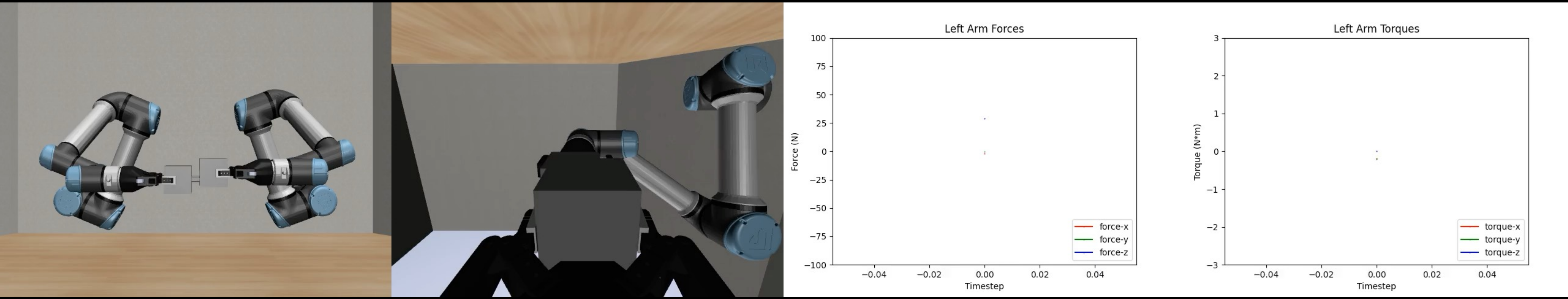


Rollouts in *Grasp Pose* environment: **0.060** mean success rate*

Videos sped up x5

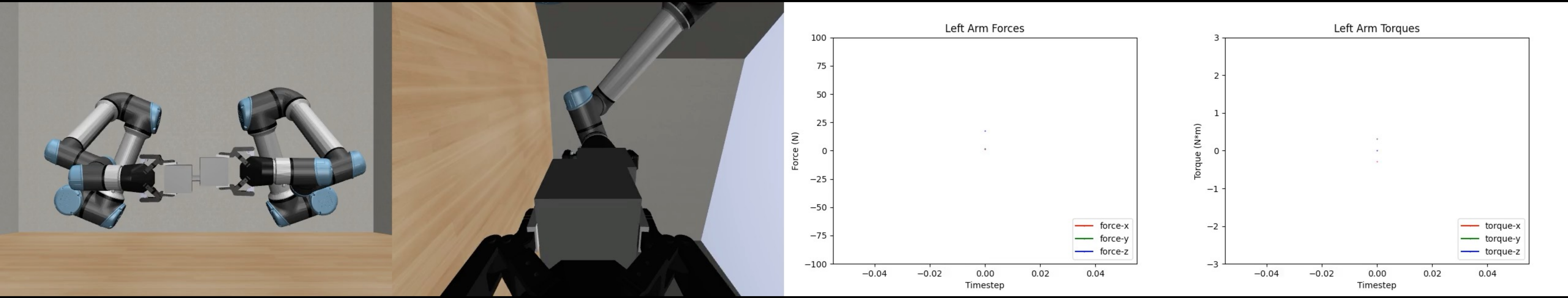
Policy Trained on Visual Variations + Sensor Noise

Frontview (Visualization) | Wristview (Part of Policy Input) | Force-Torque (Part of Policy Input)



Rollouts in *Canonical* environment: **0.973** mean success rate*

*Success rates over 50 rollouts averaged over 6 training seeds



Rollouts in *Grasp Pose* environment: **0.173** mean success rate*

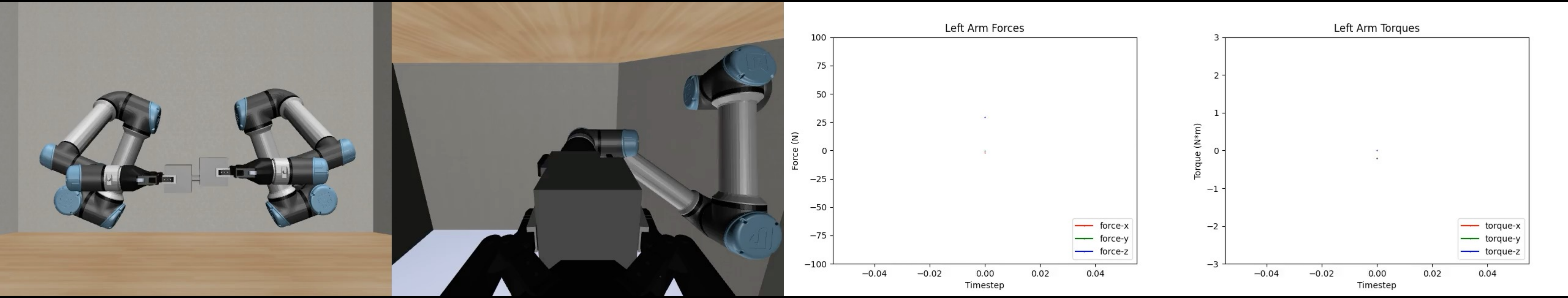
Videos sped up x5

Policy Trained on Object Shape + Grasp Variations

Frontview
(Visualization)

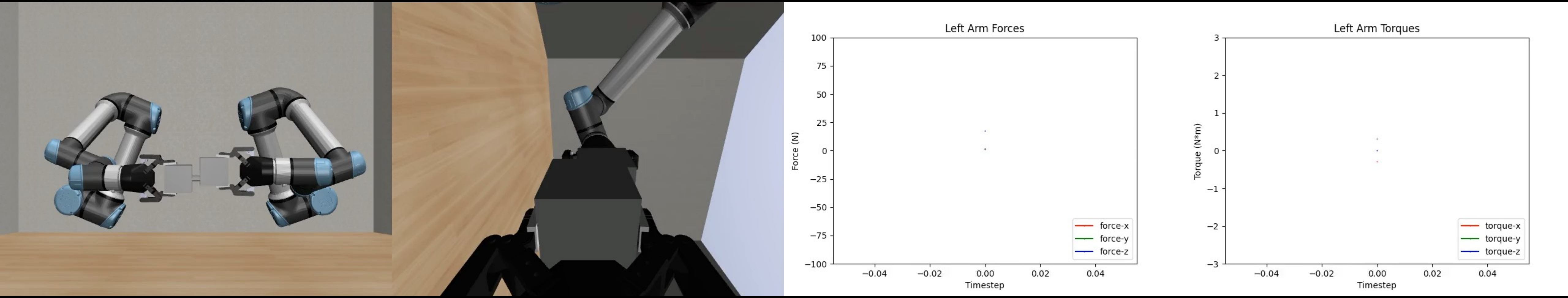
Wristview
(Part of Policy Input)

Force-Torque
(Part of Policy Input)



Rollouts in *Canonical* environment: **0.957** mean success rate*

*Success rates over 50 rollouts averaged over 6 training seeds

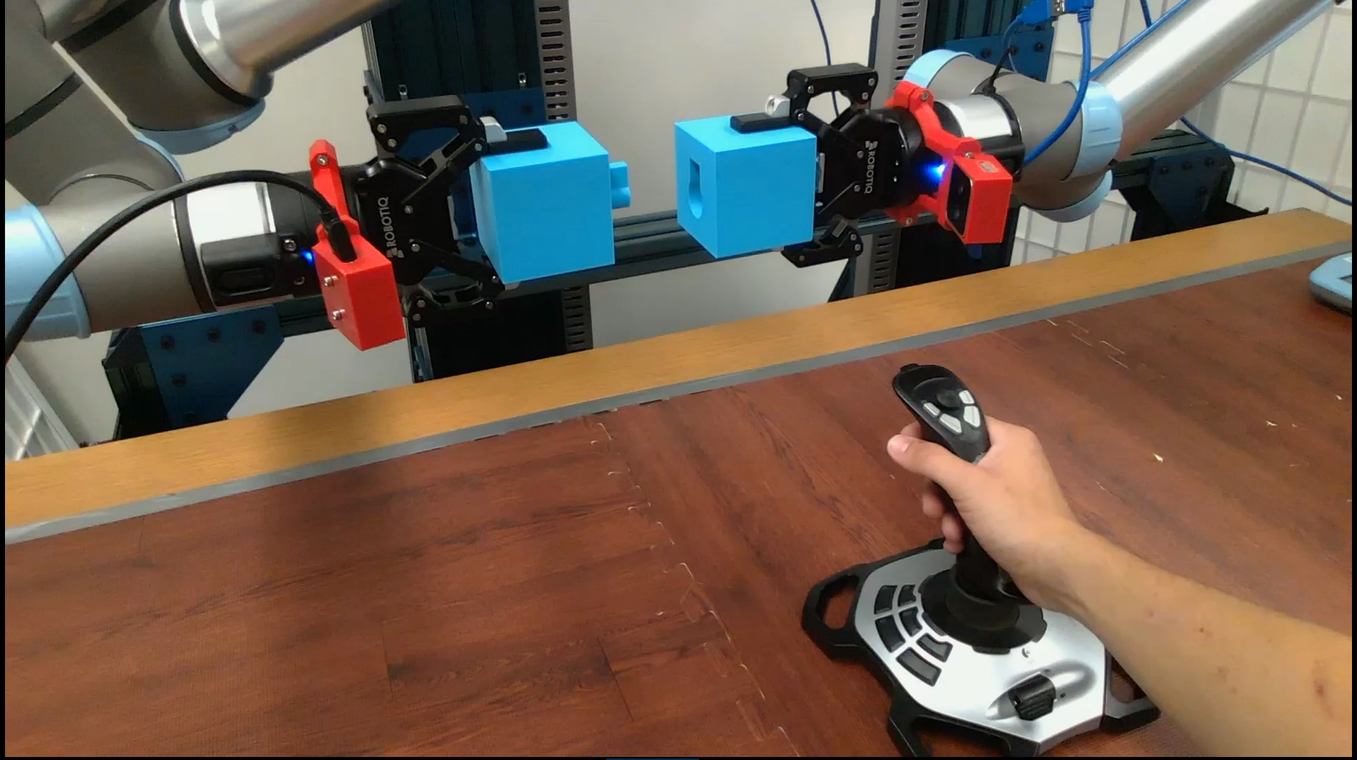


Rollouts in *Grasp Pose* environment: **0.620** mean success rate*

Videos sped up x5

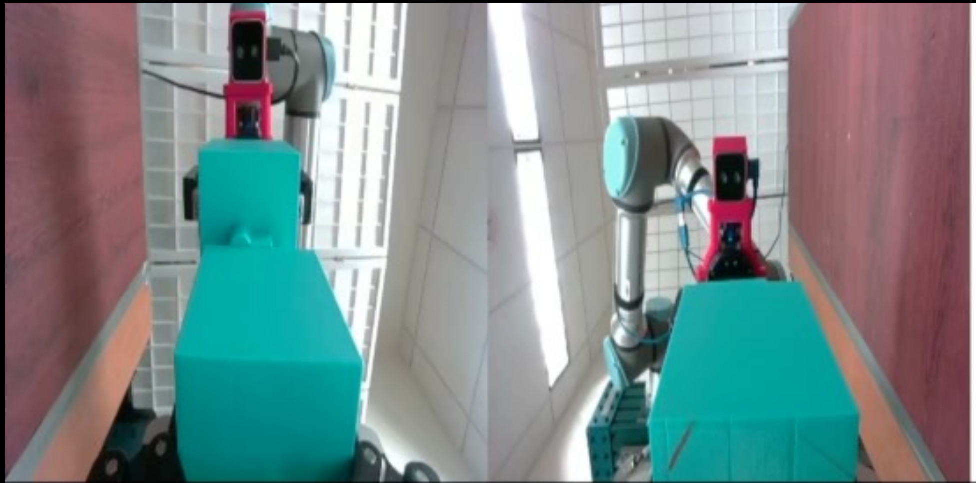
Real World Data Collection

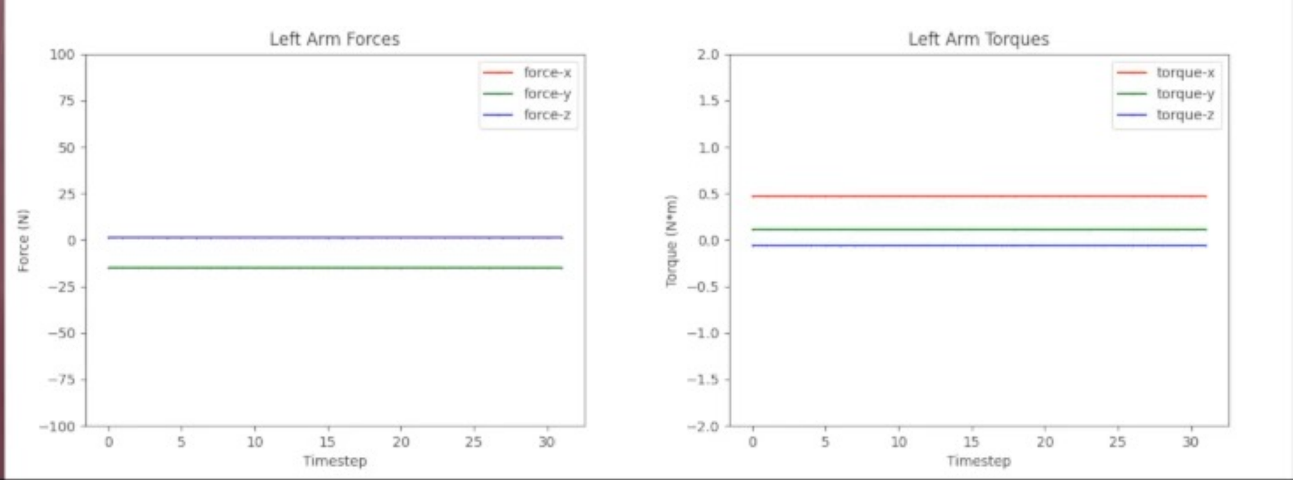
Teleoperated demonstrations collected in “canonical” (no variations) environment



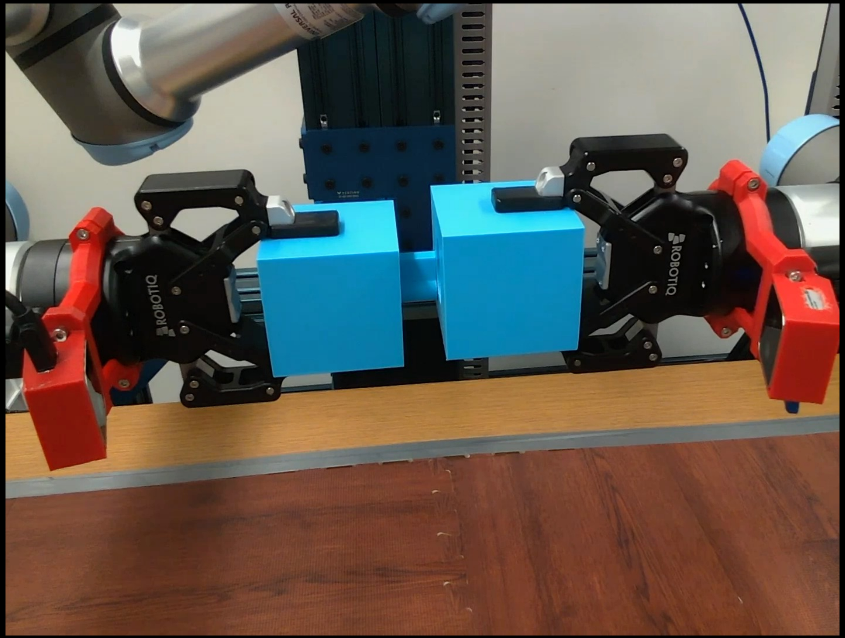
Videos sped up x4

Human Demos

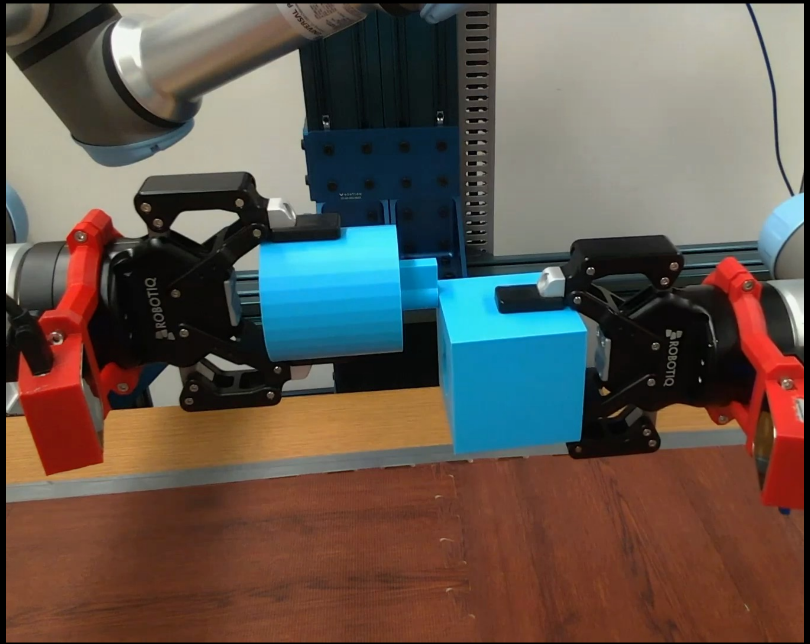




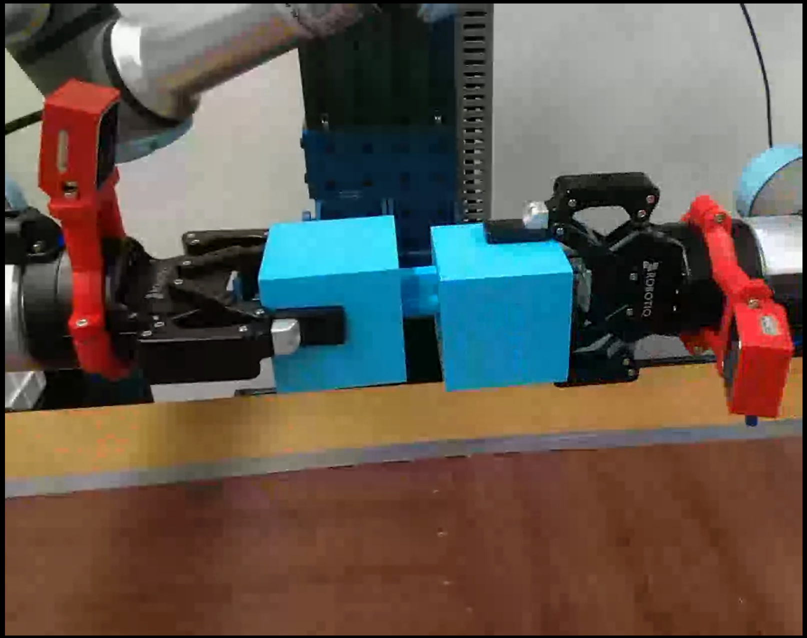
Real World Evaluation: Policy Trained on No Variations



Rollouts in *Canonical* environment
0.900 success rate*



Rollouts in *Object Body Shape* environment
0.800 success rate*



Rollouts in *Grasp Pose* environment
0.150 success rate*

Data augmentation may be necessary to improve robustness

Diaz, Ryan, Adam Imdieke, Vivek Veeriah, and Karthik Desingh. "AugInsert: Learning Robust Visual-Force Policies via Data Augmentation for Object Assembly Tasks." *arXiv preprint arXiv:2410.14968* (2024).

*Success rates over 20 rollouts

Videos sped up x2

Evidence from neuroscience and cognitive science supports the notion that humans employ spatio-geometric features, mediated by specific neural pathways and cognitive processes, to perform object assembly tasks.

We posit that robots need representations that can capture spatio-geometric features to learn novel-object assembly skills from demonstrations.

- No explicit object-geometric knowledge
- Maybe visual information is good enough ← **Requires F/T sensing**
- Maybe pretrained visual representations are good enough to give us these features.

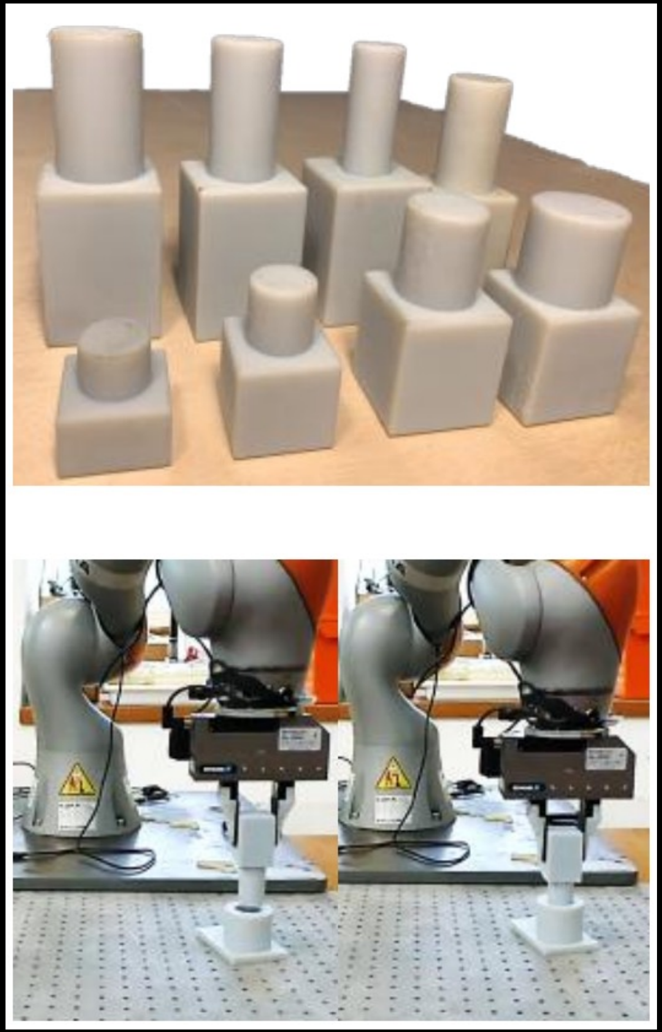
Not necessarily true due to distributional shift

Take aways!

- Grasp variations are hard to be robust to in object assembly learning.
- Visual pre-trained models on internet data is not necessarily best for object assembly – probably not the spatio-geometric features we hoped for.
- Force-Torque data is vital during the contact-rich phase of object assembly. Spatio-geometric feature learning should also incorporate tactile information.
- Data augmentation tricks can help with accommodating observation level variations in object assembly task.

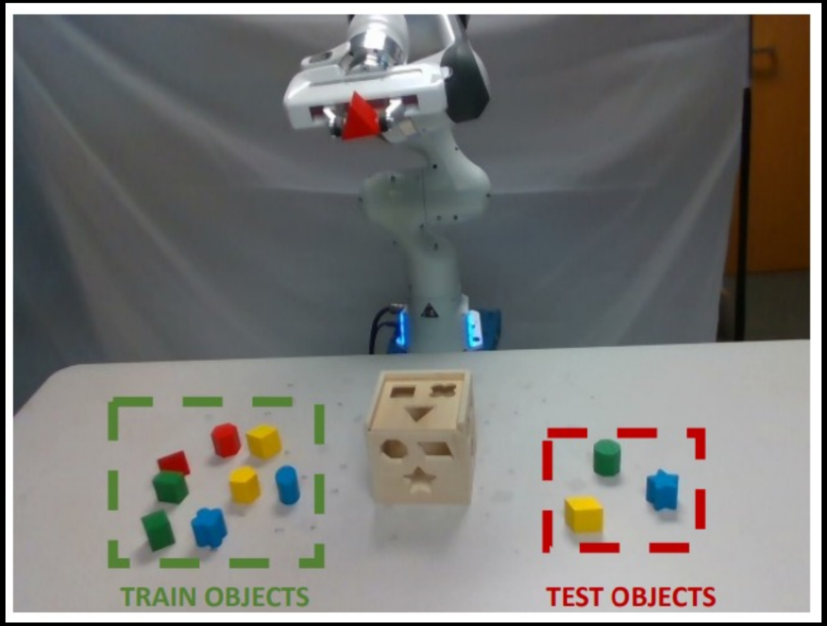
Some Related Works

Peg-hole Insertion



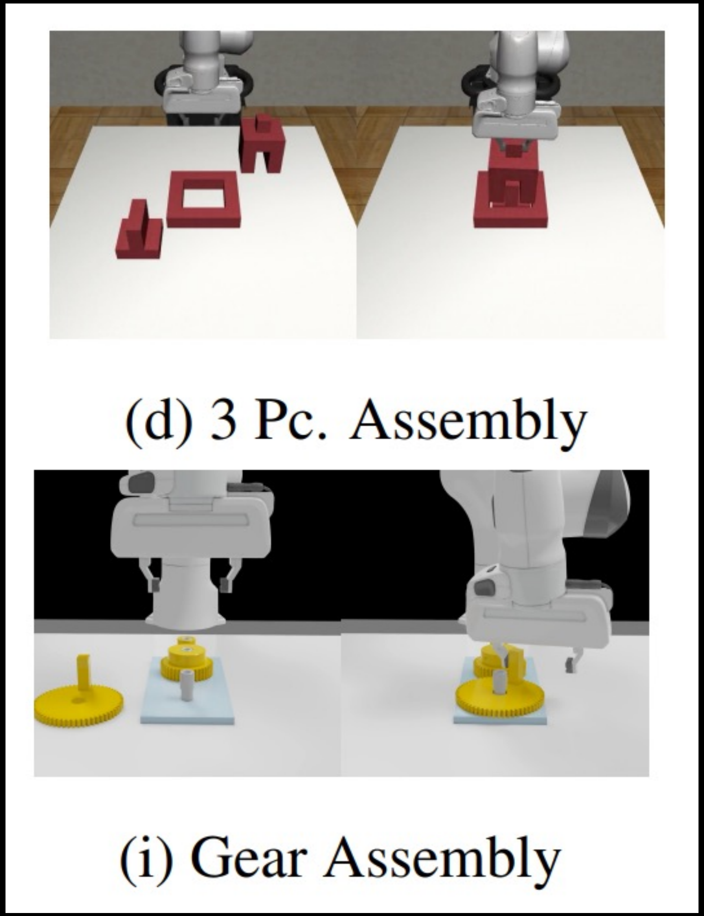
Gao, Wei, and Russ Tedrake. "kpam 2.0: Feedback control for category-level robotic manipulation." *IEEE Robotics and Automation Letters* 6, no. 2 (2021): 2962-2969.

Shape Sorting



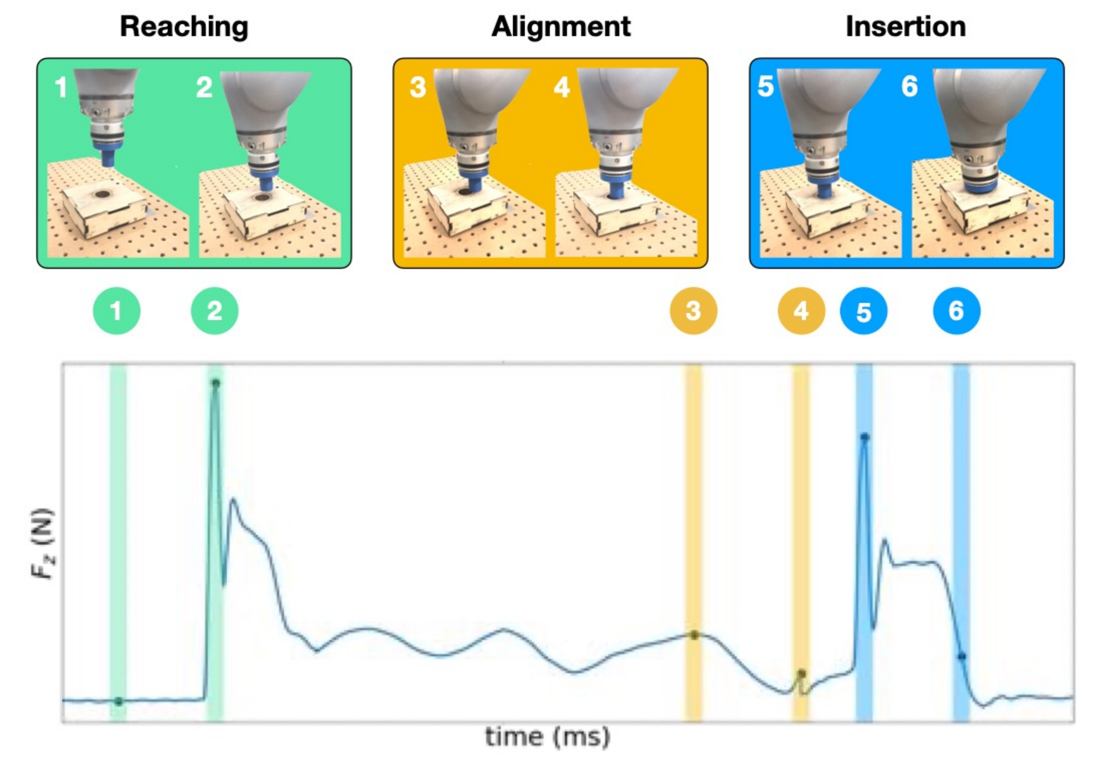
Dasari, Sudeep, Jianren Wang, Joyce Hong, Shikhar Bahl, Yixin Lin, Austin S. Wang, Abitha Thankaraj et al. "RB2: Robotic Manipulation Benchmarking with a Twist." In *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Multi-part Assembly

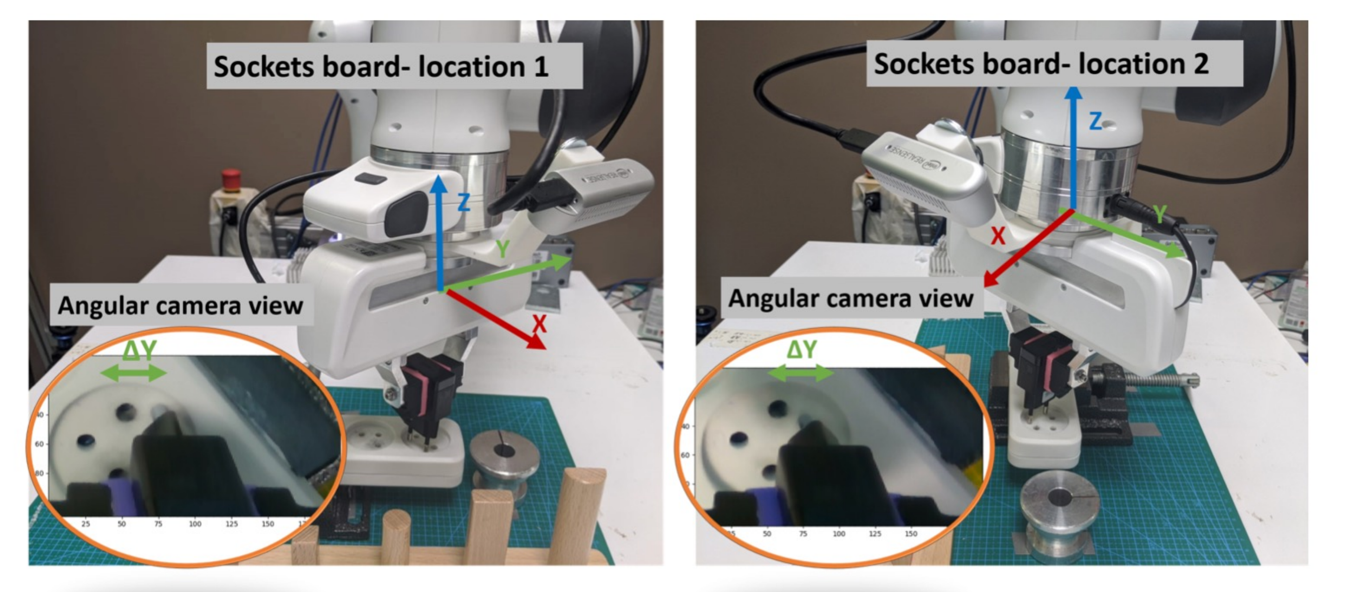


Mandlekar, Ajay, Soroush Nasiriany, Bowen Wen, Iretiayo Akinola, Yashraj Narang, Linxi Fan, Yuke Zhu, and Dieter Fox. "MimicGen: A Data Generation System for Scalable Robot Learning using Human Demonstrations." In *7th Annual Conference on Robot Learning*.

Some Related Works



M. A. Lee, Y. Zhu, K. Srinivasan, P. Shah, S. Savarese, L. Fei-Fei, A. Garg, and J. Bohg, "Making sense of vision and touch: Self- supervised learning of multimodal representations for contact rich tasks," in 2019 International conference on robotics and automation (ICRA). IEEE, 2019, pp. 8943–8950.



O. Spector and D. Di Castro, "Insertionnet-a scalable solution for insertion," IEEE Robotics and Automation Letters, vol. 6, no. 3, pp. 5509–5516, 2021.

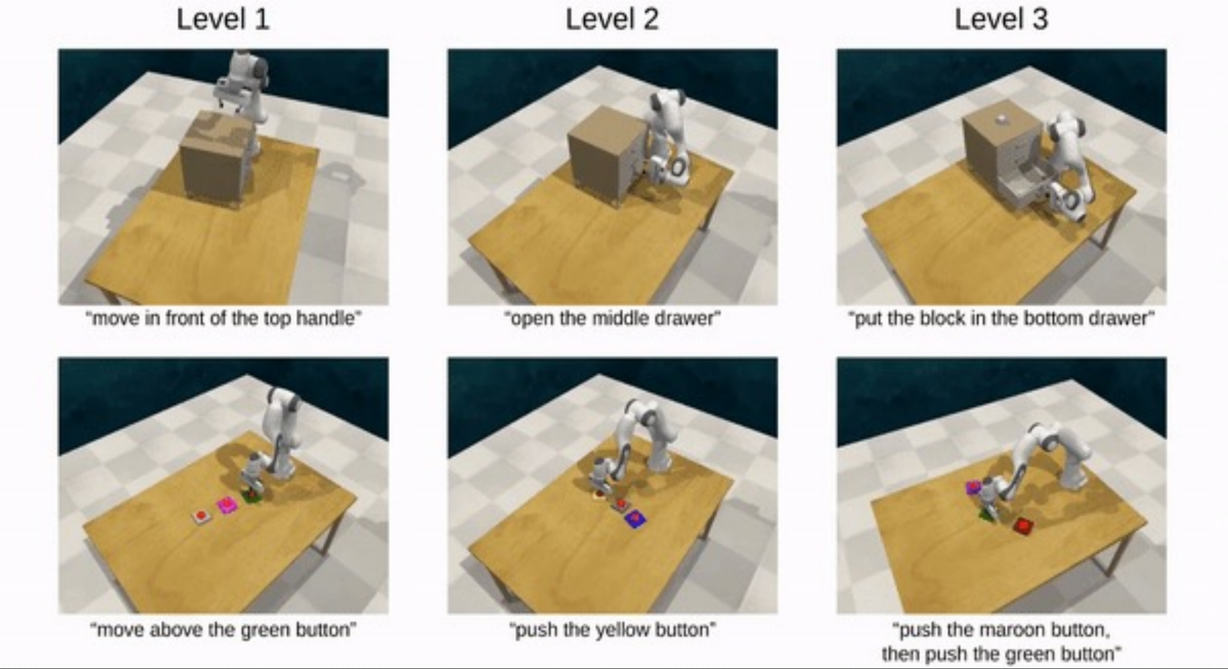
Other works from RPM Lab

Grasping for Manipulation of Larger Objects

SuperQ-GRASP: Superquadrics-based Grasp Pose Estimation on Larger Objects for Mobile-Manipulation

Xun Tu and Karthik Desingh
University of Minnesota Twin Cities

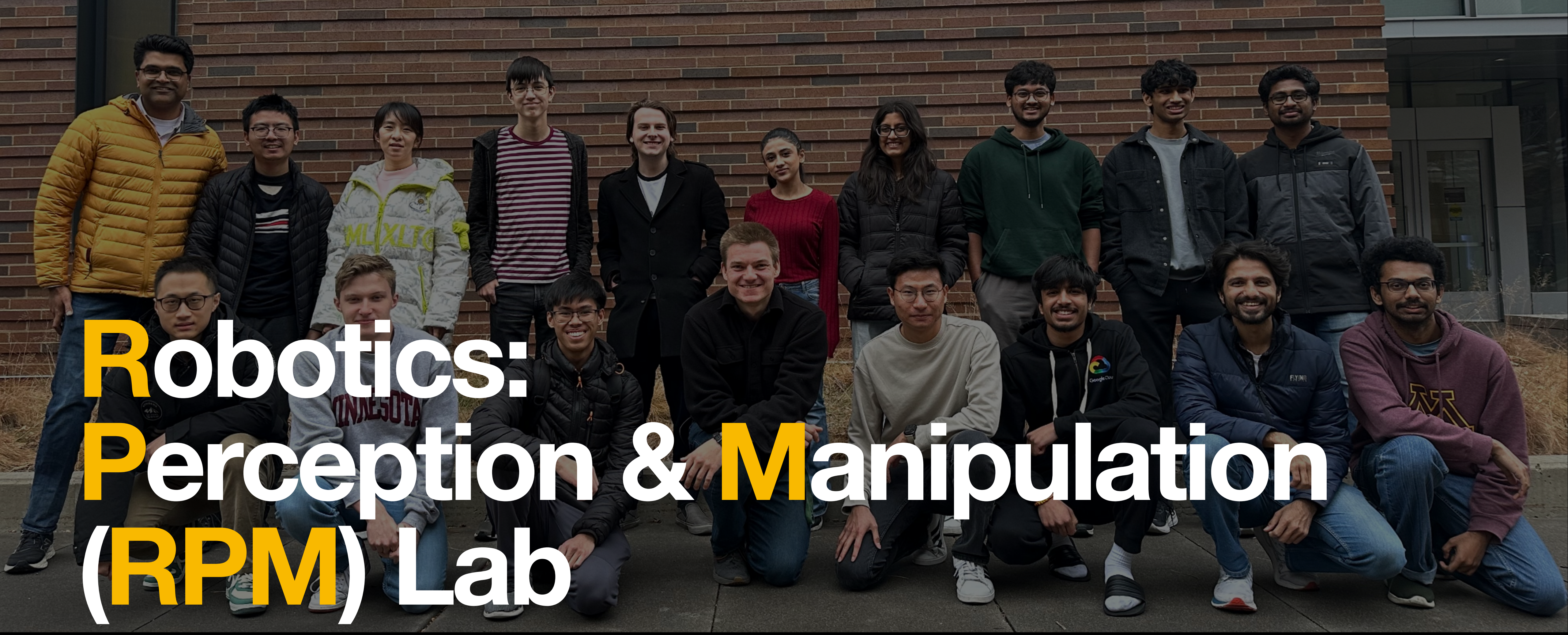
End User Directed Robot Learning Via Natural Language Based Interaction



IEEE ROBOTICS AND AUTOMATION LETTERS, VOL. 9, NO. 9, SEPTEMBER 2024 8051

Talk Through It: End User Directed Manipulation Learning

Carl Winge, Adam Imdieke, Bahaa Aldeeb, Dongyeop Kang, and Karthik Desingh, *Member, IEEE*



Robotics: Perception & Manipulation (RPM) Lab

Karthik Desingh
Assistant Professor, University of Minnesota
Minnesota Robotics Institute (MnRI)
Department of Computer Science and Engineering