Structure from Action:

Articulated Object Structure Discovery with Active Interactions

Structural and Compositional Learning on 3D Data workshop @ ICCV

Shuran Song

Columbia University
Artificial Intelligence & Robotics Lab

Structure from Action:

Articulated Object Structure Discovery with Active Interactions

Structural and Compositional Learning on 3D Data workshop @ ICCV

Shuran Song

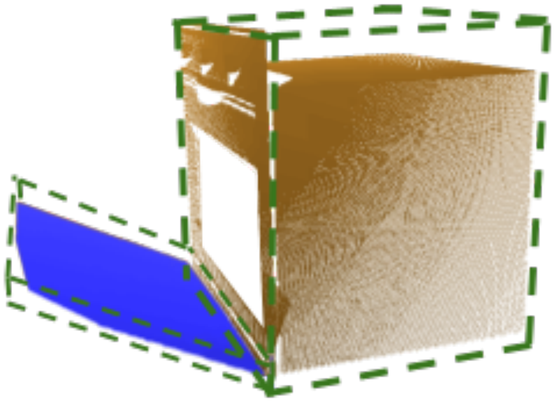
Columbia University
Artificial Intelligence & Robotics Lab

Structures in Articulated Object

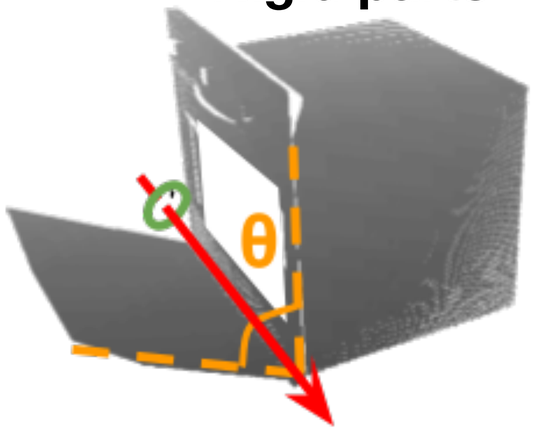
Articulated Objects

Structures in Articulated Object

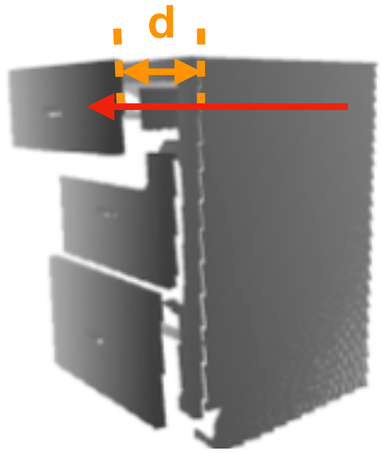
Articulated Objects



Rigid parts



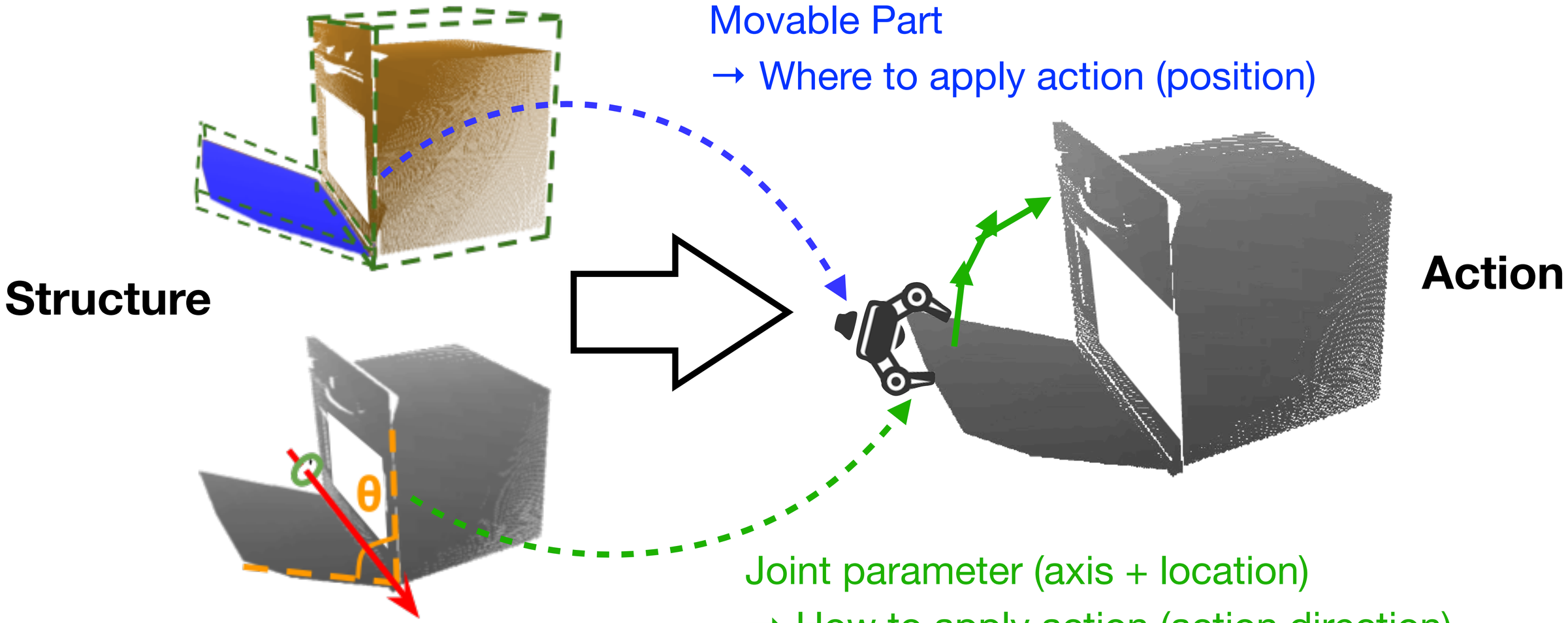
Joints (revolute)



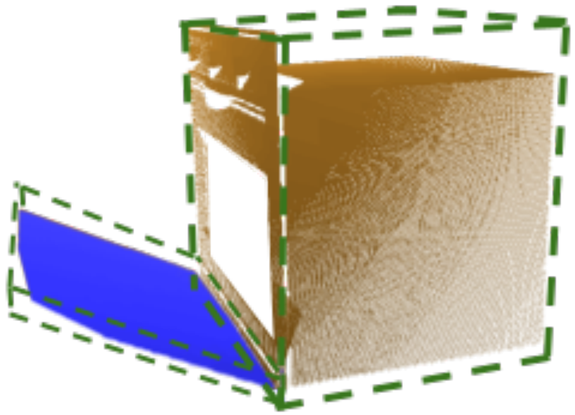
Joint (Prismatic)

Kinematic Structure

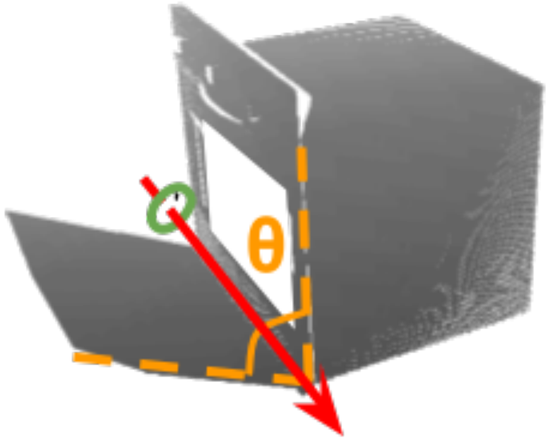
Structure → Action



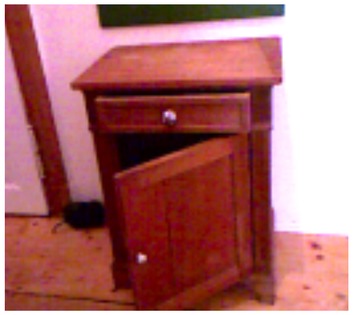
Structure → Action

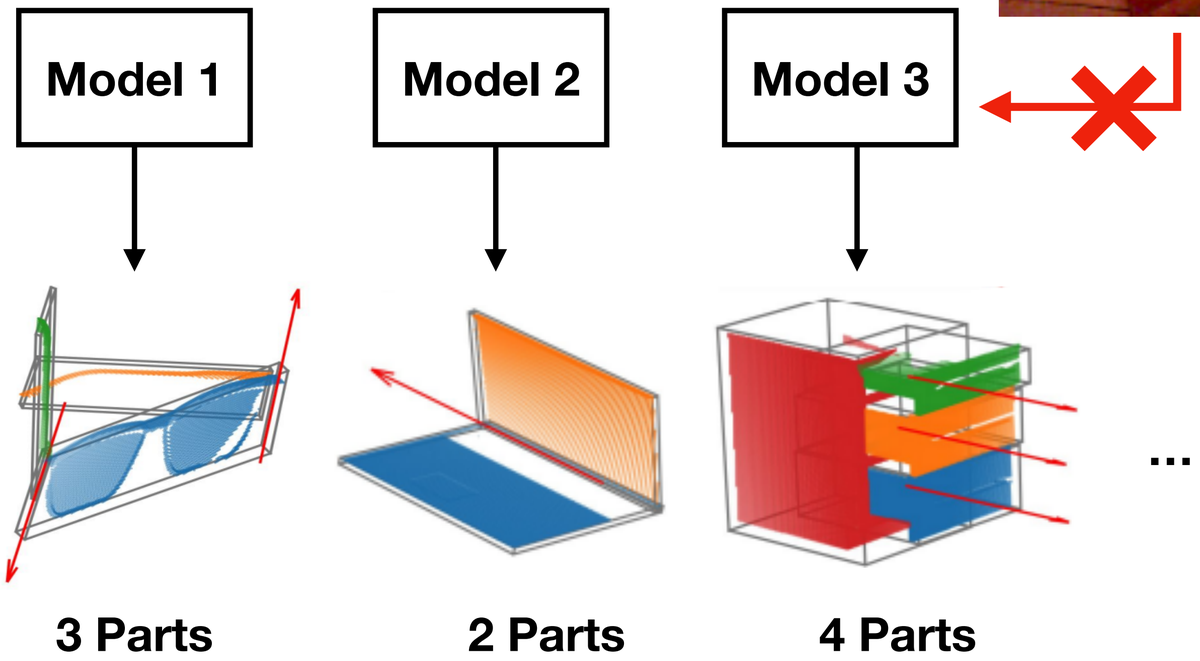


Structure



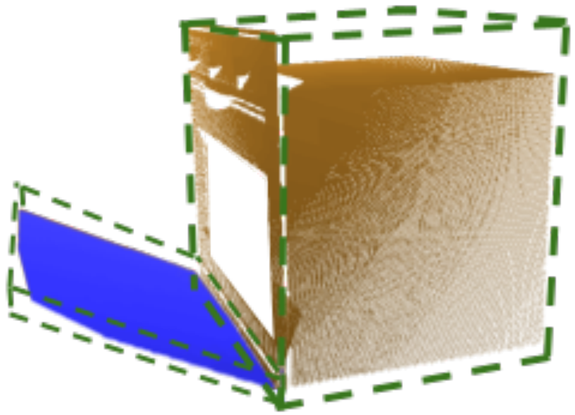
Inferring structure from a single image





* Category-Level Articulated Object Pose Estimation. Li et al, CVPR 2019

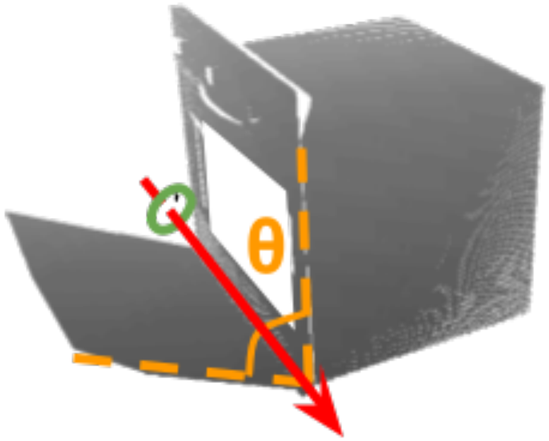
Action → Structure

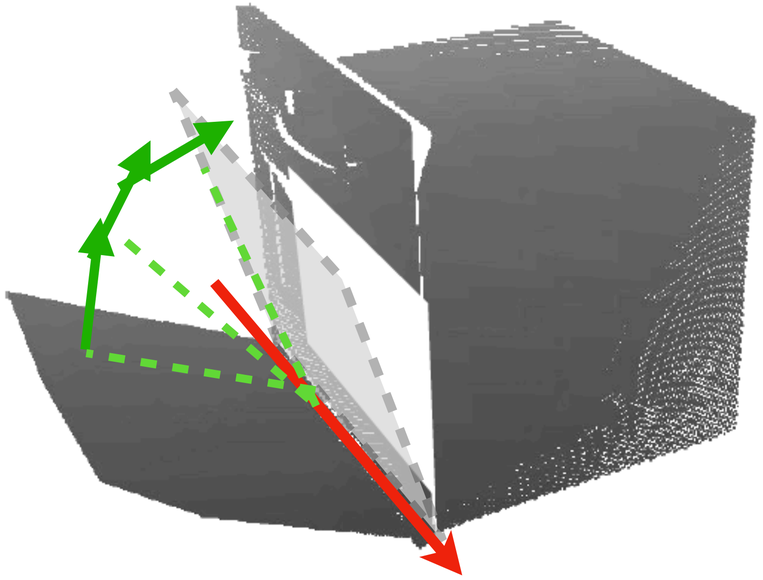


Rigid Motion

→ Part segmentation

Structure



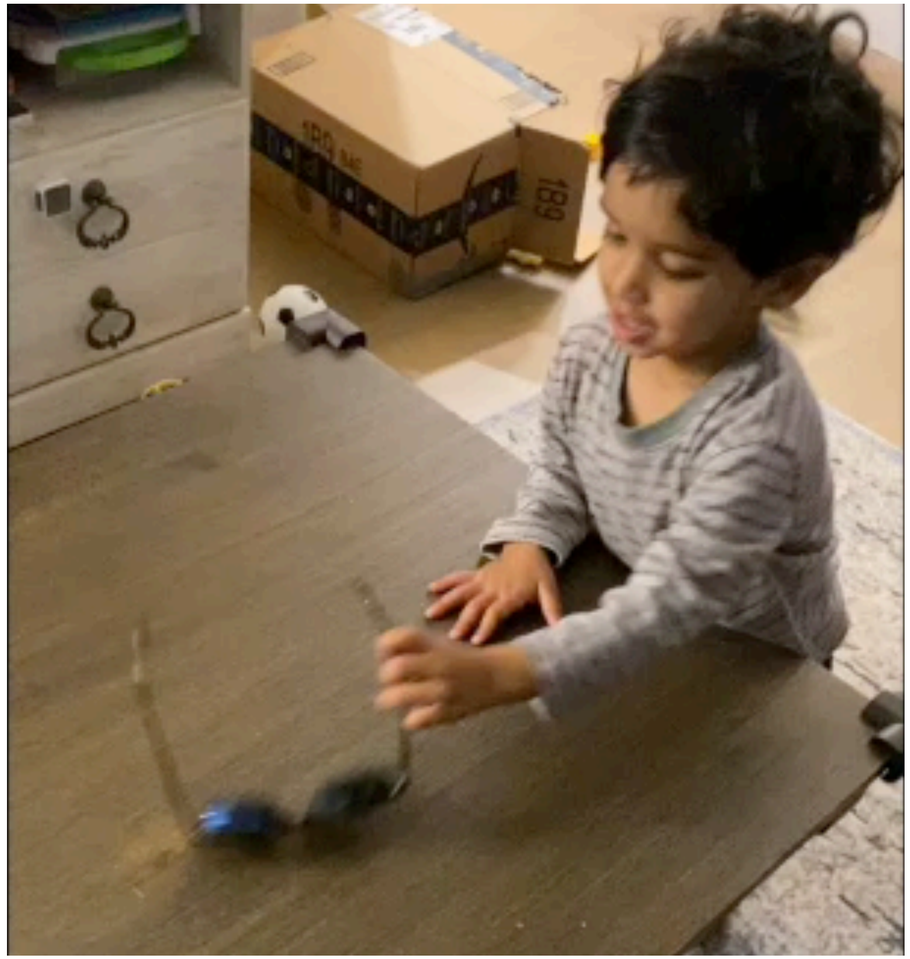


Action

Motion trajectory

→ Joint axis location and orientation

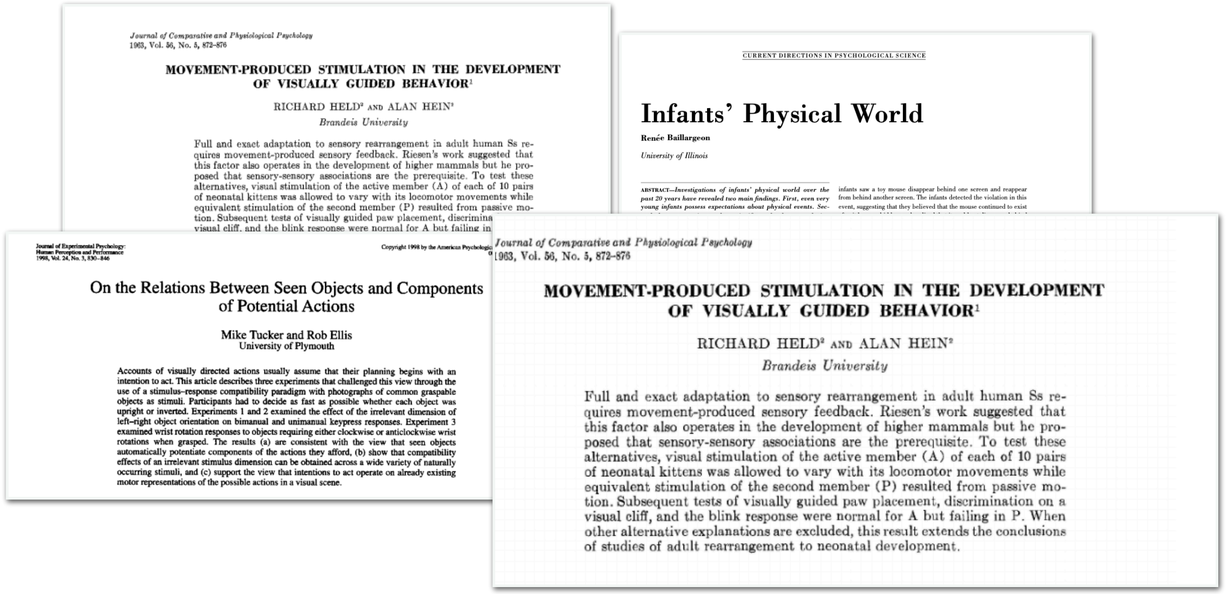
Action → Structure



Video Credit: Samir Y. Gadre

How do kids learn to understand an articulated objects?

Interacting + observing!



Journal of Comparative and Physiological Psychology
1963, Vol. 56, No. 5, 872-876

MOVEMENT-PRODUCED STIMULATION IN THE DEVELOPMENT OF VISUALLY GUIDED BEHAVIOR[1]

RICHARD HELD[2] AND ALAN HEIN[3]

Brandeis University

Full and exact adaptation to sensory rearrangement in adult human Ss requires movement-produced sensory feedback. Riesen's work suggested that this factor also operates in the development of higher mammals but he proposed that sensory-sensory associations are the prerequisite. To test these alternatives, visual stimulation of the active member (A) of each of 10 pairs of neonatal kittens was allowed to vary with its locomotor movements while equivalent stimulation of the second member (P) resulted from passive motion. Subsequent tests of visually guided paw placement, discrimination on a visual cliff, and the blink response were normal for A but failing in P. When other alternative explanations are excluded, this result extends the conclusions of studies of adult rearrangement to neonatal development.

CURRENT DIRECTIONS IN PSYCHOLOGICAL SCIENCE

Infants' Physical World

Renée Baillargeon

University of Illinois

Journal of Experimental Psychology: Human Perception and Performance
1998, Vol. 24, No. 3, 830–846

Copyright 1998 by the American Psychological

On the Relations Between Seen Objects and Components of Potential Actions

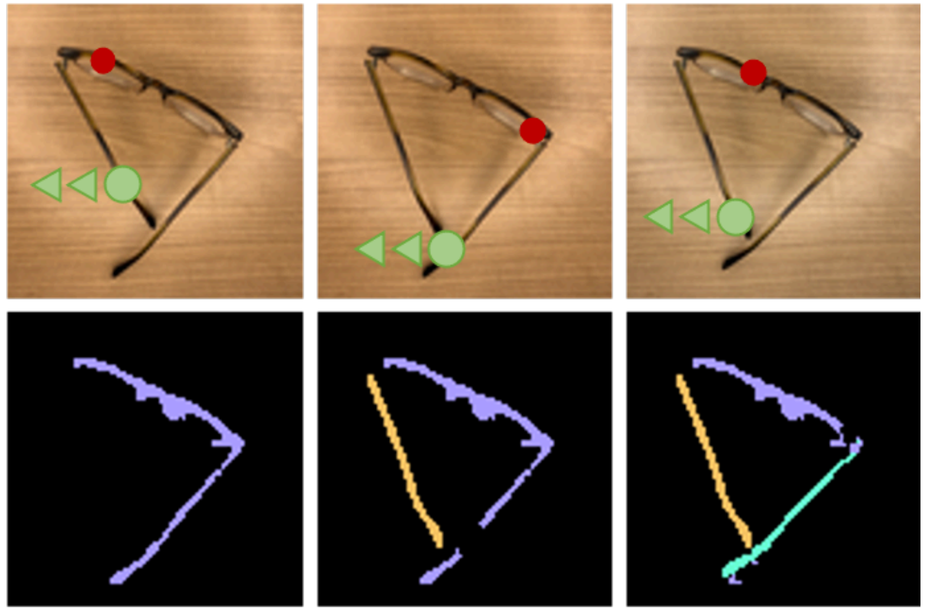
Mike Tucker and Rob Ellis
University of Plymouth

Accounts of visually directed actions usually assume that their planning begins with an intention to act. This article describes three experiments that challenged this view through the use of a stimulus–response compatibility paradigm with photographs of common graspable objects as stimuli. Participants had to decide as fast as possible whether each object was upright or inverted. Experiments 1 and 2 examined the effect of the irrelevant dimension of left–right object orientation on bimanual and unimanual keypress responses. Experiment 3 examined wrist rotation responses to objects requiring either clockwise or anticlockwise wrist rotations when grasped. The results (a) are consistent with the view that seen objects automatically potentiate components of the actions they afford, (b) show that compatibility effects of an irrelevant stimulus dimension can be obtained across a wide variety of naturally occurring stimuli, and (c) support the view that intentions to act operate on already existing motor representations of the possible actions in a visual scene.

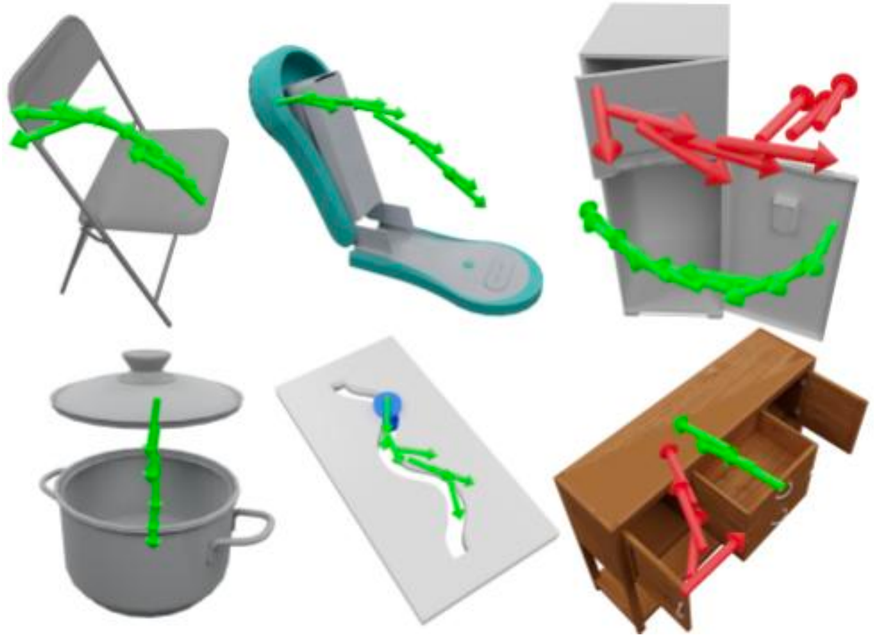
Can we allow our robot to do the same?

Structure from Action

Act the Part



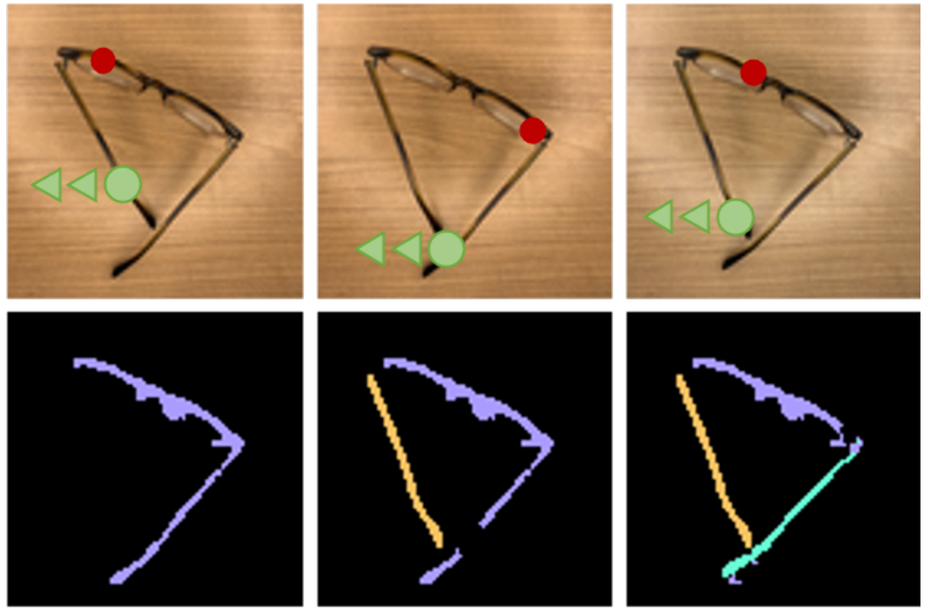
Universal Manipulation Policy



Leverage active interaction as way to discover Articulated Object Structure

Structure from Action

Act the Part



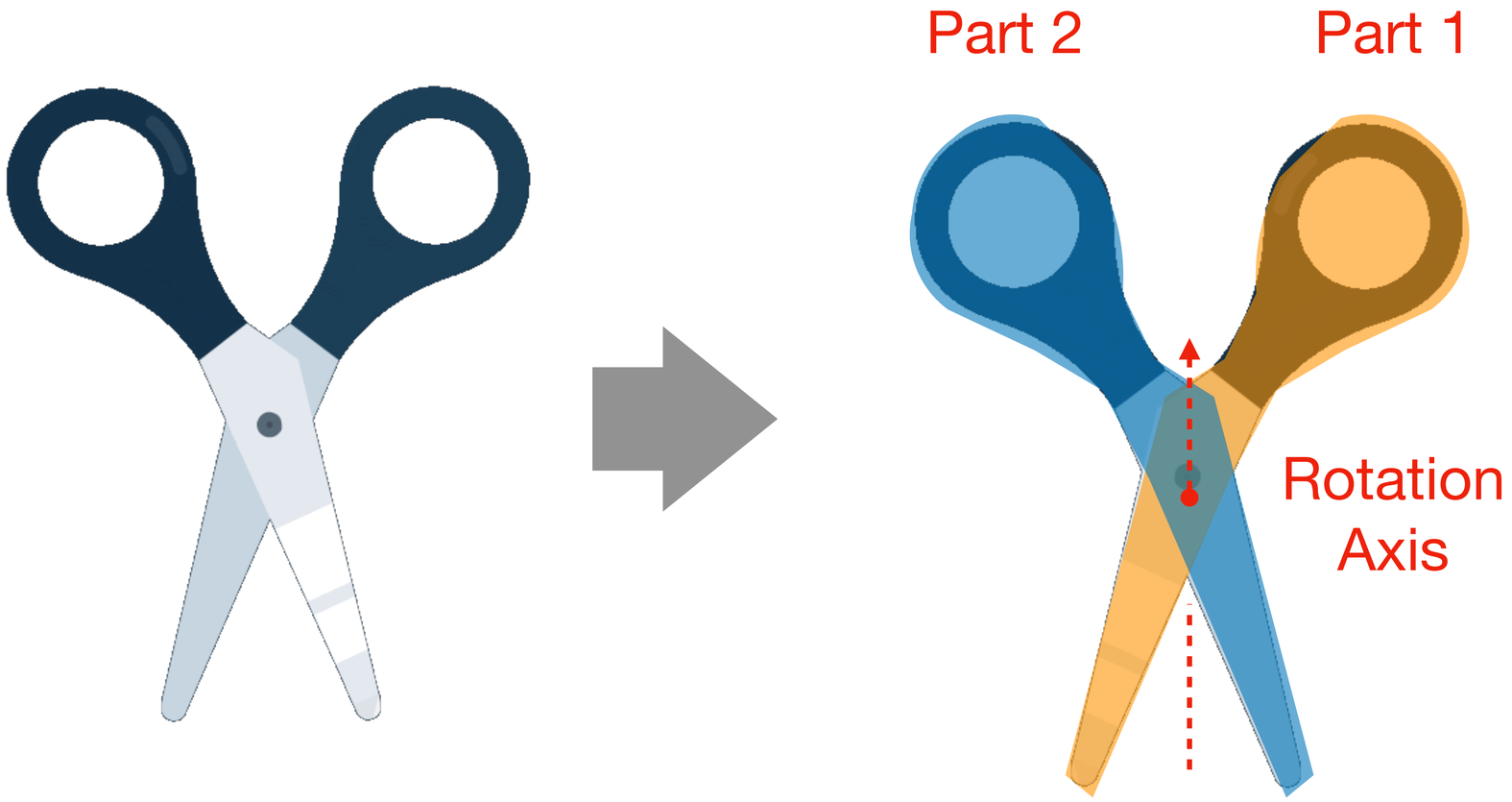
Act the Part: Learning to Interact to Discover Articulated Object Structure

ICCV 2021

Samir Y. Gadre, Kiana Ehsani, Shuran Song

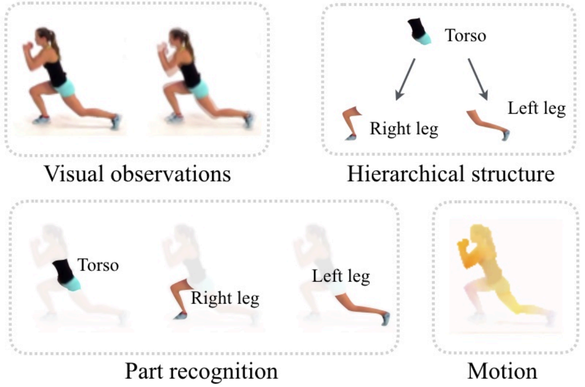
Goal

Discover the kinematic structure for articulated objects through interaction

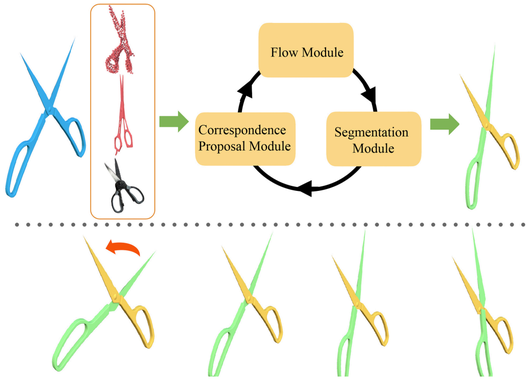


Object part discovery and segmentation with unknown objects

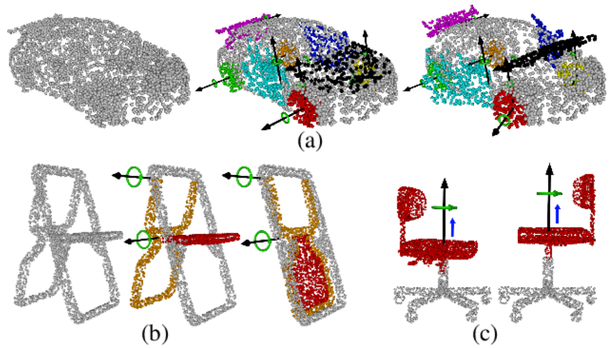
Unsupervised Part Discovery from Videos



Xu et al.

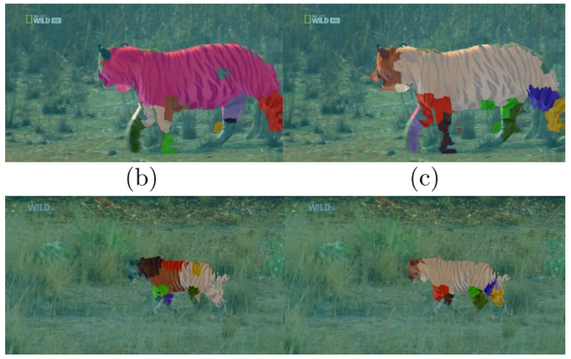


Yi et al.

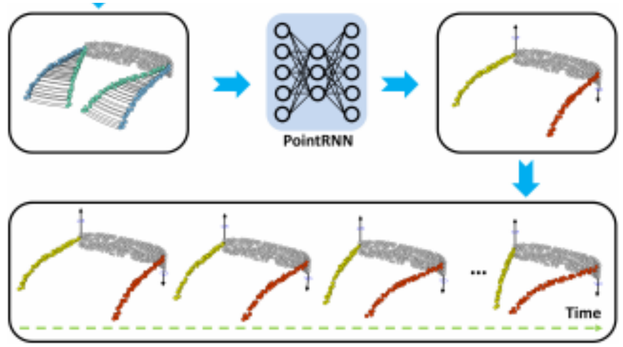


Wang et al.

Tokmakov et al.



Pero et al.

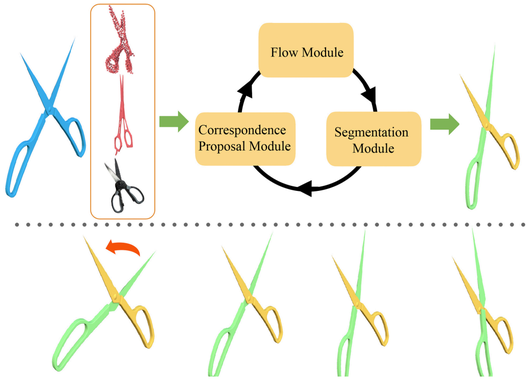


Shi et al.

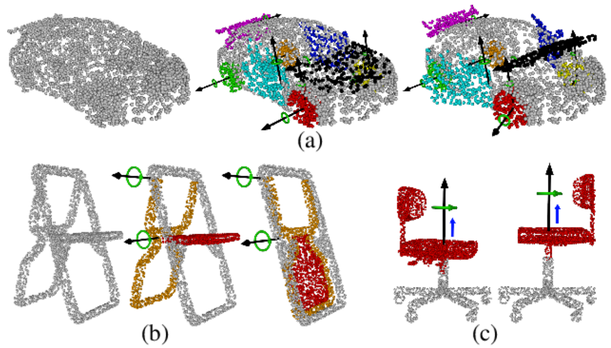
Prior works: Motion Consistency for Part Segmentation

Object Part Segmentations from Video

Who causes all these motions ?

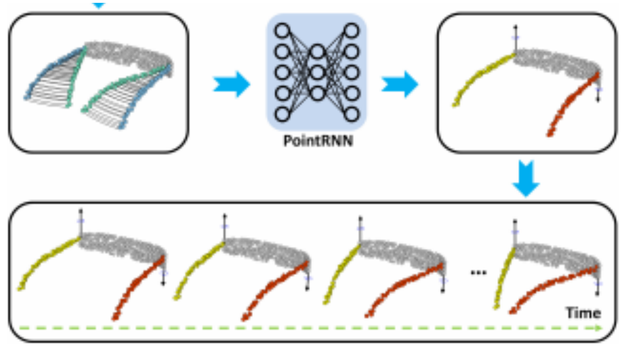


Yi et al.



Wang et al.

Most of Objects don't move by themselves ...

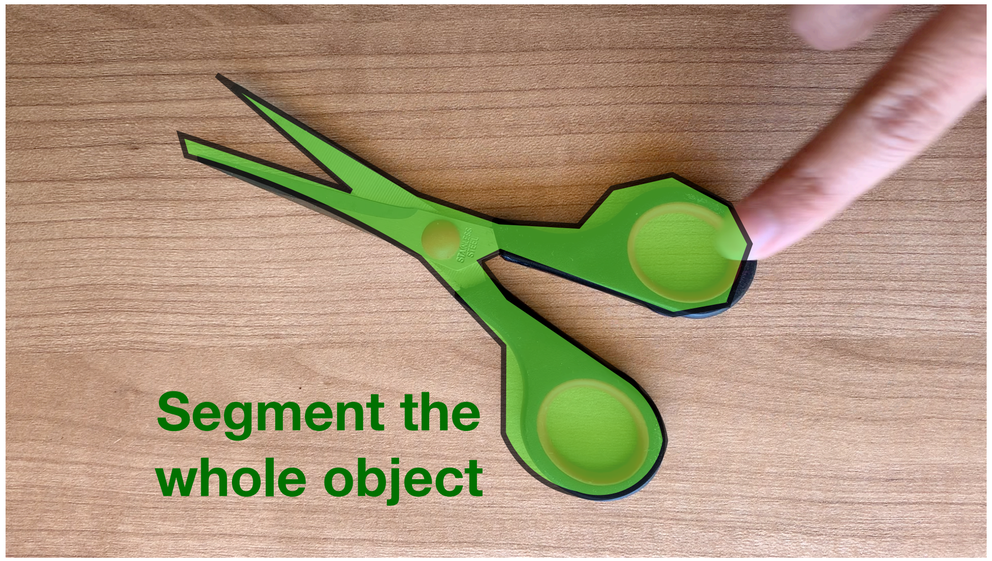


Shi et al.

Prior works: Motion Consistency for Part Segmentation

Not all Motions are Informative

Not all Motions are Informative



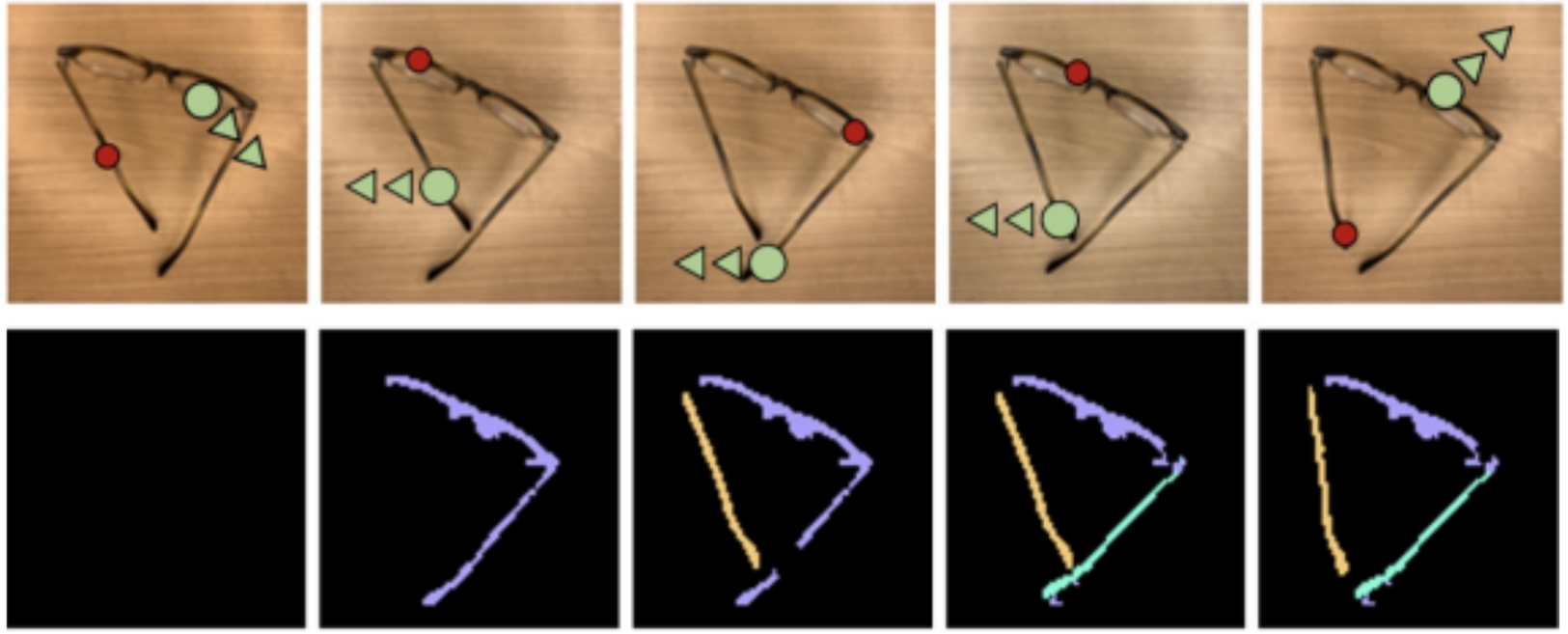
Some interactions don't give insight about articulation.

Single step interaction is not enough for multi-link object

"Informative or not" really depends on "what the system already know (i.e.,belief)"

Act the Part

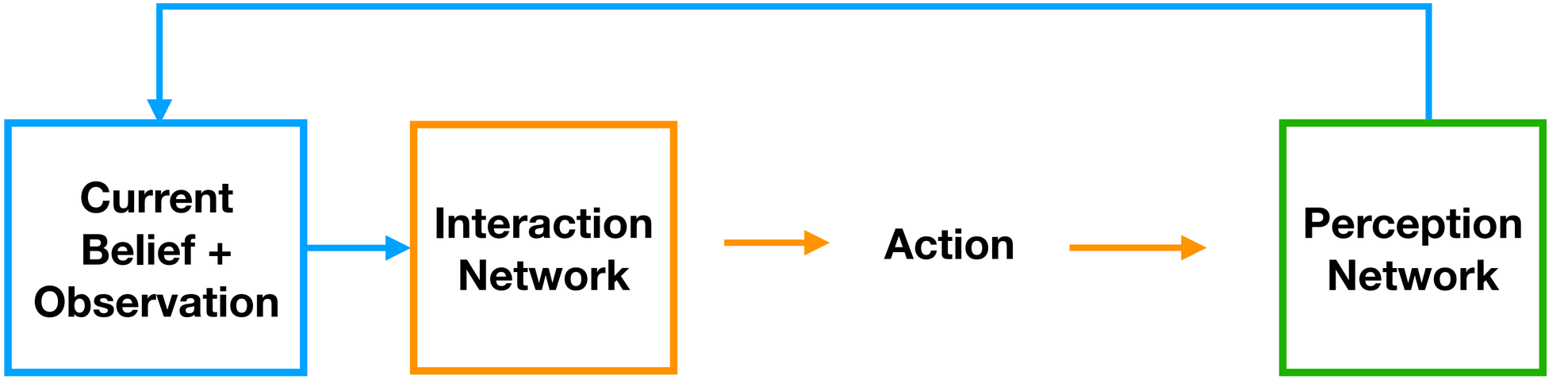
Learning to Interact to Discover Articulated Object Structure



Generate a sequence of hold and push actions that would results in informative motion.

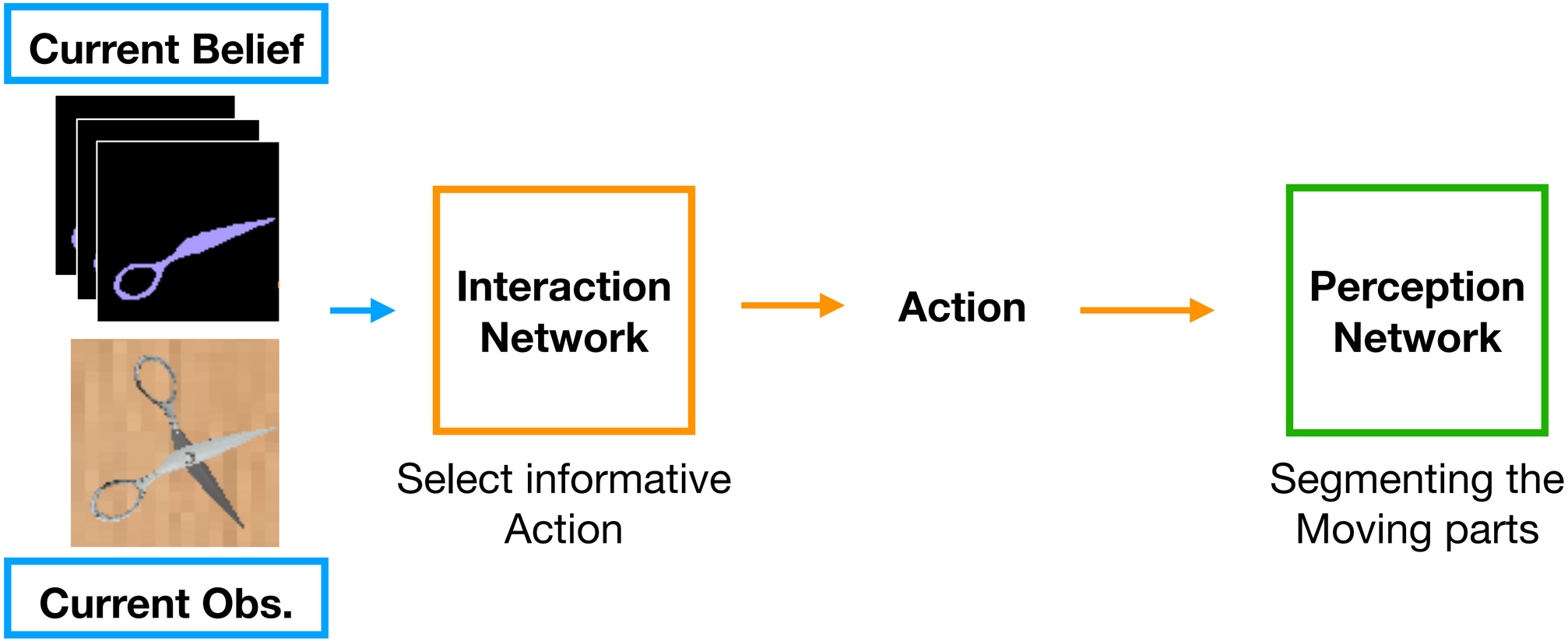
Act the Part

Key Idea: couple action selection and motion segmentation.



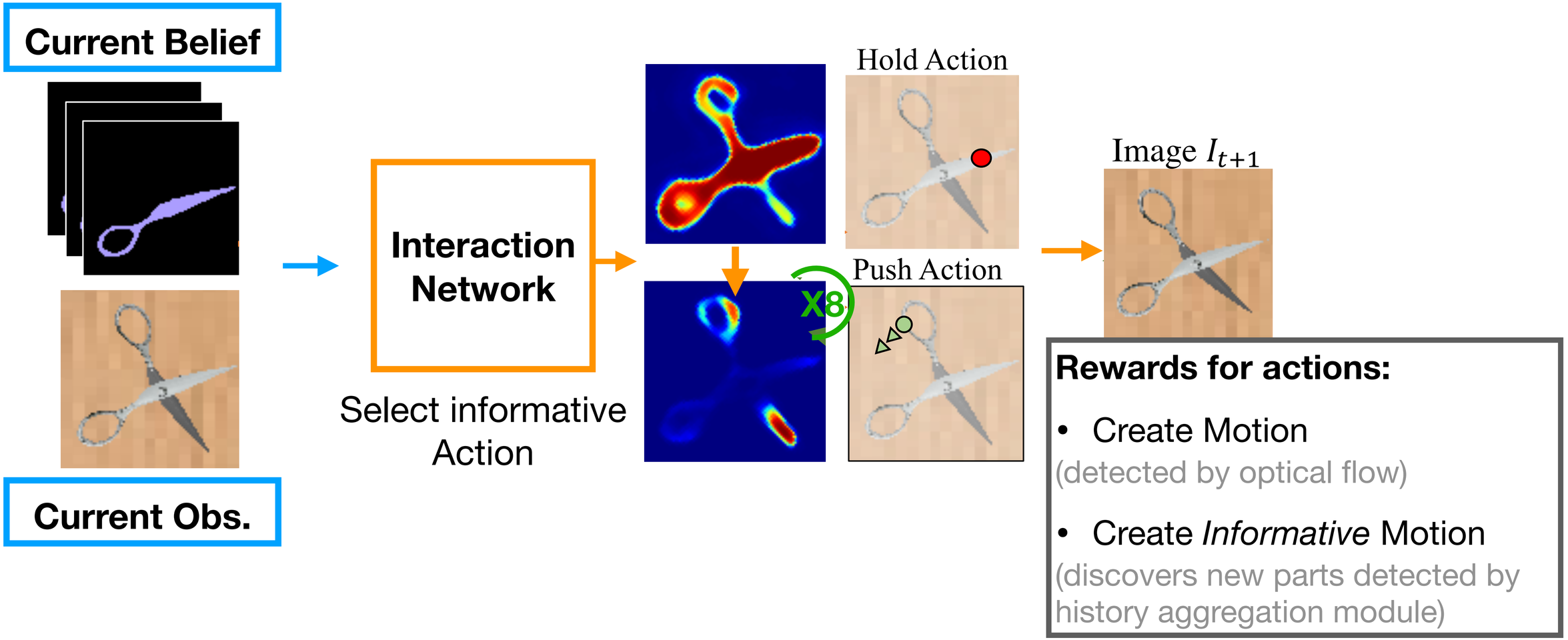
Act the Part

Key Idea: couple action selection and motion segmentation.



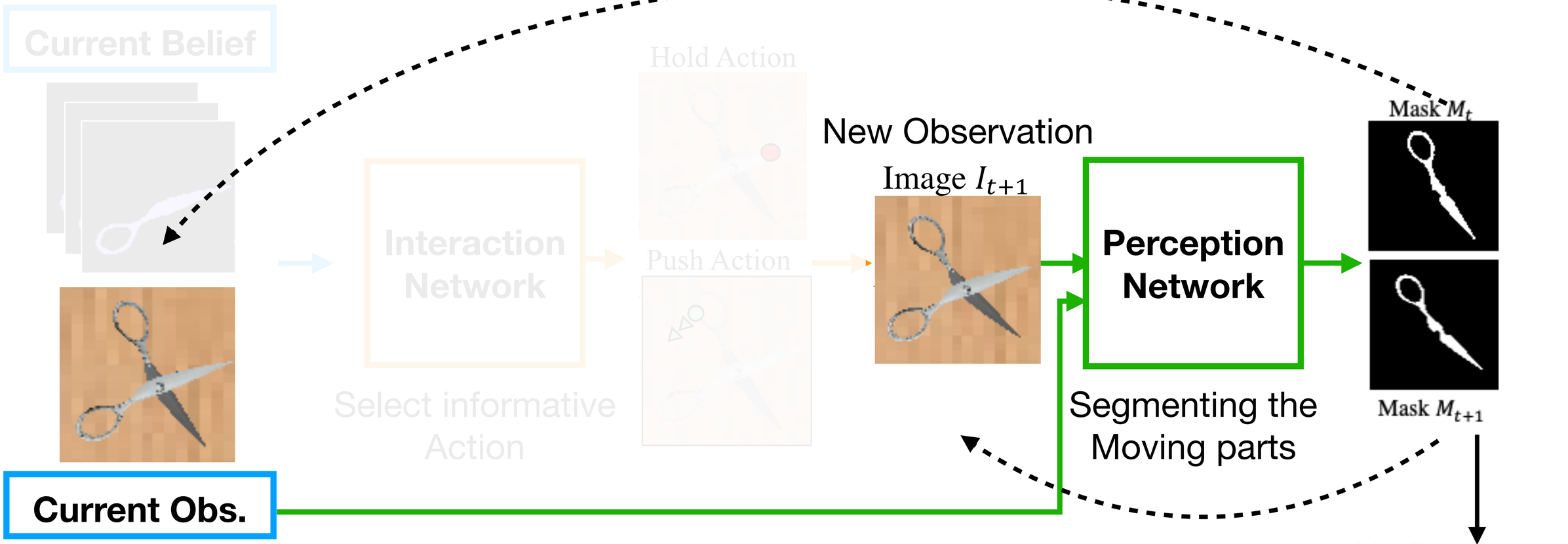
Act the Part

Key Idea: couple action selection and motion segmentation.



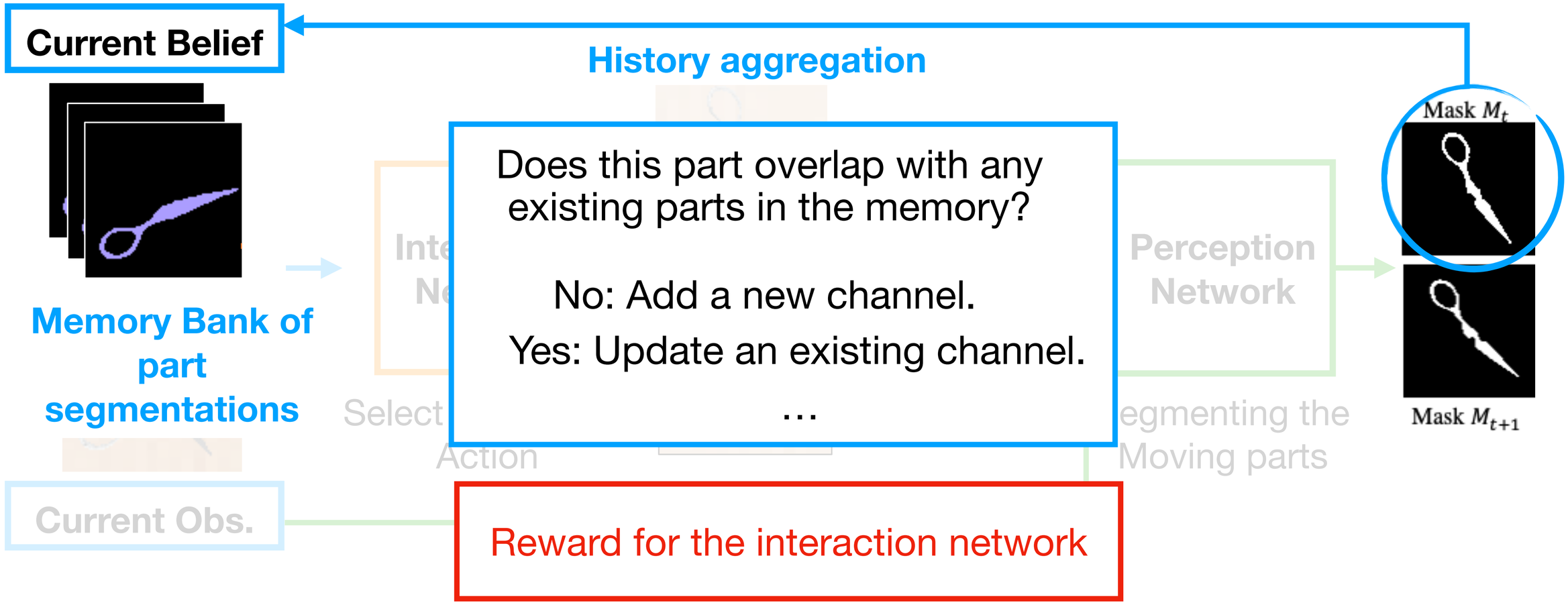
Act the Part

Key Idea: couple action selection and motion segmentation.



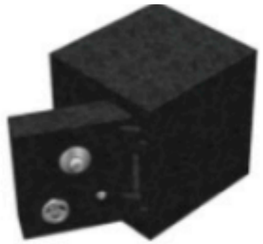
Act the Part

Key Idea: couple action selection and motion segmentation.



Training Testing Objects

Training



Safe

Scissors

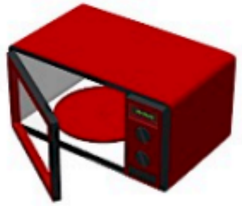
USB

Knife

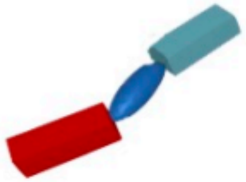
Testing Categories

Simulation

Lighter



Microwave



Multilink

Eyeglasses

Pliers

Realword

Earbuds

Eyeglasses

Tea Bag

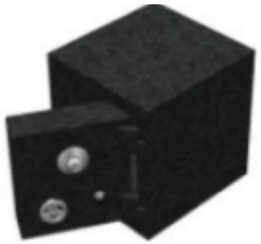
Keys
(Two Link)

Keys
(Three Link)

<u>One</u> Model for <u>All</u> Object Categories

Training Testing Objects

Training


Safe


Scissors

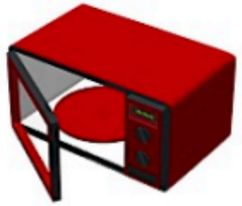

USB


Knife

Testing Categories

Simulation


Lighter


Microwave


Eyeglasses


Pliers


Multilink

Realword


Earbuds


Eyeglasses


Tea Bag

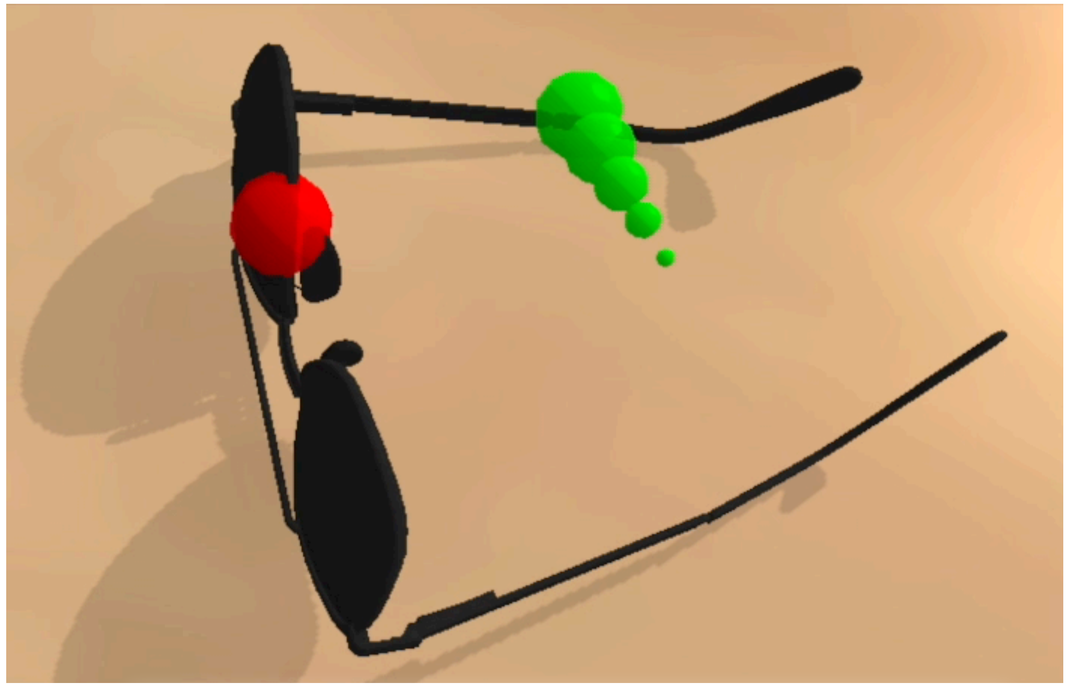

Keys
(Two Link)


Keys
(Three Link)

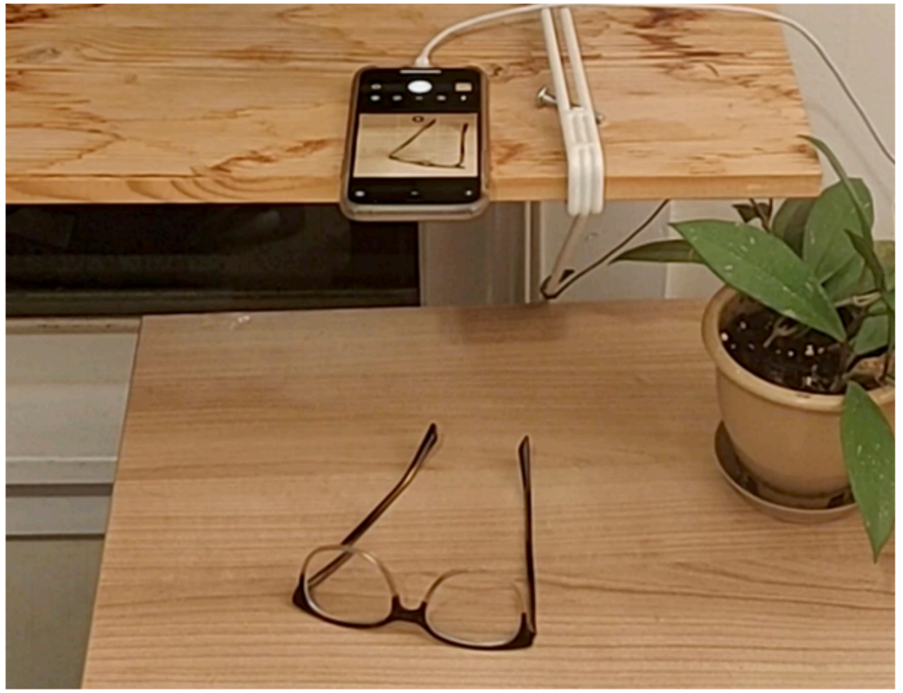
Two links

Two or three links with different kinematic structure

Experiment Results

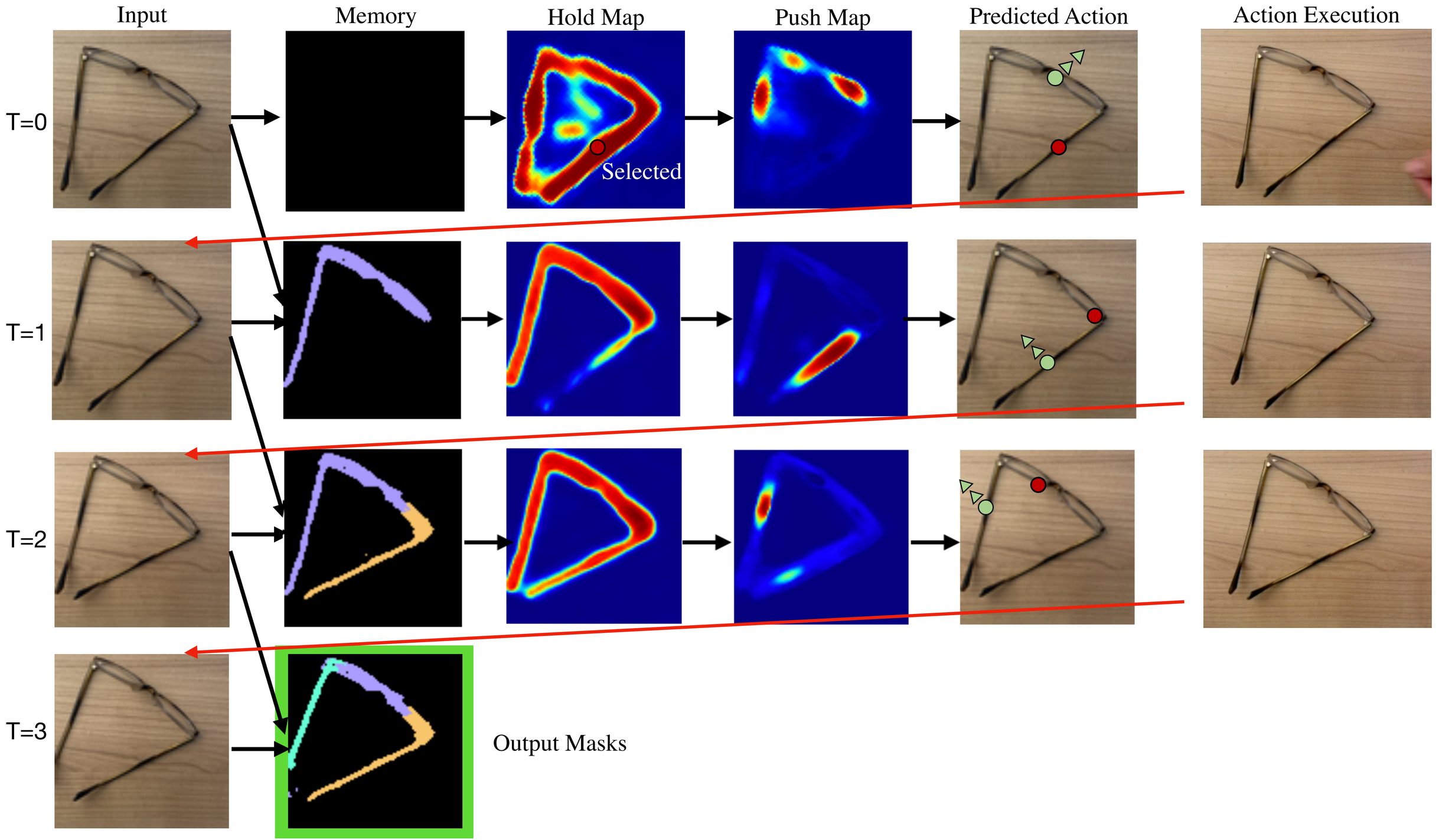


Train in Simulation

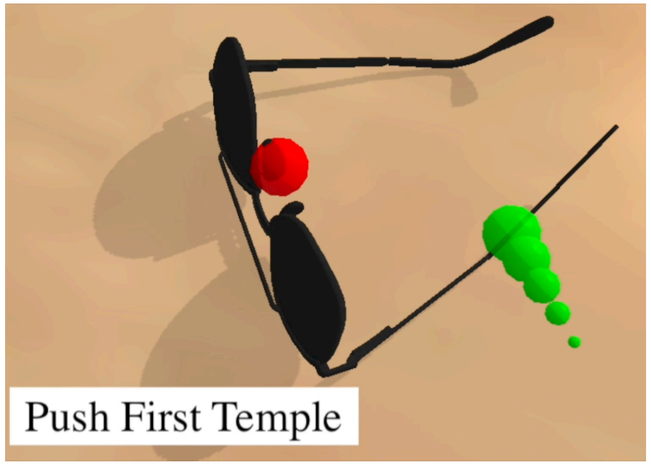


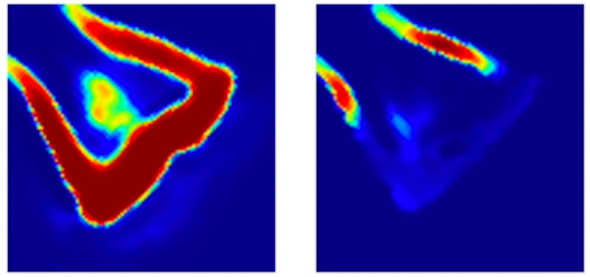
Test on Real World Objects*

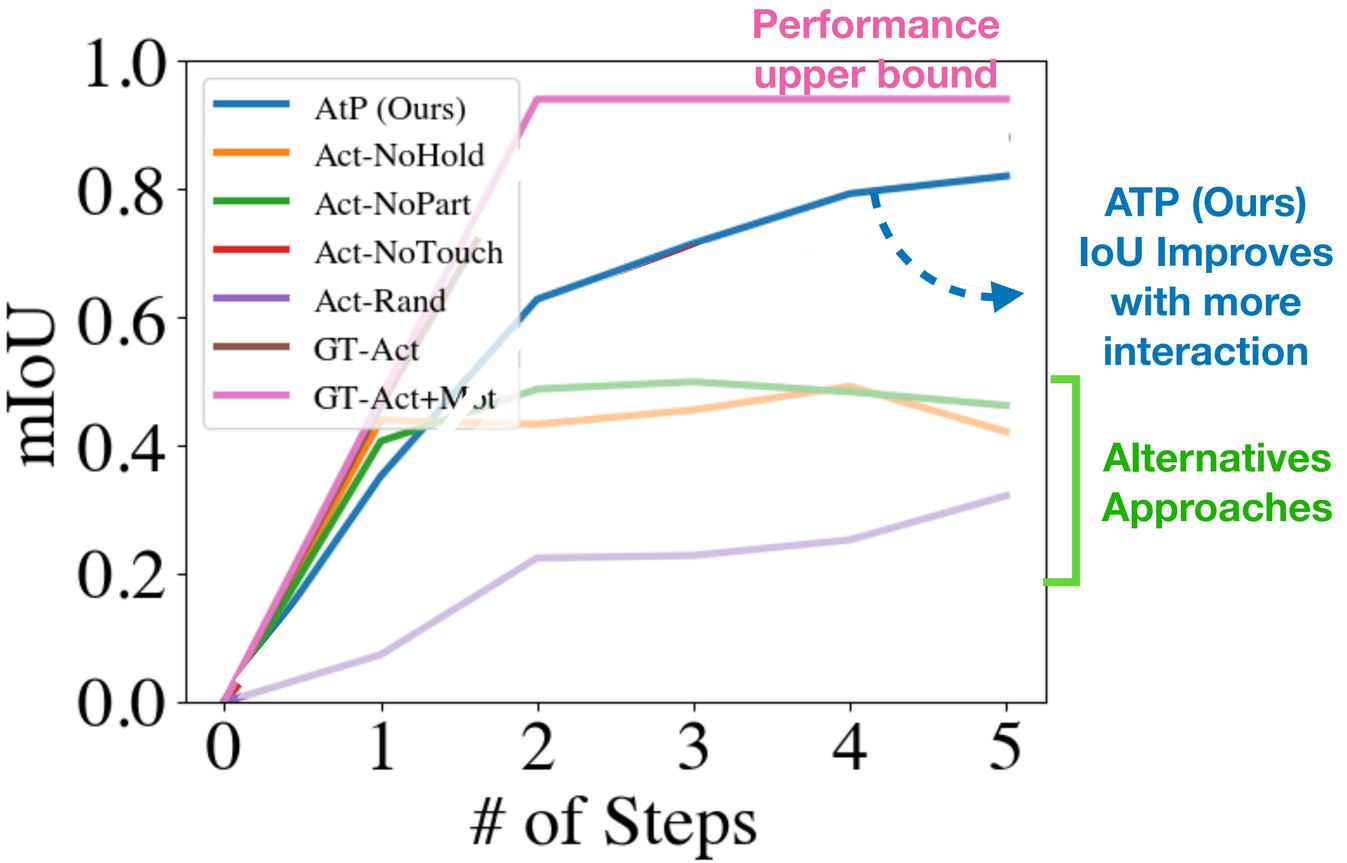
** working from home setup :)*


Input
Memory
Hold Map
Push Map
Predicted Action
Action Execution
T=0
T=1
T=2
T=3
Selected
Output Masks

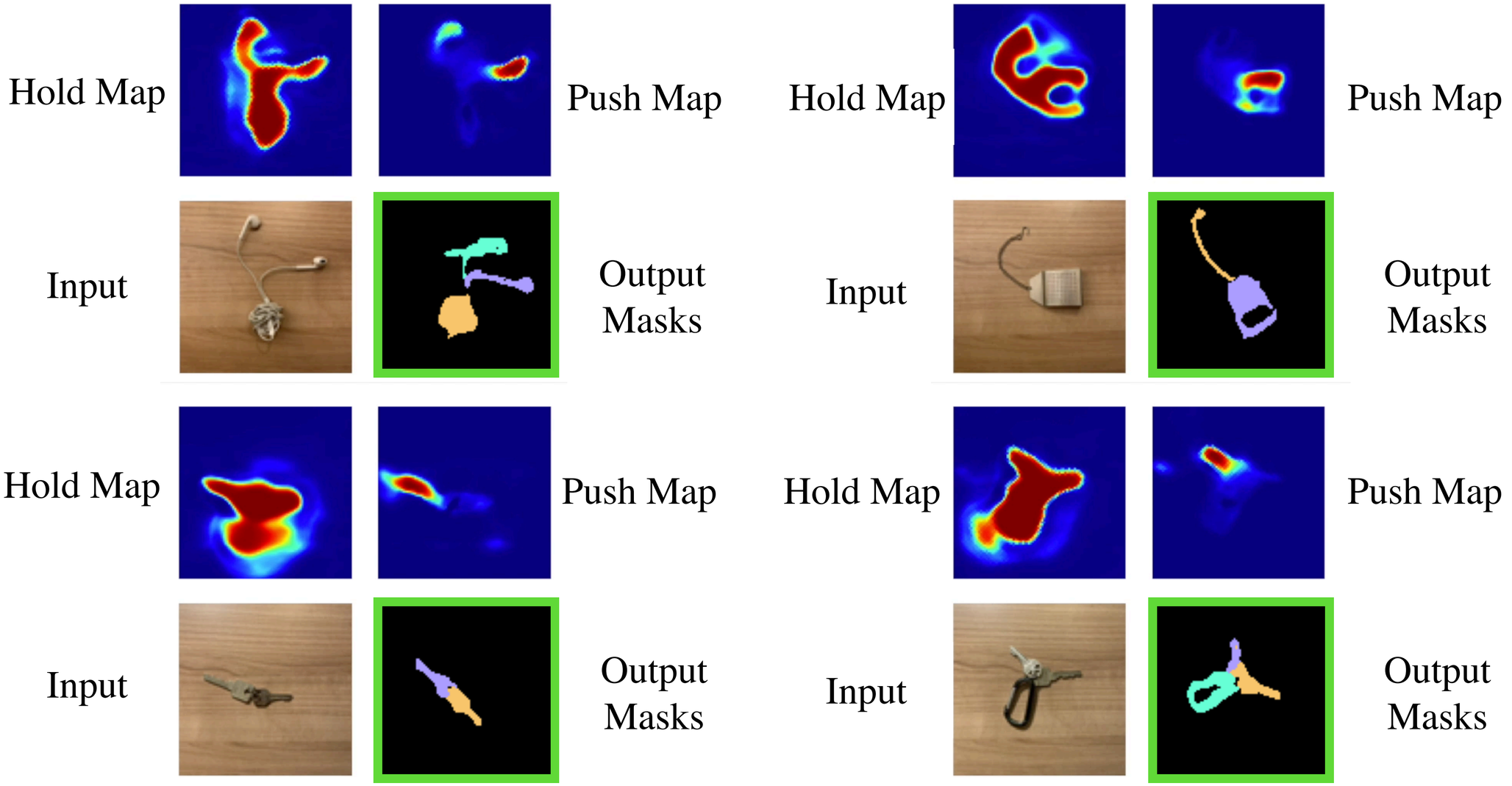
Qualitative Results in Simulation







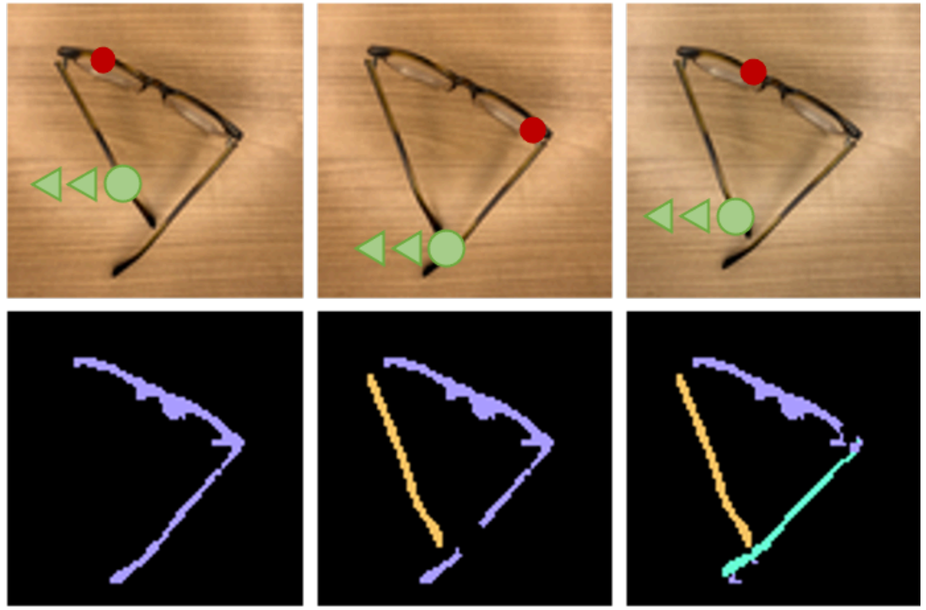
Results: Real World Generalization



Webpage: **https://atp.cs.columbia.edu/** with online Demo!

Structure from Action

Act the Part

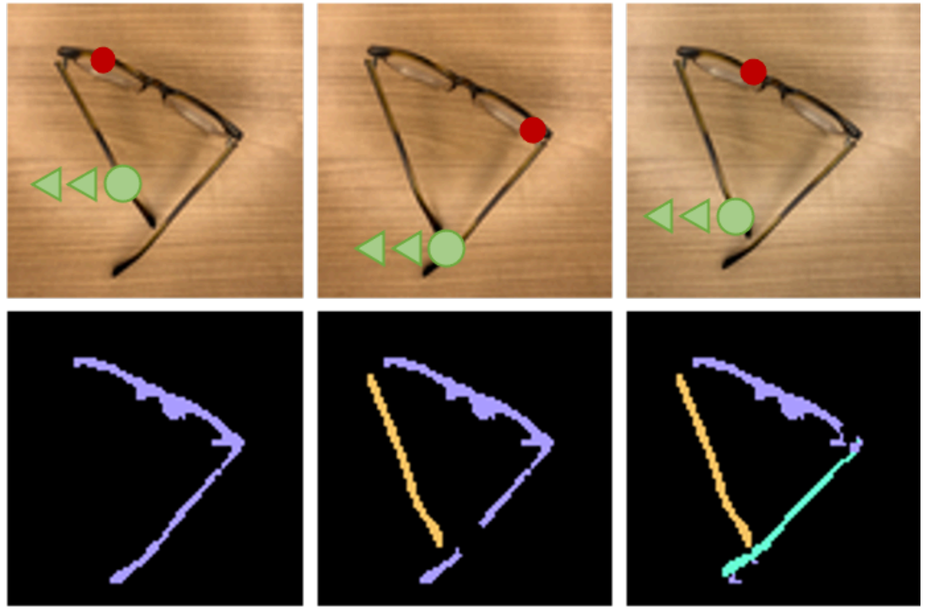


Summary:

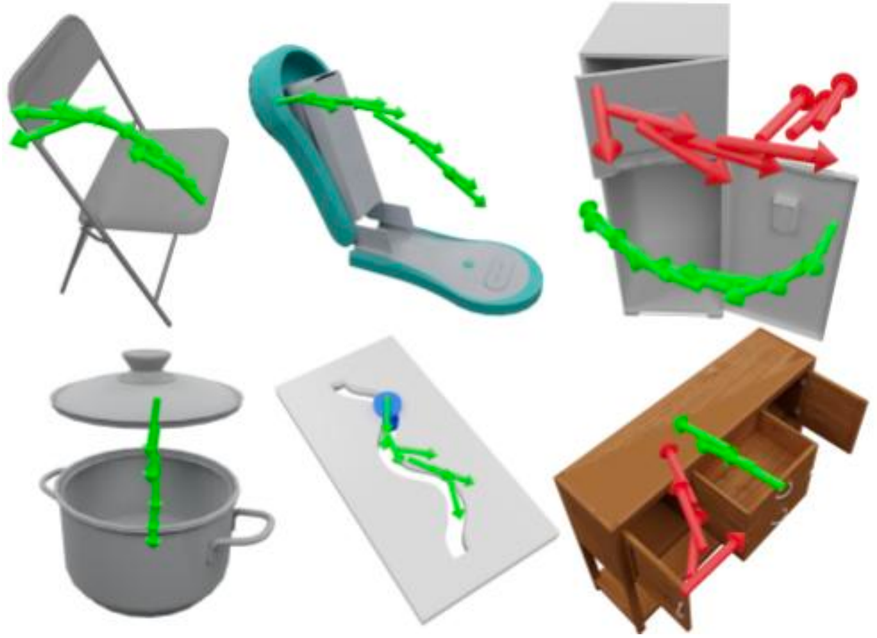
- AtP is able to learn effective interaction strategies isolating and discovering parts
- It is able to generalize to novel categories of objects with unknown and unseen number of links

Structure from Action

Act the Part



Simple 2D Action
with discrete action direction



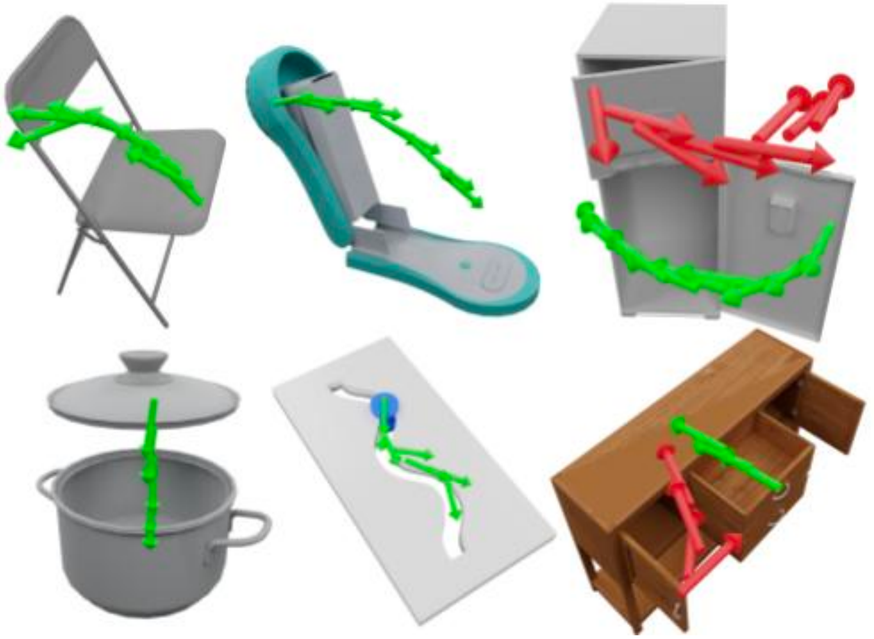
x Closed-loop, 3D Actions

x Goal-conditioned Manipulation Tasks

Structure from Action

Upgrade the Interaction Policy

- General Action Representation (Continuous Action in SE(3))
- Closed-loop Action Sequence
- Goal-Conditioned Manipulation Tasks

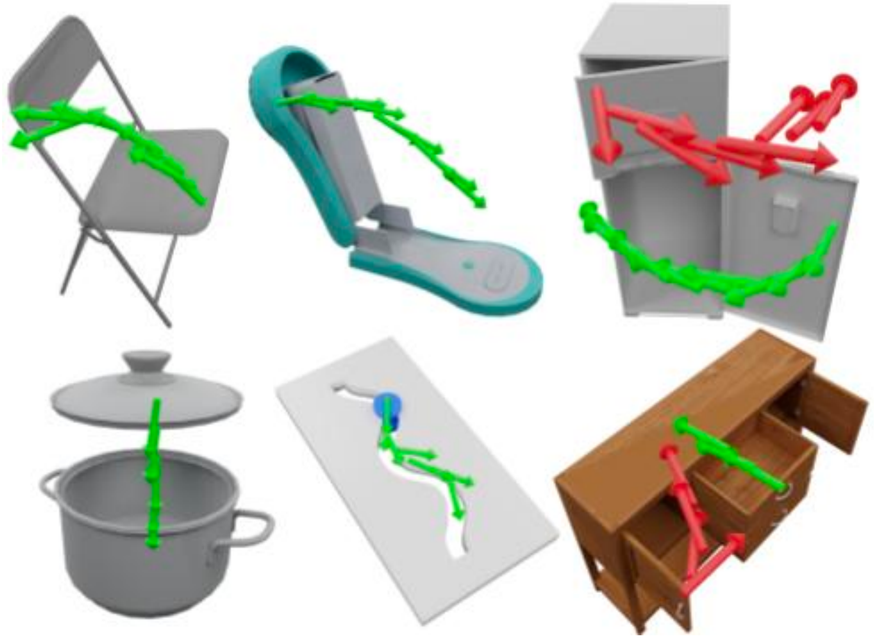


Improve the Understanding of the Object Structure

Structure from Action

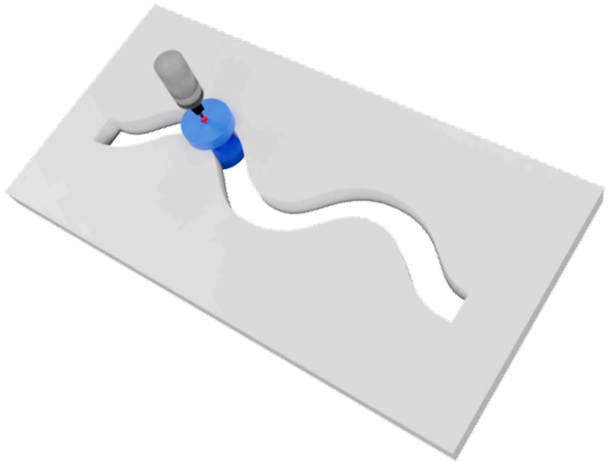
UMPNet: Universal Manipulation Policy Network for Articulated Objects

Zhenjia Xu, Zhanpeng He, Shuran Song



Universal Manipulation Policy





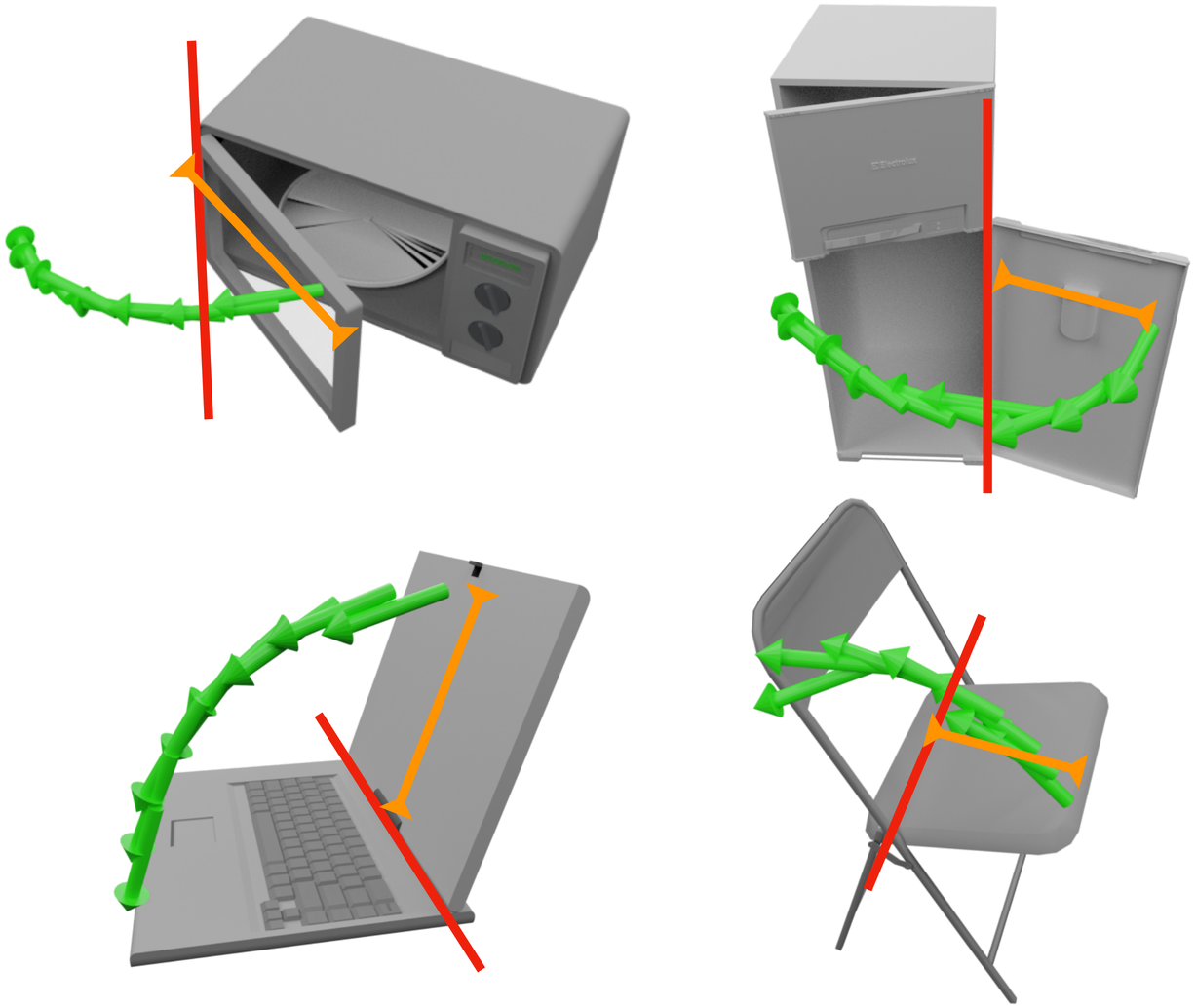
Action trajectories may vary drastically due to objects kinematic structures and geometry.



Goal: learn a **single** manipulation policy to handle **all** these objects **from visual observations.**

Universal Manipulation Policy

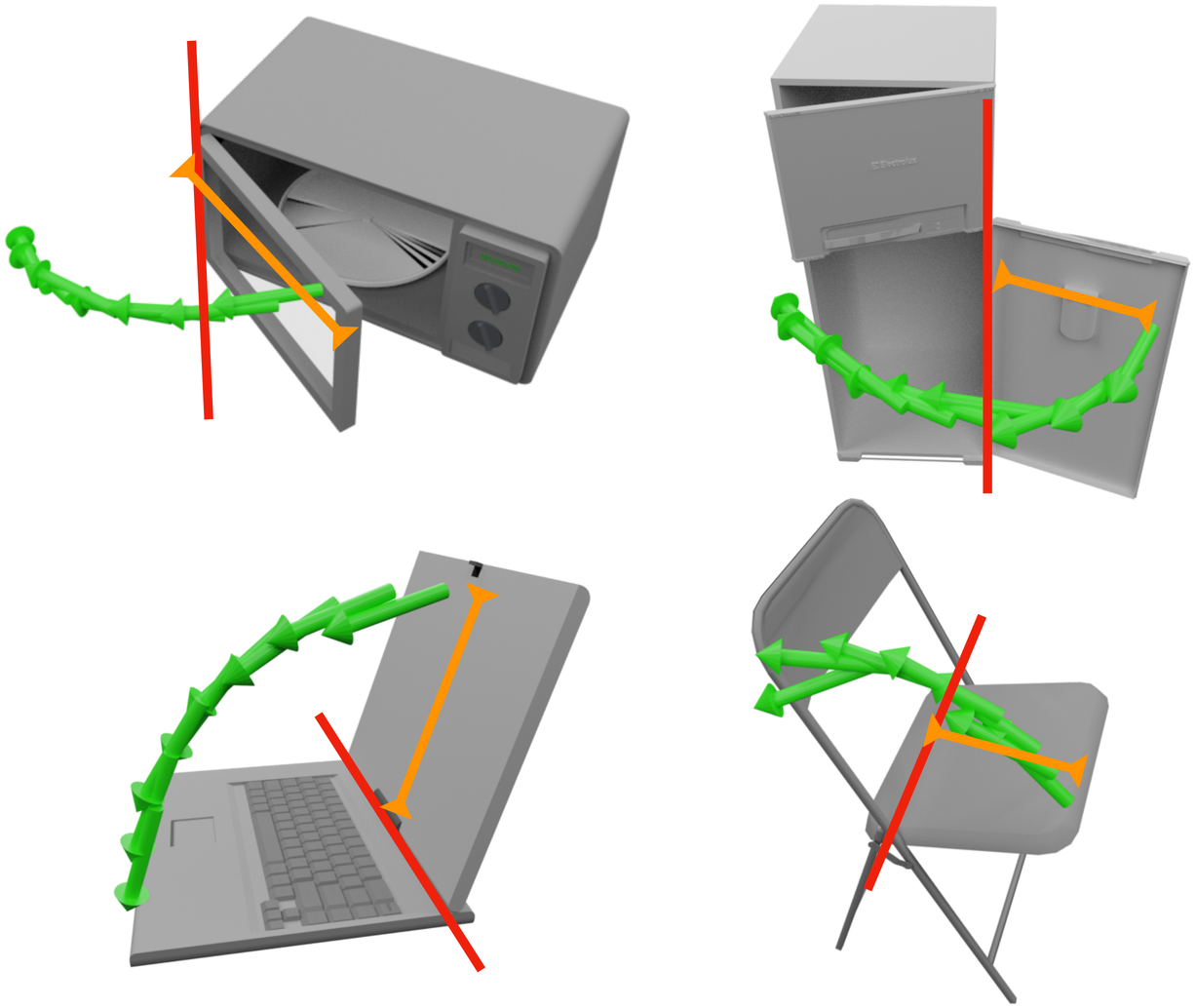
Why it is Possible ??



Can be summarized by a similar high-level function conditioned on the objects' underlying kinematic structure.

Universal Manipulation Policy

Why it is Possible ??



Learning to interact with a diverse set of articulated objects

Aquire Generalizable Knowledge on:

- Articulation structure
- How these structures would react to different actions.

Beyond a specific object instance or category

What is an effective interaction policy?

- Where to interact

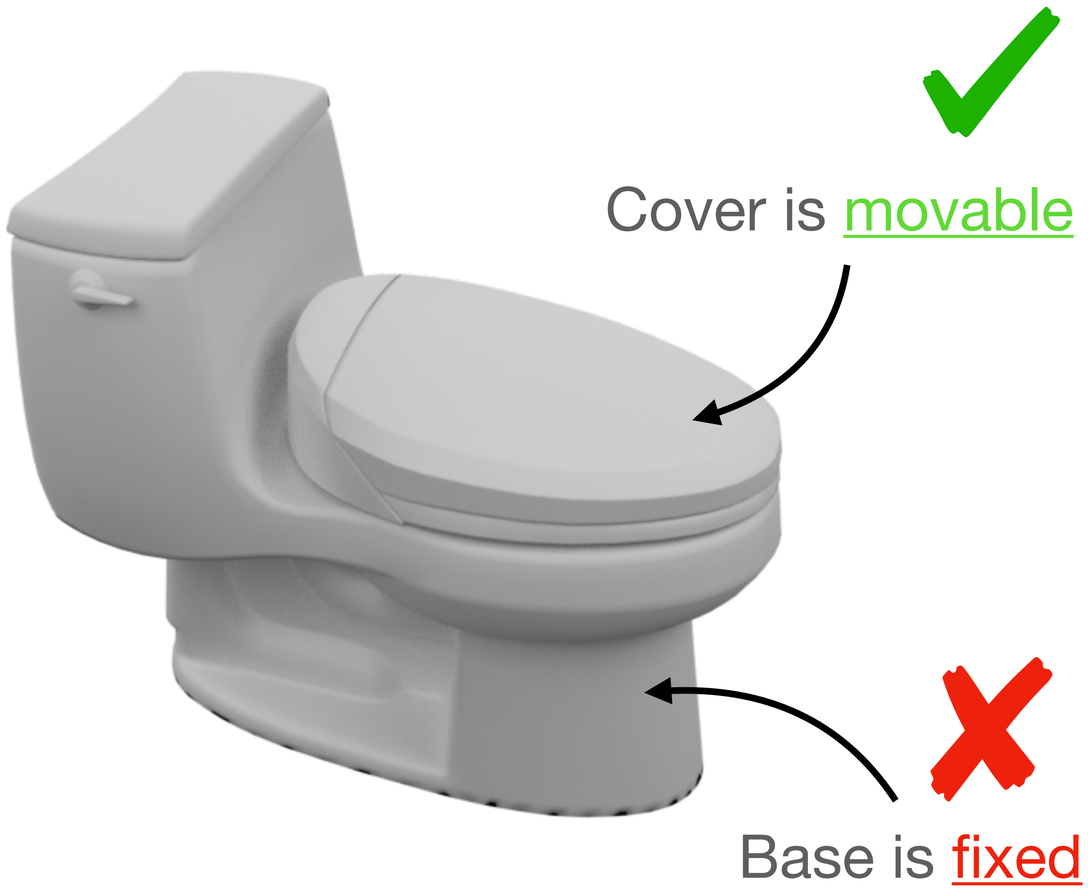
  Interacting with the movable cover instead of the base

- Effective action direction

  Pulling up instead of pushing down

- Arrow-of-Time awareness

  Keeping pulling up the cover to visit novel states instead of moving up-and-down



What does the policy need to learn?

- Where to interact

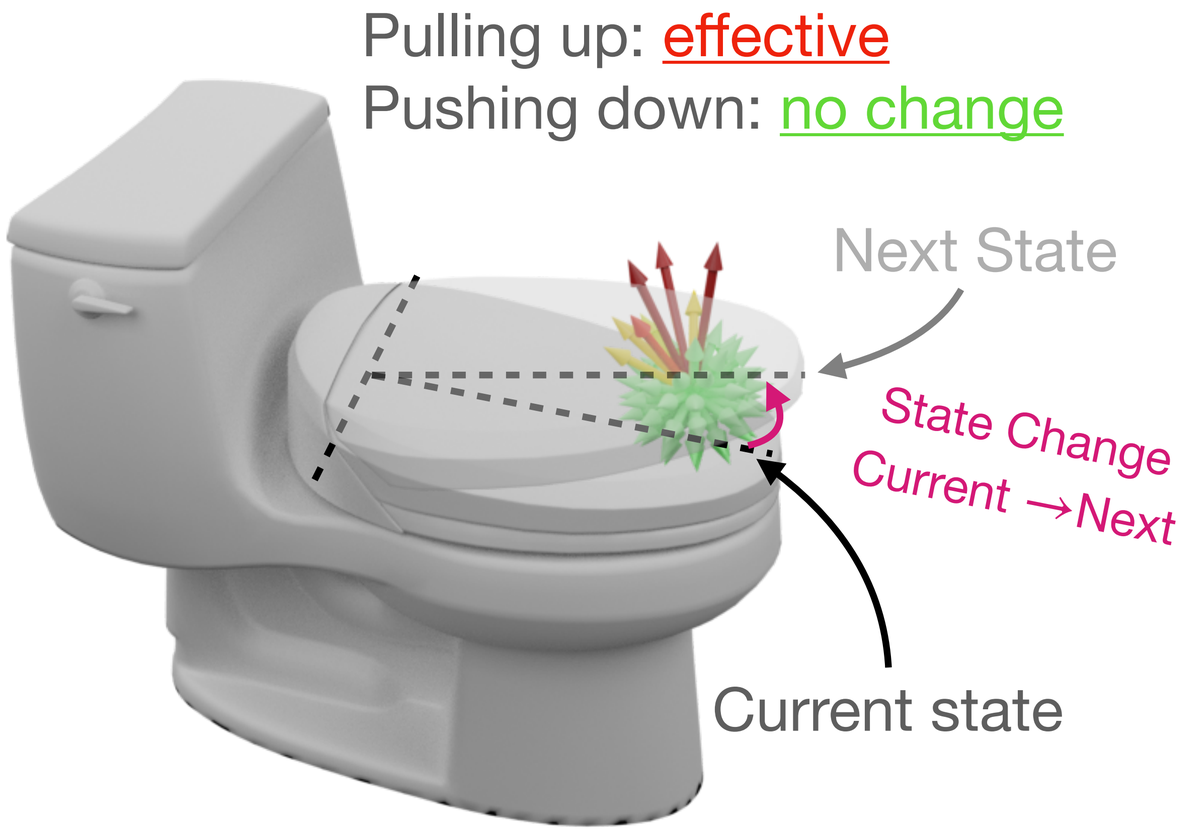
  Interacting with the movable cover instead of the base

- Action direction

  Pulling up instead of pushing down

- Arrow-of-Time awareness

  Keeping pulling up the cover to visit novel states instead of moving up-and-down



What does the policy need to learn?

- Where to interact

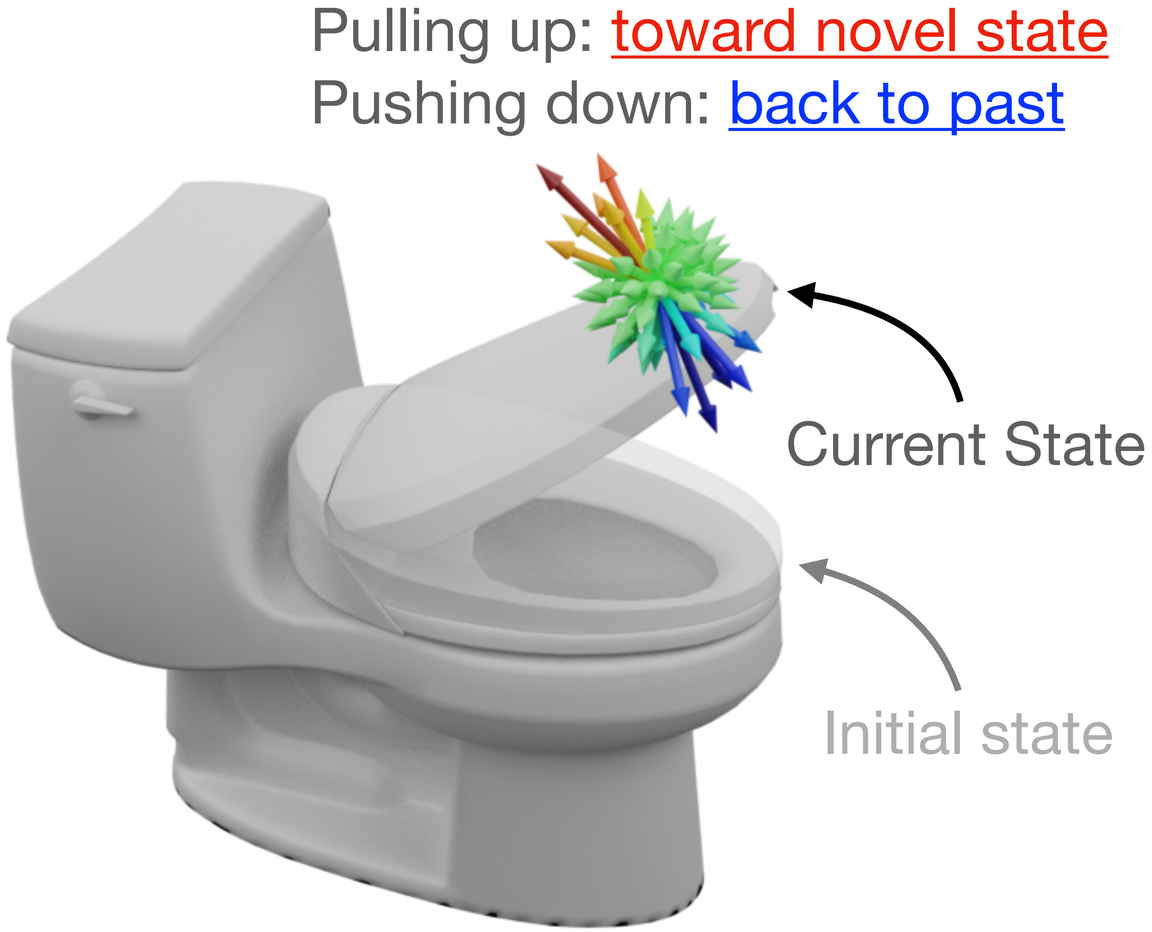
  Interacting with the movable cover instead of the base

- Effective action direction

  Pulling up instead of pushing down

- Consistent action direction

  Keeping pulling up the cover to visit novel states instead of moving up-and-down. (Arrow-of-Time awareness)



What does the policy need to learn?

- Where to interact

  Interacting with the movable cover instead of the base

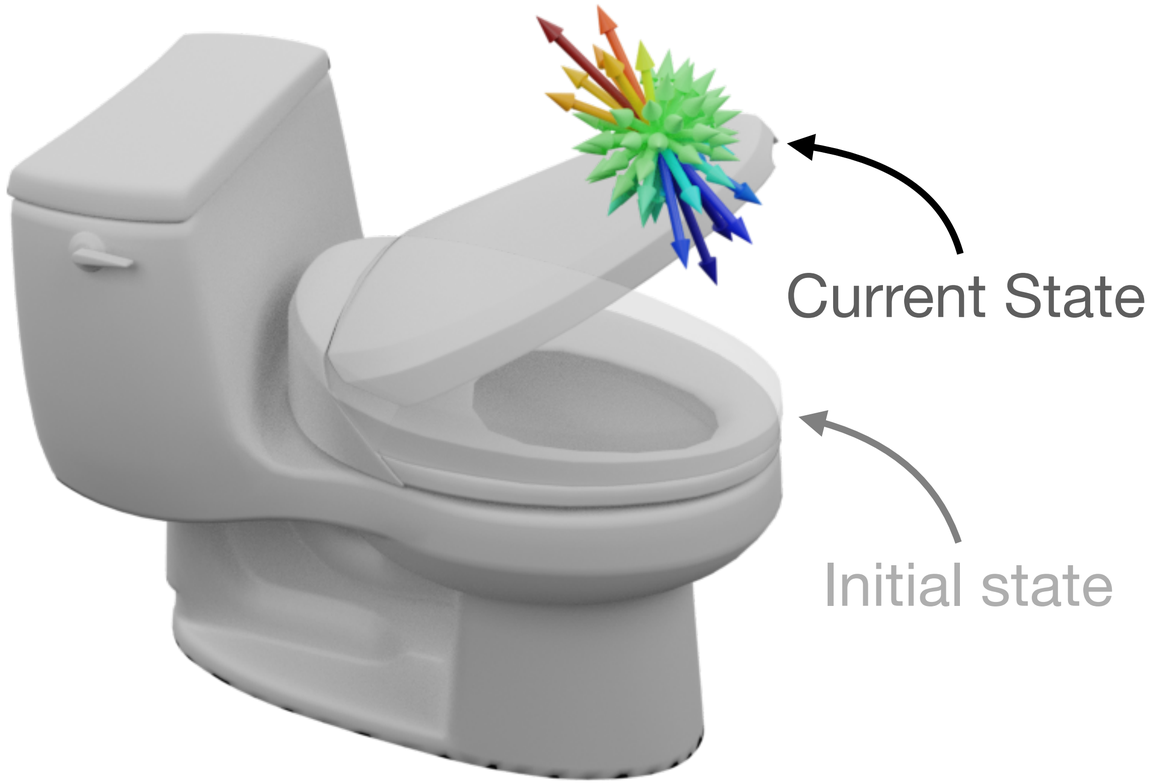
- Effective action direction

  Pulling up instead of pushing down

- Consistent action direction

  Keeping pulling up the cover to visit novel states instead of moving up-and-down. (Arrow-of-Time awareness)

Pulling up: toward novel state
Pushing down: back to past

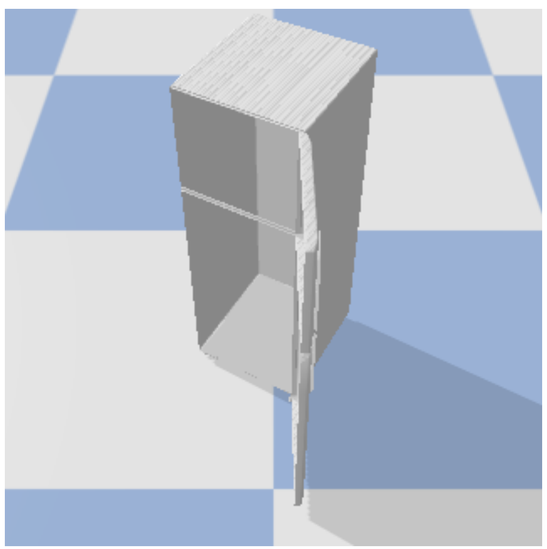


Goal conditioned manipulation, without goal-conditioned training

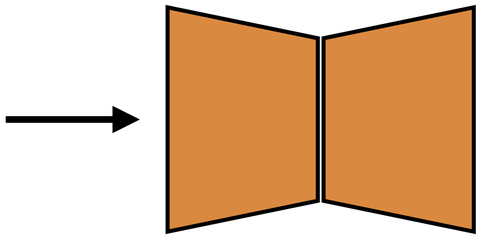
Universal Manipulation Policy Network

Universal Manipulation Policy Network

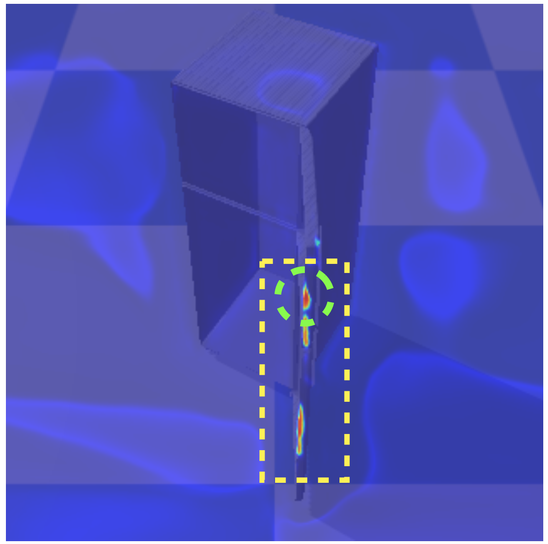
Interaction Position Inference



Visual Observation

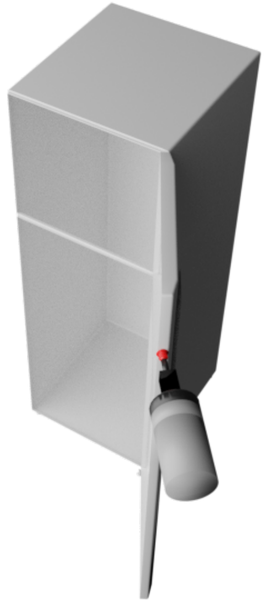


Position Net



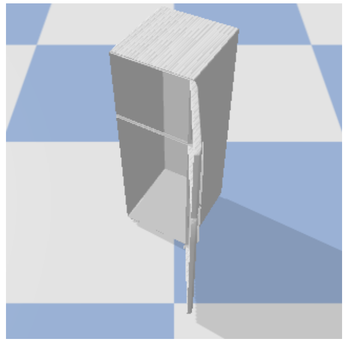
Position Affordance

Max

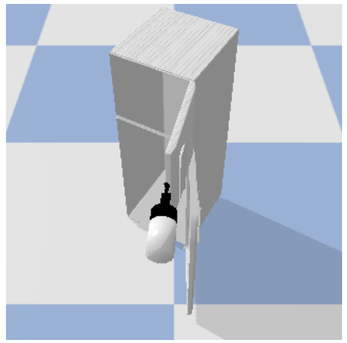


Execution

UMPNet: Direction Inference



Init



Current

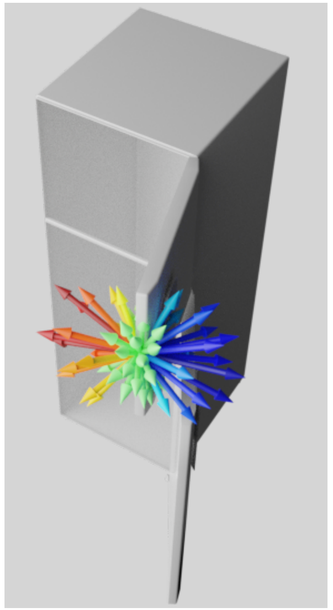
AoT Net

Dist Net

Direction Net

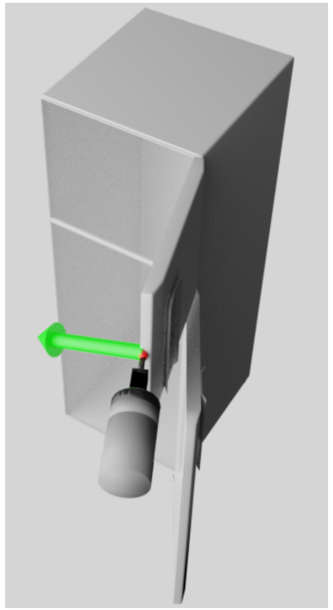
Arrow of Time
(AoT)
label
(-1/0/1)

Distance
(scalar)



Direction Score

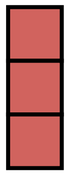
Max



Execution

Action Sampler

×N



Action

<- Continuous SE(3) Space

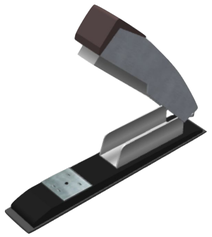
Training and Testing Objects

Training Categories (12)


Fridge


Folding Chair


Laptop


Stapler

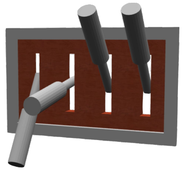

Trashcan


Microwave


Toilet


Window


Cabinet


Switch

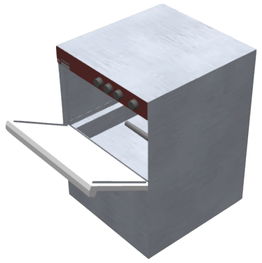

Kettle

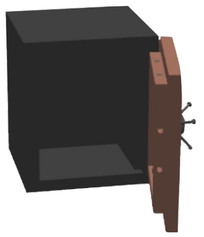

Toy

Testing Categories (10)


Box

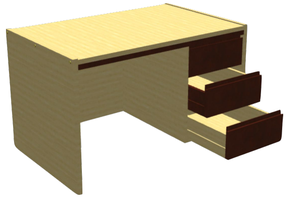

Phone


Dish Washer

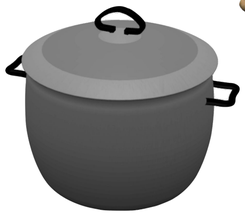

Safe


Oven


Washing Machine



Table


Kitchen Pot


Bucket

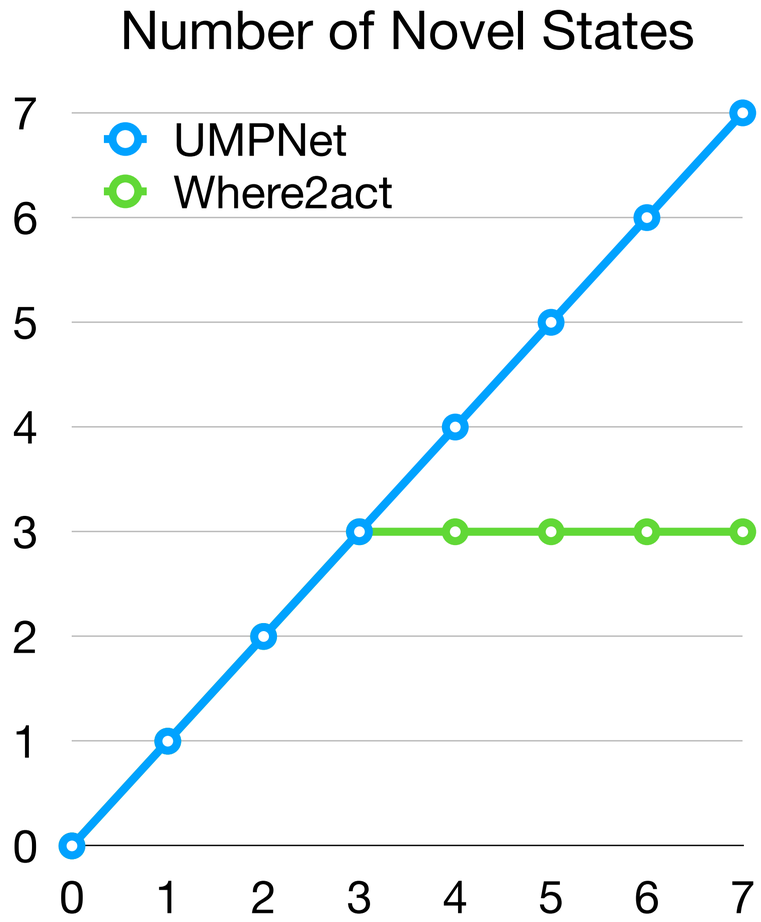

Door

The policy is trained with self-guided exploration
without any human demonstrations, scripted policy, or pre-defined goal conditions.

Open-ended State Exploration

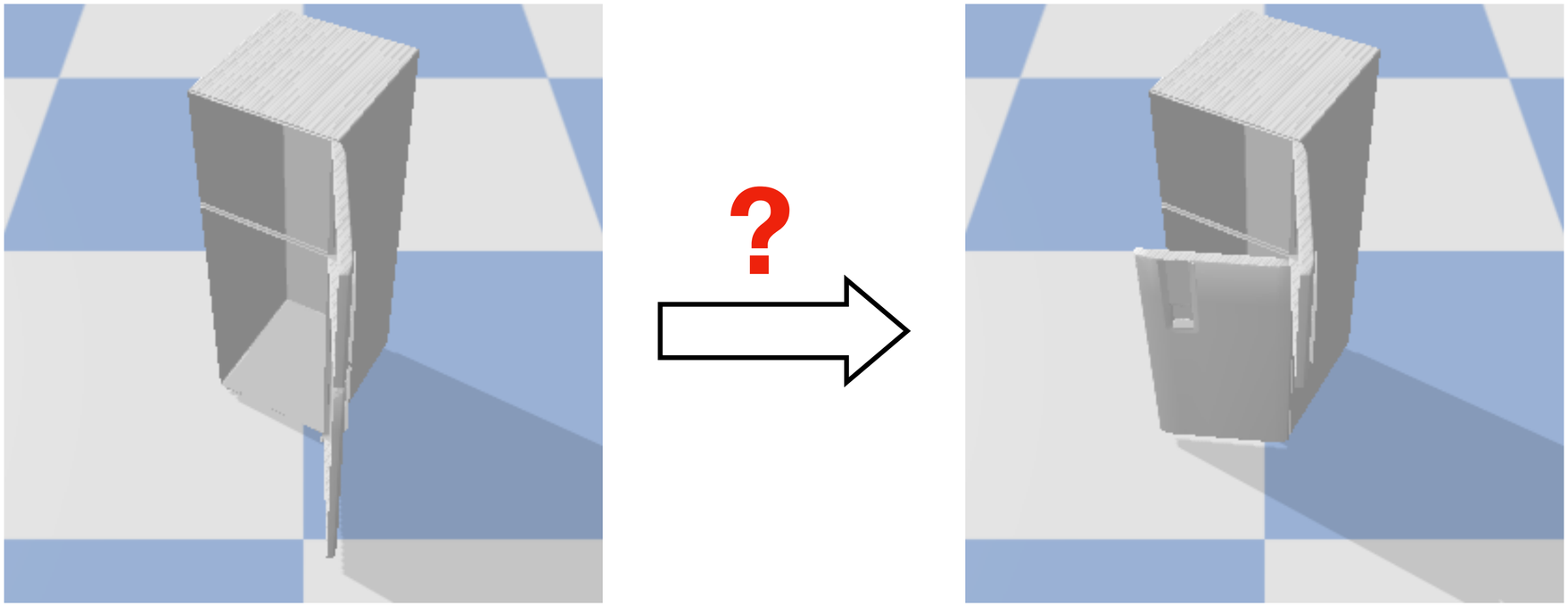
UMPNet

Where2act*
[Mo *et al.*]



*Where2act: infers a single-step interaction from current image observation

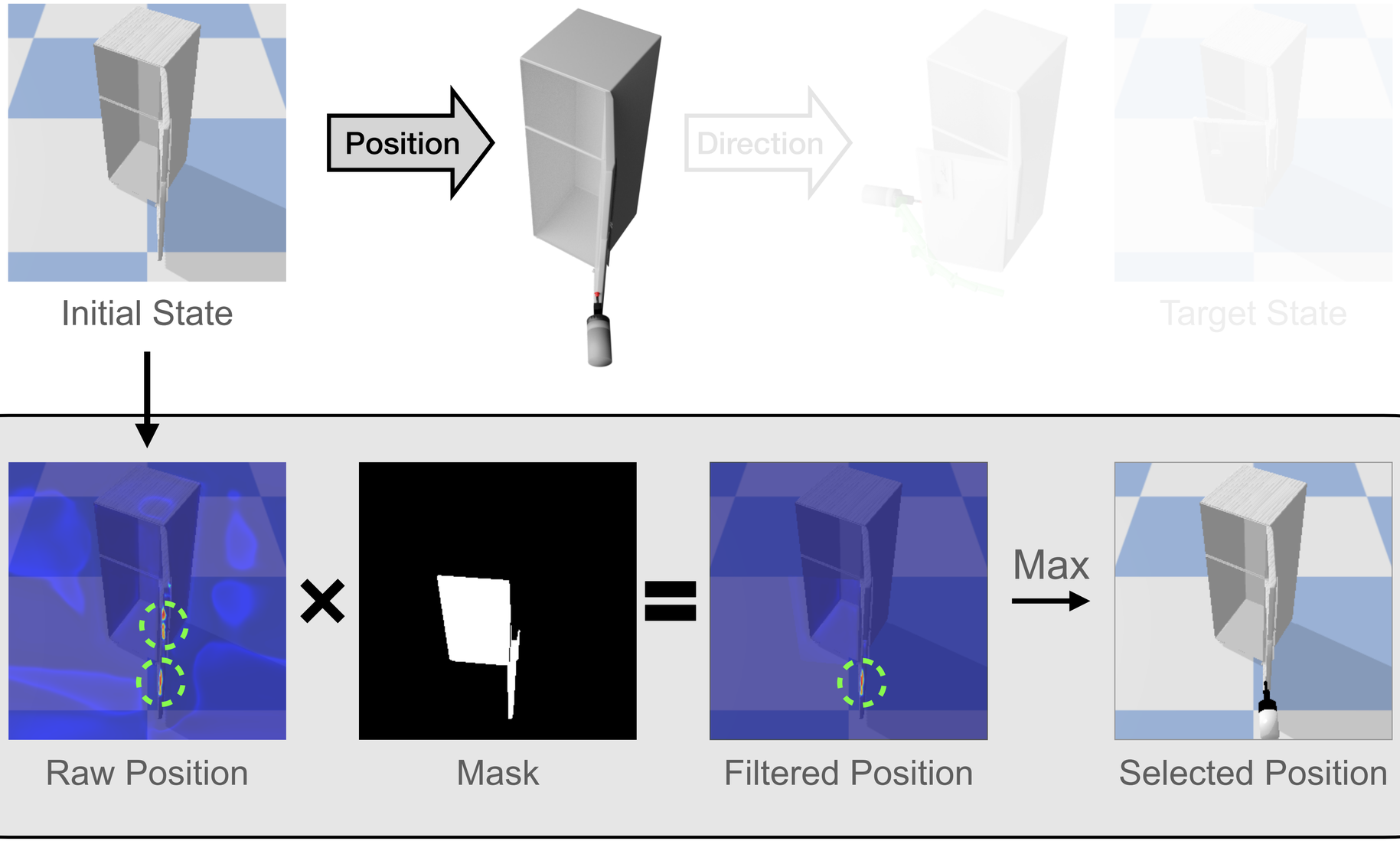
Goal Conditioned Manipulation



Initial State

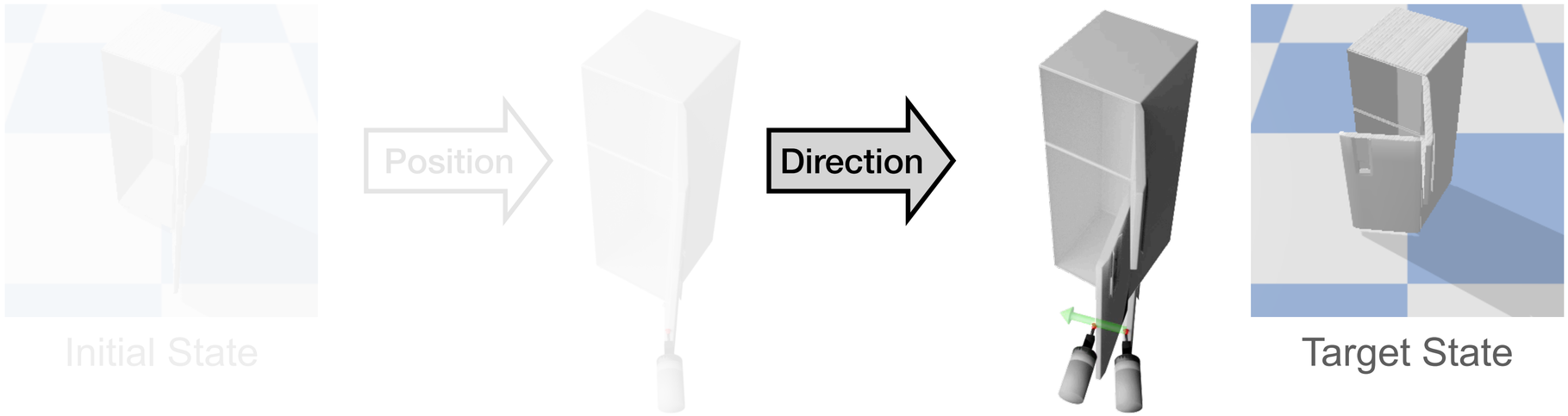
Target State

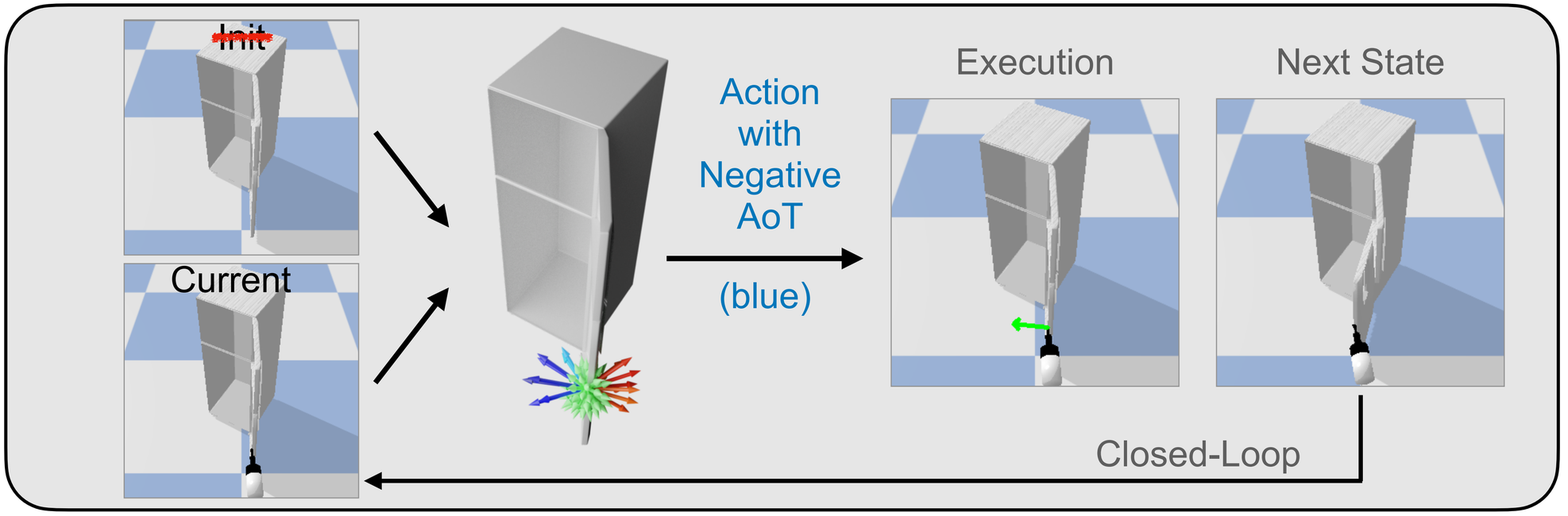
Goal Conditioned Manipulation



Position Inference

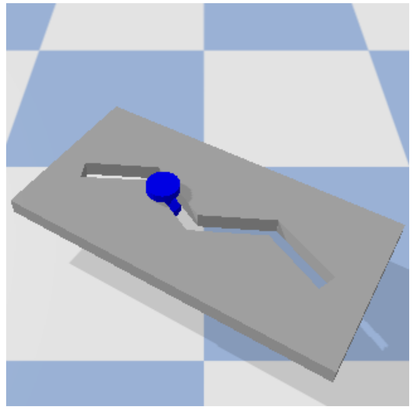
Goal Conditioned Manipulation



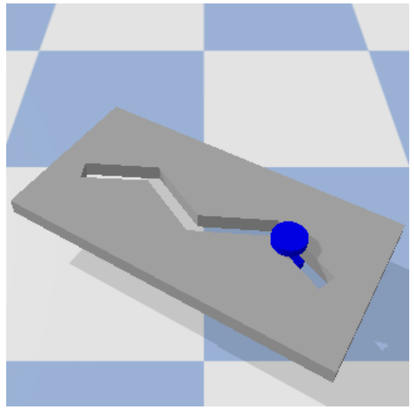


Direction Inference

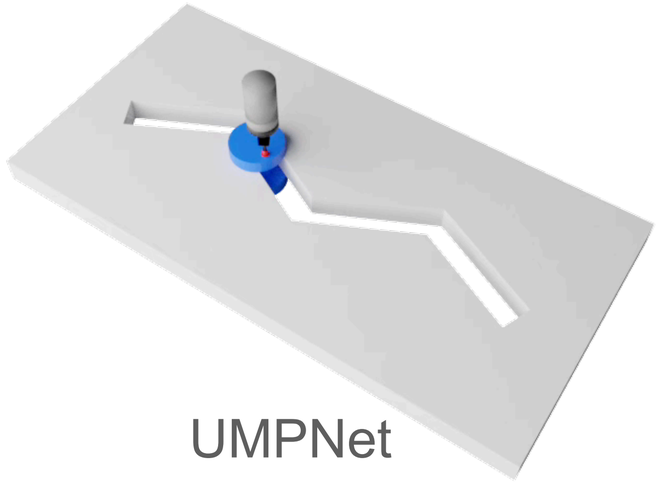
Goal Conditioned Manipulation: Results



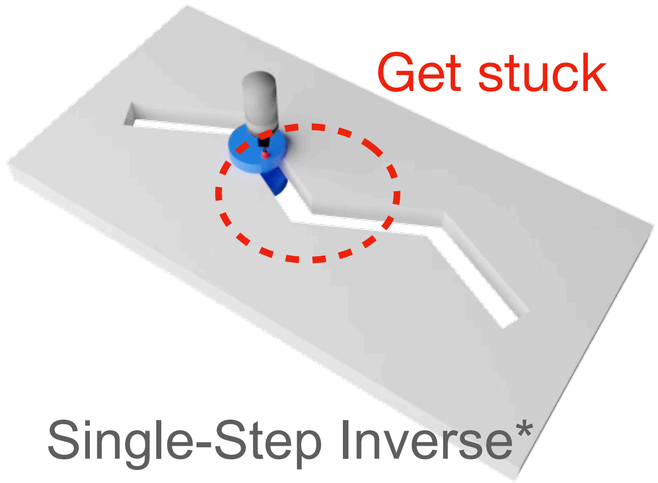
Initial State



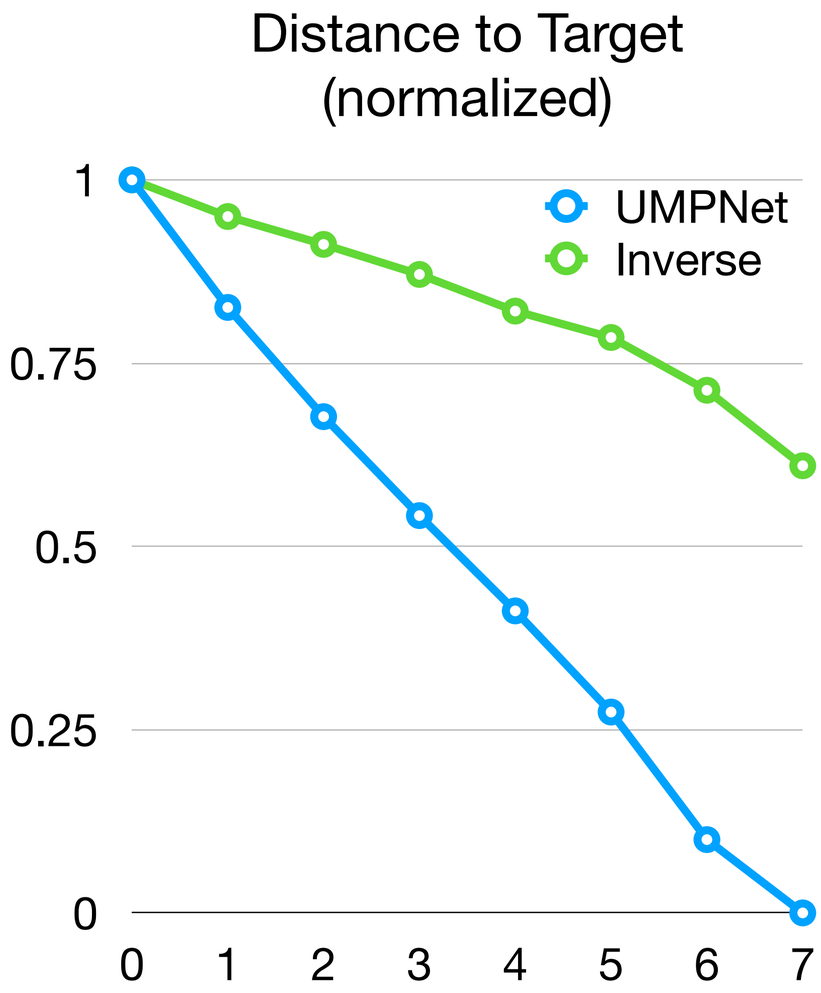
Target State



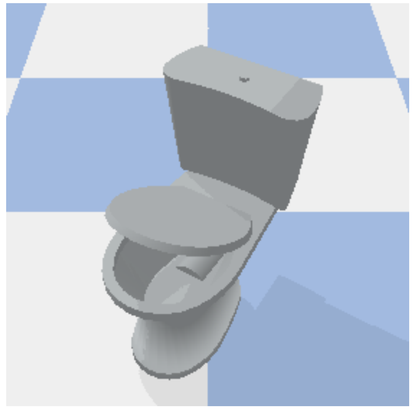
UMPNet



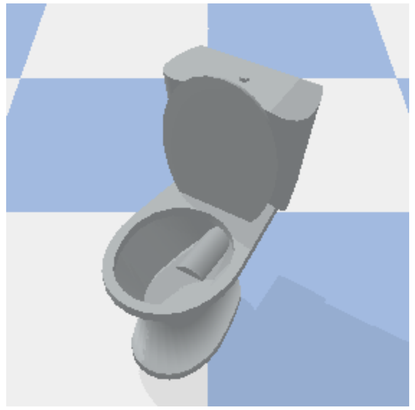
Single-Step Inverse*



Goal Conditioned Manipulation: Results



Initial State

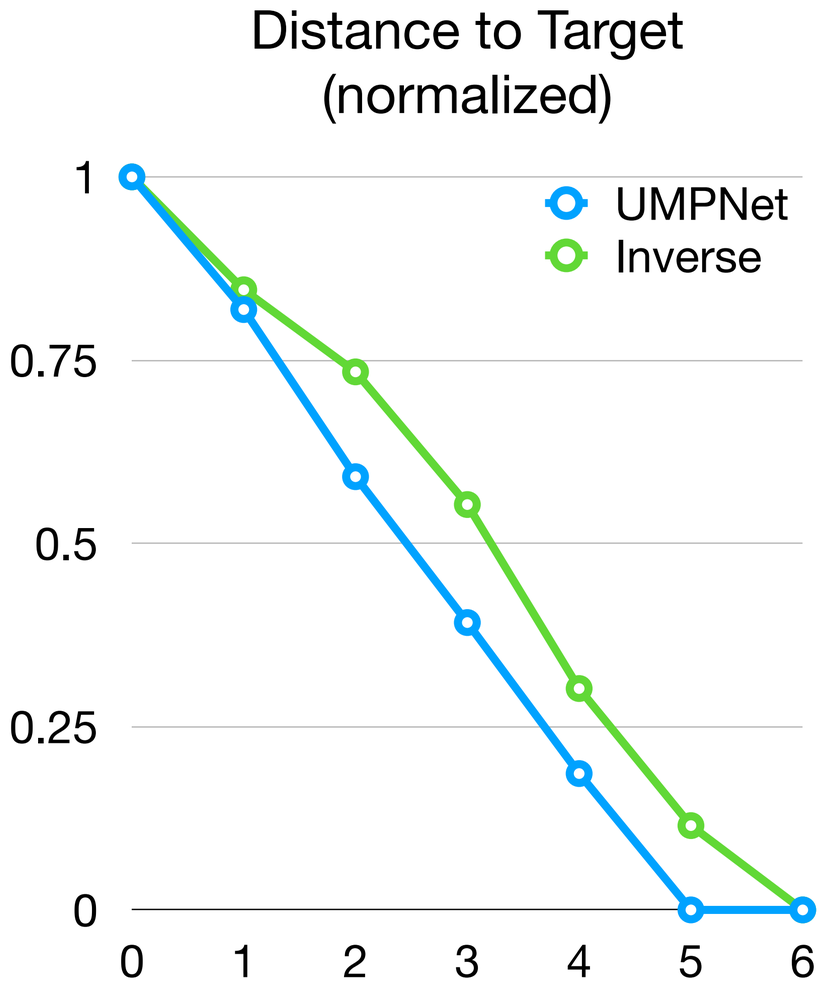


Target State

UMPNet

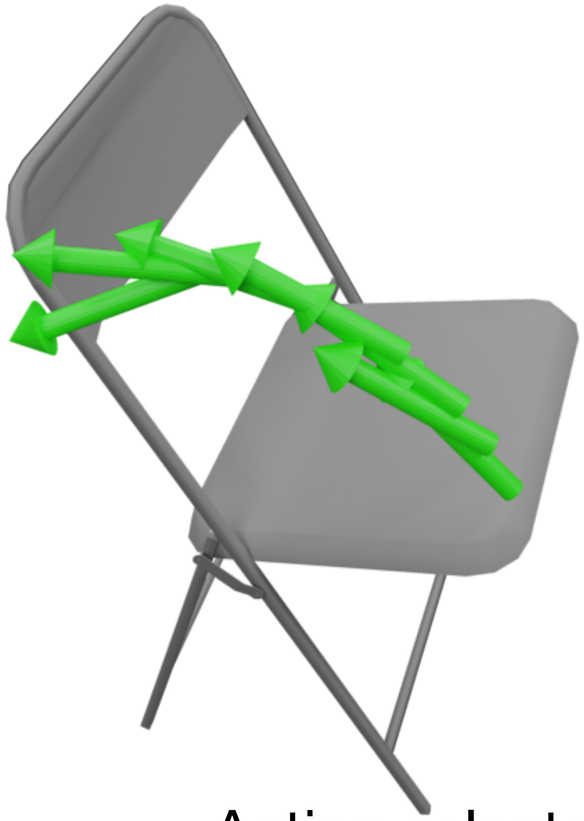
Bad direction

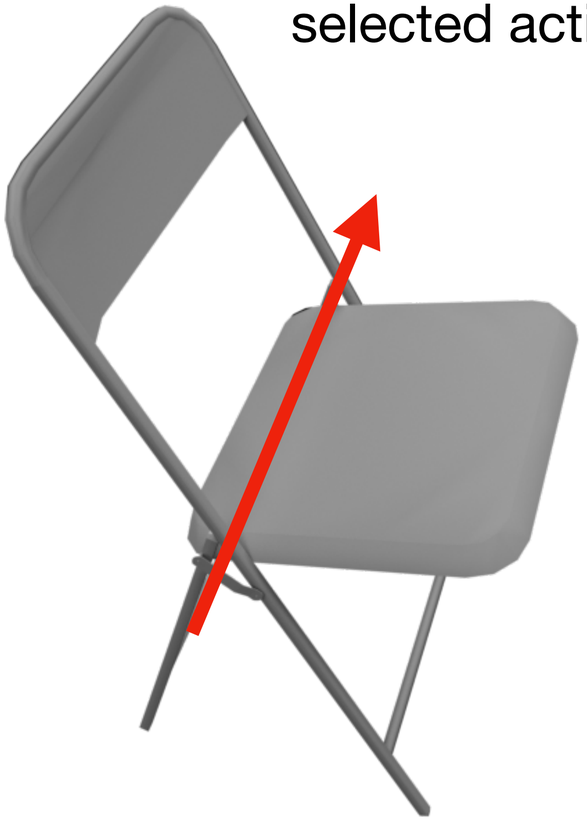
Inverse*

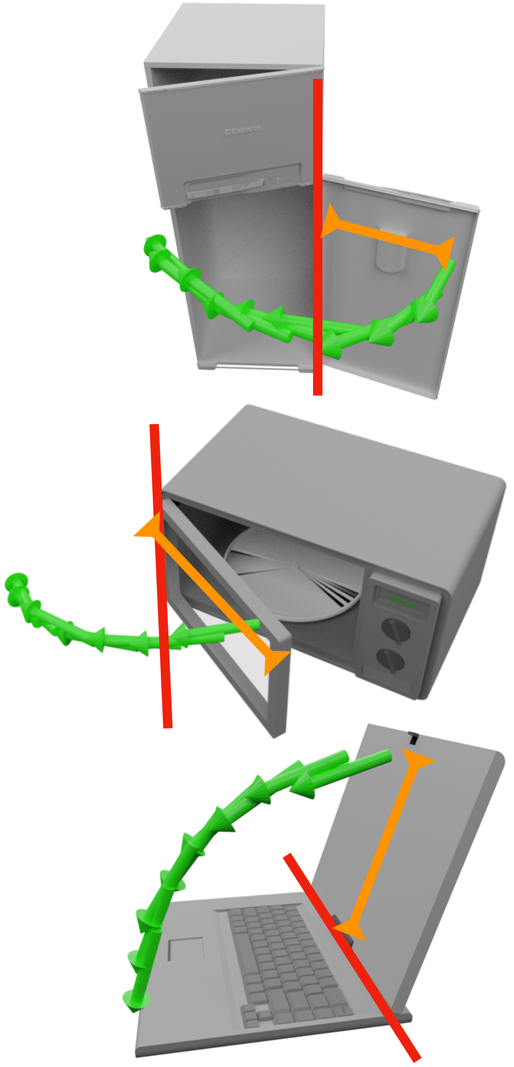


Action → Structure

Joint parameters inferred from the selected actions

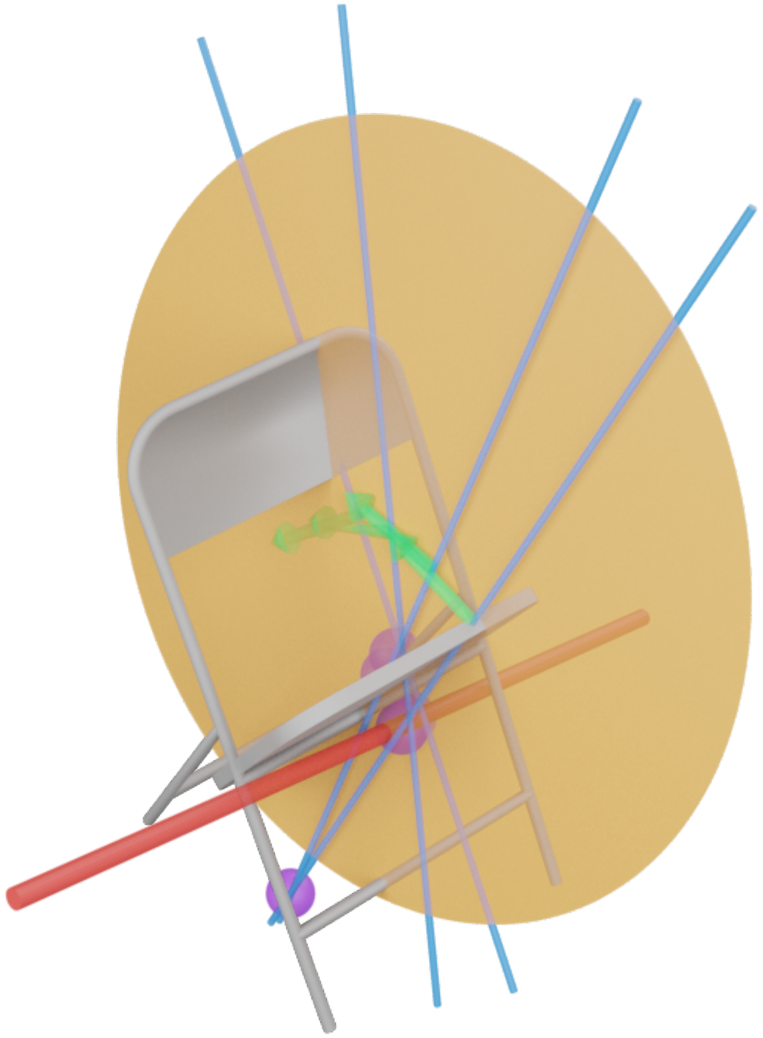


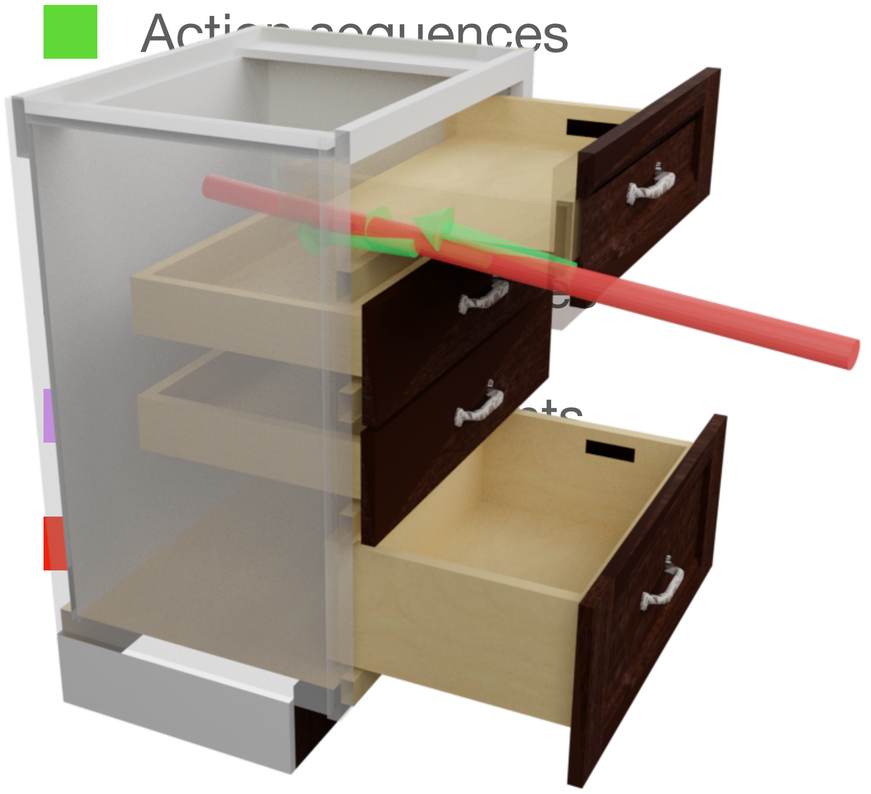




Action selected by the policy should reflects its belief on the objects' structure.

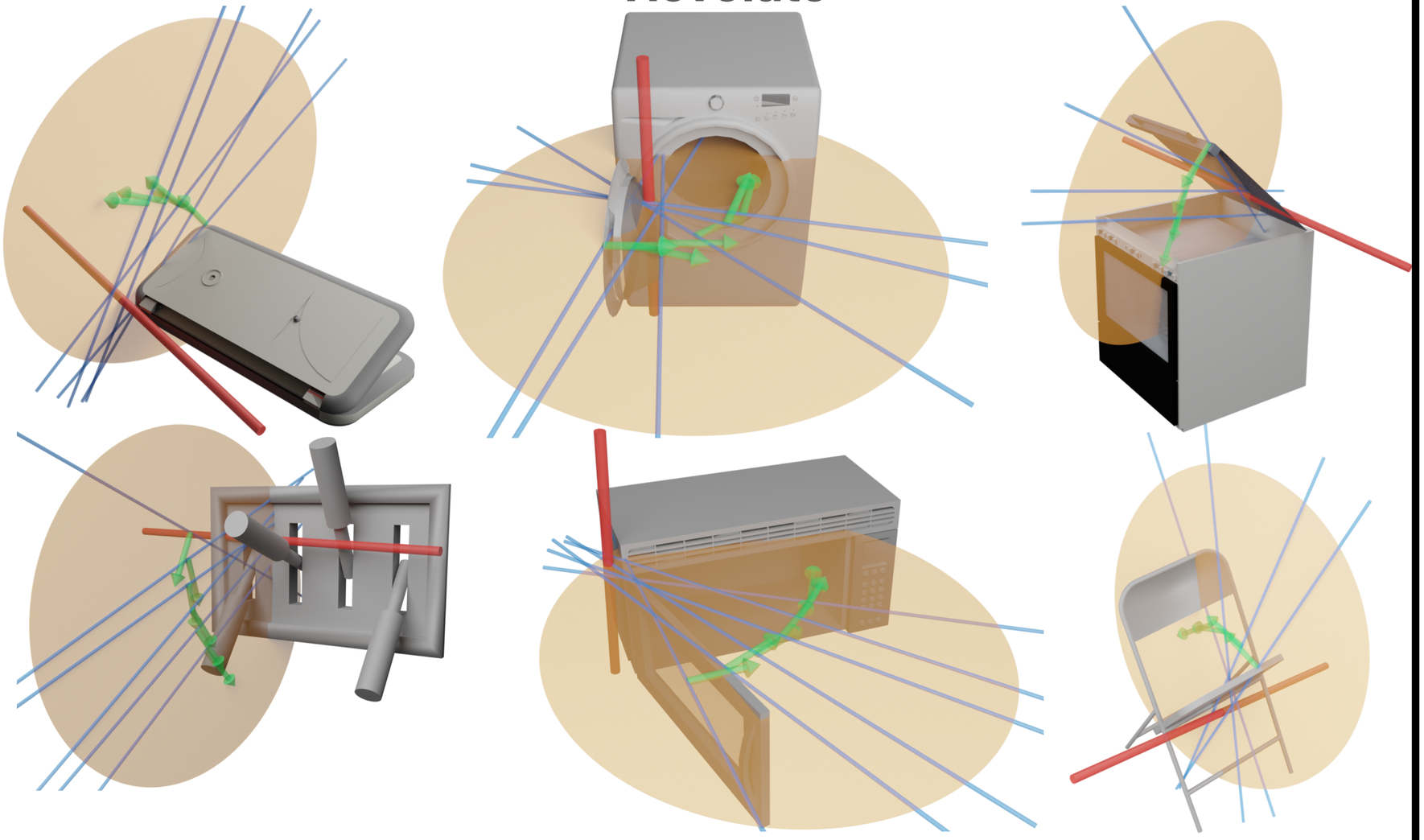
Action → Structure





Compute the joint parameters inferred from the actions selected by the policy

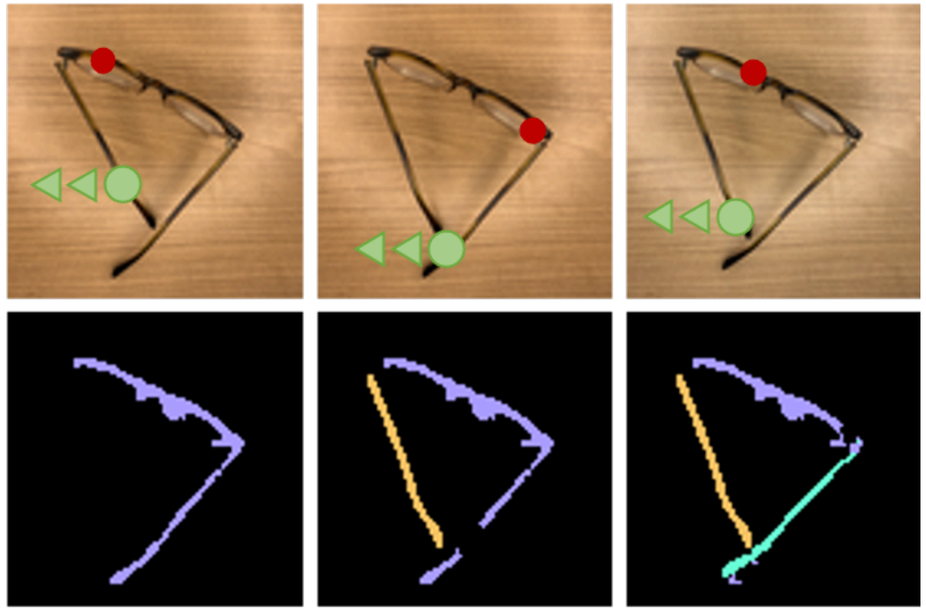
Articulation Structure Inference



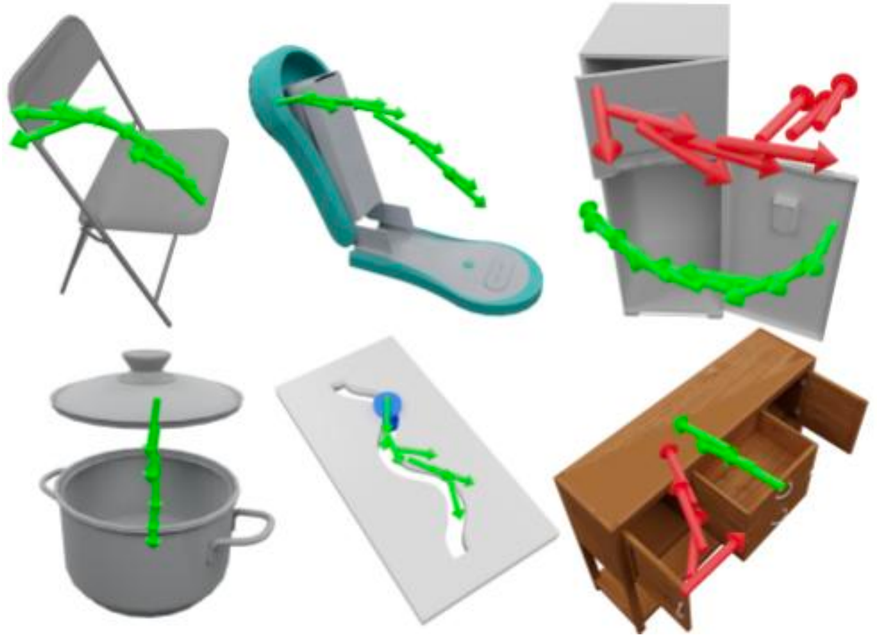
Project Webpage: ump-net.cs.columbia.edu

Structure from Action

Act the Part: Learning to Interact to Discover Articulated Object Structure



UMPNet: Universal Manipulation Policy Network for Articulated Objects



Underlaying structure of object through interaction

Generalize beyond a specific object instance or category

Acknowledgements

Act the Part: Learning to Interact to Discover Articulated Object Structure
Samir Y. Gadre, Kiana Ehsani, Shuran Song


Samir Y. Gadre


Kiana Ehsani

UMPNet: Universal Manipulation Policy Network for Articulated Objects
Zhenjia Xu, Zhanpeng He, Shuran Song


Zhenjia Xu


Zhenjia Xu

Google amazon.com

Thank You!

What’s next