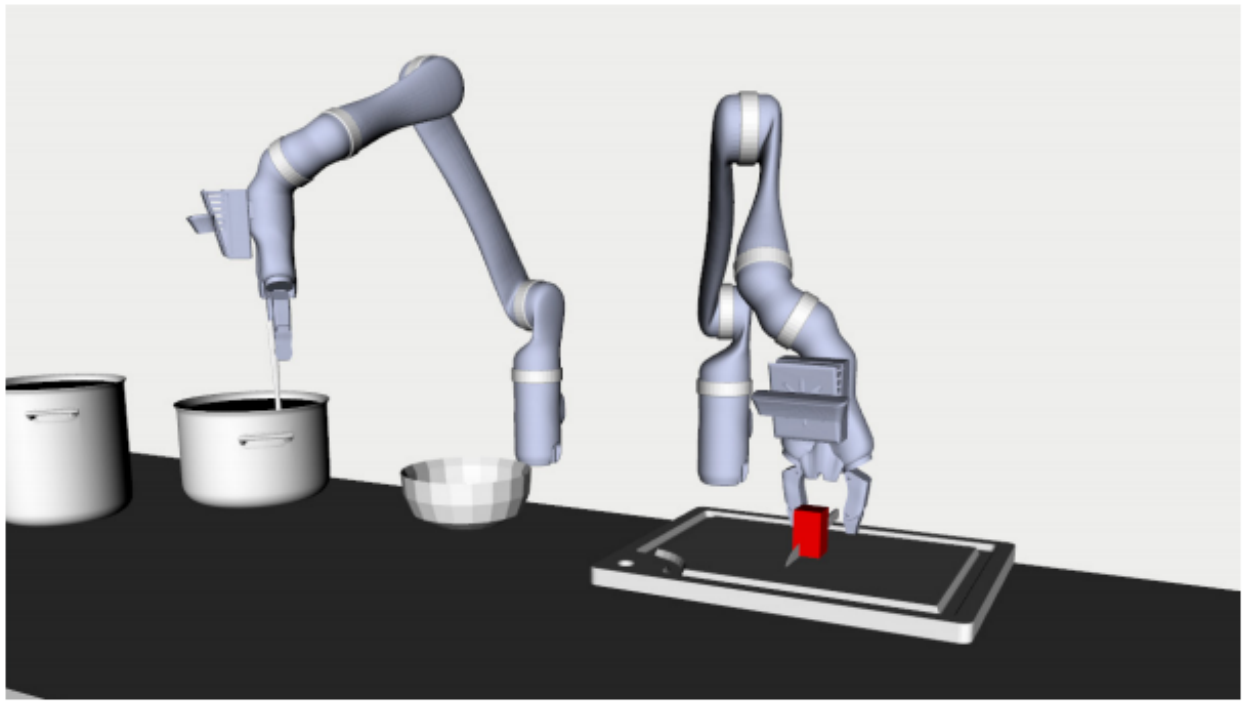
Robot Learning Collaborative Manipulation Plans from YouTube Cooking Videos



Zhang, H. and Nikolaidis, S., 2019. Robot learning and execution of collaborative manipulation plans from youtube cooking videos.

Hejia Zhang and Stefanos Nikolaidis
Department of Computer Science, University of Southern California

CARoS

USC Viterbi School of Engineering

LinkedIn
skype
LiVEJOURNAL
foursquare
vimeo
YouTube
Insta
facebook
Blogger
my
twitter
flickr
Behance
tumblr.
Like
yelp

Research Question

How can the robot **learn** and **execute collaborative** manipulation plans from online videos?

Key Insight

Hands contain temporal and spatial information of performed actions and manipulated objects

Our approach

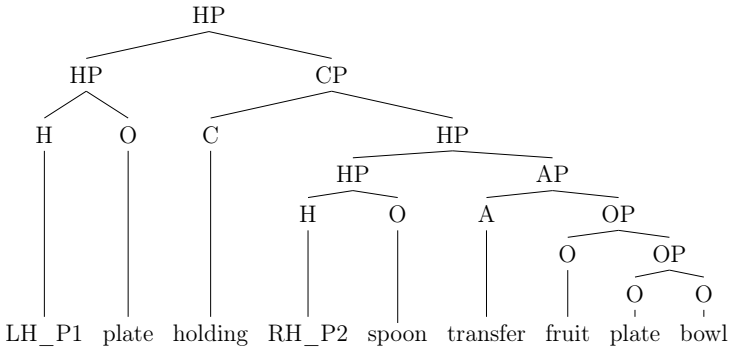
Input Video

Hand Detection and Object Association

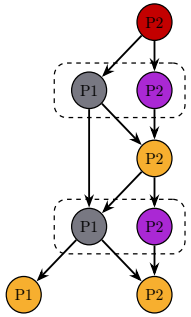
Segmentation

cut | pour | **transfer** | spread | grip | stir | sprinkle | wrap | coat | roll | **holding** | handover ....

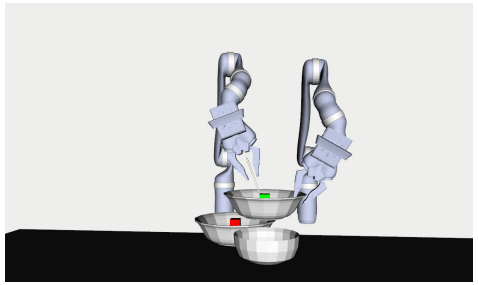
Action Recognition



Symbolic Command Generation



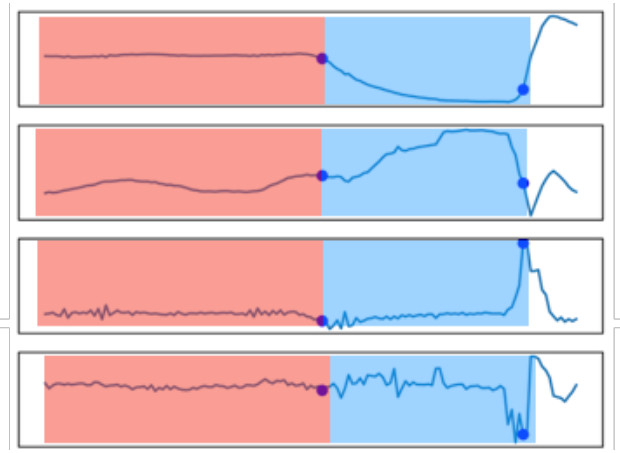
Action Graph Generation

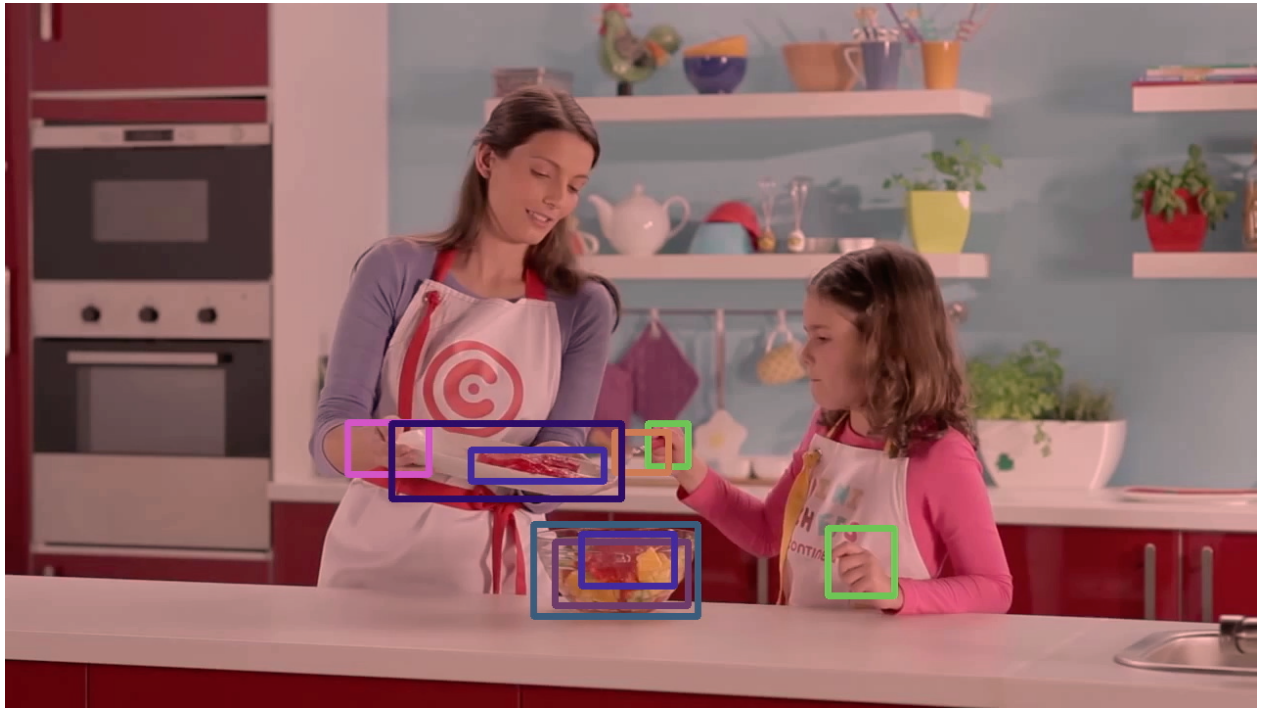


Robot Execution

We use deep learning techniques to detect hands and objects in the video.

We segment the video to be several action clips based on the hand trajectories.





We infer performed actions based on the association of hands and objects.

We infer performed actions based on the association of hands and objects.

We infer performed actions based on the association of hands and objects.

P(*action* | knife, chicken, board)

We infer performed actions based on the association of hands and objects.

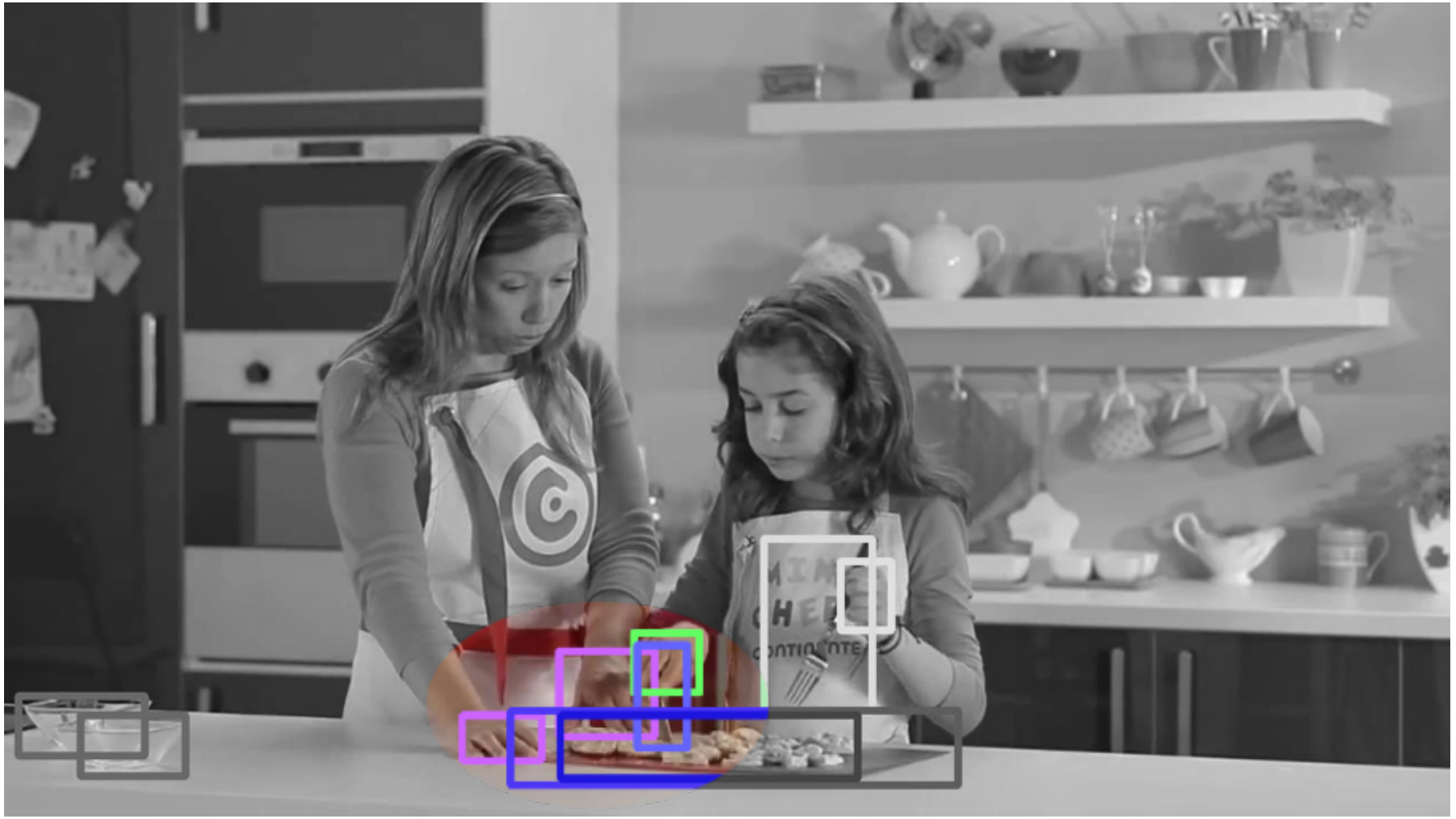
P(cut | knife, chicken, board)

Yang, Y., Li, Y., Fermuller, C. and Aloimonos, Y., 2015, March. Robot learning manipulation action plans by" watching" unconstrained videos from the world wide web. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.

We infer performed actions based on the association of hands and objects.

We infer performed actions based on the association of hands and objects.

Collaborative action: holding handover



We infer performed actions based on the association of hands and objects.

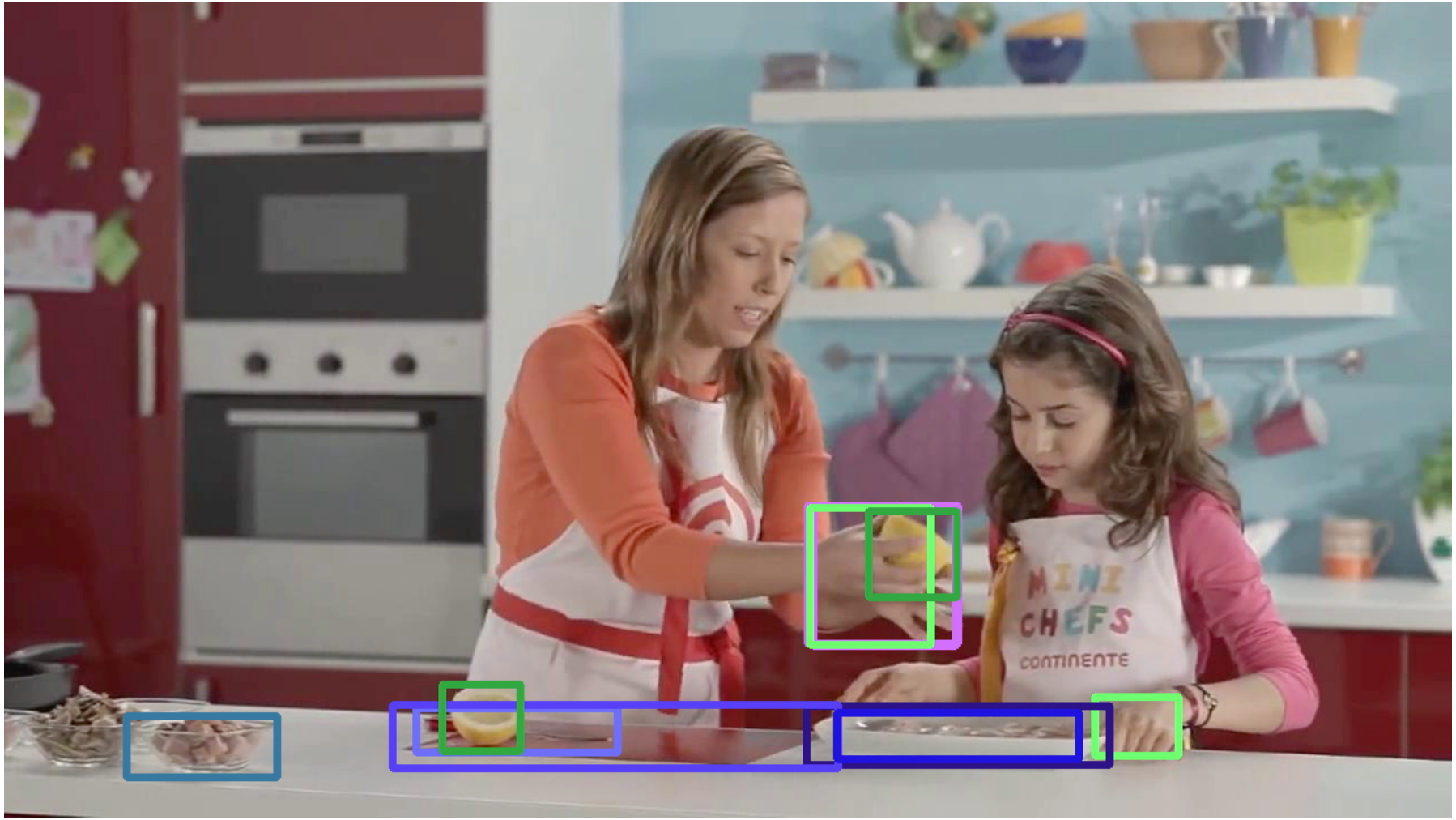
Collaborative action: holding handover

We infer performed actions based on the association of hands and objects.

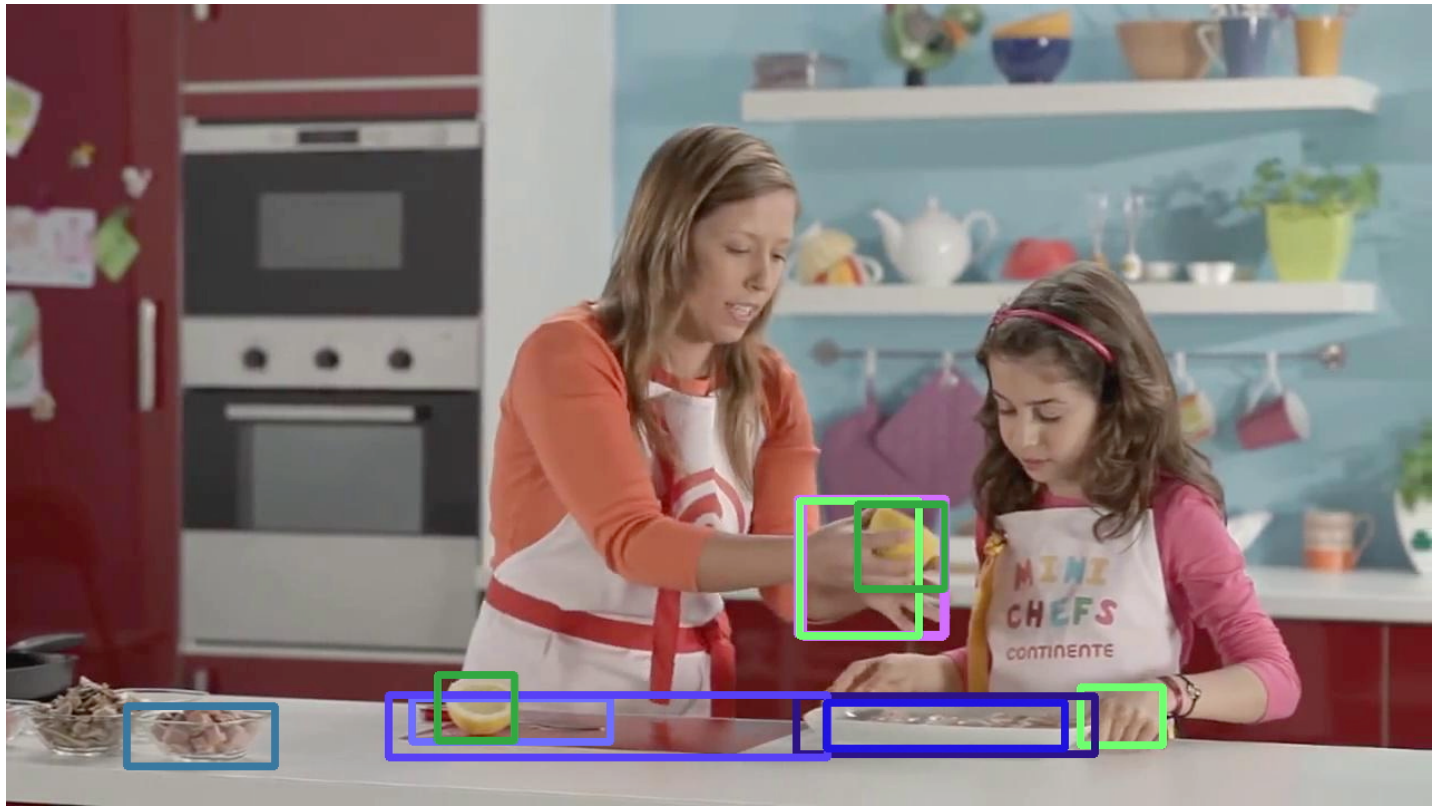
Collaborative action: holding handover

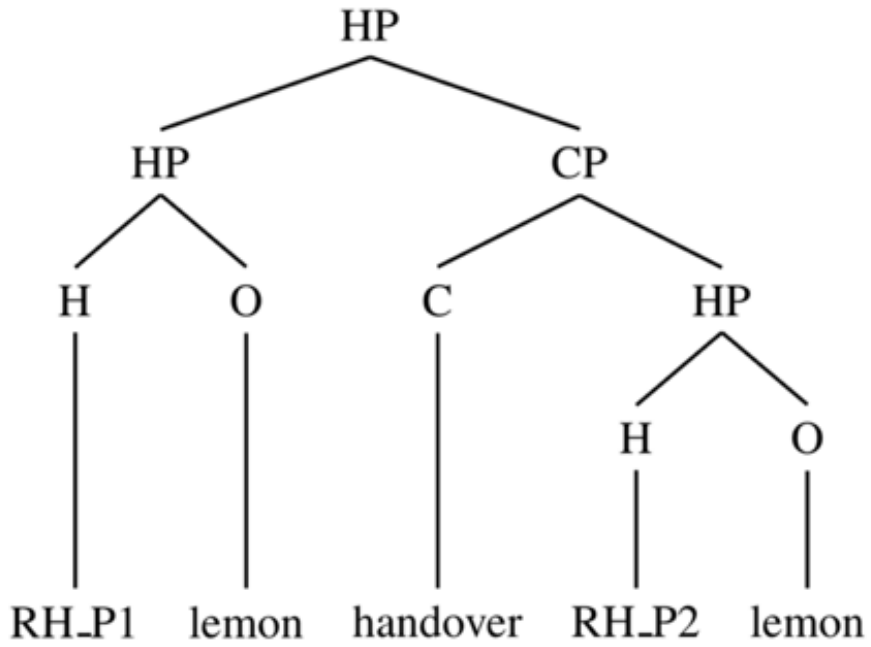
We infer performed actions based on the association of hands and objects.

Collaborative action: holding handover

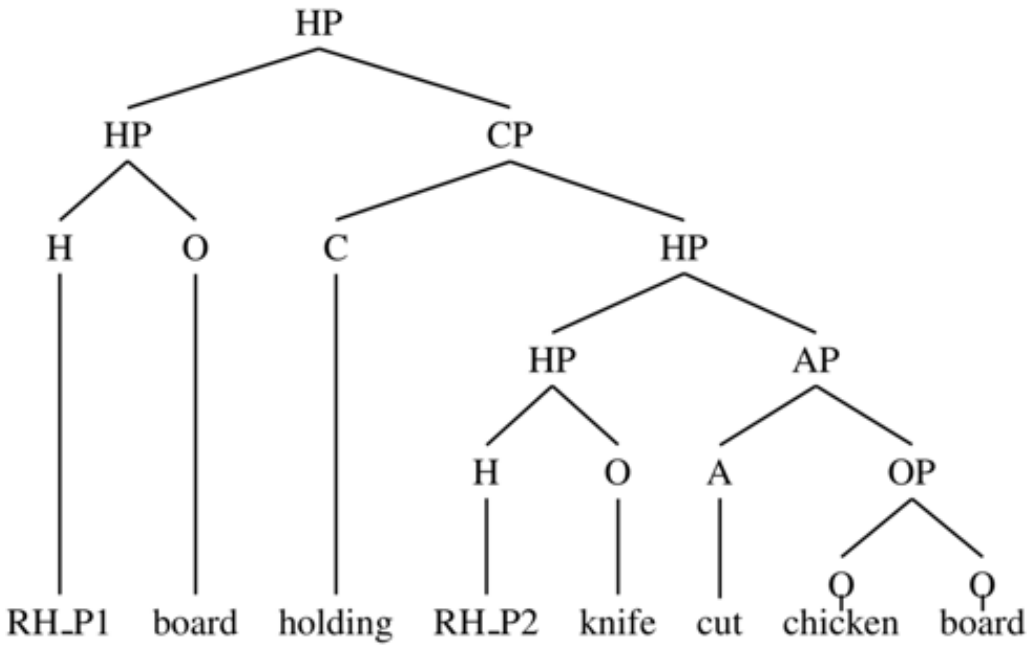


Example: Handover



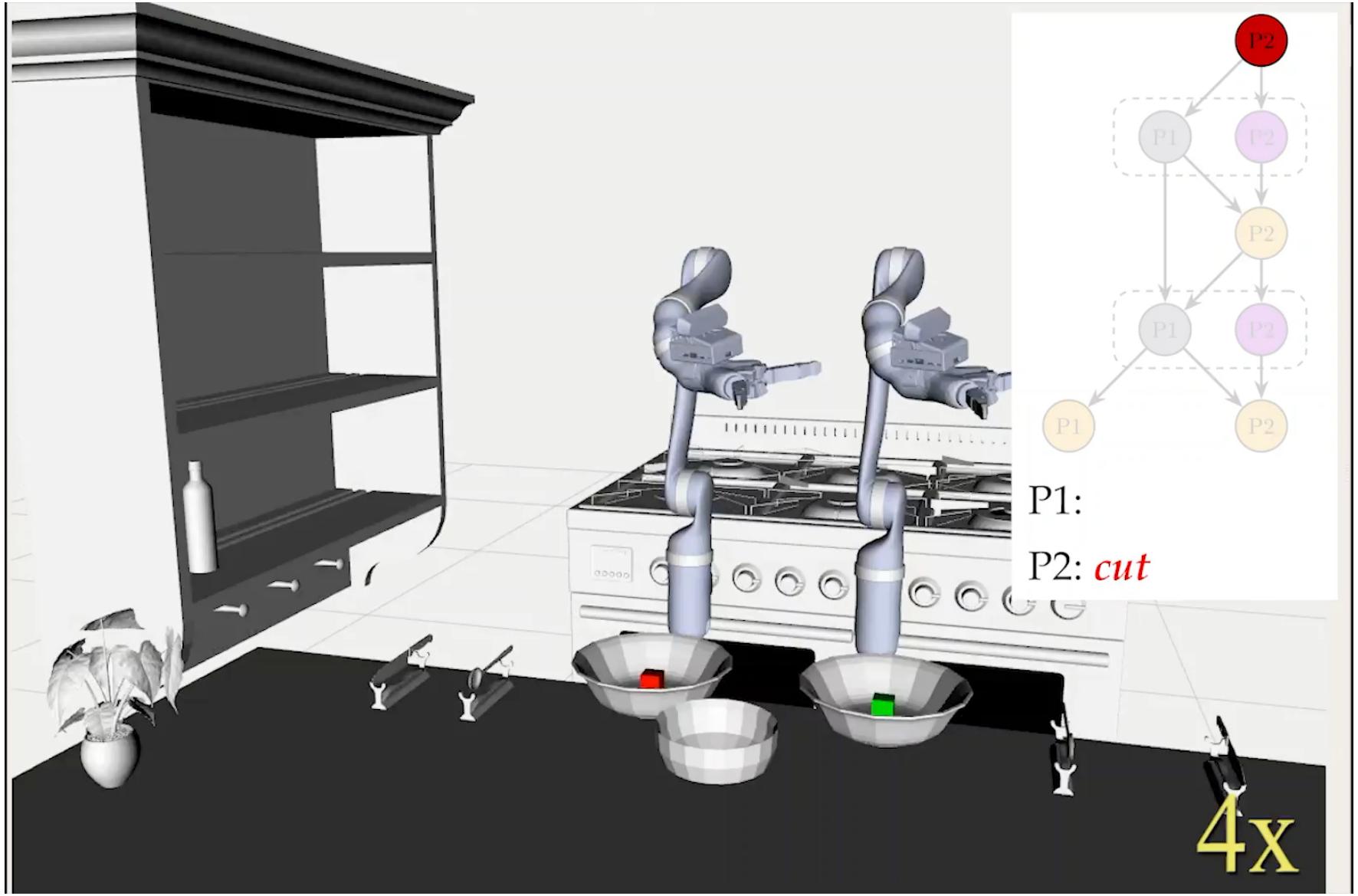


Example: Holding / Cut



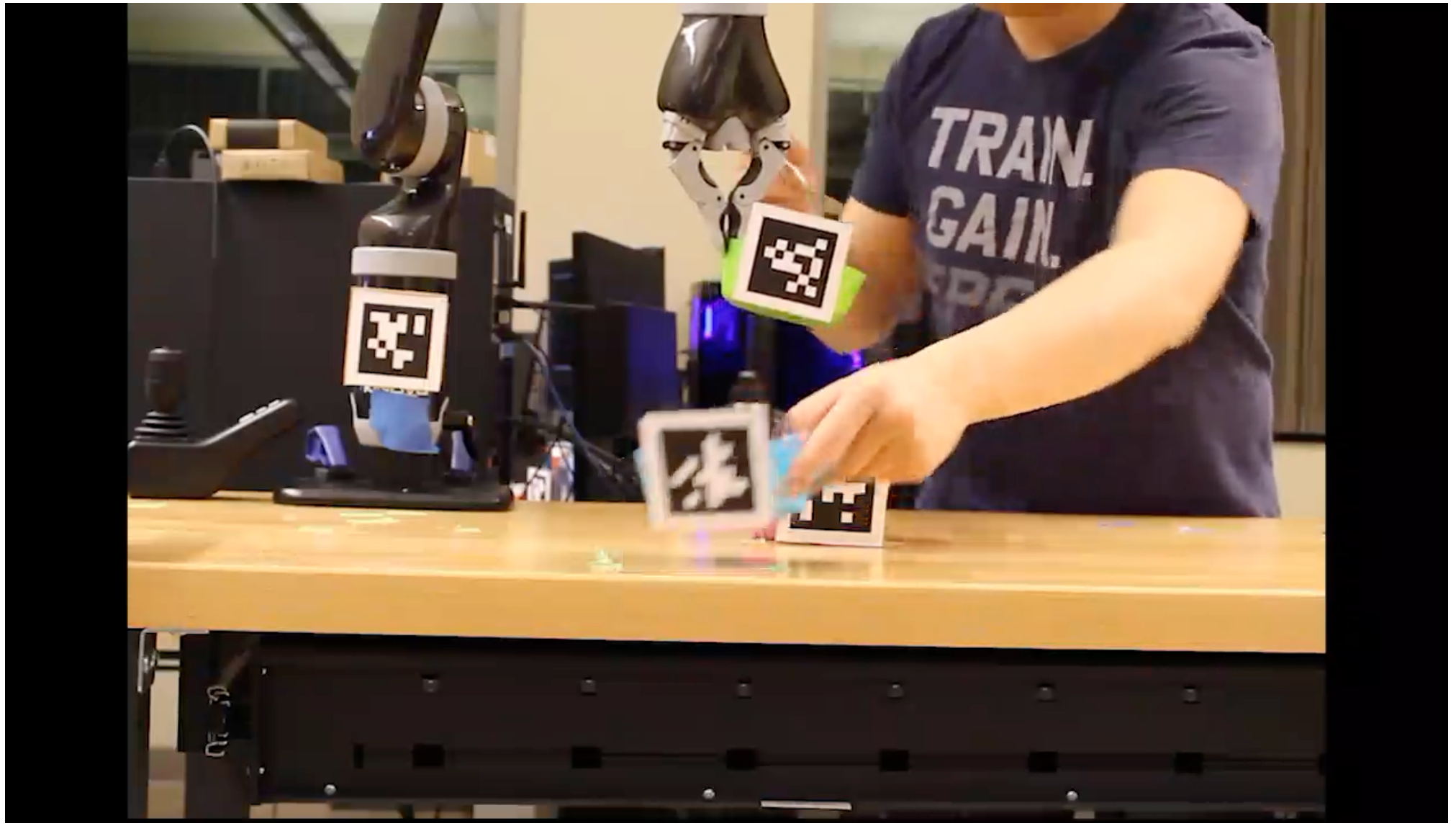
Based on the extracted action sequence, we construct a robot executable action graph.

Demo: Action Graph Execution (Simulation).



WeCook: https://github.com/icaros-usc/wecook

Demo: Action Graph Execution (Real World).



WeCook: https://github.com/icaros-usc/wecook

Future Work

When a collaborative action is about to happen?

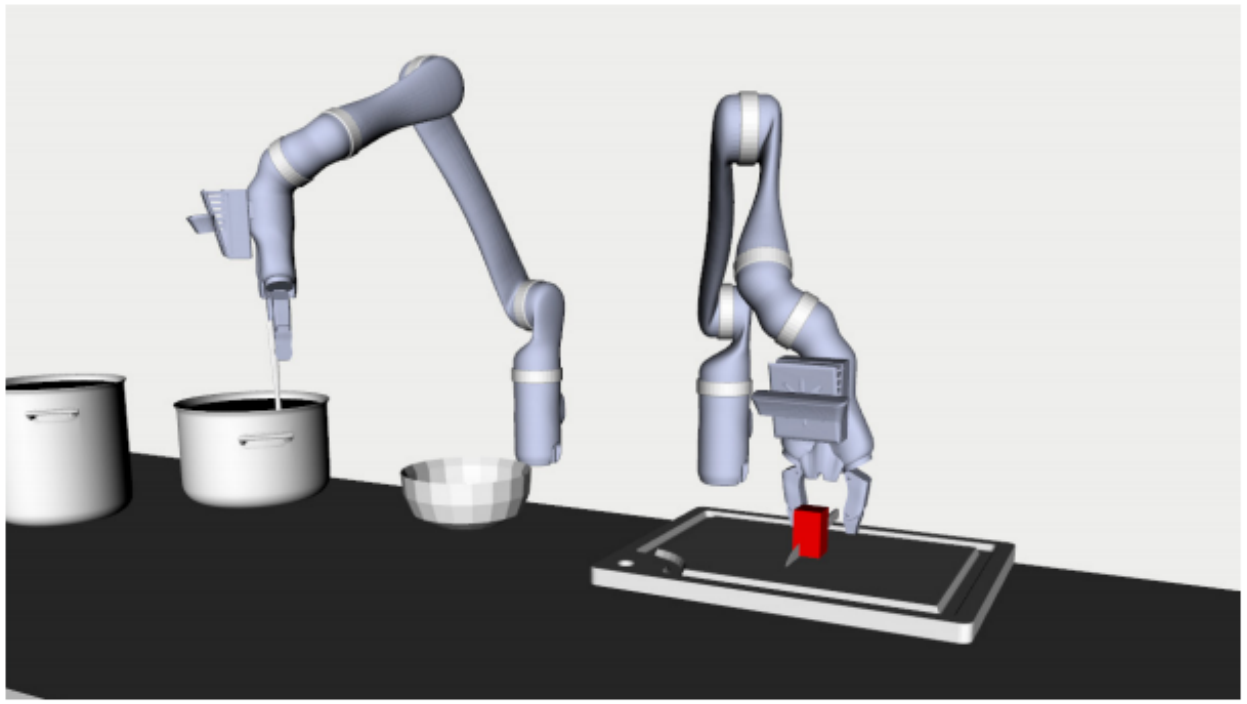
How collaborative actions emerge in close proximity interactions?

Why a collaborative action happens?

Contribution

By leveraging hand-object and object-object associations in unconstrained YouTube videos, robots can **learn** and **execute** human-interpretable, collaborative manipulation plans.

Robot Learning Collaborative Manipulation Plans from YouTube Cooking Videos



Zhang, H. and Nikolaidis, S., 2019. Robot learning and execution of collaborative manipulation plans from youtube cooking videos.

Hejia Zhang and Stefanos Nikolaidis
Department of Computer Science, University of Southern California

ICAROS

USC Viterbi School of Engineering

WeCook: https://github.com/icaros-usc/wecook