How Do I Do That? Synthesizing 3D Hand Motion & Contacts for Everyday Interactions

Aditya Prakash*, Benjamin Lundell, Dmitry Andreychuk, David Forsyth, Saurabh Gupta†, Harpreet Sawhney†

UNIVERSITY OF ILLINOIS URBANA-CHAMPAIGN · Microsoft

Motivation

Performing everyday activities require hand-object interactions of varying skill levels

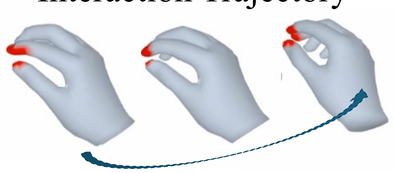
Application: Training people at scale

AR & XR devices as assistive tools

Task

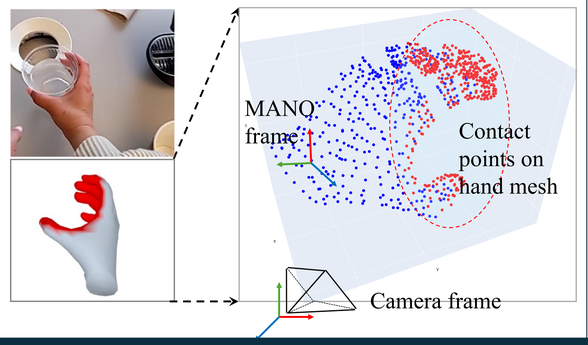
Input image + contact point

Interaction Trajectory



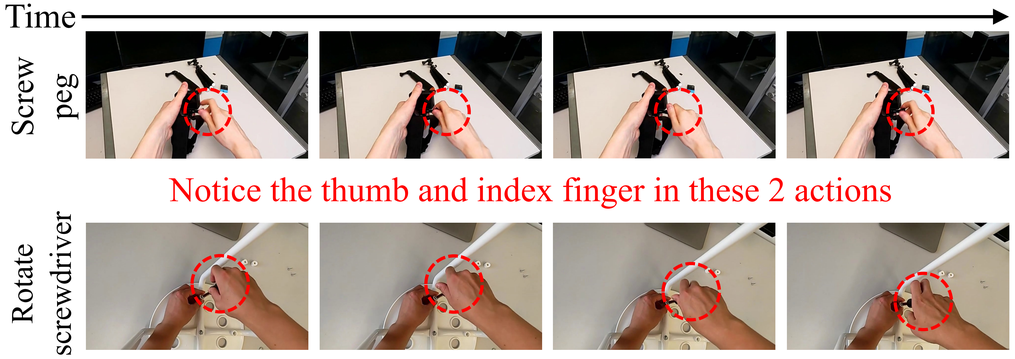
Predicting future motion: 3D hand poses + contact maps

3D models of thin, small, occluded, dynamic objects in contact is not available

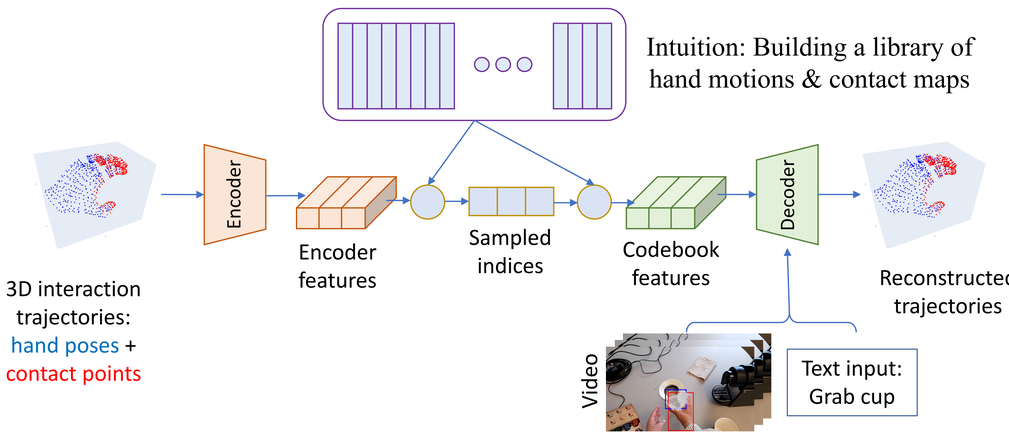


Approach

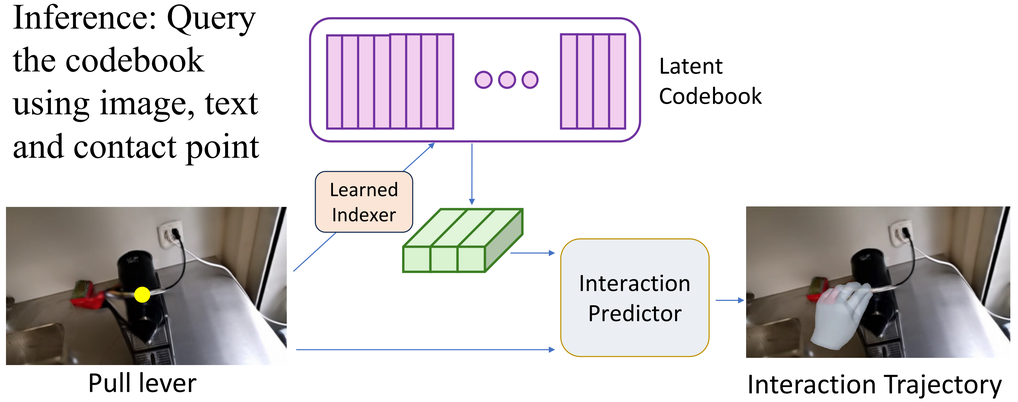
Insight: Similarities in motion & contacts across different actions



Idea: Learning a codebook of interactions using VQVAE



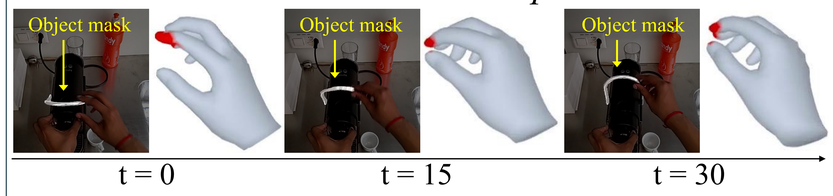
Inference: Query the codebook using image, text and contact point



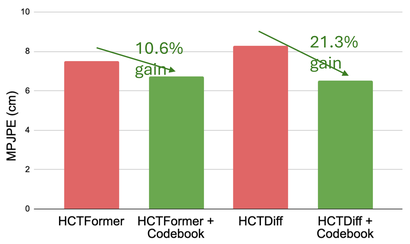
Experiments

Design a data engine to process large scale videos using off-the-shelf hand pose & segmentation models

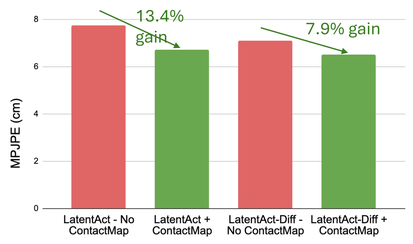
Generated annotations for the task *pull lever*



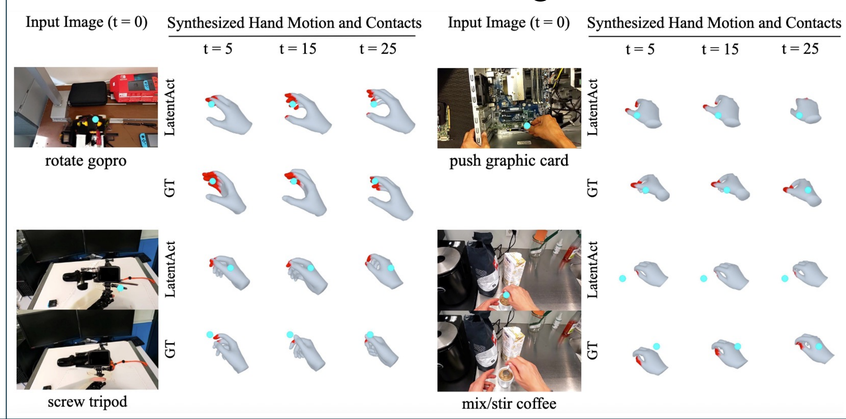
Results: Codebook leads to better generalization



Results: Contact maps help hand pose predictions



Visualizations across several settings:



*:Part of the work done during internship at Microsoft, †:Joint last authors, indicates equal contribution

Contact: adityap9@illinois.edu