

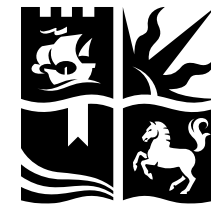
Understanding Hand-Object Interactions in Egocentric Video

Siddhant Bansal

University of Bristol, United Kingdom



PERCEIVING SYSTEMS
MAX PLANCK INSTITUTE FOR
INTELLIGENT SYSTEMS



University of
BRISTOL

What does Egocentric Mean?

Ego: A person's sense of self-esteem or self importance



What does Egocentric Mean?

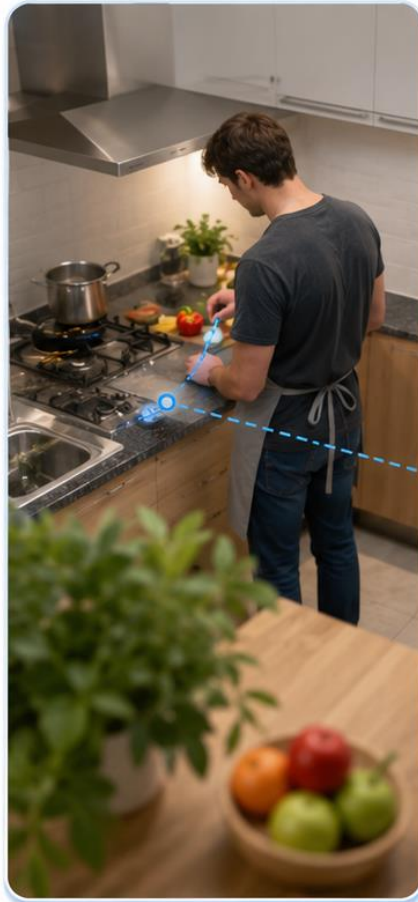
Egocentric: human is the central reference point



What does egocentric *video* mean?



Why do we care about them?



Challenges faced in Egocentric Videos

Transparent objects



Glass

Heavy Occlusion



Bottle in left hand

Challenging Natural Interaction



Glass in right hand



Bowl



Plate in left hand

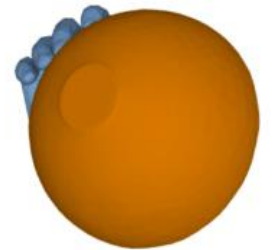


Cup

How to mitigate them while utilising Egocentric videos?

What if we could see the hand and
object in 3D?

How to mitigate them while utilising Egocentric videos?



Two steps in this direction

Reconstructing Objects along Hand Interaction Timelines in Egocentric Video



CVPR
JUNE 3-7, 2026



DENVER
COLORADO

Towards in-the-wild Egocentric 3D Hand-Object Pose Estimation



Reconstructing Objects along Hand Interaction Timelines in Egocentric Video

Zhifan Zhu¹, Siddhant Bansal¹, Shashank Tripathi², Dima Damen¹

¹University of Bristol, UK ²MPI for Intelligent Systems, Germany



PERCEIVING SYSTEMS
MAX PLANCK INSTITUTE FOR
INTELLIGENT SYSTEMS

CVPR
JUNE 3-7, 2026



DENVER
COLORADO



University of
BRISTOL

Goal

Estimate object's 3D
pose

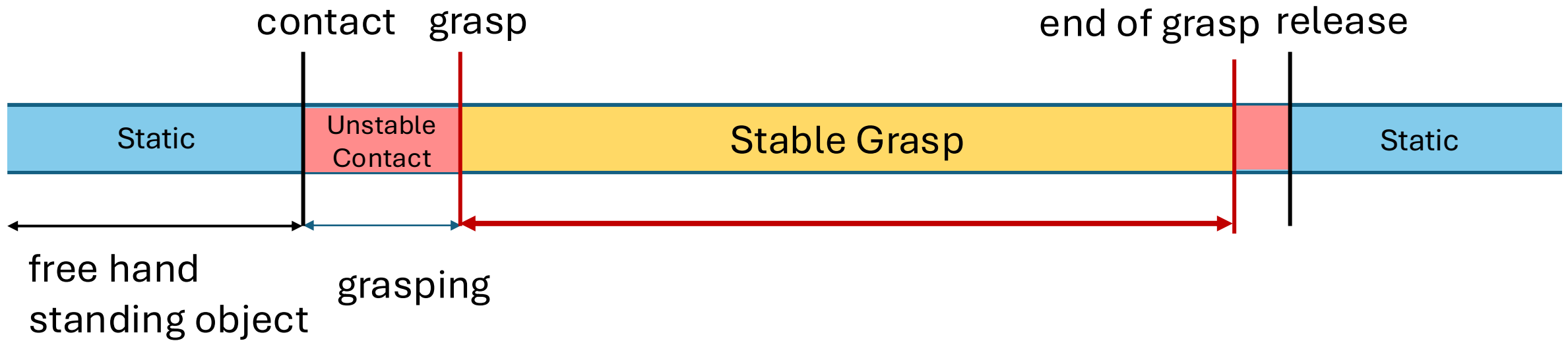
While targeting
challenging egocentric
view

But how do we select objects in this mess?



Defining Stable Grasp





EPIC-Grasps: Stable Grasp Boundaries in EPIC-Kitchens

Sequences	Instances	Categories	Subjects
2431	~390	9	31

1446 left hands



, 985 right hands



We extract Stable Grasps from 3D datasets

ARCTIC (CVPR 2023)



Zicong Fan, et al.
ARCTIC: A dataset for dexterous bimanual hand- object manipulation. **CVPR 2023**

HOI4D (CVPR 2022)



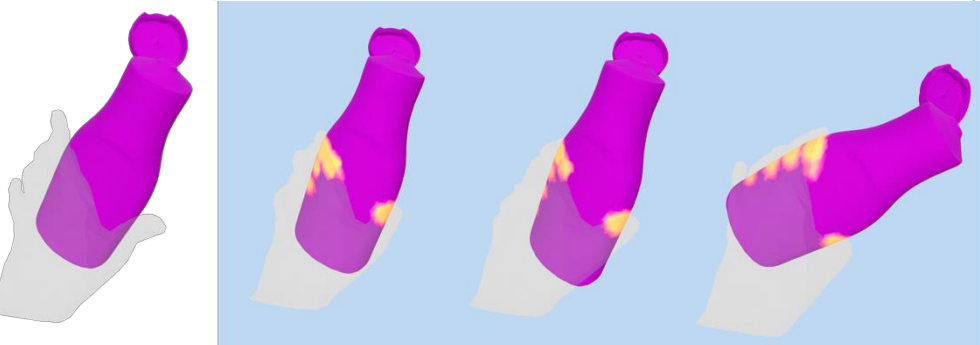
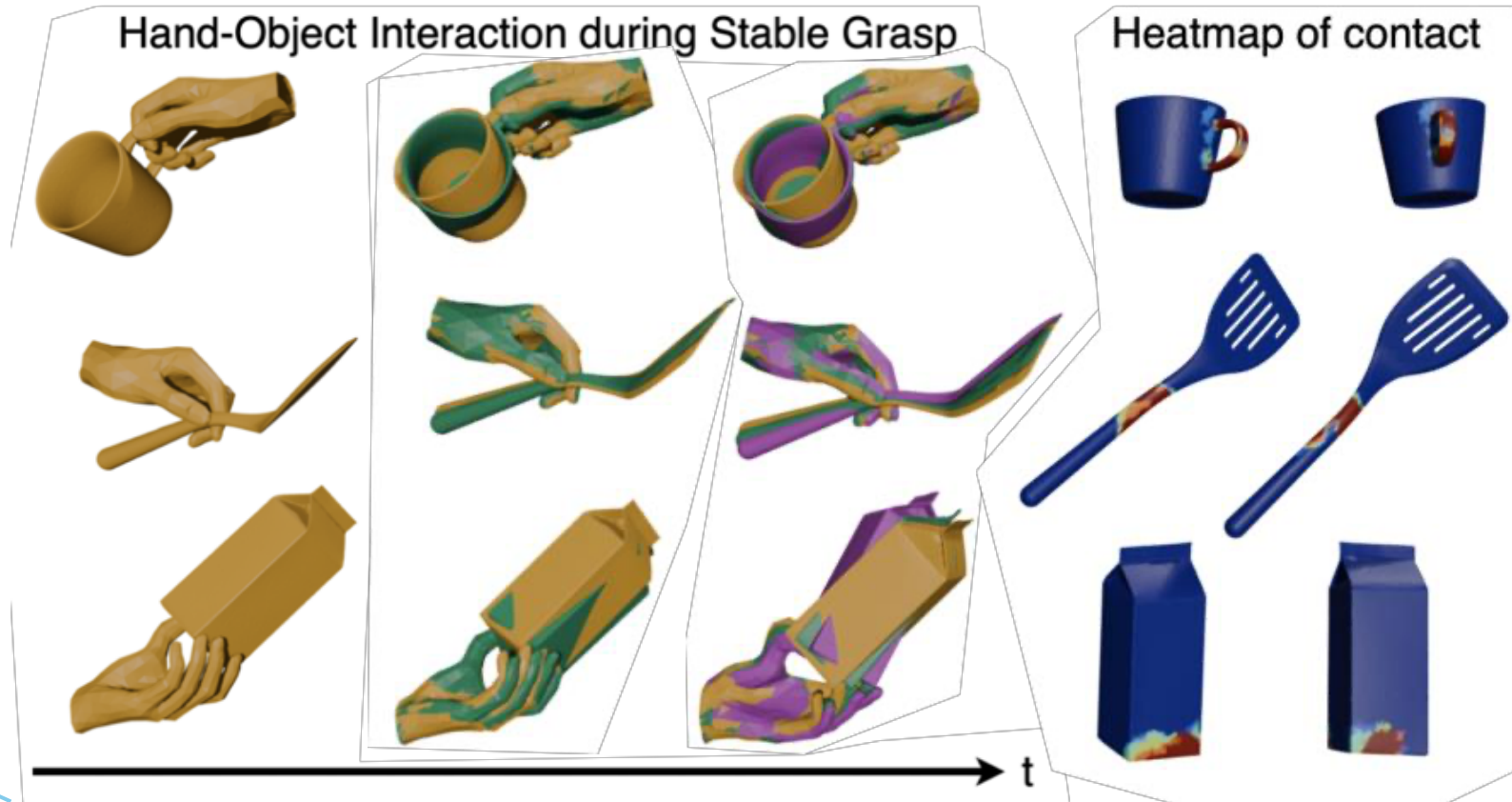
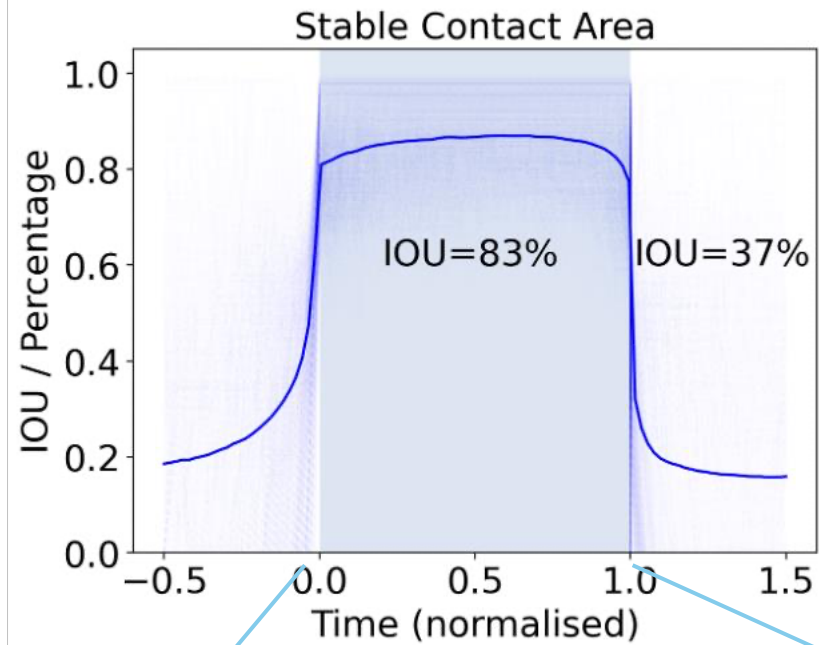
Yunze Liu, et al.
HOI4D: A 4D Egocentric Dataset for Category-Level Human-Object Interaction. **CVPR 2022**

HOT3D (CVPR 2025)

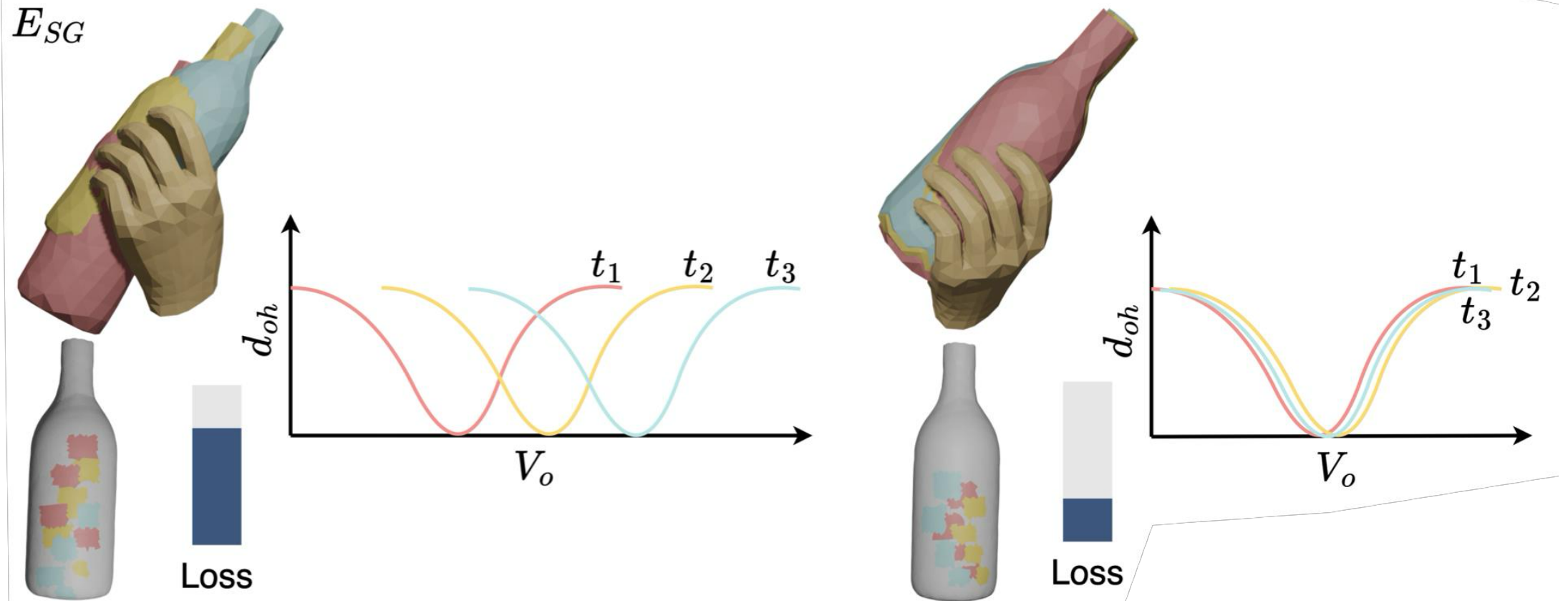


Prithviraj Banerjee, et al.
HOT3D: An egocentric dataset for 3D hand and object tracking. **CVPR 2025**

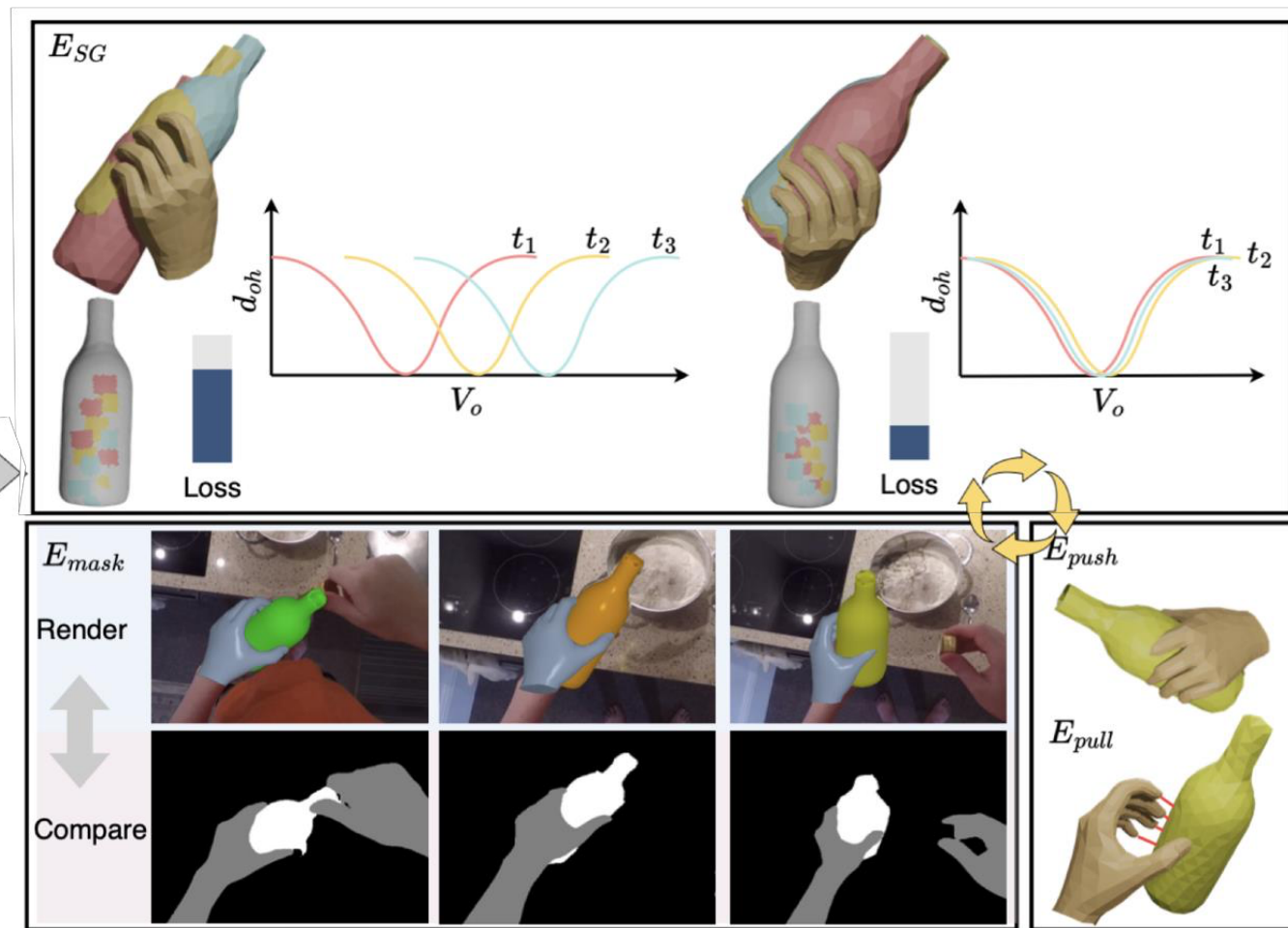
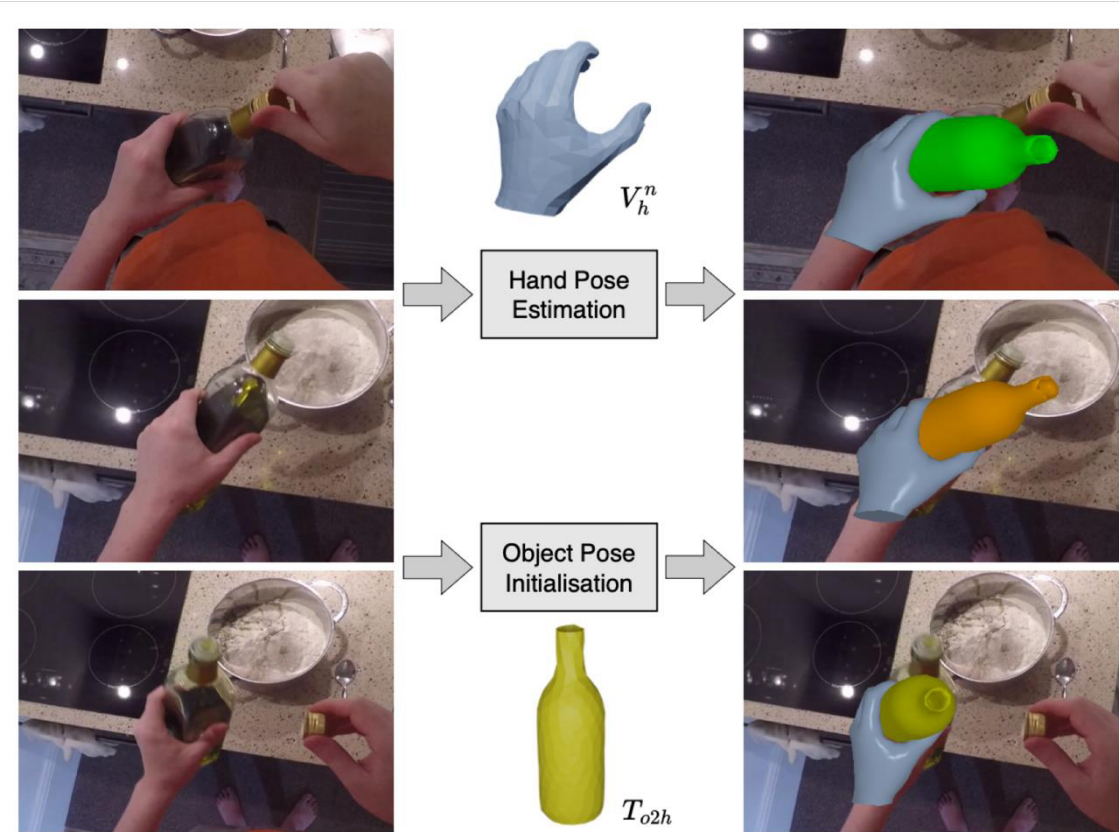
Visualising obtained Stable Grasps



The Stable Grasp Constraint



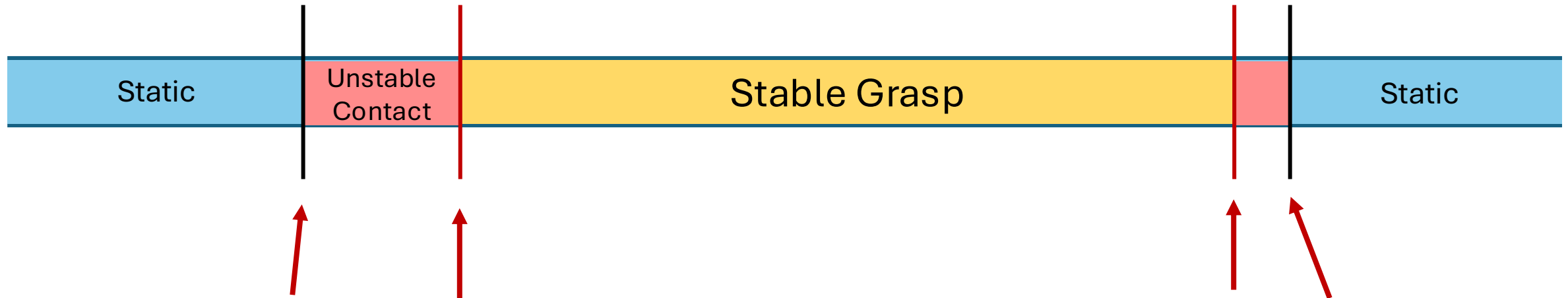
Method



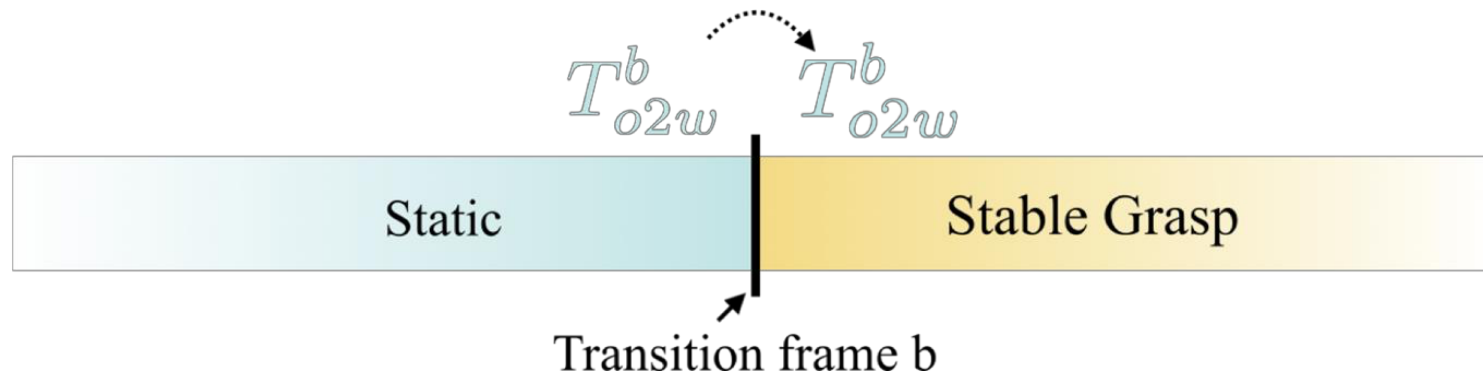
Bottle Samples



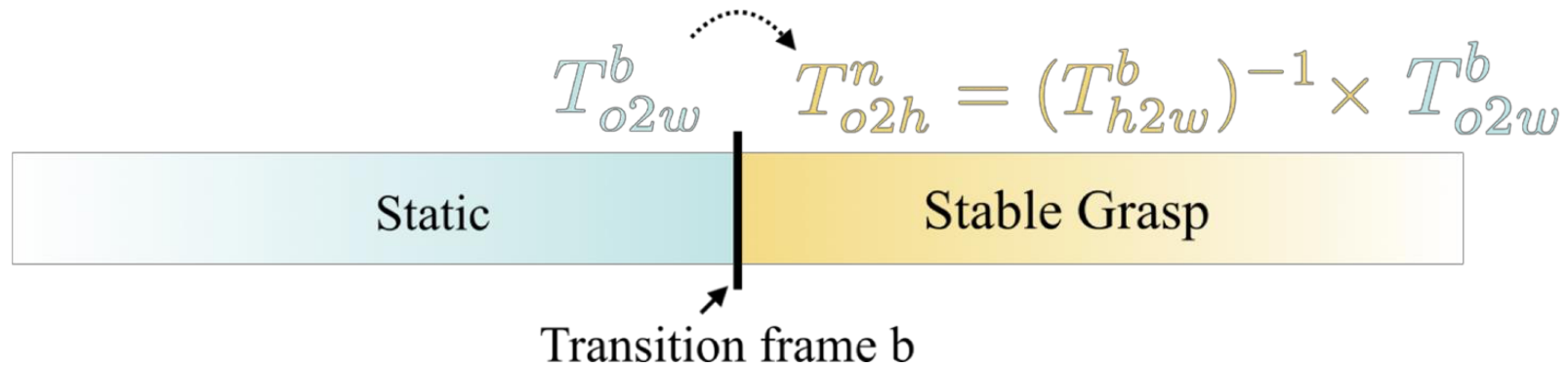
Getting Object's Pose for other Segments



Propagating poses at the transition frame



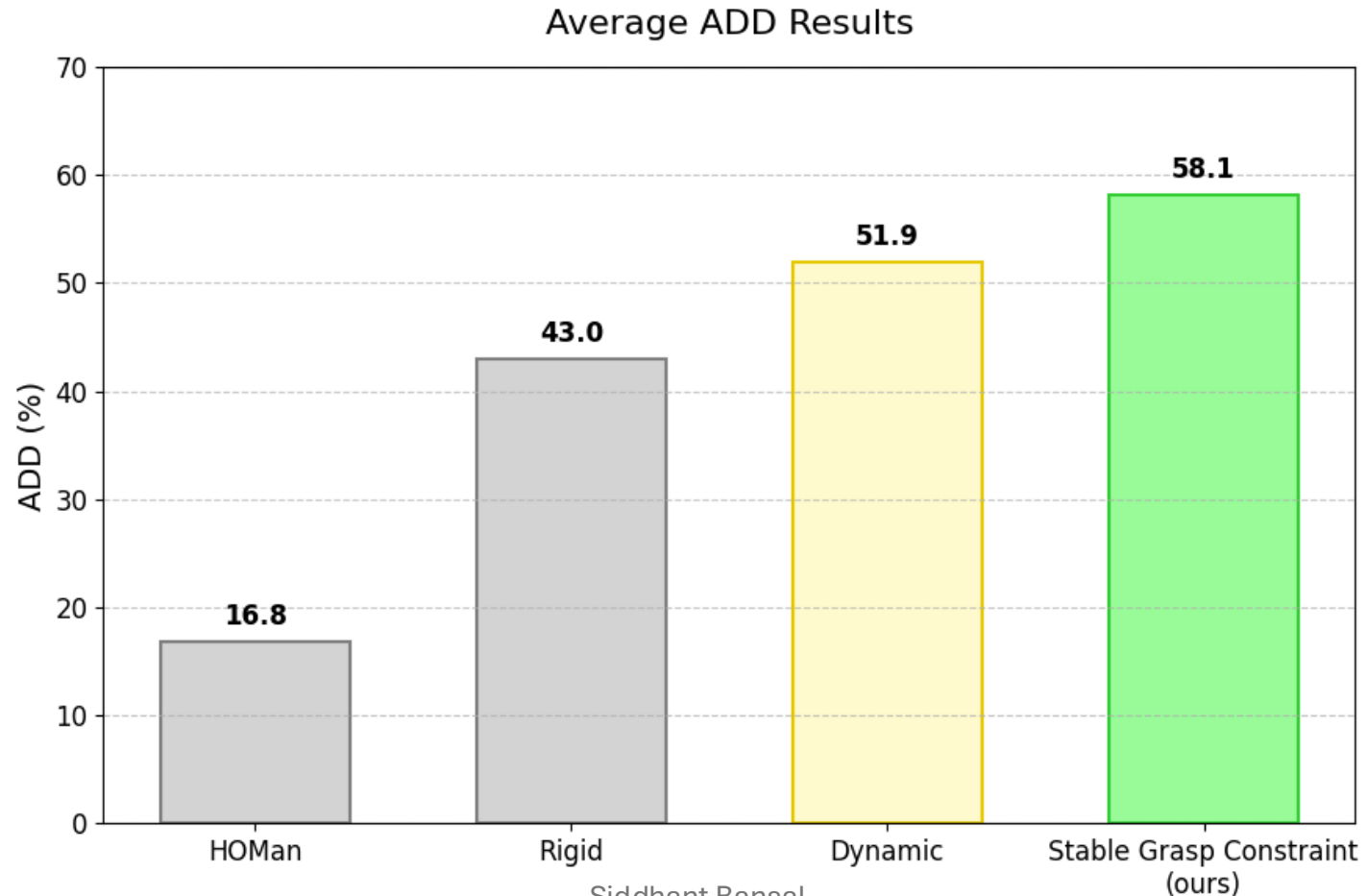
Propagating poses at the transition frame



Quantitative results

- Definition of **ADD** metric:

ADD is 1 if the average vertices displacement is less than 10% of the object diameter





EPIC-HIT



Siddhant Bansal
VISTA Lab, July 2026

This works, but there are issues

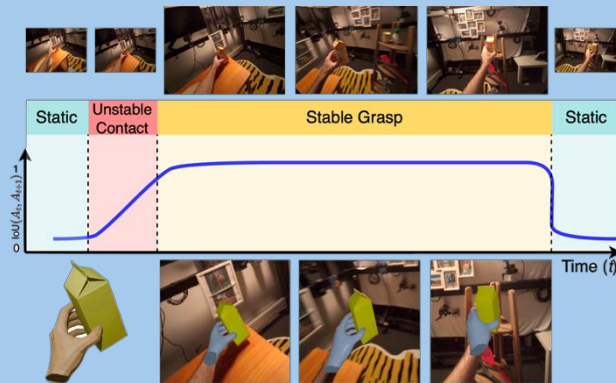
Needs dataset-level and
method-level
assumptions

Needs iterative
optimisation per-sample

There is no direct 3D
evaluation for in-the-
wild Egocentric images

Towards solving the issues

Reconstructing Objects along Hand Interaction Timelines in Egocentric Video



CVPR
JUNE 3-7, 2026



DENVER
COLORADO

Towards in-the-wild Egocentric 3D Hand-Object Pose Estimation



ECCV 2026 | MALMÖ | SEPT 8-13

Towards in-the-wild Egocentric 3D Hand-Object Pose Estimation

Siddhant Bansal¹, Zhifan Zhu¹, Shashank Tripathi², Jiahe Zhao¹, Michael Black², Dima Damen¹

¹University of Bristol ²MPI-Germany



PERCEIVING SYSTEMS
MAX PLANCK INSTITUTE FOR
INTELLIGENT SYSTEMS

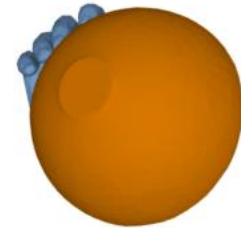
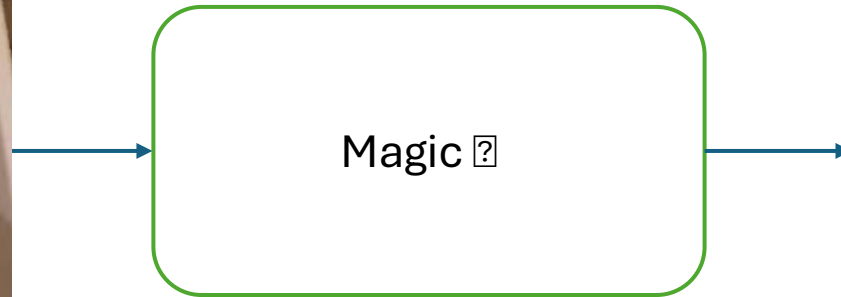


University of
BRISTOL

What are we trying to achieve?



In-the-wild Images



Single forward pass

A step towards in-the-wild

EPIC-Contact Dataset

2.3 K clips
64.2k frames

Creating the EPIC-Contact Dataset

Input to the annotators



3D MANO Hand

Hand Contact



Painted Hand Vertices

Annotator Agreement

$\kappa = 0.61$

Contact on Object

Region-wise Contact Axis



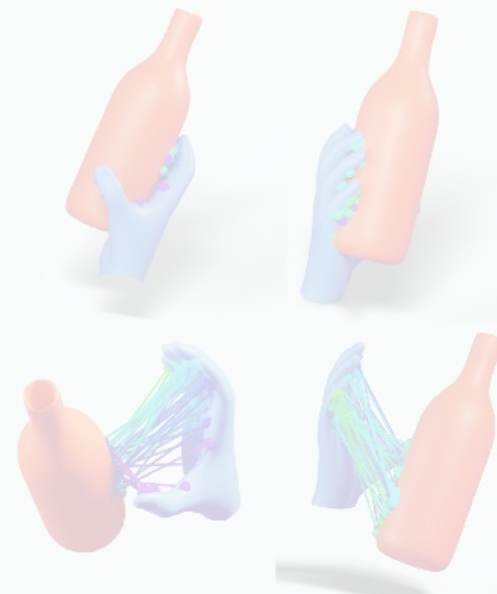
Annotator Agreement

$\kappa = 0.62$



Transferred contact

EC-fit: Posed Hand-object Mesh




Annotating Hand Contact Regions

Please annotate human contact using the **body part**:

leftHand with **MUG**

(left hand) mug



0:01 / 0:02

Left click and drag to rotate the body. Right click and drag to move the body. Use the mouse wheel to zoom in/out.



ERASE ALL RESET VIEW

Brush size:

1 2 3 4 5 6 7 8 9 10

DRAW ERASE

Selection Mode: Inactive



Contact Regions on Objects

Please annotate **LEFT** hand contact using the part:

fingers

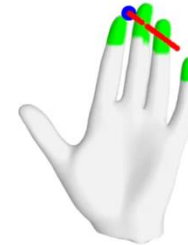
with

MUG



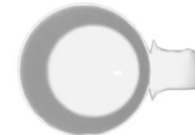
Left click and drag to rotate the body. Right click and drag to move the body. Use the mouse wheel to zoom in/out.

RESET VIEW



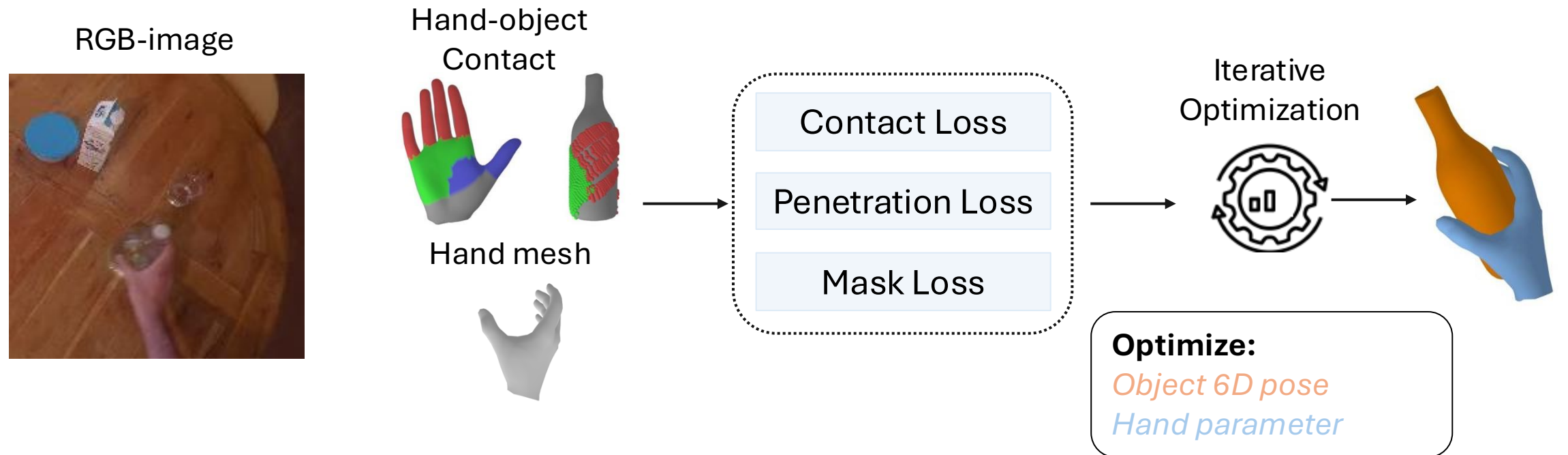
Move the contact patch as required

RESET VIEW



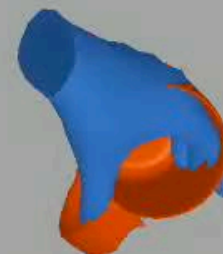
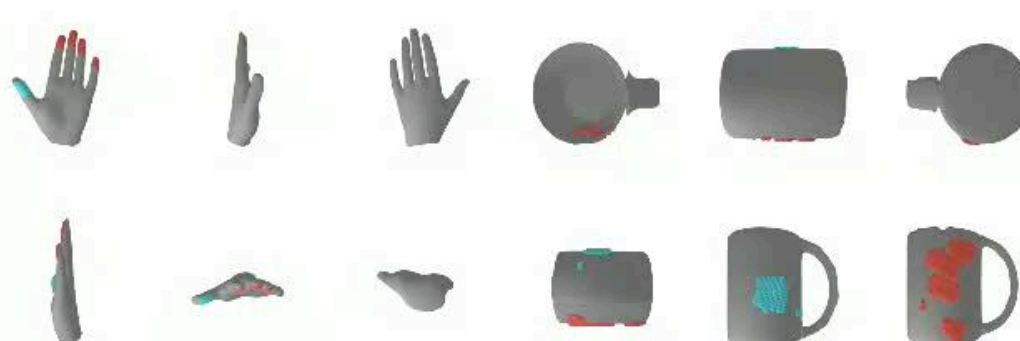
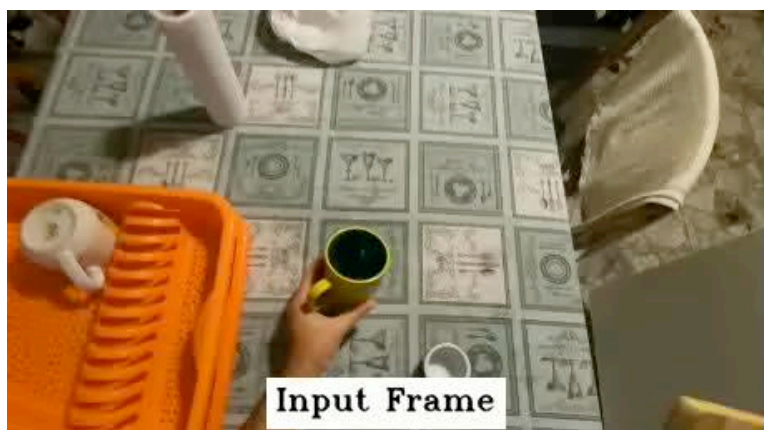
EC-Fit: Getting GT for in-the-wild data: Contact -> Pose

- PICO [CVPR25]: human-object interaction synthesis
- Adapt PICO to in-the-wild hand-object pose estimation



EC-Fit: From Contact to Posed Hand-Object Meshes

Contact-based Alignment



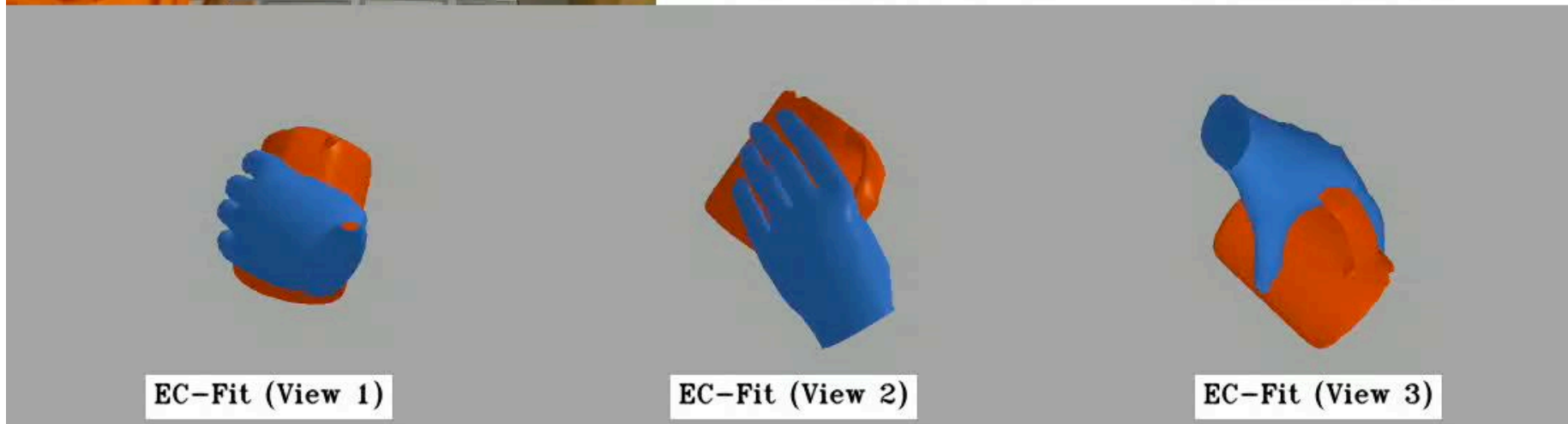
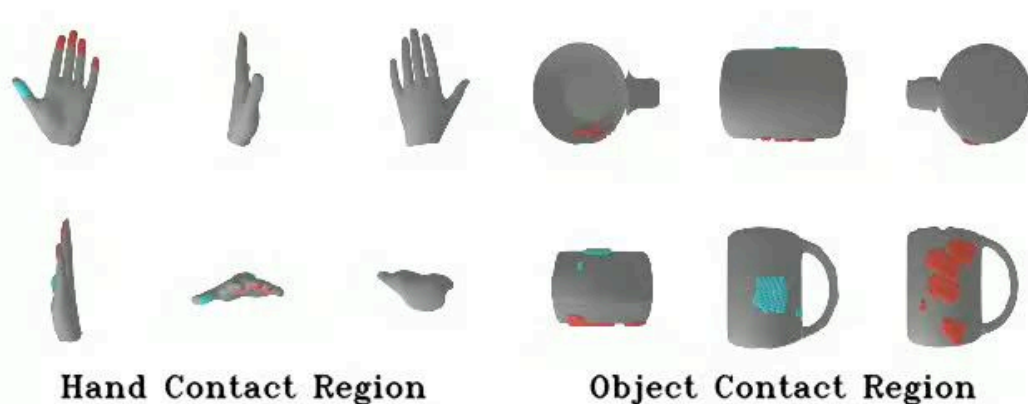
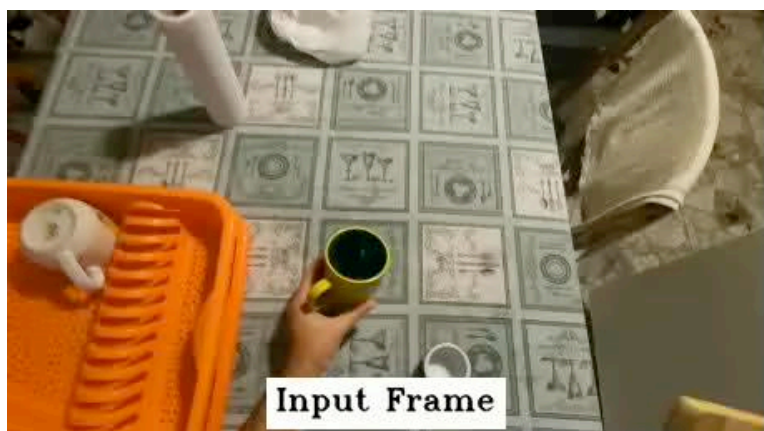
EC-Fit (View 1)

EC-Fit (View 2)

EC-Fit (View 3)

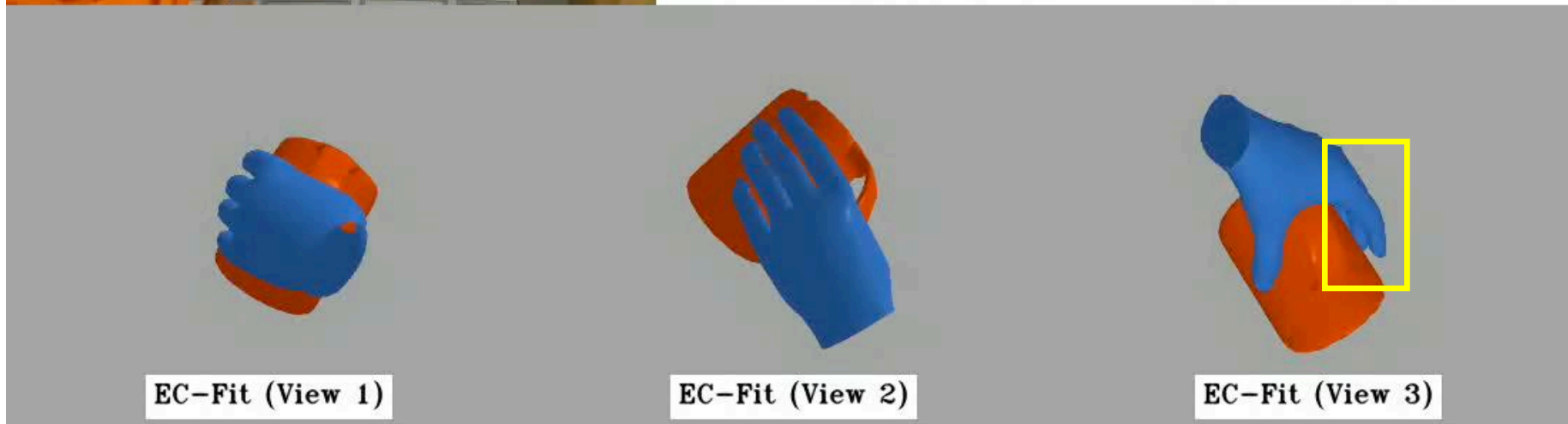
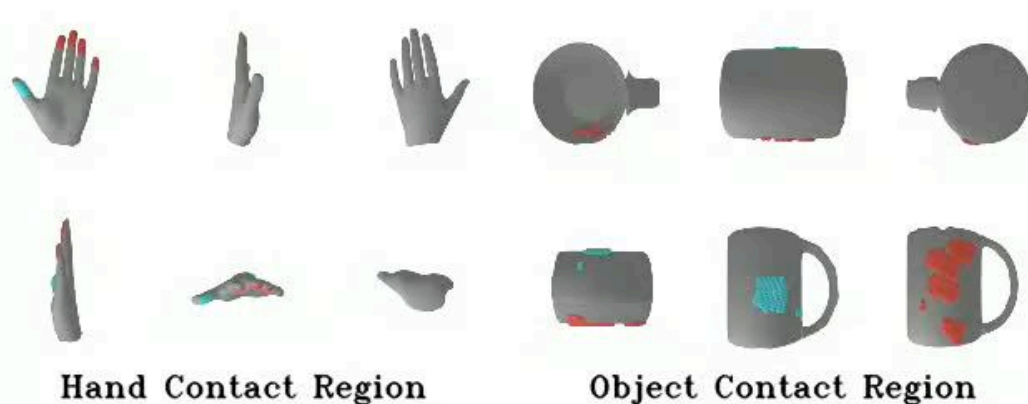
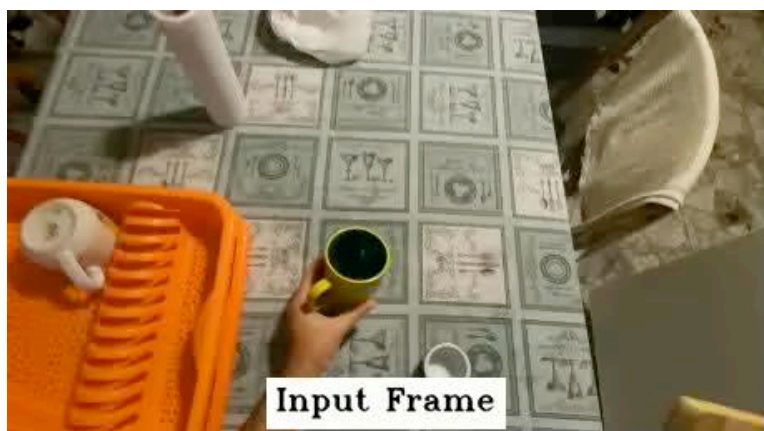
EC-Fit: From Contact to Posed Hand-Object Meshes

Image-guided Refinement: Updating object's pose



EC-Fit: From Contact to Posed Hand-Object Meshes

Image-guided Refinement: Updating object and hand's pose



EPIC-Contact Videos



An end-to-end learning-based method for in-the-wild egocentric images

EPIC-Contact Dataset

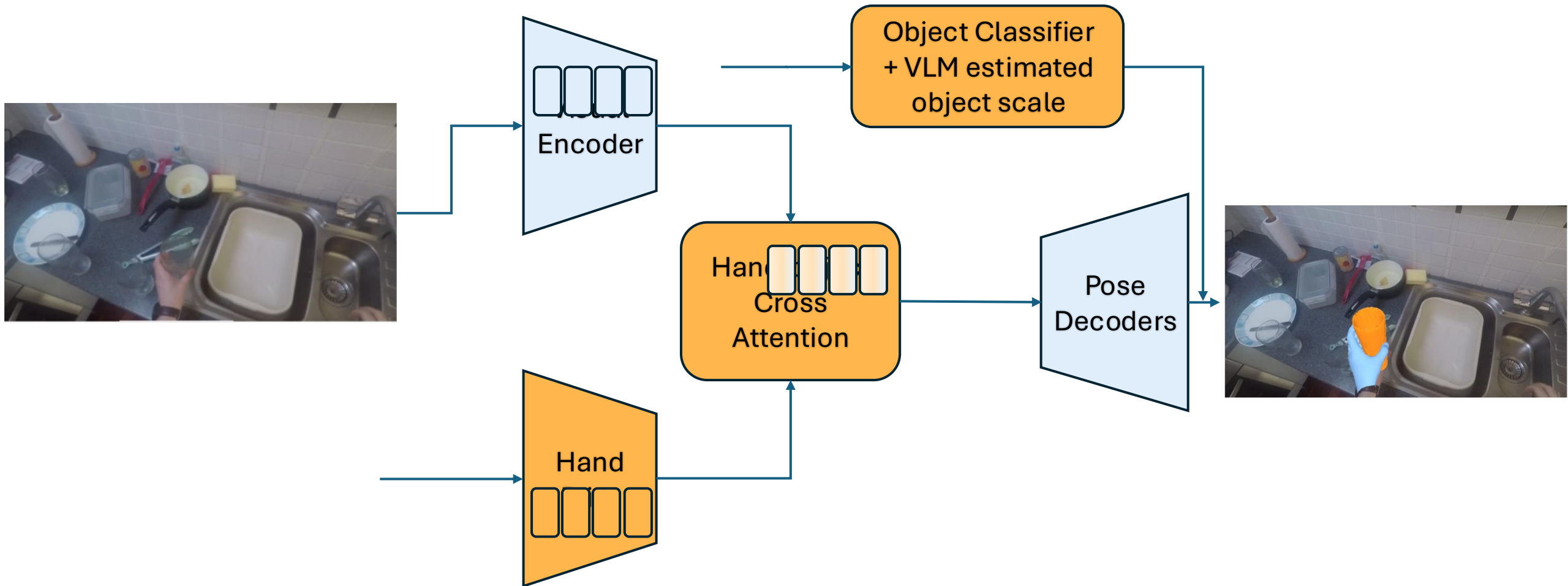


2.3K clips
64.2K frames

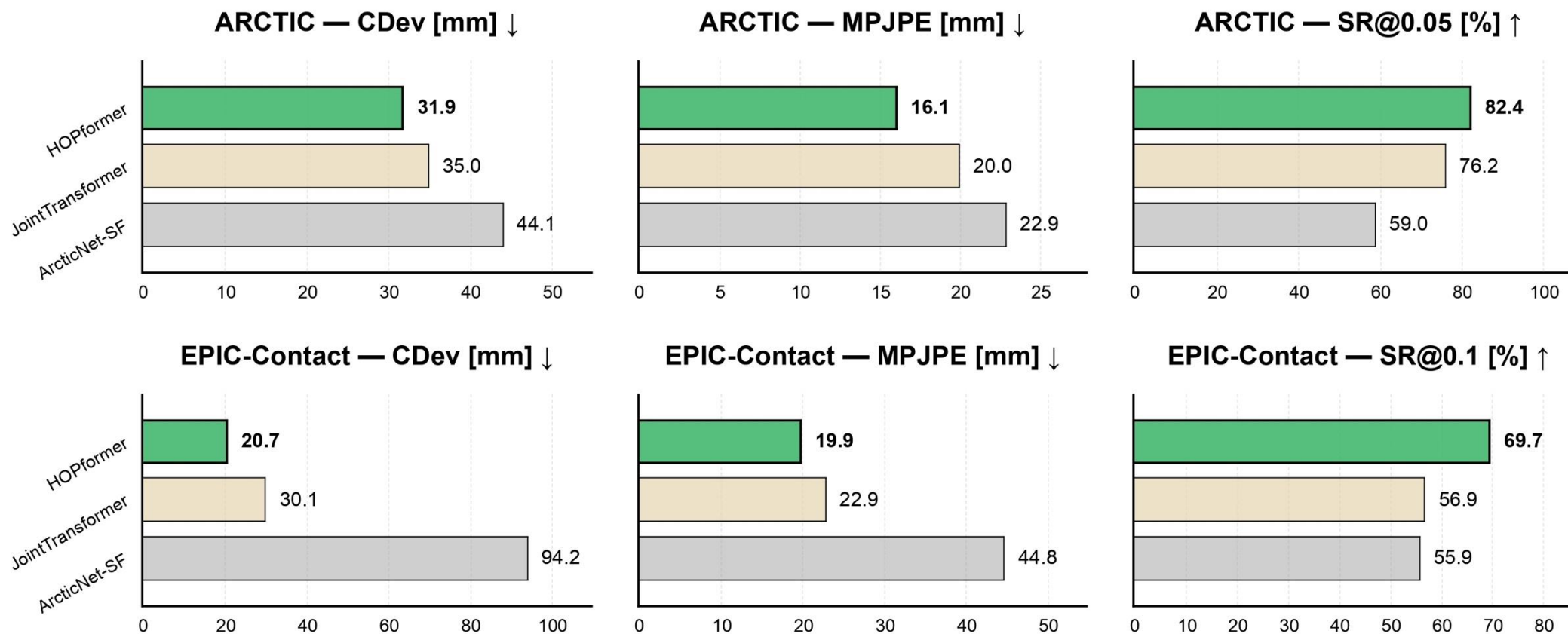
HOPformer
Model

Transformer-based
Conditioning on hand priors

HOPformer: Architecture



Quantitative Results



Qualitative Results on EPIC-Contact



Pan



Bowl



Bottle



Glass



Saucepan



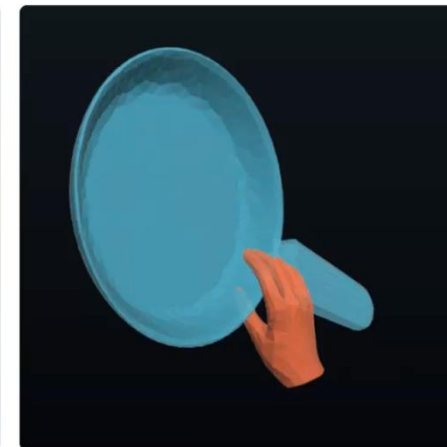
Plate

Ground truth

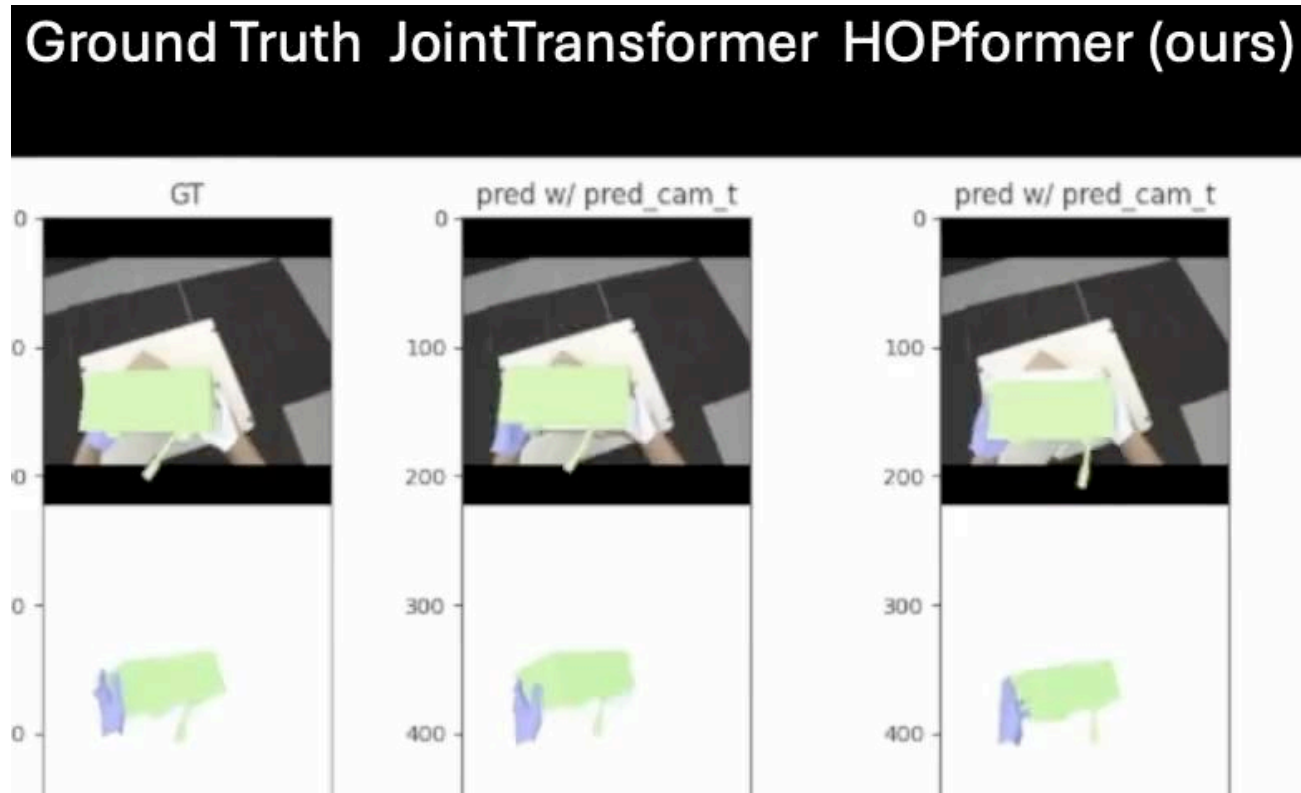
ArcticNet-SF

JointTransformer

HOPformer ours



Qualitative Results on ARCTIC



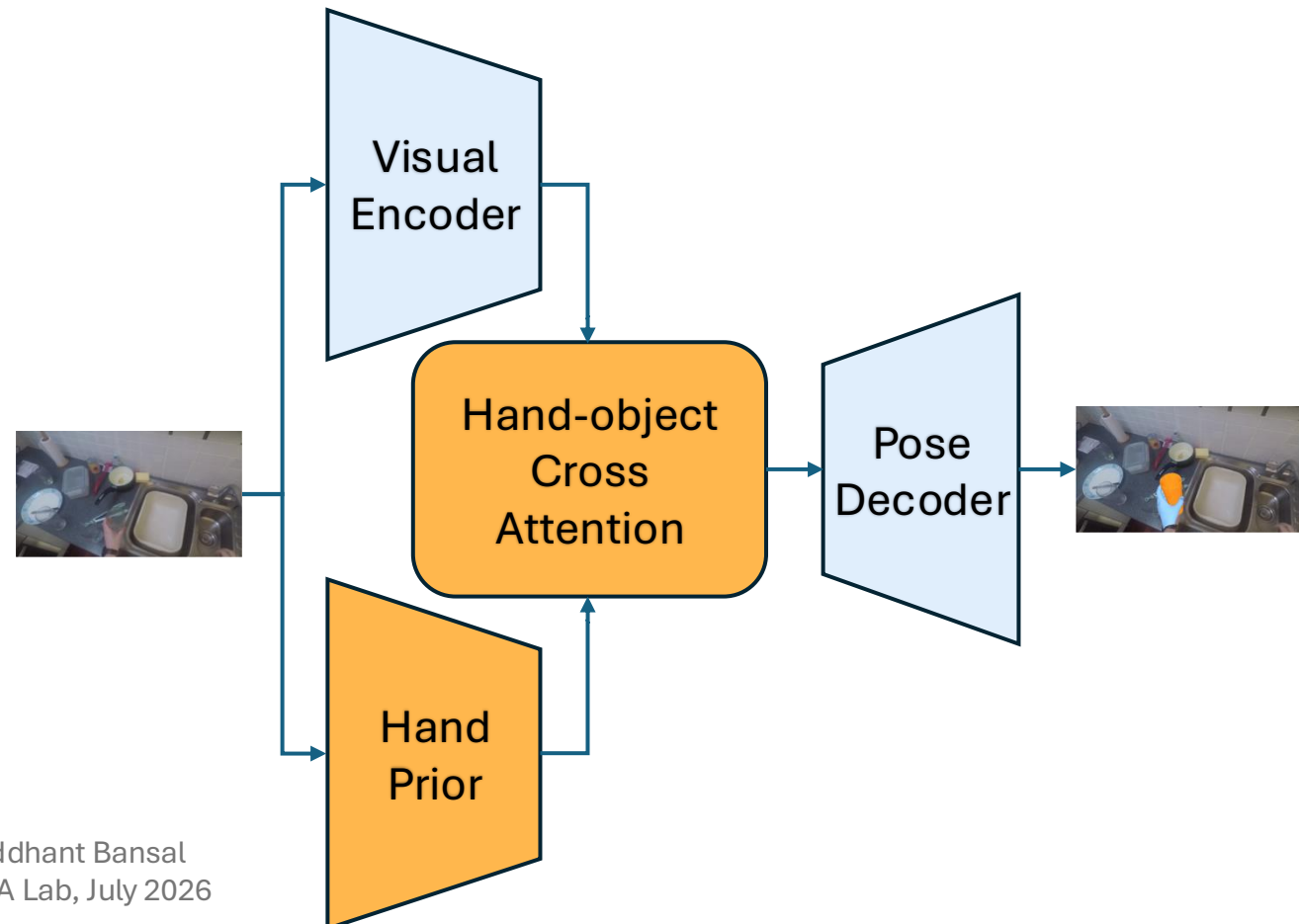
Conclusion: Two major contributions towards in-the-wild 3D Hand-object pose estimation

EPIC-Contact Dataset



2.3K clips
64.2K frames

HOPformer Network



Thanks for listening! Questions?



Code, datasets, and papers