National Czech Programme | eosc

ML for Computer Vision, Robotics and Protein Engineering

Use cases and datasets

Martin Cífka, Georgy Ponimatkin

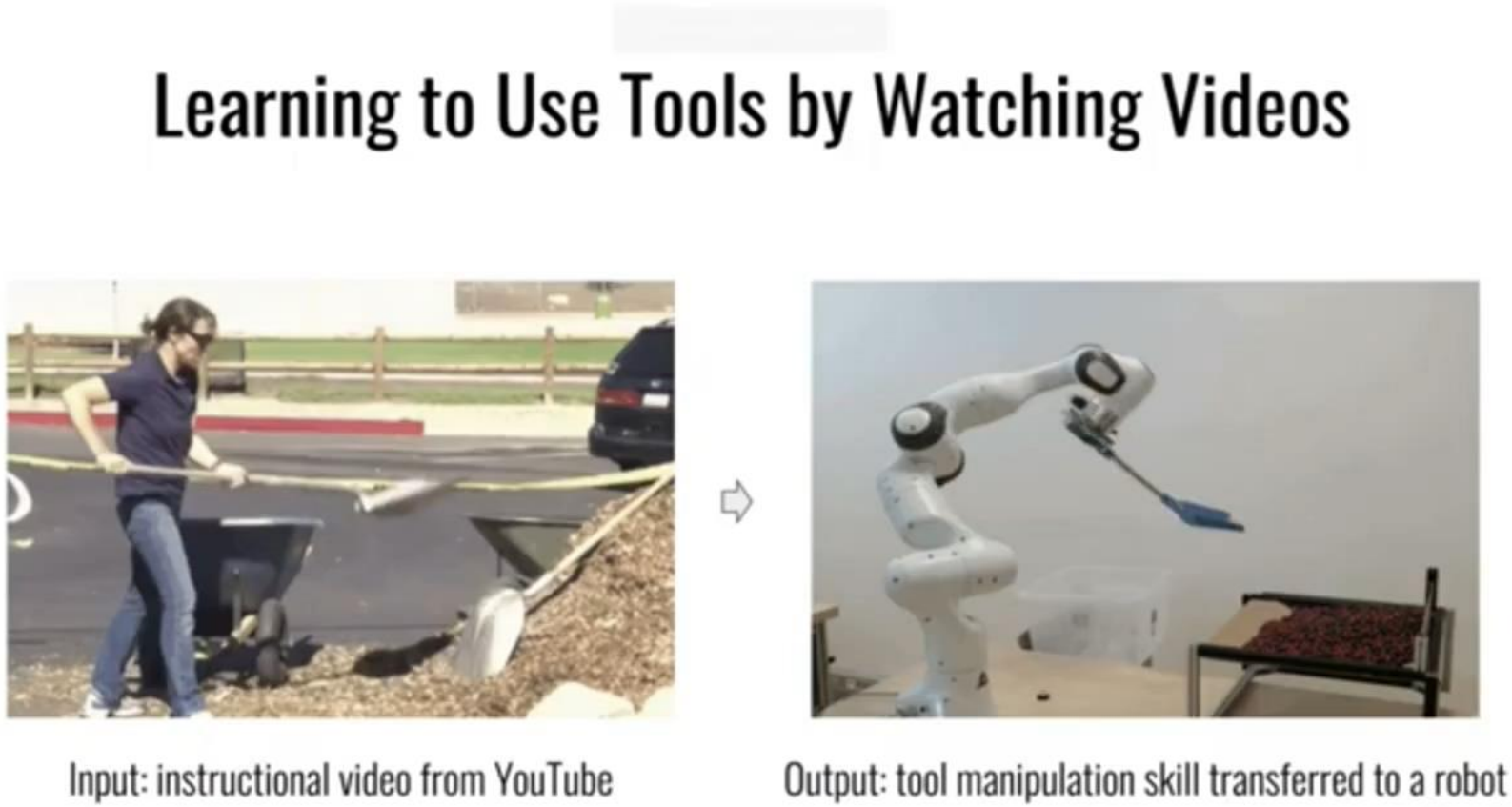
Intelligent Machine Perception, CIIRC CTU

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

MUNI ICS

Use Case: Training Robots from Internet Videos



K. Zorina, et al. "Learning to manipulate tools by aligning simulation to video demonstration." *IEEE RAL,* 2021.

Use Case: Training Robots from Internet Videos

G. Ponimatkin, M. Cífka, et al. "6D Object Pose Tracking in Internet Videos for Robotic Manipulation". In *submission*.

Dataset: Objaverse-XL

Figure 1: Objaverse-XL includes a ginormous collection of diverse 3D objects from a variety of sources. Here, we show examples of objects in Objaverse-XL rendered in a scene.

M. Deitke, et al. "Objaverse-XL: A Universe of 10M+ 3D Objects". arXiv:2307.05663

Use Case: Object Pose Estimation 'in-the-wild'

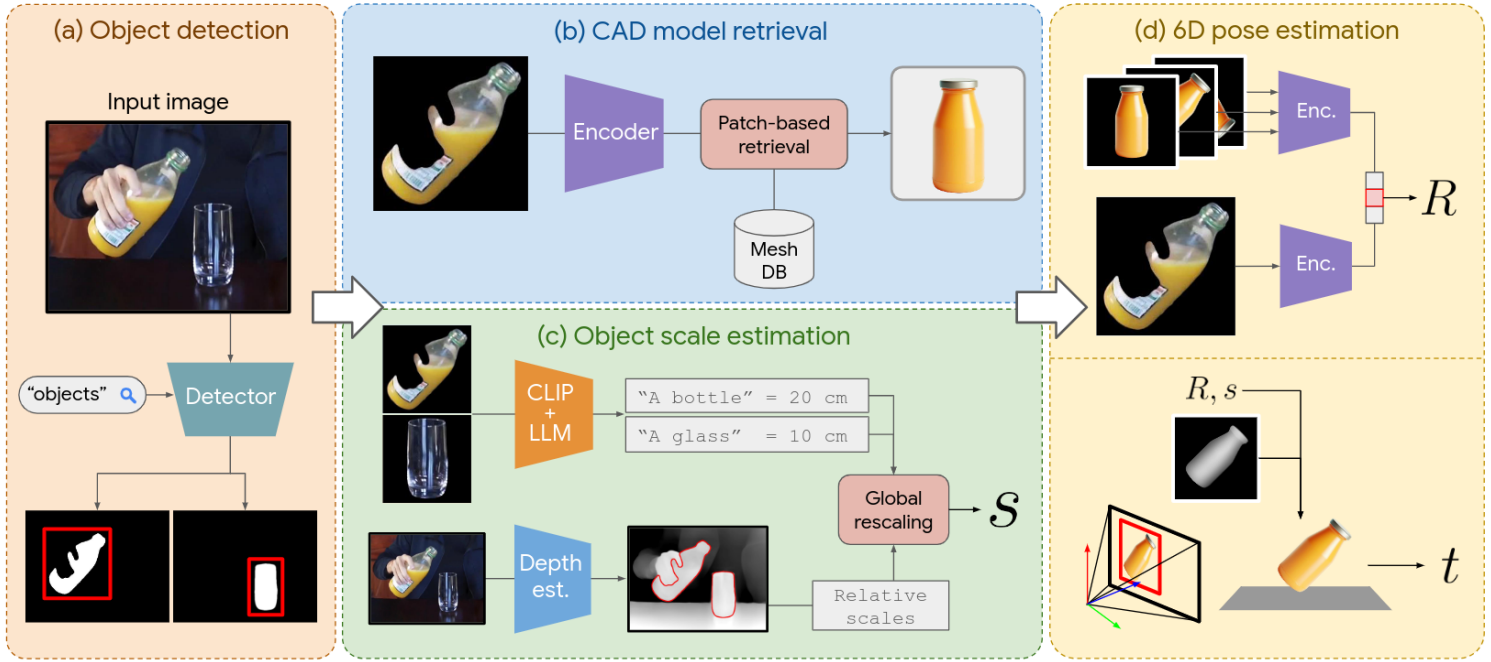


Figure 2: **Overview of 6D object pose estimation without a known 3D mesh (Sec. 3.1).** Given an input RGB image, our method: (a) detects and segments objects present in the image, (b) retrieves similar meshes from a large-scale object database via patch-based retrieval, (c) estimates the absolute scale of depicted objects in the scene via LLM-based re-scaling, and (d) estimates the camera-to-object rotation R and translation t via alignment of the retrieved (approximate) mesh.

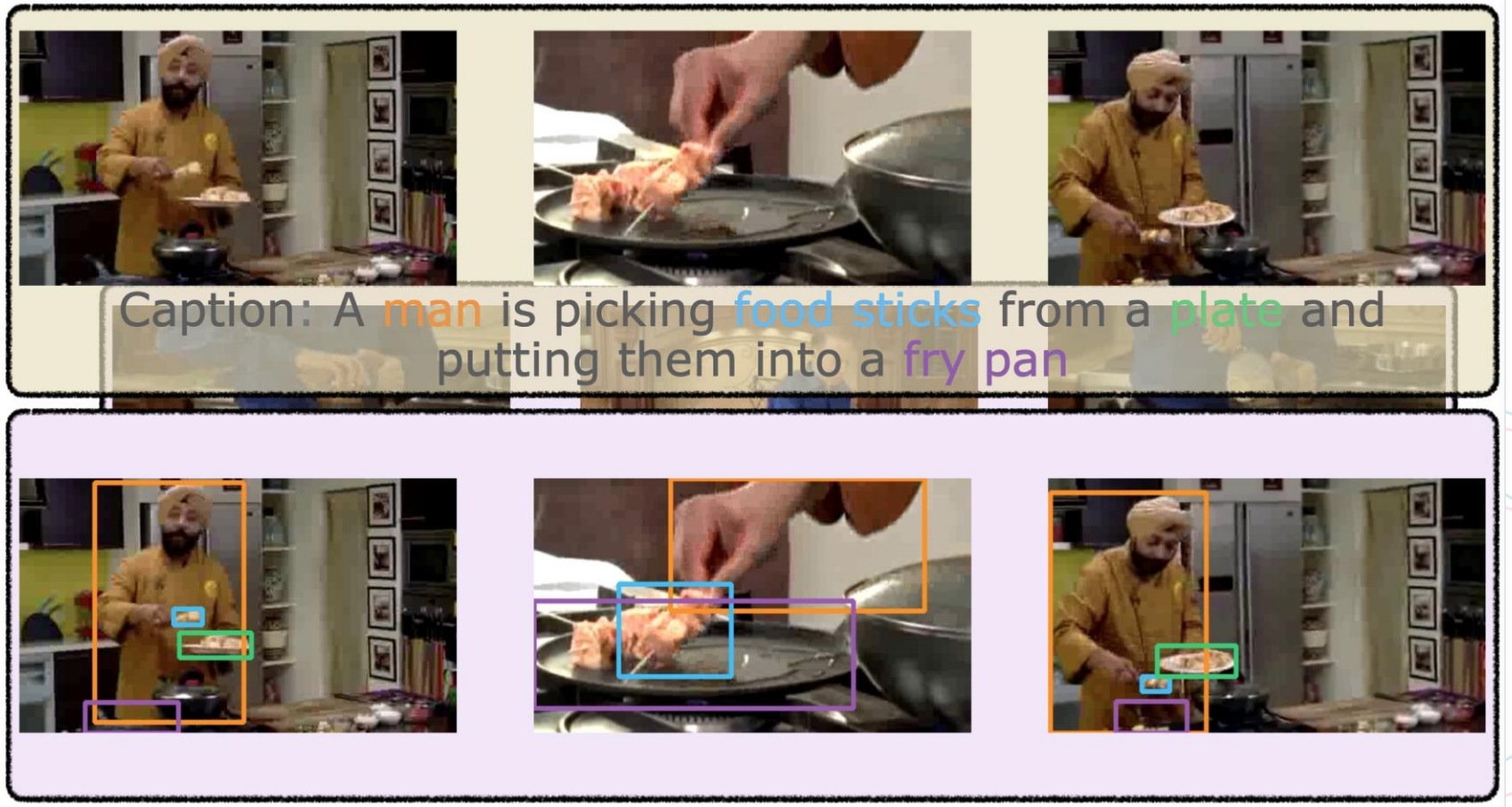
G. Ponimatkin, M. Cífka, et al. "6D Object Pose Tracking in Internet Videos for Robotic Manipulation". In *submission*.

Dataset: HowTo100M

Figure 2: Examples of clip-caption pairs retrieved with the help of our joint embedding. Pairs are selected based on the similarity between visual appearance and corresponding narration, while they are arranged based on linguistic similarity across pairs. Examples are taken from 4 distinct clusters, corresponding to *Knitting*, *Woodwork/Measuring*, *Cooking/Seasoning* and *Electric maintenance*.

A. Miech, et al. "HowTo100M: Learning a Text-Video Embedding by Watching Hundred Million Narrated Video Clips". ICCV 2019

Dataset: GROC (HowTo100M annotation)



E. Kazakos, C. Schmid, J. Sivic, "Grounded Video Caption Generation". arXiv:2411.07584.

Dataset: Open-X Embodiment

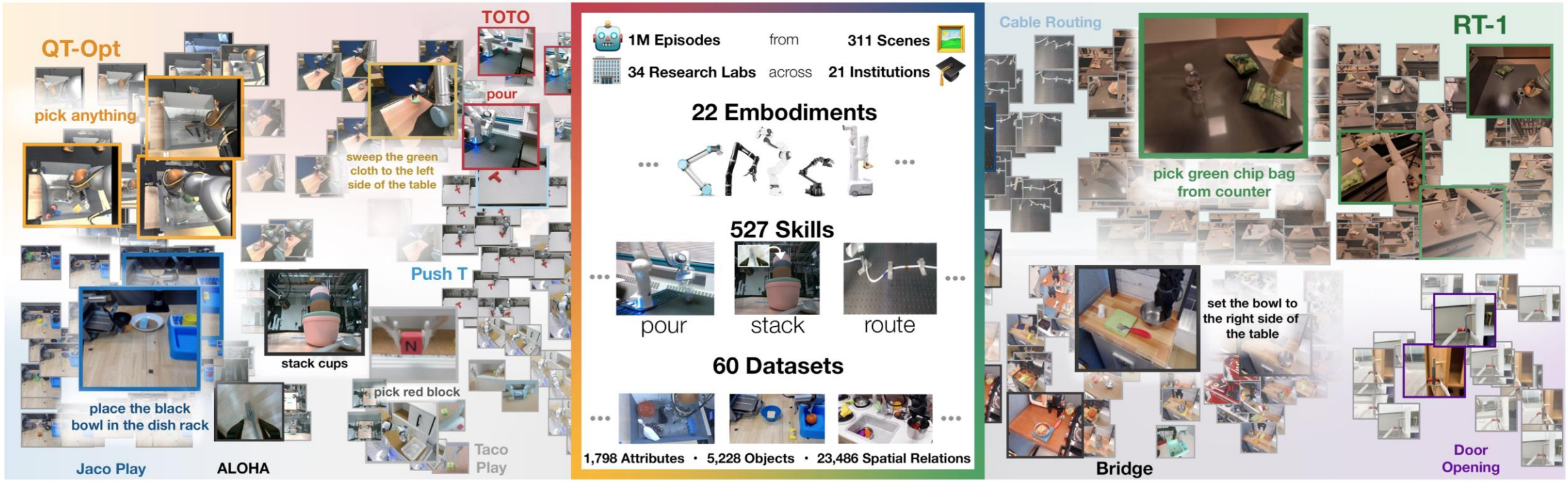


Fig. 1: We propose an open, large-scale dataset for robot learning curated from 21 institutions across the globe. The dataset represents diverse behaviors, robot embodiments and environments, and enables learning generalized robotic policies.

Open X-Embodiment Collaboration, "Open X-Embodiment: Robotic Learning Datasets and RT-X Models". ICRA 2024

What is Needed

- To achieve this we need infrastructure to search and stage data
  - We need to search subsets of the data, preprocess them
  - Combine datasets
  - Transfer to HPC facility (LUMI/Karolina) for training
  - Do the training (VLM finetuning uses 64 A100 for 14 days)
  - Save produced results and create demos

- Core issues:
  - Large number of small files (inodes are the limit)
  - Persistence of the data over long time
  - Data loading speed from the disk and subsequent pre-processing
  - Continuous updating of the data and annotations

Datasets used by IMP team at CIIRC CTU

- Computer Vision and Robotics
  - Objaverse (~10TB, 1M 3D assets, 1M .glbx files) + **renders** (~4TB, ~50 000 archives)
  - **HowTo100M** (~10TB, 1M YouTube videos, splittable into 100M clips) **+ GROC**
  - Open-X Embodiment (~10TB, 1M robotic videos, .pickles packed as webdataset, featuring video, trajectories etc)
  - NuScenes (~62 GB, ~205 000 images of autonomous driving)

- Machine Learning for Protein Engineering
  - mdCATH (~3.3TB, 135 000 MD trajectories)
  - ATLAS MD (~300GB, 4170 MD trajectories)
  - **PPIRef** (~100 GB, 300 000 protein-protein interactions in .pdb format)
  - **GeMS** (~80GB, 200M mass spectra)
  - PLINDER (~746GB, ~500 000 protein-ligand complexes)
  - ProteinGym (~65GB, ~2.5M mutations across 217 proteins)
  - **MassSpecGym** (~262MB, 231 000 MS/MS spectra, 29 000 unique molecules)

National Czech Programme | eosc

Děkujeme za pozornost!

Spolufinancováno Evropskou unií

MINISTERSTVO ŠKOLSTVÍ, MLÁDEŽE A TĚLOVÝCHOVY

MUNI ICS

cesnet

VŠB TECHNICKÁ UNIVERZITA OSTRAVA | IT4INNOVATIONS NÁRODNÍ SUPERPOČÍTAČOVÉ CENTRUM