CZECH INSTITUTE OF INFORMATICS ROBOTICS AND CYBERNETICS CTU IN PRAGUE

COMENIUS UNIVERSITY BRATISLAVA

TU Delft

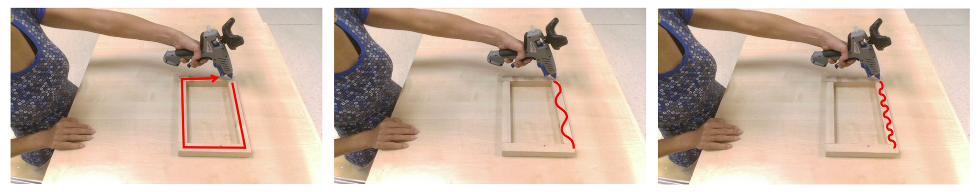
Imitrob: Imitation Learning Dataset for Training and Evaluating 6D Object Pose Estimators

Jiri Sedlar [1], Karla Stepanova [1], Radoslav Skoviera [1], Jan K. Behrens [1], Matus Tuna [2], Gabriela Sejnova [1], Josef Sivic [1], Robert Babuska [1,3]

[1] Czech Technical University in Prague, Czechia, [2] Comenius University, Bratislava, Slovakia, [3] Delft University of Technology, The Netherlands

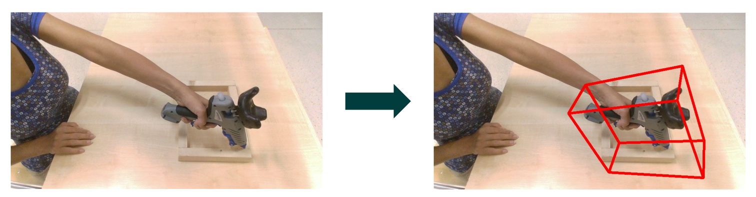
Motivation

TOOL MANIPULATION TASKS



EXAMPLES OF DIFFERENT TASKS WITH ONE TOOL

6D OBJECT POSE ESTIMATION



INPUT IMAGE — 6D OBJECT POSE

Imitation learning in industrial applications

- Robot learns how to perform a task from demonstrations
- Requires 6D pose (position & orientation) of the tool

Challenges for 6D object pose estimation

- Manipulation and occlusion of the tool by the hand
- Lack of benchmarking datasets

Contributions

New 6D object pose benchmarking dataset

- Evaluation of a given 6D object pose estimation method
  - Suitability for imitation learning of a given robotic task
  - Generalization capabilities w.r.t. various setups

Methodology and software for dataset extension

- Acquisition of data for new tools and manipulation tasks
- No need for manual annotation or 3D model of the tool

Dataset and code

- imitrob.ciirc.cvut.cz/imitrobdataset

AGIMUS INNOVATIVE ROBOTICS FOR AGILE PRODUCTION

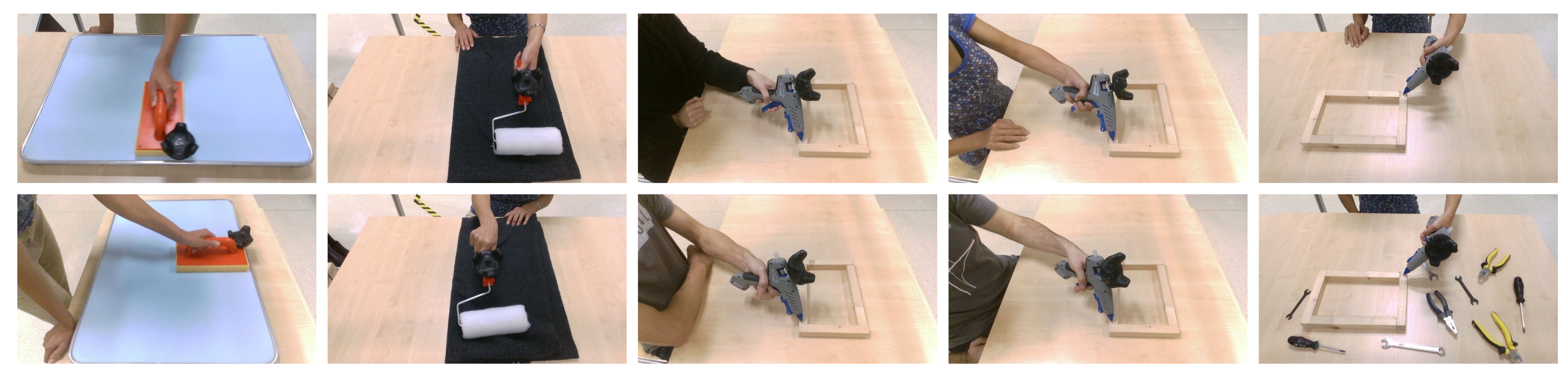
2023 IROS OCTOBER 1 - 5, 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems

Imitrob dataset

TOOLS

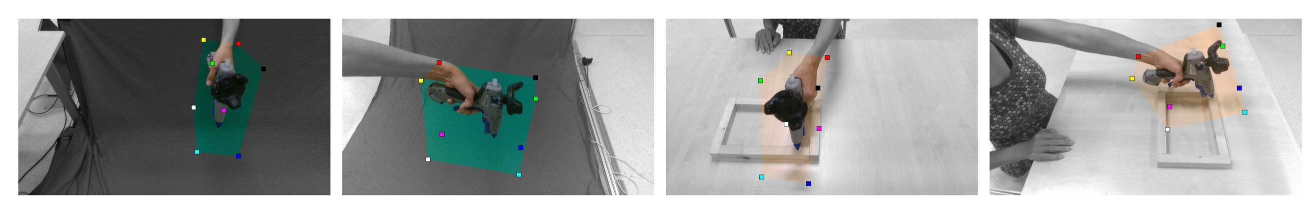
GLUE GUN — GROUT FLOAT — ROLLER — GLUE GUN 2 — GLUE GUN 3 — GLUE GUN 4 — HEAT GUN — POWER DRILL — SOLDERING IRON

SETUP VARIABILITY



FRONT / SIDE **CAMERA** — LEFT / RIGHT **HAND** — FOUR **DEMONSTRATORS** — WITHOUT / WITH **CLUTTER**

GROUND TRUTH 6D POSE



TRAINING SET — TEST SET

Test dataset

- 208 videos of 12 manipulation tasks with 9 tools
- Industrial use-cases in manufacturing-like environment
- 2 cameras, left/right hand, 4 demonstrators, (no) clutter

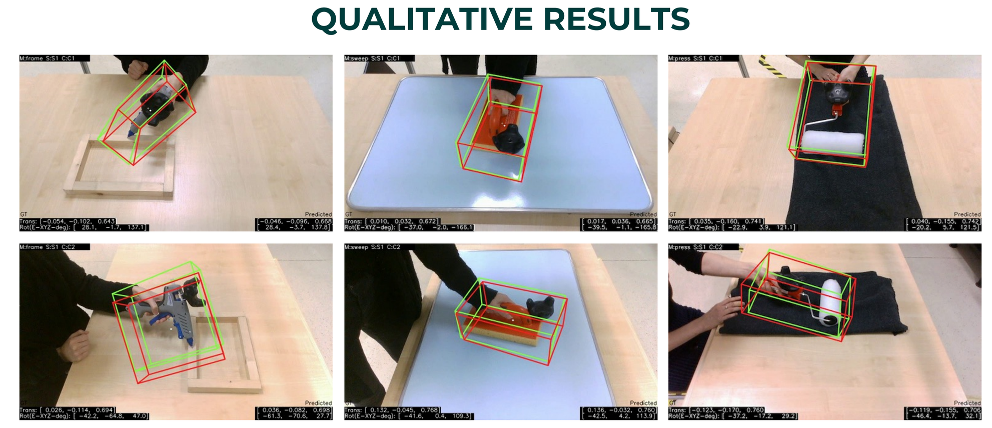
Training dataset

- 144 videos of random manipulation with 9 tools
- 2 cameras, left/right hand, 4 demonstrators

Ground truth 6D object pose

- Measured by HTC Vive tracker attached to the tool

QUALITATIVE RESULTS



PREDICTED 6D POSE (DOPE method)
GROUND TRUTH (HTC Vive)

This work was supported by the European Regional Development Fund under the projects IMPACT (CZ.02.1.01/0.0/0.0/15_003/0000468) and Robotics for Industry 4.0 (CZ.02.1.01/0.0/0.0/15_003/0000470), EU Horizon Europe Programme under the project AGIMUS (101070165), VEGA (1/0796/18), MPO TRIO (FV40319), CTU Student Grant Agency (SGS21/184/OHK3/3T/37), and the Czech Science Foundation (GA21-31000S).

Experimental results

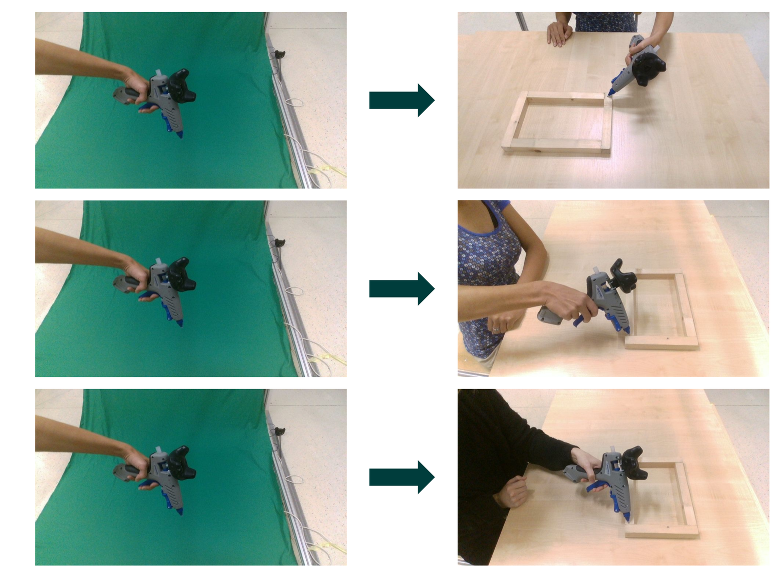
Tool	glue gun	grout float	roller	glue gun 2	glue gun 3	glue gun 4	heat gun	power drill	soldering iron	average
ADD_{5cm}	60.4	78.6	50.5	9.0	4.7	23.4	13.2	59.8	12.8	**34.7%**
E_{rot}	7.3	4.1	8.7	38.5	40.3	20.9	14.3	8.0	35.6	**19.8°**
E_{tra}	4.0	2.5	3.7	9.9	10.2	8.4	7.0	3.8	9.0	**6.5 cm**

GENERALIZATION EXPERIMENTS

DIFFERENT **CAMERA VIEWPOINT**

OPPOSITE **HAND**

DIFFERENT **DEMONSTRATOR**



TRAINING SET — TEST SET

Experiments

- Evaluation for each task and tool
- Generalization between training and test data setups

Evaluation metrics

- ADD (average bounding box distance) 5 cm pass rate [%]
- Rotation [deg] and translation [cm] errors

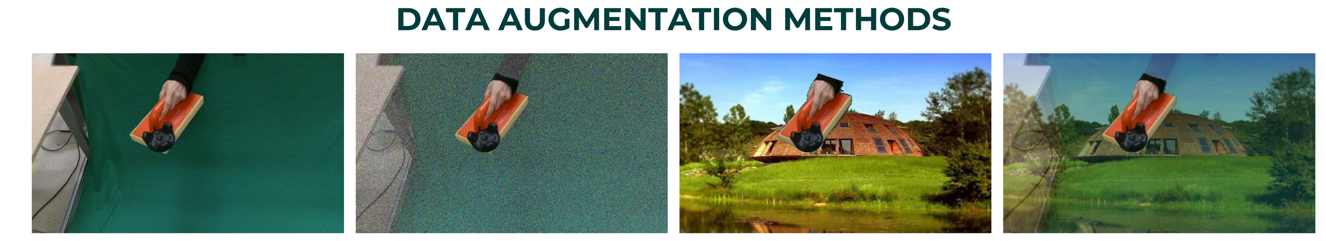
Example usage

- 6D object pose estimation method DOPE [1]
- Depth channel or 3D object model not required

Data augmentation

- Blending of original background with random images (first application in 6D object pose estimation domain)

DATA AUGMENTATION METHODS



NoAug (no augmentation) — **BgNoise** — **BgRandom / BgAlternate** — **BgBlend** (best results)

	NoAug	BgNoise	BgRandom	BgAlternate	BgBlend
ADD_{5cm}	29.2%	29.6%	45.6%	56.5%	**60.1%**
improvement	—	1.0×	1.6×	1.9×	**2.1×**

[1] Tremblay et al, Deep object pose estimation for semantic robotic grasping of household objects, CoRL 2018.