CLAMP

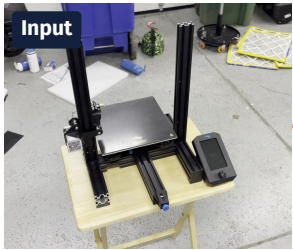
A Sim-to-Real Benchmark for Closed-Loop Kinematic Pose Estimation and Assembly Reasoning

Kevin Murray[1] Randolph Beauregard Robert III[2] Petar Z Duric[1] Zoran Duric[3]

[1]Overlab LLC [2]AVA Labs [3]George Mason University

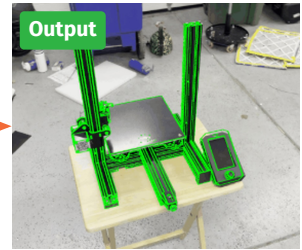
NeurIPS 2026 • Evaluations & Datasets Track

The task: one image in, the machine's state out



One RGB image of a real machine

Model
trained only on synthetic images



Its state (ground truth in green)

Pose
Rotation and translation in the camera frame.

Kinematics
Every joint angle and slide position.

Assembly
Which parts are present; here, no gantry or top bar.

Uses: AR-guided maintenance, assembly verification, and robots that work alongside equipment.

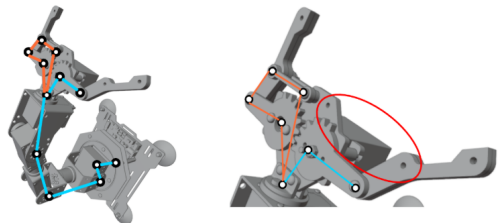
Real machines break today's benchmarks

Existing benchmarks: rigid objects or **serial revolute robot arms**, always complete. Real equipment adds:

Closed kinematic loops: four-bar linkages, parallel arms. URDF's tree *cannot represent them at all.*

Prismatic joints: linear axes of printers, routers, and saws. DREAM has revolute arms only.

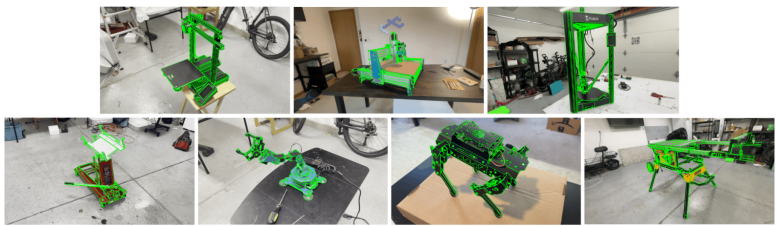
Missing parts: this changes the geometry *and* can change the degrees of freedom.



Robot-arm gripper with its kinematic graph: open-chain edges in cyan, a closed four-bar loop in orange. Removing the constraining link (red) breaks the loop and frees the finger to rotate.

CLAMP in numbers

7	210	21k	2M+	27
machines	real scenes	real test images	synthetic images	max. DOFs



Ground-truth labels (green) on real test images of all seven machines.

Two test splits: **kinematics only**, and **kinematics + assembly** with parts removed.

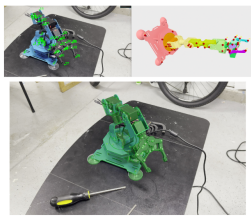
Synthetic training data, real test labels

Training: 2M+ synthetic images of valid poses



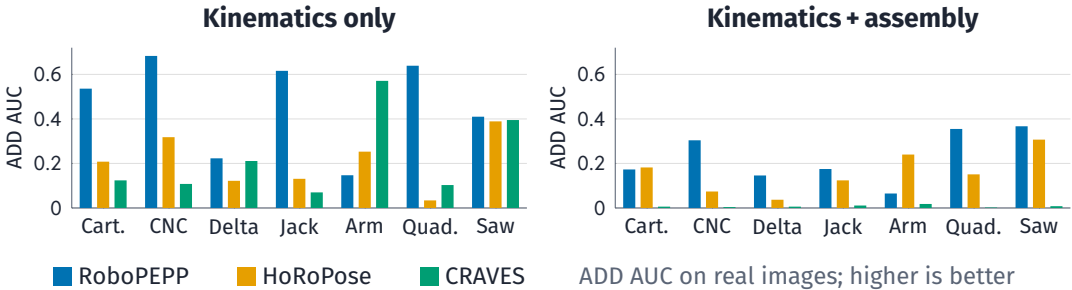
- Random joint values are **projected onto the constraint manifold**, so every pose obeys the loop closures.
- Removing parts drops their constraints: **closed loops become open chains**.

Testing: 21k real images, labeled by multi-view fitting



- 4K orbit video → COLMAP cameras → a few clicked 2D–3D points.
- Pose and joint state are **fit jointly under the same constraints**: **8.1 px** median error on 4K images.

No baseline performs consistently well



.465 vs. ∼.72
Best average here vs. RoboPEPP on DREAM.

.465 → .226
RoboPEPP with parts missing; CRAVES: .008.

0.1 vs. 207 mm
RoboPEPP's mean loop violation: CNC router vs. delta printer.

Takeaways

- First sim-to-real benchmark with **closed loops**, **prismatic joints**, and **kinematic plus assembly perturbations**: ~21k labeled real images of 7 machines.
- Released: kinematic graph, constrained data generator, multi-view labeling, and baselines.
- No adapted baseline performs consistently well, especially with **missing parts** and **parallel mechanisms**.
- Next: **constraint enforcement at inference**, domain-specific detectors, other method families.

Dataset

Code & paper

huggingface.co/datasets/clamp-benchmark/clamp-benchmark
github.com/overlab-kevin/clamp

Dataset license: CC BY 4.0